Listenership in Human-Agent Collectives: A Study of Unidirectional Instruction-Giving

By

Abdulmalik Caxton Yusuf Ofemile. MA, B.Ed. (AFHEA)

School of English

University of Nottingham

A thesis submitted for the degree of Doctor of Philosophy in English to the University of Nottingham

January 2018

Table of Contents

Abstract

Research in nonverbal listenership behaviour and instruction-giving has focused on interaction with people while paying inadequate attention to human-agent interaction even as recent research indicates that, increasing pervasive computing is significantly changing how humans interact with intelligent software agents and extending the boundaries of discourse to contexts including satellite navigation systems giving directions to drivers, self-checkout machines in supermarkets and intelligent personal assistants on smartphones. This thesis reports studies that use spontaneous listener facial actions and gestures to understand the nature and pattern of spontaneous nonverbal listenership behaviours, identification and communication in instruction-giving contexts.The research methodology used is as follows. Participants who are all L1 speakers of English (forty-eight in Study 1, six in Study 2) were tasked with assembling two Lego models using vague verbal instructions from a computer interface in Study 1 and a human instructor in Study 2 with a 15-minute time limit per iteration.The interface in study1uses three voices of which two are synthesised and one is non-synthesised human recording by a voice actor while Study 2 used a live human voice. A 24-hour long multimodal corpus was built and analysed from interactions between participants and the interface in Study 1 while a 3-hour multimodal corpus was developed from Study 2. The multimodal corpuswas annotated for marked facial actions and gestures occurring at points when participants requested that instructors repeat instructions.Participant requests were nonverbal in HAI and a combination of nonverbal and verbal instructions in HHI contexts. The repetitions were quantified and classified into nine typologies.The results reveal key findings regarding the use of spontaneous nonverbal listenership behaviours as pragmatic markers indicating listener comprehension or incomprehension of instructions, perception of instructor-identities, projection of attitudes, meaning-development, task-execution strategies and interaction management even though, the agent could not attend to them in the same way a human can. Using these results, the thesis submits that there are potentials for applied linguistics theories and research to be used to identify and understand pathways to make agents more responsive to human behaviour, make human-agent interaction more credible and provides a theoretical foundation for future multidisciplinary research.

Acknowledgements

I am immensely grateful to those that helped me while working on this thesis. At the very fore front are my PhD supervisors, Svenja Adolphs and Tom Rodden for their invaluable ongoing academic, practical and moral support.

I would also like to thank the Tertiary Education Trust Fund (TETFUND) and the management of FCT College of Education Zuba, Abuja for funding and making it possible for me to do the PhD.

My fellow PhD students for providing inspiration and feedback. Special gratitude goes to Leigh Clark for being a friend and very useful research partner whose consistent insights were critical to bringing this thesis this far.

My gratitude goes to the academic and administrative staff of the school of English University of Nottingham for their consistent support over the years. I am particularly indebted to Jo Pullen, Rebecca Peck, Tracy-Ann Stead, Denine Carmichael and Stephen McKibbin for being very accommodating throughout my studies

Finally, I must thank all my family for the support. I am particularly grateful to my wife Larai and children for their companionship, support and reassuring care.

Table of Figures

List of Tables

List of Abbreviations

3-D Space	Three-Dimensional Space
CASA	Computers As Social Actors paradigm
CHILDES	Child Language Data Exchange System
CL	Cepstral Lawrence: one of the synthesised agent voices
CLAN	Software used for coding and analysis of video compatible with the CHILDES corpus/chat format
CP	CereProc Giles: one of the synthesised agent voices
ELAN:	Software used for creation of complex annotations on video and audio resources
HAI	Human-Agent Interaction
HACs	Human-Agent Collectives
HCI	Human-Computer Interaction
HHI	Human-Human Interaction
HR	Human Recorded voice: agent created using a voice actor
HRI	Human-Robot Interaction
HV	Human Voice: Human instructor
M-E	Media Equation
Corpus	Multimodal corpus

Chapter 1 Introduction and Background Information

1.1 Background to the Study

As computation increasingly pervades the world around us, it is significantly changing the way we communicate with highly inter-connected, autonomous and intelligent components (aka agents) (Jennings et al., 2014). Specifically, agents in this thesis are simulated (see Chapter 4.2.1) and described as those run with simulated bodies and environments where a program takes in commands and returns appropriate percepts or inputs (Poole & Mackworth, 2010).

Interaction with agents is extending discourse to contexts where we interact with smartphones, intelligent personal assistants (Siri, Cortana); Computer Assisted Language Learning (CALL) devices-DuoLingo; embodied conversational agents (ECAs) in academic research or healthcare-Gretaand self-checkout systems in supermarkets.As discourse contexts expand, new discourse communitiesemerge. Discourse communities are "groups that have goals or purposes, and use communication to achieve these goals" (Borg, 2003, p.1). Although, the concept originally refers to Human-Human Interaction, (HHI), it is extended in this thesis to describe groups of humans interacting with agents in Human-Agent Interaction (HAI) contexts.

The interaction contexts engender partnerships that allow diverse actors to build dynamic relationships similar to those which exist between speakers and listeners conversing in HHI with some level of restrictions. Furthermore, these relationships in new discourse communities tie in with Clark's (1996) concept that language use is a form of joint action where participants have responsibilities and expectations. For example, the linguistic systems or conventions of English language provide speakers with a signalling system and describe what speakers normally do and expect others to do (Carston, 1999) to communicate effectively.

The persons of interest forthis study in the joint action are listeners and their communicative behaviours during interaction otherwise called listenership. Listenership is "…the active, responsive role that listeners have in conversation" (O'Keeffe et al., 2007, p.142). This implies that listeners participate actively in any discourse even when they are not talking, and it is a behaviour expected of them by speakers (see Chapter 2.2.3 for details). The listeners participating in this study are all speakers of English as a first language from any English language speaking context that were purposively selected (Chapter 3.2.2a)

One aspect of listenership is nonverbal feedback which includes spontaneous facial actions and gestures (see Chapter 2.2.3). While facial actions refer to facial expressions (Ekman, 1997) gesture covers a multiplicity of communicative movements, primarily but not always of the hands and arms (McNeil, 2005). They are commonplace in HHI; this study is interested in how these operate when people interact with agents.

As interaction with agents deepen, it is envisaged that people may increasingly share interaction space with agents because already self-checkout systems in supermarkets are sharing work space with cashiers in checkout points in supermarkets. Although presently they are independent of each other, it is envisaged that in the future there will be more collaboration between them.

To conceptualise and understand this scenario, Jennings et al (2014) proposed the concept of Human-Agent Collectives (henceforth HACs) to describe a “new class of socio-technical systems in which humans and smart software (agents) engage in flexible relationships in order to achieve both their individual and collective goals” (Jennings et al., 2014, p.80).

HACs thus will be a discourse community that is socially heterogeneous(Picard, 2015)among whom control and information will be widely dispersed between actors. Participants have flexible autonomy, in other words, neither the humans nor the agents are always in charge as roles are continuously switched because of controlled autonomy.

However, it is pertinent to note that this study uses the setting of instruction-giving in assembly tasks to engender interaction. The joint action of instruction-giving and taking between humans and agents in HACS may require reciprocal actionsbecause of the expectations inherent in interactions.

Reciprocal actions tie in with some aspects of Hymes’ (1972a) speech events because the norms of English language – as given by linguistic conventions – require that instructions be couched in a way that they will make the listener do something as distinct from advice or a request. In addition, the assembly task may not exist without the use of speech; hence, instruction-giving is also a speech event in the sense expressed by Hymes (1992, 1972b).

Looking at instructions again, they are speech events that tell the listener or instructee what to do and how to do it during the assembly task. At the same time, these instructions are not only a means of expression for the instructor, but they also define the rules of engagement between speakers and listeners.

However, while Hymes' speech events are primarily focused on the use of speech, reciprocal actions go beyond speech to nonverbal communication.These occur in what sociolinguists describe as speech situations or socially contextual situations where speech takes place (Hymes, 1972a), such as assembly tasks, a dinner party or even a family meal.Mitchell, (1957)applied the concept of speech situation to a natural setting focusing on the routines and rituals involved in buying and selling in markets (Prodromou, 2005) and how these affect language use during interaction.

Moreover, Hymes'(1972b) speech event comprises settings, participants, purpose, key, or (culture) message form, message content, tone and channel. Participants in this study are speakers of English as a first language from different parts of the world agent instructors and one human instructor who is also an L1 speaker of English. There are two contexts of interaction projected in this study. The first is HAI which represents some form of intercultural interaction. The second is HHI although all interactants are humans and speakers of English as a first language, but the fact that the instructor is from Southern England while some participants acquired English from contexts outside England suggests that to an extent, interaction between them is intercultural.

Culture is used in this thesis as operationally defined by the University of Minnesota Centre for Advanced Research in Language Acquisition (CARLA) for the purposes of intercultural studies as:

> *... the shared patterns of behaviors and interactions, cognitive constructs, and affective understanding that are learned through a process of socialization. These shared patterns identify the members of a culture group while*

> *also distinguishing those of another group* (CARLA, University of Minnesota, 2013)

This definition implies that culture is learned as part of an individual's development processwithin a society and it shows people as similar to or different from others during interaction. Thus, culture provides a basis for interlocutors' viewpoints, understanding and projection of identities during interaction as will be outlined in 2.6.

Furthermore, the notion of intercultural interaction context as expressed in this thesis istied toBhabha's (1994) seminal work on multicultural interaction which postulates that where two cultures meet, a new one is born which provides a basis for articulation of differences between interlocutors (See 2.7). In the same vein, this thesis holds that when Humans interact with agents,new contexts and forms of interaction are born because the interactants are culturally different as outlined in chapter 2.7Thus, context and culture set the norms of joint action between instructors and listeners duringinteraction as well as the norms of interpretation that are further developed within the participants' interaction cultures.

In view of this, this thesis holds that, it is important to understand how nonverbal language can be used in a form of joint action between non-responsive agents as instructors and humans as listenersduring interaction.

1.2 Statement of the Problem

Having described the background, this discussion will turn to the gaps in research in related areas. The communication that occurs within the space of HAI as envisaged by HACs is yet to be fully understood by researchers, even though linguistics has an established tradition of focusing on human interaction because language is a tool for understanding human social life. For example, linguists have used corpora of spoken English to explain what Firth referred to as the context of situation (Firth, 1957). Others have focused on how people from different cultures create better superdiverse communities in multicultural cities (Blackledge et al., 2016), actualise participatory design of systems as culturally diverse teams (Muller, 2009) or use agents as tools of communication in superdiverse human contexts (Cress et al., 2016; Baynham et al., 2015; Garcia & Wei, 2014).

Research indicates that people project their identity for others to perceive through performance during interaction (Gumperz, 1999; Knap &Hall, 1997;Goffman, 1959;Garfinkel, 1956). Considering the impact of identity on interaction, there is the need to explore how identity is projected and perceived in an instruction-giving context where the instructor is an agent as well asto examine listener use of gestures during interaction.

Furthermore, when it comes to assessing listener feedback during interaction with computers or agents the use of spontaneous listener facial actions and gestures is under-emphasised,for example, Nassand Lee's,(2000) investigation on identity used questionnaires andMonroy-Hernández et al's(2011)focusing on disourse makers used participant narratives after the task. Although, these are standard measures of data gathering, they may be inadequate for effectively assessing the role of nonverbal listenership as a feedback mechanism during interaction in HACs.

Moreover, research indicates that agent insensitivity to human nonverbal behaviour is a fundamental limitation (Esposito et al., 2015) during interaction hence the need to understand the impact of listener spontaneousfacial actions and gestures when interacting with agents. Research also suggests that instructions form a critical part of the communication process and meaning development (Carter, 2004) in human interaction,but our understanding of how instruction-based interfaces used in HCI enable meaning-development is limited.

Thus, not much is known about how humans react spontaneously with their facial actions and gestures when receiving instructions from agents during assembly tasks.It becomes necessary for this study to understand the relationship between instructions and marked spontaneous facial actions and gestures emerging from this interaction. This focus is taken because with time "HACs need to understand and respond to the behavior of people…"(Jennings et al, 2014,pp. 87-88) during interaction.

This study focuses on multimodal aspects of listenership in HACs and discusses this in the context of discourse analysis approaches within the field of Applied Linguistics. This is because in the real-world people communicate multimodally withall their senses and at the same time. For example, when we exchange greetings with people on the street, we use all our language skills as the speech is uttered, we perceive it, we see the person's reaction, we smell the street, and maintain an acceptable posture or distance during that very brief interaction. We do not perceive the words as individual phonemes, nor do we see the person's facial expressions separately from their gestures rather, we make meaning from combining these and may be more from our experience.

However, when humans interact with agents, humans perceive the interaction multimodally but the agent's ability to interact multimodally may be restricted by its capabilities. For the researcher, the aim is to provide a clearer understanding of the interaction thus, there is the need to understand each strand of multimodal communication used in HAI, reconcile these various fragments of interaction (Adolphs& Carter, 2013) and join them into an over-arching whole.

1.3 Objectives and significance of the study

This study is significant becauseinsights from this study may help researchers and end users to identify the appropriate communicative support required by interlocutors when people interact with intelligent agents in dynamic contexts like disaster zones, classrooms, traffic control, supermarkets, government offices and private vehicles.

Furthermore, as more people and organisations increasingly interact with agents,the research methodology and findingsfrom this study may help dicourse analysts replicate the study in other contexts. They may use the findings in this study as a comparison for theirs with a focus on devising strategies to better predict, generate and understand listener nonverbal responses in emerging discourse contexts.

For stakeholders like governments, designers and end users of intelligent agents, in contexts where humans who control or monitor agent activities and collaborate with them are physically separated from these agents, this study may provide linguistic data that could be used to derive rules and guidelines to govern the behaviour of such agents in the future.

In addition, findings from this study may improveour understanding of nonverbal listenership behaviour in HACs and provide solutions that are adaptable to other contexts such as language learning and healthcare.

1.4 Research Questions

This study uses spontaneous listener facial actions and gestures to focus on multimodal aspects of listenership behaviourdisplayed by participants when taking instructions from agents (representing HAI context) and humans (representing HHI context) during assembly tasks.Specifically, this research sets out to answer the following questions:

1. What spontaneous facial actions and gestures do listeners display in HAI when requesting repeats of agent instructions during assembly tasks and why?
2. What spontaneous facial actions and gestures do listeners display in HHI when requesting repeats of instructions from humans during assembly tasks and why?
3. How do spontaneous facial actions and gestures displayed in HAI compare with those in HHI?

1.5 Hypotheses

In view of the research objectives and problems identified, the study proposes that there is a relationship between listeners' nonverbal listenership behaviour and their execution of agent instruction. Specifically, the hypotheses are as follows:

H_1 There is a relationship between the listener's facial actions and whether they successfully execute the instructor's verbal instructions.

H_2 There is a relationship between the listener's gestures and whether they successfully execute the instructor's verbal instructions.

1.6 Thesis overview

This study focuses on the interaction between humans and agents in a unidirectional instruction-giving context. It also focuses on the role, impact and relationship between instructions, marked listener facial actions and gestures displayed while taking instructions to assemble Lego kits. The thesis is structured as follows:

Chapter 1: Introduction. This Chapter sets out the research focusing on the interaction context of HACs, research problem, questions, and hypotheses. It also describes the structure of the study.

Chapter 2. Literature Review. This Chapter reviews previous researchintolistening process and comprehension, listenership as facial actions and gestures, instructions as information units, emergent identities, and turn-taking and repairs emerging from interaction.

Chapter 3: Multimodal Research Framework. This Chapter focuses on the research methodology design and the annotation scheme used analysing data in the study.

Chapter 4:Study 1: Listenership behaviours in Human-Agent Interaction (HAI). This chapter reports the first study. This study aimsto assessnonverbal listenership bylisteners receiving assembly instructions from agents. The Chapter will describe the specific background to the study, review related literature on simulated agents and the essence of listener expectations during interaction. The specific study design, resultsand their implications are also outlined here.

Chapter 5: Study 2: Listenership behaviours inHuman-HumanInteraction (HHI). This Chapter reports on the second study. This will focus on the specific background to the study such as lessons from the first study, objectives and limitations. The study design describes the specific study procedures and processes. The study results are presented and discussed while drawing conclusions and highlighting implications.

Chapter 6: This Chapter compares the interaction patterns of listenership behaviours observed in Study 1andStudy 2 in view of the study hypotheses, relates these to how they answer the research questions and briefly summarises the study findings. It also outlines the implications of the study for existing theories of listenership, nonverbal communication, HACs, agent design, and discourse analysis.

Chapter 7: This Chapterconcludes the study. It outlines the contributions, applications, study limitations, areas of further research and componential changes required as a way of reviewing interaction in HACs.

Chapter 2 Literature Review

2.1 Introduction

This thesis investigates the pattern of nonverbal listenership behaviours that listeners display when taking instructions during assembly tasks. This Chapter is structured around the idea expressed in 1.1 that language use is a joint action between speakers and listeners used to build communication during interaction and it comprises speaker actions, listener actions and issues that influence how communication occurs during interaction. Speaker action includes constructing instruction as information units, instruction-giving and responding to listener feedback. The second aspect is listener action which begins with a review of listening processes and comprehension, followed by discussions on listenership as a lead in to nonverbal listenership thennarrows down to listener feedback as given by spontaneous listener facial actions and gestures as behaviours of interestthat are distinct from other nonverbal listenership behaviours.

The third aspect focuses on factors critical to interaction such as culture as a basis of human interaction with others, as well as the theories of identity and identification focusing on how they emerge from multimodal interaction where interlocutors project and perceive identities through different media. This review foreshadowsspecific discussions in subsequent Chapters where they are presented again.

2.2 Listener Joint action

Listener actions refer to active role, issues and position that listeners have to take and contend with during interaction and these include engaging an effective and efficient listening process, developing meaningful interpretation of speaker utterance as comprehension, providing feedback in form of requests as well as display of nonverbal listenership behaviours specifically, spontaneous facial actions and gestures during interaction.

2.2.1 Listening Process

As a starting point for the review of listener actions, this chapter outlines the process of listening because research suggests that, listening is vital for effective and efficient communication because it provides input (Nunan, 2002, p.238) for listeners during interaction even though, it is given the Cinderella treatment in contrast to speaking, reading and writing. As outlined in 1.1, this thesis focuses on listenership behaviours thus, this chapter provides a theoretical basis for understanding how interlocutors use one of the receptive language skillsaka listening as a systematic course of action to interpret speaker utterances during interaction.

Two views dominate research around listening focusing on how language learners as listeners process speaker input, these are bottom-up processing and top-down interpretation models (Harmer, 2007).

Bottom-up processing assumes that listening is a process of decoding speaker input incrementally with the smallest meaningful units of language (phonemes) reshaped into larger and complete texts. Listeners move progressively upwards decoding and linking phonemes to form phrases, phrases are linked to form meaningful utterances(Field, 2004) such as sentences. The listener is seen as a sequential store of information (Anderson & Lynch, 1988) while listening is the process of hearing speaker input from receiving aural stimuli and giving meaning to it (Oxford, 1993).

Top-down processing assumes that "larger units exercise an influence over the way in which smaller ones are perceived" (Field, 2004,p.364).This suggests that listeners reconstruct speaker meaning using the aural stimuli as a guide; for example, listener interpretation of phonemes depend on their knowledge of that particular word. Therefore, listening is not sequential but a framework of two strategic actions 'decoding' and 'meaning building,' with one running into the other (Field, 2008). Although these two views seem like opposites they are actually complimentary and I shall discuss these processes in 2.3.

For the purposes of this study, listening refers to interactive listening (Vandergrift, 1997; Rost,1991) and it is described as that vital aspect of interaction that provides feedback to the interlocutor in order to clarify meaning, signal understanding, or advance conversation (Vandergrift, 1997, pp.494-495). The concept of interactive listening is used in this study because research suggests that it enables listeners to assume active roles, develop cooperation with interlocutors to achieve the aim of interaction and resolve comprehension problems(Shipley, 2010; Vandergrift, 2004).

Nunan (2002) suggests that listening may be classified according to listening purpose, listener's role in interaction and the type of text listened to or any combination of these.
The purpose of listening is important because it is linked to the type of information that speakers communicate as outlined in Roman Jakobson's(1960) functions of natural language as referential, connotative, poetic, phatic, metalingual and emotive. Listening involves specific processes and strategies that meet each listening purpose relevant to the speakers language function. For example, listening to a poetry recitation for enjoyment will require appreciative thinking while listening to the same poem to pass valued judgement will require relational thinking (Oxford, 1993). At another level, listening to instructions for assembling IKEA furniture requires a different skill set from listening to a poem. Examples of purposive listening include listening for specific information(Nunan, 2002), listening for details, main ideas, appreciation, empathy and valued judgement (Oxford, 1993).

The role a listener plays during interaction also influences the type of listening. Nunan (2002) calls this reciprocal listening because in normal human communication listeners are often expected to respond as feedback. Examples include listening in action (Rost, 1991) and active listening (Shipley, 2010).Another way of characterising listening is in terms of the type of text being listened to.Type refers to the semiotic modes of texts that are used in meaning-making during interaction. The semiotic mode is used to differentiate verbal language from others (Sindoni, 2013) like written, audiovisual, multimedia and hypermedia formats. Audio texts offer only sound while audiovisuals offer sound and imaged text. Multimedia offers text, still images, animation, video, and audio presentationswhile hypermedia combines dynamic text presented on computers with interlinked nodes and multimedia (Moos & Marroquin, 2010). Three types of listening based on text modes are foundational, discriminate and comprehensive (Wolvin, 2010a, 2009; Wolvin & Coakley, 1996).

Although these variables enable categorisation of listening, the categories are not rigid because listeners may need a combination of listening strategies to be effective during interaction. For example, to be effective in instruction-giving contexts listeners have to consider the purpose of listening, their roles as instructees and the semiotic mode of the text they are listening to at the same time, they may have to draw from different strategies from all the skill sets during interaction.

Listening as an integral part of the communication process involves six elements including attending, understanding, interpreting, evaluating, remembering and responding. However, these elements may not always be present in every interaction because as outlined earlier, listening is affected by listener purpose, text type and listener role.

Attending refers to perceiving oral signals, understanding refers to deriving meaning, and interpretation has to do with giving meanings and reasons for utterances with the listener relying on schemata, context, topic and non-verbal signals from the speaker. Evaluation is a critical process where the listener separates facts from inferences in daily discourse while remembering involves recalling, re-grouping and recognising patterns of information received while responding refers to listener reaction or feedback to speaker utterance.

The success or failure of the listening process may depend on the listener's competence. Competence refers to a person's overall linguistic ability to perform (Esteki, 2014, p.1522) or do something successfully with a language which in this thesis, refers specifically to a person's ability to listening effectively and efficiently while taking verbal instructions.Research in ESL suggests that each listener has a competence or listening proficiency level, and this is described "…as the highest passage at which comprehension of the passage read aloud to the child is at least 75%" (McKenna & Stahl, 2015, p.67). The notion of highest passage describes the number and difficulty level of vocabulary items contained in the passage read aloud. The higher the density and level of difficulty of vocabulary items, the higher the passage. This is can be seen in the differences between Lamb's Tales from Shakespeare and the full edition of Shakespeare's plays. This implies that, listening may be assessed by the extent to which a listener is able to interpret speaker utterance and achieve the same level of meaning intended by the speaker during interaction.

Going further, Vandergrift (2004, p. 5) suggests that there are basic, intermediate and advanced listening proficiency levels. Basic level listeners are passive and do not need to respond to speaker input, for example, when listening to random radio broadcasts. Intermediate level listeners understand ideas but there is the risk of misunderstanding speaker input while advanced level listeners can understand

everything easily. Commeyras (2013) opines that effective listeners take the speaker's position, paraphrase utterances, adopt unbiased attitudes and respond genuinely while focusing on meaning, content, speaker intention and implied meaning.
The interactive listener is advanced with high proficiency and comprehension levels when compared to basic and intermediate level listeners as research indicates. For example, Vandergrift (2003, p.1-5) found that the more proficient listeners employed metacognitive strategies more frequently than the less proficient listeners. While Liu (2008), in a study among undergraduates in Taipei found that listeners with higher listening levels used top-down processing such as elaboration, whereas the less-proficient listeners had difficulty using the higher order strategy because they focused on unknown lexis or grammar. In another study of Iranian EFL freshmen, Bidabadi and Yamat, (2014, p.23-25)concluded there was a significant positive correlation between the listening strategies employed by listeners and their listening proficiency levels.

Having outlined the listening process, I will now focus on listening comprehension in order tooutline how listeners process speaker input during interaction.

2.2.2 Listening Comprehension

Listening process provides a means of dealing with speaker utterance as comprehensible input and as outlined in 2.2.1, this could involve top-down processing and bottom-up processing. Researchers in Teaching English as a Second Language(TESOL) and applied linguistics have tried to establish the link between listening and comprehension (Field, 2008, 1998; Gaiyan 2007; Flowerdew, 1994; Anderson & Lynch, 1988). This thesis draws influence from one suchapproaches developed by Gaiyan(2007) to identify the information sources available to listeners during interaction as a way of understanding the relationship between listening and comprehension.This chapter will explain Gaiyan's approach and extend it as a way of clarifying the the comprehension process.

Gaiyan (2007) citing Anderson and Lynch(1988) states that both L2 and L1 listeners need three information sources: schematic, contextual and systemic knowledge in every speech event(Figure 2.1) below. Gaiyan explains that the arrows leading to comprehension indicate a potential co-occurrence of "top-down expectation driven process and bottom-up data driven process" (Gaiyan, 2007, p.3). Furthermore, Gaiyan applies these processes as teaching strategies in China- where English isa Foreign Language (EFL) without explaining the intricacies of comprehension. Gaiyan's approach successfully outlines the basic features of the two views on listening comprehension and these have been adapted in this study to explain the listening comprehension processes in unidirectional instruction giving contexts (Figure 2.1) overleaf.

Krashen's (1982) comprehension hypothesis states that we acquire language and develop literacy when we understand messages as 'comprehensible input' comprising what we hear and read. What we hear during verbal interaction is speaker utterances and Krashen also stresses the importance of making utterances understandable to the listener during interaction. Listeners make speaker utterances understandable to themselves during interaction because of their knowledge and information derived from context, inferential schemata and systemic processes outlined (Figure 2.1).

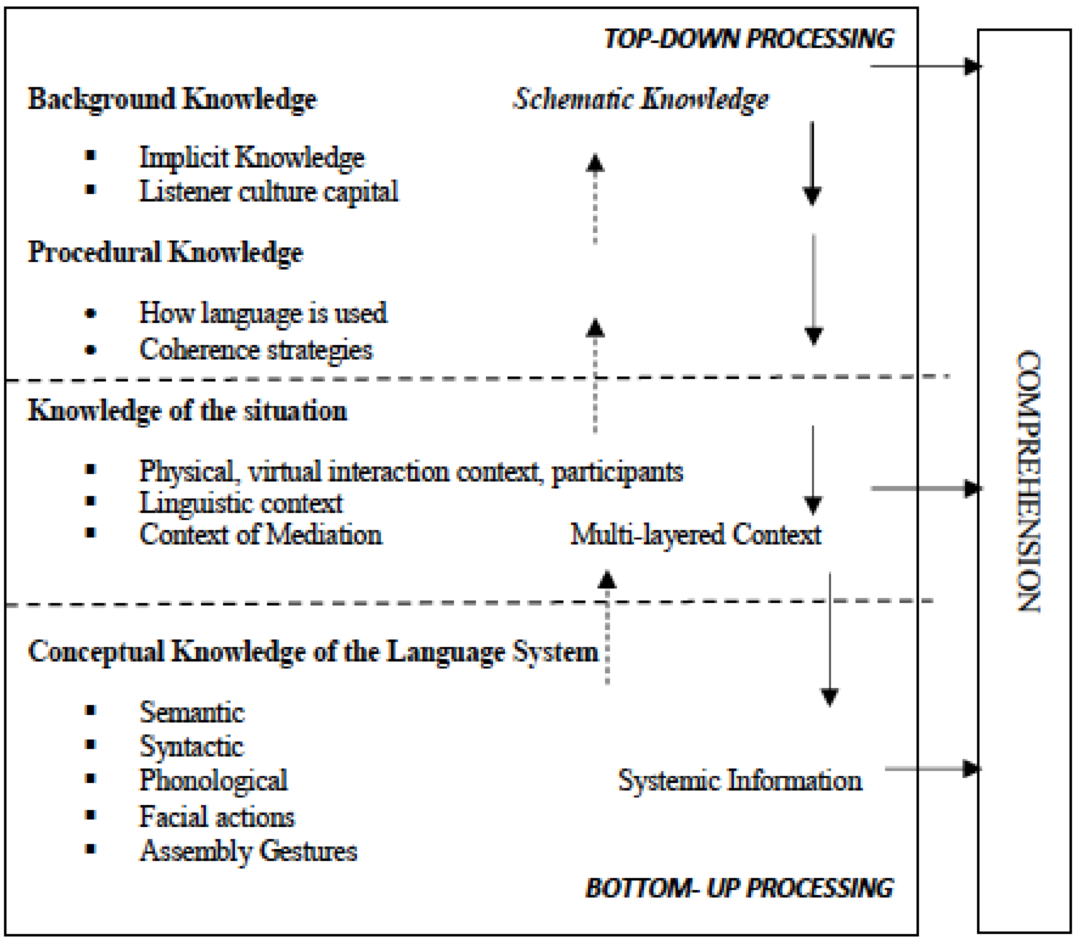


Figure 2.1: Listener's information Comprehension Process Modified from Gaiyan (2003, p.3)

The solid arrows coming down indicate the predictive nature of listening where the listener continues to guess or be in a state of anticipation about what the speaker will say or mean with each utterance also called ‘listening out’ by Lacey (2013) based on the information available. Meanwhile the broken arrows going up indicate what the listener does incrementally with the information received when giving feedback. The arrows going up and down do not signify separate routes for distinct processes rather, they suggest a continuous loop that meets at different points within seamless boundaries indicated by the broken lines across the space. The interaction between the information sources and meaning-making during comprehension may depend on listener role, listening purpose, text listened to and the speaker.

Another point of divergence is that whereas for Gaiyan (2007), arrows going towards comprehension suggest co-occurrence of top-down processing with bottom-up processing, in this thesis the arrows indicate the impact of knowledge and information sources on the comprehension processes as will be outlined later.

Harmer (2007) suggests that top-down processing is activated by the listener’s schemata. In this case schemata refers to inferential schemata described as “the ways that successive turns in talk can be interpreted” (Coulthard et al., 2016, p.10). And when listeners relate speaker input to specific interaction contexts such as instruction-giving, they interpret successive instructions using schemata and research suggests that this may support and lead to the development of procedural knowledge (Canagarajah, 2014).

Schematic knowledge comprises ‘background knowledge’ and procedural knowledge (Figure 2.1, p. 14). Background knowledge also called propositional knowledge consists of facts the listener brings into the interaction, for example, knowledge of the topic of discussion and implicit knowledge. Research suggests that strategic listeners develop implicit listening comprehension knowledge through performance (Vandergrift & Tafaghodtari, 2010) as part of background knowledge. Implicit knowledge is derived from everyday interaction unintentionally and without the subject being aware of such knowledge. It is the opposite of explicit knowledge that is deliberately and consciously developed by the subject (Esteki, 2014). Rebuschat and Williams(2012) point out that language comprehension is largely dependent upon implicit knowledge , for example, in online speech comprehension learners rely on implicit knowledge.

Background knowledge also consists of aspects of the interlocutor's listening culture capital. Culture capital is described as the embodied form of long-lasting dispositions of the mind and body (Latour, 2014) acquired and unconsciously used as a social asset, for example, turn-taking, facial actions and gestures during interaction.

The role of top-down information in comprehension is partly derived from Stanovich's (1980) Interactive Compensatory Hypothesis (ICH) which posits that, interlocutors use contextual information to compensate for inadequacies in the communication medium. ICH is relevant in this thesis because inappropriate implicit knowledge or irrelevant listening culture capital (Figure 2.1, p. 14) may rupture listeners'understanding of speaker-utterance. This ties in with Field's (2008) suggestion that listeners seem to use more top-down information to compensate for gaps in their understanding.

One compensatory measure available is the activation of schemata. Mazzone (2015) explains that listeners activate schema by making inferences from explicit meanings of speaker-utterances to the intended conclusions. For example, a frame like 'good morning' activates the schema of 'greetings' and the need to reply with the same or a more suitable language frame. The activated schema represents listener-culture capital brought into a new context. The parallel in task contexts will be that, when instructors say, "Select X", listeners need to comply by picking out "X"from other alphabets.

Procedural knowledge is knowledge of steps and actions used to achieve a goal (Rittle-Johnson et al., 2015) which in conversation focuses on how and when language is used in given contexts. Canagarajah (2014) compares procedural knowledge with Canale and Swain's (1980) strategic competence as a component of communicative competence because procedural knowledge enables listeners to manage communication, negotiate meanings, codes, and identities while achieving the aims of interaction.

The interactive listener employs various procedures in processing speaker utterances including what Hansen and Jensen (1994) call 'global and local coherence strategies' that roughly correspond with Brown's (2007) top-down and bottom-up listening macro skills.

Listeners use global coherence strategies to take in speaker-utterances such as assembly instructions and process them at different rates as outlined in bottom-up comprehension. To do this effectively, Brown (2007) suggests that listeners need to retain chunks of speaker utterances (this fore-shadows the role of focus in the structure of instructions in Chapter 2.5) and recognise word order, lexical and contextual meanings (Brown, 2007). For example, when listeners receive instructions, they may retain parts and use later as recoverable information anaphorically (Chapter 2.5) during tasks.

Furthermore, listeners use local coherence strategies to recognise links between major ideas in a discourse and the overall structure of the discourse in a top-down comprehension process. To do this efficiently, Brown (2007) suggests that effective listeners need touse body language in meaning-making, second guess meanings of utterances from use and ask for repeats or clarification when in doubt.

Another major source of knowledge is context. The notion of "multilayered context" (Chun et al.,2016, p.68) is used to project the understanding that when people interact with technology, the context has layers such as interaction context, linguistic context and mediation context. This is another point of departure from Gaiyan's (2007) proposal which described context from two views only. This thesis applies the concept of context as outlined in chapter 1.1, to assembly tasks focusing on nonverbal listenership behaviours emerging from unidirectional instruction-giving (Chapters 4 & 5).

Forms of discourse used in interaction constitute a dimension for characterising context (John, 2016) as linguistic. The linguistic context requires socio-affective strategies for building interaction (Bidabadi & Yamat, 2011) such as the ability to predict the relationships between linguistic forms and meanings also called bootstrapping (Huang et al., 2016) during interaction. This is because the linguistic context also refers to the description of the occurrence of semantic and syntactic forms of language used in interaction, such as parts of speech.

In addition, interactive listeners also require contextual cues to successfully make meaning of speaker utterance. Adolphs (2008) explains that contextualisation cues are used to analyse the relationship between surface structures and context. Contextualisation cues are surface structures that could be lexical, linguistic, paralinguistic or prosodic indicating contextual suppositions at different discourse levels and used to encode speaker-expectations about the upcoming discourse (Adolphs, 2008). For the listener, these features are used to enrich meaning, make it relevant, aid decoding processes andinfluence listener-orientation towards interactions.

Furthermore, theuse of a simulated agent-instructor with three voices (Chapter 4)is a form of technologically mediated language use. Interlocutors should give consideration to the context of mediation which is the physical medium of utterances (agent voices) and how it interacts with other types and layers of context(Chun et al., 2016) such as the physical and linguistic contexts outlined in Figure 2.1.Although the use of a human instructor (Chapter 5)is close to natural conversation, a listener's experience in both contexts is influenced by media used in interaction because as Chun et al (2016) explain, the same text in one medium may be experienced differently by listeners when presented through a different medium. For example, assembly instructions given in sythesised voices may be experienced differently when issued by a human instructor hence, the need to compare listenership behaviours emerging from HAI and HHI(Chapter 6).

The last source of information during comprehension is systemic (Figure 2.1). This relates to the listener's conceptual knowledge that provides the building blocks for bottom-up processing. Conceptual knowledge is described as knowledge of abstract and general principles (Rittle-Johnson et al., 2015, p.588)of a language system. This comprises verbal (semantic, syntactic, phonological and prosodic) and nonverbal (facial actions and gestures) components of the language that form aspects of co-text used by the listener during interaction. Conceptual knowledge is comparable to Canale and Swain's (1980, p.1-3) grammatical competence because it is critical to listener processing of speaker input as a strategic skill and for determining listener comprehension skills.

Conceptual knowledge also relates to lexical coveragewhich refers to the percentage of words in a piece of discourse known to interlocutors (Adolphs & Schmitt, 2003) in this case the listener. The role of lexical coverage in listener comprehension has been found to be critical for both L1 and L2 users of English becausevan Zeeland and Schmitt (2012, p. 475) found that although L1 and L2 speakers demonstrated adequate comprehension of spoken texts, the L2 speakers showed less variation than L1 speakers. They concluded that at 95% coverage, language users would need to know between 2,000 and 3,000 word families for adequate listening comprehension" (van Zeeland & Schmitt, 2012, p.457) during interaction.In addition, research suggests that less skilled listeners either spend more time decoding unfamiliar words or rely more on context to decode speaker utterances because they are either too engrossed with details or lack the linguistic competence to properly decode the utterances (Osada, 2004; Tsui & Fullilove, 1998, p.432-437).

Competence in turn, influences listener feedback and attitudes towards the interaction.Hendrikse et al. (2016,p.94) hold that attitudinal reactions describe the listener's willingness and ability to react and/or respond to speaker utterances and reject or accept the message verbally and nonverbally. This implies that in the first place, listener comprehension competence is dependent on their lexical coverage in the language of interaction. Secondly,the level of competence influences their ability to respond to verbal instructions and this may provide a second view to the integrative use of spontaneous listener facial actions and gestures during interaction.

Thus, this thesis will use the insights obtained from this review to design aspects of the research framework in Chapter 3 that will be used in analysis (in Chapters 4 and 5) that may make explicit the relationship between nonverbal listenership behaviours (the focus of the next chapter) with listener comprehension or incomprehension of instructions.

2.2.3Spontaneous Nonverbal Listenership as feedback

This is the third aspect of listener action and as explained in Chapter 1.1, listenership is that active role played by a listener that builds communication and provides feedback during interaction. Nonverbal listenership is used in this thesis to describe marked spontaneous facial actions and gestures displayed by participants during assembly tasks. Facial actions are facial movements in terms of component actions used in nonverbal communication while gestures are motions of limbs or body to express or help express thought feeling or to emphasise speech (Givens, 2015; Kendon, 2004; Ekman, 1997).

Nonverbal listenership falls within Yngve's (1970)backchannels just as response tokens (O'Keeffe et al., 2007; Gardener, 2002), listener response (Maynard, 1990), and newsmakers (Gardener, 1997) such as nodding of the head (Knight, 2009; White 1989) and gazing (Young 2004; Fellegy 1995).

The role of listener nonverbal listenership behaviours in interaction as expressed in this thesis ties in with White's (1983)view that the speaker and listeneruse two channels of communication during interaction. Speakers use the main channel while listeners use the back channel (–where and when spontaneous facial actions and gestures occur) to interject speakers without claiming the floor during interaction.

In addition, listenership behaviours help to sustain the flow of interaction and Knight (2009,p.35-37) stresses that they exist to reinforce Grice's maxim of cooperation in talk that enable listeners to signal attention to speakers during interaction. Furthermore, while a turn for a speaker means'I speak, you listen, while listenership for listeners implies that 'I listen while you speak'(Oreström, 1983, p.21).

The discussion will now focus on the characteristics of spontaneous nonverbal listenership behaviours. Spontaneous facial actions and gestures are generated by biologically-given processes that operate automatically, eliciting muscle reactions quickly and independent of conscious cognitive processes (Dimberg et al. 2000;Ekman 1992; Dimberg 1990), for example, faces made when a person sees a sudden flash of lightning. Meanwhile simulated facial actions are approximations of facial behaviours that occur when people make little attempt to manage facial appearance (Ekman et al.1980, p.1126-7), for example, faces deliberately made for a photoshoot.

Furthermore, associative stimuli such as voice commands, a hot object, and harsh words may elicit spontaneous emotion display during interaction. However, spontaneous display of listenership behaviour is also determined by individual differences or somatic factors in a process explained in Chapter 7.2.5. Somatic factors describe characteristics relating to the human body or trunk, skeleton, muscles and nervous system that control emotional expressions, perceptions and bodily responses to external stimuli (Mooney, 2012; Zajonc, 1984). Somatic factors have been found to contribute to the generation of nonverbal behaviour (Ekman & Friesen, 2003; Tomkins, 1984).

Furthermore, spontaneous facial actions and gestures align with other emotional behaviourslike the voice or posture showing fluency in communicationwhereas, simulated ones show people trying to mask, reduce or check one emotion by portraying another thereby causing dysfluencies in these expressions. Spontaneous behaviours are also more reflexive and smooth with fewer phases than deliberate behaviour (Hess & Kleck 1997, p. 271-273). Conversely, simulated emotions are more assymetric because their timing of muscular contraction is more irregular with missing components in the frequency of physiological movements in different parts of the face (Weiss et al.1987; Hager & Ekman,1985; Ekman et al.,1981).

Spontaneous facial actions and gestures are also mediated by neurological pathways that are different from those mediating simulated actions. For example, the chemicals released when spontaneous emotions are expressed, are never released when emotions are simulated. This is also true for changes in body temperature, fluid movements and oestrogen injection into the body system (Weiss et al. 1987).

On the issue of prevalence of facial actions and gestures in HHI, research has found differences and similarities in expressive behaviour across cultures (Mehrabian, 2007;Trandis & Lambert, 1958; Engen et al., 1957). Some researchers suggest that similarities occur due to evolution, universal innate development programmes or learning processes in the human makeup or genome (Knapp et al., 2014; Fridlund, 2014).Others suggest that differences could depend on human neuro-physiology, constants in early learning, evolution, ontogeny or development of each individualfrom conception(Ekman, 2004, 1997; Kendon, 2004; Gosselin et al, 1997).

The discussion will now turn to the categories of gestures and facial actions to clarify the evolutionary and innate factors responsible for their characteristics which Ekman (1997, 1992b) suggests, evolved to aid people to deal with fundamental life tasks such as achievement, failure and frustration. In recognition of this and the revelatory nature of interaction, the next step of this review will cover the types of facial actions and gestures existing from earlier research to serve as a guide for describing those that will emerge from this study.The interaction is task-based which implies that facial actions and gestures emerging may comprise a variety and this chapter will influence how facial actions and gestures are categorised later in this thesis.

2.2.3.1 Categorising Facial Actions

Ekman (2016,1997,1982a) identifies eight (8) classes of facial actions given as basic emotions, non-basics, emotional attitudes, moods, emotional traits, emotional disorders, emotional plots and blends. However, the categories relevant to this thesis include basic, non-basic, emotional attitudes, and moods (Table 2.1) overleaf. This does not discountenance the fact that, other categories may emerge because of the nature of interaction in a task-based context used in the study. FACs the coding scheme (Ekman and Friesen, 1978) is widely used to code facial actions and most relevant to this thesis as it provides the referential guidelines for classifying and encoding listener facial expression (Details in 3.3).

Although research into facial actions focuses on emotions, it is important to note that not all facial actions portray emotions. Nevertheless, facial actions provide the motivation, arousal, adaptive functions, control and variety necessary for effective social interaction between people and other interlocutors such as smart agents and these bring a social dimension to HCI (Bartlette et al., 2009). Ekman (2007) suggests that emotions are distinct from other facial actions because emotions have automatic appraisal, commonalities in antecedent events, presence in other primates, quick onset, brief duration, unbidden occurrence, distinctive physiology and a distinctive (universal) signal. Furthermore, it has been arguedby the neuroscientist Antonio Damasio (2001) that emotions are distinct from feelings because an emotion is an involuntary bodily response, a more complex version of a reflex; for example, when one is in a dangerous situation, one's pulse increases, blood flows away from one's digestive system. On the other hand, a feeling is the process of one becoming aware of emotion such as sensing danger and fear within us.

Basic facial actions (Table 2.1) portray basic emotions. Ekman (2016, 1997) posits that emotions are described as basic based on three premises that lead to the emergence of emotion families and their variations. These premises also inform the framework designed for analysing facial actions in Chapter 3.

Firstly, emotions differ from one another in their expression, appraisal, antecedent events, common features and functions outlined in Chapters 4.4.3 and 5.3.3. Secondly, the differences between individual emotions are attributable to the principle of ontogeny or the process of human development as an organism while universal emotions are due to conventionalisation or wide acceptability, presence and recognisability as outlined in Chapter 3.3.3. Thirdly, there is the existence of non-basic emotions which are a combination of basic emotions called blends or mixed emotional states (Ekman,1997).

Table2.1 Summary of Facial Actions (Adapted from Ekman, 2016, 2007, 2003)

S NO	FACIAL ACTION	ILLUSTRATION	
1	Basic Emotions anger, disgust, sadness, fear, happiness, others: shame, surprise, embarrassment	Surprise, felt smile, neutral	1. The differences between individual emotions due to ontogeny while universals are due to conventionalization 2. Evolutionary and innate factorsare responsible for thefeatures of basic emotions and they deal with fundamental life tasks 3. These led to the development of **emotion families** and their variations 4. Neutral is a basic emotion because facial muscle movement or lack of it is one form of expression 5. A recent survey among scientists researching around emotions by Ekman, (2016) indicates that:unlike 20years ago, there is a consensus supporting (80%) Darwin's proposal that emotions are universal 6. basic emotions includeanger, fear, disgust, sadness and happiness. 7. Other emotions range from shame, surprise and embarrassment 8. Basic emotions may be viewed through the lens of positive-negative; approach-avoidance; frameworks of emotional responses
2	Non-basic or blended emotions	Blends	These are combinations of basic emotions called blends or mixed emotional states
3	Emotional attitudes	Frown	They are more sustained and may involve more than one emotion e.g. frowns
4	Moods	Apprehension, euphoria, irritation	Each is infused into different emotions: • Moods are adaptive and do not develop from evolution; • A mood activates specific emotions , for example, when irritable we get angry; • Last a lot longer than emotions, • While we can identify the cause of an emotion, we cannot do this for moods; • Reduce a person's flexibility, • Distort our thinking; • Reduces our ability to control;

			• Although they prepare us for future experiences but they are more of a problem than beneficial;

Some basic facial actions are universal because of their presence in all human populations because we are hardwired to produce and recognise facial actions. The universal ones are basic facial actions that evolved within man to deal with fundamental life tasks, such as achievement or failure (Ekman, 1997). Examples outlined in Table 2.1, include smile-as-joy, anger, disgust, sadness, fear, happiness, surprise, and neutral expressions (Ekman 2016).

I will now briefly describe some. Ekman (1982a, 2007) says fear, surprise and interest are interrelated because they can be confused by observers since they are determined by the same internal processes (Damasio, 2001; 2008) and so look alike.

A smile is easily recognisable, however, its origin is said to be confusing because some believed that it is the opposite of all negative feelings, others like Van Hoof as reported by Ekman and Friesen (1982) argue that smiles evolve from fear but this is countered by another argument that fear and smiles are distinct. Studies focusing on smiles suggest that emotions have display rules that are applied selectively by people from different cultural backgrounds in varying ways.
As per typology, Ekman and Friesen (1982) identified three types of smiles including the felt smile, false smile and miserable smile, while Pease and Pease, (2004) and later Malik (2010) identified five types, tight-lipped, twisted smile, drop-jaw smile, sideways looking smile, and George W. Bush grin.These classifications overlap as the features in one may be used to explain the functions of the other. For example, the drop-jaw smile is the same as a felt smile yet one variation is a false smile.

As outlined above, there are non-basic emotions such as smugness and scorn (Table 2.1)occuring from a combination of basic emotions becausefacial muscles are complex enough to display a blend of emotions (Ekman 1997). The blending process involves one emotion running into another to produce a new one (Ekman& Friesen 1978a) in a manner similar to articulation of diphthongs where one sound runs into another to sound as one.

Ekman (1997) observes and Damasio(2007) agrees that the blending process is possible when the eliciting stimulus is also a blend for two reasons. The first is when the emotion eliciting circumstance elicits more than one feeling. The second is when habits (common to a group, or idiosyncratic) link one emotion to another such as, when a second emotion is generated in response to the initially inspired one during interaction.

Furthermore, Ekman (1992b) suggests that people display a blend of emotions when the eliciting stimulus is also a blend. In addition, blends sometimes create differences between the emotion truly felt and the one feigned leading to what Ekman (1982a) referred to as 'overdo behaviours', which is understood in the same sense as overacting.

Going further, emotional attitudes (Table 2.1) are found to be more sustained than basic emotions and may involve more than one emotion being elicited during interaction (Ekman,1997). Listeners have attitudinal reactions such as frowns that may indicate their willingness to accept or reject a message (Allwood, 1995,1993) in ways that are different from emotions but indicate personal preferences, affective disposition and intrapersonal stances (Scherer,2005) during interaction.

Ekman et al., (1972) observed that people employed different display for different contexts , for example, display rules used in private are different from those used in public or social situations. Moods (Table 2.1) are infused with different emotions (Ekman, 2007) and are found to be adaptive but unlike basic emotions, blends and attitudes, moods did not develop from evolution. In addition, moods activate specific emotions; for example, when one is irritated one gets angry. Thus, moods may distort our thinking, make us lose self-control, and reduce our flexibility during interaction. Research also suggests that moods last longer than emotions, prepare us for future experiences but may be more problematic than beneficial in interpersonal relationships(Ekman, 2007,1982a).

Micro-expression also called"hot spots" (Jeremiah, 2014) occur when people deliberately give off emotions that do not reflect their true feelings. Research indicates that people can control facial actions because social pressures sometimes dictate such control and the very behaviour expected in each context is concealed (Ekman and Friesen, 1969a-b). Furthermore, microexpressions provide a full picture of the concealed emotion, but so quickly that it is usually missed however, microexpressions are the greatest source of information leakage from the human face (Porter et al., 2012; Ekman, 1988).

Regarding the manifestation of microexpressions, research suggests that, when a person is experiencing emotions, the physiology takes over so, even when people try to mask their true feelings, they still leak out (Bartlett et al., 2009; Ekman, 1988). In addition, Jamine Driver in Miller's (2014) documentary argues that microexpressions are not only spontaneous but they possess the quality of consistency with the context; for example, the spontaneous reaction of one wrongly accused by others may make the accused person's innocence apparent.

Secondly, there is also consistency at the level of multimodal alignment and Miller, (2014) suggests that when people speak, all gestures and emotions expressed face the same direction, expressing the same thought or reinforcing one another. Hence any disconnect between any of these suggests that the person may not be telling the truth but the truth leaks out either way through microexpressions(Miller, 2014;Matsumoto & Ekman, 2004; Ekman, 1988) These features imply that microexpressions are consistent with the real context or focus of interaction, attitudes towards the interaction and the true feelings expressed by the person even when we try masking them.

The cue to assessing microexpressions is the duration but this has raised some debate. Ekman (2003, 1982b) holds that it lasts 1/5th of a second, Matsumoto and Hwang, (2011) suggest ½ a second while Yan et al., (2013) concluded that it may be physiologically impossible for the duration to be shorter than 100 microseconds.These confirm Pfister et al.'s (2011) observation that the length definition of microexpressions varies, which implies that the exact duration is not easy to come by. Following these, microexpressions are described in this thesis as very fast facial actions that last between 100 and 500 microseconds.

The anatomically based Facial Action Coding System -FACS (Ekman & Friesen, 1978a) is the most comprehensive method of coding facial displays and will provide the bedrock upon which the framework for categorising facial actions (see chapter3.2.4.2& 3.3.3.1) that emerge from the interactions (Chapter 4 & 5) in this thesis.

2.2.3.2 Categorising Gestures

As outlined in 2.2.4, gesture describes a variety of body movements used to communicate specific messagesor perform specific communicative functions during interaction as outlined in chapter 3.3.1. To get a sense of when between body movements become gestures, I will illustrate with the following example, hitting a friend in the arm while chatting or playing a game of scrabble may just be classified as a body movement that passes no message however, when the movement is accompanied by an utterance such as ‘stop wasting time’ it becomes a gesture used to hasten the receiver to action; when accompanied by head nods and eye movement pointing to a given direction, the same movement becomes a gesture used to show direction or indicate that something or someone is coming from a given direction or just as a warning to the receiver.

Research has shown that gesture is processed by a common neural system with other language skills (Xu et al., 2009; Butterworth & Hadar, 1989), blind people and even chimpanzees gesture so, it may be linked to motor action (Matsumoto & Hwang 2013). Gesture and language skills are thus linked temporally, structurally, pragmatically, and semantically which enables them to jointly use the same framework(Enfield et al, 2007; Kita et al, 2007, p.1213; Rauscher et al.,1996, p.227; McNeil, 1992, p.9).

While the face provides information about the occurrence and nature of emotions, body acts like posture and position provide information about the intensity of the emotion (Ekman &Friesen 1969, p.90). This agrees with the view that the ability to gesture co-evolved in humans as they were adapting to their environment, physical anatomy, cognition and communication needs (Matsumoto & Hwang 2013). Furthermore, these views imply that gesture evolved alongside speaking and listening and with the advent of literacy, reading and writing.

As per meaning generation during interaction, Goodwin (2003) adopts an integrative approach and treats gestures, body posture, gaze, talk and environmental structure as components of embodied action in contextualised human interaction. This view is linked to the notion that gestures are only meaningful and communicative when communicators contextualise them in special ways by placing gesturesexplicitly in speech situationsor fields of interaction such as HAI, HHI, instruction-giving and assembly tasks.

However, it is important to distinguish between all the actions described as gesture that occur along Kendon’s continuum (McNeil, 2006, 1996). Kendon’s continuum proceeds as follows, gesticulation, speech-framed gestures, emblems, pantomime and signs (Figure 2.2) below.

Figure 2.2: Kendon's Continuum (Adapted from McNeil, 2006, 1996)

At one end of the continuum there are gesticulations described as idiosyncratic spontaneous movements of the hands and arms accompanying speech (Kendon, 2004, p.99). Gesticulations co-occur with speech, embody meaning related to speech but do not replace speech in HHI and have been shown to precede speech and are present in most languages(McNeil, 2006, p.18).

The next are speech-framed gestures which are built into the sentence where they occupy a gap such as when interlocutors hesitate or lose words when speaking (Kendon, 2004, p.99; McNeil, 2006).

These are followed by emblems also called representational gestures (McNeil, 1992, p.7) that convey direct verbal translation or semantic meaning through hand shape, position, or motion and do not occur with speech or content (Kendon, 2004, p.5).Matsumoto and Hwang (2013) also describe emblems as spontaneous verbal messages encoded in body movements understood as a form of cultural communication. Examples include “… word ‘good’ (thumb up, hand in fist)” (Matsumoto & Hwang, 2013, p.2).Emblems have metaphoric and iconic contents that exemplify quotable gestures such as the money gesture used to ask for money. Emblems are useful for communicating in noisy arenas or in places where talking is impossible.

The next are pantomimes, which are gestures that convey narratives, messages or whole texts without speech (McNeil, 2005). And at the other end of the continuum are signs. Signs are gestures that function independent of speech and are highly conventionalised (Kendon, 2004, p.33; McNeil, 2006). They comprise lexical words in sign languages such as the American Sign Language (ASL) and British Sign Language (ESL).

In addition to Kendon's continuum, there are dimensions introduced by McNeil (1996) and McNeil and Levy (1993) based on semiotic categories that I call the McNeil dimensions, which fall within gesticulations and speech-framed gestures in Kendon's continuum. McNeil's dimensions include iconic, metaphoric, deictic, beats and Butterworth gestures. McNeil's dimension gestures co-occur with speech and are classified as 'co-speech' (Mol et al., 2012) or co-verbal gestures (Matsumoto & Hwang, 2013) and also function as illustrators (Ekman,1982a)that depict some content of a message.For the purposes of this thesis, McNeil's (2005, 1992) classification will be used as guide under which other categories will be classified in this review as described below.

Iconic gestures also called illustrators (Ekman & Friesen 1969) are interwoven with speech and illustrate aspects of the semantic content of speech.For example, when a person portrays a physical thing by showing how thin it is. Iconic gestures are distinct because they are used to show physical, concrete items. They add details to mental imageries created by communicators in the manner of nonverbal similes and metaphors. The timing of iconic gestures in synchrony with speech can indicate whether they are spontaneous or deliberately added for a deliberately calculated effect (Sekine et al., 2015;Mol et al., 2012;Kendon, 1980). When spontaneously produced, the preparation for the gesture will start just before the speech is made or along with the speech. However, when gesture is staged, McNeil (2005, 1996, 1985) says that there is a small lag between speech and gesture that can portray the communicator as manipulative.

Metaphoric gestures are the same as Ekman and Friesen's (2003) pictographs and ideographs are three–dimensional, like human vision and perception of the environment, thus, they can make an abstract entity concrete. For example, a cartoon's pictorial content is abstract while the comic strip is concrete. Narrators would present the content metaphorically as if they are offering an object to an audience using frames like 'here is xyz'. Metaphoric gestures are also used to explain abstract ideas with finger pinches, physical shaping and general waving of the arms to bring the message clearly to the listener or observer.

Beat gestures are the same as Ekman and Friesen's (2003, p.65) rhythmics. According to Cornelius et al. (2013), beat gestures are simple rhythmic gestures that do not contain semantic content and can be as short as a beat or as long as needed but the aim is to make a point and they are not dynamic even though the beat may change. Holler and Wilkin (2011) suggest that beat gestures tend to have the same form regardless of the content, for example, a simple flick of the hand or snap of the fingers. Beat gestures can create emphasis in communication like the judge striking the hammer to pass judgement, or a teacher rapping on the Table to get noisy students to pay attention in class.

Deictic gestures, also called pointing gestures(Cartmill et al., 2012, p.130) indicate real, imaginary or implied entities, directions, ideas etc. They are connected to the conversational or gestural space (Goodwin, 2003) even when there is nothing to point at and can occur with or without speech. For example, a speaker may ask 'where did that idea come from before?' while pointing at a space on the Table. The space was not, in fact, the idea's former location but, over the course of the conversation, had come to represent that point of origin called 'origo' (Goodwin, 2003, p. 273).

Butterworths (named after British psycholinguist Brian Butterworth) are used to describe gestures people make when they try to recall something, for example, closing their eyes, staring at a point, stamping their feet, putting the hand on the forehead, grinding their teeth or even shaking their heads sideways. However, it is important to emphasise that these vary from culture to culture.

This thesis believes that Kendon's continuum as outlined by McNeil (2006) may keep growing because the communication context is expanding due to technological development. Technology is changing the way we communicate and extending discourse to contexts where we interact with smart agents thus, this thesis argues that in order to understand the role of gestures in these interaction contexts, researchers may need to come up with new gesture classifications that are based on the interaction as outlined later (See Chapters 3.2.4.2; 5.1; & 7.2.3).

In addition, influences from studies by McNeil (2006), Kita (2008), Mehrabian (2007; 1967), to mention but a few provide the bases for designing a measurable guideline for sub-dividing and coding gestures emerging from interaction during assembly tasks (see 3.3.2).

2.2.3.3 Spontaneous Facial Actions and Gestures in Interaction

The discussion will now turn to the roles that facial actions and gestures play in task-based interaction as a basis for understanding their impact on unidirectional instruction-giving assembly contexts. This chapter is structured to first focus on gestures then facial actions.

Gestures structure conversation and aid mutual meaning development. Research (Mol et al., 2012) indicates that the perception of gesture and speech contributes to the mutual construction of concept representation by interlocutors. They also hold that observers, like listeners, repeat gestures and later adapt them when interacting with others. In addition, mimicking other people's gestures leads to the creation of a shared understanding and meaning between interlocutors (Holler & Wilkin 2011). Although mimicry is 'social glue', some gestures may be omitted during interaction for social reasons such as the interlocutors' self-awareness of the negative impact of solecism in social interaction which may make them to avoid doing something that is socially unacceptable.

Furthermore, research indicates that gestures aid task completion rates in listeners andwhen speakers were barred from using gestures, they became less fluent (Kipp & Albretch, 2007). Gestures may influence the course of spatial reasoning, problem solving and enhancing listener's discourse comprehension (Sekine & Kita, 2017;Kale et al., 2010). Research also found that a person's cognitive workload associated with task handling is significantly reduced when exposed to natural gestural videos (Cornelius et al,2013, p.94).

Moreover, Pineet al.'s (2010, p. 175) study indicate that gestures are critical for conceptualisation and speech development. They add that gestures facilitate pre-verbal thinking and the generation of an utterance i.e. from thinking to speech generation. Research has also suggested that the representation of meaning activated during production may activate the gesture because of the interactive alignment (Mol et al., 2013). These cyclic or interactive cause and effect relationships influence the listening process too as shown in Cornelius et al.'s (2013) studyindicating that gestures enhance listener’s comprehension of communication about spatial concepts and the feedback they provide.

Furthermore, Cartmill et al, (2012) found that gestures precede and predict linguistic development as children used co-speech pointing gestures to modify and supplement speech by creating a sentence-like meaning using formulaic expressions or language chunks (Ellis, 2005, p.73-92) like “eat food; car drive”. They concluded that these enabled interlocutors to anchor communication in their immediate environment.

In addition, research indicates that representational gestures are used to represent the form of task objects and the nature of actions to be used with those objects (Fussell et al., 2004). Instructors providing and clarifying instructions during interaction require complex coordination (Fussell et al., 2004, p.296) with instructees, however, the complexity is reduced when instructors use pragmatic gestures to scaffold conceptual development (Capone and McGregor, 2004) during interaction.

Furthermore, research suggests that conversational grounding occurs when people carry out physical tasks through gestures, actions and speech because the use of gestures facilitated task performance (Lis, 2012; Capone and McGregor, 2004; Fussell et al., 2004; Dimberg et al., 2000; McNeill, 1989).

During interaction, recognition and repetition of gestures by interlocutors is easy because "gestures are an integral part of language as much as are words, phrases and sentences, gestures and language are one system" (McNeil, 1992, p.2). This view is reinforced by later researchers indicating that gestures and speech stem from a single concept or communicative intention and possibly have the same conceptualisation (Cartmill et al., 2012;Kendon, 2004; Kita & Ozyurek, 2003).

However, there is an argument by Krauss et al. (1991) that gestures communicate to a limited extent but not as much as some theories would have us believe. However, I observe that Krauss et al’s (1991) study exposed the subjects to specific gestures in isolation. Secondly, the study did not take cognisance of the role of culture in nonverbal interaction because each culture has its system of encoding and decoding gestures.

The discussion will now focus on the role of facial actions in interaction.Research suggests that listener feelings and attitudes towards the interaction may also be distinguishable as positive, neutral or negative in both HHI and HAI (Clark 2016; Asano et al. 2014; Bonoma & Felder, 1977; Mehrabian & Ferris 1967). In addition, facial actions are important when communicating implicit feelings and attitudes (Carpenter, 2015; Ackermann & Mathieu, 2015) measuring bias (Meadors & Murray 2014) and validating behaviour towards specific communication as positive or negative evaluations, or as submissive or dominant (Bonoma & Felder 1977) during interaction.

Furthermore, expressed emotions influence outcomes that are important for an interlocutor's role in interaction. The interlocutor's role is linked to their identity (see Chapter 2.6) which is used to encode and decode positive, neutral and negative emotions. In this regard, Rafaeli and Sutton (1987) also report that expressed emotion is either positive or negative, esteem enhancing vs esteem degrading and sometimes involves physical touch. Positive emotions such as smiling and enthusiaism are esteem enhancing while negative emotions such as frowning and disgust are esteem degrading thus, the esteem enhancing-degrading sequence emphasises expressed emotions comprehended by the audience.

Research on the impact of emotions on social interactions focusing on how one's emotions influence one's own behaviour in bidirectional relationships indicate that, in general, happy people are more willing to help, to cooperate with, and to trust others in a negotiation. In contrast, people experiencing negative emotions are in general less trustworthy, more competitive, and more selfish (Dunn & Schweitzer 2005; Sanfey et al. 2003; Forgas 1998; Allred et al. 1997; Pillutla & Murnighan, 1996; Baron, 1990; Baron et al. 1990; Travis 1989; Berkowitz 1989; Carnevale & Isen 1986).

At the same time, an interlocutor's role is linked to their identity (see Chapter 2.6) which is used to encode and decode positive, neutral and negative emotions. This implies that the impact of emotions on the interlocutor's role is bidirectional during interaction which is plausible in view ofAndrade and Ho's (2009) study indicating that one's emotions influences co-interlocutors' decisions during negotiations. The study concludes that, people can intuit and strategically modify the expression of a current emotional state to improve their wellbeing in each social interaction context (Andrade & Ho, 2009).

Human emotions have been used in HAI to determine facial recognition in robots and intelligent agents (Cowie et al 2001), measure user frustration when interacting with agents (Miles et al. 2013) and monitor emotion regulation with computers (Klein et al., 2002). However, the agent's ability to use a human co-interlocutor's facial actions communicatively at human levels requires further understanding for agents to respond to human behaviour effectively.

As outlined earlier, gestures and speech have a common concept and are integral aspects of language. This implies that facial actions and gestures as components of nonverbal languagemay share the same conceptualisation that makes their repetition and recognition easy for interlocutors. And as outlined in 2.2.4.2, while the face provides information about the occurrence and nature of an emotion, gestures provide information about the intensity of the emotion which suggests that when gestures co-occur with facial actions, interlocutors may communicate better.

Regarding the co-occurrence of facial actions and gestures during interaction, it will be of interest to see how this relationship affects their functions in task-based interactions and informs assessment procedures to be outlined in chapter 3. This is important in view of research indicating that, gesture production processes are bidirectionally linked with speech production processes (Kita, 2013), speech and gesture interact and are linked in language production and perception during interaction (Wagner et al., 2014); gestures co-occur with speech and thought (Kita et al, 2016; Kendon, 2004) and gestures share the same computational similarities with overt language skills (Wagner et al., 2014; McNeil, 1985) such as active listening (see Chapter 2.2).

2.3Speaker Action: Constructing Assembly Instructions as Information Units

This is the speaker's aspect of the joint action during assembly tasks. As outlined in Chapter1.1, increasing human interaction with digital devices isextending discourse to contexts where we take unidirectional instructions from smartphones, intelligent personal assistants (Siri, Cortana); Computer Assisted language learning (CALL) devices-DuoLingo; Embodied Conversational Agents (ECAs) in healthcare or self-checkout systems in supermarkets. However, there is the need to understand how instructions gets packaged as information units sent from the instructor to the instructee during interaction.

The first aim here is to discuss the conception of instructions as information units. The second is to discuss how interlocutors use them in interaction. The Merriam-Webster online Dictionary says an instruction “is a statement that describes how to do something” and these have been variously classified. Types of task-related instructions include big instructions, spatial instructions and procedural or work instructions (Shriraman et al., 2010).

Big or compositeinstructions fuse many operations and encode relationships among large numbers of low level operations that enable users to efficiently and concurrently execute many tasks at once, such as used in programming (Zhao et al., 2014; Pappu &Rudnicky, 2012; Mourgues et al., 2012;Eirisdottir & Catrambone, 2011;Shriraman et al, 2010).

Spatial instructions give route directions that tell the listener where and how to move along a plane from one physical or virtual location to another (Zhao et al., 2014; Pappu &Rudnicky, 2012). Procedural instructions proposed by Eirisdottir and Catrambone, (2011) also called work instructions (Mourgues et al., 2012) are defined as, stepwisedescriptions of how to carry out a task that guide people by describing the condition for carrying out procedure, actions required and how the states of the system change because of these actions (Eirisdottir&Catrambone, 2011).

Assembly instructions used in this thesis are similar to procedural instructions but, they combine aspects of big and spatial instructions. Assembly instructions usually tell users the condition for actions, what action to take and expected consequence, and they are task oriented because they provide information in the form of procedures(Eirisdottir & Catrambone, 2011) or sequences of actions that move a listener from one level of action to a higher one.

There are two types of assembly instructionsdistinguishable on the basis of their orientation. Those that provide information in the form of procedures are task-oriented while those that provide information in the form of examples are instance oriented (Eirisdottir & Catrambone, 2011). The assembly instructions used in this research are task-oriented because they describe the procedures for assembling Lego models sequentially.

Word order and prosody are used to convey the information status of instructions (Breenet al., 2010) and their meaning development. Word order holds that given information generally precedes focused information. In this thesis, given information refers to old information in the discourse while, focused material refers to new information (Hilpert, 2014, 2013; Breen et al., 2010).Prosody is the study of the tune and rhythm of speech and how these features contribute to meaning and it comprises frequency, duration and loudness that give rise to phrasing, stress and intonation (Iyiola, 2010).

In addition, I have argued elsewhere that instructions are discourse actions used for carrying out tasks (Ofemile, 2015a) and to this end, instructions are Information Packaging Constructions (henceforthIPC) structured"…by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a context or discourse"(Vallduvi & Engdhal, 1996, p.3). This means that communicative and psychological articulation of information in instruction-giving contexts involves a process that uses word order to organise meanings, then relate new meanings to old ones to aid listener comprehension.

Going further, Common Ground (henceforth CG)descibes information that is mutually known, shared and continuously modified in communication (Krifka, 2008, p.243-245). Word order and CG enable interlocutors to build rapport and mutually develop meaning during interaction.However, it is important for instruction-giving process to recognise that CG could change with accommodation when interlocutors accept uncontroversial facts (Krifka, 2008) because instructions could be modified to meet listener's expectation asoutlined in Chapter 4.2.2.

Another concept that is linked to CG is the notion of focus in interaction. Focus "…indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions"(Krifka, 2008, p.12). There are cases of multiple foci, in one sentence where one expression introduces alternatives that are exploited in one way, and another expression introduces alternatives that are exploited in a different way as seen in the case of vague expressions. For example, the vague instruction"Pick the black thing"in a context where there are many black things, is subject to many interpretations of its structured meanings depending on the background and focus of interaction. This may be clarified if the interlocutors have established a CG regarding

“…the black thing”. In addition, the focus established by the CG influences how listeners deal with recoverable information during comprehension (Chapter 2.5).

At another level, foci are determined by the physical composition of instructions such as big or composite instructions described earlier. The amount of information to be processed might overwhelm normal human capacity (Cowan, 2001) because such information is held in the memory for an extended temporal duration, which may lead to errors in interpretation.This ties in with the view in dynamic semantics thatthe structure of instructions influences their meaning-making function during tasks.

Furthermore, instruction-giving and taking is viewed as “… an approach to meaning representation where pieces of text or discourse are viewed as instructions to update an existing context with new information, with an updated context as result”(Van Eijck & Visser 2012,p.1). This implies that instruction-giving is viewed as a composition of successive steps or stages of updating information and meaning during tasks-based interaction. This is linked to the listening comprehension process outlined in Chapter2.2.1 and 2.2.2because listeners rely on several sources of knowledge to interpret input or proffered content while constantly updating focused information with given information and making accommodation for reliable information before giving feedback.

As outlined earlier, prosody is used to convey information status in verbal instructions and Breenet al. (2010) suggest that differences in prosody such as accent placements lead to successful discrimination in attention paid and speaker perceptionamong listeners. For example, Clark et al., (2014) and Ofemile (2015a) separately agree that prosodic features of the agent’s voice made human listeners perceive verbal agent-instructors as unfriendly and pretentious during interaction.

Prosody has also been found to have strong links with comprehension. A study comparing students’ reading prosody in first and second grades with their reading comprehension at the end of third grade by Miller and Schwanenflugel (2008) found that, early acquisition of an adult-like intonation proficiency enabled better comprehension. Another longitudinal study of about 1,750 fourth graders found a strong correlation between prosody and their overall reading achievement (Daane et al., 2005).These agree with Breen et al.’s (2010) conclusion that, where there are no prosodic contrasts, for example, at the level of phonological articulation in speaker accents then poor acoustics occur. Poor acoustics may lead to poor listener discrimination of focus and comprehension of utterances because humans can discriminate languages by hearing the prosody of speech(Komatsu et al., 2004). These imply that, while poorlistener discrimination of prosody may lead to meaning difference between speakers and listeners, lack of prosodic contrast in speaker utterances may lead to lack of rapport with listeners at the level of meaning-building and attitude perception.

At another level, assembly instructions as IPCs enable interlocutors to organise and arrange meanings, relate new meanings to old meanings and facilitate the hearer's job of integrating new information into old. This is done using four instructional primitives (Vallduvi & Vilkuna, 1998; Engdhal & Vallduvi, 1996). Instructional primitives are frames used to assess the linguistic structure and functions of assembly instructions andthese include:

I. Instruction 'add information'- focus will make listeners to move from a lower level of knowledge-K to a higher level; K1 to K2 where K2 is a superset of K1.
II. Instructional Primitive Link 'Go to address X' this instruction frame tells listeners what to doand it is also used as a link to repeat ideas in previous instructions using pronouns, or other forms of anaphoric references.
III. Instructional Primitive Tail 'Go to entry under some given address' under a specific address. Tails are optional like Links
IV. Contrast with X alternatives, where X is an informal characterisation of focus.

Focus indicates the presence of "…alternatives that are relevant for the interpretation of linguistic expressions" (Krifka, 2007, p.20) but within the context of structuring instructions, the 'add information' frame enables the speaker to present given information before new information. This enables the listener to move from a lower stage of a task to a higher one.

In addition, the instructional link 'go to' frame is used by instructors to specify where and what to look for and do during tasks. This instructional frame is linked to the principle of discourse iconicity(Herring, 1990) which states that, information is placed close to the part of discourse to which it relates thus, instructions begin with the part already referred to and end with what is being talked about, for example,

Instruction 1: identify the red wing-like bit and slide it inside the blue square. The pronoun “it” in the second part of the instruction is within the ‘go to’ link frame referring to “the red wing-like bit”. This principle also holds for the instructional primitive go to a given address; however, the reference is specific so instead of saying “identify the red wing-like bit and slide it inside the blue square”, the instructor will say “pick the red wing-like bit and slide it sideways into the blue square”.
Going further, contrasting alternatives is related to information structure in each language as it concerns the notion of opposites within a set of alternative elements. Contrast in instruction construction is achieved by placing different alternatives before listeners within instructions.

As outlined earlier, assembly instructions have aspects of spatial instructions which include landmarks and actions (Hilpert, 2014; Fontaine & Denis, 1999).A landmark is any element in an environment that is external to the observer and serves to define the location of other objects or locations (Sorrows & Hirtle, 1999) during tasks. Landmarks in assembly instructions are represented by a milestone, point or stage in the process that define other processes and aids listener monitoring of instruction execution. A landmark is achieved in an assembly task when a listener completes Stages 1, 2 3, and 4(Chapter 3.2.1-assembly procedure). Action is built into instructions and denoted by verbal phrases like; lift it; join the, move the, or take.

Landmarks and actions used to encode messages to listeners enable researchers to assess how listeners cognitively navigate through the information maze during interaction and the way information is packaged in instructions affects the listener successful execution or otherwise of tasks. The review will now focus on listener’s use of instructionsas given by how users employ them in relevant interaction contexts.

Generally, instructions are beneficial for brainstorming, reducing error, facilitating performance, task completion (Eirisdottir & Catrambone, 2011; Paulus et al., 2011), and guiding users or listeners to use products or carry out a task when it cannot be communicated through the product or task design (Li et al., 2011). Furthermore, Erhel and Jamet’s (2013) study of the role of instructions in educational gamesfound that, deep learning is indeed compatible with instruction provided there is room for feedback. They also found that the choice of instructions may hinder learning however, negative impact can be cancelled by adequate feedback. In another research assessing the impact of instructions on usability tests, Zhao et. al. (2014) found that instructions led to dialogue between interlocutors and aided layout navigation but also created functionality problems for participants.

Following this, this study believes that there is a need for appropriate instructions that can aid listeners to focus on specific items of information or establish inferential links between task items and their schemata (Erhel & Jamet 2013). Therefore, this study uses verbal assembly instructions to elicit listener responses because Schneider et al. (2011) found that people retain more of verbal instructions and are able to transfer them to other contexts when the neeed arises.

Eirisdottir and Catrambone (2011) posit that procedural instructions are organised as successive steps needed to complete a task, for example, the procedural instructions needed to gain access into a secured building areas follows:

1) Swipe your identity card on the keyhole
2) Press the black bell then turn the door handle down
3) Step on the door mat and the door will open inwards for you to go inside.

These procedural instructions enable the user to access a building just as assembly instructions help listeners get tasks done quickly and efficiently when this cannot be communicated through the product or task design (Li et al., 2011). Moreover, appropriate instructions can aid learners' focus on specific items of information or establish inferential links between items of the task and their learned knowledge and experience (Erhel & Jamet, 2013).

Although instructions have been found to be beneficial, research indicates that pragmatic people are not motivated to read to instructions but would rather have a go or try things out (Kolb, 2014; Honey & Mumford, 1992) while others found that, people resist using instructions and only resort to them when they do not know what to do (Erhel & Jamet, 2013; Eirisdottir & Catrambone, 2012). These attitudes were seen with people using telephone systems (Szlichcinski, 1979), consumer products (Wright et al., 1982), household appliances (Schriver, 1997) and computer software (Carroll, 2002).

However, instructions have also been found to have a negative impact because Zhao et al., (2014) argue that explicit instructions could create problems for the user in dialogues, layout navigation and tool functionality during usability tests. McCrudden et al., (2010) employed a mixed method in a study assessing user comprehension and concluded that instructions hampered the achievement of reader goals, processing and comprehension of information.In addition, Erhel and Jamet (2013) found that the choice of instructions could hinder learning however, such negative impact may be cancelled by adequate feedback from interlocutors.

In HHI, either as encoders or decoders people often repeat instruction for various reasons that may include compliance with standard procedure (Stack, 2012; Schneider et al., 2011), enabling the listener to deal with information in manageable chunks (Field, 2007; White, 1998) and as a coherence strategy (Hansen & Jensen, 1994) during comprehension.Stack (2012) suggests that listeners repeatinstructions to acknowledge receipt of instruction, confirmthe purpose of the message, clear doubtsand understand the urgency of the message.

This review suggests that repetition is not only natural in human interaction,but repeating instructions may enhance interaction. The principles of information packaging reviewed in this chapter inform the design of instructions(Chapter 3) with a focus on how repetition of assembly instructions shape the interaction between instructors and instructees in HAI and HHI during assembly tasks in chapters 4 and 5.

2.4Culture and Emergent Identities in HACs

The third aspect of language use as a joint action relates to culture as a pervading influence on interaction, negotiating differences and the process of identification during interaction. This chapter outlines how identities emerge during interaction and the role of culture in creating interaction contexts as well as influencing interaction.

2.4.1 Emergent Identities in HACs

Identity has been a subject of concern for researchers, for example, G. H. Mead proposed the “I” and the “Me”; Sigmund Freud’s Id, ego and superego and Burke’s (2007) role identities, social identities and person identities have all attempted to understand the place of identity in human interaction. This may be related to the accepted view that one of the interpersonal functions of communication is the assertion of identity because this is what portrays us as individuals and members of a group (Fearon, 1999). For the purposes of this thesis:

> "Identity can been (sic) seen and defined as a property of the individual or as something that emerges through social interaction; it can be regarded as residing in the mind or in concrete social behaviour; it can be anchored to the individual or to the group" (DeFina, 2006, p. 265).

This definition is simplistic and infers that identity like personality comprises properties an individual has that emerge performatively from interaction with others.

Views differ on how identity is created thus, this thesis draws on influence from research around identity as developed by several researchers such as Farr & Riordan, (2015); Taylor, (2009); Hollway, (2009); Goffman (1959) and Garfinkel, (1956), as well as the linguistic frameworks developed by Bucholtz and Hall, (2005) and adapted by Clark (2016) to understand the relationship between linguistics and identity.While the ethnographic research provide a general framework, sociolinguistic research provide a linguistic approach used to analyse identity in HAI.

Language use has a strong link with identity because language is used to describe our own identity and those of others just as it enables us to identify someone as an individual that belongs to a given social group (Clark, 2016, p.25). For example, a speaker's voice and choice of words may shift their intention and consequently change the way a listener percieves them. These views are interwoven in this thesis to get a balanced view of the emergence and roles of identity during interaction.

One view holds that identity is created through the process of identification which is "the act, conscious or not of accurately imagining oneself in another's place" (Holloway, 2009, p.255). This refers to the feeling or sense of close emotional bond or association with something or someone to achieve a better understanding, for example, a teacher thinking like a learner or a writer thinking like the audience.

In addition, identity is created through a person's socialisation process from education, family upbringing within differing cultural and social rules (Farr & Riordan, 2015; Erikson, 1959). This implies that identity is developed as part of the general process of acquiring the culture of one's parents and immediate society in preparation for the roles society has for them.

Holloway's definition and the view that identity is a product of socialisation tie in with Bucholtz and Hall's (2005)ontological principles to an extent. The first being that identity is an emergent product rather than a pre-existing source of linguistic and other semiotic practice that is fundamentally social and cultural (p.588). They thus re-echo Goffman's (1959) position that identity is constructed performatively because in ordinary everyday situations, people behave in ways that tell others who they are, what they do and their life expectations. This view is further illustrated by Hartford, Carr and Pope's(2010) concept of performativity which is useful in understanding identitycreation and perception in online interactions because apart from showing what each interactant can do, it holds that our identities are created by what we do every day using our skills and capabilities. These skills and capabilities have been specified as linguistic and semiotic practices used by interlocutors as they take up temporary roles during interaction as expressed in Bucholtz and Hall's positionality principle (Bucholtz & Hall, 2005, p.591). Such assumed roles include speakers as instructors and listening instructees that have macro-level demographic categories in HHI.

Demographic categories in HHI include age, sex, social status, educational background, or marital status, this thesis is of the view that demographic categories in HAIrefer to the affordances of agents such as voice (Chapter 3) and as outlined below. Linguistic markers used by listeners generate discourse and to project their identity and perception of agent identity are nonverbal (facial actions and gestures) even though the interactional positions are fixed in this case.

In addition, just as in HHI, listener perception of instructor identity may be multi-layered because they may not conformwith the listener's expectation developed from previous interactions which may affect their perception of the emerging identities however, Bucholtz and Hall (2005) suggest that this gives a bigger picture of the emerging identities and by extension the listener's attitudes towards the interaction.

The three processes have been used to explain the emergence of multiple identities that may include but are not limited to personal, social and relational identities (Farr

& Riordan, 2015; Taylor, 2009; Hollway, 2009; Garfinkel, 1956)and my interpretation is illustrated in Figure 2.3below.

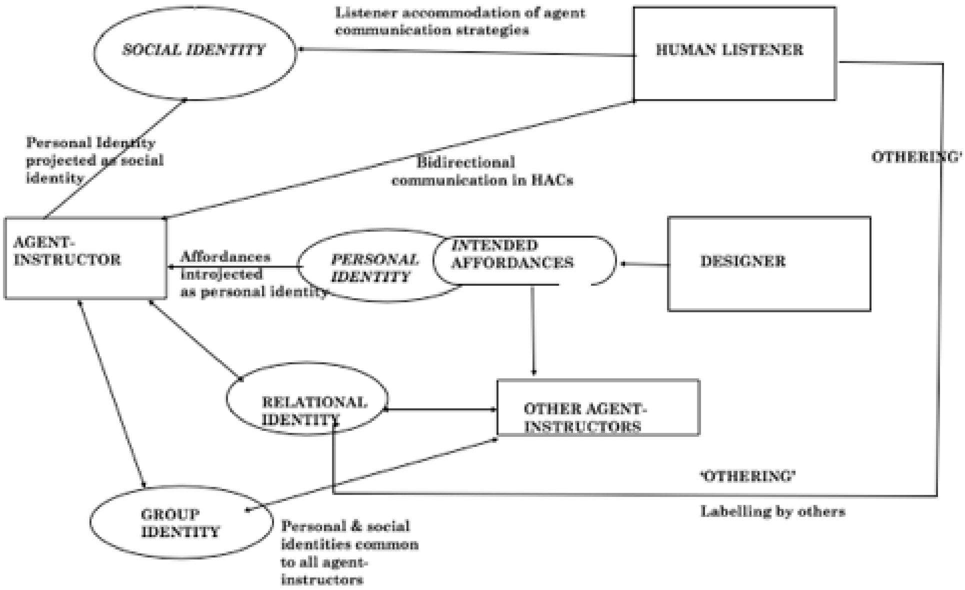


Figure2.3 Agent Identities Emerging from HAI

Among people, personal identity is a person's own idea or view of who one is ''the real me" (Taylor 2009, p.170) which emerges from a person's multiple social contexts, connections and interaction with other people and contexts. However, when it comes to agents, the designer, sometimes with users determines its personal identity (Figure 2.3). The agent's personal identity is linked to its affordances or proposedcapabilitiesand introjected by designers into agents;for example, Apple's Siri and Microsoft' Cortana offer verbal interaction across platforms (Jiang et al. 2015; Kiseleva et al., 2016). In addition, Siri's designer intended it to interact verbally with users thus, programmed it to recognise voices, understand terms that are required for it to fulfil the tasks it supports and give appropriate feedback. Thus, Siri's personal identity is linked to its verbal capabilities and this may make users to categorise agents as having a particular linguistic or vocal property. This is linked to the notion that agents conforming or not conforming to listener expectations may inform researchers on how they socially position agents during interaction.

The agent's personal identity is projected or performed to be received by othersin a specific setting as theenacted or social identity (Farr & Riordan, 2015; Garfinkel, 1956) (Figure 2.3). Thearrow linking the agent to social identity represents the agent's discursive role during interaction, such as an instructor in the assembly context. Research suggests that discursive roles and practices enables interlocutors to shape, reshape and make sense of their identities during interaction (Farr & Riordan, 2015).This implies that personal identities are not separatefrom social identities because the former provides the tools to develop the latter. For example, the agent's identity of being simulated and verbal (Chapter 4.2.1) enables it to carry out its role as instructor during interaction thus, the personal identity of being simulated and verbalis both a social identity and an important part of who the agent is(Taylor, 2009; Hollway, 2009; Zimmerman, 1998).This ties in with the notion of demographic categories outlined in Bucholtz and Hall's (2005) positionality principle which states that identity and macro categories contribute to emerging identities. In addition, the inseparability of interlocutors' personal identity from their social identity also ties into to the notion that identities emerge in relation to one another where categories are linked and not independent (Clark, 2016 p.27). However, this thesis is of the view that while in HHI, interlocutors can register their identities with co-interlocutors, in HAI it is not clear how this can be done particularly with the nonverbal responses from the listenersto agent instructions.

When two or more subjects share the same personal and social identities they are said to have a group identity(Figure 2.3)in which members get the same experience and are seen as being of the same genre within an imagined community(Taylor, 2009)..Furthermore, Bucholtz and Hall (2005) suggest that linguistic forms and structures are used to mark out personal and social identities of individuals and group members that are sufficiently similar for interactional purposes and can thus be classified as speech communities in the sense expressed by Gumperz (2009, p.66).For example, academicians are also seen as lecturers and researchers just assome agents could be thought of as instructors or personal assistants because they are seen as having the same identities.

When it comes to setting social boundaries in human interaction, people sometimes have labels conferred upon them by others called relational identity (Raban, 1991). It is necessary to understand how this plays out within the context of HACs because research suggests that some humans still think of agents as simply machines (Clark et al.,2016) to be used which sets them apart from humans as "them not us" (Taylor, 2009, p.178).

Relational identity is ascribed to a group by others through the process of 'othering' (Chauhan & Forster, 2014). Othering is a process of engaging with others we perceive as mildly or radically different from ourselves (Canales, 2000). The process engages us in some kind of space purification that serves to mark and name others as different from us (Tope et al., 2014, p.451). Chauhan and Forster (2014) opine that othering in human society is achieved through representational absence, representation of difference and representation of threat. Othering ties in to Bucholtz and Hall's (2005) relationality principle particularlyinstitutional and structural of identity formation process. They illustrate this using the link between 'authorisation and Illegitimisation (Bucholtz & Hall, 2005, p.598). Authorisation refers to the imposition of an identity through institutionalised power and ideology while illegitimization relates to the ways in which identities are dismissed or ignored by the same structures(Bucholtz & Hall, 2005, p.598). Such identities are reinforced by by sustained objectification of personalities and reducing themto a preconceived image (Chauhan & Forster, 2014) that is usually negative.

Relationality works in diverse ways through discourse that produces genres that are interactional and, in a way, connected to power relations in interaction.The view of this thesis is that, distinction enables interlocutors to identify insiders and outsiders in any interaction context based on accents, diction and other preconceptions.While this is possible in verbal interaction specifically speaker utterance, there is the question of how this comes out in nonverbal interaction however, it may be useful in classifying nonverbal listenership behaviour and deriving meanings for similar or different behaviours during interaction.

Human othering of agents in social interaction may have been sustained by the perception that they constitute a threat to an already existing system just as the poor are considered a threat by the affluent (Chauhan & Forster, 2014). For example, Brynjolfsson and McAfee's (2012)"Race Against the Machines" re-echoes John Maynard Keynes' (1930) "The Machines Are Coming: The Spectre is coming"and forcasts that technological advancement will create technological unemployment by encroaching into hitherto strictly human spheres (McAfee, 2011). These suggest that computers represent perceived threat and loss of privilege (Topeet al., 2014,p.452) that may have "precipitated the greatest othering" (Chauhan & Forster, 2014, p.400) of agents.

However, with increasing pervasive computing in human life, identity in HAI may be co-constructed via negotiation with others (Farr & Riordan, 2015) from a position of difference as is done in HHI. This agrees with the partialness principle (Bucholtz & Hall, 2005) which holds that

> *"Any given construction of identity may be in part deliberate and intentional, in part habitual and hence often less than fully conscious, in part an outcome of interactional negotiation and contestation, in part an outcome of others' perceptions and representations…" (Bucholtz & Hall, 2005, p.605)*

This principle suggests that identity can in part be repetitiously and deliberately constructed, become part of an interlocutor's perception of co-interlocutors without being stable. This adds the notion of interlocutor intention, consciousness and deliberateness in the formulation, assignment, projection and perception of identities while, instability refers to mobile identities that emerge from interlocutors interaction in different contexts.

While deliberateness in interaction is easily evident in HHI due to human nature, it is still a bit different for agents even as recent research in artificial intelligence indicate that new techniques exist that enable agents such as Facebook's chatbots, to successfully navigate negotiations in a human-like way (Lewis et al., 2017) and robots such as NASA's Pepper have emotional intelligence to sense and think about their environment, then act based on their perceptions during interaction almost like humans (BBC News, 2017). These agents are programmed by designers to act is specific ways including language use and interaction management –turn taking but they are yet to attend communicatively to nonverbal feedback.

The challenges for this thesis is to assess whether listeners will respond nonverbally to the agent instructor's use of face-saving communicative strategies in its diction and how listeners will make allowances for a non-human instructor in comparison to a human instructorduring interaction.These challenges, Bucholtz and Hall's (2005) framework as well as insights from Clark (2016, p.) provide some guidelines for designing the frameworkoutlined in chapter 3 and for assessing the implications for the study as a whole.

2.4.2 Culture and the Articulation of Differences in HACs

In the joint action between speakers and listeners, language is used to give and receive instructions as well as to describe their own identities and those of others from a foundation and context provided by culture. This chapter argues that just as people from different backgrounds have different cultures, agents are culturally distinct from humans so their interactional competences and approaches differ and as outlined in 1.1, Bhabha (1994) postulates that where two cultures meet, a new one is born within a zone of hybridity.

Bhabha articulates this theory from two descriptions namely, the genealogy of difference and the idea of translation.Difference refers to culture difference that comes from the awareness that such differences cannot be accommodated within any Universalist framework (Rutherford, 1990, p.209) as it is very difficult, even impossible and counterproductive to fit together different forms of cultures and believe that they can easily coexist (Bhabha, 1988) because recognisable variances exist between them. Allolio-Näcke, (2014) citing Hall (1996:4f) while discussing hybridity stresses that, it *is "*only through the relation to the other that the positive meaning of any term and its identity can be understood" (Allolio-Näcke, 2014, p.927). This implies that it is only by contrasting cultures that differences become apparent in any interaction.

Bhabha (1988) also uses the concept of translation to refer to the articulation of cultures during interaction. Bhabha in Rutherford (1990, p. 935) describes translation as way of imitating an original without reinforcing the priority of the original. This opens up the possibility of articulating different cultural practices "…because all cultures are symbol-forming and subject-constituting, interpellative practices" (Rutherford, 1990, p.210). Moreover, culture lives through some form of signification such as icons, symbols, myths and metaphors that enable members to conceptualise and understand their social experience. Often, signification (as in language use) is only representative of specific cultures thus making them become self-alienating in relation to other cultures. Furthermore, this thesis believes that translation is best expressed in language use during interaction that requires some level of competence as will be outlined later.

The articulation of differences in culture is further clarified using Johan Fornäs (2012) example as cited in Kaun and Fast, (2013, p.8) that distinguishes between ontological, anthropological, aesthetic and hermeneutic understandings of culture.

The ontological understanding is based on the distinction between nature and manmade culture (Kaun & Fast, 2014). Natural things are "Designoid"(Dawkins, 1986) meaning that they may appear designed, but they are not, and have come into being without an external designer or conscious design process. While manmade things are explicitly designed by an external designer for a purpose. This is a fundamental distinction between humans and agents as per origin. The interactional implication is inferred from Dautenhahn’s (2013) observation that, agents’ body gestures do not evolve over many years like humans and they do not have natural gestures, face or limbs. This suggests that agents may not use their bodies to communicate the way humans do.

The aesthetic understanding of culture refers mainly to human artefacts like literature, art, music, theatre and film and constitutes a specific sector in society (Kaun & Fast, 2014). This implies that humans have a ready culture that is operationalised by artefacts used in the socialisation process to imbibe and pass on cultural norms and values that differentiate one group of people from another. Furthermore, agents cannot be said to have a ready culture although, they come with certain physical qualities such as component parts but because these are not independently created like human artefacts, it can be argued that agents do not in the strict sense have an aesthetic culture like humans and this difference may affect their social norms of interaction (Hartley, 2002, p.87).

The hermeneutical understanding of culture as used in this thesis indicates that just as it is among people from different cultural backgrounds, agents also have different meaning generating and communication practices that are different from those of humans as outlined in 2.4.1. For example, the agent’s linguistic behaviour is realised within its internal structure pre-determined by the designer and in some cases externalised using five different languages for five different purposes (Ferber, 1999). These briefly put include L1; implementation languages used for programming agents, L2; communication languages used for interacting between agents such as data transmission and reciprocal requests; L3 languages for describing behaviour and the laws of the interaction environment; L4 languages for representing knowledge such as describing the world in which they operate and make predictions; L5 formalisation and specification languages used to specify and explain what is understood by the agent through interaction such as stating the conditions that have to be met when implementing a multi-agent system. In contrast, human can use the grammar of a single language to make ideational and intrapersonal meanings (Halliday & Matthiessen, 2009) in different interaction contexts depending on the user's competence.

Research indicate that humans differ from machines in competences (skills and capabilities) employed in various interaction contexts. Fitts (1951)developed the Men Are Better At vs. Machines Are Better At list (MABA-MABA) to assess roles in human interaction with machines (Fitts, 1951, p.10). Later, Laundry (2009) developed the Humans Are Better At vs. Machines Are Better At (HABA-MABA) list to assess HCI. Recently, Jennings et al. (2014, 2012) have also outlined the expected capabilities of smart agents when interacting with humans in HACs.

To this end, Fitt (1951), Laundry(2009) and Jennings et al.(2012) separately suggest that humans are capable of sensory functions like the detection of stimuli, perception, flexible and ingenuous innovation, inductive and deductive reasoning, and judgement, In addition, humans possess full autonomy and will power, better agile teaming abilities that aids the achievement of individual and group goals, capacity for incentive engineering (– agents are created by humans in the first place), and are better at developing auditing standards and mechanisms to manage and regulate crowd generated content. Agents on the other hand are better at speedy and powerful responses to control systems, repetitive work, reasoning deductively, agile teaming focusing on specific goals, and gathering crowd generated content.

Although such heuristic role allocation has been criticised for attempting to create points of substituting humans with agents in economic centres, they provide essential ethnographic information that will help researchers and designers to understand the nuances of human interaction with automated systems (Cummings, 2014, p.62-66). In addition, heuristic role allocation enables researchers to assess how humans develop experiences that form a culture that is distinct from agent culture just by what they do every day. However, it must be recognised that agent behaviour, culture and identity within HACs are emergent (Spiliotopoulos & Carey, 2005) because they manifest from interaction with others.

These differences can be summarised as indicating dissimilarity in patterns of thought, behaviour, creation of artefact and level of transformation done in an environment. They do not only prepare the ground for intercultural communication in HHI and HAI but, they provide the ingredients for evolving the third space of hybridity or inbetweeness as expressed by Bhabha (1994) outlined in 2.4.2.1. In addition, the notion of negotiating differences from the background of cultures will inform the design of the multimodal research framework in Chapter 3 used to assess nonverbal behaviour in Chapters 4 and 5.

2.4.2.1 Interacting in the third space of hybridity in HACs

This sub-chapter will outline the features of the third space as a zone for intercultural communication in HAI from a background of HHI. Following the analysis of differences in chapter 2.4.2 above, this thesis extends Bhabha's (1994) theory to the contexts of HACs and believes that humans and agents do not have shared values and norms as well as rituals and these define them as having different interactional cultures and by extension interactional competences.

As outlined in 1.1 and illustrated in 2.4.2, when humans interact with agents during collaborative tasks, a third space of hybridity is created (Figure 2.4) below.

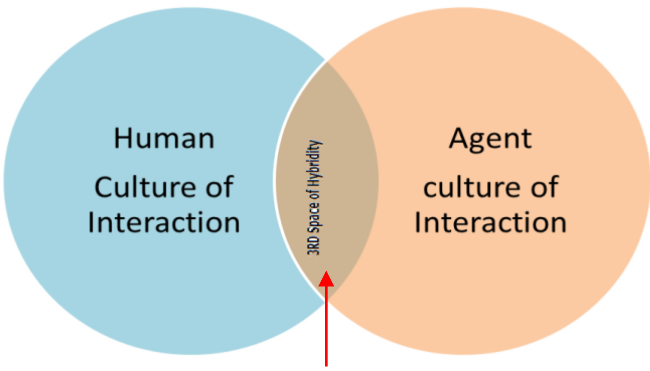


Figure2.4: The 3rd Space of Hybridity in HAI

This area of inbetweeness is created at two interwoven levels. The first level is as a physical or virtual interaction space- is created by the mere presence of interlocutors while the second is at the level of interactional approaches.This is because this space must exist as a "precondition for the articulation of cultural difference" (Bhabha, 1996, p.209) during interaction as interlocutors negotiate meanings, views and identities from the stand point of differences with co-interlocutors.In addition, these differences may influence how interlocutors articulate symbol-forming and subject constituting interactive practices as well as identification during interaction across different cultures (Allwood, 1985).

One feature of the 3rd space is that there is an overlap of characteristics between human field and agent fields creating an ‘inbetweeness ’or overlap. Bhabha’s (1994) area of concern was colonization in which native people found themselves caught between their own traditional culture and the new culture of the colonisers. Their continuous negotiation and creation of their identities led to the creation of a new third or hybrid culture. Thus, Bhabha (1994) argues that this space contains an ever-changing and unpredictable combination or aggregation of attributes from each of the cultural fields. For example, HACs isa combination of agents and humans working together to achieve one goal. The team composition is neither exclusively human or agents rather it combines both where members have their individual competences stirred togethermaking the hybrid team distinct from exclusive teams.

In addition, the 3rd space provides a context where interactants question hitherto held ideas then negotiate and deconstruct the series of inclusions and exclusions on which a dominant culture is premised. The mere entry of formerly excluded subjects and ideas into mainstream discourse further deconstructs and opens up the hybrid space. This openness and presence of many distinct stakeholders suggests that participation is motivated by a broad range of incentives (Jennings, et al., 2014).

However, stakeholders may also require reformulation, translation, rethinking and extension of old principles to reflect the new model, paradigm or alliance (Jennings, 2012; Rutherford, 1990) during interaction. For example, humans have always used a one-way form to communicate with passive machines but HACs is providing a two-way communication channel with intelligent machines that will be able to give and take orders using multiple channels and interfaces.

Regarding communication during interaction, our speech is often accompanied by facial actions, gestures and other body movements but as outlined earlier (2.4.2) agents do not have regenerative forms of nonverbal language. It is the view of this thesis that, understanding how human facial actions and gestures occur may provide an insight on how agents could attend to them communicatively during interaction.

Furthermore, there is (Re-) negotiation of identities of self and others to conform to the new culture and context. Allolio-Näcke (2014) citing Hall (1996:4f) stresses that it is only through the relation to the other that the positive meaning of any term and its identity can be understood (Allolio-Näcke, 2014, p.927). Thus, when stakeholders project their identity, they also perceive other's identity from interaction this may shatter hitherto held stereotypes as the idea of pure identity or pure culture no longer holds. While negotiation is possible in HHI, it becomes a challenge in HAI in view of the types of agents currently existing however, self-negotiation by humans as well as implied negotiation with the agent may be inferred from listener interpretation of agent instruction, management of turn-taking and execution of instruction during interaction.

When meaning is co-constructed within the third space, stakeholders negotiate from the background of difference such that the new mutually developed meaning becomes successfully translated across that bar of difference and separation between(Rutherford, 1990) participants. However, in HAI agent vocabulary and phraseology is pre-determined and not dynamic compared to human language use and this may influence listener comprehension and meaning-making during interaction. This thesis focuses on listener facial actions and gestures which present agents cannot attend to communicatively, however the listener has to achieve the same level of meaning with the instructor to successfully execute instructions. The joint action may thus be limited.

One major impact is that team members mutually develop a working culture that enables them to collaborate effectively and achieve set goals. Jennings, et al., (2014) suggest that this may require a definable information infrastructure that allows the authenticity and accuracy of seamlessly blended human and agent decisions, data, and content to be confirmed and verified. This infrastructure is made effective by systematic packaging of information to serve their communication needs (Hilpert, 2013) such as, labelling or naming things, giving commands or describing things. This in turn enables the development of a working language based on mutually agreed assumptions, collective actions, relationships and dynamics within the hybrid space such as the turn-taking management outlined in chapter 3.

Muller (2009; 2003) posits that homogeneity becomes the norm within the hybrid spacebecause stakeholders have with time developed shared authority, accentuated interpretation and collectivism in the process of group formation and interaction which according to Bhabha (1996) is the locus of all cultural interaction. While this is quite obtainable in HHI, the case of HAI is a bit different because present agents do not have the same level of perception with humans. This thesis agrees with Fowles (2000) who believes that interaction within the 3rd space enables participants to progress from the symmetry of ignorance to symmetry of knowledge and transformation because of exposure to mutually developed, accurate and audited ideas during interaction. The third space epitomises the multicultural interaction with multiple smart devices that HACs envisages will become the norm with time thus, understanding the unique dynamics of the space will inform the design of the multimodal research tools in used to assess instructor-instructee interaction Chapters 4 and 5.

2.5 Summary

This Chapter has outlined research in listening and listenership using spontaneous facial actions and gestures as lenses for assessing how listenership behaviours signal comprehension or lack of it when taking instructions during assembly tasks. It also describes the construction and role of instructions as information packages passed from instructor to instructee during interaction.It also outlined how culture distinguishes one group from another which forms a basis for negotiating differences during interaction. Furthermore, the notion of language use as a joint action provides an overarching frame work for this review while aspects of this interaction are discussed as speaker action, listener action and the role of culture and identification provide the bedrock for designing a multimodal research methodology in Chapter 3, agent deployment in Chapter 4 and the assessment of listener nonverbal listenership behaviours in Chapters 4 and 5 as bases for comparing HAI with HHI. This will combine existing measures with novel operationalisation that will enable systematic and reliable data gathering and analysis as will be outlined in Chapters 3, 4 and 5.

Chapter 3 Multimodal Research Framework Design

3.1 Introduction

This study usesspontaneous listener facial actions and gestures to understand how nonverbal listenership behaviour is projected when people take instructions during assembly tasks. To this end, this Chapter sets out the study's research procedure in two main chapters. The first chapter traces the multimodal research procedure used to organise experiments and develop the multimodal corpus. The second chapteroutlines the annotation scheme developed and used to make the emerging corpus readable.

3.2 Multimodal Procedure

Basic corpus research procedure comprises the following stages recording, transcribing, coding and mark up, management and analysis (Adolphs & Carter, 2013) and this is a useful general guideline for developing the framework. Sindoni (2013) outlines a six-step multimodal procedure focused on online interactions, which are multimodal as follows; using purposely built multimodal corpora, selecting different semiotic resources, analysing multimodal corpus using specific software, annotating multivariate language use and variation using problem-oriented approach, categorisation and discussion of the most frequent lexical bundles considered as indicators of register variation, and separate but combinable analysis of various distinct semiotic resources like verbal language, images, videos and others. This study combines both approachesin conceptualising the research framework for building and analysingthe multimodal corpus emerging from the interactions in Chapter 4 and 5 as outlined below.

3.2.1 Creating the Multimodal Corpus

The multimodal corpuswas developed from video recordings of interactions between instructors and participants. In the first study, a simulated agent gives assembly instructions to human participants (HAI) while, in the second study, the interaction is between a human instructor and human participants (HHI) assembling Lego models. Creating the corpus involves setting design focus, eliciting circumstances and recording interactions.

Step 1 Setting the CorpusObjective

O'Keeffeet al., (2007) recommend that a corpus must have a design rationale or principlethat enables thecorpus designer to decide what it should cover in addition to what is legal and ethical. In view of these, the corpus aimed to record participants receiving instructions from simulated agents and humans to assembletwo Lego modelsuntil either all the instructions have been taken and executed or within a maximum of 15-minutes depending on which comes first. Thiscreates a basis for comparison (Adolphs, 2008)and gives a sense of direction to the study.

Step 2: Setting the Eliciting Context

The second stage in the creation of the corpus is setting the eliciting context. The eliciting context (Figure 3.1 below) is the physical and/or virtual interaction situation and environment (West & Turner, 2013) set aside for interaction between instructors and instructees.

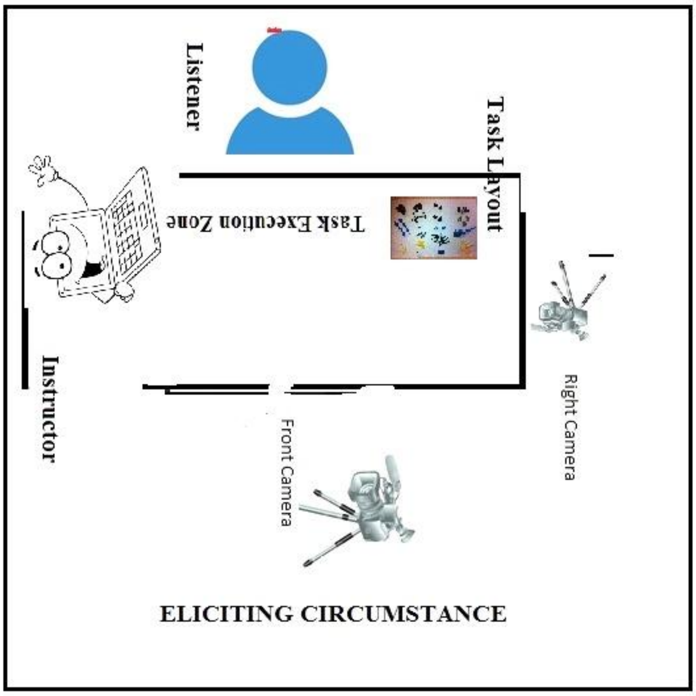


Figure3.1 Layout of Eliciting Circumstance

The eliciting context is an ecosocial system (Lemke, 1997) consisting of humans, agents, their things, their behaviour within the system, and the overall system dynamics, which depends on the principles that govern the flow of interaction and communication and what the flow means to interlocutors. The eliciting contextis used in this thesis to capture and describe the types of interaction that occur between instructors and instructees by linking interactions to the elements of which they are composed. Elements refer to everything in the physical location of interaction, such as the participants, Lego task layout, camera positions and task execution zone (Figure 3.1).To infer judgement and accurately predict facial actions and gestures displayed, this study uses the side and full views of a participant's face as outlined in Chapters4.3 and 5.3.

Step 2a. Voice continuum

This study uses a voice continuum or clineof instruction-giving voices as stimulus to elicit spontaneous listenership behaviours from participants. This is unlikeEkman and Friesen (2003,1969)whouse descriptive contexts to elicit posed facial actions; for example, asking participants how they will react if they saw their friends and they responded with smiles.

A cline of voices is used because research indicates that acoustic phonetic features determine how well a listener decodes input from speakers during interaction (Babel et al.,2014). Thus, the cline provides this study with the opportunity to assess listener comprehension of instructions as well as listener attitudes towards different voiced instructors during interaction.

The voice continuum is used to describe and characterise voices and not to produce a ranking order that labels one voice as better than the other. The range of voices used in instruction-giving include Cepstral Lawrence (henceforth CL), style 2- CereProc (henceforth CP), human recording (henceforth, HR) and real human voice (henceforth, HV) shown (Figure 3.3)below.

The heterogenous voices used by instructorsbelong to three broad ranges of voice progression in the continuum namely, synthesised, human-like and the target voice(Figure 3.2). The simulated agent has a range of voices within the voice continuum (Chapter 4.3.1.) while the human instructor uses a natural voice (Chapter 5.2.1.

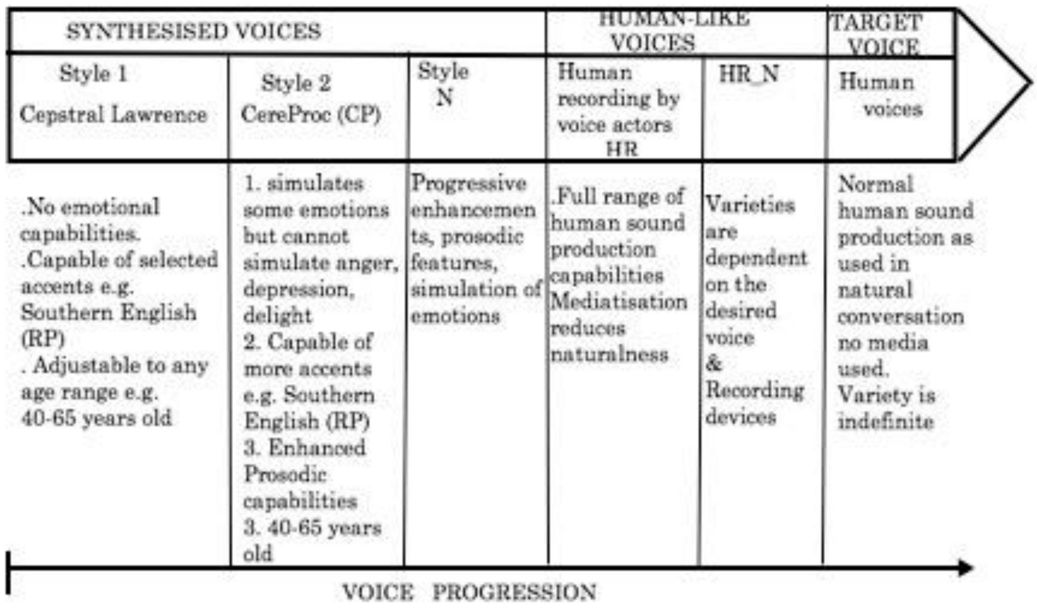


Figure3.2 The voice continuum

Within the continuum, there is the progressive organisation of voice form-function relationships whereby certain features remain constant, like age and language variety,while emotive and prosodic capabilities are progressively manipulated to become more human-like in agents.All the voices are in a Southern English accent Received Pronunciation (RP) because it has few regional overtones and is more widely understood than other varieties (Crystal, 2002). However as outlined in (Figure 3.3)above, each voice possesses a unique combination of stress and intonation to determine the quality of the speaker’s voice as a sound wave perceived by the listener (Iyiola, 2010).

Step2b. Assembly Instruction Design and Task Procedure

Interaction in this study is designed around assembly tasks. Assembly task was used in this study because it demands that participants cognitively process instructions as information packages as well as engages them in the physical or kinesthetic interpretation of instructions as assembly gestures. In addition, Lego assembly task is relatively safer for participants in comparison to tasks such as cake making or following driving or walking instructions on a highway.

The manufacturer's pictorial booklet contains visual building instructions[1] that guide users through the Lego model[2] assembly process. Research suggests that visual instructions have been found to enhance information carrying capacity and user comprehension (Kress & van Leeuwen, 2006). However, in their visual modes Lego instructions are not suitable for the purposes of this study because the interaction requires vague verbal instructions.

The manufacturer's visual instructions are transformed into written instructions using a process comprising preparation, lexisification and Assembly instructions as final product outlined (Figure3.3) below.

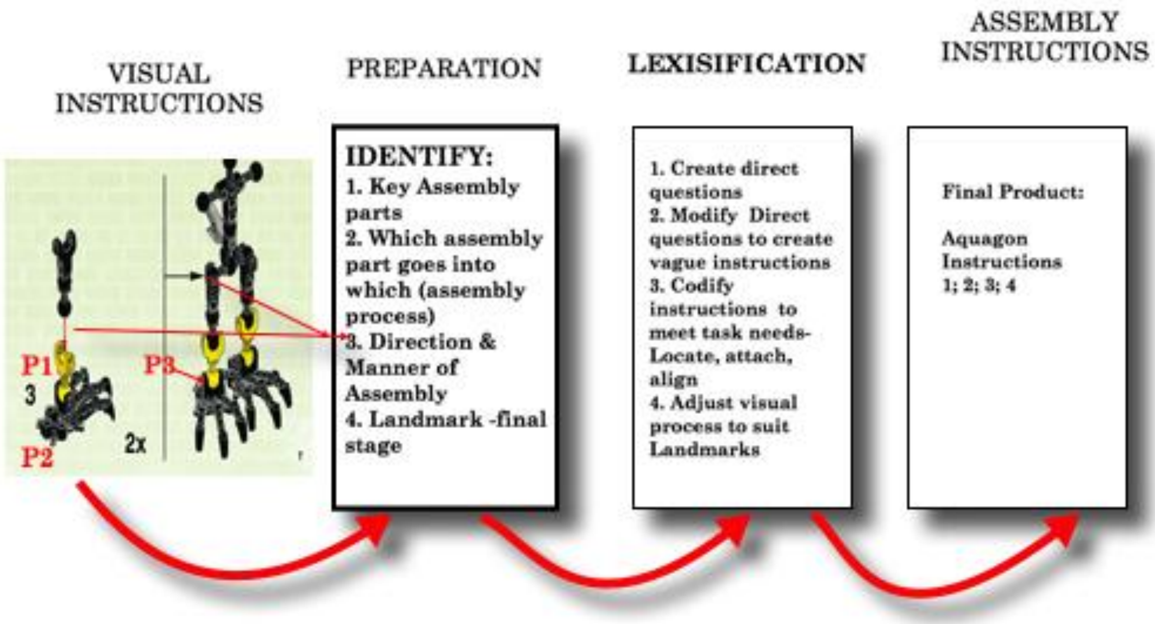


Figure3.3: Transforming visual Assembly Instructions to verbalised Instructions

.

Preparation aims to enable the researcher identify words that describe assembly parts, processes and associative actions that are critical to a successful assembly task. Two

---

[1]Aquagon https://www.lego.com/en-us/service/buildinginstructions/search?ignorereferer=true#?search&text=44013

Nex https://www.lego.com/en-us/service/buildinginstructions/search?ignorereferer=true#?search&text=6221

[2]*Aquagon* and *Nex* models have been retired. *Aquagon* has 41 pieces and *Nex* 39 pieces

experimental sessions were held with two participants who instructed the researcher on how to assemble the Lego kits. The participants were given only the manufacturer's pictorial manual and were asked to verbalise the visual instructions. This provided the opportunity to compare how people spontaneously transformed visual information as vague instructions. Although the sample size is small but, some patterns of vague language use emerged from the interaction. Such vague language use patterns include discourse markers (*so, now*) were commonly used at the start of new stages of location or attachment in the task process. Hedges and minimisers such as "twist it a bit" and "just sort of attach it" were frequently used with alignment and attachment stages. In addition, vague nouns specifically "thing" was used for describing various assembly pieces, with frequent reference to their quantity, size, shape and colour.

In addition, research indicates that other vague language categories exist. Cutting, 2007, p.5) outlines Channell's (1994) vague language categories such as vague additives (*around ten*); vague implicature and vague quantifiers (*15,000 died*); vague placeholders (*thingy, whatsisname*); tags (*or something, and so on*). Although Channell's categorisation has been criticised for inconsistency and lack of clarity in some respects, these do not affect the usefulness of her work in guiding this study to set out the interaction tool i.e. assembly instructions.

Furthermore, Clark (2016, p.55) outlines other classes of such as approximators subdivided into rounders (– estimators of measurements) and adaptors (-creates imprecision by reduction of class membership) illustrated as follows- rounders (approximately two kilometres; about there, around quarter to nine) adaptors (somewhat, sort of, kind of, a little bit, a smidgen).

Based on the guidance provided by the experiments and research on the vague categories relevant to instruction-giving, the researcher identifies all assembly parts, for example, P1 is "yellow piece with ball joints" while P2 is "black foot with claws"; P3 "small black piece that is Y-shaped" (Figure 1). In addition, additional actions required for the task to be successful are identified, for example, a participant needs to select assembly pieces before attaching them, hence words such as *Locate, take, pick,* are used to build instructions in this regard.

The next step is to identify the assembly progression specifying which assembly part attaches with the other sequentially. For example, P3 is first inserted into P2 then P1 is attached to P3. This step leads into the next identifying the direction and manner of attaching assembly parts. The arrows in the manufacturer's visual instruction provides directional and orientation information. Visual mapping of direction and orientation leads to landmarking the assembly process as complete or not.

Having identified assembly pieces, progression and associative actions, the next stage of transformation is lexisification. Lexisification is derived from the word 'lexis' refers to the selection and use of words and phrases to replace visual/pictorial texts in such a way that the resulting sentences effectively describe the original message to an approximate level of meaningfulness. (outlined in Figure 3.4 below)

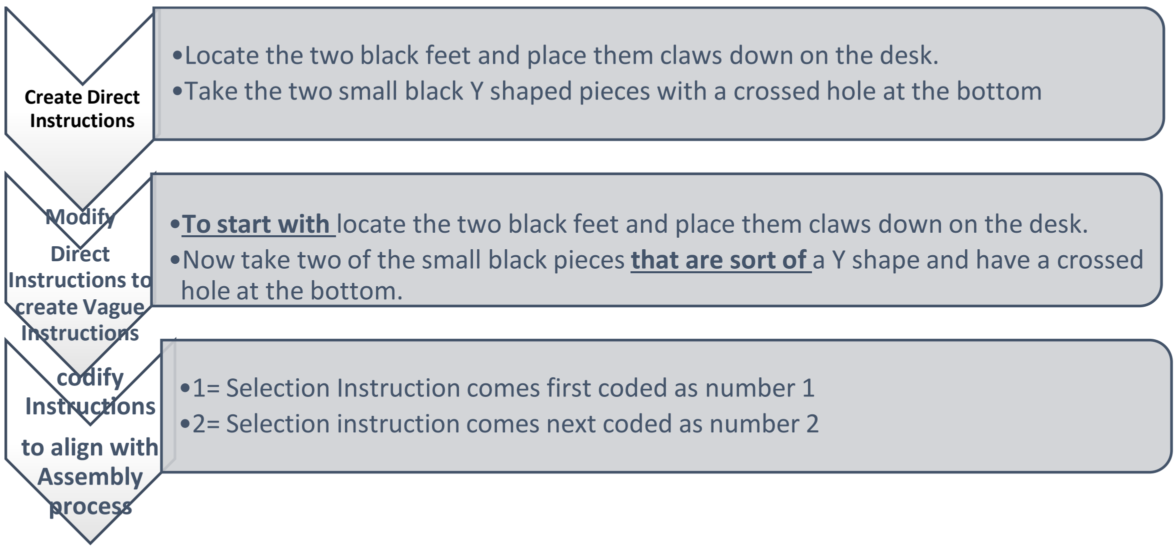


Figure3.4 Lexisification process

The first step is to create direct instructions as shown in Figure3.4 above. The direct instructions are modified to create vague instructions by inserting vague discourse markers, minimisers, hedges and references to nouns as appropriate (bold and underlined in Figure3.4).

Codifying instructions occurs at two levels. The first specifies what each instruction hope to achieve in the assembly process (locate, attach, align). And the second is related to task progression i.e. which comes first in the interaction as deduced from the visual instruction for every stage. The examples (Figure 2) above are used for selecting assembly pieces thus, they fall within the general class of 'Locate'. The instruction *"To start with locate the two black feet and place them claws down on the desk*" is coded as step 1 because it has to take place before step 2- *"Now take two of the small black pieces that are sort of a Y shape and have a crossed hole at the bottom".*

The final stage in lexisification is to adjust the overall visual process and progression to suit research-based landmark. As outlined in chapter 2, a landmark is any element in an environment that is external to the observer and servers to define the location of other objects or locations. In this case landmarks define when an assembly stage is completed, and another begins. As earlier outlined, Aquagon model has 41 pieces and *Nex* 39 pieces and the assembly process begins from the feet upwards to weaponising the models. This is remarkable shift from the manufacturer's approach. For example, the illustration in Figure 3.4 is in stage 7 of the manufacturer's visual instructions but these have been moved to stage 1 in the study.

The end products are vague instructions that enable participants to identify and select the appropriate assembly pieces; attach the selected piece to another in the model and position the pieces in a 3-D space, for example rotating it 180^0.

In order to reduce tediousness in task, some steps in the instructions contained multiple stages of the assembly process. A shorter step containing one stage in the assembly is shown in the Aquagon instruction below

> *Step 2: Now take two of the small black pieces that are sort of a Y shape and have a crossed hole at the bottom.*

Step 2 involves the location process only while providing description of the piece (size, colour, shape, and condition). Other steps such as the example from Aquagon instructions below contains one instance of location stage and one of orientation.

> *Stage 1: To start with locate the two black feet and place them claws down on the desk.*

The first relates to picking up the assembly piece while the second given by "place" relates to positioning the pieces in a given location. These examples indicate that the instructions were varied and included all the stages of the assembly process. In addition, this structure reduced the predictability that instructors may have projected.

The instructions were designed to describe assembly pieces and processes in simple terms to aid listener comprehension. This was done using familiar words and phrases from every day use such as claws; ball joint; socket; spikes; cylinder; arms; orange piece to mention but a few, with a view to focusing on listener feedback during interaction.

Furthermore, research indicates that instructions are useful when they aid the user to navigate through the information maze and aid listener task execution thus, should contain actions (Hilpert, 2013; Fontaine& Denis, 1999), landmarks (Sorrows & Hirtle, 1999), points, positions and milestones.However, in this thesis, a landmark in procedural instructions, is represented by milestone, point or stage in assembly instructions and milestone is an assembly stage that defines other assembly stages, for example, moving from assembly stage one to stage two.While Action toolkits will be denoted by verbal phrases including, *lift it, join the, move the, insert; break the; take*.

This thesis adapts Fontain and Denis' (1999) four protocols classes (underlined)used to guide spatial directions to illustrate useful information-carrying capacity of assembly instructions (illustrations are in bold letters).

Class 1: <u>Actions without any reference</u> to any stage, point, position of execution.
*To start with**locate** the two black feet and place them claws down on the desk*

Class 2: Actions associated with stages, <u>point, position</u> and milestone in execution.
*Just **put** these in the **rear gap of each** foot and attach them to the **cross-shaped connector***

Class 3: Non-procedural aspects of the stage, point or position executed.
The model is complete

Class 4: Stage, point and position mentioned with reference to other stage, part, point and position. For example:

Now *to make the arms just take two grey pieces with ball joints,* *two yellow pieces with ball joints and the two small black pieces that look like fists.*

These instructions (Appendices V-VII)did not only build on existing information but they hierarchically guide listeners through assembly stages and were built into the multimodal framework designed to assess the interaction and implemented in Chapters 4 and 5.

Furthermore, Hierarchical Task Analysis (henceforth HTA) encompasses principles and ideas through which one can find more efficient ways of carrying out tasks thus, HTA was used to match instructions to assembly stages as outlined(Table 3.1).HTA is a " … systematic method of describing how work is organised in order to meet the overall objective of the job" (Embrey et al., 1994, p.162).It describes tasks in a hierarchical structure of main goals, sub-goals, operations and plans (Stanton, 2006; Lane et al.,2006), and has been used, modified and applied in measuring medication administration errors(Lane et al., 2006), identification of error variance in system performance, assessing personal communication among systemsand assessing expert language learner procedures in self-management (Annette,2008; Stanton, 2006).

Following these examples, this study is using the match between instructions and assembly stages to develop a hierarchy of task goals and processes(Table3.1) below.

Table3.1 Task Procedure

Major Task Goal	**Sub-Goals**	**Assembly Stages**		**Instructions per Models**	
		Stage	**Description**	**Aquagon**	**Nex**
Assembling two Lego models using verbal instructions from an intructor	1. Decoding verbal instructions 2. Assembling Lego Models in 15 minutes	*Assembly Cycle times*		*Kit constructional units*	
		1	Legs and Lower body	1 -11	1 – 10
		2	Middle body	12 - 30	11 - 18
		3	Upper body and head	21 - 38	19 - 26
		4	Weaponising and finishing	39 – 48	27 - 49

The major goal of interaction from the listener's view point is to assemble two Lego models. The task goal is sub-divided into taking instructions and assembling Lego models. The first sub-goal is linked to listeners'comprehension or incomprehension of theinstructor's utterance andassociated spontaneous as facial actions and gestures displayed. The second sub-goal is the listener's practical execution of instructionswhich generates assembly gestures.

The process of assembling Lego modelsinvolves four major stages listed (Table 3.1) above. Eleven instructions enable listeners to assemble the legs and lower body of the Aquagon model while ten assemble Nex model. The rest of the stages also have sets of instructions that enable the listener to execute themsuch as making the mid-body;

making the upper body and head; and weaponising the model. Each execution of instruction by the listener within an assembly cycle time is considered a kit constructional unit. Assembly cycle time in this study refers to the time from start of an assembly stage until the stage is finished.Each major stagecomprises many Kit constructional units that correspond with the relevant number of instructions;for example, assembling the legs and lower body has eleven kit constructional stages for Aquagon and ten for Nex. In this way, HTA becomes an economical method of describing tasksand enabling the study to focus on aspects of the interaction that have impact on assembly tasks(Annette, 2008; Stanton, 2006, 2004; Embrey, 2000).

Step 3 Recording

The laststep in raw corpus creation isvideo recording interactionsto produce purposely built multimodal corpora (Adolphs & Carter, 2013; Allenwood, 2008) from different groups of data sets and access to specific communicative contexts such as assembly tasks. The cameras are placed on tripods for stability, tuned to get a clear focus then the shooting begins. For synchronization of the timings of the two cameras, I use a hand clapto mark when interactions beginin preparation for aligning the videos later. A clear record is obtained of crucial components observed for measurements using two digital video cameras like Knight (2009) and stored in Raw video folderhowever, the cameras in this study focus on one participant from the front and the left hand side (Figure 3.1).Although, it was difficult to obtain the same quality of data from the two cameras because of differences in model specifications, the raw videos are saved in MP4 format to save space and compromise less on the quality of videos recorded.Going further, QuickTime Player 7[3] is used to align the raw videos of each interaction from the two views using the clap as a clue to marking the timings. The cameras record the listener's sequences of body movements from different positions, while the QuickTime allows for the synchronised videos to be used for the examination of coordinated movement i.e. across each view (Figure 3.1).

3.2.2Sampling Procedure

Different sampling techniques were used in the selection of participants, task assignment and other sampling units in this thesis as outlined below.

[3]https://support.apple.com/kb/dl923?locale=en_US

3.2.2a Population Sampling

Purposive sampling was used to select speakers of English as a first language (Crystal, 2010) as the target population because research indicates that such people naturally respond to vague language (Channell, 1994). Speakers of English as a second language or foreign language were not selected as this may introduce other dynamics into the study that may bebetter dealt with in future studies.

Participants were recruited for eachexperiment via email from the university to students, personal contact, and online recruitment site[4]. Thereafter, participants self-selected themselves by replying to emails, and accepting online invitations. Those who used the recruitment site, must indicate that they are age 18+, are L1 Speakers of English and must be able to attend sessions in person before proceeding while the same conditions were specified in the emails. Participants are then directed to doodle[5]to decide on the best time and date for them to attend sessions. Self-selection allows participants to participate in the study on their own accord and has been assessed to be an effective strategy in experimental settings (Aitamurto et al., 2014) such as the assembly task. However, self-selection sampling has high mortality rates (Dörnyei , 2007)or participant failure to turn up for sessions which was responsible for the low participant turnout in the second experiment (Chapter 5).

3.2.2b. Task and corpus sampling

The task assignment was randomised to make sampling systematic (Dörnyei, 2007) usingRandom.org from www.random.orgto allocate slots, voices, tasks and timings to participants because it is straight forward, simple and eliminates clustered selection(Bryman & Bell, 2007; Altman, 1974). This is an external control mechanism designed to make population and corpora balanced without researcher bias.

The corpus contains different semiotic resources from the videos that cut across gestures, facial actions and media in every recorded interaction which make the corpus homogenous and useful for quantitative and qualitative analysis (Jewitt,

[4]https://www.callforparticipants.com/

[5]https://doodle.com/en_GB/

2014,2013, 2012;Lemke, 1990). However,there are concernsabout whether the multimodal corpus created adequately represents the language or means of communication used in each context (Adolphs, 2008) such as assembly tasks. To this end,the focus of the corpus covers a selection of nonverbal listenership behaviours specifically facial actions and assembly gestures because the study tries to understand them as a system of signs and symbols and how a listener uses them to communicate during interaction. Thus, representativeness in terms of nonverbal listenership behaviours as interaction modes in the corpusaccording to Adolphs (2008),sets a categorisation standard that will also be explicitly used when reporting the results of this study.

Stratified purposive sampling as explained in Onwuegbuzie and Leech, (2007) requires that a sample frame be divided into strata first then, a purposeful sample is selected from each sample stratum.Furthermore, Biber (1993) observed, in corpus design "stratified samples are almost always more representative than non-stratified samples"(Biber, 1993,p. 244).

Following this, the sample frame selected in this study is shown in specific occurrences when participants requested for instructions to be repeated during tasks. These critical points were selected as the general sample frames to assess interaction because as Clark et al. (2014) suggest, when analysing multimodal corpora in instruction-giving contexts, random selection of individual instructions or specific time marking are not used because there is no assurance that either could be adequately covered during interaction.

This general sample frame of repeats is divided into strata. The first stratum is selected using a minimum benchmark or cut-off mark of number of times instruction repetition occurs within each assembly stage which was purposively set to be at least fourteen (14) times in the first study and six (6) times in the second study. This significantly reduced the effects of overwhelming data (Jewitt, 2014, 2012).The second stratum focused on participants that did not repeat any of benchmarked instructions thus, the instructions most repeated by these participants during the tasks were purposively selected. The third stratum focused on task stages (see Table 3.41) and assembly process and samples were purposively selected from repeated instructions demonstrating the organisation of the assembly task.This level of systematic stratified purposive sampling ensured that every participant was included just as every instruction repeated was analysed.However, specific approaches for each are outlined in 4.3.6 and 5.2.3 respectively.

3.2.3 Study Measures

Before each of the studies took place, ethics approval was applied for and granted by the university and the experiment conditions specified in the approval were communicated to participants vide consent forms (appendix IIa and IIb) as outlined in 4.3.3 and 5.2.3.

In study one, the agent uses two voices in two in separate tasks from the continuum to give instructions to one participant. In the second study, the human instructor gives one set of instructions per task. Instructions perform the dual functions of testing participants' comprehension and Lego assembly skills. The study measures listener repetition of instructions and tries to understand how, why and when they occur.

Secondly, the study measures user nonverbal behaviours that are observable and measurable when they ask instructors to repeat instructions. While repeats may indicate purpose and comprehension levels, nonverbal feedback may indicate the participants' emotions, assembly strategies and attitudes towards the instructor. Specifically, the study measures objective and behaviouralvariables (Mutlu, 2011).

The objective variables can be determined through direct observation across participants (Mutlu, 2011; Shriberg, 2001) and these include:

1. Demographics - age, sex, languages spoken
2. Time - how long it takes participants to complete tasks, measured using a timer, video recorder and field notes logging their interaction with instructors from start to end.
3. Task performance as indicated by repeats of instructions associated with assembly cycle times. Participants request instructors to repeat instructionsassociated with specific assembly cycles when they need redelivery of instructionsasoutlined inChapters 4.4.2.1-2 and 5.3.2.1-2
4. The study uses a mixed method approach thus, data obtained from each study will be analysed to give within and between comparisons based on 4.3.2. and

5.3.2. Furthermore, data obtained fromHAI is compared with that obtained from HHI.

For this study, behavioural measures refer to observable and measurable nonverbal listenership behaviours.

1. The experiment will identify and classify marked facial actions and gestures displayed within specific occurrences of instruction repeats during tasks
2. The data obtained fromStudy 1(HAI)is compared with Study 2(HHI).

3.2.4Data Analysis

Data analysis involves two steps the first outlines corpora management procedures and the second focuses on linguistic segmentation.The study analyses corporain a principled way through systematic codification of data (LeCompte & Goetz, 1982) based on criteria outlined in the annotation scheme (Chapter 3.4) using specific software. This approach enables researchers to sort through and make sense of complex, detailed and interactive data which in turn, paves the way for easily relating content and expression planes captured in sophisticated media requiring low-level analysis (O'Halloran & Smith, 2012).

3.2.4.1 Corpora management

This chapter outlines how the study manages the corpora developed as data. The first step devises a way for managing data as information while the second processes the corpora using specific software.

Thestudy’smanagement information system provides the tools to organize and efficiently manage separate aspects of the corpora as they are being produced. The system ensures that the corpus is saved in properly labelled folders where access is restricted. These folders include:

1. Raw videos folder: video recording as they are;
2. Edited works folder:contains folders for each participant labelled P1, IH1

- Each sub-folder containsvideos the annotator has edited by cutting, joining, removing voice,etc and kept for future use. This is the study’s database

- Annotated video folders: Many programmes require that copies of the aligned video or audio be in the same folder with their own output or documents, for example, this is the case with CLAN's chat files, but ELAN does not require this except if it is linked to chat files.

3. Strip folder: this folder contains picture frames used for illustrating facial actions and phases that make up each gesture phrase displayed.

Organising the folders makes it easier to comply with the practical steps suggested for data collection and analysis (Jewitt, 2013; Bezemer & Jewitt, 2010; Kress, 2011) such as viewing videos with video logs to organise them into themes in line with research questions. This enables researchers to focus on how participants interact, for example, how they execute instructions when they get to critical points

Data analysis begins with the use of specific software and schema for corpus analysis. As outlined in 3.2.1-step 3, QuickTime Player 7 was used to align the videos recorded from the two angles for ease of analysis.

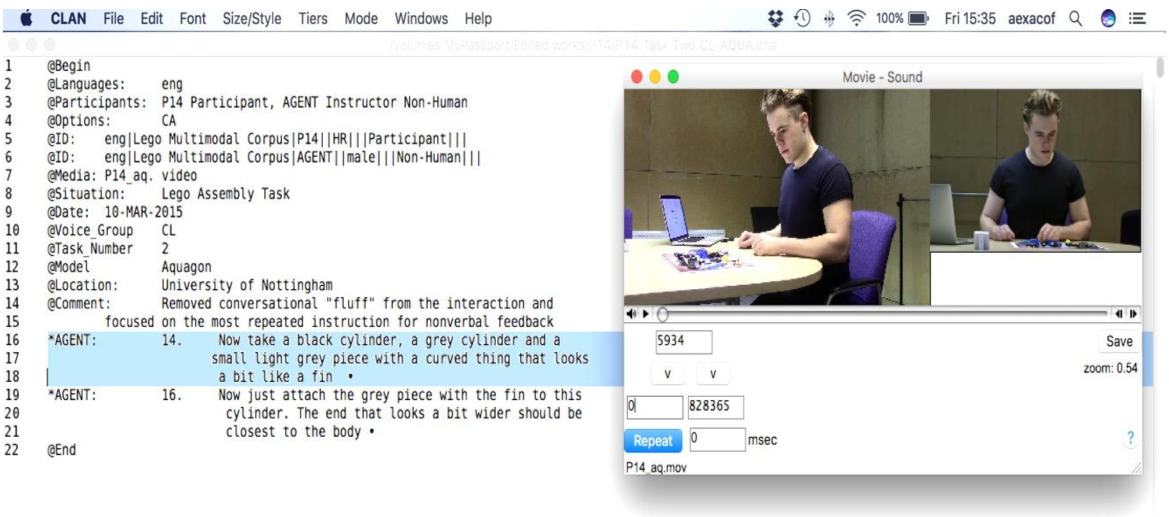


Figure3.5CLAN Annotation Interface

The annotator labels the aligned videoclip, for example as P14_aq.mov and saves it in .mov or MP4 format in P14 sub-folder. The aligned videois uploaded as media unto CLAN[6] (Computerized Language ANalysis). CLANis software designed specifically to analyse data transcribed in the CHAT format. The chat file is saved in the same

[6] More information http://childes.psy.cmu.edu/clan/

folder as the media file. CLAN is used to time-stamp and demarcate all significantly repeated instructions as bullet points on the chat document earmarked as *Agent: 14 and 16 shown (Figure 3.5) above.

The processed chat file is exported toELAN[7]-EUDICO Linguistic Annotator 4.9.0 (Lausberg & Sloetjes, 2009) for further analysis bearing the same file name.

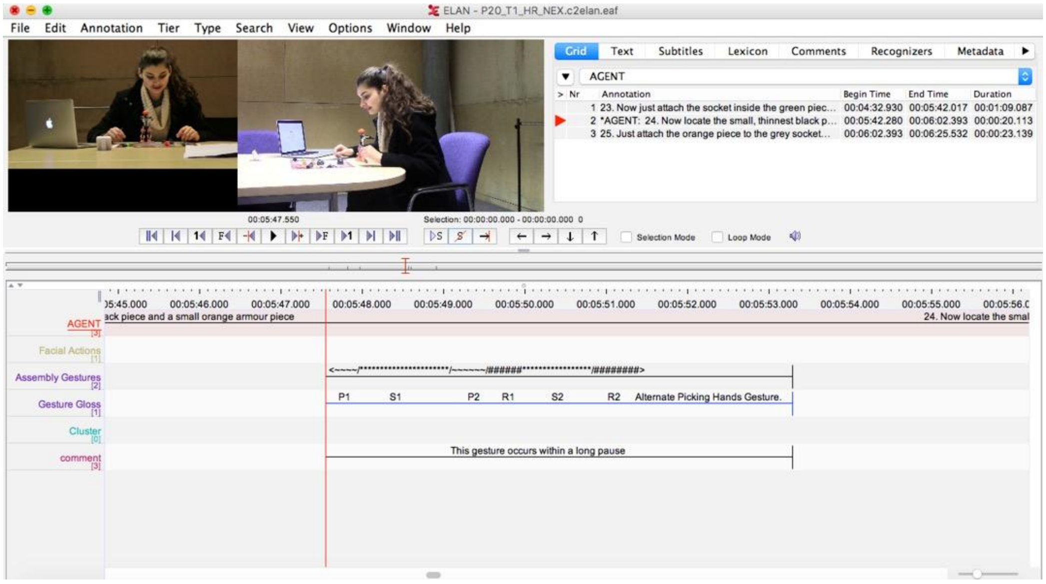


Figure3.6ELAN Annotation Interface

The ELAN annotation interface (Figure3.6) above is calibrated in tiers to show agent instructions, facial actions, assembly gestures, clusters and comments then saved in the participant's sub-folder. Each tier is independently time-alignedwithin the timeline of agent instruction tier that was marked by CLAN. The assembly gesture tier has a child tier called gesture gloss. The assembly gesture displays the structure of the gesture while the gloss tier labels the structures and names the gesture (Figure3.4). The other tiers are demarcated manually to show where and when they occurby the researcher using a mouse. Each tier is annotated manually using adapted schemas (Feng & O'Halloran 2013; Ekman, 2003).

Here the researcher observes how spontaneous nonverbal behaviour changes to give a thick description. The annotator watches video excerpts with sound to match instructions with facial actions and gestures. The annotator watches the videos in mute mode to reduce distraction and verify the annotation with sound and reduce bias

[7]https://tla.mpi.nl/tools/tla-tools/elan/

when annotating nonverbal listenership behaviour.The earmarked tier is watched frame by frame to get a detailedobservation of the interaction. For example, gestures and facial actionsare demarcated to show duration and identify the frame containing each facial action and gesture phrase as they unfold (see Chapter3.3). This process is repeated as many times as possible until annotators have identified the relevant framealso called the current frame containing the relevantfacial action or gesture phrase.

To save this frame as a picture, you right-click the mouse on it and a dialogue box appears with optionssuch as detach; save current frame as a picture; player info; forced aspect ratio, zoom, and copy non-adjusted media time (ignores aspect). Select 'save current frame as a picture', label it, for example as 'P20_aq_joining Hands_prep' for the preparation phrase of the joining hand gesture, then save the frame in the Strips folder.

3.2.4.2 Linguistic Segmentation

The multimodal corpora developed provide language data for qualitative analysis in the first instance and as Dörnyei (2007) points out although, such language data is not gathered with the aim of being objectively counted however, it may yield categories that can be quantified (Dörnyei, 2007, p.38). To this end, the aim of linguistic segmentation and annotation of the corpusin this thesis is to identify multivariate nonverbal language use, variation and nuances as communicative devices within the listener's implicit channel of expression during interaction.

This study is not an ethnographic study in the strict sense because the researcher did not have to move into any society, live with them for a given period and study aspects of their culture in their context. However, the studyapplies aspects of ethnographic research such as the intensive study of either a small group (L1 speakers of English), phenomena (spontaneous facial actions and gestures) and the findings tend to be unique to the interaction context (assembly task).

Regarding, approach to language data analysis, the study applies the ethnographic principle of thick descriptionsdefined as rich accounts of details of groups or phenomena (Geertz, 1973) in qualitative analysis of the multimodal corpora developed. To this end, this thesis will provide detailed descriptions of each marked occurrence of facial actions and gestures as well as meanings generated and attached to them as outlined below.On the usefulness of this strategy, Guba and Lincoln (1994) argue that a thick description provides other annotators with a database for making informed judgement thus creating the potential for the transferability of findings to other contexts. In addition, thick descriptions have enabled this study to use almost every relevant information relating to listener facial actions and gestures displayed during interaction for qualitative data gathering and analysis.

Furthermore, annotation is done manually using CLAN and ELAN because nonverbal behaviour is context dependent and may communicate different things to different people during interaction that may not be taken notice of by automatic recognition software (Gunes & Hung, 2016). Linguistic segmentation relies on the detailed coding scheme (Chapter3.3) that guides the annotation used in this study.

Bearing in mind that the study does not use post-task interviews with participants, a systematic approach still needed to be used to determine when facial actions become emotions and body movements become gestures. The process of transforming actions and movements into meaningful units of nonverbal language is outlined in 3.3.1.3 as part of the annotation scheme.

Segmentation of the corpus focused on assessing the nonverbal behaviour of the subject by identifying facial actions and gestures made. In doing this, the researcher used the facial action coding systems (Ekman & Friesen, 1978a-b), and derived gesture hierarchies as done by Kendon(1980); Kita et al., (1998) andKipp (2003) to mention but a few.Three basic steps are used in linguistic segmentation and annotation of facial actions in this thesis illustrated using a smile (Figure 3.7 below) and aligning hand gesture (Figure 3.8) overleaf.

Figure3.7Annotating Facial actions- a Smile

Preparation P1n Initiates the gesture by picking the assembly piece, compares them and tries to mentally and visually place them on the Lego model.

Stroke P1n performs the gesture by bringing the hands together thus physically taking the piece close to the part of the Lego kit and sizing them up

Retraction begins as P1n's hands start to go apart still holding the piece without fixing it on the Lego kit.

Retraction/Recovery (dying) is completed when P1n's hands are finally at rest on the table.

Figure3.8 Vignette for Annotating Gestures

The steps proceed as follows. The Facial Action Coding Scheme (henceforth FACS)(Ekman & Friesen, 1978a-b) is used to describe the neuro-biological processes generating facial actions. As outlined in chapter 2.2.3.1, FACS is useful for discovering and coding facial actions because it provides an objective and comprehensive way to analyse facial actions as components. These components as biological processes,describe facial muscles responsible for an expression as Action Units (AU) (see 3.3.3.1).The components as neurological processes relate to the five senses and research suggests that a smile may be elicited by positive physical, verbal and tactile stimulation (Ekman & Friesen, 1982), for example, when people enjoy touching or experiencing something they smile.Thus, using FACSthe facial muscles responsible for the smile (Figure 3.5)include AU 6+ AU 12+ AU7 cheek raiser, lip puller and lid tightener.

Regarding gestures, description focuses on form and function. Form is analysed through segmentation. Kita et al (1998) explain that in gesture analysis, segmentation is the recognition of a stretch of movement with a certain directionality (path of motion) as a phase and the identification of the phase type (manner of motion). Phase type in this study refers to preparation, stroke, failed stroke, retraction, partial retraction, or hold stages of gestures. For example, the picking hand gesture in Figure 3.6 is like a beat gesture (Kirk et al., 2005) with two basic moves (out-in). This is made more explicit with a vignette, directional arrows and description going on side by side from top to bottom as illustrated in Figure3.6. Sometimes, the structure of the gestureis shown below the vignette to illustrate when and howthe gestureoccurs(3.3.2.1).

The next stage is to determine the classification of the nonverbal action using the model outlined in the annotation scheme(Chapter3.3 & section 3.3.1.3). Facial expressions are classified using FACS thus, the facial action in Figure3.5 is described as a 'felt smile' (Ofemile et al., 2016; Ekman, 2007, 2003). Thereafter, the researcher locates the communicative function of the nonverbal action within the interaction context. For example, the facial action –felt smile is a positive face (Tipples et.al., 2002) and suggests that the participant is enjoying the interaction with the instructor. Gestures also have task functions;Figure3.6, for example, is used for sizing up assembly bits and measuring their fit but, this may also indicate the participant's thought processes during interaction.

The annotation system adopted in this study is close to standard orthography and capturesas much fine granularity as possible. The level of granularity in this study is influenced by Potts, (2016),Harrison, (2014), Bressem, (2014), Kita(2013), Matsumoto and Willingham, (2009), Allenwood, (2008), Kendon (2004, 1988), Ekman and Friesen, (1978b), andMcNeil, (1985).It aims to cluster, quantify and identify the variety of spontaneous assembly gestures, facial actions and clusters that emerge from interaction.

To enable easy assessment of the impact of components of multimodal interaction on communication and emotion, the annotation system involves identifying annotation tiers, values and their descriptions. Tiers will look at values obtained with descriptions for sequence, phrases, practice, and referent from a listener's perspective. Furthermore, whileKita et al (1998) used numbers to depict different gestures on rows and columns, this study like Zwitserloodet al (2008) uses descriptive tiers that group nonverbal behaviour under major headings such as, hand shape, finger, wrist position, palm orientation, and Back of Hand (BoH) (Allenwood, 2008) under gesture while lip shapes are annotated under facial actions usingFACS.

One aspect of linguistic segmentation focuses on categorisation and discussion of the most frequent linguistic and multimodal bundles considered as indicators of agreement and variation in listenership behaviours as seen in listener comprehension (Field, 2008; Oxford, 1993) and information construction (Hilpert, 2014, 2013; Sorrows & Hirte, 1999; Vallduvi & Engdhal, 1998, 1996). In this study, categoriesof nonverbal behaviour include facial actionsand gesture are recorded, classified and compared across categories with the aim of identifying emerging relationships useful for rule setting (Ekman, 2016; Sekine & Kita, 2015; Bressem, 2014; Kita, 2013; Matsumoto, 2012; Kendon, 2004; McNeil, 1985; Goetz & LeCompte, 1981).

Furthermore, categorisation enhances systematic analysis beginning at the micro-level and proceeding upwards until the whole interaction is captured. The data is analysed as a form of typology , for example, repeated instructions are classified into typologies based on needs, gestures and facial actions are categorised according to families (Kendon, 2004; Ekman, 1997) in Chapters 4 and 5.

As outlined in Chapter 2, evolutionary and innate factors are responsible for facial action characteristics in people (Ekman, 2007) and these led to the development of emotion families and their variations. Similarly, assembly gestures are classifiedas families because they have one or more kinesic or formational characteristics in common (Kendon, 2004). Within each family there are varieties of facial actions (Tables 4.5, 5.4) and gestures(Tables 4.6& 5.5) that are identifiable as distinct because of their manner of execution and communicative functionsduring interaction.

Another analytical focus deals with understanding the communicative practices of participants as listeners, for example instruction-taking strategies, process and the motivation for communicative behaviour during interaction.

Regarding results presentation in Chapters 4 and 5, samples that are representative of families of facial actions and gestures displayed within the Chapters because they have all the basic features of each family. Furthermore, specific or unique occurrences are explained while the others are in the appendices.

3.2.5 Inter-rater Reliability (IRR)

Statistical tests such as correlational, Pearson correlation, Spearman correlation, and chi-square are used in social science research to look for the association between categorical variables. Knight (2009) points out that statistical tests allow analysts to examine relationships between research populations and to ascertain whether patterns observed between populations occur by chance as a measure for accepting or rejecting hypotheses.

It has also been proposed that reliability and validity in research can be assessed using criteria of trustworthiness(Bryman & Bell, 2007;Guba & Lincoln, 1994; Lincoln & Guba, 1985). Trustworthiness means that results are credible, transferable, dependable and confirmable and these parallel internal validity, external validity, reliability and objectivity criteria respectively.

On-going research findings are credible if they meet existing good practice benchmarksand are submitted to other researchers for confirmation (Lincoln & Guba, 1985).This thesis believes that credibility is linked to inter-rater agreement becauseInter-Rater Reliability (IRR) is a process by which one set of coders assess theobjectivity, reliability and validity of another coder's interpretation of phenomena thus, establishing the degree of agreement between the coders.

I am the first annotator and the second annotator was another PhD student of English and Applied linguistics researching sign language in a UK university who had submitted his thesis after defence and has a lot of experience in annotating nonverbal interaction using ELAN. In conformity with established research practice, a skilled second annotator watched interactions on video and analysed them (Appendix IV) using ELAN and the annotation scheme developed in six sessions across three days (Chapter 3.3) as suggested by Kita et al. (2013). After each session, the videos were stored in the same secured location as the originals by the researcher until the sessions were completed. The second annotator's results were compared to mine to ensure reliability and validity of data analysed in this study. This also minimised the effect of rater bias on study because the second annotator was emotionally detached from the study and was able to provide a second view that was compared to mine. The inter-rater analysis ideally established some levels of agreement among annotators using the same consistent method which made the results more reliable (Gwet, 2008; Dörnyei, 2007).

Facial actions are analysed using Action Units (AU) thus, IRR is measured based on the number and similarity of AUs identified when categorising facial actions. Equally, gesture annotation is done through segmentation as outlined by Kita et al. (1998) thus, the basis of IRR is the number and type of phases identified within a given gesture by annotators.

Since there is no concensus among researchers regarding the single best index for IRR, Wongpakaranet.al. (2013) and Lombard et.al. (2010) agree that using multiple methods enables researchers to account for agreements by chance and make a reliably stable interpretation of interrater indices. Thus, IRR in this study is calculated using two broad approaches (Appendix III) percentage agreement or the rule of the thumb and analysis of first and second annotatorsperceptions to establish coefficient correlation. Percentage agreement computes agreement in percentagesand results obtained are traced to comparative benchmark scales for their Kappa values (Gwet, 2012; Lombard et al., 2002)

The second approach assesses IRR using an inter-rater agreement calculator[8] available online(Geertzen, 2012)to measure Fleiss' Kappa (*k*) and Krippendorff's alpha (α). Gwet (2012) and Wongpakaran et al., (2013) suggest that the benchmark for acceptable per cent agreement is 75% for data with more than 5 categories.

3.3Annotating Nonverbal Listenership Behaviour

3.3.1 Introduction

Annotation is the process of adding interpretive, linguistic information to corpora(Leech, 2013, p.2)using a scheme put together by researchers to ensure that rigour is established through codification. Codification leads to systematic analysis of data and creates a reference material for assessingIRRand making the study reliable and valid. A detailed annotation scheme ensures that data codification is uniform

[8] See https://nlp-ml.io/jg/software/ira/

(Bentahar et al., 2006), clearly stated, comprehensible (LeCompte & Goetz, 1982) and easy to use by other researchers.

There are many gesture transcription schemes available such as Birdwhistell's, (1952),Laban's, (1975)proposals for posture and large-scale gestures, McNeil's (1979) proposals, McNeil-inspired proposals including Kita et al, (1998), Mansoon, (2003), the MUMIN multimodal coding scheme(Allwood et al., 2005)and Harrison, (2014). Following these examples, the corpus in this thesisis made useful and searchable through annotation based on comparative coding schemes for making annotations.

The corpusannotation scheme has combined and extended insights from Kita et al. (1997), McNeil (1999) Bressem, (2014) and Kendon (2004) to mention but a few. These conventions enablea structured and systematic practical analysis of nonverbal listenership behavioursand provide a framework of implementation for theories of nonverbal listenership emerging from interaction.

However, since holistic transcription is time consuming and almost impossible, Harrison (2014) observed that researchers normally decide on a system of annotation and coding that is oriented towards specific communicative function such as McNeil's (2005,1985, 1979)speakership skills. Thus, this annotation is multimodal and geared towards spontaneous nonverbal listenership behaviours- (facial actions, assembly gestures, repetition requests)displayed during tasks.

While assessing nonverbal behaviours, Ekman and Friesen (1969) posit that, in order to understand a person's nonverbal behaviour, there is the need to understand how that behaviour became normalised or taken for granted or natural in that person's everyday life(May & Finch, 2009; McGill & Emerson 1992). The three considerations that enable us to understand the normalisation process of nonverbal behaviour in people are origin, usage and coding which provide the basis for understanding how a movement becomes gesture or a facial action becomes an emotion.

3.3.1.1 Origin

Nonverbal behaviours take root from three sources. The first is the relationship between stimulus event and the responding nonverbal activity existing within the framework of the specie such asa person’s reflex reaction to hot or cold things.

Secondly, there is experience common to all members of a group deriving from interaction with their environment and the human body , for example,legs are used for walking in all human cultures. Thirdly there is experience which varies according to age, family and individuals that may also be personal, group, or vicarious. While some nonverbal behaviour could be learned as part of mastery of a particular act like urinating, and eating others according to May and Finch (2009) become routinely embedded or normalised during interaction like speech.

3.3.1.2 Usage

Usage refers to the regular and consistent occurrence of a nonverbal act and it has to do with the following: 1) the external conditions found whenever the act occurs; 2) the relationship of the act to the associated verbal behaviour; 3) the person's awareness of emmiting the act; 4) the person's intention to communicate; 5) feedback from the person observing the act; 6) the type of information conveyed (Ekman & Friesen, 1969, p.53). Usage is useful because it enables the annotator to focus on nonverbal behaviours that perform communicative and/or interactive functions. External conditions refer to the context of interaction.Ekman and Friesen (1969) describe external feedback as direct verbal comments by paricipants during interaction but in this thesis, external feedback refers to nonverbal externalisation of thought processes by listeners either as gestures or facial actions. Awareness refers to obvious visual attention paid to specific nonverbal behaviours that are reactive to the listener's actions and clearly inform interlocutors that their nonverbal actions are perceived and evaluated.

3.3.1.3 Coding

Coding refers to the correspondence between the act and its meaning. This is done based on a model showing how listener multimodal action such as a movement becomes a gesture (Figure3.9) below and earlier outlined in chapter 2.2.3.2. Coding describes how meaning is contained in a verbal act and it shows the relationship between the act and its referent(Jeremiah, 2014). For example, using two hands to show the size or shape of an object the relationship is pictorial and iconic, while bringing two palms together shows prayer mood, or intimacy.

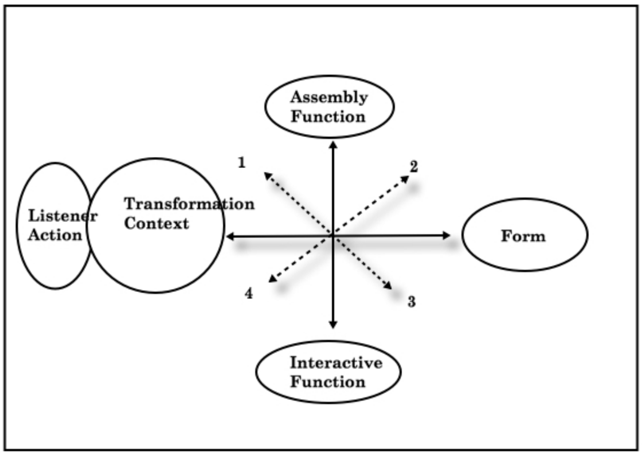


Figure3.9Representation of Listener Multimodal Action by Form and Function (Influenced by Mozdzenski, 2013,p.184)

The provisions of the model (Figure 3.9) above show the process through listener action either as facial action or body movement becomes a meaningful facial expression or gesture through coding. The model is premised upon the notion that listener action on its own may be meaningless but, coding shows the ways in which listener nonverbal acts (other than arbitrary ones) are related to their significants. For example, when listener movement executes all or part of an act in a performance as a rendition of part or all of its meanings the relationship is kinetic and iconic e.g. fist waving or hitting action. However, when placed within a given context these movements begin to take on meanings and interpretations that were hitherto not there or envisaged when these listener actions first occurred in isolation.

The transformational context provides a basis for actions to move from the level of mere symbolism or meaninglessness to the threshold of meaningfulness in interaction. The transformation context includes interaction context outlined in chapters1.1.; 2.4.2.1 and the specific eliciting context (Chapter 3.2). The transformation context also refers to the communicative context within which listener action occurs and thus shapes its meaning because as Ekman and Friesen (1969) explain, typical nonverbal behaviour may combine elements of more than one code (Ekman and Friesen, 1969).

In addition, the effect of transformation context on listener facial action and gestures is similar to meaning-making in verbal communication, where there is a marked difference between the meaning of words when they occur in isolation (lexical meaning) and when they occur with other words in use (contextual meaning).

In this sense, the context includes the type of verbal or nonverbal actions that co-occur with the facial action and assembly gesture being analysed such have been earlier been called co-verbal gestures, co-thought gestures (Kita, 2013; Mehrabian, 2007; McNeil, 1985; Kendon, 1980). The communicative context enables researchers to establish congruence or incongruity of nonverbal behaviour as well as its meaning from context (Pease & Pease, 2004).

Transformation is achieved through the normalisation process outlined above however, this model becomes operational when it comes to coding and interpretation. At this point, the model (Figure 3.7) is used for synthesizing the representation of facial actions and gestures through form and functions.

Form refers to the AUs responsible for the occurrence and appearance of facial actions (See 3.3.3.1) and the structure or configuration of the gesture phrase. The form of facial action is directly linked to their typologies, for example, facial actions can either be basic or non-basic. However, the form of gestures is indicative of their structure as basic or concatenated (See 3.3.2.1) and may suggest its typology.

Form is expressed in coding, for example, when part of the body like the hand, head, lips, eyes is used to point to somebody, an object, part of the body, direction, or location,(i.e. the referent) thus, pointing is always intrinsically coded because the act shows something which is the referent or some nonverbal deixis.

The functions describe the interactive purposes and assembly roles (numbered 1,2,3,4) that listenership behaviours serve during the task as outlined in chapter 2 (2.2.3.1 & 2.2.3.2). Assembly functions refer to nonverbal actions directly linked to the execution of the task while interactive functions relate to actions that build intrapersonal and interpersonal communication during interaction. These functions emerge performatively, may intersectat some point and are given meaning based on the context of interaction, instruction type and context of co-occurring actions.

Accuracy of interpretation is a potential issue of concern because nonverbal behaviour interpretation is influenced by interaction context, culture and personal experience.However, this model when implemented using the annotation scheme (3.3.2) provides a standard interpretation manual that ensures some level of accuracy in the analysis of interaction, facial actions and gestures done in Chapters 4 and 5.

3.3.2 Annotating Gesture

3.3.2.1 The Gesture Phrase

Kendon's, (2004, 1990, 1988) detailed kinesis structure with hierarchies for analysing gestures defined terms required for implementation and set the context for analysing gestures. The Gesture Unit (G-Unit) begins with a rest pose then several gestures consecutively occur in succession and it ends with another rest pose. Kita et al., (1998) describe the same procedure as a hierarchical process that begins when the limb is lifted away from the body and ends when it is back to a resting place such as the work space, Table, and lap or arm rest of a chair.

The Gesture Phrase (G-Phrase) comprises several mandatory and optional phases. The common gesture phrase comprises (preparation+ stroke + retraction). While multiple concatenated gestures may comprise (preparation + pre-stroke hold +stroke + post-stroke hold + retraction)

Preparation and retraction are focused on reaching destinations while strokes are the content-bearing part of gestures that focuses on the form of the movement itself or its trajectory, shape and posture (Kita et al., 1998; McNeil, 1992).

However, in view of the peculiar nature of assembly tasks, this structure has been expanded in this thesis to include 'failed strokes'(Figure3.10) to enable annotatorsidentify occasions when participants fail to execute gestureas against when they succeed.

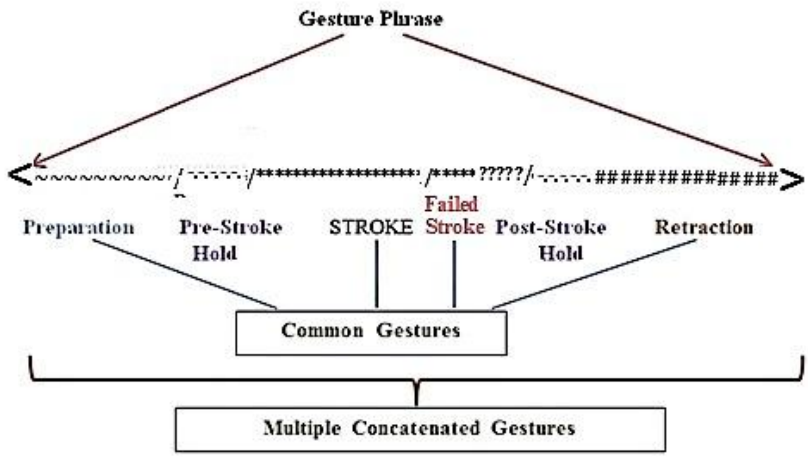


Figure3.10 The Expanding G-Phrase (Adapted from Kendon, 2004; Kipp, 2003; Kita et. al., 1998; McNeil, 1992)

The phases are summarised below:

1) Preparation: in this position, body parts move from rest position to the start or initial position of gestures and this is an optional phase that could go in any direction when executing gestures.

2) Stroke: this phase is mandatory, contains the expression of the gesture and requires more effort.

3) Failed stroke: this phase may be accidental, and it describes when the expression of the gesture does not occur after due effort has been made.

4) Retraction: the body parts move to a rest position in the case of a common gesture. However, when executing composite gestures participants may omit this stage by executing single or multiple concatenated gestures

5) Holds: are generally temporary cessations of movement without exiting the gesture hierarchy. There are pre-stroke and post-stroke holds that are optional and occur shortly before or after the gesture stroke.

6) Recoil phase occurs after a forceful stroke as a reaction to gesture execution force.

The kinetic structure enables systematic annotation, transcription and classification of gestures within the expanding boundaries of the gesture phrase. The transcription convention(Table 3.3 below) used for this study has been influenced by McNeil’s (1992) aggregation of the works of others particularly Karl-Erik McCullough’s full version of convention.

However, some additions have been made for clarity and to cater for the unique nature of assembly gestures. For example, there is a high number of holds occuring before and after strokes during tasks,making it necessary to differentiate them in signage as pre-stroke hold and retain McNeil’s post-hold stroke.

Table3.2Gesture Phrase Transcription Conventions

S.NO	SIGN	Meaning Within the G-Phrase	Short Tag
1	~	Preparation	P
2	\\\\	Pre-stroke Holds	PsH
3	/**/	Stroke phase	S, S1; S2; S3; Sn
4	.-.-.	post stroke hold	psh
5	<>	G-Phrase duration	Not applicable
6	///	Silent pause with duration in seconds	Not applicable
7	/**??/	Failed stroke phase	S, S1; S2; S3; Sn
8	‘~’	Recoil	r
9	##	Retraction	R

Following the examples of Bressem, (2014); Kendon, (2004), Kita et al. (2008), and McNeil (1992), this study developed the annotation scheme for assembly gestures usingfour parameters that include hand shape, orientation, location and movement summarised (Table 3.4)below.

Table3.3 Gesture Annotation Scheme

Gesture Parameters	Class Description	Notation
Hand Shape	Configuration	Fist; Flat;Single fingers; Combination of fingers
	Name of digits	Thumb; Forefinger; Middle finger; Ring finger; Little finger
	Shape of digits	Stretched; Bent; Crooked; Flapped down Connected; Touching; Coupler-shaped
	Activation of digits	Wriggling/vibrato
		Flick
Handedness	Operational hand	Both hands; **RH**-Right hand; **LH**-Left hand
Orientation	Palm	Palm(s) up/supine; Palm(s)down; Palm(s) Lateral
	Grasp	Hold up-; Pull up-
Location/Gesture Space	Sagittal axis	Upper right; Right; Centre; Centre-centre; Left; Upper left
	Task area	Front right-; Front-Front Left
Movement	Hand movement	straight
		Circular clockwise
		Circular anticlockwise
	Movement 1: along vertical and horizontal axis of the gesture space	Up-; Down; Right; Left; Front
	Movement 2: to and from the sagittal axis/ body of the listener	Diagonal right Towards sagittal Axis Diagonal left Away sagittal axis

3.3.2.2 Location-Gesture Space

Gesture space is described as a shallow disk visualised in front of the interlocutor (McNeil, 2005) or listener in this thesis. The description of location and movement of hands in existing systems such as McNeil's have been criticised for either being impossible or too crude and lacking adequate details (Zwitserlood et al., 2008). Following this and the absence of gestures done backwards or behind participants because of the nature of tasks, McNeil's (2005) depiction of gesture space has been modified to describe assembly gestures (Figure 3.11) below

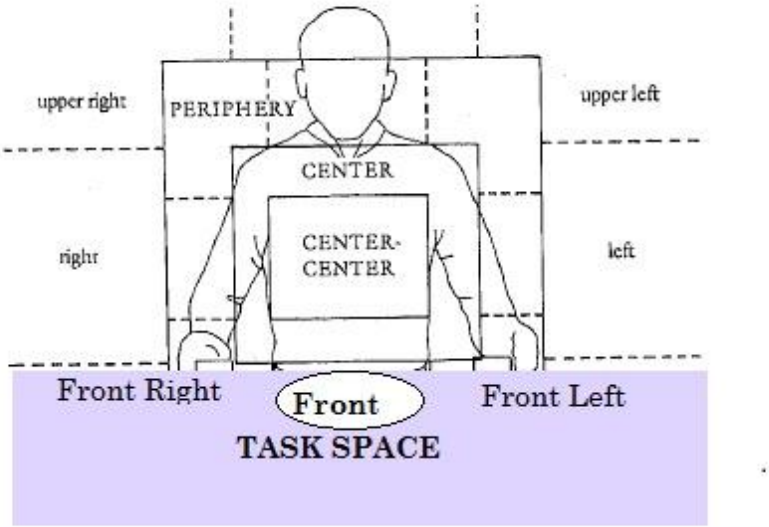


Figure3.11Listener Gesture Space in Assembly Tasks (Adapted from McNeil, 2005,p.89)

The space is divided into a system of concentric squares. Directly in front of the listener's torso is the "centre-centre". There is the periphery divided into the upper right and right; the upper left and left (Figure 3.11). The adaptation used in this study further reduces the gesture space of the listener because the task does not require gestures above the head. In addition, listenersare expected to sit on a chair behind a Table and assemble Lego kits thus removing the need for lower gestures that occur from the knees downwards. The task space is also divided into 'front' depicting the space on the Table directly in front of the listener. Following McNeil's (2005) example, the periphery of the task space is labelled 'front right' and 'front left'. These will justify the classification of hand movements across the task space and will be used in conjunction with the movement depicted in the sagittal axis.

3.3.2.3 Hand Shape

Hand shapes as described in Kirk et al., (2005), Kendon (2004), Harrison (2014), and Bressem, (2014) have been extended in this study for the purposes of clarity. Bressem, (2014) developed a hand configuration describing thefist, flat hand, single fingers, and combination of fingers that are used in this annotation scheme. The shape of digits is described as stretched, bent, crooked, flapped down, connected, or touching.

Name of Fingers

Bressem, (2014) numbers fingers beginning from the thumb as 1,2,3,4 and 5, in contrast, this study uses everyday ordinary names of fingers, thumb, forefinger, middle-finger, Ring-finger and Little-finger (Figure 3.12).

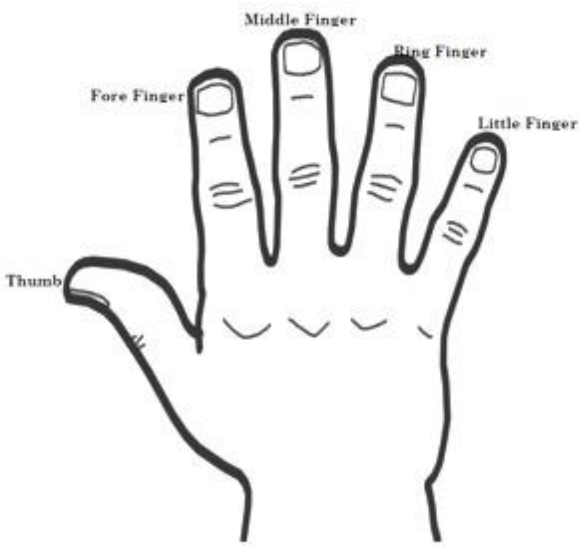


Figure3.12Name of Fingers

Shape of Digits

Bearing in mind the differences between Bressem's interaction contexts with the unique demands of assembly tasks, the shape of digits when all fingers touch at the tips is described as coupler-shaped as depicted (Figure3.13)overleaf.

Figure3.13 Coupler-shaped Picking Hand (Right- Left)

Coupler-shaped handslook like the fingers bunch family of gesturesbut it is the opposite of grappolo as depicted in Kendon (2004). However, it is used for precision grip (Kendon, 2004) when picking up assembly bits up as shown in Figures 4.30, 4.43, and 4.66 later.

Finger Movement

This movement could be horizontal or lateral as indicated by arrows and could be used for selecting pieces before picking up, visual assessment and locating pieces mentioned in the instruction duringassembly tasks (see Figure 3.14) below

Figure3.14One-finger flick

The finger movement in Figure 3.14 occurs in the front-left location of the task space. Flicks occur when participants propel assembly pieces with a sudden but gentle movement of one or more fingers on the surface of the workspace. This is different from the flick used to remove ash off a person’s sleeve or the way a horse flicks its tail because it is not as intense in speed or suddenness.

Combination of Fingers

In normal interaction fingers play a critical part during conversations because they perform functions within a space like pointing, negation, and agreement (Kendon, 2004). In assembly tasks a combination of fingers is used for assistive gestures as shown (Figures 3.15, 3.16, 3.17)below.

Figure3.15 Thumb-forefinger Combination

Figure3.16 Thumb-middle-Ring FingerCombination

Figure3.17 Crooked Digits

Figures 3.15, 3.16, and 3.17, above describe the shape and combination of fingers used in picking up small things in the Lego kits like flat and rounded ones (Figures 4.32b, 4.65a, 4.65b). Figure3.17 is a bit different because it is crooked (Bressem, 2014) and used in readiness for picking up just as people take aim before shooting.

3.3.2.4 Orientation

Orientation indicates the palm's position during tasks. These include, palm up, palm down (Bressem, 2014) and palm lateral orientations.

Palm Up-Supination

This is part of the palm open or palm up supine family (Bressem, 2014; Kendon, 2004). Supination is the movement of the forearm into a palm up orientation used in receiving, giving and presentation functions. The arrows indicate the direction of movement and it may be extended towards the space immediately before the interlocutor and can be done with one hand or both hands as shown (Figures 3.18and 3.19)below.

Figure3.18 Two-hands Palm up

Figure3.19 One-hand Palm up

The palm up gesture (Figures 3.18& 3.19) is used to place and display something on the palm before self and others as “Palm presentation” (Kendon, 2004,p.265) and it is used in assembly tasks to confirmlistener comprehension of instructionsand assesstasks executed.

Palm Down-Pronation

Pronation is the movement of the forearm into a palm down orientation. It is similar to McNeil’s gesture 122 (McNeil, 1992, p.10) depicting an iconic gesture as palm facing down however, the ones depicted (Figures 3.20, 3.21) below do not make a flat surface, illustrate an event or convey a location.

Figure3.20 Palm down front and side views

Figure3.21 Two-hand Palms Down

In assembly tasks, the palm down gesture is used to pick, select or drop assembly parts by listeners and it may involve one or two hands (Figures3.18&3.19 above) depending on the task and listener’s skills

Palm Lateral

In this orientation, the palm is open, facing the body with the knuckles pointing away (Kita, 2000, p.172) from the sagittal axisas depicted (Figures 3.22 & 3.23)below.

Figure3.22 Palm Lateral

Figure3.23 Two-Hand Palms Lateral

The lateral palm at 45^0 or less could be either half-supination or half-pronation depending on the direction of wrist movement as outlined in supination and pronation.Figure3.20 depicts palm lateral that is centre-left while Figure3.23shows both hands in the upper centre regionused to move and place assembly parts before listeners for inspection.

Grasping Hands

Grasping was first described by Drillis (1963) later six different ways of grasping a device were distinguished (Wimmer & Boring, 2009). Grasping shows how people hold up or pull out things from right, left or central positions, for example, from a Table. Wimmer and Boring (2009) posit that graspingmanifests in hold-ups, pull-ups and graspingand these orientations are used in this thesis to identify user-handedness in assembly tasks.

Hold-up

Hold-up could be from the right or leftdepending on the task and function during interaction,but the object is always held from underneath (Wimmer & Boring, 2009) with the rest of the body protruding out and upwards.

Figure3.24 LH-Hold-up

The hold-up(Figure3.24) aboveis from the left executed with an up-side-down movement and it is used like palm presentations to examine a piece picked up or wait for the next instruction or assembly step.

Figure3.25LH_Hold-up

In Figure3.25, the hold-up is also from the left with LH (the receiving hand) holding up the body to receive a bit from the RH (the giving hand) and this is used in joining hands and aligning hand gestures.

Co-occurring grasping gestures

Sometimes hold-up co-occurs with grasps during assembly tasksto aid listeners in executingsome task stages.

Figure3.26 RH Hold-up Co-occurring with LH Grasp

In Figure3.26two grasping gestures co-occur and the RH is holding up an assembly piece while the LH grasps the body from the left. Note that the feet of the Lego kit are on the Table unlike in Figure 3.23 where the feet are off the Table however, like the hold-up, grasping is used to prepare for joining and aligning hands gestures.

Pull Up

This is another form of grasping that occurs at the end of the picking gesture from the top (Wimmer & Boring, 2009; Drillis, 1963)(Figures 3.27 and 3.28) below.

Figure3.27 Grasp_pull up

Figure3.28 Grasp_pull up RH

Figure3.29 Grasp_pull up with both hands

Participants use one hand(Figures 3.27 and 3.28) and both hands (Figure3.29) to grasp and pull up assembly pieces from the task area.

3.3.2.5 Movement

The movement phrase is divided into continuous discrete phases. A phase is divided into two if there is an abrupt change in direction of the hand movement (McNeil, 1992). A movement with two segments but without anabrupt change in speed within a 3-D spaceis regarded as one phase even if there is a sudden change in direction (Kita et al., 1998, pp 25-32). While a multi-segment phase is said to have occurred when a hand reaches for an object and pushes it to a desired position as shown in the movement by people using a top stone or muller and a grinding stone to grind seeds (Figure4.68).

In addition, semi-multi-segment phases occur if a hand stops suddenly in the first segment then bounces back in the second segment in the opposite direction tracing the trajectory of the first segment either fully or in part. When the same movement is repeated without a hold between, the phase is said to be repetitive (Kendon, 2004; Kita et al., 1997; McNeil, 1992) as shown in a repetitive opening hands gesture (Figure4.51).

The movement classifier positions the arm, wrist or fingers within a specific direction (Kendon, 2004). Types of arm or finger movements include straight, arched, spiral/rotation, bending to pull or rising while the direction of movement could be circular indicated with circular arrows (Bressem, 2014; Prillwitz et al., 1989). The following classifiers for assembly gestures are used in this study.

Type 1

Type 1 is the straight direction of movement as the crow flies in any direction and indicated by straight solid arrows (Figure 3.30below).

Figure3.30 Type 1 Movement Straight

Type 2

Type 2movements are circular in nature.They include Bressem's (2014) arched movement with the range measuredin degrees like 180^0 and described in this thesisas 'clockwise and anti-clockwise'(Figure3.31) below.

Figure3.31 Type 2 Movements Circular (Clockwise &Anticlockwise)

Direction of Movement

Movement 1

Movement 1 follows Bressem's (2014) example but with a modified notation indicating movements along the vertical and horizontal axis (Figure 3.32) below.

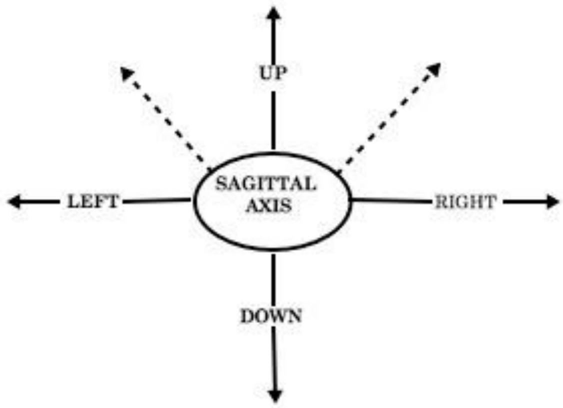


Figure3.32 Movement along Vertical and Horizontal Axes (Adapted from Bressem, 2014, p.17)

As explained in Chapter 3.3.2.2, this annotation focuses on forward movements as there are no backwards or downwards movements however, downwards movements are placed to show contrast. Left and right movements correspond with 'front left and front right' while the circlerepresents 'front' in location (Figure3.11).

Movement 2

Movement 2 depictshand movementsfrom the listener's sagittal axis (Bressem, 2014), which is either towards or away from the sagittal axis. These movements have been modified in this study to suit the listener's body(Figure 3.9) and view it from above. The movements are described (Figure 3.33) belowas follows.

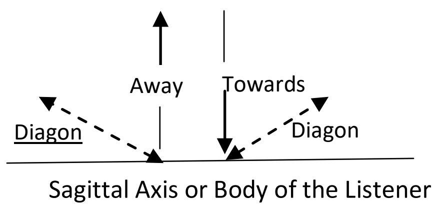


Figure3.33 Movement to-from the Listener's Sagittal Axis (Adapted from Bressem, 2014:27)

The horizontal line indicates the listener's sagittal axis. The solid arrow on the left indicates movement away from the body towards the task space and kit layout while the one on the right shows movement towards the sagittal axis. The bidirectionalbroken arrows indicate diagonal movements to and from the sagittal axis from right and left.

During annotation, the coder may need to identify movements that have not taken place to show participants' intentions indicated by a broken arrow (Figure 3.34) below.

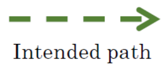


Figure3.34Intended Direction

The broken arrow goes in every direction and is used for both hands like the queen in a chess game.

3.3.2.6 Colour Coding

Research indicates that colour coding reduces confusion on the part of readers and aids object detection (Dalal & Triggs, 2005; Papageorgiou et al., 1998).To make gesture analysis clearer, this study uses colour and letter coding consistently because this draws attention to specific aspects of the annotation. The following colour and letter codes(Table 3.5) are used in this thesis

Table3.4Colour codes

S.NO	Colour Code	Meaning
1	RH →	Right hand code and movement
2	LH →	Left hand code and movement
3	↔	Bidirectional right hand
4	↔	Bidirectional left hand
5	- - - - ▶	Intended movements

When the participant's clothes are the same or similar colours with 2 and 4 above a contrasting colour is used.

3.3.3 Annotating Facial Actions

3.3.3.1 Facial Action Coding Scheme

As outlined in chapters 2.2.3.1 and 3.2, Ekman and Friesen's (1978a) FACS is used to identify all functional anatomic facial muscle movements orAction Units(AU),head and eye positions (Matsumoto & Willingham, 2009) as well as tomeasure the relaxation or contraction of each muscle (Figure 3.35).Specific movements and positions of the face relate to given impressions in normal interaction as research indicates that there is an association between specific facial muscles and specific emotions (Ekman et al., 1972).

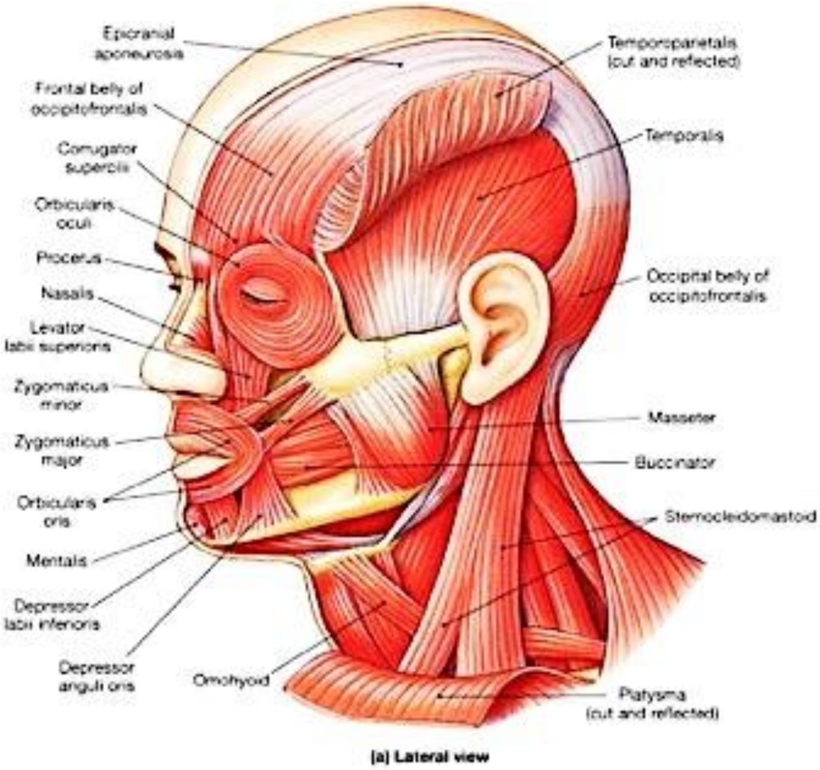


Figure3.35 Muscles for Facial Expression[9]

---

[9]more informationhttp://droualb.faculty.mjc.edu/Lecture%20Notes/Unit%203/muscles%20with%20Figures.htm

Table3.5 Ekman's Basic Emotions by FACAID[10]

EMOTION	Action Units (AU)
Happiness	6 + 2
Sadness	1 + 4 + 15
Surprise	1+2+5B+26
Fear	1+2+4+5+7+20+26
Anger	4+5+7+23
Disgust	9+15+16
Contempt	R12A+R14A

Facial muscles (Figure 3.35) create facial actions however, more than onemuscle can be grouped into oneAU(Table 3.6 above)while musclescan be divided into separate AUsbecause of duration, intensity and asymmetry (Ekman, 2007,1997).Ekman (1997) categorises intensity of facial actions by appending letters A – E to AUsindicating minimum to maximum intensity as outlined as follows; A) Trace; B) Slight; C) Marked or pronounced; D) Severe or extreme; and E) Maximum.However, Ekman's (1997) scheme does not specify the difference between 'severe or extreme' and 'maximum' intensities, this may create ambiguity, make interpretation subjective and without a standard.In view of these, this thesis adopts only three levels of intensity given as A-Trace; B- Slight; and C - Pronounced or maximum to correspond with classifications of emotions as 'slight, Partial and full' by Ekman and Friesen (2003). In addition, the default AUis without a corresponding alphabet and automatically classified as maximum. For example, using a hypothetical emotion AU (1+2A+5B) to illustrate; AU1 indicates pronounced and maximum intensity, AU5B has slight intensity and AU2A shows a trace of intensity. The schemes designed by Ekman and Friesen, (2003, 1978a-b) and Dragoi (2015) influence the facial actions annotation scheme indicating muscles responsible for each AU used in this thesis outlined (Table 3.7)below.

Table3.6 Facial Action Annotation Scheme (Adapted from Ekman & Friesen, 2003,1978a-b; Dragoi, 2015)

[10]https://en.wikipedia.org/wiki/Facial_Action_Coding_System

Annotation Group	Facial Action	Action Unit Number/Code	Muscular Basis
Face	Neutral face	AU0	
	Inner brow raiser	AU1; AU1B	frontalis (pars medialis)
	Outer brow raiser	AU2	frontalis (pars lateralis)
	Brow lowerer	AU4; AU4B	depressor glabellae, depressor supercilii, corrugator supercilii
	Upper lid raiser	AU5; AU5B	levator palpebrae superioris, superior tarsal muscle
	Cheek raiser	AU6; AU6B	orbicularis oculi (pars orbitalis)
	Lid tightener	AU7; AU7B	orbicularis oculi (pars palpebralis)
	Pulls or tighten lips towards each other	AU8; AU8B	the orbicularis oris muscle
	Nose wrinkler	AU9; AU9B	levator labii superioris alaeque nasi
	Lip corner puller	AU12; AU12B; R12A	zygomaticus major
	Dimpler	AU14; R14A	buccinator
	Lip corner depressor	AU15; AU15B	depressor anguli oris (also known as triangularis)
	Lower lip depressor	AU16; AU16B	depressor labii inferioris
	Chin raiser	AU17	mentalis muscle
	Lip stretcher	AU20	risorius w/ platysma
	Neck tightener	AU21	platysma
	Lip funneler	AU22	
	Lip pressor	AU24	orbicularis oris
	Lips part	AU25	depressor labii inferioris muscles
	Jaw drop	AU26	masseter; relaxed temporalis and internal pterygoid
	Mouth stretcher	AU27B	pterygoids, digastric muscle
	Dimpler	R14A	buccinator
	Lip bite	AU 32	
	Swallowed/compressed lips	AU8B	orbicularis oris
	Glabella lowerer	AU41	separate strand of AU 4:depressor glabellae (Procerus) muscle
	Inner eyebrow lowerer	AU42	separate strand of AU 4: depressor supercilii
	Eyebrow gatherer narrows eyes	AU44; AU44B	corrugator supercilii (Darwin's muscle of difficulty)
	Wink	AU46	Orbicularis oculi
	Head downward movement	AU54	
Intensity of facial action	Trace Slight Pronounced or maximum	A B C	
Eye/ocular motor movement	Saccade-left Saccade-right	AU61 AU62	Lateral rectus and medial rectus muscles; AU46 orbicularis oculi muscle;

3.3.3.2 Lip Action

The annotation for emotions also includeslips action and eye movement in relation to facial actions and task completion. Although theymay not,on their own project emotion, they may enable observers to make a better judgement. Lips may be described as puckering, protruding and pouting(Bell, 2015). The relevant lip action is licking the lips in stronger forms such as biting (Figure 3.36) and swallowing or compressing (Figure 3.37). Biting the upper lip could be a stronger form of licking while swallowing lips indicate self-restraint. Generally, biting lipssuggests nervousness, a habit, anxiety, anger, frustration or thinking.

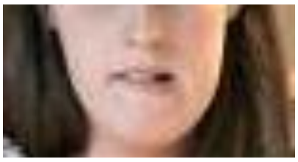


Figure3.36 Biting Lower Lip

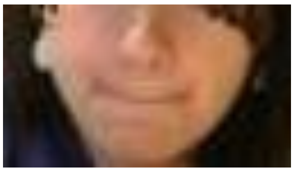


Figure3.37 Swallowed or Compressed Lips

3.3.3.3 Eye Action-Ocular Motor Movement

Eye action refers to positioning and movement used to characterise and classify how the eyes track an object when focusing on an image or object of interest (Dragoi, 2015). The relevant eye movement in this thesis is the guided saccade, described as ‘short, rapid, jerky (ballistic) movements of predetermined trajectory that direct the eyes towards visual objects or targets (Dragoi, 2015; Dell’Osso, 1994) left or right without moving the head.

3.4 Conclusion

This Chapter has outlined an overview of the framework used in Chapters 4 and 5. It examined the processes of developing corpus, addressing issues that ranged from the multimodal procedure of creating the corpusto development of an annotation scheme. The procedureinvolves setting the objective and context, recording the corpus, sampling procedures, study measures, analytical representation and re-use. Although this methodology is byno means conclusive, it has developed a research context for multimodalcorpus analysisandresearch by bringing to the fore someprimary practical and technological concerns faced when working in emerging areas such as HACs.The analytical frameworkenables actual examination of communication between instructors and instructees in HAI and HHI. The Chapterprovides a framework that enables specific, accurate and relevant analyses of interactions outlined inChapters 4 and 5.

Chapter 4 Assembly Instructions in Human-Agent Interaction

4.1Introduction

This study tries to understand the type of communication that occurs when humans take assembly instructions from agents during assembly tasks. Specifically, it uses spontaneous facial actions and gestures as lenses to understand the patterns of listenership behaviour in Human-Agent Interaction(HAI) in a unidirectional assembly instruction-giving context.This Chapter tries to answer two compound questions. What facial actions do listeners display when they ask agents to repeat instructions during assembly tasks and why?What gestures do listeners display when they ask agents to repeat instructions during assembly tasks and why?

4.2 Related Literature

This chapter covers literature specific to how listeners use instructions and spontaneous nonverbal behaviour elicited, as well as the roles they play during interaction.

4.2.1Simulated Agents

As outlined in Chapter 1.1,the agent used in this studyis simulatedwhich researchindicates isuseful for testing design options, cheaper to make when compared to embodied agents and enablesus to test conditions that may be difficult to arrange in the real world (Gomes et al., 2014; García et al., 2000).

Simulated agents use object-oriented application frameworks or modelling systems to define their behaviour and interaction patterns (Lin et al., 2015; Becket & Badler, 1993). Object-oriented framework is technology used for reifying software design and implementation at reduced costs with greater efficiency and specialisation (Fayad & Schmidt, 1997).For example, the agent in this study is developed on a specific platform (MacApp) and can only be applied to the specific domainof instruction-giving.

Simulated agents have goal-oriented behaviour meaning that, they are designed at task levelsto exhibit specific behaviours in specific contexts(Douvilleet al., 1996).Goal-oriented behaviour also refers to the designation of specific agents within a framework to simulated specific processes (Lin et al.,2015) that make it impossible for one simulated agent to function in a different context. Thus, there are trust agents(that sell ideas or provides customer service e.g. Lebara phone top up voice)and agent-instructors in business contexts (e.g. self-checkout), but trust agentscannot function as agent-instructorsor vice-versa.

García et al. (2000) suggest that simulated agents react to input with regulated spontaneity because the agent's internal state modulates its reactive behaviour and structures external behaviour.This confirms Granieri and Badler's (1995) study indicating that an agent's internal motor process nodes influence its operational motion process during interaction. For example, clicking the repeat button makes agent-instructorsrepeat instructions.

Furthermore, Granieri and Badler(1995)suggest that where there is a discrepancy between the agent's internal state and external input, the agent relyingon modes within it may display an undesired behaviour. For example, the agent-instructorin this study cannot be made to skip instructions even if users click'next instruction" when it is delivering a current instruction,because its internal mode is regulating the command.

Moreover, simulated agents are stabile in selection and persistence when executing normalised external behaviours (Douville et al., 1996). Regarding the agent used in this study, these behaviours are normalised internally as a pre-recorded database from the wav files linked to voices used to give instructions consistentlyin this study.Having described the simulated agent the next step is to outline communicative and interactional actions that occur when listenerstake assembly instructions from agents.

4.2.2 Listener Expectations in HACs

As outlined in Chapter 1.2, language use is a joint action between interlocutors where each has responsibilities and expectations. Tannen and Wallat (1987) had earlier described this phenomenon as interlocutors'prior assumptions that shapetheir knowledge schema and embody expectations before interaction. Similarly,Grice(1999),presenting the cooperative principle and maxim as it relates to logic and conversation, argues that talking is a joint cooperative transaction or quasi-tacit contract between interlocutors.

Going further, Grice (2006, 1999, 1975),viewing interaction from the current speaker's perspective, prescribes maxims of quantity, quality, relevance and manner of passing information to a co-interlocutor. However, when Grice's maxims are viewed from a listener's positionthis thesis holds that theybecome the listener's unique communicative expectations during interaction similar in some aspects to Tannen and Wallat's (1987) knowledge schema.

At another level, the Computers Are Social Actors (CASA) paradigmand Media Equation (M-E) (Nass & Moon, 2000; Reeves & Nass, 1996) focus on how people interact with computers and computerised spaces. The CASA paradigm holds that during interaction,people treat computers and computerised spaces as real people and spaces because computers call to mind similar social practices as humans (Nass et al. 1994). Similarly, M-E holds that an individual's interactions with computers and new media are fundamentally social and natural, just as in real life (Reeves & Nass, 1996)because people process interactions with computers and computerised spaces in the same way they would do in real life interaction contexts.

To demonstrate how people use social rules in HAI, CASA/M-E focus on the locus of the self or other (Nass et al., 1994). Findings suggest that, people applied social norms and notions of self and other to computers or verbal agents, just as people in HHI perceive and project emergent identities during interaction (Chapter 2.6). Furthermore, even small changes in creating perceived agent personalities could elicit social behaviours from people (Lee et al., 2012, 2000); for example, people can identify themselves as teammates with computers (Nass et al., 1996) and can be flattered by them much like they would be with other people (Fogg & Nass, 1997).

CASA/M-E have been applied in assessing human traits in automated systems like politeness (Nass et al., 1999), gain-loss theory (Moon & Nass, 1996), social facilitation (Rickenberg & Reeves, 2000), social presence, principles of similarity-attraction (Lee & Nass, 2003; Gong & Lai, 2001; Nass & Lee, 2001; 2000) role assignment (Nass et al.,1996), identity stereotyping (Straitet al., 2014;Torrey, et al., 2013;Mitchell, et al., 2011;Nass et al., 1997) self-serving bias (Moon & Nass, 1998), emotion and active listenership (Klein et al., 2002)and providing non-judgmental feedback(Johnson & Gardner, 2005).

Social behaviours are encapsulated in language use during interaction becausehumans have an elaborate language for describing people, objects, places, ideas and situations they encounter (Reeves & Nass, 1996). These include basic human communication devices such as spontaneous facial actions and gestures that are not only effective tools of social behaviour (Chapter 2.2.3) but are also powerful automatic social responses in HAI (Reeves & Nass, 1996; Nass et al., 1994). This implies that, social responses may project an interlocutor's expectations during interaction. This foreshadows the discussion on the relationship between M-E and Gricean principles later.

Furthermore, M-E holds that the natural rule stating that the presence of motion demands attention in the real world also applies to media (Reeves & Nass, 1994)because the human brain evolved into a world where only humans displayed rich social skills and all perceived objects are real physical objects, for example, if you see a face it is a real person you are looking at. Going further, Reeves and Nass, (1996) posit that the human brain has not evolved for 20th century technology, however, it is evolved to do the best things that ensure our survival. This level of evolution may be why people respond to computers in the same way that they respond to human beings during interaction.

As outlined earlier, M-E holds that people treat computers as real people and spaces during interaction however, Luger and Sellen (2016) counter Reeves and Nass (1996) and argue that humans do not respond to agents in the same way that they do to humans during social interactions. Luger and Sellen (2016) based their argument on Shechtman and Horowitz's (2003) study suggesting that during interaction with agents, people have three conversation goals, to be outlined later. While Luger and Sellen's (2016) argument is plausible however, the premise that people engage more with other people than with agents.It is the opinion of this thesis that their conclusion ties in more with Shechtman and Horowitz's (2003) finding suggesting that people treated agents differently from other people because they perceived agents as different from them, which is othering (Chapter 2.6).And as outlined in chapter, attitudes and identities are developed performatively so, it is also the opinion of this thesis that people may not interact with agents as they do with humans, but this may depend on factors beyond attitude formation that are context- dependent.

However, Shechtman and Horowitz’s (2003) finding that interlocutors have conversational goals ties in with Grice’s (1975) principles of conversational implicature and illustrates the relationship between M-E and Gricean principlesbecause both recognise that discourse is organised around aims and the strategies interlocutors develop to achieve these aims.

Furthermore, Shechtman and Horowitz’s (2003) hold that during interaction human behaviour has three main goals including task, communication and relationship goals. This agrees with Grice’s (2006, 1999, 1975) cooperative principle stating that, in conversation interlocutors have common immediate goals.

Task goals specify the purpose of an activity that interlocutors are jointly involved in. For example, in this thesis, listener expectation may include a reciprocal understanding from the instructor (at least in HHI) that the aim of interaction is to assemble Lego models.

Communication goals ensure that exchanging ideas and information between interlocutors goes on smoothly and without misunderstanding and this to an extent exemplifies Grice’s (1999) maxims for listeners. Listeners expect that instructions are adequately informative to aid task execution (Grice’s maxim of quantity); genuine (Grice’s maxim of quality); clearly stated without ambiguity (aspects of Grice’s maxim of manner) and relevant to the purpose (Grice’s maxim of relation).

Relationship goals ensure that interlocutors set and keep the tone of interaction. The relationship goal exemplifies Grice’s (1999) assertion that there is a tacit understanding between interlocutors that transactions should continue in an appropriate style. Listener expectation is that interaction may be orderly and performed with reasonable dispatch (Grice’s maxim of manner) where interlocutors’ contributions may dovetail but remain mutually independent because of their discursive roles (Chapter 2.6).

There is another relationship between M-E and Griceanprinciples based on the suggestion that people perceive computers as having human-like personalities and slight changes in personality elicits social behaviour from people. This relationship is intertwined in the link between Grice’s Cooperative Principle (CP), maxims and conversational implicature as explained by Grice (2006).

M-E suggests that humans react socially to changes in computer personality and in parallel, Grice's explanations are viewed from listeners' perspectives. When a speaker fails tofulfil a maxim, for example, not being as informative as required or givingmore information than is needed, comprehension is hampered, and listeners may become uncooperative thus violating the principle of relationship.

In addition, when listeners are faced with a clash in interpreting speaker utterance because of obscure, unclear or irrelevant expressions, the flow of discourse becomes disrupted and listeners may become unable to carry out responsibilities required (such as executing instructions)to achieve interaction aims. The listener's inability to makethe same meaning with the speaker and subsequent disruption of interaction are social behaviours occasioned bya speaker's lack of clarity which creates conversational implicature (Grice, 2006, 1999).

Conversational implicature describes what is meant by a speaker's utterance even though not explicitly said or inferred. Clark et al (2016a) observed that speakers using vague expressions may create obstacles such as interaction distance for co-interlocutors. However, the listener, during social interaction as implied by M-E resolves conversational implicature using what Grice (2006) describes as conventional meaning of words, linguistic contexts, background knowledge, and interaction assumptions as outlined in Chapters 2.2.1 and 2.2.2.

M-E has been criticised for poor handling of scientific issues, not considering the demand characteristics of the interaction context, and making inadequate disclosure regarding its research procedures thus, creating ethical issues for the study (Brotherton, 1999). However, Zillner, 92000) suggests that M-E provides a methodology and technique for making user interactions with automated systems a better experience both in terms of task efficiency and user satisfaction. Following these, this thesis believes that it is reasonably safe to suggest that M-E may present a framework for characterising and understanding listener expectations in HACs.

4.3 Study Methodology

4.3.1 Agent Design

A simulated agent was created on a computer interface for the study instead of a real agent because it provides users with experiences similar to that which actual agents provide (Clark et al, 2014). The interface (Figure4.1 below) issued instructions to participants from a Hypertext Mark-up Language (HTML) file connected to a database of assembly instructions on .wav files.

Figure4.1 The user interface for both models

The interfaces show the Lego models 'Aquagon and Nex' obtained from the Hero range of Lego models to help the listener keep focus on the model. The progress indicator shows how many instructions the participant has taken and how many are remaining, for example, step 3 of 47. At the beginning, the participant presses the 'start button' and subsequently, this becomes the 'next' button used to get succeeding instructions. The repeat button used to repeat instructions and the interface keeps record of the number of repeats; however, there is no button to go back to a previous instruction as a condition imposed to ascertain self-propelled behaviours in participants.

The instructions for both models and synthesised voices were inputted into a text-to-speech program (Text2SpeechPro) and exported as .wav files. As outlined in 3.2.1, the agent uses a range of voices that include Style 1 CL, Style 2 is CP, and HR. HR or a voice actor file was sent as one package and edited with Audacity[11].

CL is produced by Cepstral an online Text-To-Speech (TTS) voice builders from www.textspeechpro.com. CL is advertised as more cultured, personal and sophisticated than other TTS voices[12] CL is the synthesised voice of a man aged between 40 – 55 years using RP accent without capacity for expressing emotions.

From CereProc[13] voices another online TTS vendor that offers a range of voices in many accents, CP was chosen from https://www.cereproc.com. CP is the synthesised

---

[11]More information http://www.audacityteam.org

[12] More information http://www.cepstral.com.

[13] More information https://www.cereproc.com/

voice of a man aged between 40 – 55 years, using RP accent that can simulate a range of emotions except anger, fear, depression, despair, exhilaration, serenity and bliss. The human-like voice is HR produced by a professional voice actor hired from http://voicebunny.com. The voice is identified as Mark[14] a man aged between 40 – 55 years, using RP accent with a full range of human sound production however, its humaness is limited because it is recorded or mediatised.

4.3.2 Participants

Forty-eight speakers of English as a first language were recruited for this study as outlined in Chapter 3.2.2 and rewarded with £10.00 Amazon voucher remuneration for participationafter the sessions. Twenty-one participants were male (23.7%) and twenty-seven were female (56.3%) with a mean age of 24.2 years.

Table4.1 Participants Randomised Groupings

Grouping and Counter Balancing Checklist					
A - CP + CL		**B - CP + HR**		**C - CL + HR**	
A1 - CP > CL		*B1 - CP > HR*		*C1 - CL > HR*	
A1 AN (1)	A1 NA (2)	B1 AN (5)	B1 NA (6)	C1 AN (9)	C1 NA (10)
IIII	IIII	IIII	IIII	IIII	IIII
A2 - CL > CP		*B2 - HR > CP*		*C2 - HR > CL*	
A2 AN (3)	A2 NA (4)	B2 AN (7)	B2 NA (8)	C2 AN (11)	C2 NA (12)
IIII	IIII	IIII	IIII	IIII	IIII

Each group was divided into two sub-groups(Table 4.1)shown as A1 and A2. Participants in A1 interacted with voice CP in the first task and CL in the second while those in A2 interactedwith voice CL first and CP later. The sub-groups are split into two cells with four (4) participants each making 12 cells in all. The task is further counter-balanced using Lego models, for example, cell ‘A1 AN’ will assemble Lego model Aquagon (A) first then Lego model Nex(N).

4.3.3 Experiment Procedure

[14] Voice can be heard http://voicebunny.com/projects/add_booking/5KKMKQC/856091

The study began with a pilot study involving one participant executing 46 instructions. Although, using one participant for pilot study may seem inadequate but, it enabled me to have a step-by-step view of interaction in instruction-giving contexts (Zelkowitz & Wallace, 1998), test preliminary hypothesesto develop more precise hypotheses and fine-tune data-gathering instruments and process (Thabane et al., 2010) in the main study.

The study used a personalised approach where each participant received instructions from the agent-instructor and the assembly task procedure is as follows:

Pre-Task Activities: these are all the preparatory activities that take place before interaction between the agent and participant occurs.Cameras and agent interface-laptop are set up as required. One camera is set to record the session from the front and the side angle camera set to record from the participant's left (Chapter 3.2).

As participants enter the laboratory, they are introduced to the researcher, and seated at the location they will be using for the task. They are briefedon what the entire session entails and given details on the smaller elements such as these are verbal instructions that they can repeat as many times as they want but they cannot go to a previous instruction. In compliance with the ethics approval granted participants are informed that video-recording equipment is used with their permission granted later in the signing of the consent form (Appendix IIb).

The information form (Appendix I) is presented to the participants for reading. Once they have read this they are asked if they have any more questions or need clarification regarding their involvement in the study.

The consent form (Appendix IIb)for participation is then presented to them for signing and it is explained to them that they are free to withdraw from the experiment at will.All but one participant gave consent for their interaction to be recorded on video but all of them agreed that the data obtained could be used showing their faces in the thesis, conferences, and academic publications. The task time-limit is divulged to participants and they are told that when it expires, the interaction will stop even if the participant has not completed the task.

The actual task takes place here. Participants are introduced to Task 1 and the interface (Figure 4.1) they will be interacting with and all available options they can use within tasks.

The participant undergoes Task 1 to completion or when time is over, and Lego Model 1 is taken away.The participant is then introduced to Task 2 by repeating the same explanation for Task 1. Then participants undergo Task 2 to completion or when the time is over, whichever comes first.

Participants are then told the details of remuneration, thereafter any other questions or business that may arise is taken care of and participants are free to leave.

Post-Task procedures begin thereafter. First data from recordings is uploaded to the secure ORCHID server then wiped from the equipment, then the Lego models are disassembled and pieces put back into place for the next study or taken back to storage in the office and pictures of the models as assembled by participants are also uploaded unto the ORCHID server and filed away in secured cabinets.

4.3.4 Data Collection and Analysis

Each interaction is recorded from two different angles. A Panasonic HDC-SD900 captured eye level shots of participants and a Canon Legria HFR306 recorded interaction shots from a slanted left angle side to capture both nuances of interaction with the interface and model assembly. Although each camera has the capability to record in full high definition, early trials showed the file sizes to be too large for storage so the smaller .MP4 format was used as a substitute without compromising greatly on quality.

Figure4.2 Front and side level shots showing participant during assembly task

Marked Interaction Points

As already outlined in 3.2.4, marked interaction points are specific occurrences when listeners asked the agent to repeat instructions during task. The instructions with the highest number of repeats are selected. For example, Aquagon, instruction 14 and 16 emerged and for Nex, it was 24 and 25 as shown (Table 3.) below.

Table4.2 Most Repeated Agent Instructions as critical interaction points

Lego Model	Instruction Number	Instruction	Action Required
Aquagon	14	Now take a black cylinder, a grey cylinder and a small light grey piece with a curved thing that looks a bit like a fin	Participants are expected to pick assembly pieces
	16	Now just attach the grey piece with the fin to this cylinder. The end that looks a bit wider should be closest to the body.	Participants are expected to Join or attach assembly pieces. Locate the pieces in a specific 3-D position
Nex	24	Now locate the small, thinnest black piece and a small orange armour piece.	Participants are expected to join or attach assembly pieces together
	25	Just attach the orange piece to the grey socket on the right arm. The end with the holes should be closest to the head	Participants are expected to attach pieces in a specific location with a specific orientation

And where a participant did not repeat any of these, the instructions most repeated by that participant during the tasks were picked. This level of systematic sampling ensured that every assembly stage was included just as every participant within the sample frame was represented and analysed.

4.3.5 Decisions on Research Hypotheses

For IRR analysis six (6) facial action categories and thirteen (13) gesture categories were sampled respectively for the Inter-Rater Reliability assessment (IRR). The percentage agreement between the two annotators for spontaneous listener facial actions displayed during interaction is 100% and gestures is 86.67%. The resulting Kappas indicating very good or almost perfect agreement,fall within Landis and Koch's benchmark of .81- 1.00 and Altman's benchmark of .81- 1.00 respectively (Gwet, 2012).

The IRR coefficient correlation for gestures displayed by participants indexed by Fleiss Kappa (α) is 0.79 while the Krippendorff alpha (α) is 0.80 implying substantial and excellent agreement which agrees with the percentage index of .87. Equally, the IRR coefficient for facial actions displayed by participants indexed by the Fleiss kappa (*K*) and Krippendorff alpha (α) is 1.00, implying an almost perfect agreement which is the same as the percentage index of 100%. The values obtained from percentage agreement, the Fleiss Kappa and Krippendorff alpha indicate that there is a very high agreement between the first and second annotators' perceptions of facial actions and gestures displayed by listeners during interaction.

Regarding H_{1a}, spontaneous listener facial actions analysed from HAI indicate that changes in listener attitudes and moods during tasks may be occasioned by how far an instructor's utterance promotes or hinders a listener's ability to successfully decode instructions during tasks. Thus, when listeners successfully decode instructions and execute tasks, they display positive facial actions suggesting positive attitudes but when there is a hindrance leading to failure to decode instructions and execute tasks, listeners display negative facial actions representing negative attitude. When listeners need to concentrate on either task or instruction comprehension, they display neutral faces; however, listeners also display positive facial actions even when they experience failure.

These suggest that there is potentially a cause-effect relationship between listener facial actions and listener successful execution of verbal assembly instructions except where listeners get sudden inspiration within the Aha! Moment. In that exceptional moment, listener facial actions may be driven by their own abilities to think outside the box and solve unexpected problems during the task. Thus, except for instances where listeners experience Aha! Moments or smile in the face of failure, there is sufficient evidence to support the claim that there is a relationship between listener facial actions and their successful execution of the instructor's verbal instructions.

Regarding H_{2a}, the analyses conducted suggest that listener gestures have a multifaceted relationship with their successful execution of instructor's verbal instructions. The first relationship is interactional with the listener being in control, particularly when requesting repeat instructions and managing turn-taking. The second is task-oriented where listener gestures focus on doing what instructors say. The third is communicative where listeners externalise their thought processes as they cognitively process instructions and initiate self-initiated self-repairs during tasks. Given these relationships, there is sufficient evidence to support the hypothesis that there is a relationship between the listener's gestures and whether they successfully execute the agent's verbal instructions and thus the hypothesis is accepted.

4.4Results

As outlined in the introduction, this study tries to understandthe pattern of interaction, facial actions and assembly gestures emerging from the joint action between the simulated agent and participants during assembly tasks.

Presentation of resultsis structured to begin with quantitative then qualitative results. Quantitative results cover the number of repeats and facial actions per interaction.Qualitative results cover interaction patterns (repeat process, typology of repeats), families of facial actions and assembly gestures displayed during tasks.As outlined in 3.3, presentation of results involves detailed analysis of emerging phenomena that are representativeof facial actions and gesture families stating their characteristics, communicative or interactional functions, implications for the Chapter as well as comparison with other phenomena as done in Chu and Kita (2016), Kendon (2004), Ekman and Friesen (1978a-b).

4.4.1. Quantitative Results

Instructions Repeated

As outlined in 3.2.3, a random selection of individual instructions or specific time marking was not used because there was no assurance that either could be adequately covered during interaction. For example, some participants could not cover all the assembly steps within the time limit but a good percentage repeated instructions. Table 4.3below shows that participants received a total of 291 instructions from agents in the first instance but asked that some of the instructions be repeated and the repeats totalled 667 times. There are 144 instructions for Aquagon and 147 for Nex models; 48 participants with two iterations each.

Table4.3 Agent Instructions Repeated

	INSTRUCTIONS				**REPEATS**			
VOICE	CP	CL	HR	TOTAL	CP	CL	HR	TOTAL
AQUAGON	48	48	48	144	116	109	75	300
NEX	49	49	49	147	138	117	112	367
TOTAL	**97**	**97**	**97**	**291**	**254**	**226**	**187**	**667**

KEY: CP= Cere Proc; CL= Cepstral Lawrence; HR= Voice Actor

Based on these, the ratio of repeats to instruction for each voice/agent is CP 3:1; CL 2:1; and HR 2:1 while the ratio for total repeats and instructions is approximately 2:1. This implies that for every instruction given, participants requested that it be repeated twice. Of these, Aquagon had 300 repeats representing approximately 46% while Nex had 56%. In addition, the voice with the most repeats is CP with 254 repeats or 38%; CL with 226 repeats or 34% and HR with 187 repeats or 28% of the distribution. These results suggest that the most comprehensible voice is HR with the least repeats while CP is the most incomprehensible. However, it is the opinion of this thesis that the number of repetition observed here may not be matched in HHI because of the effect of face threats are absent in HAI since the agent is simulated and unresponsive. Face wants specifically the desire to avoid any action that may lead to lose of self-esteem or appreciation form others is ever present in face-to-face interactions with humans (outlined in 6.3).

Table 4.3 above shows that the most repeated instructions are 14, 15, 16, 25, and 26 for Aquagon while 8, 24, 25, 27, 28, 29, 38, 43, and 46 are most repeated for Nex. Instruction number 25 is the most repeated across the models and voices, which indicates that it may be the most difficult to decipher in the distribution.

As outlined in 4.4.3, listeners displayedtwo hundred and forty-seven (Table 6.2) facial actions and these reflect positive, neutral, and negative attitudes towards the agent during interaction (Figure4.3)below.

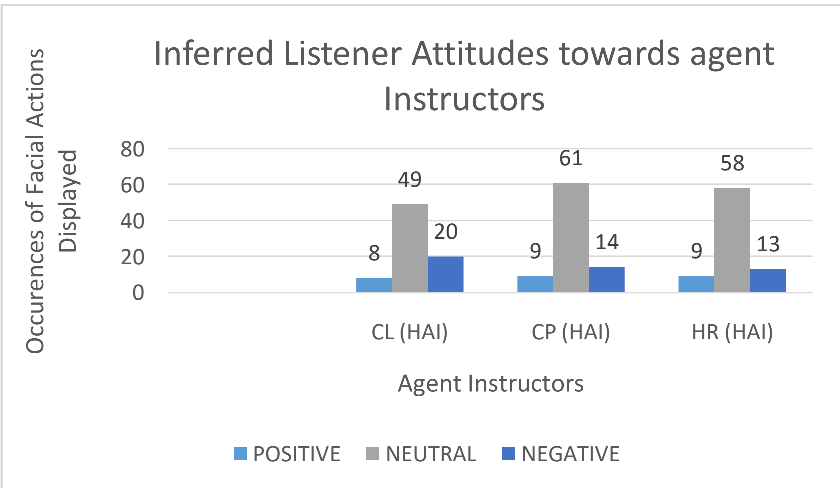


Figure4.3 Inference of Listener Attitude from Facial Actions

Positive facial actions include felt smile, slight smile, George W Bush grin, tight-lipped smile, and micro-smile representing 11% of the distribution. Neutral includes neutral, neutral concentration-Instructions, neutral concentration-task, neutral intense, neutral face –hard, neutral face -down, and workman face representing 73%. Negative facial actions represent 16% of the distribution which includecontrolled fear, disgust, slight disgust, angry-disgust, micro-anger, micro-sadness, compressed or swallowed lips,frown and nervous smiles.

The preponderance of neutral facial actions in the distribution suggests that participants for most of the time may not be judgmental towards the agent or interaction rather, they were concentrating on the task, cognitively processing the instructions or were just indifferent. The neutral face may also suggest that the listener is stoic by nature so can maintain a blank face during interaction however, this is beyond the scope of the present study.

The next dominant attitude is negative with 16% suggesting that participants had issues with the interaction that generated fear, disgust, and sadness however, participants were able to make the best of a bad situation as suggested by nervous smile. Going further, positive facial actions suggest that participants enjoyed interacting with the agent instructors. There is the need to compare these facial actions with those that occur in human-human interaction to get a sense of the level of similarities and differences.

4.4.2 Qualitative Results

Although, interpretation of nonverbal behaviour is influenced by interaction context, culture and personal experiences, asoutlined in chapter 3, the annotation scheme (3.3.2) provides a standard interpretation manual that ensures some level of accuracy. Furthermore, only samples that are representative of facial actions and gesture families are presented within this Chapter because they have all the basic features of each category while the rest are in the appendices (Appendix VIV) and are cross-referenced in the analysis and discussions. The results presented at this stageare descriptivewith some specific conclusions regarding individual nonverbal interaction behaviours, butgeneral conclusions and implications are presented later with comparative results (Chapter 6).

4.4.2.1 The Process of Repeating Instructions

Interaction is assessed using listener request repeats of instructionsbecause repetition is an important aspect of human interaction that is tied to linguistic theories of repairs (Schegloff et al.,1977),politeness (Brown & Levinson, 1987) task performance (Kim& Tracy-Ventura,2013) and listenership (Tsuchiya, 2013). Repetition is a pattern that is potentially present in language in various forms that users employ to project their way of seeing things and create greater mutuality between them (Carter, 2004); however, language users have a choice whether to form such patterns.

The results suggest that the process of repeating instructions in HAI is a discernible, systematic, consistent and continuous chain of events that has three major stages; initiation, execution and resolution (Figure 4.4overleaf). When participants execute an instruction, for example, No.1, they move on to No.2 which is the next one as 'current instruction' but before they move to the current one, they must initiate a request using the 'Next' button.

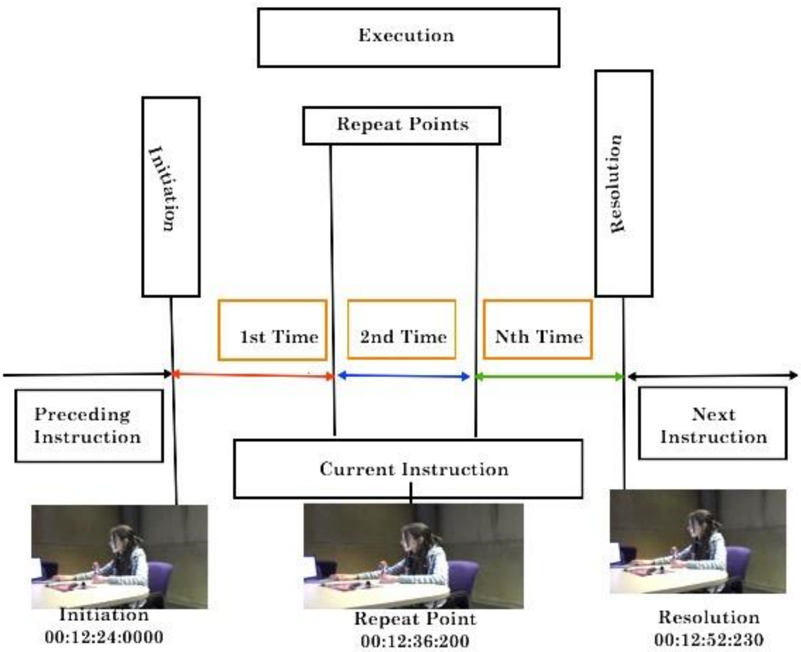


Figure4.4 The Process of Asking Agents to Repeat Instructions

When participants ask agents to repeat instruction number 2 the first repeat point is reached thereafter, everyrepeat of instruction 2 creates new repeat points until the instructionis executedor said to have been resolved when they ask for instruction number 3. The results indicate that when instructions are repeated there is a process with distinct features outlined overleaf.

Repeats are discernible: repeats are noticeable and observable when they occur during interaction. For example, the user pressing the ‘Repeat’ button on the interface or the instructor re-vocalising the same instruction more than once makes such instruction audible to listeners.This is similar to printing an electronic document to make it readable ashardcopy.

Repeats are systematic: this means that repeats are organised, well-ordered and are planned actions by listeners to achieve a given aim. For example, initiation comes before execution not vice-versa.

Repeats are consistent in nature: this means that repeats are unchanging, reliable and rational communication strategies that have been found to promote active listening (Oxford, 1993) and set a communication standard (Stack, 2012). For example, when listeners request for a repeat, the agent repeats the whole instruction without adding or subtracting anything such that even when repetitions occur in error, they always serve communication functions as will be seen in the typology of repeats (Chapter 4.4.2.2).

Repeats involve a continuous chain of events: The repeat process in HAI begins with initiation, which occurs when the listener presses the ‘Next button’ for the agent to give them a subsequent instruction for the first time. Execution is done with current instructions and relates to the actual demand for repeats by the listener. Repeats start to count after the current instruction has been given and the listener presses the ‘Repeat button’ at intervals called ‘repeat points’. The repeat point is achieved when the agent finishes relaying the current instruction and the listener asks for it to be repeated and the agent starts all over again. The instruction is said to be resolved when the listener asks for the ‘Next instruction’ after executing the current instruction.

4.4.2.2Sequence and Typology of Repeats

As outlined in Chapter 4.4.1, repetition is a pattern that is potentially present in language that language users need to project their way of seeing things and creating affiliations between them. The resultssuggest that people repeat instructions for various reasonswhich enable classification of repeats based on timing, listener action and reasons deduced from observing and analysing the interaction context during interaction using the model outlined in 3.3.1.3.

Timing is linked to the chain of events or repeat sequence and presupposes thatdifferent types of repeats occur at different timesduring interaction in assembly contexts.Furthermore, timing definesthe sequence of repetitionsalong the task timelinewhile the sequence(Figure 4.4) indicates that repeats occur at three different times during the task. Some repeats occur before assembly action, some co-occurring with assembly action while others occur after the initial instruction or the assembly action has taken place. I will come back to this after outlining repeat typologies.

Listeners' actionsare linked to timing and include the activities going on and listeners'nonverbal feedback while the agent-instructor is repeating instructions such as expressing surprise (Figure 4.12). It may have some verbal components and social signals, but these may be covered by future research. Listener action is observable,whichmakes it possible to see ifa listener is examining work done, working as the instruction is issued, delaying instructions, or just listening during interaction.

A listener's own reason for repeating instructions usually comes in the form of nested narratives, semi-structured interviews and written feedback (Clark, 2016) after the task. However, research indicates that people have a tendency to forget exact details of interaction after they have taken place and they tend to re-construct events and this may lead to inaccuracies (Bach & Goncalves, 2004; Norman, 1988). Thus, this study relied more on listener facial actions, gestures, general body language task at hand and interaction context for information during classification using the model outlined in 3.3.1.3.

The results indicate that nine (9) typology of repeats emerged from the interactions(Table 4.4overleaf). Each type is given a code for easy identification; for example, R1 is interpreted thus 'R' = repeat; 1 = serial number; 1a = subclass 'a'. R1, R3, and R4 have two, five and six subclasses respectively while the rest do not have any subclass. The typology suggests that listeners use repeats for finely grained purposes ranging from self-correction, through clearing doubts, separating the assembly process, mitigating the effects of information overload to simple errors as outlined below and (Table 4.4) overleaf.

R1 type of repetitionis meant for clarification that may lead to user self-correction and has two sub-classes (R1a & R1b). With R1a, clarification occurs before listeners execute the current instruction/assembly task leading to self-correction. With R1b, after repeating instruction for clarification, a task error occurs which listenerscorrect immediately. For example, a listener repeats the instruction, then attaches pieceswith the wrong orientation, then detaches them then sets the orientation right.

Table4.4 Typology of Repeats in HAI

Type	CODE	GENERAL FUNCTION	SUB-CODE	SPECIFIC FUNCTION
Clarification	R1	For self-correction	1a	Clarification leading to self-correction
			1b	After clarification, user makes a task error then self–corrects, for example, picks the wrong piece the drops it for the correct one
Clarification	R2	Used to reduce confusion and clear doubts. Usually occurs after listening to the instruction.	2a	Clarification for comprehension
Confirmation	R3	Usually occurs after listening or executing an assembly process, it affirms or disproves user interpretation of agent instruction, assembly action taken and helps user self-assurance	3a	To confirm assistive action taken, for example, if the correct piece was picked
			3b	To confirm if the correct operative action was done or if the correct assembly procedure was followed like joining pieces together.
			3c	used for confirmation of assistive action taken leading to self-correction, for example, picking pieces confirming then self-correcting the action
			3d	Used for confirmation of assistive action taken leading to self-correction of operative action, for example, picking up pieces, confirming then attaching them.
			3e	clarification then confirmation of assembly procedure e.g. task stages
Composite	R4	Multiple repeats for different purposes but with a focus on executing one assembly step	4a	Repeats confirm actions when they occur after the action has taken place and the listener/user can be seen inspecting the work done while the instruction is being repeated
			4b	Used for clarification, then self-correction and step by step confirmation
			4c	The first for clarification; the second for self-correction and the third is for confirmation of assembly procedure like joining pieces together
			4d	Used to demarcate and clarify the assembly procedure and refocus listener's attention for easy compliance then confirm action taken like joining pieces together
			4e	Used to demarcate the assembly process then confirm the action

				taken
			4f	Used to confirm action and self-correct
Demarcation	R5	Used to break up one assembly stage into chunks		
Confirmation and demarcation	R6	Occurs intermittently, the first may affirm or disprove listener interpretation or action while subsequent ones break up one assembly stage into chunks		
Clarification and demarcation	R7	Used the first to reduce confusion, clear doubts and break up one assembly stage into chunks		
Refocusing	R8	Redirects user attention to critical aspects of the task		
Error	R9	A simple listener mistake of repeating instructions		

R2 types are clarification repeats that usually occur after listening to first time instructions. They are characterised by lack of any follow-up assembly action,which suggests that they are used to clear doubts, reduce confusion and enhance listeners’ comprehension.

R3 are confirmation repeats that usually occur after listening to an instruction or carrying out an assembly task. R3 enable users to assess their comprehension and interpretation of agent instruction in comparison to the instructor’s intended meaning.R3 is also used to assess whether the action taken by the user is correct or wrong. If these are affirmed, then user self-assurance and confidence level may improve.

R3a) are used to confirm assistive actions taken; for example, if the correct piece was selected; R3b) enable listeners to confirm if the correct operative action or assembly procedure was followed like joining pieces together; R3c) are used for confirmation of assistive action taken leading to self-correction; for example, picking uppieces, confirming then dropping if the wrong piece is selected; R3d) are used for confirmation of assistive action taken leading to self-correction of operative action. For example, picking pieces confirming then assembling; R3e) areused for clarification then confirmation of assembly procedure such as moving from assembling the legs to lower body (see Table 3.1).

R4s are multiple repetitions that can serve six different purposes but with a focus on executing one assembly step. These types of repeats are used strategically to aid listener’s execution of instruction. R4a) This subclass of repeats confirm actions when they occur after the action has taken place and the listener can be seen inspecting work done while the instruction is repeated; R4b) listeners use this subclass sequentially for clarification, then self-correction and step-by-step confirmation; R4c) are multiple repeats where the first is used for clarification; the second for self-correction andthe third is for confirmation of assembly procedure like joining pieces together; R4d)are used to demarcate and clarify assembly procedure in the first instance. Thereafter they are used to refocus a listener's attention for easy compliance and finally to confirm listener action taken; R4e) are used to demarcate assembly processesthen confirm action taken; R4f) are used to confirm action and carry out self-correction.

R5 are repeats used for task demarcation. Listeners use these types of repeats to separate each assembly stage into finely grained chunks to check information overload (Bomann & Jones, 2003) and simplify assembly task.

R6 are confirmation and task demarcation repeats. They occur intermittently between listening and task execution. The first assesses listener interpretation of agent instruction and meaning or task action taken. The others separate different stages of the assembly process into finely grained, manageable chunks.

R7 repeats involve clarification and task demarcation. The first repeat occurs without any listener assembly action while subsequent ones occur as needed by listeners during assembly tasks. They tend to reduce or check listener confusion and separate different stages of the assembly process as listeners try to mitigate the effects of information overload.

R8 repeats refocus and redirect listener attention to critical aspects of instructions. Listeners repeat instructions but listen for specific information in the instruction.

R9 repeats occur in error and enable co-interlocutors to observe listener reaction.

As outlined when discussing timing, repetition is sequential, andFigure 4.5below indicates that users ask for some types of repeats before the assembly action or receiving the next instruction; for example, R1a, R4b, R7, and R9.Some repetitions co-occur with the assembly action in which case they are strategic; for example, R1b, R4d, R4e; R5, R6, R7, R8, and R9 (Figure 4.5)below.

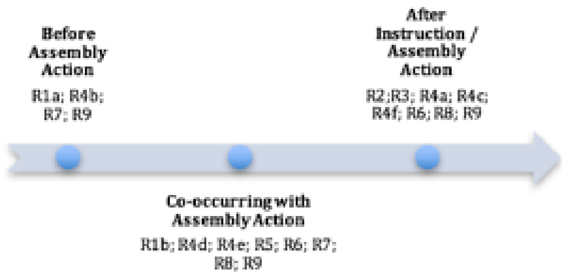


Figure4.5 The sequence of Asking for Instruction repeats

Others are confirmatory and occur after the assembly action has taken place or the current instruction has been received; for example, R2, R3, R4a, R4c, R4f, R6, R8, and R9 (Figure 4.5) above. Furthermore, some are just ubiquitous errors -R9 type that occur at any point in time during the interaction.Although, repeats occur individually but they can be and are often combined due to timing, and listener actions.As per functions, repeats that occur before execution are meant for clarification and checking confusion. Those that co-occur with execution are used as strategic guides by listeners especially when instructions are complex, while those that occur after execution are used for assessing listener actions and some are errors that serve no purpose.

4.4.3Listener Facial Actions

This section answers the question about what facial actions listeners display when receiving instructions from agents during assembly tasks.The results show that listeners displayedtwo hundred and forty-seven (Table 6.2) facial actions. The chapteridentifies, describes and classifies facial actionsusing FACSoutlined in Table 3.5.As outlined in 3.2.4, description entails stating the AUs as action descriptors with underlying facial muscles responsible for facial actions.In addition, the circumstances sorrounding the occurrence of facial actions determine its meaning and functions (see 3.3.1.3). Presentation shows the facial action family and its illustration. Illustration includesfacial action child, sibling buckets (e.g. variations of neutral faces) and theirinteractive functions as outlinedbelow.

4.4.3.1 Basic Emotions

Participants displayed basic facial actions such as neutral, surprise, smile, fear, and disgust.

Neutral Faces

Neutral face (AU0, Table 3.7) is expressionless with all muscles relaxed (Ekman, 2007). There are five forms of neutral expressions described (Figure 4.6)becausethe results suggest that the listener's face is never expressionless because of involvement in task and information processing.

Figure 4.6: Neutral Face

The first type of neutral face (Figure 4.6) above shows listeners'faces without any emotion because facial muscles are at rest as AU0(Ekman, 1982b) when they are about asking for repeat or next instructions. The second type of neutral face co-occurs with listeners leaning forward (appendix VIV.2).Another sibling is the neutral face with a hard expression (appendix VIV.3) while another occurs when the participants hands are busy (appendix VIV.4) sometimes with partially furrowed forehead produced by Darwin's muscle of difficulty. These neutral faces suggest that the participants are concentrating on instructions, task or multitasking.

Surprise

Surprise (Figure4.7) overleafis often triggered by something unexpected such as a sudden loud noise or sighting an unexpected phenomenon and Ekman (2003)says that it is very brief probably lasting a few microseconds. For example, surprise is probably triggered by CP’s voice when giving instruction to the participant for the very first time. The following muscles (Figure 3.33; Table 3.7) make the expression possible:AU1 raised the inner brow, AU2 raised the outer brow, AU5B slightly raised the upper lip while AU26 made the jaw to drop down (Ekman, 1997).

Figure 4.7: Surprise

Research suggests that, direction of gaze may indicate where the source of stimulus is located (Hadjikhani et al., 2008; Ambadaret al., 2009). In Figure 4.7 above, the participant displayed fixed gaze towards the probable source of surprise in the task layout.Another variation is the slight surprise that occurs with a lower intensity (Appendix VIV.6b).

Smiles

The results indicate that participants displayed felt (Figure 4.8), nervous, G.W. Bush grin and tight-lipped smiles during interaction. (Appendix VIV.7, 8, & 9)

Felt Smiles

The felt or full smile also called the Duchenne smile is a positive emotion (Figure 4.8) elicited by positive visual, gustatory, kinaesthetic, physical, verbal and tactile stimulation(Ekman et al., 1972) such asamusement, delight, contentment, satisfaction, beatific experiences, relief from pain, pressure or tension and enjoyment of anotherperson(Ekman,1982a-b). In addition, a felt smile suggests that participants are enjoying and liking the interaction, (Pease & Pease, 2004) or just found the agents' voices funny.The felt smile is the opposite of a false smilethat deliberately attempts to project a positive emotion that does not exist (Ekman & Friesen, 2003).

Figure 4.8: Felt smile

The facial movements (Figure 3.33; Table 3.7) responsible are as follows. AU 6, raises the cheek, gathers the skin around the eyes inwards, narrows the eyes apertures and produces crow's feet wrinkles. AU12 pulls the lips sideways exposing the teeth and AU7 tightens the eye lids, raises the lower eye lid creating wrinkles below the lower eye lid. The felt smile is a positive face (Tipples et al., 2002) that pulls back both the mouth and the eyes (Malik, 2010) with maximum intensity and in some cases leading to laughter.The nervous is superimposed on negative emotions; G.W. Bush grin is a smirk indicated by the sad element becomes visible when the upper part of the face is covered and tight-lipped smile is signalled by stretched lips during interaction (Appendix VIV.7, 8, & 9).

Controlled Fear

Laboratory studies suggest that fear may be inborn in organisms not learntbecause laboratory rats that had never seen a snake displayed fear when confronted by one for the first time (Ekman & Friesen, 2003).The clue for controlled fear (Figure 4.9)below is stretched and downwards-shaped lips. The action units (Table 3.5) responsible for this emotion include AU 1 Inner brow raiser, AU 2 the Outer brow raisers, AU4brow lowerer, AU5upper lid raiser, AU7 lip tightener, AU20lip Stretcherand AU26Jaw dropper (Ekman, 1997).

Figure 4.9: Controlled Fear

However, Ekman (2003) suggests that people can control their emotions if they successfully take charge of the triggers for aspecific emotion. For example, the participant in Figure 4.9above may have experienced a similar situation-handling fragile materials- in the past and developed an adaptive affective style to control fear of destroying them because conditioned fear learning is a resilient indelible form of learning (Ekman, 2007) that relies on emotional triggers.

Slight Disgust

Ekman (2003) describes disgust (Figure 4.10&Appendix VIV.10)as "... a feeling of aversion"(Ekman 2003, p.190*)* that may be triggered by smell, taste, sight, thought and touch of something offensive.Disgust may also be ignited by encountering a hated person, location or experience.

Figure 4.10: Slight Disgust

This emotion is given by the raised upper and lower lips with the lower lip protruding slightly. Nostril wings are raised slightly with wrinkles appearing on the sides and bridge of the nose with eyebrows pulled down. These actionsare produced by the slight contraction of AU9 that wrinkles the nose, AU15B the lip corner depressor, AU16B the lower lip depressor and AU44B a separate strand of AU 4 that visibly narrows the eyes when the head is bent (Table 3.7).

4.4.3.2 Emotional Attitude

Frown

Ekman (2003) describes a frown (Figure 4.11 & appendix VIV.11) as an emotional attitude produced primarily by the furrowing of the eyebrows through the action of AU 44brow gatherer also called the muscle of difficulty because frowning occurs with many kinds of difficulties, mental or physical (Mutlu, 2011). AU 46 lowers eyelids as done in a wink,frowns could be executed with the neck jerking forward.

Figure 4.11: Frown

Additional muscles that produce the frown include AU 24 lip pressor, AU 21 that produces a slight wrinkling of the surface of the skin of the neck in an oblique direction(Standring et al., 2008), AU 41that creates a horizontal wrinkle across the bridge of the nose, AU17 chin raiser, and AU15 lip corner depressor.Participants in Figure 4.20above,are looking down while one has fingers on the chin suggesting that shemay beconcentrating(Ekman, 1977) or cognitively processing the instruction.Ekman (2007) explains that when people frown, they are often perceived by others to be feeling unpleasant, resentful, or angry, although this may not be the case. This is because awrinkled forehead may also indicate displeasure, sadness or worry, anger or less often confusion or concentration.

4.4.3.3 Non-Basic Emotions

Non-basic emotions mostly comprise different blends of emotions. The process of formation is described as one emotion running into another to produce another (Ekman, 1977) in a manner like diphthongs where one sound runs into another and sounds as one at the point of articulation.As outlined in Chapter 2.4.1,people can display blends of emotions because facial muscles are sufficiently complex and independent enough to create discrete emotional patterns observable in one face (Ekman et al., 1980).

Smug Expression

The participant (Figure4.12)below displays a blend of enjoyment and contempt creating a smug expression (Ekman,2007) because of repeating instructions in error(R9 inTable 4.7).

Figure 4.12:Smug Expression

The muscles responsible (Figure 3.33; Table 3.7) for smugness are as follows: AU4 working in conjunction with AU42 a separate strand of AU4 to lower and gather the eyebrows together. At another level, AU8brings the lips toward each other and compresses them while AU12 in conjunction with AU6B tighten the lip corners and raisethe cheeks to form a slight smile.

Angry-Disgust

This is a blend of two emotions triggeredby irritation, experience of a hated phenomenon or frustration during interaction.

Figure 4.13:Angry-Disgust

The participant (Figure 4.13) above experiences a combination of anger and disgust. Although, observers may sometimes confuse one for the other because of the intensity (Ekman &Friesen, 2003), the emotions are basically characterised by lowered brow, raised upper lip, flaring nostrils, and the open mouth curved downwards. The facial muscles responsible (Table 3.7)are as follows: AU4 brow lowerer, AU7 eye lid lowerer, AU9 nose wrinkler, AU15 lip corner depressor and AU16 lower lip depressor.

4.4.3.4 Eye Action

Static Searching face

Static searching facedescribes saccadiceye or oculomotor movementdisplayed during interaction.The saccade eye movement is used to track objects without head movement (Dragoi, 2015; Liversedge & Findlay, 2000; Dell'Osso, 1997). The participant (P4) leans forward and listens to the repeated instruction while searching for the piece described with her eyes sweeping across the Table in the three continuous saccadic movements (left-right-left). The stages of saccadic movement are illustrated in the vignette (Figure 4.14) overleaf.

Initiation: Saccade eye movement from right to left.

Return: Saccade eye movement back from left to right

Final: the locating movement from right to left again.

Figure 4.14:Static Searching Face- Saccade Eye Movements

Ekman (1977) suggests that saccadic movements are produced by AU61 for turning left and AU62 for turning right;AU54 enables P4's static head to bend down when she leans forward and AU12compresses her lips (Ekman, 1977). However, Ekman(1997, 1977) does not specify the muscles responsible so, I had to rely on research around neuroscience and human physiology indicating thatlateral rectus and medial rectus muscles[15]are responsible for horizontal saccadic movements (Dragoi, 2015; Dell'Osso, 1994; Robinson, 1964). The saccadic eye movements suggestthat P4 is searching for pieces and only stops when she locates the right one.

4.4.3.5 Moods

Aha! Moment

Aha! Moment is "a moment of sudden realisation, inspiration, insight, recognition, or comprehension" (Merriam-Webster, 2012)that leads to discovery. The form and function of the Aha! Moment is illustrated in the vignette (Figure 4.24) below.

---

[15] For details see http://droualb.faculty.mjc.edu/Lecture%20Notes/Unit%203/muscles%20with%20Figures.htm

Phase 1-Impasse: P19 is in a period of mental fixation (Mai et al., 2004) indicated by an expressionless face thinking outside the box Time: 00:06:16:327

Phase 2: Break in the mental fixation and the insight or solution appears suddenly and elicits positive effect (Shen et al. 2016; Topolinski & Reber; 2010) -the full smile occuring at 00:06:17:862

Figure 4.15:Aha! Moment

The Aha! Moment is realised through the experience of processing fluency (Shen et al., 2016; Topolinsky & Reber, 2010) in two phases. In the first phase the participant reaches an impasse, which is a period of mental fixation (Mai et al., 2004). This is when P19 tries to execute instruction 26(Appendix VII), she comes to an impasse and becomes stuck and unable to execute the expected task (Figure 4.15: timer: 00:06:16:327). The second phase occurs suddenly when a break in the mental fixation occurs as P19 experiences a sudden insight the instant at which the solution to the problem becomes clear, eliciting a positive effect - P19's full smile(Figure 4.15: Timer: 00:06:17:862).

Workman's Face

Workman's face is a blend of disgust showingall over the face andcontempt with a trace of anger.Ekman et al, (1997) report that people often press their lips together when doing anything that requires physical exertion such as pulling things apart or trying to lift heavy things.

Figure 4.16:Workman's Face

This suggests that participant (P10) is not disgusted, angry or contemptuous rather the expression is showing due to force exerted during task. Workman's face (Figure 4.16) is realised by AU9 the nose wrinkler, AU15 lip corner depressor used to raise the upper lip, AU16 lower lip depressor used to raise the lower lip and make it protrude while lid tightener slightly pulls lip corners.

Tense Mouth

The tense-mouth is an aggressive sign in our nearest primate relative, the pygmy chimpanzee or bonobo (de Waal, 1997) where the face may show obvious muscular tension (i.e., with the lips held tightly together) such as compressed lips, lip biting and pouty mouth.

Compressed or Swallowed Lips

The compressed or swallowed lips action (Figure 4.17)belowis produced by compressing, in-rolling, and narrowing the lips to a thin lineposition in which the lips are visibly tightened and pressed together. This is done using AU 8 which pullsor tightensthe lips towards each other (Ekman, 1988).

Figure 4.17:Compressed Lips

In assembly tasks, lip compression may indicate cognitive processing such as pondering, thinking, or feeling uncertain about the instruction. The action is more pronounced in the top-right and lower participants while the one top-left of the vignette displays only a slight compression without swallowinglips enabled by AU8B(Table 3.7).

Another form of the tense mouth occurs as lip biting (Appendix VIV.13) with the lower lip over the teeth.

Pouty Face

Givens (2010) observes that children pout in sadness, frustration, and uncertainty while adults spontaneously pout or display shades of the pouting cue such as contractions of the chin muscle when disagreeing with comments during interaction.As outlined in Chapter 3.3.3.2, pouting is part of lip action which Morris (2015) also calls pointing lips used by people in certain cultures to indicate direction when their hands are busy.

Figure 4.18:Pouty Face

When people pout AU17 and AU25 contract the chin at the same time, AU21 stretches the neck, while AU20 stretches muscles of the lower lip(Givens, 2016). AU22 funnels the lips making the pout more prominent and shaped to that used in whistling. Pouting suggests that the participant (Figure 4.18) is either concentrating on the task, cognitively processing instructions or self-comforting herself during the interaction. These inferences differentiate pouting face from the lip-pout that often accompanies the shoulder-shrug display, which is a sign of resignation.

4.4.3.6 Microexpressions

As outlined in chapter 2.2, microexpressions occur when people deliberately give off emotions that do not reflect their true feelings such as the samples analysed (Figures 4.19; 4.20; 4.20 and Appendix VIV. 15-17 below.

Micro-anger

Ekman (2007) suggests that anger and fear often occur together with fear preceding anger while anger galvanises people to action, thus they have similar features but anger could also be confused with perplexity.

Figure 4.19:Micro-Anger

Although, the participants are trying to remain calm, anger leaks out probably due to perplexity (Ekman& Friesen, 2003) during the task. The participants (Figure 4.19) above display different forms of anger realised by AU4B the brow lowerer, AU5 the upper lid raiser, AU7B the orbicularis oculi muscle or lid tightener, and AU20B the lip stretcher making lips thinner as a clue to anger, while AU25B opens the mouth and keeping the lips wide apart as if in speech. The combined action of the AUs narrows the lips forming a square mouth which suggests anger.

Micro-disgust

Some microexpressions observed in this study are prominently displayed on one side of the face such as the one-sided disgust(Figure 4.20)with the mouth bending to the right. It is realised by AU9B the nose wrinkler, AU15B lip corner depressor raising the upper lip, while AU16B lower lip depressor raises the lower lip making it protrude while the eyelids are relaxed.

Figure 4.20:One-sided Disgust

The participant (Figure 4.20) above displays a disgust that Ekman and Friesen (1978a) said is not like the disgust a wife directs at a husband when in conflict or when an adult perceives a bad smell; rather, P10 directs this emotion involuntarily towards the effort she is putting into the task.

Micro-sadness

Participants' emotional leakage of sadness (Figure 4.21 below) is signalled by lips turned down very slightly. The muscles (Figure 3.33) responsible include AU1Braising the inner brow, AU15B pullingthe lip corner, sometimes AU25B slightly parts lips, AU20B stretches the mouth and AU6 raises the cheeks(Table 3.7).

Figure 4.21:Micro-Sadness

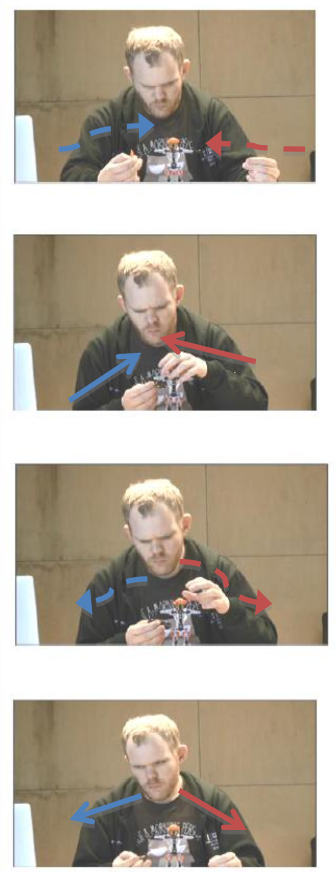
The participant on the right displays slightly oblique eyebrows, very slightly lowered lip corners, stretched mouth and slightly raised cheeks (Ekman, 2007). Both participants' eyes are focused but oblique eyebrows suggest being intrigued while drooping lips may suggest dejection. Sadness may be triggered by failure to achieve a goal (Ekman, 2007) such as successfully executing the agent's instruction. However, the participant on the left is not showing sadness in the brows while the emerging pout suggests determination.

4.4.4Assembly Gestures

The results of the study indicate that participants displayed two hundred and forty-seven (Table 6.3) assembly gestures classified as familiessuch as aligning, picking, presentation, joining hands, communicative and a mixed collective or sacrum gestures. The process of determining how an action transforms into a gesture with meaning and functions is outlined in chapter 3.3.1.3. The details are outlined as follows.

4.4.4.1 Aligning Hands

Fussell et al (2004) demonstrated that gestures that go beyond mere deictic reference are the most crucial in terms of task performance. The aligning hands is similar to Kirk et al's(2005, p.11) 'mimicking hands' in an assisted assembly task because they enable listenersto order and discover the fit of assembly pieces. The role of aligning hands is illustrated in vignettes (Figures4.22) below.



Preparation P1n initiates the gesture by picking up the assembly piece, comparing them and trying to mentally and visually place them on the Lego model.

Stroke P1n performs the phase by bringing the hands together thus physically taking the piece close to the part of the Lego kit and sizing them up

Retraction begins as P1n's hands start to move apart still holding the piece without fixing it on the Lego kit.

Retraction/Recoveryis completed when P1n's hands are finally at rest on the Table.

Figure 4.22:Aligning Hands

Aligning hands enable listeners to discover compatible assembly bits when theyvisually and mentally assess their fit in a four-stage gesture phrase. In Figure4.22, P1n mentally compares oneassembly piece with the Lego kit it will fit into then brings them together to size them up physically. The gesture is completed in a two-staged process that culminates with both hands retracting towards the sagittal axis to rest on the table palm up.This gesture may occur with extra strokes (appendices VIV.18a-c), with twists, used for repositioning (Multiple aligning hands), and may last only one microsecond like microexpressions (Micro-aligning hands) (Appendices VIV.19a-c).

4.4.4.2 Presentation

When participants want to examine selected pieces, assembly actions or finished work they use presentation gesture (Kendon, 2004) to display these in a manner similar to one making presentation to co-interlocutors.The form and function of the gesture are shown(Figure4.23)below.

Preparation: P42 initiates the gesture with both hands holding the assembled kit's feet down on the Table.

Stroke: P42's hands lift the Lego kit up then turn it upside down about 180^0 inwards with head down and the feet up. The RH is holding the head while the LH holds the legs

Retraction 1: This takes place as a reversal of the stroke. The kit is turned about 180^0 outwards or away from the sagittal axis to make a presentation.

Retraction 2: P42's hands now go downwards and this puts the kit back on its feet.

Figure 4.23:Basic Presentation

Another variation of presentation is the checking hand that participants use to search for assembly bits before selecting the desired one. The checking hand gesture is enacted by picking up a piece in a seemingly random manner then ending with a horizontal presentation. The functions and role of the checking hand areoutlined in the vignette (Figure4.24)below.

Preparation: The **LH** rises and goes towards the workspace palm down vertical, with digits curved

Stroke1. P3’s **LH** moves towards kit layout palm down to pull up one piece.

Stroke 2. P3’s **LH** performs the key stroke by horizontally holding up the picked piece for examination

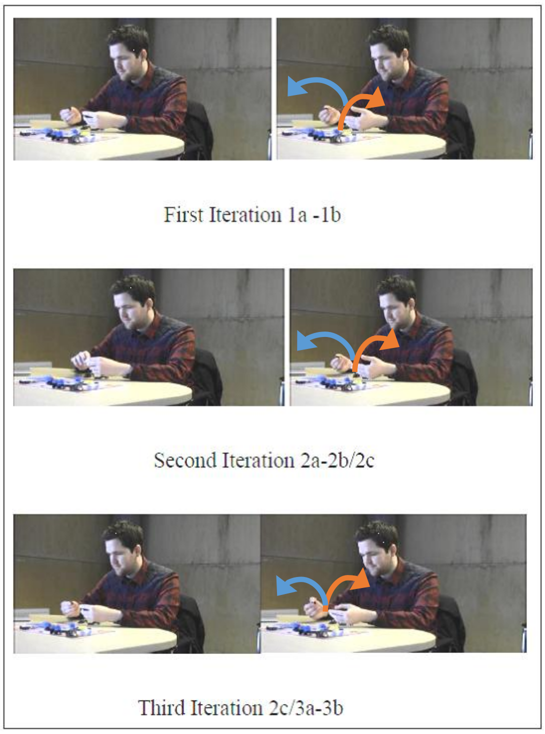
Figure 4.24:Checking Hand

The checking hand gesture co-occurs with the G.W. Bush grinhowever, the context within which the gesture occurs suggests that presentation enables users to assess their comprehension of the instruction or confirm the suitability of the action taken.

Presentation may be done ambidextrously for comparison (appendix VIV 20a), right hand (appendix VIV 20b), left hand (appendix VIV 20c), in a full circle (appendix VIV 20d), semi-circle (appendix VIV 20e), with multiple retractions (appendix VIV 20f) in combination with other gestures (appendix VIV 20g) and with the fingertips (appendix VIV.20h).

Repetitive Opening Hands

In assembly tasks, opening hands gestures are like beat gestures (McNeil, 1999) or rhythmics (Ekman& Friesen, 1978a-b) because they occur with two movements of either up and down; inside-outside or outside-inside from the sagittal axis. However, to avoid confusion with beat gestures, I have classified the emerging gesture as repetitive opening hands. The form and function of the repetitive opening hands are shown in the vignette (Figure4.25) below.



P2 **prepares** in 1a by picking up the pieces in both hands with fingers curled in a fist. The **stroke(S1)** occurs in 1b. The hands open sideways/inside out, palm up with digits stretched out presenting the pieces picked before P2's eyes for evaluation.

P2 repeats the gesture again while listening to the instruction. 2a is **preparation** with hands held as fists/ fingers curled while 2b is the **stroke (S2)** open sideways/inside out, palm up with digits stretched and two-handed post-stroke hold (2c).

This prepares for the **stroke (S3)** open sideways/inside out, palm up with digits stretched out placing the pieces in 3b while listening to the repeated instruction.

Figure 4.25:Repetitive Opening Hands

Repetitive opening hands gesture is said to be repetitive because the gestures occur in continuous successions called iterations with one leading seamlessly into the next (Chapter3.3.2.5). The gesture is enacted in three iterations with two beat phaseseach. The first phase (a) is preparation while the second (b) is the stroke. The gesture has three preparations and strokes outlined in the text box.Repetitive opening hands enable listeners to perform task-aiding functions like self-evaluation of comprehension levels and task progress that may lead to self-correction.

4.4.4.3 Picking Hands

When participants want to select assembly pieces they use picking gestures. The samples presented here include the knowing hand, searching hands and two-hand setting aside.The other variations of these three are in the appendix VIV 21 a-n.

The Knowing Hand Gesture

This is a beat gesture that is the opposite of the wavering hands gestures (Kirket al., 2005) and searching hands (Figure 4.26) because the listener correctly decodes instructionsand goes straight ahead to pick up the required piece without hesitation. The form and function of the knowing hand gesture areoutlinedin the vignette (Figure4.26) below

Preparation: P2 initiates the gesture with the **LH** in a hold, palm down with cupped digits. The **RH** is palm lateral and in a hold too.

Stroke: Here P2's **LH** zooms downward, palm down then grasping the piece with coupler-shaped digits and pulling it up.

Retraction: The **LH** goes backwards and up then turns palm up to do a presentation.

Figure 4.26:The knowing Hand

This is an assistive assembly gesture that aids the execution of operative gestures. It can be done with a hold, a supine post stroke hold, multiple gesture phrases, an aligning hand, both hands at once, both hand simultaneously or with radial movements, a concluding hold, the wrist doing all the work, two fingers, and one hand selecting pieces one after the other (See appendix VIV.21b-n).

Searching Hands

This is similar toKirk, et al's (2005) wavering hand but it is done with both hands alternativelyand may indicate listener's incomprehension of instructions. The form and function of the gesture are shown in the vignette (Figure4.27) overleaf.

Preparation: P34 initiates the gesture with both hands in a hold.

Stroke 1: The **LH** moves forward, palm down with digits held coupler-shaped. The **RH** comes down lower without touching the Table.

Retraction 1: The **LH** withdraws halfway towards the sagittal axis

Stroke 2: The **LH** still palm down moves forward then uses a combination of the thumb, forefinger and middle finger to pull up an assembly piece.

Retraction 2: P34's **LH** withdraws again towards the sagittal axis to drop the piece picked up in S2.

Initiation/Retraction 3: The RH moves away from the sagittal axis towards the kit layout palm down. The **LH** completes the retraction 2 with a hold

Stroke 3: Here P34 executes two strokes at once. The **RH** palm lateral grasps the upright body then moves leftwards. The **LH** palm down pulls up another piece and moves leftwards too.

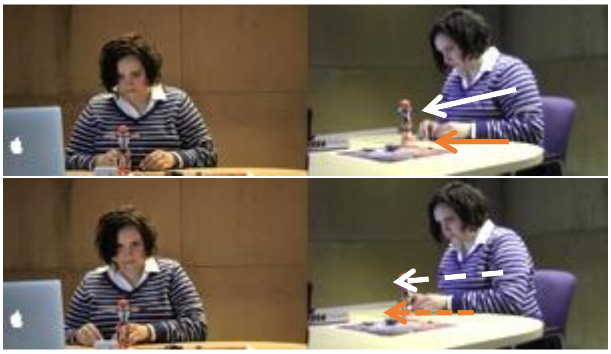
Stroke 4: P34's **RH** now lifts the upright body up then moves to the right. At the same time, the **LH**, picks up another piece, moves rightwards and comes to a post-stroke hold too.

Figure 4.27:Searching Hands

Another form of the searching engages more gesture phrases as illustrated in appendix VIV.21a

Two Hand Setting Aside Gesture

This is also like beat gestures but instead of the back and forth movement, it is a one directional movement that takes place in two steps. The form and function are shown in the vignette (Figure4.28) below.



Preparation: P4 prepares for by using two hands to pick up the piece using a joining gesture.

The **stroke** is executed with two hands holding and pushing the piece forward indicated by the broken arrows until the desired position is achieved

Figure 4.28:Two-Hands Setting Aside

The first part of this gesture grasps the assembly bits without lifting them off the Table and in a continuous motion, the hands push them on thetable towards the desired location. This movement is similar to the one used by people grinding seeds on a grinding stone with a top stone, or muller that is indigenous to many parts of Africa, Australia and Asia, where they crush, grind and shift seeds forward in one move. Another version is segmented which contains alternate selection and ends with re-positioning of the assembly piece (Appendix VIV.22)

4.4.4.4 Joining Hands

After picking up assembly pieces then may be aligning them for fit, the next stage is attaching them to the appropriate location with the correct orientation using the operative gesture aka joining hands gestures. They are iconic representations of parts of speechwithin instructions such asactive verbs indicating what to do;prepositions indicating where to do it, and adverbs indicating how to do it, for example, join, attach, insert, or place. One variation of the joining hand is used to attach fragile pieces (Appendix VIV.23a) and may also be enacted to show adjustments made during the task as multiple joining hands (appendix VIV.23b-c). The form and function of the basic joining hand are shown in the following vignettes (Figure4.29) below

Preparation: P20 initiates the gesture from the **RH** picking up a piece and the **LH** resting on the Table. P20's **LH** goes up, palm down, with digits held like claws and holds the Lego arm from the top. The **RH** moves up, palm lateral with sideways grappolo-shaped digits and brings the piece to join the Lego arm.

Stroke: The **LH** spins the body around in 180^{0}to position the arm(indicated by the curved broken green arrow). The **RH** attaches the piece to the arm with a bottom-up stroke.

Retraction: The **RH** retracts towards the agent while the **LH** retracts to left backwards.

Figure 4.29:Basic Joining Hand

The structure of the gesture phrase is as follows

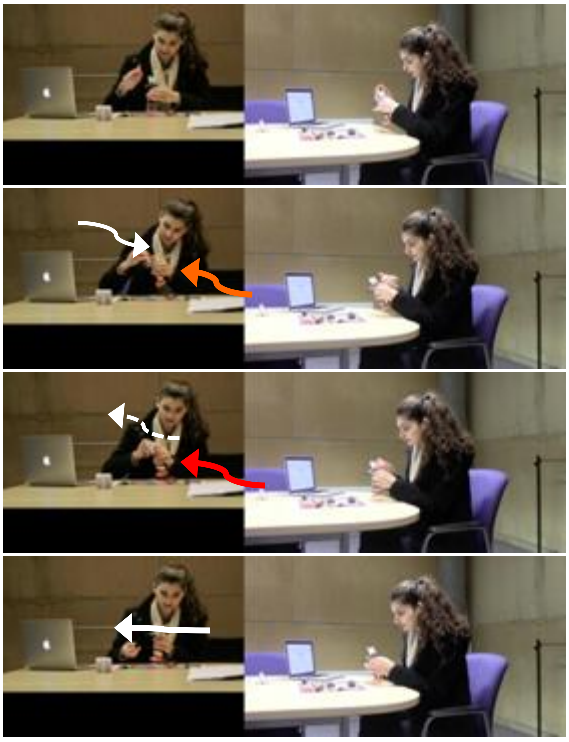
<~~~/ ***/-.-.-.->

P S Post stroke hold

… attached onto the top two black joints of the body

FailedJoining Hands

The failed joining hand gesture is said to have occurred when participants cannot successfully attach,insert, turn, or place an assembly bit in the required position and orientation within a3-D space. It may be enacted with the repositioning of assembly bits (Appendix VIV 23d). The form and function of a failed joining hands gesture are shown in the vignette (Figure4.30) below.



Preparation: P20 initiates the gesture by picking the piece to be joined with the **RH** and the body with the **LH**

Stroke 1: P20's **RH** brings the piece picked from the top as the **LH** brings the body from beneath to attach the socket inside the green piece. **Failure:**The socket slides down to the **LH** instead (point indicated by red arrow)

Stroke2: P20 uses the second stroke to recover the dropped piece with the **RH** palm down and two fingers sliding upwards (indicated by the blue broken arrow)

Retraction:the **RH** retracts rightwards to a hold with the **LH** holding the body.

Figure 4.30:Failed Joining Hands

The gesture phrase structure is shown below.

<~~~~~~~~~~~//***???????????/ /**********/-.-.-.-.-#####>

P S1 (Failed) S2 Post stroke hold and R

Now just attach the socket inside the green piece to the top black joint of the body.

The eyes again should more or less… …be facing…forwards…

4.4.4.5 Others/Sacrum

These are gestures that perform functions but do not belong to any specific class because of their unique features; these include crooked aim, top-down keeping hand, twisting hands, vibrato dropping hands, and concatenated mixed hand.

The crooked aim is executed with a crooked hand (Bressem, 2014; Drillis, 1963) as a preparatory move akin to one taking aim before shooting or hitting an object used to aid picking and presentation gestures in assembly tasks. The form and function of the gesture are shown in the vignette (Figure4.31) below.

Preparation: P25 initiates this gesture in two moves. The first is the rest position where the **LH** is in a hold while the **RH** is on the workspace. The second phase of preparation occurs when P25 uses the forefinger and thumbin a semi-circle to form a crooked **LH**that moves back and forth taking aim

Crooked Aim

Stroke 1: P25 executes the stroke from the aiming position with the **LH** coming down, palm down using the crooked hand to graspand pull the piece out from the kit layout.

Stroke 2: The **LH** retracts to a hold palm up to display the piece before P25's eyes for examination. The **RH** retracts to the initial rest position.

Figure 4.31: Crooked Aim

The structure of the gesture phrase is as shown below

<~~~ ~~~~~/**************/********/#####>

P S1 S2 R

Small orange armour piece

The crooked aimappears to be a more cautiousgesturebecause of the repetitive phase (Kendon, 2004; Kita et al., 1997; McNeil, 1992) of back and forth movement of the picking handindicated by two blue arrows. The crooked aim should not be confused with Kendon's (2004) ring hand recognised today as a classic President Trump co-speech gesture where the thumb andforefinger are joined at the tips to form a circle. In the crooked aim, thefingers do not meet, rather they form a semi-circle.

The Top-Down Keeping Hand

Participants use the keeping hand gesture to execute vague and incomplete instructions. The form and function of this gesture are shown in the vignettes (Figure4.32) below.

Preparation: P13 begins by holding his hands out in the readying position to go in any direction.

Stroke 1: P13 executes the first stroke by moving the **RH** forwards, palm down with coupler-shaped digits to pull up the specified piece.

Retraction 1: P13 moves the **RH** back (indicated by the green broken arrow) then hesitates in preparation for another move

Stroke 3: P13 then moves the **RH** rightwards (indicated by the green broken arrow) and downwards, palm down with coupler-shaped digits to pull up another piece while still holding the first piece.

Stroke 4: P13 retracts the **RH** then moves leftwards palm down horizontal and drops the pieces into the **LH** from the top.

Retraction 2: The **RH** retracts to the right (indicated by the broken blue arrow) to initiate the instruction giving process.

Figure 4.32:Top-Down Keeping Hand

As outlined above, some assembly instructions are partly vague while others are specific, as illustrated with these instructions below:

1. *To start with locate the two black feet and place them claws down on the desk.*
2. *Now take two of the small black pieces that are sort of a Y shape and have a crossed hole at the bottom.*

The vagueness in instruction 2 enables researchers to identify self-propelled behaviours (Chapter 4.3) that listenersuse to deal with ambiguity during interaction, for example, (Figure 4.32), P13 picked the pieces up with the right hand as instructed then dropped them into the left hand from the top a self-propelled technique not stated in instruction 2.

Twisting Hands

During the assembly task, sometimes pieces get stuck together and listeners pull them apart. The form and function of the twisting hand gesture are explained in the vignette (Figure4.33) below.

Preparation: P10 initiates the gesture by performing a lateral joining hands gesture with the **RH** joining the immobile **LH.**

Stroke 1: P10 performs the stroke by first turning the **LH** clockwise and the **RH** anticlockwise. These movements loosen the bits ready for the next stroke

Stroke 2: This occurs when hands are pulled apart in opposite directions and the joined bits are separated with a force.

Recoil: The force of the pull brings the hands together again in a recoil indicated by the broken arrows.

Figure 4.33:Twisting Hands

Participants use twisting hands gesture to forcibly pull apart two pieces that were hitherto joined together by twisting and pulling their hands simultaneously in opposite directions.

Vibrato dropping hands

This is an unconscious gesture that occurs when people carry out tasks in much the same way as people miss steps when walking or have mishits when hammering a nail. Vibrato is borrowed from the description of asaxophonist's wriggling fingers on a woodwind instrument to producestronger or richer tonesso that it seems to shake slightly. The form and function of the vibrato dropping hands gesture are shown in the vignette (Figure4.34) overleaf.

Preparation: P22 initiates this gesture after picking up a piece and repeating the instruction. P22 wriggles her fingers in a gestural hedge ‘the vibrato’ with the **LH**

Stroke: The drop occurs while P22 is still wriggling the fingers of her**LH.** The piece held drops down sharply on the workspace, indicated by the broken green arrow.

Retraction: This occurs in two stages. The first is a reflex action used by P22 to retrieve the piece with the **LH.** The second stage completes the retraction when the **LH** closes around the piece in a post-stroke hold while the **RH** goes towards

Figure 4.34:Vibrato-Dropping Hand

Vibrato-dropping hand is a compensatory gesture in the sub-class of hedge (Gullberg, 1998; Fraser, 2010) that listeners use to probably mitigate the effect of mistakenly dropping assembly pieces or information overload.

Concatenated Mixed Hand

During the assembly task, participants sometimes must employ a combination of gestures to execute an instruction. This eclectic approach produces concatenated gestures comprising joining, dropping, picking and displaying hands that flow seamlessly into one another as participants repeat instructions. It is used to break instructions or task into manageable chunks, confirm if the instruction was correctly executed and aid the next operative gesture. The form and structure of the gesture are outlined in the vignette (Figure4.35) overleaf.

Preparation: P24 initiates the gesture from an aligning hand position. The hands are in front examining the fit of the pieces

Stroke1: P24 enacts a dropping hand with the **LH** palm down and coupler-shaped digits in the front-left location. This may indicate a failed joining hand gesture in the preparation

Stroke2: P24 picks up another piece from the dropping position by just stretching the digits forward and using the thumb and forefinger to pick up another.

Stroke 3: P24 performs another joining hands gesture but this fails and it is more of an aligning hand.

Stroke 4: P24 continues by performing a two-handed displaying hand gesture by turning the **RH**clockwise and the **LH** anticlockwise

Retraction: P24 retracts by moving the **RH**diagonally to the right towards the agent to ask for a repeat. The **LH** is placed in front on the Table palm up with the thumb and forefinger holding the unassembled piece.

Figure 4.35:Concatenated Mixed Hand

The structure of the gesture phrase is as shown below:

…the crossed connector of the orange piece/ The other half should stick out at the other end
<~~~~~~~~~~~~~~~/************************/************/##############>
P /S1_S2_ S3 picking-drop-Joining hands/ S3 leading to S4 JH to-Display/ **R**etraction.

The concatenated mixed hand gesture occurs over a long stretch but the most expressive points were picked for illustrating the vignette and these include phases of joining, dropping, picking and displaying hands that flow into one another seamlessly as the instruction is repeated. However, the preparation stage co-occurs with part of the instruction from the second word while S1-S3 co-occurs with the last part of the agent's instruction. S4 and retraction occur during the pause after the whole instruction has been given.

4.4.4.6 Communicative Gestures

These are discourse-oriented gestures that offer a channel for observing the psychological activities that take place during interaction (McNeil, 1985) and interactional practices between interlocutors. Examplesincludeself–repairs, turn-takingand deictic gestures.

Turn-Taking and Repairs

Turn-taking is a process by which interlocutors allocate the right or obligation to participate in interactional activities such as conversations, classroom interaction and HCI (Carroll, 2002; Sacks et al., 1974). Repair organisation is a self-righting mechanism in social interaction that describes how participants deal with issues arising from speaking, listening, hearing or understanding in every conversation (Clark 2012; Schegloff, 1997; Schegloff et al., 1977).

The organisation of turn-taking is essential to human communication like conversation (Sacks et al., 1974), Computer Mediated Communication (Garcia & Jacobs,1999) and has been found to give insight into language processing and explain the character of language (Levinson, 2016). Furthermore, Garcia and Jacobs, (1999) report that participants in normal conversation do not concurrently inhabit speaker and hearer roles but during arguments, they do so. Participants may interrupt each other frequently and speak simultaneously yet display an orientation to what the other said while they were talking. This may be accounted for by the listener's ability to process speaker utterances in real time such that it looks automatic (Field, 2009).

This study presents turn-taking between the agent and the participant as kit constructional units. Kit constructional units comprise a unit of the agent's instruction as a speaking turn where the instructor holds the floor, while the participant's compliance with the instruction is the listener's speaking or interaction turn where the instructee holds the floor. An example of a kit construction unit transition is shown below:

*AGENT INSTRUCTION: To start with locate the two orange feet and place them on the desk

*LISTENER COMPLIANCE: Listener is expected to pick the orange feet and place them down on the desk

A kit construction unit transition could be long or short depending on the length of instruction and pace of listener compliance. The default compliance position is for the listener to execute instructions; however, the results indicate that listener-action may include asking for repeats, executing instruction, carrying out repairs or even doing nothing.

Some participants altered thekit constructional unit as a turn-taking violation. For example, when P35 received instruction 36(Aquagon) but did another thing outlined below

*AGENT INSTRUCTION: 36. Now take the tail end and place it inside the back hole of the grey fin at the back of the body

> *LISTENER COMPLIANCE P35: Listener repeats the instruction thenpicks the yellow face piece instead of inserting the tail inside the fin so, does not execute the instruction.

Another example of turn-taking violation occurred when P15 repeated one instruction but executed another as illustrated below

> *LISTENER COMPLIANCE P15: Repeated Instruction 44 twice then executed 43 and 44 while listening.

P15 listened first to instruction 43 but did not execute the instruction thus, delaying the transition. P15 then listened to and repeated instruction 44 but executes instructions 43 and 44 at the same time.According to Sacks et al. (1974),these actionsviolate the expected transition sequence specifying that when the agent gives instruction, the participant executes it; rather, P15 has given the agent two turns in a row while altering her own turns. This suggests that it might be a strategy transferred from human interactionand a comprehension strategy.

Positive Turn Violating Hand

As outlined earlier, a person taking assembly instructions like P14 may positively violate turn-taking protocol by pre-empting instructionsbycorrectly executing them before they are issued. This action may suggest that participants have the correct orientation towards instructions given. The form and function of the gesture are shown in the vignette (Figure 4.36) overleaf.

Preparation: P14 has already picked up pieces then initiates the gesture by striking a repeat point to get the agent to repeat the instruction

Stroke 1: P14 performs a joining hands gesture with the **RH** while the **LH** is in a hold. This pre-empts the agent's next instruction {*8. these should attach to the front joint of each black piece. The ends with the holes should be closest to the grey piece.*} And violates the turn-taking arrangement in place

Stroke 2: P14 picks up another piece with the **RH**

Stroke 3: P14 keeps it in the **LH**

Retraction: P14 retracts the **RH** towards the agent (indicated by the green broken arrow) to initiate instruction no 8.

Figure 4.36:Positive Turn Violation

Repairs in HACs

Repair is an effective interactional tool used to put interaction back on the right track during conversation (Frenečik, 2005) and have been found to indicate that participantsare learning in HCI (Jordan et al., 1995).

Furthermore, Schegloff et al., (1977, p.365) suggest that an organisational amplification which focuses on distinguishing between approaches involving (1) initiatingrepair, described as "methods for indicating trouble and making its management the focal activity within the interaction until either the trouble is resolved or efforts to do so are abandoned" and (2) actual repair, which are "practices for resolving whatever trouble of speaking, hearing or understanding has arisen or been indicated" (Schegloff et al., 1977, p.361). In summary, repair organisation comprisesTrouble Source (TS) identification and location, repair initiation and actual repair (repair-execution) (Frenečik, 2005).

Although, this organisational amplification and TS location provides fourrepair trajectories to participants(Hayashi et al., 2013; Frenečik, 2005), the simulated agent giving instructions creates a unidirectional interaction pattern and thus affords the listeners only the Self-Initiated Self-Repair (SISR) trajectory. SISR refers to the first opportunity within the first turn or immediately after turn construction to make repairs (Frenečik, 2005). Self-initiated joining repairs occur during assembly taskswhen participants attach assembly bits wrongly they then proceed to effect corrections. SISR may occur in two stages (Appendix VIV 24). The basic repair process is outlinedin the vignette (Figure 4.37) below.

Trouble Source (TS)

Preparation1: P23 initiates the gesture by holding oneassembly piece in the **RH** palm down

Stroke 1: P23's **RH** moves to the right and turns clockwise 180^{0} to examine it while listening to the repeated instruction. P23 identifies the TS as wrong choice of assembly piece

Repair initiation

Preparation 2: P23 begins repair from stroke 1 with the **RH** in a post-stroke hold.

Stroke 2: P23's **LH** moves rightwards, palm down with coupler-shaped digitsto pull out the piece from the **RH**

Repair refinement

This move continues from stroke 2 in the next frame where P23's **LH** goes to the right palm down to drop the piece in her front then moves towards the workspace. This is another self-repair within a single repair process and may indicate hesitation

Actual repair execution

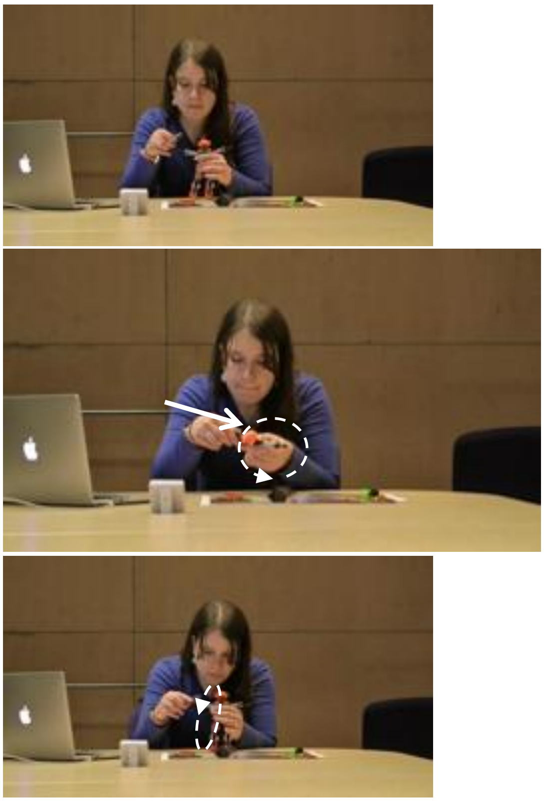
Preparation 3: P23 continues the process after refinement with the **LH** moving diagonally towards the kit layout.

Stroke3: The repair action of returning the wrong piece is executed when P23's **LH** moves diagonally palm down with digits clutching the wrong piece to drop it in the layout. Broken green arrows indicate intended paths.

Figure 4.37:Self-Initiated Joining Repairs

Self-Restraint

Self-restraint is an aspect of one's ability to deal with one's environment. Self-restraint involves cognitive, social, behavioural and continuous creative operative skills that will enable one execute courses of actions (Bandura, 2006, 1986)such as assembly tasks.Operative skills are related to the concept of physical guidance or restraint- described as the application of physical contact that induces individuals to go through the motions of a desired physical activity(Martin & Pear, 2016)such as clapping one's hand over one's mouth or using a bridge hand to steady one's aim when taking a pool shot. During the assembly task, itwas observed that P15 used a combination of facial actions and gestures as pragmatic skills to enact and externalise self-restraint as illustrated in the vignettes (Figure 4.38) below.



Preparation: The cluster is initiated at the beginning of the joining hands gesture. P15's **RH** is holding one piece while the **LH** is holding the body of the Lego kit upright.
Stroke: P15's **RH** attaches the piece to the body of the Lego kit now turned on its back with the **LH** (indicated by broken arrow). Her mouth is tensed with swallowed lips suggesting physical and mental self-restraint probably because of the need to balance the force required for assembling fragile assembly pieces with the fear of breaking them with too much force.
Retraction: P15's **RH** moves back to a post-stroke hold with the **LH** now holding the kit upright again (indicated by the broken arrow). Her face becomes neutral and relaxed with the lips closed.

Figure 4.38:Self-Restraint Gesture

Pease and Pease (200420-24) suggest that to get accurate reading, gestures should be analysed in clusters and to ascertain congruence within interaction contexts.Thus, in the initial picture(Figure 4.38) the participant's face is neutral with lips closed and relaxed in a position however, in the second picture when the participant executes a stroke, she now has swallowed lips. Considering the task context, swallowed lipsmay be attributed to the controlled force she is exerting when handling fragile bitsbecause when the participant stops what she is doing her face and lips become relaxed again.

False Starts

Maclay and Osgood, (1959) identified False Starts (FS) as a hesitation phenomenon in spontaneous English language use described as incomplete utterances or self-interrupted utterance. Furthermore, existing work in psychology suggests that hesitationmay be externalised as hesitant nonverbal behaviours- the jerky hand movement when two people reach for an object at the same time(Moon,et al., 2010) where the reflexive withdrawal makes it incomplete or a self-interrupted gesture.People hesitate when they perceive risk in their actions or considerhow these will deal with their feelings of self-esteem and self-conceptandhow useful these activities are during interaction(Wong & San Hu, 2011).Following these, this thesis identifies and classifieslistener nonverbal behaviour that demonstratessimilar hesitation as false start gesturesshown in the vignette (Figure.4.39)below.

Preparation: P36 initiates the gesture with the **RH** resting on the Table. The **LH** is palm down and pulling up an assembly piece

Stroke 1: P36's **RH** moves forward diagonally, palm down. It then hovers a bit in a brief hold. The **LH** palm lateral moves backwards towards the sagittal axis to a post-stroke hold

Stroke 2: P36's **RH** moves backwards quickly to another brief hold- the jerky-hands move

Stroke 3: P36 moves the **RH** forward and diagonally again, palms down with digits held coupler-shaped. Then using a combination of the thumb and forefinger, to pull up the assembly piece.

Figure 4.39: False Start Gesture

False start gesture is enacted with several movements given as follows. At preparation P36 rotates the left wrist in ulnar pronation-movement of the wrist towards the pinky side (Henley, 2010) to pick up one assembly piece but she drops it (FS1) in the second stage, the right-hand moves forwards seemingly to pick up, only to stop midway(FS2). In the third picture, the right hand moves back to a hold and, in the fourth, the right hand moves forward to now execute the picking gesture. FS1 and FS2suggest hesitation on the part of P36 probably due to perceived functional and psychological risks (Wong & Hu, 2011) such as picking up ofthe wrong piece, which may lead to a wrong assembly move or failure and probable loss of face.

Deictic Hands

Pointing gestures or gestural-deictic reference indicate an object, location or direction (Foster, et al., 2008; Kendon, 2004), for example, the helper's decisive pointing movement (Kirk et al., 2005) used to direct the performer's wavering hand to a specific piece and location in assisted assembly tasks. The deictic pointing hand is a steering gesture used by participants to locate 3-D points and positions mentioned in instructions during assembly tasks. The form and function of the gesture are shown in the vignette (Figure 4.40) below

Preparation: P20 initiates this gesture from a repeat point. The active hand in the gesture is the **RH.** The broken arrow indicates the projected path

Stroke: P20's **RH** move towards the upright body palm lateral with the forefinger pointing and touching the assembly point as the agent repeats the instruction.

Retraction: P20's **RH** retracts to a hold while still listening to the instruction.

Figure 4.40:Deictic Hand

<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~/*******************************/-.-.-.-.-.-#####>
P S Post stroke hold/R
The grey socket of each arm should be attached onto the.../ top two black joints of the body

The structure of the gesture phrase shown above indicates that the gestural stroke occurs when the agent mentions the assembly point *'...top two black joints of the body'.* Foster et al. (2008) argue that pointing to support a mutual task makes the entity referred to more accessible and, in this case, pointing enables the participant to assess the comprehensibility of assembly instruction and appropriateness oftask execution as the agent repeats the instruction.

When deictic gestures measure and give iconic interpretation to lexical items it is said to be co-interpretational (Appendix VIV.25a). when the gestures are used to portray measurement with fingers, they are classified as deictic representation (Appendix VIV 25b). In addition, when fingers are used to enumerate things as done in a tally sheet, deictic enumerator (Appendix VIV.25c) occur.

Deictic Hedge

The deictic hedge gesture is produced by vibrato and pointing actions occurring together indicated by the hairpin bend arrows (Figure 4.34) overleafand,just as outlined in Figure4.34,vibratodescribes the listener's wriggling forefingers (flexion and extension) opposite the extended thumbduring the task.Furthermore, research in psychology suggests that wriggling fingers may be used to control anxiety during instruction-taking activities (Williams et al., 2016) and in combination with pointing, they becomea compensatory gesture in the sub-class of hedge (Fraser, 2010; Gullberg, 1998).

The results suggest that listeners mitigate information overload using deictic hedge while processing instructions to locate the assembly piece during interaction. The form and function of the gesture are illustrated in the vignette (Figure 4.41) overleaf.

Preparation: P34 initiates the gesture with the **RH** requesting instruction while the **LH** is in a hold

The **RH** rises palm lateral, then moves rightwards as the **LH** rises and moves towards the kit layout.

Stroke 1a: This occurs in two stages as P34’s **RH** wriggles the forefinger, palm down and oscillates to the right. This is the first part of the vibrato.

Stroke 1b: P34’s **RH** comes closer as the finger oscillates back to the left. This is the second and final part of the vibrato where the stroke is complete.

Preparation 2: This is a brief hold continued from S1b. P34 now raises the**LH** to the level of the **RH**. Both hands are palm down with the forefingers and thumbs stretched out like clasps in readiness for the next move.

Stroke 2: P34’s both hands dive down towards the centreof the kit layout to pull up assembly bits at once.

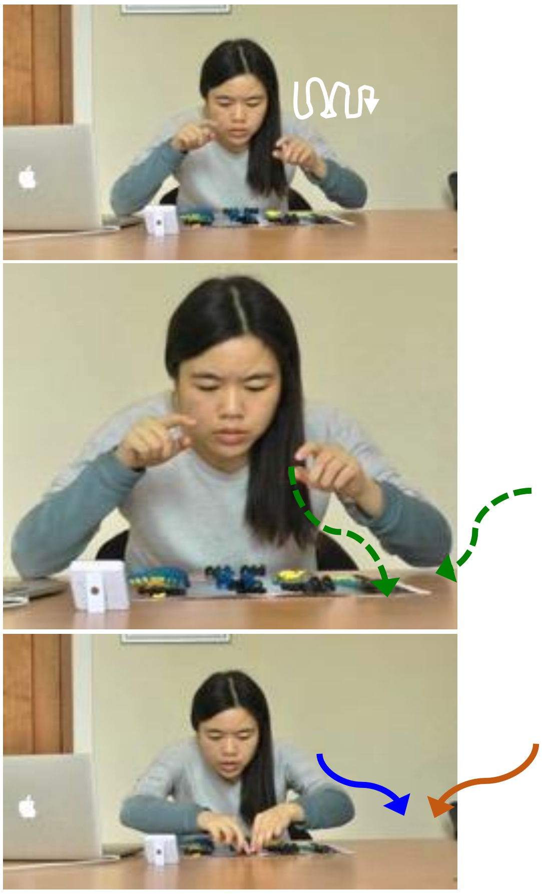


Figure 4.41:Deictic Hedge

4.5 Discussion

The results of this study indicate that listeners displayed a robust pattern of nonverbal listenership behaviour during interaction with simulated agents. The discussion will focus on the emerging interaction pattern (given by repetition process and typology), facial actions and assembly gestures.

4.5.1 Emerging Hybrid Interaction

In the joint task between agent and listeners, interaction is built around listener requesting agent instructors to repeat instructions. The interaction is described as hybrid because listeners make requests by pushing a button while agents respond by giving verbal instructions thus making the interaction both nonverbal and verbal.

Furthermore, the results suggest that interaction in HAI discernible and systematic which is partly attributed to agent design (4.3.1). In addition, hybrid interaction is consistent where agents always repeat instructions in the same exact form without adding, deleting or paraphrasing words or ideas. This is one area of agent competence and advantage over men that serves to create their individual and group identities (Chapter 2.4).

Hybrid interaction involves a chain of actions and events that are mainly reposed in the listener because the simulated agent has limited abilities. Listeners use a chain of events to manage the interaction thus, they initiate interaction, create repeat points and resolve every impasse reached (4.4.1). Secondly, listener facial actions and assembly gestures represent listener perception and interpretation of verbal instructions.

Listener perception and interpretation of instruction signal listener comprehension levels and communication needs during tasks. To this end a typology of repeats corresponding with listener interaction needs have emerged with their positions along the assembly process timeline (4.4.2.2).

4.5.2 Facial Actions

Facial actions analysed indicate that when people take instructions from agents, they project basic emotions, non-basic emotions, attitudes, moods, and context-based actions (Table 4.5)below. They are also able to manage interaction using repetition systems for different reasons (Figure 4.4; Table 4.3) and execute instructions using gestures (Table 4.6).

Table 4.5Listener Facial Actions

	Facial Action Family	**Illustration**		
SN		**Child**	**Sibling Buckets**	**Interactive Function**
1a	Basic Facial Actions	Neutral	Neutral	May indicate relaxed composure, multitasking, and initiation of turn-taking
1b			Neutral face down	Interpreting instruction when visually searching for assembly bits
1c			Neutral concentration on task	concentration on task
1d			Neutral hard	Difficulty in cognitive processing
1e			Neutral face intense	May mental agony, perplexity, confusion and

				determination
1f			Neutral concentration on instruction	cognitive processing
1g		Surprise	Surprise	Indicates disruption of listener activity either by sudden stimulus
1h			Slight surprise	or realisation leading to self-correction
1i		Smile	Felt smile	May indicate positive interaction
1j			Nervous smile	Making the best of a bad situation
1k			Tight-lipped	Positive mask for real feelings
1l			G.W Bush grin	Smirk tending to a negative emotion
1m		Fear	Controlled fear	Used for adaptive control of emotion
1n		Disgust	Slight disgust	Indicates feeling of aversion towards people, things or experience
2a	Non-Basic Blends	Smug	Smug	Indicate a mix of joy and contempt
2b		Angry-Disgust	Angry-disgust	Indicates irritation, negative experience or frustration
3	Emotional Attitudes	Frown	frown	May indicate concentration
4a	Moods	Eureka!	Aha! moment	Indicates self-discovery and problem solving
4b		Workman Face	Work face	Suggests exertion of force when under pressure
4c		Tense mouth and Lip Action	Compressed lips	May indicate concentration, cognitive processing, anxiety, nervousness, mood shift, change of heart
4d			Biting lips	May indicate frustration or concentration
4e			Pouty face	May be used for concentration or self-comfort
5	Eye Action	Static Searching face	Saccade	Used to find or locate pieces in a small space without moving the head
6a	Microexpressions	Hot spots	Anger	Emotional leakages Indicating true emotions depending on the context
6b			Smile	
6c			Tight-lipped smile	
6d			disgust	
6e			sadness	

Basic emotions expressed include felt smile (Figure 4.8) to project enjoyment or happiness while the nervous, partial, tight-lipped smiles (Appendix VIV.7-8) and George W. Bush grin (Appendix VIV.9) suggest that the participants experienced varying degrees of happiness even when embarrassed.In the same vein, others were afraid, annoyed and repulsed by aspects of the interaction as expressed by controlled fear (Figure 4.9), angerand disgust (Figure 4.10) displayed.

Furthermore, participants displayblends of facial expressions (Figures 4.12& 4.13)such assmug and angry-disgust expressions (Ekman & Friesen, 2003) to indicate a mixture of emotions during interaction (Figure 4.21). Sometimes they expressed surprise (Figures 4.7; Appendix VIV.10) during interaction.

Moreover, participants display facial actions that extended Ekman's existing classifications of facial actions to include searching face(Figure 4.14), Aha! Moment (Figure 4.15) and workman's face (Figure 4.16). Participants visually locate assembly pieces using saccadic eye movements in the searching face, project problem-solving strategy using Aha! Moment and when exerting effort, they display a workman's face.

During interaction, participants display unpleasant moods such as frowning (Figure 4.11) as if they were interacting with real people. Participants also compressed lips (Figure 4.17; Appendix VIV.12), or bit lips (Appendix VIV.13) when they were pondering, thinkingor feeling unsure about instructions and/or task.

Although participants' faces were sometimes expressionless (Figures 4.6, 4.7,Appendices VIV.2, 4a, 3, & 4b) during interaction, this does not mean that they were non-challant towards the task, rather, this suggests that theywere paying attention and concentrating as active listenerswhen interacting with instruction-giving agents.

The results also indicate that it is possible for humans to transfer culturally learned skills such as the pouty face (Figure 4.18) to non-human interaction spaces. In addition, the pouty face is used for self-comforting in HAI as against being used to point directions when the hands are busy in HHI.

Participants also tried to hide their emotions during interaction but, their true feelings leaked out as microexpressions of anger, disgust, and smiles (Figures 4.19, 4.20, 4.21, Appendix VIV.14, - 17) respectively.

As outlined in 4.4.1, participants were able to nonverbally project their attitudes towards the instructor (that could not attend to them) and the interaction itself as positive, negative and neutral using the facial actions(Figure 4.3) analysed above.

4.5.3 Assembly Gestures

As outlined in chapter 2, the gestures used by participants to execute instructions are called assembly gestures because they emerged from assembly tasks. They represent listener interpretation of verbal instruction as well as interaction management strategies. The measurable guideline designed (Chapter3.3.1.3; 3.3.2) was used for coding and categorising assembly gestures(Table 4.6a, 4.6b &4.6c)into a hierarchy of six gesture families. Below is a brief description ofassembly gesture families, their forms and specific functions as outlined in 3.2.4.2.

To select assembly bits, participants used different strategies that will assist them in the operative stage. When they decoded the agent's instructions correctly, they use the knowing hand gesture (Figure4.26) without hesitation but when they were not sure, they used the searching hand gesture (Figure4.27) to look for pieces.

Sometimes they picked up assembly pieces and retained them in their hand (Appendix VIV.21b, 21c, 21d) as self-propelled tactics when instructions did not specify what to do with assembly pieces. Furthermore, when listeners picked up more than one piece at a time, they used both hands at once (Appendix VIV.21g), one hand after the other (Appendix VIV.21h, 21i, 21j), one hand to pick upmore than one bit, (Appendix VIV.21m) and both hands to select and set aside (Appendix VIV.21n) assembly bits.

They also used any combination of gestures in succession as concatenated picking hands (Appendix VIV.21a&21d)to select assembly pieces.When picking up from a very tight space they used the two-fingers (Appendix VIV.21m) and when they needed to drop assembly pieces, they used the segmented repositioning(Appendix VIV.22) gesture.

When the assembly pieces are near, participants did not need to raise their hands so they picked ulnarly (Appendix VIV.21l) by turning just the wrist and, when they just needed to push the assembly pieces to a desired location without picking them up, they used the two-hand setting aside gesture (Figure4.28) in a movement similar to one grinding seed on a grinding stone with a top stone.

After selection, listeners measured the fit of assembly bits using aligning hand gestures. There is the basic aligning hand (Figures4.22) which is the default gesture that enableslisteners to visually and mentally assess the fit of one or more assembly parts with others.Listeners measure fit by making assembly piecestophysically touch one another very quickly (Appendix VIV.19a), using multiple strokes (Appendix VIV.19b), and repositioningassembly bits (Appendix VIV.19c).While these gestures enablelisteners to assess how assembly pieces fit, they also externalisedlistener's thought processes for co-interlocutors to observe.

After aligning assembly bits, participants attach them using joining hands gesture (Figure4.29). Participants used a more careful approach (Appendix VIV.23a) to attachand adjusttinyassembly pieces.Participants also used multiple joining hands gestures with embedded adjustments (Appendix VIV.23b; 23c) to measure fit and attach assembly bits. Listeners use twisting hands gestures(Figure4.78) toposition and give assembly bits specific orientation.However, when participants are unable to attach pieces together, failed joining hands gestures (Figures4.30;Appendix VIV.23d)are said to have occurred.

Listenersuse presentation gestures to placeassembly piecesand workdone before their eyes for assessment. When simply done laterally or supinely it is the basic presentation (Figure4.23; Appendix VIV.20a-20c) but when done with the fingertips randomly pulling and horizontally holding up assembly pieces, this is a checking hand (Figure4.24).When listeners turn their hand 180^0 with the thumb and fore-fingerlaterallyholding assembly pieces while other digits arecupped, this is presentation with fingertips (Appendix VIV.20h). In addition, when listeners turn around a piece to give it a 360^0inspection they usecircular presentation (Appendix VIV.20d). They combine dual picking and presentation in one gesture (Appendix VIV.20g) to select and examine assembly pieces concurrently and use repetitive opening hands (Figure4.25) to assess comprehension. Participants enact presentations with retractions or adjustments (Appendix VIV.20f) and use their fingertips to bring pieces close enough to measure their fit while turning them (Appendix VIV.20e).

Participants can communicate, manage interaction and build rapport using assembly gestures. As outlined in Chapter4.4.4, when participants preempt and forcefully seize ground from agent-instructors they use positive turn-taking violation gestures (Figure 4.35) in a way similar to how people interrupt others in discussions.

However, when they make mistakes in any of the assembly stages, they carry out self-initiated self-repairs (Figure 4.37,Appendix VIV.24) during interaction. Participants also use decisive pointing movements or deictic hands (Figure 4.40) gesture to locate 3D points and positions on the assembly kit when executing assembly tasks such as the twisting hand (Figure 4.33).

When the agent gives instructions, participants are sometimes able to interpret and mentally process them with iconic representation of key words using the deictic co-interpretation gestures (Appendix VIV.25a). Participants locate, select, and set aside assembly bit using the deictic enumerating hand gestures (Appendix VIV.25c). Participants produce iconic representation of the agent’s instruction, locate assembly points and measure the kit within a 3-D space using the deictic representation hand gesture (Appendix VIV.25b).

During the assembly process, listeners use some gestures to externalise the psychological processes involved in their decision-making and error correction in an observable manner during interaction. For example, participants can terminate gestures half-way or intentionally drop and replace assembly piecesusing false start gestures (Figure 4.39). Furthermore, when participants unconsciously make errors and they try to compensate for these using nonverbal hedging gestures (Figures 4.41) and unconsciously mitigate information overload when processing instructions using the vibrato-based gesture (Figure 4.34) When participants became anxious they used the deictic hedge gesture to regain self-control and composure (Figure 4.38).

Table4.5SpontaneousAssembly Gestures in HAI

Gesture family	Sub-Group	Description	Function
Aligning Hand	**Basic aligning hand**	**Assembly bits are brought close but they do not touch**	**Enables listener to visually and mentally assess the fit of one or more assembly parts into others**
	Micro-aligning hand	**Assembly bits touch**	**Enables listeners to physically and visually measure the fit and size of the assembly pieces**
	Extended aligning hand	**Executed with four strokes**	**Enables listeners to measure fit of assembly pieces. It externalizes the user's mental strategy in processing information.**
	Aligning hand with repositioning	**Done with simultaneous clockwise and anticlockwise twists**	**Enableslisteners to measure fit physically within a 3-D space**
Picking Hand	**knowing hand**	**A beat gesture executed without hesitation**	**Enables listeners to select an assembly piece and suggests correct decoding of instruction**
	Searching hand	**Executed with hands wavering. It is the opposite of the knowing hand**	**Indicates that listeners are unable to select assembly kit due to incomprehension ofagent's instruction**
	Basic picking hand	**Executed with one hand**	**Enables listener to select an assembly piece**
	Picking with a hold	**One hand pulls up and keeps the assembly piece without dropping it or dropping it into the other hand**	**Enables listener to select and retain assembly pieces with one hand**
	Concatenated picking hand	**Executed with multiple gesture phrases**	**Enables listener to identify, isolate and pull up assembly pieces**
	Picking and aligning hand	**Pulls up assembly pieces then move them close**	**Enables listener to select and assess their fit**
	Dual picking hand	**Executed with both hands at the same time**	**Enables listener to select two pieces at once**
	Alternate picking hands	**Executed using both hand sequentially with one going after the other. As one picks and is retracting, the other goes forward to pick**	**Enables listener to select multiple pieces in an orderly manner**
	Segmented repositioning hand	**Done with alternate picking hands and dropping hand**	**Enables listener to pick and drop assembly pieces with both hands then place them with one hand in the desired location**
	Picking ulnarly	**Picking movement is executed with the wrist movement towards the pinky side of the palm**	**Enables listener to select pieces that are near**

Table4.6 SpontaneousAssembly Gestures contd.

Gesture family	Sub-Group	Description	Function
Picking hands	**Two-finger picking**	**Executed with a combination of any two fingers of the hand**	**Enables listener to select tiny assembly pieces from a small 3-D space that may be too small for the whole hand**
	Double picking hand	**One hand is used to pick more than one piece sequentially or one after another**	**Enables listener to select more than one assembly piece using one hand**
	Two- hand setting aside	**Executed with both hands without picking the pieces up using a movement similar to that used when people grind seeds using a muller on an indigenous grinding stone**	**Enables listener to locate and push assembly pieces to the desired position.**
Presentation	**Basic presentation or displaying hand**	**Executed with the supine hand**	**Enables listener to display the assembly piece on the flat of the hand**
	Checking hand	**Executed with fingertips randomly selecting and holding upassembly pieces horizontally**	**Picking and showing listener assembly pieces for confirmation**
	Repetitive opening hand	**Done with palms moving in cycles to remain supine or palm up more than once**	**Continuously placing the same assembly parts before listener for initial inspection many times**
	Fingertips presentation	**Executed with palm adduction fingertips lateral, cupped and facing listener**	**Enables listener to assess assembly pieces**
	One- hand presentation	**Executed with the hand holding the piece turning lateral or supine**	**Enables listener to place assembly bits for inspection using either the right or left hand**
	Circular presentation	**Done with the fingertips holding and turning or twirling the piece around continuously**	**Enables listener to inspect a cylindrical assembly bit at 360 degrees**
	Presentation for alignment	**Executed with fingertips of both hands turning and bringing the pieces close**	**Enables listener to examine and measure fit of assembly bits at the same time**
Communicative Gestures	**Positive turn-taking violation**	**Done with pre-emptive execution of an instruction not given correctly**	**Enables listener to preempt the agent by forcefully seizing ground albeit positively**
	Self-initiated self-repairs	**Proceeds in these stages:trouble spot identification, repair initiation, repair refinement and actual repair**	**Enables listener to correct joining errors during interaction**
	Extended self-initiated self-repairs	**Proceeds in three stages: trouble source identification contains two full gesture phrases before repair initiation and actual repairs take place**	**Used to correct picking errors**
	Deictic hands	**Decisive pointing movement**	**Enables listener to locate 3D points and positions**
	Deictic Co-interpretational	**Fingers measuring parts of the kit when the instructor is uttering the adjective**	**Enables listener to interpret and mentally process instruction using iconic representations of specific words in the instructor's utterance.**
	Deictic enumerator	**Executed with fingers flicking assembly pieces on the Table**	**Enables listener to locate, select and set aside assembly bits**
	Deictic representation	**Flexing the forefinger and thumb bidirectionally while moving up and down**	**Enables listener to produce an iconic representation of the agent's instruction, locate assembly points and measure the kit within 3-D space**
	Deictic hedge	**Enacted with wriggling of the fingers or vibrato and pointing fingers**	**Enables listener to control anxiety during interaction**

Table4.6 SpontaneousAssembly Gestures concluded

Gesture family	Sub-Group	Description	Function
Joining Hands Operative Gesture	**Basic joining hand**	**Executed with both hands placing, pushing, sliding one piece into another**	**Enables listener to fix assembly pieces together within a 3D location and appropriate orientation**
	Multiple joining hand	**Same as basic joining hand except that the process is repeated more than once**	**Enables listener to adjust assembly techniques when attaching parts that come in pairs such as the legs.**
	Failed joining hand	**Occurs when attempts at enacting joining hand are unsuccessful**	**Enables the co-interlocutor to observe the listener's inability to attach any assembly bit. Enables listener to initiate repairs.**
Others /Sacrum	**Crooked aim**	**Executed with the forefinger and thumb forming a crooked hand**	**Enables listener to prepare for picking. Probably a comforting move similar to the vibrato and self-touch**
	Self-restraint	**Given off by the swallowed lips and highly controlled movement of the hands.**	**Enables listener to exercise care in handling fragile assembly bits**
	False starts	**Termination of picking gestures halfway and/or picking and dropping pieces akin to the jerky hand movements used by two people trying to pick an object at once.**	**Hesitation by listener to assess perceived functional and psychological threats. Externalises the user's self-interruption of the assembly task.**
	Top-down keeping hand	**One hand selects the assembly piece then drops it into another , for example, left hand picks and drops into the right hand.**	**Enables listener to select specific pieces from among many when destination is not specified**
	Twisting hands	**Participants' hands hold already assembled bit then move simultaneously in opposite directions , for example, right hand moves clockwise while left hand moves anti-clockwise**	**Used to wrench assembled bits apart in the repair process**
	Vibrato dropping Hand	**Occasioned by mistakes similar to mishits when hammering a nail. It begins with wriggling of fingers (vibrato-like movement) then the piece drops unintentionally**	**Nonverbal hedge (vibrato) externalises listeners' attempts to mitigate communication issues and accidental dropping of assembly bits in much the same way that observers notice when objects drop from people's hands unintentionally**
	Concatenated mixed hand	**An eclectic gesture that combines many gestures to execute one instruction**	**Used to break tasks into manageable chunks, confirm listener cognitive processing of instructions and assists the next assembly stage to mention but a few**

Thesacrum family is a mixed bag of gestures because they perform various functions that cannot be classified in any other family. There is the crooked aim gesture(Figure4.31)from crooked hand used by participants to prepare for picking up using a movement akin to one taking aim before throwing a stone at something. In addition, participants exercise self-restraint (Figure 4.38) with controlled pulling or pushing during the joining hand gesture signalled by swallowed lipsand others use eclectic gestures(Figure4.35) to break-up tasks into manageable chunks.

4.7Conclusion

This Chapter assessesspontaneous listener nonverbal listenershipbehaviour whenpeople receive instructions from agents during assembly tasks.The results suggest that humans can use nonverbal listenership behaviour to manage interaction, indicate their perception of and attitudes towards instructors and indicate their comprehension or incomprehension of agent-instructions in HAI even though, the agent does not attend to these.

Specifically, humans use emotions to provide information to co-interlocutors about events, responses and probable next behaviour (Clark et al., 2016a) because emotion enables organisms to deal with interpersonal encounters as they emerge and adapt to deal with them in the future.In addition, findings suggest that the causal relationship between humans and agents in HACs may be effectively managed when humans display nonverbal listenership behaviours. For example, participants' facial actions may project their perceptions of the agent as positive or negative. This confirms earlier research indicating that emotions are crucial to the development and regulation of interpersonal relationship because people that are incapable of facial expressions may have great difficulty in managing causal relationships(Ekman& Oster,1979).

Although humans sometimes display neutral faces that ordinarily may indicatelack of emotion, the results suggest that people may not be completely neutral even when their faces are seemingly blank because they have traces of other emotions. The results also indicate that even though humans try to conceal their real feelings about the interaction, their feelings still leak out as microexpressions during interaction with the agent.

The gestures (Table 4.6) emerging from the study indicate that it is possible for humans to show their intentions to agents in a finely grained nonverbal manner (Kirk et al., 2005). The taxonomy of repeats may also indicate the listener's interactional needs from time to time when dealing with instruction giving-agents because repetition may signal and resolve communication problems.

In addition, humans can operationalise agent instructions using gestures just as they can communicate with the agent and specify their preferences nonverbally during interaction. For example, communicative gestures such as self-repairs support the processes of meaning-making and rapport-building during interaction.

The results presented suggest that humans interact with agents probably as they do with people so, there is the need to observe instruction-giving between humans to assess the extent to which HAI and HHI are similar or different. This may enhance our understanding of the communication, interaction and listener nonverbal feedback during interaction in instruction-giving contexts.

Chapter 5 Assembly Instructions in Human-HumanInteraction (HHI)

5.1 Introduction

Study 1 tried to understand to the type of communication that takes place when humans take assembly instructions from agents within the context of HAI. Forty-eight speakers of English as a first language were recruited for the study. Their ages ranged between 18-35 years. They were instructed by a simulated agent to assemble two Lego models using a continuum of three voices –CL, CP and HR – the first two are synthesised and a professional voice actor recorded the last one. Each of these voices had an educated Southern England accent and an age range of 45-55 years. Ninety-six interactions were recorded on video thus producing twenty-four hours of multimodal corpora.

Results from Study 1 indicate that, in HAI listeners ask for instructions to be repeated for several communicative and interactional reasons using a systematic process that is sequential with distinct characteristics. The results also suggest that assembly gestures are actionable moves that may occur alone or co-occur with facial actions aimed at successful execution of agent instructions, while facial actions may indicate listener perception of the agent's identity and listener attitudes towards it.

As outlined in Chapter 4.7, there is the need for another study that will look at interaction from a purely human angle. The first study did not use a 'real' human voice and followed the trend of researchers in HCI to focus on assessing the use and role of synthesised voices, human voices, recorded voices (Esposito et al., 2015; Litman & Forbes-Riley, 2006; Nass & Lee, 2001) as individual voices or any two combinations of the three but none of these used a continuum that cuts across four voice strands.

Thus, there is the need to expand the continuum by including a human instructor (HV) during interaction (Figure5.1overleaf) thus creating the context ofHHIbeside HAI. This study may provide an experimental basis for understanding spontaneous listener nonverbal feedback when taking instructions from a human instructor during assembly tasks. In addition, the study will hopefully enable comparison of listenership in HHI with HAI along a voice continuum.

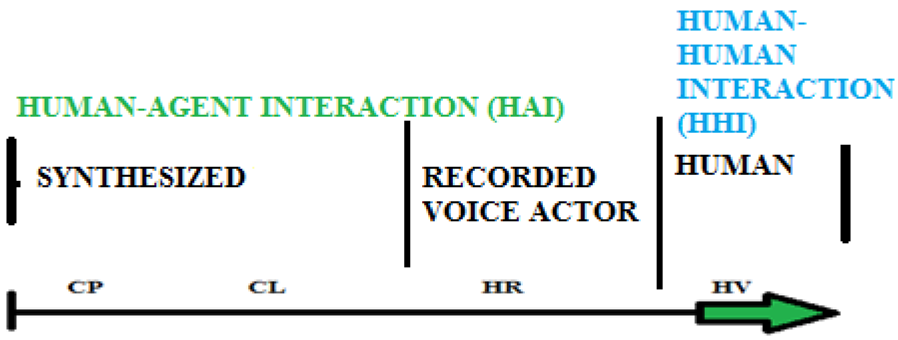


Figure5.1 Expanded Voice Continuum

The continuum presumes that interaction within HACs may take any of these forms HAI; HHI and Agent-Agent Interaction (AAI); however, the focus of this thesis is on the first two (Figure 5.1). In addition to the three voices used in Study 1, a human being giving instruction expands the continuum to four voices. This makes interaction natural and familiar because participants are used to taking instructions from people in various contexts.

5.1.1 Research Questions

The study hopes to answer the following questions:

1. What facial actions do listeners display when requesting repeats of instructions during assembly tasks and why?
2. What gestures do listeners display when requesting repeats of instructions during assembly tasks and why?

5.1.2 Statements of hypothesis

In view of the research objectives and problems identified, the study hopes to test the following hypotheses:

H_1 There is a significant relationship between the listener's facial actions and their execution of verbal assembly instructions

H_2 There is a significant relationship between the listener's gestures and their execution of verbal assembly instructions

5.2 Instruction-giving in verbal interaction

Interaction in HAI is characterised by instructors giving verbal instructions while listeners respond nonverbally, in contrast the HHI context has a higher element of verbal interaction because listeners request for instruction repeats verbally. In view of this, this thesis briefly outlines the features and processes involved in verbal interaction in contrast to nonverbal interaction. This will enable an informed comparative analysis of HAI and HHI in instruction-giving contexts.

While verbal interaction comprises written and oral communication, for the purposes of this thesis verbal interaction refers to oral communication only.Verbal interaction in the context of this thesis is developed around listener request, speaker reading of written instructions aloud, repetition of instructions-(when listeners need clarification they ask instructors to repeat the instructions) and listener execution of the instruction.

On the part of the speaker, readability of written documents is essential.Badrudeen & Sabhawarwal, (2010, p.2574) argue that there is no acceptable standard for assessing readability even though many formulas have emerged including software packages such as Microsoft® word® package (Microsoft Corp, Redmond, WA). This implies that readability depends on the structure of the text and context of use. The structure of instructions outlined in chapters 2.3 and 3.2.1 improves text readability and clarifies information conveyed during interaction.

Ribes & Rodriguez, (2001, p.309) found that humans may not pay attention to instructions unless they are unable to execute them successfully which makes them to adjust their performances. This implies that participants handling tasks may initiate responses that enable them to correct initial task errors just as in verbal errors during interaction.

Visual communication includes texts such as painting, cartoon, coins (Gross, 2009, p.149-152) and this thesis extends this to facial actions and gestures in comparison to audio mode of interaction. Regarding the mode of interaction in this thesis there is a continuous shifting of modality as the same text is passed from speaker to listener. For the instructor, the text (written instructions) is visual while the listener recieves it as a verbal text (spoken instructions) but responds visually (using facial actions and gestures).

Gross (2009,p.149) suggests that there are marked differences between verbal and visual interaction as briefly summarised as follows. Verbal text is processed sequentially while the nonverbal are processed not only sequentially, but also in parallel and simultaneously.

Secondly, words never lose their identities in contrast the identity of images can easily be submerged in larger wholes, as eyes and ears get submerged in a familiar face. This suggests that when facial actions and gestures co-occur with verbal communication and a nest of multimodal communication, their individual characteristics pale into insignificance. It is the view of this thesis that, analysing individual facial action and gesture emphasises their individuality although interpretation requires that they be considered within the context of other communication modes they co-occur with (Chapter 3)

Thirdly, unlike words, nonverbal communication can undergo meaningful spatial transformation and manipulations. This is evident in the transition of emotions and gestures that may occur during interaction , for example, the role and effect of emotional leakages outlined in chapter 3.

Furthermore, there are internal connections existing among the fundamental components of verbal communication as shown in language use, in contrast, visual texts are organised spatially into synchronuous hierarchies. As per facial actions and gestures, this thesis argues that they also have internal consistency among their components , for example, gesture phrase has a standard structure (Chapter 3) even though they may be organised spatially.

However, when the same verbal message is relayed, the speaker's differing expressions may alter the meaning , for example, saying "I'm sorry" with a smirk, sadness, or a smile will mean different things to an interlocutor (Gottman et al., 977) just as the meaning of gestures may alter with the listener's differing expression and the interpretation of facial actions may also be altered when the circumstances sorrounding their occurrence are congruent or incongruent with the context of interaction.

Verbal interaction is characterised by turn-taking (indicatingGrice's principle of cooperation in a conversation) and has definite and pragmatic inter-party moves (the nature and type of interaction characterising a turn i.e apology, denial, or excuse extended in this thesis to requests for repetition of instructions

Intertextuality i.e. the intensity and quality of verbal interaction between interlocutors(Brennan et al., 2013, p.665) is exemplified in this thesis by the patterns of repetition initiated at the behest of the listener. This follows research indicating that repetition have been used to reframe arguments in public and private (Tannen, 2006, p.597-600)

Tannen(2007) makes a distinction between synchronic and diachronic repetition. Synchronic repetition is the recurrence of words and collocation of words, within a conversation or text. Diachronic repetition is the recurrence of words in discourse which occurs at a later time (Tannen, 2007, p.2-9). Tannen used this distinction to explain the notion that texts are often embedded in specific contexts and related to other texts in situations where the same topics appear in subsequent discussions.

In another study, Brennan et al(2013) used synchronic and diachronic repetition to analyse the interactive and dynamic nature of organisational response to stakeholder activism as a way of managing conflicts. Their study concluded that communication in an interactive process that is a power relation between a firm and a specific stakeholder that is akin to a conversation.

Following these, it is the view of this thesis that, the relationship between an instructor and instructee may not only be conversational but, involve power relations that can be made apparent with the patterns, role and functions of repetition during interaction.Verbal communication as a primary tool of communication beteween people is used pragmatically to express ideas, desires and concepts during interaction. However, the focus of this thesis is nonverbal communication as projected through listenership behaviours and it is of interest to see how listeners use facial actions and gestures as pragmatic markers indicating listener desires, ideas and conceptualisation of interation in instruction-giving contexts.

5.3. Study Methodology

5.3.1 Instructor

As outlined in Chapter 5.1, the instructor is the voice of a real human being at the end of the continuum (Figure 5.1). The voice is produced as a normal human sound used in normal conversation by a manand speaker of English as a first language from Southern England.

Instructions are delivered without the use of any intermediary media during interaction in this study.

5.3.2 Participants

Six participants that are speakers of English as a first language were recruited for the study as outlined in Chapter 3.2.2. Participants were remunerated with a £10.00 Amazon voucher for their time. There were two males and four females together they have an average age of 23.7 years.

Table5.1 Grouping Checklist HHI

S.No	Participant Ref	**Model**
1	IH1_H_H_Aq_N	Aqua/Nex
2	IH2_H_H_N_Aq	Nex/ Aqua
3	IH3_H_H_Aq_N	Aqua /Nex
4	IH4_H_H_N_Aq	Nex/Aqua
5	IH5_H_H_Aq_N	Aqua/Nex
6	IH6_H_H_N_Aq	Nex/Aqua

The task is counter-balanced using the Lego models; for example, Participant IH1 will assemble Lego model Aquagon (Aq) first then Lego model Nex (N) last but, participant IH2 will assemble Nex first and Aquagon thereafter. Counterbalancing enabled all participants to have equal interaction with the instructor.Although, participants are few which limits the extent to which the study results can be applied, but their interactions provided a purely human angle to understanding the dynamics of unidirectional instruction-giving in assembly tasks.

5.3.3 Experiment Procedure

The task procedure is the same as the one used in Chapter 4 except for the presence of a human instructor during interaction. Participants were informed that video-recording equipment was used with their permission granted later in the signing of the consent form (Appendix IIa). The information form (Appendix I) was presented to participants for reading. Once they had read this they were asked if they had any more questions or clarification requests regarding their involvement in the study.

Before each session, participants were requested to fill out a short consent form (Appendix IIa) and were informed that they were to construct two different Lego models using spoken instructions from a human instructor and they could ask the instructor to repeat instructions as many time as they want but, they could not go back to previous steps.Andin line with the approved research ethics, they were free to withdraw from the experiment at will. All participants gave their consent for their interactions to be recorded and the data used in the thesis and academic purposes only. Participants were also informed they had 15 minutes to get as far as they could in each task. A timer was set so participants could keep track of this time limit. After the debriefing, the first model was presented and the task began. Once the time limit expired or the model was complete, the task was deemed to be over. This procedure was repeated once more with the second Lego model and participants were free to leave. During interaction, the researcher took notes on the instructions repeated, time taken, pieces not used and organising the data.

Post-task procedures involved uploading data from recordings to the secure ORCHID server then wiping the same from the equipment. Lego models were disassembled, and pieces put back into place for the next study (if another one that day) or taken back to storage in the office. Field notes taken were also uploaded and filed away in secured cabinets.

5.3.4 Data Collection and Analysis

Each session was recorded from one angle instead of two because the side-view Canon Legria HFR306 camerafailed to store recordings into the memory source. The researcher did not notice this until later and several attemptsto make it function failed. All the other laboratory cameras were fully engaged so, there was no camera replacement available. Rescheduling time and dates with participants was out of the question because it was difficult enough to get them in the first place so, to avoid loss of venue and disappointing participants that have already been given time and dates, the remaining sessions just had to go ahead as planned.

The inability to capture interactions from the left side angle led to data loss. However, Easton et al.(2000) suggest that when faced with equipment failure back-up plans should be used. The back-up plans included alternatively panning and close up shots with the Panasonic HDC-SD900 to captureeye level shots of the participant's face and the workspace. This made it possible to see the participants' hands assembling the Lego kit.Another back-up planadjusted for data loss by ensuring that in addition to repeated instructions, every assembly stage was analysed thus finding a balance between the depth in this study with the breadth in the first because the participants were fewer in the second study. As already explained in Chapter 4.3.5, the MP4 format was used without compromising greatly on quality.

Figure5.2 A Participant Taking instructions to assemble one Lego model

A multimodal corpus of three hours of video recorded interactions between participants and the instructor was built. Each interaction was labelled with a proper reference like this IH (participant No.) _Lego kit Model_Task Number_HH_file format and stored in a raw videos folder. For example, a raw video was thus referenced, IH1_Aqua_Task 1_HH.MP4. The corpus was analysed systematically as already outlined in Chapter3.2.4.

5.3.5 Decisions on Research Hypotheses

As outlined in Chapter 1 the study set out to test two hypotheses as follows:

> H_{1b} There is a relationship between the listener's facial actions and whether they successfully execute the instructor's verbal instructions
>
> H_{2b} There is a relationship between the listener's gestures and whether they successfully execute the instructor's verbal instructions.

Six (6) facial action categories and thirteen (13) gesture categories were sampled respectively for the inter-rater reliability assessment. The percentage agreement between the two annotators for spontaneous listener facial actions and gestures displayed during interaction is 96% and 88% respectively. The resulting Kappas indicating very good or almost perfect agreement,fall within the Landis and Koch benchmark of .81- 1.00 and the Altman benchmark of .81- 1.00 respectively (Gwet, 2012).

The percentage agreement values tally with the following coefficient correlation value. IRR coefficient correlation for gestures displayed by participants indexed by Fleiss Kappa (α) is 0.79 while the Krippendorff alpha (α) is 0.80 implying a substantial and excellent agreement thus agreeing with the percentage index of .87. Equally, the IRR coefficient for facial actions displayed by participants indexed by Fleiss kappa (K) is 1.00 while the Krippendorff alpha (α) is also 1.00 implying an almost perfect agreement which is the same as the percentage index of 100%. The values obtained from percentage agreement, Fleiss Kappa and Krippendorff alpha indicate that there is a very high agreement between the first and second annotators' perceptions of facial actions and gestures displayed by listeners during interaction.Thus, the study results could be accepted as trustworthy and reliable for use in testing the hypotheses of this study.

As per H_{1b}, spontaneous listener facial actions displayed in HHI indicate that changes in listener emotions during tasks may be related to a listener's ability to successfully decode instructions during tasks. Thus, when listeners are successful, they display positive facial actions representing positive attitudes but, when unsuccessful, they display partial negative facial actions suggesting negative attitude and when paying attention to task or instruction, they remain neutral.

The patterns of listener attitudes towards the interaction as expressed through basic facial actions, facial task strategies (oculomotor movement) and emotional transitions suggest that there is potentially a cause-effect relationship between listener facial actions and their successful execution of verbal instructions. Thus, there is sufficient evidence to support the claim that there is a relationship between listener facial actions and their successful execution of the instructor's verbal instructions and the hypothesis is hereby accepted.

Regarding H_{2b}, the results suggest that listener gestures have task-oriented and communicative relationships with their successful execution of the instructor's verbal instructions. This is because listener gestures focus on assembling the Lego kit and are used to externalise listener-thought processes as they cognitively process instructions and initiate self-initiated self-repairs during tasks. Thus, there is sufficient evidence to support the hypothesis that there is a relationship between the listener's gestures and whether they successfully execute the instructor's verbal instructions and the hypothesis is hereby accepted.

5.4. Results

The results of the study cover repeats of instruction as a marked phenomenon during interaction, facial actions and gestures.As outlined in Chapter 4.4.2, only representative samples of nonverbal behaviours are presented within the Chapter while the rest are in the appendices (Appendix X) however, they are cross-referenced in the analysis and discussions. As per facial actions and gestures the circumstances sorrounding their occurrences and how they transform to meaningful multimodal codes with linguistic and interactional functions is outlined in 3.3.1.3. Similarly, the results are descriptive with conclusions limited to individual occurrences of nonverbal behaviour but, these results are compared with Chapter 4 to get study conclusions and implications in Chapter 6.

5.4.1 Quantitative Results

There are 48 instructions for Aquagon and 49 for Nex models; 6 participants had two iterations each.Table 5.2 below shows that the participants received 582 instructions from the instructor but asked for repeat instructions 67 times shown(Table 5.2) below.

Table5.2 Instructions Repeated

	INSTRUCTIONS	REPEATS
VOICE	HV	HV
AQUAGON	288	30
NEX	294	37
TOTAL	582	67

Table 5.2 supports Carter’s (2004) assertion that repetition is an ever-present aspect of human interaction even in unidirectional instruction-giving interaction. Table 5.2 above indicates that Aquagon had 30 while Nex had 37. Although, the number of participants in this Study 2 is lower than those in Study 2, the most repeated instruction in both studies for the Aquagon model was instruction number 14, while Nex was number 24, indicating that these may be the most challenging instructions for participants to comprehend.

The results (Figure 5.3) below indicate thatthe participants displayed facial actionsduring interaction. This re-echoes earlier research findings suggesting that the capacity to experience emotions may be a fundamental human strength (Fredrickson, 2001).

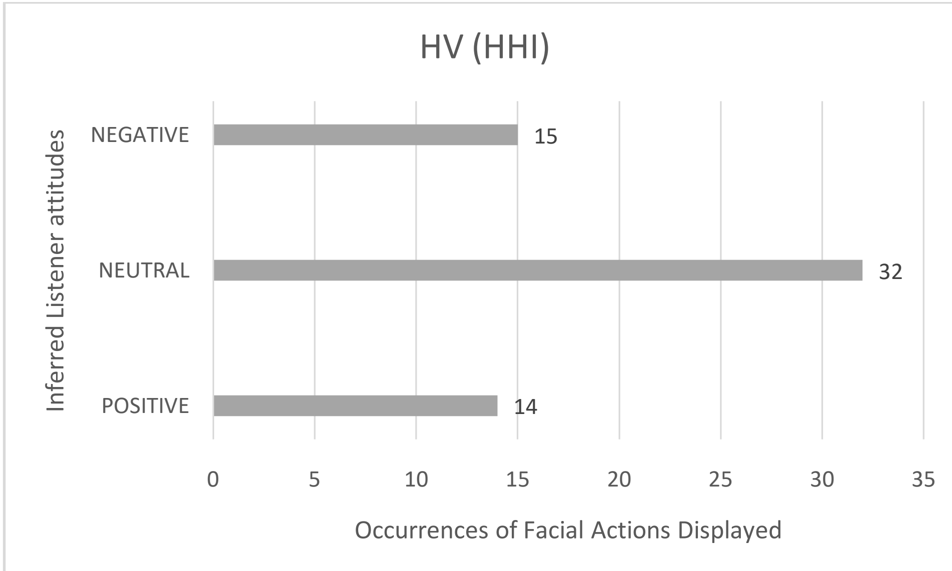


Figure5.3 Listeners' Inferred Attitudes to Interaction

The results indicate that positive facial actions include felt smile, slight smile, and controlled laughter representing 18% of the general distribution. Neutral facial actions include neutral, neutral concentration, workman effort face, and static searching head representing 57% of the facial actions displayed. Negative facial actions represent 25% of the distribution and include puzzled face, compressed or swallowed lips, disgust, slight disgust, slightly compressed lips, micro-frown and nervous smile.The results suggest that attitudes might be reliably detected and measured through facial expressions just as Meadors and Murray (2014) measured and classified bias through body language. Listener attitudes towards the interaction may also be distinguishable as positive, neutral or negative (Mehrabian & Ferris, 1967).

5.4.2 Qualitative Results

5.4.2.1 Emerging verbal interaction Pattern

The instructor was asked to read out instructions while listeners could ask for the instructor to repeat instructions as many times as required but they cannot ask for previous ones (Chapter 5.3). The emerging interaction pattern observed in HHI-context indicates that the instructor reading out the instructions and listeners executing instructions displayed features of spontaneous conversation during task. This chapter will outline the features relating torepetition as a joint action, listener verbal actions and instructor verbal actions from the interaction.

As outlined in Chapter 4.4.2, repetition is an important aspect of human interaction that is ever-present in human language and tied to linguistic theories of repairs, taskexecutionand listenership. Results from this study suggest that the process of repeating instructions in this study is discernible, arbitrary, and involves diverse events. The following are the features of the repeat process observed in this study.

Repeats are discernible: Repeats are noticeable when they occur, for example the instructor or speaker re-vocalising same instruction more than once makes it audible to listeners.

Repeats have variety: Repeats in HHI come in many forms and varieties but the listener uses them consistently and reliably during the assembly process as rational communication strategies that promote active listening and set a communication standard (Stack, 2012; Oxford, 1993). The results suggest that repeats always serve a communication function as will be seen later in the typology of repeats (Chapter 5.3.2.2).

Repeats involve diverse events: As outlined in Chapter 1.6, interaction between the instructor and instructees is a form of joint action with the repeat process being an integral part. The repeat process in HHI involves the listener making a request and the instructor repeating the instruction. However, unlike in HAI where there is only one way of asking for a repeat (pressing the repeat button), participantsin HHI use four language frames (a selection of words and sentence construction that fit specific pragmatic functions of language use during interaction such as making requests) to request repeats.

1. The most common method is for the listener to speak out using the most common verbal phrase, for example, "Could you please repeat that?" or any utterance within this framework in structure or meaning. Then the instructor repeats the given instruction.
2. Another verbal approach is the completion drill. This term is used in the sense that it is applied in the teaching of grammar. Pattern drills are used to provide adequate repetition of language structures in meaningful contexts to establish correct habitual response in learners (Paulston, 1971). The completion drill in any form requires that the teacher utters one part of the grammatical structure and the learner completes it (Harmer, 2007). Similarly, the listener repeats part of the instruction in the form of a

question, for example, IH3 *"Just place the big spikes?"* And the Instructor repeats the whole instruction "*26. Just place the big spikes into the holes that are closest to the edge of each piece."* Another variety of the completion drill ends with a conjunction, for example, IH4 says *"Just place the big spikes into the holes…and"* then the instructor repeats the instruction by completing the remaining part like this: *"…and… that are closest to the edge of each piece."* Both have reduced the instruction that should have been rendered as *"Just place the big spikes into the holes that are closest to the edge of each piece."*

3. Another version is the question tag. Here the listener repeats part of the instruction but ends it with a Wh-question tag. For example, IH5: *"Just place the big spikes where?"* and the Instructor repeats the instruction.
4. Nonverbal requests: it was also observed that listeners use backchannels like 'umhmm', smiles or head nods to indicate that they are ready to take the next instruction. This is like pressing the 'next button' on the interface that makes the agent give listeners the next instruction.

Unlike in HAI where instructor action prompted by the push of a button generates the repetition of whole instructions(Chapter 4), instructor verbal action in HHI involves a variety of ways to repeat instructions that are partly couched by listener language frames and characterised with speaker disfluencies as outlined below

Instructor responses to listener request for instructions also have variety: the results indicate that the instructor repeats instructions in different ways even though they were written down for him to read from. The basic and most common response to a request is restating the whole instruction as it is written down on paper. Sometimes the instructor just repeated the critical aspects focused on by the listener by paraphrasing. For example, when a participant asked, "Just place the big spikes where?" The instructor replied, "Into the big holes" instead of repeating the whole instruction "Just place the big spikes into the holes that are closest to the edge of each piece."

Speaker disfluencies: disfluencies are hesitations, repairs, insertions or deletions that occur when people read or speak (Shriberg, 2001, p.156). Although the instructions were written and the instructor was expected to read them out, results indicate that speaker disfluency occurs during interaction and this may have posed comprehension challenges for listeners.

I. Disfluency occurred due to tracking errors (Shriberg, 2001, 1994). Tracking error occur when instructors forget where they stopped reading in the text to continue from there. For example, speaker disfluency occurs as the instructor could not articulate the vague Nex instruction 39 for IH5 when requested due to a tracking error. The instructor had to go back to instruction 38 before repeating instruction 39.

II. There was also insertion as the instructor inserted the word ‘Then’ into the instruction when giving Nex instruction 29 to IH4 in the first instance. However, this was corrected when repeating the instruction as speaker repairs.

III. Another form of disfluency observed is that of omission of parts of the instruction. In one instance, the instructor gave only the first part of the instruction in the first instance as if that was the whole. When asked to repeat the instruction, he then read out the complete instruction.

Listener expectation has been set in the pattern of preceding interaction where the instructor gives the complete instruction before the listener reacts. A violation of this pattern by the speaker may constitute information overload for listeners processing verbal instructions in real time.

5.4.2.2 Typology and Sequences of Repeats

As outlined in Chapter 4.4.2, the bases for classifying repeats include timing, listener or user action and reasons deduced from listener listenership behaviour. Timing refers to when the repeat occurred whether before, during or after execution.

Listener action is linked to timing whichdescribes what listeners are engaged in while instructions are repeated. As outlined in Chapter 4, the listener’s reason for repeating instructions is deduced from spontaneous nonverbal listenership behaviours displayed, timing, listener action and reasons deduced from observing and analysing the interaction context during interaction using the model outlined in 3.3.1.3. Thisbecausepeople have a tendency to forget exact details of interactions after they have taken place and they tend to re-construct events that may lead to inaccuracies (Bach& Goncalves, 2004; Norman, 1988).

Table5.3 Typology of Repeats in HHI

Type	CODE	GENERAL FUNCTION	SUB-CODE	SPECIFIC FUNCTION
Clarification	R1	For self-correction	1a	Clarification leading to self-correction
			1b	After clarification, listener makes a task error then self–corrects, for example, picks the wrong piece then drops it for the correct one
Clarification	R2	Used to reduce confusion and clear doubts. Usually occurs after listening to the instruction.	2a	Clarification for comprehension
Composite	R4	Multiple repeats for different purposes but with a focus on executing one assembly step	4a	Repeats confirm actions when they occur after the action has taken place and the listener can be seen inspecting the work done while the instruction is repeated
Demarcation	R5	Used to break up one assembly stage into chunks		

Table5.3shows that four (4) types of repeats emerged from the study. These typologies are described below:

- R1 types of repeats are meant for clarification that may lead to listener self-correction. There are two sub-classes of R1: R1a) clarification occurs before the user executes the current instruction/assembly task leading to self-correction; R1b) after repeating instruction for clarification, listener makes a task error then self-corrects; for example, listener asks for a repeat of the instruction, then picks up the wrong piece (black instead of yellow) then drops it and picks up the correct one (yellow). Sometimes the dropping and picking action is accompanied by private speech like 'wrong piece' muttered under the listener's breath.

- R2 types are clarification repeats that reduce confusion and enhance user comprehension. They usually occur after listening to the instruction for the first time and the listenerwill be seen listening again without doing anything.

- R4a is used strategically to aid user execution of instructions. For example, IH5 uses R4a to confirm actions that have already taken place and listenerscan be seen inspecting the work done while the instruction is repeated.

- R5 are repeats that are used for demarcation. Listener use R5 to break assembly process into chunks to simplify assembly tasks and check information overload (Bomann & Jones, 2003).

When instruction repeats are placed within the framework of a timeline of task execution,i.e. the time the repeats were made in comparison to when the participant carried out the assembly action, the results indicate that different types of repeats occur at different times and for different purposes during the assembly process. The sequence shown (Figure5.4) below indicates that repeats occur at three different times during the task.

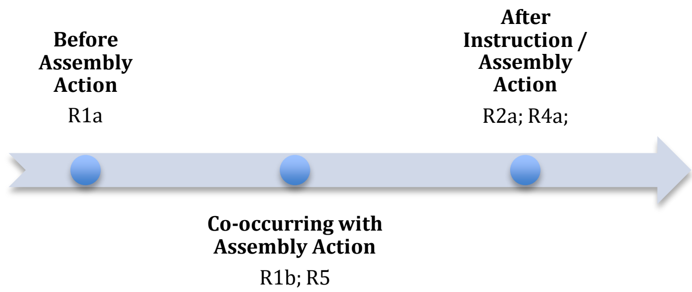


Figure5.4 The sequence of Asking for Instruction Repeats in HHI

Timelines include before the assembly action, during the assembly action and after the initial instruction or the assembly action has taken place. The results suggest that listeners ask for confirmation that may lead to self-correction (R1a) before the assembly action. Listeners also seek clarification and break up tasks during the assembly action (R1b, R5) and assess their actions and comprehension after instruction and assembly action (R2a, R4a).

5.4.3Listener Facial Actions

This chapter outlines listener facial actions as response and the results indicate that listeners display the following categories of facial actions when requesting repeats of verbal instructions during the task: basic emotions, emotional attitudes, moods and eye actions.

Table5.4 Listener Facial Actions in HHI

	Facial Action Family	Illustration			
		Child	Sibling Buckets		Interactive Function
1a	Basic Facial Actions	Neutral	Neutral		May indicate relaxed composure, multitasking, and initiation of turn-taking
1b			Neutral face down		Cognitive processing
1c			Neutral concentration on task		Concentration on task
1h		Smile	Felt smile		May indicate positive interaction
1e			Nervous smile		Making the best of a bad situation, masking embarrassment
1m		Disgust	Slight disgust		Indicates feeling of aversion with lower intensity towards people, things or experience
1n			Disgust		Indicates feeling of aversion with a comparatively higher intensity towards people, things or experience
3b	Emotional Attitudes	Transition of emotions	Tense mouth to neutral face		Indicates transition from one emotional state to another within a unit of task
4c	Moods	Tense mouth and lip action	Compressed lips		May indicate concentration, cognitive processing, anxiety, nervousness, mood shift, restraint
5	Eye Action	Static searching face	Saccade		Used to visually find or locate pieces in without moving the head

5.4.3.1 Basic Emotions

The basic emotions expressed in this study include the following

Neutral Face

The neutral face (Figure 5.5)below is expressionless and all the muscles are relaxed as depicted by AU0 (Ekman, 1977) showing no emotion.

Figure5.5 Neutral Face

Neutral face suggests that participants are relaxed tending towards comfort with the interaction. When the head is bending down with eyes focused on the task space, the neutral expression may indicate concentration on the task while listening to the instruction. This was the dominant expression in this study and always co-occurred with the stroke in the gesture phrase.

Smile

Felt Smile

As outlined in 4.4.6 felt smiles (Figure 5.6)below are elicited by positive physical, verbal and tactile stimulation. The facial muscles responsible for this emotion include cheek raiser (AU6), lip puller (AU12) and lid tightener (AU7). This smile is a positive face (Tipples et al., 2002) that becomes laughter.

Figure 5.6:Felt Smile

The felt smile may indicate that the participant is enjoying and liking the interaction with the instructor.

Nervous Smile

IH5 has a bitter-sweet experience during the task (Ekman &Friesen, 1982), due to the realisation that the assembly processshe used is wrong. The situation worsens as it becomes impossible for IH5 to assemble the next stage due to preceding errors.

Figure 5.7:Nervous Smile

The morphological characteristics indicate that the muscles responsible for the nervous smile (Figure 5.7) include AU6 cheek raiser, AU12 lip corner puller. IH5 may be trying to mask embarrassment with a smile.

Disgust

Disgust

Participants experience a combination of anger and disgust (Figure 5.8). Although, observers may sometimes confuse one for the other, disgust is characterised by lowered brow, upper lip raised, flaring nostrils, and (the sometimes open) mouth curved downwards. The facial muscles responsible (Figure 3.33; Table 3.7) include AU4 brow lowerer, AU9 nose wrinkler, AU5 upper lid lowerer, AU7 lid tightener, AU15 lip corner depressor, and AU16 lower lip depressor.

Figure 5.8:Disgust

The slight disgust has a lower intensity than the one displayed (Figure 5.8) above. It is the slight disgust (Appendix X.5).

5.4.3.2 Mood

Tense Mouth Action: Compressed or Swallowed Lips

As outlined in 4.4.6, the tense-mouth is an aggressive sign of man's nearest primate relative, the bonobo (de Waal & Lanting, 1997)indicated by the face displaying obvious muscular tension with the lips held tightly togetheras compressed or swallowed lips (Figure 5.9) below.

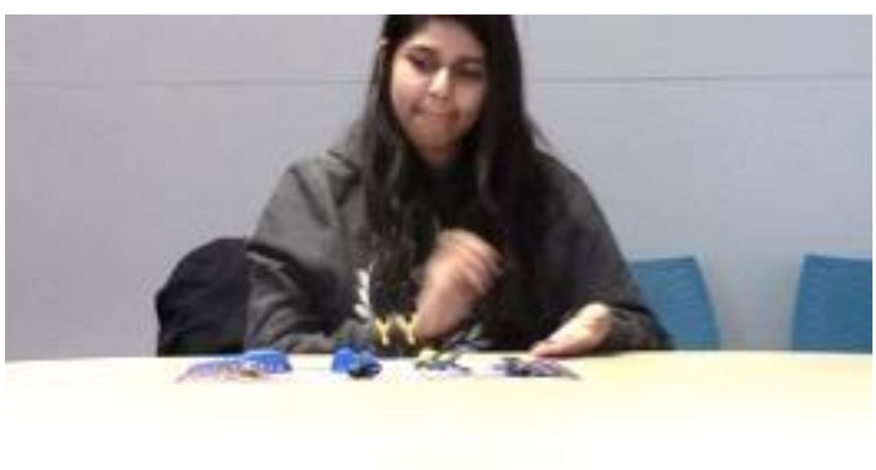


Figure 5.9:Compressed or Swallowed Lips

Compressed or swallowed lips (Figure 5.9) aboveis a position of the mouth in which, the lips are visibly tightened, pressed together, rolled inwards and narrowed to a thin line through contraction of AU8 the lip tightener (Ekman, 1978). The tense-mouth may be a sign of frustration, threat, determination, cognitive processing (Givens, 2016; Ekman, 1997) and self-restraint duringassembly tasks.

5.4.3.3 Emotional Attitudes

Transition of Emotions

Transitions in emotion occur quickly and daily during interaction (Ekman et al., 1988) indicating the dynamic nature or demands of interaction on a person. The illustrations below indicate one participant's transition from one emotional state to another within the same task space (Figure 5.10)below.

A: Compressed lips B: Slightly compressed lips C: Uncompressed lips

Figure 5.10:Transition of Emotions from L-R (A B C)

Looking at Figure 5.11 from right to left,one observes the participant (IH2) experiencing a transition from concentration with swallowed or compressed lips to an expressionless concentration on a neutral face. Swallowed or compressed lips in 'A' may indicate self-restraint (Ekman& Friesen, 2003) as IH2 is twisting the Lego bit to the desired position. As the task is being completed, the compressed lips start becoming uncompressed as shown in 'B'. The task is completed in 'C'and the lips are completely uncompressed, the face is expressionless, but the eyes still suggest concentration.

5.4.3.4 Eye Action

Static Searching Face

As IH2 listens to the repeated instruction during the assembly task her eyes darting across the table searching for assembly pieces. This action, oculomotor movement sub-class ‘saccade’ is used to track an object without head movement (Dragoi, 2015; Dell’Osso, 1994). The orbicularis oculi muscle makes eye movements possible but lateral and medial rectus muscles make saccades possible. The action units responsible include AU57making IH2’s head lean forward (Ekman, 1977);AU61 turns the eyes right and AU62 turns the eyes left (Figure5.11)below.

Figure 5.11:Static Searching Face-Saccade Eye Movement

IH2 struggles to attach the bits unsuccessfully thus, she may be searching with her eyes for other pieces that may fit. This facial action is neither a negative or positive face (Tipples et al., 2002).

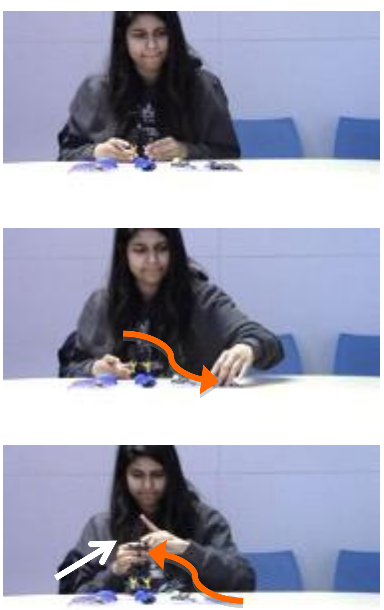
5.4.4 Assembly Gestures

This part of the thesisfocuses on listener action as response and will answer the research question "what gestures do listeners display when requesting repeats of instructions during assembly tasks and why?" Eighty-three assembly gestures (Table 6.3) emerged from this study

5.4.4.1 Picking Hands

The knowing Hand

In the assembly process one main hierarchy of tasks is picking up kits before joining them together using the picking hand gesture. Picking up is done with one hand and enacted like a beat gesture that goes in two moves (out-in). Out is movement away from the sagittal axis(Figure 3.6) while in represents movement towards the sagittal axis. The form and function of the gesture are shown in the vignette (Figure 5.12) below



Preparation: IH2 initiates the gesture with both hands at rest and palms lateral on the Table.

Stroke: IH2's **LH** moves forward, palms down with coupler-shaped digits and uses a combination of the thumb and forefinger to pull up the piece described in the instruction

Retraction: IH2's **LH** moves backwards towards the sagittal axis to drop the picked piece in the **RH**

Figure 5.12:knowing Hands

Another form of the picking hand gesture occurs without a drop with two basic beat-like movements outlined in three steps in the vignette (Figure 5.13) below.

Preparation: IH3 begins this gesture from a pensive position. The **RH** is at rest on the Table while the **LH** is held in a fist covering the mouth

Stroke: IH3's **LH** moves to the left-hand side of the task area, palm down with coupler-shaped digits to pull up the assembly piece

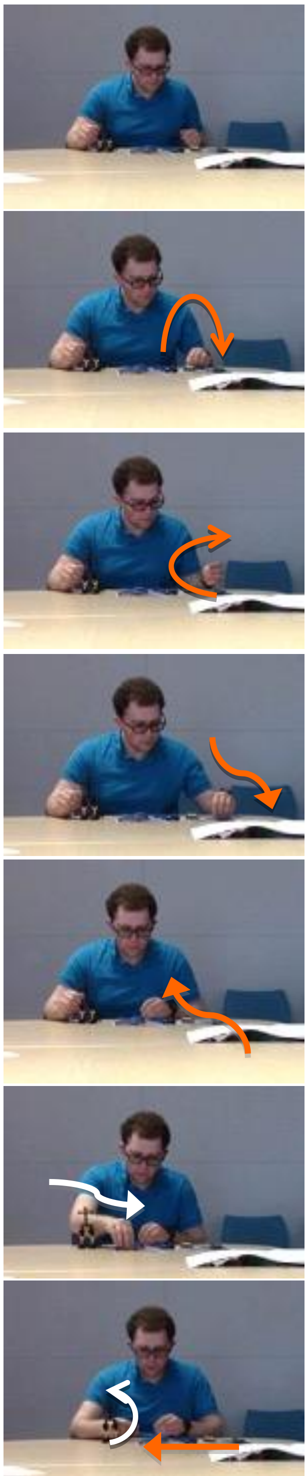
Retraction: IH3's **LH** move back towards the sagittal axis with the piece picked and stops in a post-stroke hold, while the **RH** grasps the upright body,holding it down to receive the assembly bit from the **LH**

Figure 5.13:Basic Picking Hand

Another variation of picking with one hand is executed with a rightward flip of the palm without lifting the wrist from the Table- ulnarly(Appendix X.5a).

Double and Dual Picking Hands

Picking is done with both hands in a convoluted sequence where the left-hand picks up two pieces in quick succession then rests, and the right-hand picks upone piece. The form and function of the gesture is shown in the vignette (Figure 5.14) overleaf.



Preparation: IH3 initiates the gesture with both hands at rest.

Stroke 1: The **LH** moves forward in a curve, palm down, digits held coupler-shaped and pulls up the first piece while the **RH** is at rest.

Retraction 1: IH2 withdraws the **LH** with a 90^0 supination leftwards to presentation the piece held.

Stroke 2: IH3's **LH** moves a bit to the left, palm down with coupler-shaped digits to grasp the second piece while the first is inside the palm.

Retraction2 1/Preparation 2: IH3's **LH** moves backwards towards the sagittal axis while **RH** prepares for the stroke.

Stroke 3: IH3 executes another strike by moving the **RH**. Diagonally towards the kit layout, palm down with coupler-shaped digits to pull up the third piece. The **LH** lifts remains at rest on the Table.

Retraction3: the **RH** moves backwards towards the sagittal axis in retraction to rest on the right-hand side. The **LH** moves rightwards to get close to rest beside the **RH**.

Figure 5.14:Dual and Double Picking Hand

Multiple selection is also enacted as double picking hand that sometimes externalises listener private speech (Appendix X.5b). This could also be done with both hands at once as dual picking hands (Appendix X.5c & 5d). Participants also selected pieces with inbuilt repairs (Appendix X.5e) and also to keep in a holding hand (Appendix X.5f).

5.4.4.2 Joining Hands

Multiple Joining Hands

This gesture is used to assemble the shins of the legs of the Lego kit by attaching the ball joints of the yellow pieces to the black sockets. The form and function of the gesture are shown in the vignette (Figure 5.15)below.

Preparation: IH1 initiates the gesture with both hands in a hold.

Stroke 1: IH1 executes the first stroke with two co-occurring moves. The first is done as the **RH** moves leftwards, palm down with coupler-shaped digits to hold down one foot. The **LH** moves rightwards, palm lateral holding the yellow piece then in a top-down move attaches the yellow piece to the socket.

Preparation 2: IH1 retracts the **RH** upwards palm down with digits held in a fist to pass another yellow piece to the receiving **LH**

The preparation continues as the **LH** moves palm lateral with the thumb, forefinger and middle finger combining tocollect the yellow piece from the **RH.** These prepare for the second stroke.

Stroke 2: IH1's **RH** moves downwards, palm down and uses a combination of the thumb and forefingerto hold down the second foot, while the **LH** moves upwards, palm down with a combination of the thumb, forefinger and middle finger, bringing the yellow piece down in a top-down joining action to fix it into the socket of the feet until a click sound signals success.

Retraction: The **RH** rests on the Table while the **LH** comes to a rest too.

Figure 5.15:Multiple Joining Hands

Joining and Displaying Hand

This blended gesture combines joining with another form of presentation called the displaying hand gesture. Participants like IH4 use this gesture to examine and confirm the action taken. The form and function of these gestures is shown in the vignette (Figure 5.16) below

Preparation: IH1's **RH** is palm up holding the assembling bit while the **LH** is palm down with the receiving bit.

Stroke 1: IH2's **RH**rotatesinto a palm down position with coupler-shaped digits to insert the bit into the one in the **LH**.

Retraction: IH2's **LH** transfers the assembled kit to the **RH**. The **RH** rotates to the right into a palm up position to perform a presentation for examination while the **LH** withdraws towards the sagittal axis and disappears under the Table.

Figure 5.16:Joining and Displaying Hand

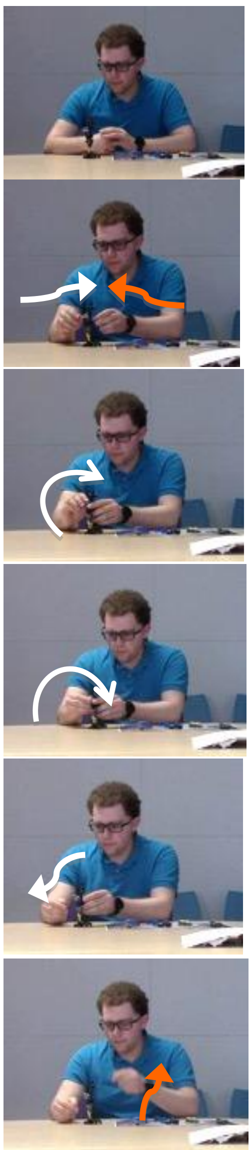
The structure of the gesture phrase (Figure 5.27) during the pause is shown below

<~~~ ~~~~~~~/***************/.-.-.-.-.-.##### >
P S Post stroke hold and R
00:04:08:754/00:04:09:714 / 00:04:10:864

Participants also attach assembly pieces with either the left hand (Appendix X.6a) or the right hand (Appendix X.6b). This study did not find out if these participants were either right-handed or left-handed, the focus is just to identify which hand was the giving and which was receiving. In addition, participants also attached assembly pieces by twisting the bits clockwise and anti-clockwise to get the desired orientation (Appendix X.6d) and they used dual joining hand (Appendix X.6e) which contains semi-multisegment phases to assemble parts that come in pairs such as yellow sockets.

Twisting Hand

The twisting hand gesture is used to manipulate assembly bits and it combines a grip with a twist. Research indicates that people use different types of precision grips when manipulating objects. These include the pinch grip and the three-jaw chuck/ three-finger chuck. The pinch grip is done with the thumb and one finger gripping while the three-finger or jaw chuck has a lateral orientation used for making a straight pull up (Wronski & Daum, 2014; Hedge, 1998)The form and function of the gesture are shown in the vignette (Figure 5.17)below.



Preparation: IH3 initiates the gesture with both hands folded and at rest.

Preparation continues as IH3 stretches both hands with palms lateral out to initiate the stroke.

Stroke: The **LH** holds the upright body steady to receive. The **RH** moves towards the sagittal axis and uses the three-finger chuck grip to hold the small piece, then rotates 90^0 inward pronation to turn the assembly piece as a first step in executing the stroke.

The stroke is completed with the second 90^0 rotation of the **RH** away from the sagittal axis (outwards pronation). The stroke is fully executed and the **RH** is now palm lateral.

Retraction: IH3 withdraws the **RH** first by putting it down on the Table. The retraction is completed when the **LH** releases the body and moves upwards towards IH3's face while the **RH** is at rest on the right side of the task area.

Figure 5.17:Twisting Hand

The three-fingerprecision pinch grip is used in this gesture (Figure 5.12g) to position assembly pieces in 3-D space, for example, rotating 90^0 while the other hand acts as a stabilising pivot.

Failed Joining Hand-Wrong orientation

A failed joining hand occurs when IH5 correctly assembles the kit but with a wrong orientation given that the second part of the instruction that describes the orientation. The form and function of the failed gesture are outlined in the vignette (Figure 5.18)below.

Wrong orientation: At the end of the assembly process, the legs are standing at an angle of about 280^0 instead of 90^0 specified in instruction 6. Indicated by the intersection of orange lines indicated by the blue arrow in the first picture

Wrong correction: IH5 now bends the first leg forward until it is flat. The same thing is done to the second leg.

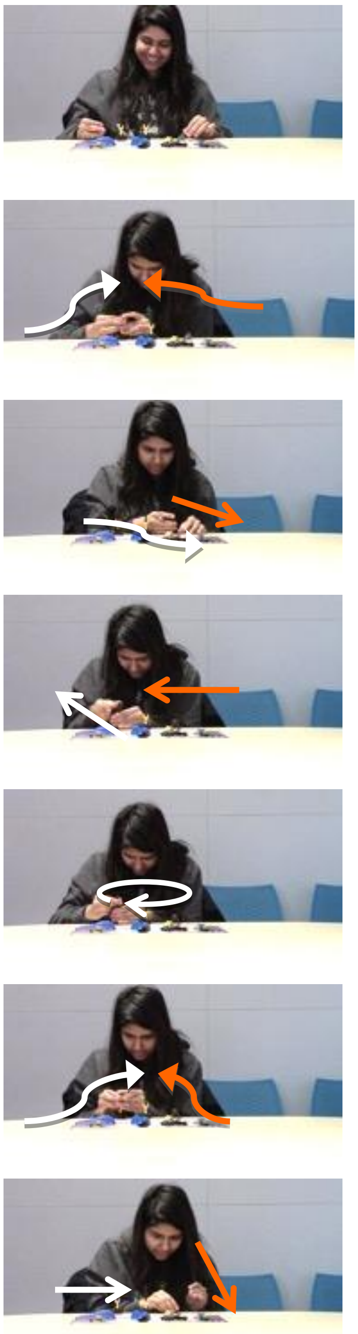
Figure 5.18:Failed Joining Hand - Wrong Orientation

The participant seems not to comprehend the aspect of instruction describing orientation however, this is soon realised as IH5 smiles saying, “this can’t be correct”. This smile is nervous and may be a defence mechanism to cover up embarrassment.

5.4.4.3 Communicative Gestures

Self-Initiated Self-Repairs

As outlined in Chapter4.4.4, repair is an effective interactional tool used to put interaction back on the right track in normal conversation (Frenečik, 2005)and suggests that listeners are learning (Jordan & Henderson, 1995). Repairs occur in trajectories (Greiffenhagen & Watson, 2009) as shown in the vignette (Figure 5.19)below.



Trouble Source (TS)

Preparation1: IH2 initiates the gesture with both hands on the Table. The **RH** is holding a piece.

Stroke 1: IH2's **RH** moves forward palm lateral to grasp the upright between the thumb, forefinger, middle finger and ring finger. This action attaches the piece to the upright in the same way we fasten cuff links with one hand. The **LH** moves rightwards, palm lateral and grasps the kit and holds it firm for the **RH** to fix the assembling bit. IH2 picks up the next assembly piece but while listening to the instruction, IH2 identifies the TS as wrong orientation of assembly piece.

Repair Initiation

Preparation 2: IH2 begins repair by dropping the piece on the left-hand side of the workspace with the **LH** in preparation.

Stroke 2: IH2's **RH** moves to the right, palm down with coupler-shaped digits to pull out the piece from the body while the **LH** holds it firmly.

Actual Repair Execution

Preparation: This is the preparation for the next stroke and IH2 grasps the top of the upright and twist it clockwise

Stroke3: The repair action of fixing the bit in the correct orientation is executed when IH2 uses the thumb to push the piece unto the body while the other digits of the **RH** as well as the **LH** hold the body firmly.

Retraction: IH2 retracts the RH diagonally to the left to pick up the second black piece in preparation for the next gesture.

Figure 5.19:Self-Initiated Self-Repairs

The structure of the gesture is as shown below

<*************************************/

These should attach to the remaining black ball joints around the outside of each leg

Trouble Source (TS)

/***/
The widest end should be closest to the feet.
Repair Initiation
/*********************************>
Actual repair execution
00:04:17:245 – 00:19:460

IH2 identifies the TS and initiates repairs as the instructor describes the orientation. The actual repair takes place within a pause lasting about 2 microseconds.However, unlike Figure 4.37 (Chapter4.4.4), the repair trajectory(Figure 5.19) is structurally similar to that used during conversations.

Backwards Repairs

Backward repairis another trajectory used by participants to execute a backward reintegration of assembly stages. The form and function of the backwards repairs gesture is shown in the vignette(Figures 5.20& 5.21) overleaf.

Trouble Source (TS)

Preparation1: IH2 initiates the gesture with both hands enacting a presentation of the piece picked. The **RH** is holding a piece while the **LH** is palm down

Stroke 1: IH2's **LH** moves down, palm lateral and grasps the kit and holds it firm for the **RH** to fix the assembling bit. The stroke occurs when the **RH** pushes the upright body in a top-down joining action to attach the two pieces. Here the hands are palm down covering the Lego kit.

Preparation 2: IH2 begins work on the second leg and this occurs in two stages. In the first stage, IH2's **RH** moves back while the **LH** is holding down the receiving bit in readiness for the next stage. IH2's **RH** then goes towards the remaining foot on the Table palm down, with digits held coupler-shaped to grasp it. This action prepares the way for attaching the two pieces.

Failed Stroke:IH2 uses a combination of the thumb, middle, and ring fingers of the **RH** to push the joining piece leftwards towards the receiving bit held by the **LH** to insert the assembly bit in a top-down movement unto the receiving bit but, the bits slide apart and the joining hand fails. IH2 identifies the **TS** as incorrect assembly of the lower foot due to wrong positioning of assembly piece.

Repair Initiation

Preparation 3: IH2 begins repair by pulling the body from the Leg with the **RH** and dropping the piece on the right-hand side of the workspace while the **LH** is holding the legs upright.

Stroke 2: IH2's **RH** moves to the left, palm down with coupler-shaped digits to grasp the top layer that is 'y' shaped and twists it about 90^0 degreestowards the sagittal axis to give it the correct orientation.

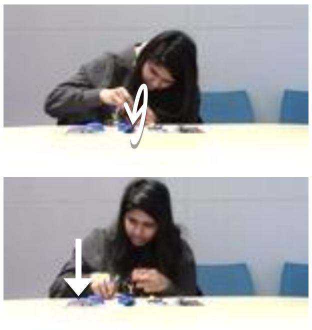
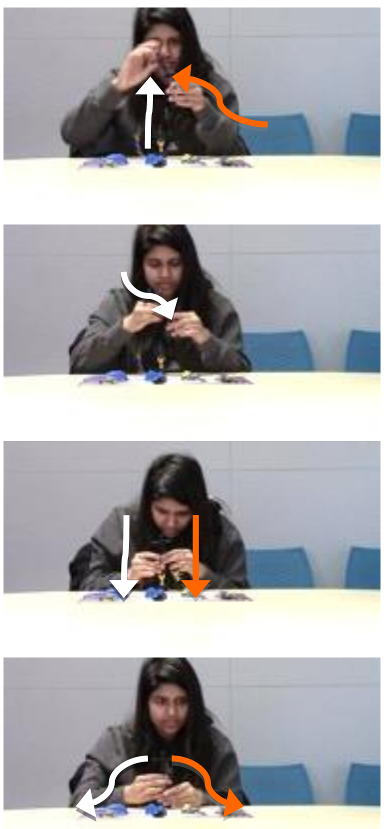


Figure 5.20:Repairs Backwards_ Trouble Spot and Repair Initiation



Actual Repair Execution

Preparation: Preparation is a picking hand as IH2's **RH** goes rightwards, palm down to retrieve the body dropped earlier. The **LH** is in a hold. Preparation continues as IH2's **RH** goes up to drop the lower part of the piece in the **LH** while the **RH** grasps the top. The final stage of preparation occurs when both hands take a firm hold of the body piece.

Stroke3: The repair action for fixing the bit in the correct position is executed when IH2 uses both hands to push the body into the legs

Retraction: IH2's hands come to rest in a post-stroke hold while listening to the next instruction.

Figure 5.21:Backwards Repairs - Actual Repairs Segment

This occurs after the instruction has been given and overlaps into the next instruction. The structure is shown below.

<***/
Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs
Trouble Source
/***/
Repair Initiation
/****************************/*************** **********>

00.04.55.603 – 00.04.60.333 / Now take a black cylinder,
Actual Repair

When participants detect errors when executing a current instruction (Figure 5. 20) they attempt to correct these with basic repair mechanismexecuted within extended repair processes(Figure 5.21). Unlike Figure 5.30, here listeners identify TS retrospectively, then move one step backwards in the assembly process to begin the repairs and proceeding to the current stage just like the speaker's remedy for tracking errors (seeChapter 5.3.2).

Deictic Hands

As outlined in Chapter4.4.4 (Figure 4.85) deictic gestures are decisive pointing movementsthat indicate object and location in assembly tasks as illustrated in the vignette (Figure 5.22) below.

Preparation: IH4 initiates the gesture with the **RH** palm up holding the assembly bit while the **LH**is holding the receiving bit.

Stroke: IH4 rotates the**RH**in full pronation to a palm down position with the forefinger stretched while the other digits curl around the assembly bit. IH4 uses the forefinger to touch and rub the receiving bit up and down (bidirectional arrow)

Retraction: IH4's hand withdraws sideways to the right.

Figure 5.22:Deictic Pointing Hand

The structure of the gesture is shown below

<~~~ ~~~~~~~~~~/****************/#####>

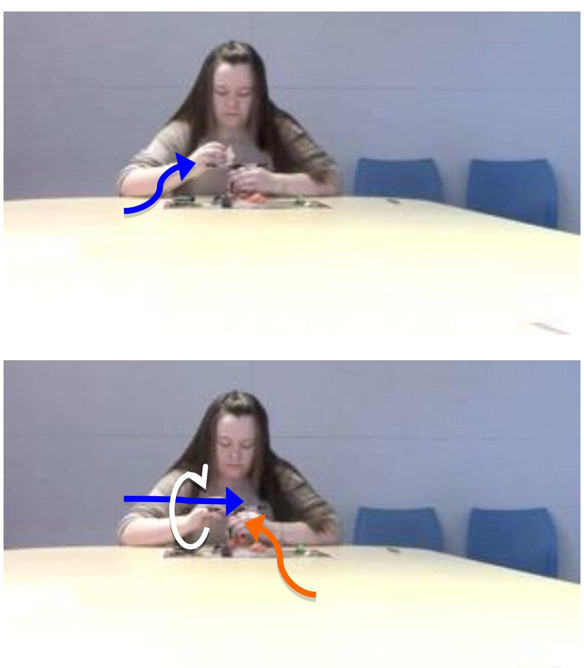
P S R

…middle ball joint on the front of the body using

Deictic gestures provide two simultaneous views of the same process(McNeil,1985) because the structure of the gesture (Figure 5.33) above indicates, the stroke co-occurs with the “front of the body” part of the instruction. This suggests that the listener uses the gesture to identify and locate the assembly point“front” on the Lego model as a spatial deictic reference just as speakers pointing at a location will utter a word like “there”. This combines the instructor’s utterance with the listener’s interpretation into a single event from two perspectives.

Aligning Hand

As outlined in Chapter 4.4.2.2 (Figures 4.34-36) aligning hands gestures go beyond mere deictic reference in task performance enable listeners order and discover the fit of assembly pieces as illustrated in the vignette(Figure 5.23) below.



Preparation: IH4 initiates this gesture with the **RH** rising, palm down with digits clutching and presenting the assembly bit for examination. The **LH** is holding the receiving bit.

Stroke: This involves two major and two minor simultaneous movements. IH4’s **RH** moves forward then the wrist rotatesanti-clockwise to place the assembly piece on the receiving bit. In a simultaneous move IH4 raises the **LH** palm lateral clutching the receiving bit, then bends the top of the receiving bit towards the sagittal axis to meet with the assembling bit held in the **RH**.

Retraction: IH4 rotates the wrist of the **RH**clockwise thereby turning the assembly bit inside out. This move enacts a lateral presentation of the assembly piece for examination. The **LH** simultaneously moves away from the sagittal axis then down with the palm still holding the receiving bit.

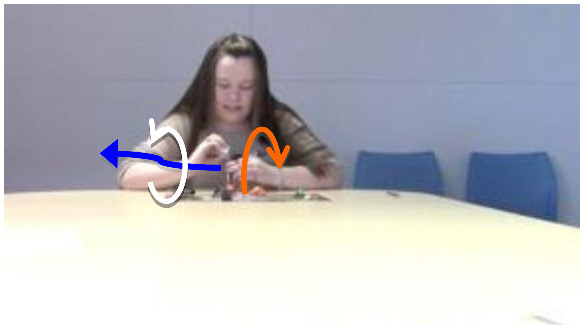


Figure 5.23:Aligning Hand

5.4.4.4 Presentation

Picking for Presentation

As outlined in Chapter4.4.4, participants examine assembly pieces and tasks completed. One form of presentation begins with picking hand gesture and ends with two presentations. The form and function are shown (Figure 5.24) below.

Preparation1: IH2 initiates the gesture from a rest position, palms lateral with both hands holding the Lego kit by the feet.

Preparation2: IH2's **RH**moves diagonally to the kit space, palm down with coupler-shaped digits to pull up a piece

Stroke1: IH's **RH** moves up, backwards, palm lateral and stops briefly at the face level to examine the assembly bit picked up

Stroke 2/Retraction: This is a stroke combined with a retraction. IH2's **RH**goes down and stops midway down while the **LH** rises up to meet it and holds one end of the piece. A two-hand presentation is enacted as IH2 examines the piece

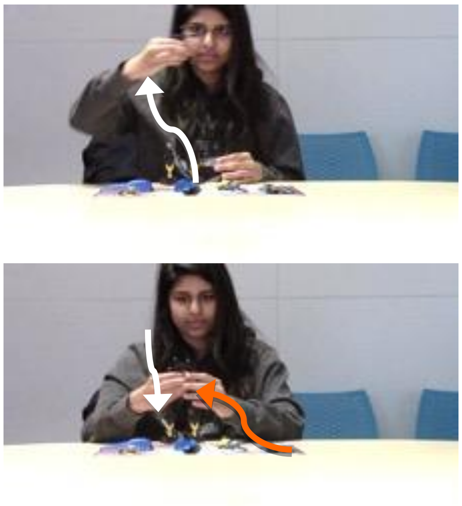


Figure 5.24:Picking for Presentation

Double Presentation

Participantsalso use double presentation gestures to compare multiple assembly objectsduring tasks. The form and functions of the gesture is illustrated in the vignette (Figure 5.25) below.

Preparation:IH1initiates the gesture with the **RH**at rest while the **LH** is holding one assembly piece. IH1's **RH** moves up diagonally left away from the sagittal axis, palm down with coupler-shapeddigits to pull-up one assembly piece from the kit layout.

Stroke 1: The **RH** holds up the piece twirling it around making a presentation while the **LH** with palm lateral is in a hold.

Stroke 2:the **RH** retracts backwards towards the sagittal axis and the turns palm up holding the piece. Concurrently, the **LH** moves diagonally rightwards, palm down holding up another assembly piece. A double presentation is made as IH1 examines the pieces in both hands at once.

Retraction 1: This takes place as IH1's **RH** moves diagonally left, palm down towards the task area to drop one assembly piece. The **LH** has already retracted diagonally to the left in a palm-up position holding one assembly piece

Retraction 2: Retraction is complete as IH1's **RH**moves diagonally rightwards towards the sagittal axis and now goes downwards to rest on the Table.

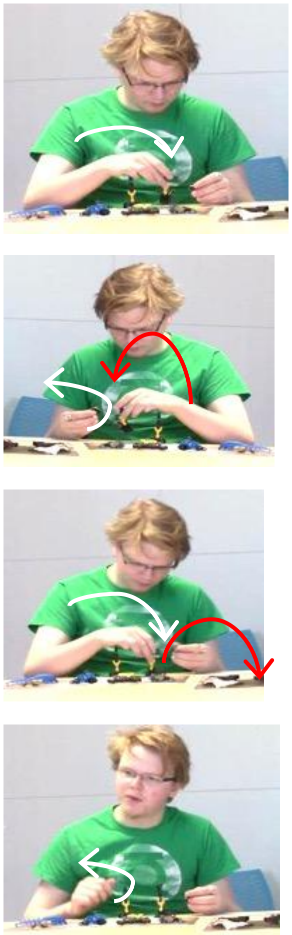


Figure 5.25:Double Presentation

5.5 Discussion

This study assesses listener nonverbal behaviour when taking verbal instructions to assemble Lego kits ina HHI-context. The interaction is between people and verbal because participants make requests for instructions and subsequent repetition verbally and the instructor reads out the instructions. Participants in this study displaypatterns of listenership behaviours that were similar and different in some aspects to those expressed in the HAI study.

The verbal interaction between instructor and instructees is shown in instruction-giving and takingand developed around instruction repetition patterns. The process of repeating instructions that emerged from the interaction is discernible, arbitrary and involves diverse events(Chapter 5.3.1).

The results indicate that listeners requested for instructionssixty-seven timesusing language frames in a variety of ways such as the basic pragmatic request frame, completion drills, question tags and backchannels that to some extent shaped the instructor's verbal action even though the instructor was supposed to be principally reading out prepared instructions.

Furthermore, the instructor displayed features of spontaneous speech-giving such asparaphrasing instructions, repeating only key words, insertingand deleting words into and fromwritten instructions perhaps, due to tracking errors and the natural tendency of people to respond contextually to requests using the co-interlocutor's language frames such as question tags (See Chapter 5.4.2.1).

In addition, listeners manage interaction using repetition sequences for different reasons (Figures 5.4; Table 5.3) just as people engaged in conversations or any verbal interaction listen for specific reasons (Chapter 2) and as shown in HHI, listeners asked instructors to repeat instructions because they needed to self-correct assembly errors and reduce incomprehension. They use composite repetitions to confirm task done, interpretation of instructions and break-up one assembly stage into manageable chunks.

Another aspect of listener action relates to suggesting probable state of mind and task strategies given by spontaneous facial actions (Table 5.4)such as basic emotions comprising anger, fear, surprise, smile, disgust and neutral expression (Ekman et al., 1981). Facial actions analysed indicate that when people take instructions, they project basic emotions, non-basic emotions, attitudes, moods, and context-based actions (Figure 5.3). Thus, when people take instructions they may remain expressionless while concentrating on task or instructions(Figure 5.5; Appendix X.1);smile fullywhen enjoying the interaction(Figure 5.6; Appendix X.2)but when embarrassed or unable to execute instructions, they smile nervously (Figure 5.7).

Furthermore, participants display varying degrees of aversion such as disgust (Figure 5.8; Appendix X.3) and slight disgust (Figure 5.9)during interaction. Moreover, when participants felt frustrated and needed to show determination while cognitively processing instructions, they displayed swallowed or compressed lips (Figure 5.10) but these moods could shift as participants displayed transitions in moods (Figure 5.11).

In addition, participants use saccadic eye movements (Figure 5.12) as a task strategy to search for assembly pieces without moving their heads. This facial action is similar to eye movements used for reading texts, but it functions like our hands used to search for something.

The other aspect of listener action relates to task execution and communication strategies(Table 5.4a & 5.4b)-These are assembly gestures (Chapter 5.4.4; Tables5.5) that are classified into fivegesture familiesand their functions are as follows.

Participants pick assembly pieces up in several ways. When they want to select assembly pieces they use variations of knowing picking hand gesture (Figure 5.13, 5.14) however, when the assembly piece is nearby they use ulnar picking hand (Figure 5.15). To sequentially select multiple assembly pieceswith one hand, they use double picking hands (Figure 5.16) but, participants use both hands to select two pieces at once as dual picking hands (Figure 5.17, 5.18) and pick more than two pieces using dual and double picking hands (Figure 5.19). Furthermore, participants concurrently pick pieces and carry out repairs (Figure 5.20), pick and drop assembly pieces before continuing the task (Figure 5.21).

After selecting assembly pieces, participants visually, physically and mentally assess their fit before attaching them to one another usingan aligning hands gesture (Figures 5.34, 5.35).Thereafter,participants use operative gestures to attach assembly pieces together using joining hands gesture (Figure 5.23, 5.24). When they need to attach more than one pair of assembly pieces, they use multiple joining hands (Figure 5.22) and dual joining hands (Figure 5.26). Furthermore, they use circular twisting joining hands (Figure 5.25) to screw assembly pieces together using pinch grip twist (Figure 5.28) to pull up and manipulate assembly pieces 90 degrees towards the sagittal axis (Figure 3.6).

Going further, participants attach and display the finished product for assessment (Figure 5.27). However, participants attach assembly pieces in the appropriate location but a wrong orientation within the 3-D space makes it a failed joining hand gesture (Figure 5.29). After joining assembly pieces, participants sometimes examine the work done using the double presentation hand (Figure 5.36).

Participants use communicative gestures (Table 5.5) to identify errors and rectify them such as dialogic repair (Figure 5.30) and retrospective backward repairs (Figure 5.31, 5.32). In addition, participants use deictic hands (Figure 5.33) to establish joint focus with the instructor and make task handling easier.

5.7 Conclusion

This study assesses the patterns of nonverbal listenership behaviour that participants display when they take instructions from a human during assembly tasks. Participants in this study displayed nonverbal listenership behaviours that are in some ways similar to those displayed in HAI.

The results indicate that the listener’s interpersonal relationship with the instructor is one factor influencing the types of gestures, facial actions and other listenership behaviour. It was found that facial actions may express listener attitudes towards the instructor and interaction. On the other hand, the frequency of repeats clearly increased when listeners could not comprehend instructions or execute instructions effectively.

Meanwhile the typology of repeats during interaction is dynamic; for example, it is interspersed with completion drill-like exchanges between instructor and instructees.These findings support the established view in linguistics that clarification practices, repetition and completion of utterances on behalf of the original speaker may resolve communication problems (Cogo & Dewey, 2012).

Table5.5 Spontaneous Assembly Gestures in HHI

Gesture family	Sub-Group	Description	Function
Picking Hand	**Knowing hand**	**Executed with one hand extending forward or away from the listener's body, palm down to pull up assembly pieces then return like a beat gesture**	**Enables listenerto select an assembly piece**
	Ulnarly picking hand	**Executed with wrist movement towards the pinky side while the other hand is in a hold**	**Enables user to select objects that are very near or within a very narrow periphery**
	Double picking hand	**Picking is executed with one hand using two movements with hesitation when selecting assembly pieces**	**Enables listenerto select more than one assembly bit sequentially and externalise mental decision-making indicated by the hesitation between the two picking moves**
	Double and dual picking hands	**Executed with both hands in a convoluted sequence**	**Enables listener to select two pieces with one hand and one with the other**
	Dual Picking Hands	**Done with both hands selecting pieces at once. Sometimes occurs with a drop then retraction or retraction without dropping what has been picked**	**Enables listener to select two pieces with both hands in one go**
	Picking handwith repairs	**Executed with one hand picking and dropping pieces (repairs) until the correct one is finally picked up**	**Enables listenerto select pieces and make self-corrections in the process.**
	Keeping and holding hand	**Executed with one hand picking and dropping the piece in another hand then picking another**	**Enables listenerto select and retain assembly pieces in the hands**
Joining Hands	**Basic joining hand**	**Executed with both hands placing, pushing, sliding one piece into another**	**Enables listenerto fix assembly pieces together within a 3D location and appropriate orientation**
	Multiple joining hand	**Same as basic joining hand except that the process is repeated more than once**	**Enables listener to attach parts that come in pairs such as the legs.**
	Twisting joining hand	**Executed with a twisted or arched 90^0 clockwise movement**	**Enables listener to attach and screw assembly pieces from the side with accuracy**
	Dual joining hand	**Executed as two complete assembly cycles. Each cycle has preparation, stroke retraction. They are linked by semi-multisegment phases used to exchange assembly pieces between hands**	**Enables listener to simultaneously assemble the lower parts of the Lego kit**
	Joining &displaying hand	**Executed as a combination of joining and presentation hands**	**Enables listener to attach pieces together and examine the success or failure of completed task**

Table5. 5 SpontaneousAssembly Gestures in HHI

Gesture family	Sub-Group	Description	Function
Joining hands	**Failed joining hand**	**These are errors of placing assembly pieces is wrong orientation within a 3-D space. Such errors occur in multiples as the participant corrects the first, the correction itself is shown to be another error. Participant's spontaneous vocalised swearing further confirms the failure**	**Enables listener to identify errors and may be correct them. It also enables the observer to assess the participant's progress while vocalisation project's the participant's thought processes**
	Pinch grip twist	**Executed with the thumb and one finger taking hold or clasping the assembly pieces then rotating them 90^0 towards the sagittal axis**	**Enables listener to pull up, position and manipulate assembly pieces within a 3-D space such as 90^0**
Communicative Gestures	**Self-initiated self-repairs**	**Executed in three major steps that begin with identification of trouble source, repair initiation and actual repair execution.**	**Enables listener to spot errors within a current assembly stage and rectify them.**
	Backwards repairs	**Executed with an extended repair process. Trouble spot is identified retrospectively when proceeding s not possible because of a previous error. The user disassembles the kit, then carries out actual repairs across all the stages to the current one.**	**Enables listener to spot errors retrospectively and carry out repairs in multiple stages.**
	Deictic hand hands	**Enacted by listener pointing to a specific location or piece as the instructor mentions it.**	**Enables listener to establish a means of self-direction and a joint focus with instructors which makes task handling easier.**
Presentation	**Double presentation**	**Enacted with both hands first with the right hand then with both hands at once.**	**Enables listenerto compare two or more assembly pieces at once.**
	Picking for presentation	**Begins with picking hand gesture and ends with two presentation gestures, the first presentation with one hand while the second is done with two hands**	**Enables listener to examine assembly piece selected then compare it with another**
Aligning hands	**Basic aligning hands**	**The assembly bits are brought close but they do not touch**	**Enables listener to visually and mentally assess the fit of one or more assembly parts into others from various positions**

While these results indicate thatlistener-facial actions are ever-present during interaction, the suggestion that they may have some level of relationship with listener execution of verbal assembly instructions will be further tested later (Chapter 6.2; 6.7).Clark et al (2016a) suggest that using human rather than synthesised voices for instruction-giving agents may further our understanding of the communication issues surrounding HAI; findings from this study do not only support this view but add that a comparison of HAI and HHI contexts may provide the data-driven grounds for this decision.

Clearly, listeners in HHI have been able to use verbal (requests for repetition) and nonverbal listenership behaviours (Facial actions and gestures) as pragmatic markers indicating their comprehension and incomprehension of verbal instructions. This suggest that as with interaction such as conversations, the use of dual communication modes (verbal and nonverbal) may enhance listener comprehension during assembly tasks. However, these results have to be compared with the HAI study (Chapter 4) in ChapterSix with a view to assessing the pragmatic and interactional implications for communication in HACs.

Chapter 6 Nonverbal Listenership Behaviour in HAI and HHI

6.1 Introduction

This Chapter contains two major chapters. The first chapter answers the question regarding ***how the use of spontaneous facial actions and gestures in HAI compare to those in HHI*** using the patterns emerging from the two interaction contexts outlined in Chapters 4 and 5. The second chapteroutlines the implications and extent to which the findings have contributed to understanding nonverbal listenership in unidirectional instruction-giving contexts.

To answer the questions presented above, two studies were conducted: in Study 1 (HAI), forty-eight participants received instructions from a combination of two voices from a cline of three voices and in Study 2, six participants received instructions from one human instructor. HAI recorded 96 interactions while HHI had 12 which in turn led to the differences in the number of instructions repeated by instructors during tasks. This difference is reflected in the quantity of some nonverbal behaviours displayed. It must be pointed out that the analysis done in this Chapter is not meant to identify which interaction is best or more successful rather, it aims to outline the ranges and/or trends of interaction approaches emerging from the tasks and a lens to understanding the communication that occurs in both contexts.

The Chapter is structured as follows: 6.2 compares hybrid with verbal interaction as shown in instruction-giving and taking, repetition patterns used by participants in the two contexts; 6.3 outlines listener attitudes projected and inferred from facial actions; 6.4 focuses on listener gestures used in task execution; 6. 5 analyses gestures that listeners use forintrapersonal and interpersonal communication; 6.6 outlines key nonverbal behaviours emerging from the study; 6.7 discusses the implications followed by a Chapter summary and conclusion in 6.8.

6.2 Interaction and Repetition Patterns in HAI and HHI

As outlined in chapter 4, language use is a joint cooperative transaction between interlocutors (Grice, 1999) and this principle is shown in this thesis in the agreed interaction pattern guiding instruction-giving and taking in assembly tasks. In both HAI and HHI, there is the quasi-tacit contract that instructors will give instructions while listeners will execute them however, listeners can ask instructors to repeat instructions as many times as required while instructors cannot go back to instructions that has been taken. Furthermore, instructors and instructees also have what Grice (1999, p.81) calls some common immediate aim which in this case is to assemble Lego models. These agreements provide a common task focus and interaction framework shown in turn-taking.

Although interaction in HAI is hybrid (listener's nonverbal requests and instructor's verbal response) and wholly verbal in HHI (4.5.1 & 5.5), the processes of asking for repeats in HAI and HHI are commonly discernible and observable because co-interlocutors can perceive, hear or see when listeners make verbal requests, press the repeat button or nod their heads during interaction. As predicted in Chapter 4.4.1, the results indicate that listeners in HAI asked agents to repeat instructions 667 times while listeners in HHI asked the human instructor for repetitions 67 times (Tables 4.3 & 5.2) due to having more participants and interactions in HAI than HHI.

The interaction process viewed through request for repetition in HHI, involves diverse events and characterised by features of spontaneous conversations on the part of instructors and instructees. On the other hand, interaction in HAI involves a continuous chain of events without variation (Chapters 4.4.2.1 & 5.3.2.1). Specifically, listeners in HHI initiate interaction in four ways which include the use of a common framework employed in making requests "Please repeat that",completion drills, sentences ending with WH-question tag and nonverbal requests such as smiles, head nods and backchannels e.g. 'umhmm' (see 5.4.2.1). In contrast, those in HAI use one continuous nonverbal process (Figure 4.4) that begins with initiation (the first time a participant asks for an instruction), then execution embedded with repetitions (every time participants press the repeat button a repeat point is reached until the instruction is executed) and finally resolution (the point when participants request the next instruction).

Instructor response is the counterpart of listener action during interaction and results indicate that this differs between contexts. Instructors in HAI respond to every request by uttering whole instructions. Although, the instructions were written for the instructor to read aloud in HHI, instructor utterances are characterised with repairs, hesitations, insertions and deletions just as in spontaneous speech. For example, instructors in HHI repeated whole instructions in 12 instances thereafter, the instructor in HHI used verbal responses that are determined by the listener's language frames and characterised by disfluencies as outlined in 5.4.2.1. This implies that while interaction in HAI goes on as planned, interaction in HHI is more spontaneous like impromptu speech.

These differences in interaction strategies used by instructors and listeners are partly attributable to the agent's affordances in HAI context (See Chapter 4.3.1) however, they also indicate listener's need for adequate information (Grice's maxim of quantity, see 4.2.2) for specific purposes even in HHI where human agency enables listeners to make verbal and nonverbal requests for instructions. On the other hand, Instructor response reflects attempts to provide information that is adequate and appropriate to the listener's immediate task needs (Grice's quantity and relation principles-see Chapter 4) as seen in the use of completion drill requests (Chapter 5.4.2.1).

Although, the interactions occur in unidirectional instruction-giving contexts, the Gricean notion that speaker-information must be clear, adequately informative to aid task execution (Grice, 2006) remains relevant because when these are flouted, meaning-making is affected, and listeners request for repetition of instructions. In addition, the purpose of listening is linked to the type of information speakers communicate and listeners expect to use during interaction (See Chapter 2.2) thus, when these expectations are not met listeners request repeat instructions for various purposes. Eighteen repeat typologies were identified in HAI and five in HHI. Of these four are common to both interaction contexts, when listeners in HAI and HHI needed clarification to carry out self-correction, they used R1 type of repeats. They used R2 for clarification, reducing confusion and enhancing cognitive processing of instructions. R4a was used to focus on instruction execution and assessing task actions, and R5 was used to break up assembly stages into manageable chunks during interaction (See Tables 4.4 and 5.3).

As pointed out earlier, listeners in HAI used a wider variety of requests than those in HHI context. These include composite repeats (R4b-f) used for clarification, self-correction, confirmation of listener responses, demarcation of assembly processes and assessment of listener actions. R6 were used to assess listener interpretation of instructions, and R7 to reduce confusion, clear doubts and break-up assembly task into manageable chunks. They use R8 types to refocus attention to critical aspects of tasks when instructions are cumbersome and require multiple simultaneous assembly actions. Participants in HAI mistakenly pressed the Repeat instead of Next button when requesting new instruction as an R9 type of repeat. Listeners interacted with agents using three voices along a continuum (Chapter 3) and the quantitative results suggest that CP with 254 repeats and CL with 226 repeats were not as comprehensible as HR (which closer to human voice) with 187 repeats (Chapter 4; Table 4.3). The non-human quality of CP and CL voices and prosody may have constituted a hindrance as the agent's voice did not match listener experiences with everyday agents in self-checkout systems, banks, smart PAs (Siri) or even language learning devices (DuoLingo). This ties in with earlier research suggesting that speaker voice or verbal identity may be linked to listener linguistic processing ability in areas such as the ability to understand speech (Besser, et al., 2013) and by extension, assembly instructions.

Furthermore, participants' strategic use of self-generated nonverbal cues such as repetition typologies as illustrated in the sequences (Figures 4.5 & 5.4) suggest that listeners were not only creating private discourse but, also self-developing meanings from interaction contexts during tasks. This is supported by research indicating that people often use cues and context to create meanings during interaction (Petrou et al., 2016)

6.3 Listener Attitudes in HAI and HHI

As outlined in chapter 4, research suggests that attitudes are deducible from body language (Chapter 4) thus this chapter compares listener facial actions displayed and attitudes during interaction. Listeners displayed facial actions (247 in HAI & 70 in HHI) that fall within six facial action families divided into 13 children groups which is further sub-divided into 31 sibling units outlined (Table 6.1) below.

Listeners in both contexts displayed neutral facial actions (156 in HAI & 37 in HHI-Table 6.1) suggesting that they were relaxed, able to initiate turn-taking, concentrate on tasks and interpret instructions. However, only listeners in the HAI context displayed six neutral faces that had in-built cues (Figures 4.7; Appendix VIV 2-4a-b) which reinforces research suggesting that the face is never a blank canvas (Hess et al., 2009). This study extends Ekman and Friesen's (1978a-b, 1969) and Ekman et al's. (2016) single classification of neutral face to include neutral hard, neutral concentration, workman effort, and static searching faces.

Furthermore, this study reduces Ekman and Friesen's (1978a-b) five-step classification to three to avoid the ambiguity of differentiating between unclear categories. For example, listeners in HAI and HHI context displayed varying degrees of aversion during interaction (Figures 4.10; 5.8; AppendixVIV.10; V.3; V.5). Specifically, listeners in HAI had 4 and those in HHI recorded 2 incidences of slight disgust. Similarly, listeners in HAI recorded 7 occasions of expressing intense disgust while HHI had 4 during interactions.

The agent may have placed more demands on the listeners' cognitive skills considering the greater incidence of confusion recorded in HAI compared to HHI perhaps due the non-human quality of the voice. The listeners' blended emotional and gestural responses to this point to their capacity to emotionally adjust to changes in interaction with other interlocutors in different ways thus, they use a blend of emotions when interacting with agents and a transition of emotions in HHI.

Table6.1: Comparison of Spontaneous Listener Facial Actions Displayed in HAI and HHI

S.NO	FACIAL ACTIONS			FREQUENCY		
	FAMILY	CHILD	SIBLINGS		HAI	HHI
1	BASIC FACIAL ACTIONS	Neutral	Neutral		74	24
2			Neutral face down		26	2
3			Neutral concentration on Task		23	4
4			Neutral concentration on Instruction		27	7
5			Neutral Hard		1	0
6			Neutral Intense		5	0
7		Surprise	Surprise		1	1
8			Slight Surprise		1	0
9		Smile	Felt Smile		4	8
10			Nervous Smile		1	1
11			Slight Smile		9	2
12			Tight-Lipped Smile		4	1
13			GWB Grin		3	1
14		Fear	Controlled Fear		2	0
15		Disgust	Disgust		7	4
16			Slight Disgust		4	2
17	NON-BASIC FACIAL ACTIONS	Blends	Smug		1	0
18			Angry-Disgust		1	0
19	EMOTIONAL ATTITUDES	Frown	frown		10	3
20		Transition of emotions	Tense mouth to neutral face		0	1
21	MOODS	Eureka!	Aha! moment		1	0
22		Workman face	work face		8	1
23		Tense mouth & Lip action	compressed Lips		15	5
24			Biting lips		2	0
25			Pouty face		1	0
26	EYE ACTION	Static Searching face	Saccade		5	1
27	MICROEXPRESSIONS	Basic Hot spots	Anger		4	0
28			Smile		3	2
29			Tight-lipped smile		1	0
30			Disgust		1	0
31			Sadness		2	0
		TOTAL			247	70

Listeners in both HAI and HHI contexts frowned (10 & 3 times respectively-Table 6.1) suggesting concentration or confusion.Furthermore, those in HHI displayed a transition of emotions from a tensed mouth action (compressed lips) to a relaxed neutral face (Figure 5.10) suggesting changes in mental states. In addition, only listeners in HAI context displayed an observable blend of emotions during interactions such as a mixture of joy and contempt (smug) and a mixture of anger and disgust (Chapter 4.4.3). These differences suggest that interaction

There are similarities and differences in moods displayed by listeners in HAI (27 siblings) and HHI (6 siblings) contexts. For example, listeners in both contexts compressed their lips (Figures 4.17; Appendix VIV.12& 5.9) suggesting that they may be concentrating on tasks, cognitively processing instructions, were anxious, or nervous during interaction. However, some moods were specific to listeners within HAI; for example, they bit their lips (Figure 4.19; Appendix VIV.13), displayed pouty faces (Figure 4.18), showed exertion of force (Figure 4.16) while a sudden resolution of a challenging task registered instantly as an Aha! Moment (Figure 4.15). As outlined in chapter 2, these moods prepare interlocutors for future experiences which is very useful in handling tasks.

As outlined in Chapters4.4.3.4 and 5.4.3.4, listeners in both contexts used saccadic eye movements (Figures 4.14 & 5.11) to search for assembly pieces without moving their heads. However, saccadic eye movements in HHI involved one left-right movement but, in HAI left-right-left movements were used. These differences observed agree with Rayner's (1998) findings that the difficulty of a search task influences eye movements during interaction.

Research suggest that listener attitudes might be reliably detected and measured through positive, neutral and negative facial expressions (Figure 6.1) overleaf. A descriptive analysis indicate that positive facial actions represent 18%, neutral face represents 57% while negative facial actions represent 25% of the distribution.

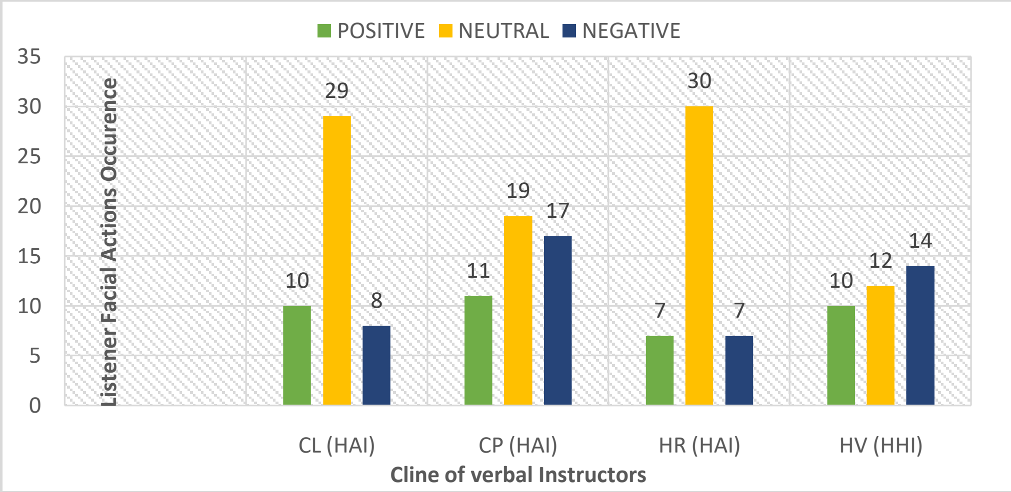


Figure6.1 Listener Facial Actions Displayed During Interaction

Key:CL= Cepstral Lawrence; CP= CereProc; HR= Human Recording; HV= Human Voice, HAI = Human-Agent Interaction; HHI =Human-Human Interaction

Furthermore, listeners' attitudes towards each instructor along the continuum differ and may impact on the interaction in different ways. For example, positive attitudes have been found to encourage user association with interlocutors, contexts, or ideas just as negative attitudes may discourage users from continued interaction in a meaningful event (Wakefield & Wakefield, 2016; Liska, 1974). Listener perceptions are developed performatively and linked to the personal identity of the instructor (verbal for agents in HAI and human presence in HHI) projected as social identity during interaction (See Chapter 2)

Brown and Levinson (1987) opine that people in all interaction contexts have face wants. The notion of face wants describes two kinds of desires- the right to territories and desire to be unimpeded in one's action (negative face) and the desire to get approval from co-interlocutors (positive face). The rational actions people take to preserve both kinds of face, for themselves and other people they interact with is referred to as politeness thus, politeness is used promote harmonious interaction. As outlined in chapter 3.2.5, the researcher was present during interactions, which may have constituted a face threatening act for listeners in HAI and both interlocutors in HHI because it impeded on the interaction context. This created the observer's paradox where listener nonverbal behaviours during interaction may have been altered by the researcher's presence, thus compromising the accuracy of the results. However, the degree of observer's paradox in this study is considerably low becausethe researcher did not interfere in the interaction except when called to re-start the computer when there is need to pick up an assembly piece that has fallen far from the listener otherwise, the researcher sat with his back to the interaction space. In addition, observer paradox is mitigated by the systematic and accurate analysis of data collected as outlined in chapter 3.For example, listeners tried to conceal their real feelings in some instances but, during analysis, these microexpressions were identified and properly accounted for (see Chapter 4).

The display of similar facial actions and attitudes by listeners in HAI and HHI confirm earlier studies suggesting that, although people interacting with agents are aware that they are not interacting with real people in real-life contexts, they still respond socially to them as if they are real people and spaces (Reeves & Nass, 2014, 1996). The differences observed in the patterns of facial actions displayed in HAI and HHI seemingly agrees with Luger and Sellen's (2016) argument that people do not respond to agents as they do people.

However, it is the view of this thesis is that a deeper look at the context suggests that the results tend to agree more with M-E rather than with Luger and Sellen's (2016) stand. Firstly, the task is a collaboratively managed social activity where interlocutors develop common communicative and relationship goals. In addition, the hybrid interaction context enables listeners and agents to build dynamic relationships just as in HHI (Chapter 1) where listeners can project their attitudes and perception of instructors' identities. These are purely social circumstances largely shaped by the cultures of the interlocutors that foregeound the interaction and prepare the ground for performatively constructed responses as seen in the listener's projected perceptions regarding instructors and interaction. Furthermore, listener responses may be due to their spontaneous regulation of emotions (Gross & Levenson, 1997) to meet interaction needs which again is linked to culture imbibed during socialisation.

Thus, it is reasonable to conclude that human responsive behaviour may not only depend on whether the co-interlocutor is a smart agent or human, rather whenever a situation or context promotes participants' needs, positive emotions result but, when they are hindered, negative emotions are displayed however, these are influenced by the culture of interlocutors as well as the culture of interaction within a specific context.

6.4 Task Execution Strategies in HAI and HHI

Gestures represent listener interpretation of verbal instructions and projection of how the instructor's verbal text is translated into the listener's visual text (Chapters 2, 4 and 5) as expressed within task hierarchies during interaction. These hierarchies provide a means of describing gestures as operative and assistive actions depicting actual task operation and interaction processes (Lane et al., 2006). Listeners used 225 gestures as task execution strategies in HAI and 76 in HHI (Table 6.2) and the trends, similarities and differences observed in both contexts are outlined below.

Assistive gestures (Table 6.3) common to HAI (34 occurrences) and HHI (13 occurrences) contexts include knowing hands (20 occurrences in HAI & 6 in HHI) used to select assembly pieces without hesitation, (Figures 4.26 & 5.12); 7 and 3 occurrences respectively of double picking hand -one hand selecting more than one piece, (Appendix VIV.21n & X.6b); dual picking hands gesture (two hands selecting two assembly pieces at once, -2 and 3 occurrences respectively; Appendix VIV.21g & X.6c); and picking keeping and holding hand gesture (selection and retention of assembly pieces in one hand, -13 and 8 occurrences respectively Appendix VIV.21c & X.6f).

However, participants in HAI displayed a variation of picking and holding hand gesture executed with a palm up or supine hold (Appendix VIV.21d). Similarly, two picking gestures were used by only listeners in HHI: the double and dual picking hand (Figure 5.14) and picking hand with repairs (Appendix X.6e).

Although the same kit layouts (Appendix, VIII) were used in HAI and HHI, there are differences in operational hands used to execute gestures or handedness (Table 3.3). Handedness in this thesis does not refer to whether a participant is left or right handed because that is outside this study's scope rather, it simply describes the hand used in executing instructions.

Table6.2Frequency of Assembly gestures displayed in HAI and HHI

	GESTURE		FREQUENCY	
S.NO	**Gesture Family**	**Sub-Class**	**HAI**	**HHI**
1	ALIGNING HANDS	Basic	24	9
2		Micro-Aligning	1	0
3		Extended Aligning Hand	6	1
4		Aligning hand with Repositioning	3	0
5	PICKING HAND	Knowing Hand	20	0
6		Searching	17	0
7		Basic Picking Hand	13	8
8		Picking with Hold	5	1
9		Concatenated Picking Hand	11	4
10		Picking & Aligning Hand	1	0
11		Dual Picking Hand	2	3
12		Alternate Picking Hand	12	0
13		Segmented Repositioning Hand	2	0
14		Picking Ulnarly	2	0
15		Two-finger Picking	4	0
16		Double Picking hand	7	3
17		Two-hand setting aside- Muller	1	0
18		Dual & Double picking Hands	0	1
19	PRESENTATION HAND	Basic presentation	1	0
20		Checking hand	6	2
21		Repetitive Opening Hands	17	0
22		One-Hand Presentation	1	0
23		Fingertips Presentation	2	0
24		Circular Presentation	2	0
25		Dual Presentation	2	1
26		Presentation for alignment	4	0
27	COMMUNICATIVE GESTURES	Positive Turn-taking violation	1	0
28		Self-Initiated Self-Repairs (S-I S-R)	2	6
29		Extended S-I S-R	14	1
30		Deictic hands	2	1
31		Deictic Co-interpretational	2	2
32		Deictic Enumerator	1	0
33		Deictic Representation	2	0
34		Deictic Hedge	2	0
35	OPERATIVE GESTURES	Basic Joining Hand	24	11
36		Multiple Joining Hand	6	4
37		Failed Joining Hand	9	14
38	OTHERS/SACRUM	Crooked Aim	1	0
39		Self-restraint	1	0
40		False Starts	9	0
41		Top-Down Keeping Hand	1	0
42		Twisting hand	1	4
43		Vibrato Dropping Hand	1	0
44		Concatenated Mixed Hand	2	0
	TOTAL		247	76

For example, only participants in HAI used two hands to alternatively select pieces 12 times (Appendix VIV.21h-21k); select, place and segmentally reposition assembly pieces (Appendix VIV.22) and pick and set aside assembly pieces in one motion without lifting them off the Table (Figure 4.28).

Furthermore, some selection strategies were exclusively used by participants in the HAI context. The first relates to the shape of the digits when selecting assembly pieces – the crooked aim (Figure 4.31) is a co-thought gesture (Kita, 2013) shows the listener taking aim thus externalising the listener's mental preparation when selecting assembly pieces.

The second is searching hands gesture (Figure 4.53), an unsuccessful attempt to locate or select assembly pieces. The third is the false start (Figure 4.39) enacted like a jerky hand normally exhibited when two people reach for an object at the same time. However, in this context, the false hand is enacted by one participant reaching out to select an assembly piece. Theresults indicate that participants used the false start nine times thus, suggesting that on these occasions, listeners hesitated or were not sure about the assembly piece to select.

The fourth is the vibrato dropping hand gesture (Figure 4.34), a nonverbal hedge used nine times to mitigate the accidental dropping of assembly pieces during the task. These strategies suggest that one's ability to execute instructions is linked not only to comprehension skills but also to the level of spontaneous preparation, tracking and recovery skills brought to bear during interaction.

When assembly instructions are not specific, a communication and task vacuum is created as listeners are left to their own devices during interaction. Participants in HAI context resort to self-propelled skills within gestures (top-down keeping hand, Figure 4.32) to select and keep pieces where they deem them fit. This reinforces Shea et al's (2013) finding that when people have very little control over an interaction context, process, or phenomena they tend to develop special skills that enable them to cope pick up self-control cues from co-interlocutors and context.The results suggest that participants in HAI may have, through assembly strategies, developed such self-control and taken cues from practice to creatively build specific interaction skills.

To ensure that assembly pieces have a good fit before joining them together, participants in HAI used 34 aligning hands while those in HHI used 10. However, participants in both contexts commonly used basic aligning hand gestures (HAI-24; HHI-9, Figures 4.34 - 4.37 & 5.34) without allowing the pieces to make contact. However, unlike participants in HHI, those in HAI used micro-aligning (Figure 4.38) and multiple aligning (Figure 4.39) gestures that involve assembly pieces making contact to assess how they fit.

Listeners used operative gestures to carry out the attachment of assembly pieces of which 39 were observed in HAI and 29 in HHI. However, there are similarities and differences in operative gestures used by participants in HAI and HHI. The three operative gestures common to both contexts include basic joining hands gesture (24 in HAI & 11 in HHI) used to attach pieces together; multiple joining hand (6 in HAI & 4 in HHI) used to assemble parts that come in pairs. In addition, listeners in HAI failed to successfully attach assembly pieces in the right position with the right orientation 9 times while those in HHI were unsuccessful 14 times (Failed joining hands- Figure 4.30; 5.8; Appendix VIV.23d). This suggests that listeners in HHI were less successful during interaction than those in HAI even though, the agent's voices appeared to be the most difficult to interact with.

While listeners in HAI used only three variations of the joining hand, those in HHI used four varieties. Specifically, listeners in HHI employed the pinch grip twisting joining hand (Figure 5.17) to screw assembly pieces together from the side, dual joining hand gestures (Appendix X.7d) to attach pairs of assembly pieces and on one occasion, a participant in HHI combined joining hand gesture with a displaying hand (Figure 5.16).

There were differences observed in object-manipulation strategies used in HAI and HHI contexts although the same fragile assembly pieces were assembled during interaction. While participants in HAI enacted self-restrain through gestures (Figure 4.38) when detaching assembled pieces, those in HHI enacted self-restrain through three transitions of facial actions (Figure 5.10) when attaching assembly pieces. These differences in enactment and use of self-restraint suggest that participants may have used task-specific self-restraint that suitably reduced task difficulty. This supports research indicating that in the manipulation of an environment, some responses are easier to physically execute while others are more difficult thus, individuals use guidance or restraint to enable them execute courses of actions (Martin & Pear, 2016).

Participants in HAI contexts on two occasions used eclectic strategies that produced the concatenated mixed hand gestures to execute joining hands (Figure 4.35) with many gestures flowing seamlessly into one another within one assembly stage. In contrast, participants in HHI use dual joining hand used to segment assembly tasks. Furthermore, while participants in HHI use precise hand movements to strategically attach and reposition assembly pieces (Figure 5.17), those in HAI use the same gesture to gauge the fit of assembly pieces (Appendix VIV.19 a-c).

The position of this thesis is that, the ability to create and use task execution strategies is invariably linked to the listener's comprehension process (Figure 2.1) because gestures as visual interpretations of assembly instructions depend on the listener adaptive use of individual comprehension skills developed from performance, implicit knowledge, listener culture capital, conceptual knowledge of the language of the multilayered interaction context (Chapter 2.2).

Furthermore, differences may be attributed to listener communicative expectations (Chapter 4.2.2) and task expectations (Table 3.2) in the joint action with instructors. Participants are expected to successfully assemble the Lego kit using verbal instruction and being aware of such expectations may also affect the listener's approach to the interaction. As new challenges emerge during tasks, participants instinctively deal with their interaction context to ensure that the aim of interaction is achieved using approaches and knowledge that materialise as continuous creative operative skills during interaction in HAI and HHI contexts, such as operative gestures.

6.5 Gesture as Communication

Although, the interaction context is unidirectional, the results indicate that listeners are still able to connect with instructors using interpersonal assembly gestures and engage in some form of internal communication within themselves using intrapersonal assembly gestures during interaction.

Interpersonal assembly gestures include gestures used for requesting new instructions, repetitions and managing turn-taking. While participants in HAI use the gesture to make the agent repeat instructions, those in HHI accomplish the same by using head nods accompanied by the backchannel ‘umhmm?’ or simply putting down the piece they picked up and looking at the instructor. Furthermore, while agent design forces this on instructor and instructees in HAI, the pattern in HHI is influenced by experience in other HHI interactions.

Moreover, results indicate that even in unidirectional instruction-giving contexts, people may still self-allocate turns during interaction similar to hearer interruption or interjections of speaker-utterance during conversations. For example, participants in HAI, on five occasions, violated the turn-taking transition sequence by delaying in the execution of current instructions, combining two or more instructions and pre-empting by executing instructions that have not been issued (see Figure 4.36). In contrast, participants in HHI interacted like people holding a conversation and did not disrupt the turn-taking arrangement.

Although Clark (1996) argues that comprehension ends when listeners achieves the same meaning with speakers, the results of this study suggest that comprehension goes beyond this point because listeners often go on to give their own contextual interpretations to guide their use of self-propagated gestures to execute instructions. For example, listeners use intrapersonal gestures to carry out Self-Initiated Self-Repairs (S-I-S-R), focus on the task and interpret instructions correctly in on 2 occasions in HAI and 6 in HHI contexts. However, while participants in the HAI context used repairs to correct joining and picking errors, those in the HHI context carried out repairs retrospectively by disassembling previous stages to re-assemble the kit.

In a context where the only mode available to communicate for interlocutors is gesture, the results suggest that participants in HAI and HHI commonly used the deictic gesture (Figures 4.40; 5.22), such as spatial reference to locate specific 3-D points on the assembly kit These gestures combine the instructor's utterance (encoding) with the listener's nonverbal interpretation(decoding)into a single communicative event from verbal and visual perspectives.

However, participants in the HAI contexts used 7 deictic gestures while those in HHI used 4 to concurrently produce iconic representations of lexical items ('widest end') in instructions and measuring the size of assembly bits while pointing to the 'feet' to identify assembly points (deictic co-interpretational gesture, Appendix VIV.25c). They also use the ticking gesture in deictic enumerating hands to create an iconic representation of a pen checking off items on tally sheets to select assembly bits (Appendix VIV.25b). Furthermore, when participants in HAI experience information overload, they may use co-interpretational gestures such as deictic hedge (Figure 4.41) enacted by wriggling the forefinger like a violin player doing a vibrato while pointing towards the kit layout to mitigate the situation.

The results suggest that, in both contexts, the kinesic structure of a gesture may affect its meaning; for example, unlike decisive pointing hand gestures, the deictic co-interpretational hand contains representational and deictic phrases that occur concurrently. The representational phrase is enacted with the digits spreading back and forth to mimic an iconic representation of the words "...widest end…" in the instruction. At the same time the digits form a span measuring the widest end of the Lego kit which is deictic. These differences confirm Kendon's (2004) suggestion that the way and manner a pointing gesture is done may make a difference to its meaning.

Unlike other gestures that were directly referred to in instructions, presentation gestures are not mentioned in the instructions; however, participants in HAI and HHI use presentation gestures (35 and 3 respectively) to scaffold operational stages of the assembly process. For example, participants use them to examine pieces selected and assembly stages completed. However, when presentation gestures are situated within the task timeline (Figure 4.3) participants in HAI used presentation hand gestures to display assembly pieces with supine hand(s) once (Figure 4.23; Appendix VIV.20a) or several times (Figure 4.25), and circular presentation (Appendix VIV.20d) for cylindrical objects. In contrast, participants in HHI extended pre-operational gestures by using both hands to compare two or more assembly pieces at once (double presentations in Figure 5.25). Following these, presentation gestures are described in this thesis as contextualised display gestures enacted to meet specific self-initiated iterative task-assessment needs during interaction.

6.6 Key Nonverbal Listenership Behaviours in HAI and HHI

The specific findings are summarised in Table 6.3overleaf andoutlined as follows. Interaction management: listeners in HAI and HHI contexts manage interaction by using repetition when receiving instructions, employing the same repeat sequence, and carrying out repairs during tasks. However, while listeners in HAI were forced by agent design to use only nonverbal modes to request instructions and be in control of the interaction, those in HHI had a more flexible approach because they interacted with a human instructor.

Listener-attitudes during interaction (Table 6.3): listeners in HAI and HHI contexts displayed facial actions suggesting that they had positive, neutral and negative attitudes towards the interaction and/or interlocutor during assembly tasks. However, Face-saving attempts due to researcher's presence may have influenced the display of microexpressions to hide real emotions and attempts to seize ground or violate turn-taking to assert territoriality over agents. At another level, Face threats could also have been responsible for the restricted display of emotions when interacting face-to-face with human instructor in HHI context.

Task execution strategies (Table 6.3): as per task execution strategies, listeners in both contexts relied on route knowledge acquired and facial actions to provide critical visual support to physical task execution. Additionally, listeners used gestures to nonverbally interpret verbal instructions as composite action such as assistive and operative assembly gestures during interaction. However, while listeners in HAI enacted gestures using simple and concatenated kinesic structures with unique shapes using both hand alternatively, those in HHI predominantly used simple structures with precise movements and multi-segment phases.

Table6.3 Key Nonverbal Listenership Behaviour Patterns in HAI & HHI

Nonverbal Listenership Behaviour		Interaction Context	
Key		HAI	HHI
Interaction management	Request modes	Nonverbal	Verbal and nonverbal
	Repeat process	Involves one continuous and integrated process: initiation, execution (repeat points) & resolution	Involves four diverse and independent events such as completion drills and instructor disfluencies
	Repeat sequence	Before assembly action: R1a; R4b; R7; R9 Co-occurs with assembly action: R1b; R4d; R4e; R5; R6; R7; R8; R9 After instruction / assembly action: R2, R3 R4a; R4c; R4f; R6; R8; R9	Before assembly action: R1a Co-occurs with assembly action: R1b; R5 After instruction / assembly action: R2a; ; R4a
	Managing turn-taking	Agent design-based, listener in control, with listener self-allocating turns at will. Leads to kit construction transition violations by listener-where listeners execute instructions at will instead of following agreed sequence	Implies agreement between instructor & instructee. Control is equally shared. No violations of turn-taking arrangement
	Repairs	Self-initiated self-repairs	Mutually allocated with permission from the instructor; repairs, backward repairs
Attitudes towards instructor & interaction	Facial actions	Displayed a wider variety of facial actions with more neutral faces. Showed surprise, frowns and a blend of emotions. Compressed lips, pouty faces, lip bite & work face- tensed mouth action. Aha! moment & Microexpressions	Narrower variety of facial actions with more neutral faces. Transition of emotions. Tense mouth compressed lips
	Perception	Positive, neutral and negative	Positive, neutral and negative
Task execution strategies	Facial actions	Facial actions: Three-way saccadic eye movement	Two-way saccadic eye movement
		Relied on route knowledge in task 2	Relied on route knowledge in task 2
	Assistive assembly gestures	knowing hand, dual picking hands, picking keeping hands & double	knowing hand, dual picking hands, picking keeping hands & double
	Shape	Crooked aim, searching hand, false starts; Aha! moment	None used
	Handedness	Alternate picking hands, repositioning, two-finger grasp; top-down keeping hand	None used
		None used	Combination of double and dual picking hand, picking with repairs
	Kinesic structure	Simple & Concatenated gesture structures	Simple structures
	Operative gestures	Joining & multiple joining hands.	Precise Joining, twisting and twisting hands. Dual joining hands with semi-multi-segment phases. Joining & presentation
		Basic aligning hands (sometimes physically touching) Twisting hands	Basic aligning hand (never touching); micro-aligning
	Presentation	Pre-operational (basic presentation, continuous opening hands, circular,); Post Operational (one-hand, lateral presentation)	Pre-operational (basic presentation, double presentation)
	Communicative gestures	Self-restraint-given by vibrato hand probably necessitated by fragile assembly bits and convoluted execution of instructions	Self-restraint indicated by compresses lips
	Intrapersonal	Deictic co-interpretational, ticking enumerating gesture, deictic hedge;	Deictic co-interpretational gesture used to locate specific parts of assembly bit.
Meaning-making	Interpretation	Iconic Interpretation of key words in instructions	Iconic Interpretation of key words in instructions
	Typology of repeats	R1a-b; R2; R3a-e; R4a-e; R5; R6; R7; R8; R9	R1a-b; R2; R4a; R5

Meaning-making (Table 6.3): regarding meaning-making, listeners in both contexts used iconic gestures to interpret key words in assembly instructions and repetition as vital meaning-making devices, their intrapersonal gestures such as nonverbal hedges, deictic co-interpretational gestures may offer co-interlocutors a view into the psychological processes involved when decoding instructions. However, listeners in HAI contexts use nonverbal private talk in ways similar to children's private talk to self-negotiate meanings because as outlined in Chapter 6.2, they had more reasons to ask for repeats than those in HHI.

As envisaged in Chapter 2.2.3.3, a bidirectional relationshipexists between nonverbal listenership behaviours asinferred from the results indicating that, facial actions and gestures perform overlapping and coordinated functions during interaction. These tie in with McNeil's (1985) argument that gestures are produced by the same internal systems that produce overt language skills and thus share the same computational similarities.

Gestures and facial actions are said to overlap when they perform the same pragmatic or task function as shown in the case of static searching face-saccadic eye movements and searching hand gesture used to locate assembly pieces.In addition, gestures and facial actions are said to perform coordinated functions when the display of one behaviour sheds light on why the other occurs at the same time. For example, Aha! moments, repairs, and presentation gestures are used to externalise a participant's ongoing chain of thoughts, while the co-occurrence of about 90% of the neutral faces with stroke phrases in assembly gestures signals concentration during interaction.

Following these, this thesis argues that since gestures and emotions are controlled by the same system and their communicative functions overlap orare integrated, it can be safely concluded that assembly behaviour is projected as an ensemble of gestures and facial actions concurrently used to concurrently display emotion, determine turn-taking during interaction, guide the instructor, carry out nonverbal private talk, execute assembly tasks, and strategically break tasks into manageable bits.

6.7 Implications of the Study

Following the comparative analysis of nonverbal listenership behaviours done above, this chapter outlines the implications of the studyfor current sociolinguistic theories of nonverbal listenership, communication, HAI and HACspartnerships.

As outlined in Chapter 6.2, listeners successfully use facial actions and gestures as pragmatic markers to indicate their comprehension or incomprehension of assembly instructions. These listenership behaviours were present even though the agent could not perceive them just as they were present when interacting with a human instructor who could. The display of similar nonverbal behaviour in both contexts confirms earlier research in HCI that people treat agents as real people and spaces (Clark et al., 2014; Nass et al., 1999). However,listener display of context–specific pragmatic markers suggest that meaning-making is strongly influenced by interaction contexts. Following these, this Chapter outlines the implications of this study for sociolinguistic theories of nonverbal listenership as they relate to the following:

- Pragmatics of interaction in HACS
- Identity projection and perception in HACs.
- Given the context of interaction, there are implications for applied linguistics research methodology regarding multimodal corpus development and analysis.

6.7.1 Implications for Pragmatics of Interaction management

Pragmatics is concerned with the interpretation of utterances as expressed within the interaction context, in accordance with the active rules of interaction within that context and how utterances aid interlocutors to achieve the aim of interaction.

As outlined in Chapter 2, language use is a form of joint action between instructor and instructee where each has responsibilities and expectations. Responsibilities are linked to interlocutors' roles as instructors and instructees who give and receive instructions with the joint aim of assembling a Lego kit. Expectations relate to what interlocutors expect co-interlocutors to do and what the context will provide to ensure that the interaction succeeds which, as outlined in 6.2 includes clarity, adequacy and relevance of instructions to the interaction needs of listeners.

Furthermore, the results suggest that interaction between instructors and instructees evolves from two concurrent pragmatic processes. The first process relates to repetition where repeat typologies emerge as listener meaning-making strategies that portray listener interaction expectations (see Chapter 6.2).The second process relates to the order of interaction or turn-taking emerging as interaction management strategies. While turn-taking in HHI is governed by social rules, it is dictated byagent design in HAI. In addition, participants in this study use facial expressions and gestures as pragmatic markers to manage the interaction in HHI and task execution in HAI contexts.

Although repetition and turn-taking occur concurrently, turn-taking process determines how the interaction turns out while the repetition focuses on the strategies employed to achieve the aim of interaction (see 3.2.1, step 2b). The results indicate that people receiving instructions during assembly tasks use repetition and turn-taking to organise discourse, develop meaning and design task execution strategies in both HAI and HHI contexts (Chapter 4 & 5).

As outlined in Chapter 2.2-2.3, listeners may come to an interaction context with preformed ideas from their background knowledge and experience regarding their roles. These ideas influence the listener's objectives and approach to information processing during interaction (Chapter 2.3).To this end, this thesis holds that both instructors and instructees have objectives tied to the task and interaction context that may in some ways, enable or hinder the interaction. While instructors aim to issue instructions, listeners aim to use the instructions to assemble the Lego model. However, the structure of discourse as determined by the treatment of given and new information by both interlocutors during turn-taking may further shape their actions and role performance as successful instructors and instructees in task-based contexts envisaged in HACs.

The actions refer to nonverbal listenership behaviours, for example, listeners in HAI seize and allocate turns during interaction even though the pattern set distributes equal turns for interlocutors. Turn-taking violation is linked to the notion of territoriality (Sindoni, 2013) which describes how and where interactants seize and defend interaction spaces. In the assembly task, turn-taking violation disrupts the interaction and later manifests as kit constructional unit violations, a task action that may disrupt or enhance the assembly process (outlined in Chapter 4). It may provide insight into the character of language dynamics occurring within HAI and reinforces the notion that there is a potential for all talk to be embedded in a power relationship as a force that enables or inhibits some (Jaworski & Coupland, 2006). Thus, pragmatics needs to define how HACs teams and agents initiate and incorporate turn-taking into interactions as a basis for calibrating turn-taking schema for HAI.

Going further, the interaction process is tied to the listener's meaning-making process as signalled by repetitions that listeners strategically use at different points in the assembly process for clarification, chunking tasks, confirmation or any combination of these (see Chapters 4 and 5).These suggest that when people interact with agents, they do not only rely on cognitive processes, they also use other strategies in making sense of interaction (Murdoch et al., 2013; Nieto, 2013), such as combining repetition with an Aha! Moment to self-negotiate learning and mediate meaning through performance.

Although this is a unidirectional instruction-giving context, the Gricean notion that speaker-information must be clear and adequately informative (Grice, 2006) remains relevant because when these are flouted, listener communication expectations may not be met. To meet communication expectations, listeners use repeat sequence to manage interaction and reduce task difficulty, which suggests that, listeners may not only be self-negotiating meaning internally, but when listeners take the floor they are nonverbally asking the agent "What does that mean?" with every repetition for clarification.

Listener self-initiated self-repairs are focused on making meaning of instructions, identifying errors, and repositioning task execution using experiences gained from route knowledge. Route knowledge in this thesis describes the spontaneous know-how a participant develops from carrying out repetitive assembly tasks. Route knowledge parallels Tannen and Wallat's (1987) knowledge schemas in that they both refer to an interlocutor's use of prior knowledge in current interaction contexts. These conclusions indicate that listeners mediate meaning using facts they bring to interactions (Vandergrift, 2004) as well as implicit knowledge developed unintentionally.

Furthermore, while self-initiated self-repairs provide a self-correcting mechanism for the organisation of language during interaction (Schegloff et al., 1977), there is still the question of how agents may take on repair organisation. Moreover, the use of repairs during tasks suggests that listeners are becoming more self-regulatory and making interaction more meaningful when receiving instructions. However, there is the need for a linguistic theory that will account for organisation of nonverbal repairs in HAI and devise ways of handling them when humans take instructions from agents in HACs.

The study indicates that facial actions are not restricted to emotions but can be used by participants as assembling strategies during task execution. Facial actions such as compressed lips and biting lips are used to indicate physical effort deployed by participants when handling assembly bits. These foreshadow discussions on task execution strategies later.

Explicit and implicit attitudes play an important role in explaining human behaviour such as in consumer choices and learners (Ackermann & Mathieu, 2015; Neto, 2009) and have also been displayed as emotions. To this end, the results confirm earlier studies suggesting that spontaneous facial actions displayed during interaction may be useful in assessing listener attitudes (Sherman et al., 2003). For example, a positive face like the felt smile (Figure 4.8) may indicate a listener’s positive disposition towards the instructor just as anger (Figure 4.13) may suggest a negative disposition. In this way, listener attitudes may subjectively organise interaction situations and orientate them towards agents that are able to respond to such nonverbal behaviourseffectively.

Furthermore, these findings reinforce the view that, emotions are crucial to the development and regulation of interpersonal relationships in any interaction context (Clark et al. 2016b; Ekman & Rosenberg, 1997). Although, facial actions are easily accounted for in HHI as indicated by studies in human only interaction, it is the position of this thesis that there is a potential for linguistics to provide an informed system that may account for the communicative use of listener facial actions in HAI.

These conclusions foreshadow discussions on the need for agents to respond to human behaviour later.Repetition, turn-taking, self-initiated self-repairs and task execution form parts of the listener's pragmatic toolkit for managing interaction and may provide a linguistic barometer for measuring listener comprehension or lack of it. However, these aspects of interaction are not adequately covered by existing rules of discourse competence (Canale & Swale, 1980) even when violating turn-taking disrupts interaction. The challenge for pragmatics is to look beyond grammatical rules and cohesion to context-specific appropriateness in language use that will enable humans and agents develop a flexible relationship (Jennings et al., 2014). Hence, this thesis opinesthat current theories of pragmatics and listenership need to provide guidelines for structuring interaction in HACs to serve the interlocutors' needs for comprehensible inputthat may make communication within HACs effective.

6.7.2 Identity Projection and Listener Attitudes

While the instructor in HHI is human with a physical presence, the instructor in HAI is a simulated agent whose personal identity is linked to its voices(see Chapter 2.6). The voice is a distinguishing mark because most people are cognisant of the fact that individuals have different voices and can accurately recognise them from their utterances without seeing them. Research suggests that listener recognition may be due to distinctions in speaker verbal features such as intensity; pitch; nasal co-articulation; spectrum; predictor coefficients; spectral coefficients; formant frequencies; timing and speaking rate (Atal, 1976). As outlined in Chapter 3, the agent used a cline of voices (CL, CP & HR) from the same age bracket and dialect but each agent voice has a personal biometric identification that is linked to its personal verbal identity making it distinct from others.

During interaction, the agent's verbal identity is projected as a social identity (Chapter 2.6) through assembly instructions while listeners engage in two automatic and simultaneous processes when taking instructions. The first is speaker recognition using some features of agent verbal identity to perceive and determine who is making the utterances (speaker identity) and the second process relates to processing what has been said as a way of further shaping listener perception and attitudes towards the speaker and interaction.

Regarding speaker-voice identityrecognition, Kersta (1978) and Doddington (1985) separately found that speaker recognition among people is very accurate even when voices are disguised. This suggests that listeners can accurately differentiate the projected identity of one instructor from another using their voices.

In addition, a recent study suggests that speaker recognition may be linked to listener’s linguistic processing abilities such as the ability to understand speech (Besser et al., 2013). This ties in with the view outlined in Chapter 2.3 indicating that listener comprehension is influenced by linguistic competence such as making meaning of texts.The results also indicate that, listeners displayed a range of facial actions when interacting with the different instructors. This implies that listeners did not only perceive instructors as different because of their voices, they also externalised their perceptions through facial actions,which may represent their attitudes towards instructors and interaction as positive, neutral or negative(Mehrabian & Ferris, 1967).

Hence this thesis argues that facial actions potentially perform the pragmatic function of projecting a listener’s emotive position regarding the instructor’s projected identity during interaction.The caveat here is that interpreting facial actions is to some extent, personally, culturally and contextually determined.

Going further, pragmatics focuses on appropriate language use in any given context however, due to the arbitrary nature of nonverbal language the concept of competence may need to be re-appraised because competence in one form of nonverbal language may not automatically be transferred to other contexts due to cultural factors.

In addition, while it is possible in HHI to carry out mutual self-monitoring during interaction because this enables people to be better attuned to the expectations and reactions of others (Boyd, 2002), it is still unclear how agents might do this spontaneously. Moreover, self-monitoring can be linked to deliberate listener feedback during interaction; however, it is not clear whether this also occurs with spontaneousfeedback.

The interactive implication for listeners is the need to connect with agents through accommodation (Figure 2.3) of agent communicative strategies (Tope et al., 2014) just as first language speakers accommodate second and foreign language speakers or mothers do their babies that are just beginning to acquire their first language during interaction. Furthermore, current theories of discourse identities need to carry out a careful analysis of talk-in-interaction (Schegloff, 1997) to develop a better understanding of active role-play in the construction of social identities in HACs.

6.7.3. Multimodal Corpus Research Methodology

Having problematised the issues surrounding nonverbal listenership behaviours in unidirectional instruction-giving contexts, the results were analysed in two layers - facial actions and gestures- which is suitable for analysis and follows established research procedures as used by Ekman and Friesen (2003, 1978a-b) McNeil (2005,1992), Kendon (2004) and Kita et al. (2013), with a focus on separate but combinable semiotic resources in nonverbal interaction.

The implication for applied linguistic research is to devise a multimodal corpus linguistics coding matrix for annotating various co-occurring nonverbal listenership behaviours with a view to reconciling these fragments into a coherent discourse during analysis. This may clarify how multimodal fusion is achieved in the perception and understanding of communication. For example, reconciliation at the level of analysis may enable researchers to understand how interlocutors integrate co-interlocutor’s facial actions, gestures, voice, utterances, and posture to arrive at a multimodal interpretation of the information exchanged during interaction.

In addition, assembly gestures emerging from this research suggest that linguists may have to take a more contextualised view of gesture phrases rather than the generalised views propounded by Kendon (2004), Kipp (2003) and Kita et al. (1998) to cover unique task-based interactions. Furthermore, results indicate that the kinesis structure of gesture phrase may have more than the hitherto perceived phases. Some assembly gestures in HAI contain phases within other phases such as in the multiple retractions and strokes within the deictic enumerating gesture (Appendix VIV.25c). Some contain gestures within the phrase such as picking in preparation for alignment (Appendix VIV.25f). These phases and gestures perform roles that if absent will render the whole gesture phrase incomplete, non-functional or meaningless thus, there is the need to devise context-specific ways of analysing gestures in task-based interaction.

6.8. Summary and Conclusion

This Chapter compares listenership behaviours in HAI and HHI contexts and outlines the implications. What seems clear is how the participants used pragmatic markers (such as facial actions and gestures) and other back-channelling moves to indicate their comprehension orincomprehension of instructions.

These gestures and expressions were present even though their instructor in the HAI context (the simulated agent) could not comprehend or orient to them like the human instructor inthe HHI context. This suggests that even in unidirectional communication, humans may still express themselves in a way that shows their communicative competence to maintain their interactional identities.

The results also suggest that although every mode is meaning-making, the interaction context may be the strongest determinant of meaning in multimodal interaction as indicated by the context-specific pragmatic markers used by participants in HAI and HHI. Furthermore, participant display of similar nonverbal listener behaviours in HAI and HHI suggests that data from existing applied linguistics research focused on HHI may be useful for deriving rules and heuristics that will drive agent behaviour (agentiquette) in the future. However, listener display of context-specific nonverbal behaviours suggests that agentiquette should be interaction- and context-specific. These findings will foreground the thesis conclusions in Chapter 7.

Chapter 7 Conclusions

7.1 Thesis Overview

The thesis tried to answer the following research questions:

1. What spontaneous facial actions and gestures do listeners display in HAI when requesting repeats of agent instructions during assembly tasks and why?
2. What spontaneous facial actions and gestures do listeners display in HHI when requesting repeats of instructions from humans during assembly tasks and why?
3. How do spontaneous facial actions and gestures displayed in HAI compare with those in HHI?

To answer these questions, this thesis reviews relevant literature in Chapter 2, focusing on related existing knowledge such as; theories of nonverbal listenership as given by facial actions and gestures; human social behaviour towards computer; verbal identity and its relationship with identity perception; and constructing instructions as information packages and their roles in unidirectional instruction-giving. Chapter 3 addresses the multimodal study design and annotation framework. The first section of Chapter 3 outlines the general procedures used to develop the multimodal corpus and execute the study. The second section introduces and describes the annotation scheme used for coding nonverbal listenership behaviours emerging from the interactions, while Chapters 4 and 5 use context-specific methodologies. Chapter 4 examines the pattern of nonverbal listenership behaviours when taking instructions from a simulated agent (HAI) using a cline of three voices. Chapter 5 examines listenership behaviour patterns in human-human interaction (HHI). Chapter 6 compares the patterns of nonverbal listenership behaviours in HAI context with those in HHI and outlines the implications of the study.

7.2 Contributions to Knowledge

The main goal of this thesis is to understand the role and impact of spontaneous nonverbal listenership behaviour that occurs when humans take instructions from agents and humans in unidirectional instruction-giving contexts. Specifically, the study uses the dual lenses of spontaneous facial actions and assembly gestures to assess interaction in HAI and HHI contexts.

The main contributions of this thesis include the creation of a multimodal corpus, identifying and extending the classes of spontaneous facial actions, conceptualising and classifying assembly gestures; introducing task-oriented assembly gestures; and conceptualising and classifying nonverbal repetition process and typologies as interaction strategies. Another is the design and use of an annotation scheme suitable for coding aspects of the corpus in a reasonably systematic way. Finally, it proposes a pathway for the development of responsive agents using applied linguistics data.This may lead to the development of agentiquette and agent response to human nonverbal behaviour.

7.2.1 Corpus Creation and Annotation

This study has resulted in the creation of a 27-hour computer-assisted multimodal corpora from 108 interactions to provide material for more complete studies of interactive sharing and construction of meaning and understanding in assembly contexts. This enables multimodal analysis to focus on the form, functions and lexis of semiotic modes people use to communicate with each other and agents (Allwood, 2010; O'Halloran, 2009). With increasing digitisation, Kress and Van Leeuwen, (2006) opine that different modes have to a large extent technically integrated. Thus, multimodal analysis needs analytical techniques that are capable of tracing and analysing these integrative patterns of multiple modes used in interaction. To this end, the thesis developed a detailed annotation scheme for systematic codification, classification and analysis of marked listener facial actions and gestures displayed during assembly taskswhich makes multimodal corpora readable, valid and comprehensible to users.

7.2.2 Facial Actions

This thesis presents a multimodal corpus that has been annotated for spontaneous listener facial actions. It extended the existing classification to include different types of neutral faces, moods (aha! Moment, work face), and static searching face. The study also modified Ekman and Friesen's (1978a-b) five-step classification of emotional intensity to only three (Table 3.7) specific levels to avoid ambiguity.This is partly in recognition that multimodal interaction between human andsimulated agentsinvolve several types of facial expressions that can replace sequences of words or accompany them in meaning-making during interaction.

7.2.3 Assembly Gestures

This thesis introduces the concept of spontaneous assembly gesturesto redefineand extend Kendon's continuum (McNeil, 2006). Assembly gestures are body movements used to perform tasks, manage interaction and communicate in instruction-giving contexts.They are task-oriented as assistive, operative and presentation gestures used to locate, select, join and assess assembly pieces or stages. Assembly gestures are used to manage interaction when used tosignal self-turn allocation techniques such as self-initiated self-repairs and aid listener comprehension when they co-occur with instructor utterance as co-interpretational gestures. Assembly gestures are communicative as interpersonal gestures used by listeners to connect with instructors and as intrapersonal gestures used to conduct nonverbal private talk, self-negotiate or simply externalise listener's thought processes. In addition, this study expands the gesture phrase to include the failed phase because failure occurs during task execution and existing annotations have not paid attention to this aspect of gesture analysis. This has led to the emergence of failed gestures as a way of identifying errors during tasks. These summarise the concept, characteristics and functions of assembly gestures.

7.2.4 Repetition Process and Typologies

This study introduces a systematic discernible process of repeating requests framed within task timelines. Although listeners in HAI ask instructors to repeat instructions by pushing a button however, this action transforms into a meaningful gesture when situated within the contexts of interaction, the goal of the interaction and other interactional factors as outlined in 3.3.1.3. The process also presents a codified typology of repeats (R1-R9) indicating the pragmatic function of each repetition emerging from interactions. The repeat process and typologies can be used to identify the spontaneous communication principles and task strategies used to orchestrate meaning during interaction.

7.2.5 Responding to Human Behaviourin Multimodal Interaction

Multimodal analysis involves the analysis, search and retrieval of sign systems and investigate the relationships between them. One of the contributions of this thesis is the creation of a multimodal corpus that could be used for infusing personality and emotional traits in agents that could make interaction with humans credible (Bevacqua et al., 2010).

There are examples illustrating this point starting with Bouissac's (1976) recommendation that human body movement be studied in terms of the 3-D volume occupied. Bouissac's approach provided a basis for Ning, et al., (2006) and Huazhong et al. (2006) to develop a computerised motion detector and a real-time shoulder shrug detector. Boker, et al., (2009) also used the same study to develop a computerised avatar used for measuring head movements.In addition, Maatman et al. (2005) identified and used three categories of natural listener behaviours such as head nods, para-verbals like 'uh-huh, mm-hmm', and disfluency indicating information overload to derive rules that drive a listening agent's behaviour. Furthermore, while focusing on personal space as an aspect of respect in interaction, Dautenhahn (2007) used child play to develop Robotiquette which is a code of social behaviour in human-robot interaction. Similarly, Hu (2014) used the behaviour that people walking display when following other's footsteps to derive rules guiding the behaviour of anticipatory agents.

Recent research in artificial intelligence indicate that new techniques exist that enable agents, such as Facebook's chatbots, to successfully navigate negotiations in a human-like way (Lewis et al., 2017) and robots such as NASA's Pepper have emotional intelligence to sense and think about their environment, then act based on their perceptions during interaction almost like humans (BBC News, 2017).

However, getting agents to accurately understand and respond communicatively to human nonverbal listenership behaviours at a human level remains a challenge in view of the increasing interaction between humans and agents in diverse contexts. These still re-emphasise Jennings et al's (2014) observation that there is the need for intelligent agents in HACs to respond to human behaviour during interaction. Agent response is described in this thesis as the agent's ability to accurately perceive and decode human nonverbal behaviours and provide communicative feedback at a human level.

Following these, and in response to the need for agents to recognise and use human behaviour communicatively, this thesis proposes a pathway for using applied linguistics data to derive rules that will drive the communicative behaviour of agents. For example, facial actions and gestures may be dynamically modelled and interactively coded using technologies for visualisation, tracking, 3-D modelling and reconstruction. In addition, this study aims to engender further discussion around the potential collaboration between applied linguists and agent designers in the development of agents that understand human nonverbal behaviour at a human level.

As to the usefulness of applied linguistics, this thesis believesthat data from applied linguistic studies may be used to derive the behaviours of agents and to enhance the capabilities of existing agents. To this end, the facial actions and assembly gestures analysed (Chapters 4 & 5) created an ensemble of nonverbal listenership behaviours that could be used to develop communicative rules for agent behaviour ('agentiquette'). Agent etiquette or agentiquette is a set of heuristics and guidelines on how future agents should recognise and use human nonverbal behaviour communicatively during interaction.

Furthermore, the trends outlined in Chapter 6 indicate that people display similar and context-specific pragmatic competences when they interact with agents or humans during assembly tasks. Similarities imply universal recognition and interpretation of such behaviours and these reinforce the view that, existing data from applied linguistics may be applicable in deriving agentiquette for agents with wide applications where multiple contexts share common interactional behaviours. However, context-specific nonverbal behaviours suggest that contexts-specific applied linguistics data may be used to derive the behaviours of context-specific agents or adapt existing agents to meet specific interaction needs.Thus, agentiquette may be used to create a relational bond in HAI although there is the challenge of accurate discernment of emotions and some gestures.

Regarding the potential collaboration between applied linguists and agent designers in the development of agentiquette, the procedure for deriving social rules may begin with developing a corpus and its annotations as done in this study. Computer programmers use the corpora annotated for marked facial actions and assembly gestures to analyse listener nonverbal behaviour into logical syntactic components in the same way that applied linguists use tree diagrams to analyse sentences. The syntactic components may be used to derive rules that will drive agentiquette in HACs, potentially making them more interactive and responsive to human behaviour. For example, with enhanced perception, agents will scaffold user experience and through improved self-awareness regarding error, agents may have the capability to identify and interpret self-initiated self-correction by humans at human level just as in HHI.

7.3 Limitations and Further Research

This thesis addressed existing issues in the field of listenership in HAI and HHI using speakers of English as a first language, exploring the same issues using populations from other contexts of English language usage such as speakers of English as a second language, may enable researchers to validate this study.

This thesis principally aimed to develop a framework for understanding nonverbal listenership behaviours in instruction-giving contexts but, the samples of this study consisted of unequal population sizes, HAI context had 43 participants, while HHI had 6 participants. Although, these contexts provide a basis for comparison, it must be recognised thatthe generalisability of the findings is limited. It would be desirable for future studies to use the framework to replicate the findings with equal samples to enhance generalisability.

This study was done in experimental settings; however, there is a need to assess interactions between humans and agents in the wild (Adolphs et al., 2016) such as the use of Siri or amazon echo in contexts outside laboratories because this will liberate the researcher and provide opportunities to investigate nonverbal listenership behaviours and instruction-giving in real-life contexts.

In addition, the interaction context used in this study is unidirectional instruction-giving where one instructor gives verbal instructions to one participant. Exploring two-way communication paradigms within other activities, such as play, may enable researchers to translocate the study. Further work may also explore non-instructive and functional contexts such as healthcare, education, transport, and family leisure time interactions. However, researchers should be aware of the ethical and implementation challenges surrounding such studies.

The study used a simulated agent as the instructor but using a more dynamic and responsive agent may provide a more realistic interaction and provide results closer to real life contexts. In addition, exploring contexts where one instructor gives instructions to a group of listeners or even one listener receiving instructions from two alternating instructors may address issues around group dynamics that this study does not focus on such as, team work, collaboration in tasks, or meaning development within groups attending to the same text.

This study used on facial actions and gestures as sources of data however, interpreting them may be very subjective due to impact of culture and context. Future researchers may consider using additional data-gathering tools such as questionnaires and interviews to enable participants express their views in writing and as narratives. When these are combined with nonverbal feedback, the data analysis may become more comprehensive and objective.

In addition, this thesis indicates that listeners use nonverbal private talk for self-regulation (Montazeri et al., 2015)during tasks but this thesis was not able to give adequate attention to its occurrences in every interaction observed. Future research may focus on nonverbal private talk in assembly tasks as this may enable researchers to identify, classify and assess its role in listener comprehension or incomprehension of speaker input during interaction

In view of the coordinated and overlapping functions of the ensemble of nonverbal listenership behaviours outlined in chapter 6.6, there is the need to investigate how this relationship may reshape HAI in interaction contexts that emerge because of new technology; for example, banks are now using eye and palm recognition technology to replace passwords and it has been proposed that the London underground service will use facial recognition technology to check tickets instead of train tickets or oyster cards. These changes may have a profound impact on how people use their banks, manage transactions or organise travel. Moreover, investigating how this relationship impacts on the communication process in these emerging interaction contexts may enable researchers to calibrate the algorithm for determining how agents understand human behaviour.

7.4 Summary

The research presented in this thesis stresses the importance of taking listener behaviour on board becausepeople can use nonverbal listenership behaviour as a pragmatic indicator of their comprehension or incomprehension of speaker utterances. As well as programming agents to respond to human nonverbal listenership behaviour at human level, attention should be paid to ways in which listeners modify their behaviours to accommodate agents becausethese behaviours may be context-specific as in the case of assembly gestures. Facial actions, on the other hand, may be universal as in the case of basic facial actions (Ekman & Rosenberg, 1997). Although, interpretation to some extent may be context-specific, consideration should also be given to the ways in which interaction needs modify agent design.

References

Ackermann, C., & Mathieu, J. (2015). Implicit attitudes and their measurement: Theoretical foundations and use in consumer behavior research. *Recherche et Applications en Marketing RAM (English Edition), 30*(2), 55-77. doi:10.1177/2051570715579144

Adolphs, R., Baron-Cohen, S., & Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *Journal of cognitive neuroscience, 14*(8), 1264-1274.

Adolphs, S. (2008). *Corpus and Context: Investigating pragmatic functions in spoken discourse.* Amsterdam: John Benjamin.

Adolphs, S., & Carter, R. (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal.* New York, London: Routeledge.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage and spoken discourse. *Applied Linguistics, 24*(4), 425–438.

Adolphs, S., Knight, D., Ofemile, A. Y., Clark, L. M., & Rodden, T. (2016). Crowdsourcing new communities of discourse: Analysing human-computer interaction in different contexts. *American Association for Applied Linguistics (AAAL).* Florida, USA.

Aitamurto, T., Galli, J. S., & Salminen, J. (2014). Self-selection In Crowdsourced Democracy: A Bug Or A Feature?. *Computer Supported Cooperative Work (CSCW '14).* Baltimore, MD, USA: ACM.

Allen, R. (2007). *The Penguin Dictionary.* London: Penguin Group.

Allenwood, J. (2008). Multimodal Corpora. In A. Lüdeling, & M. Kytö, *Corpus Linguistics : An International Handbook (vol 1)* (pp. 207 - 224). Berlin: New York: Walter de Gruyter GmbH & Co. KG.

Allolio-Näcke, L. (2014). Hybridity . In *Encyclopedia of Critical Psychology* (pp. 925-928). New York: Springer .

Allwood, J. (1985). Intercultural communication. *Papers in anthropological linguistics, 12*, 1-25.

Allwood, J. (1993). Feedback in second language acquisition. *Adult Language Acquisition. Cross Linguistic Perspectives 2*, 196-236.

Allwood, J. (1995). Reasons for management in spoken dialogue. *NATO ASI SERIES F COMPUTER AND SYSTEMS SCIENCES, 142*, 241-241.

Allwood, J., Cerrato, L., Dybkjaer, L., Jokinen, K., Navarretta, C., & Paggio, P. (2005). The MUMIN multimodal coding scheme. In *NorFA yearbook 2005* (pp. 129-157).

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All Smiles are not created equal: Morphology and timing of smiles perceived as amused, polite and embarrassed/nervous . *Journal of nonverbal behaviour , 33*(1), 17-34.

Andrade, E. B., & Ho, T. H. (2009). Gaming emotions in social interactions. *Journal of Consumer Research, 36*(4), 539-552.

Annette, J. (2008). Hierarchical Task analysis. In D. Diaper, & N. A. Stanton, *The Handbook of Task Analysis for Human–Computer Interaction* (pp. 67-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Atal, B. S. (1976). Automatic recognition of speakers from their voices. *Proceedings of the IEEE*, *64*, pp. 460 - 475.

Babel, M., McGuire, G., & King, J. (2014). Towards a More Nuanced View of Vocal Attractiveness. (D. Reby, Ed.) *PLoS ONE, 9*((2): e88616). doi:10.1371/journal.pone.0088616

Bach, J. R., & Goncalves, M. R. (2004). Noninvasive ventilation or paradigm paralysis? *European Respiratory Journal, 23*(4), 651-651.

Badrudeen, S., & Sabhawarwal, S. (2010). Assessing Readability of Patient Education materials: Current role in Orthoaedics. *Clinical Orthopaedia Related Research, 46*(8), 2572-2580.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory. .* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*, 307-337.

Bartlett, M., Littlewort, G., Vural, E., Whitehill, J., Wu, T., Lee, K., & Movellan, J. (2009). Insights on Spontaneous facial expressions from automatic expression Measurement. In M. C. Giese, & H. Bulthoff, *Dynamic Faces: Insights from Experiments and Computation.* Massachusetts: MIT Press.

Baynham, M., Bradley, J., Callahan, J., Hanusova, J., & Simpson, J. (2015). *Language, Business and superdiversity in Leeds.* Birmingham: Translation and Translanguaging Project. Retrieved /02/ 12, 2016, from http://www.birmingham.ac.uk/generic/tlang/index.aspx

BBC News. (2017, june 18). Five Robots that are changing everything. *BBC News*. London, UK: BBC News Channel. Retrieved June 17, 2017, from http://www.bbc.co.uk/news/video_and_audio/headlines/40306617/five-robots-that-are-changing-everything

Becket, W., & Badler, N. I. (1993). Integrated behavioral agent architecture. *Proceedings of the Third Conference on Computer Generated Forces and Behavior Representation*, *3*, pp. 57-68.

Bell, S. S. (2015). *Reading body language, read body language of men and women.* Retrieved May 05, 2015, from Body language expert: http://readingbodylanguagenow.com/readingbodylanguageofbitinglips/

Bentahar, J., Bouzoubaa, K., & Moulin, B. (2006). A Computational Framework for Human/Agent Communication Using Argumentation, Implicit Information, and

Social Influence. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)* (pp. 372-377). Hong Kong: IEEE. doi:10.1109/WI-IATW.2006.9

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends in amplification, 17*(2), 75-93.

Bevacqua, E. E. (2010). Building Credible Agents: Behaviour Influenced by personality and Emotional TraitsL. *INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH (KEER 2010).* Paris.

Bezemer, J., & Jewitt, C. (2010). Multimodal Analysis: Key Issues. In L. Litosseliti, & (ed), *Research Methods in Linguistics.* (pp. 180-197). London: Continuum.

Bhabha, H. (1988). The commitment to theory. *New formations, 5*(1), 5-23.

Bhabha, H. K. (1994). *The Location of Culture.* New York, London: Routledge.

Bhabha, H. K. (2006). The Commitment To Theory: Cultural Diversity and Cultural Differences. In G. G. B. Ashcroft, *The Post-Colonial Studies Reader,* (pp. 155–157). New York: Routledge.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing, 8*(4), 243-257.

Bidabadi, F. S., & Yamat, H. (2014). Strategies employed by Iranian EFL freshman university students in extensive listening: a qualitative research. *International Journal of Qualitative Studies in Education, 27*(1), 23-41.

Birdwhistell, R. L. (1952). *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture.* Washington, DC: Foreign Service Institute.

Blackledge, A., Creese, A., & Trehan, K. (2016). *Enterprising Communities: How Everyday Entrepreneurs Create better Communities.* Papers in Translanguaging and Translation (WP.11). Retrieved from http://www.birmingham.ac.uk/generic/tlang/publications/index.aspx

Boker, S. M., Cohn , J. F., Theobald, B. J., Matthews, I., Brick , T., & Spies, J. R. (2009). Effects of Damping Head Movement and Facial Expression in Dyadic Conversation Using Real-Time Facial Expression Tracking and Synthesized Avatars. *Proc. Philosophical Transactions of the Royal Society Bv, 364*, 3485-3495.

Bomann, M. H., & Jones, Q. (2003). Information Overload. In K. Christensen, & D. Levinson, *Encyclopedia of Community: Fromthe village to the Virtual World* (pp. 661 - 664). Thousand Oaks, CA: SAGE Knowledge. doi:10.4135/9781412952583

Bonoma, T. V., & Felder, L. C. (1977). Nonverbal communication in marketing: Toward a communicational analysis. *Journal of Marketing Research*, 169-180.

Borg, E. (2003). Discourse communities. *ELT Journal, 57*(4).

Bouissac, P. (1976). *Circus and culture: A semiotic approach.* Indiana University Press.

Boyd, D. (2002). *Faceted Id/Entity: Managing representation in a digital world.* Massachusetts: Unpublished Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Media Arts and Sciences, Massachusetts Institute of Technology. Retrieved /02/ 20, 2015, from http://smg.media.mit.edu/people/danah/thesis/danahThesis.pdf

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Accoustic correlates of information Structure. In *Language and Cognitive Processes* (Vols. 25. 7-9, pp. 1044-1098). Taylor & Francis. doi:http://dx.doi.org10.1080/01690965.2010.504378

Brennan, N. M., Merkl-Davies, D. M., & Beelitz, A. (2013). Dialogism in corporate social responsibility communications: Conceptualising verbal interaction between organisations and their audiences. *Journal of Business Ethics, 115*(4), 665-679.

Bressem, J. (2014). A linguistic perspective on the notation of form features in gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendor (Eds.), *Body Language -Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38.1.).* Berlin, Boston: De Gruyter Mouton.

Brotherton, C. (1999). The Media Equation - How People Treat Computers, Television, and New Media like Real People and Places, B. Reeves and C. Nass. *Reviewing Sociology, 11*(2).

Brown, D. H. (2007). *Teaching by principles: An interactive approach to language pedagogy.* White Plains, NY: Longman.

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language usage,.* Cambridge: Cambridge University Press.

Bryman, A., & Bell, E. (2007). *Business Research Methods* (Second ed.). New York: Oxford University Press.

Brynjolfsson, M. E., & McAFEE, A. (2012). In *Race Against the Machine. How the Digial Revolution is Accelerating Innovation Driving Productivity, and Irreversibly Transforming Employment and the Economy. Research Brief.*

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse studies, 7*, 585-614.

Burke, P. (2007). Identity Control Theory. In G. Ritzer, *Blackwell Encyclopedia of Sociology.* Oxford: Blackwell Publishing Co.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*(1), 168-174.

Canagarajah, S. (2014). In search of a new paradigm for teaching English as an international language. *TESOL Journal, 5*(4), 767-785.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

Canales, M. K. (2000). Othering: Toward an understanding of difference. *Advances in Nursing Science 22.4*, 16-31.

Capone, N. C., & McGregor, K. K. (2004). Gesture Development: A Review for Clinicaland Research Practices. *Journal os Speech, Language and Hearing Research, vol 47*, 173-186. doi:1092-4388/04/4701-0173

CARLA, University of Minnesota. (2013, May 27). *What is Culture: CARLA University of Minnesota.* Retrieved from CARLA: Center for Advanced Research in Language Acquisition, University of Minnesota website: http://carla.umn.edu/culture/definitions.html

Carroll, J. M. (2002). Introduction: Human-Computer Interaction, the past and the present. In J. M. Carroll, & (Ed), *Human-Computer Interaction in the New Millenium.* New York: Addison- Wesley.

Carston, R. (1999). Review: Using Language By Herbert H.Clark. *Journal of Linguistics*, 167-171.

Carter, R. (2004). *Language And Creativity: The Art of Common Talk.* London: Routledge.

Cartmill , E. A., Beilock, S., & Goldin–Meadow, S. (2012). A word in the hand: action, gesture and mental representation in humans and non–human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences , 367*(1585), 129–143.

Channell, J. (1994). *Vague Language.* Oxford: Oxford University Press.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*(2), 257-270.

Chun, D., Ker, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal, 100*(S1), 64-80.

Clark , L. M., Ofemile, A. Y., Adolphs, S., & Rodden, T. (2015). Language and Identity in Human Computer Interaction: A multimodal corpus approach. *British Association of Applied Linguistics (BAAL) Conference.* Aston University, Birmingham, UK.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Clark, L. (2012). *On the Lingua Franca Core and L1 influenced pronunciation variations- a repair based study on the speech of ELF users.* York St John University, Department of English. Unpublished.

Clark, L. M. (2016). *Exploring Vague Language Use in Human-Agent Interaction.* University of Nottingham, UK. Unpublished.

Clark, L. M., Bachour, K., Ofemile, A. Y., Adolphs , S., & Rodden, T. (2014). Potential of imprecision: exploring vague language in agent instructors. *Proceedings of the second international conference on Human-agent interaction (HAI '14)* (pp. 339-344). New York NY, USA,: ACM. doi:http://dx.doi.org/10.1145/2658861.2658895

Clark, L. M., Ofemile, A. Y., & Adolphs, S. (2016b). "But, it’s just a machine” – Examining the social boundaries of computer speech and politeness. *British Association for Applied Linguistics (BAAL) Conference,.* Anglia Ruskin University, Cambridge.

Clark, L. M., Ofemile, A. Y., Adolphs, S., & Tom, R. (2016a). A Multimodal Approach to Assessing User Experiences with Agent Helpers. *ACM Transactions on Interactive Intelligent Systems(TIIS), 6*(4), 1-31.

Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation.* London and New York: Continuum International Publishing Group.

Commeyras, M. (2013). Listening Comprehension Assessment. *https://www.slideshare.net/mcommeyras/listening-comprehension-assessment-how-to-make*. Slideshare.

Cornelius, C. J., Nguyen, M. A., Hayes, C. C., & Makena, R. (2013). Supporting virtual collaboration in spatial design tasks: Are surrogate or natural gestures more effective? *IEEE Transactions on Human-Machine Systems, 43*(1), 92-101.

Coulthard, M., Johnson , A., & Wright, D. (2016). *An introduction to forensic linguistics: Language in evidence* (Second edition ed.). London and New York: Routledge.

Crystal, D. (2002). *The English Language: A guided tour of the language.* London, UK: Penguin .

Crystal, D. (2010). *The Cambridge Encyclopedia of English Language. Second Edition.* Cambridge: Cambridge University Press.

Cummings, M. M. (2014). Man versus machine or man+ machine? . *IEEE Intelligent Systems, 29*(5), 62-69.

Cutting, J. (2007). Introduction to Vague Language Explored. In J. (. Cutting, *Vague Language Explored* (pp. 3-20). Palgrave Macmillan.

Dörnyei , Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies.* Oxford, UK: Oxford University Press.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *A.Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading .* U.S. Department of Education , Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

Dalal , N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005.1.* IEEE.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*(1480), 679-704.

Dautenhahn, K. (2013). Human- Robot Interaction. In M. Soegaard, & R. F. Dam, *The Encyclopedia of Human- Computer Interaction, 2nd Edition* (p. Chapter 38). Aarhus, Denmark: The Interaction Design Foundation.

de Waal, F. B. (1997). The chimpanzee's service economy: food for grooming. *Evolution and Human Behavior, 18*(6), 375-386.

de Waal, F. B., & Lanting, F. (1997). *Bonobo: The Forgotten Ape.* California: University of California Press

Dell'Osso, L. F. (1994). Evidence Suggesting Individual Ocular Motor Control of each eye (Muscle). *Journal of Vestibular Research, 4*(5), 335-345.

Doddington, G. R. (1985). Speaker recognition—Identifying people by their voices. *Proceedings of the IEEE,, 73*, pp. 1651-1664.

Douville, B., Badler, N. I., & Levison, L. (1996). Task-level object grasping for simulated agents. *Presence: Teleoperators & Virtual Environments, 5*(4), 416-430.

Dragoi, V. (2015). Ocular Motor Control Section3: Motor Systems. In J. H. Bryne, *Neuroscience Electronic Textbook* (p. Chapter 8). Retrieved 12 14, 2015, from http://neuroscience.uth.tmc.edu/s3/chapter08.html

Drillis, R. J. (1963). Folk norms and biomechanics. *Human Factors, 5*(5), 427-441.

Easton, K. L., McComish, J. F., & Greenberg, R. (2000). Avoiding common pitfalls in qualitative data collection and transcription. *Qualitative health research, 10*(5), 703-707.

Ekman , P. (1977). Biological and cultural contributions to body and facial movement. 34-84.

Ekman , P. (1997). What We Have Learned by Measuring Facial Behaviour. In P. Ekman, & E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (pp. 469-486). Oxford: OUP.

Ekman , P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of personality and social psychology, 39*(6), 1125-1134.

Ekman, P. (1982a). *Emotion in the Human Face.* New York: Pergamon Press.

Ekman, P. (1982b). Methods for measuring facial action. In K. R. Scherer, & P. Ekman, *Handbookof Methods in Nonverbal Behaviour research* (pp. 45-135). New York: Cambridge University Press.

Ekman, P. (1988). Lying and nonverbal behavior: Theoretical issues and new findings. *Journal of Nonverbal Behavior , 12*(3), 163-175.

Ekman, P. (1992a). An argument for basic emotions. *Cognition & emotion, 6*(3-4), 169-200.

Ekman, P. (1992b). Are there basic emotions? *Psychological Review, 99*(3), 550-553.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In J. M. Larrazabal , & L. A. Pérez Miranda (Ed.), *Language, knowledge, and representation: Proceedings of* *the Sixth International Colloquium on Cognitive Science (ICCS-99)* (pp. 39-50). Netherlands: Springer.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life.* New York, USA: Macmillan.

Ekman, P. (2016). What Scientists who study Emotion Agree About. *Perspectives on Psychological Science, 11*(1), 31-34.

Ekman, P., Wallace, F. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of personality and social psychology, 54*(3), 414.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal behavior, 6*(4), 238-252.

Ekman, P., & Friesen, V. W. (1969). Repertoire of Nonverbal Behaviour: Categories, Origins, Usage, and Coding. *Semiotica, 1*(/1), 49-98.

Ekman, P., & Friesen, V. W. (1969a). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88-106.

Ekman, P., & Friesen, V. W. (1978a). *Facial Action Coding System(FACS):A technique for the measurement of facial action.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, V. W. (1978b). *Facial Action Coding System(FACS):Part Two.* Palo Alto, CA: Consulting Psychologist Press.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face.* Cambridge: Major Books.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology, 30*(1), 527-554.

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). ,.* USA: Oxford University Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in The Face: Guidelines for research and an Integration of findings.* New York: Pergamon Press Ltd.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer , K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology , 38*(2), 270-277.

Ekman, P., Hager, J. C., & Friesen, W. V. (1981). The symmetry of emotional and deliberate facial actions. *Psychophysiology, 18*(2), 101-106.

Ellis, R. (2005). Analysing learner language. In R. Ellis, & G. Barkhuizen, *Analysing learner language* (pp. 73-92). Oxford: Oxford University Press.

Embrey, D., Kontogiannis, T., & Green, M. (1994). *Guidelines for Preventing Human Error in Process Safety* (Vol. 1). New York: Center for Chemical Process Safety.

Engdhal, E., & Vallduvi, E. (1996). Information Packaging in HPSG. In C. Grover, & E. Vallduvi, *Studies In HPSG: Edinburgh Working Papers In Cognitive Science* (Vol. Vol 12, pp. 1 - 32). Edinburgh: Centre For Cognitive Science, University of Edinburgh.

Engen, T., Levy, N., & Schlosberg, H. (1957). A new series of facial expressions. *American Psychologist, 12*(5), 264.

Erikson, E. H. (1959). Identity and the life cycle: Selected papers. *Psychological issues*.

Esposito, A., Esposito, A. M., & Vogel, C. (2015). Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognition Letters, 66*, 41-51.

Esteki, B. (2014). The relationship between Implicit and Explicit Knowledge and Second Language Proficiency. *Theory and Practice in Language Studies, 4*(7), 1520-1525.

Fayad, M., & Schmidt, D. C. (1997). Object-Oreiented Application Frameworks. *Communications of the ACM , 40*(10), 32-38.

Fearon, J. D. (1999). *What is identity (as we now use the word).* Unpublished manuscript, Stanford University, Stanford, California.

Feng , D., & O'Halloran, K. L. (2013). The multimodal representation of emotion in film: Integrating cognitive and semiotic approaches. *Semiotica, 197*, 79 – 100. doi:DOI 10.1515/sem-2013-0082

Field, J. (1998). Skills and Strategies towards a New Methodology for Listening. *ELT Journal*, 110 - 118.

Field, J. (2002). The Changing Face of Listening. In J. C. Richards, & W. A. Renandya, *Methodology in language Teaching: An Anthology of Current Practice* (pp. 242-246). Cambridge: Cambridge University Press.

Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down. *System, 32*, 363-377.

Field, J. (2008). *Listening In the Language Classroom.* Cambridge: Cambridge University Press.

Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930-1955. In J.R. Firth et al., *Studies in Linguistic Analysis. Special volume of the Philological Society.* Oxford: Blackwell.

Fitts, P. M. (1951). *Human Engineering for an Effective Air Navigation and Traffic Control System.* National Research Council. Washington DC: National Research Council.

Flowerdew, J. (1994). *Academic listening: Research perspectives.* Cambridge, NY: Cambridge University Press.

Fontaine, S., & Denis, M. (1999). The Production of Route Instructions in Underground and Urban Environments. In C. Freska, & D. M. Marks (Ed.), *Spatial Information Theory:Cognitive and Computational Foundations of Geographical Information Science* (pp. 83-94). Verlag Berlin: Springer.

Foster, M. E., Bard, E. G., Guhe, M., Hill, R. L., Oberlander, J., & Knoll, A. (2008). The roles of haptic-ostensive referring expressions in cooperative, task-based human-robot dialogue. *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 295-302). ACM.

Fowles, R. (2000). Symmetry in design participation in the built environment: Experiences and insights from education and practice. *In Proceedings of CoDesigning 2000.* London: Springer.

Fraser, B. (2010). Pragmatic competence: The case of hedging. In *New approaches to hedging* (pp. 15-34).

Fredrickson, B. L. (2001). The Role of Positive Emotions in Positive Psychology: The Broaden-and-Build Theory of Positive Emotions. *Am Psychol. Author Manuscript, 56*(3), 218-226.

Frenečik, M. (2005). Organization of Repair in Talk-in-Interaction and Politeness. *Theory and Practice in English Studies 3:Proceedings from the Eighth Conference of British, American and Canadian Studies*, (pp. 1-10). Brno Masarykova univerzita.

Fridlund, A. J. (2014). *Human facial expression: An evolutionary view.* San Diego, London, New York, Boston, Sydney, Tokyo, Toronto: Academic Press.

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, D. I. (2004). Gestures Over Video Streams to Support remote Collaboration on Physical Tasks. *Human-Computer Interaction, 19*, 273-309.

Gaiyan, W. (2007). *Understanding Listening Comprehension process and effectively organising Listening class.* Retrieved /01/ 05, 2010

Garcia , O., & Wei , L. (2014). *Translanguaging: Language, Bilingualism and Education.* Basingstoke: Palgrave Macmillan. doi:10.1057/9781137385765

Garcia, A. C., & Jacobs, J. B. (1999). The Eyes of The Beholder: Understanding the Turn-Taking system in Quasi-Synchronous Computer-Mediated Communication. *Research on Language and Social Interaction, 32*(4), 337-367. doi:10.1207/S15327973rls3204_2

García, C. G., Pérez, P. P., & Martínez, J. N. (2000). Action selection properties in a software simulated agent. *Mexican International Conference on Artificial Intelligence* (pp. 634-648). Berlin Heidelberg: Springer.

Garfinkel, H. (1956). Conditions of successful degradation ceremonies. *American journal of Sociology*, 420-424. Retrieved from http://www.jstor.org/stable/2773484

Geertz, C. (1973). Thick Description: Towards an Interpretive Theory of Culture. In C. Geertz, *The Interpretation of Culture.* New York: Basic Books.

Geertzen, J. (2012). *Inter-Rater Agreement with multiple raters and variables.* Retrieved April 4, 2016, from https://nlp-ml.io/jg/software/ira/

Givens, D. (2016). *The Nonverbal Dictionary of Gestures, signs and body language cues.* Spokane Washington: Centre for Nonverbal Language Press.

Givens, D. B. (2010). *Center for Nonverbal Studies.* (D. B. Givens, Ed.) Retrieved 10 18, 2015, from center-for-nonverbal-studies.org: http://center-for-nonverbal-studies.org/tensemou.htm

Givens, D. B. (2015). Measuring Gestures (Chapter Three) . In D. Chadee, & A. Kostić, *Social Psychology in Nonverbal Communication* (pp. 66-91). New York: Palgrave-Macmillan.

Goetz, J. P., & LeCompte, M. D. (1981). Ethnographic Research and the Problem of Data Reduction. *Anthropology & Education Quarterly, 12*(1), 51-70.

Goffman, E. (1959). *The Presentation of self in Everyday Life.* Harmondsworth: penguin.

Gomes, L., Faria, P., Morais, H., Vale, Z., & Ramos, C. (2014). Gomes, Luis, et al. "Distributed, agent-based intelligent system for demand response program simulation in smart grids. *IEEE Intelligent Systems, 29*(1), 56-65.

Gong, L., & Lai, J. (2001). Shall we mix synthetic speech and human speech? Impact on users' performance, perception and attitude. *CHI Letters 2001, 7*(3), 158-165.

Goodwin, C. (2003). Pointing as situated practice. In S. Kita , *Pointing: Where Language, Culture, and Cognition Meet* (pp. 217-241). Mahwah, NJ: Erlbaum.

Gosselin, P., & Kirouac, G. (1997). Afterword: Component and Recognition of Facial Expression in the communication of Emotion by Actors. In P. &. Ekman, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (pp. 268-286). Oxford: OUP.

Gottman, J., Markman, H., & Notarius, C. (1977). The Topography of Marital conflict: A sequential Analysis of Verbal and Nonverbal behavior. *Journal of Marriage and Family, 39*(3), 461-477.

Granieri, J. P., & Badler, N. I. (1995). *Simulating humans in VR. Virtual Reality and its Applications.* London, UK: Academic Press.

Greiffenhagen, C., & Watson, R. (2009). Visual repairables: analysing the work of repair in human–computer interaction. *Visual Communication, 8*(1), 65-90.

Grice, H. P. (1999). Logic and conversation. *The Discourse Reader*, 76-88.

Grice, H. P. (2006). Logic and Conversation. In A. Jaworski, & N. Coupland, *The Discourse Reader* (Second ed., pp. 66 - 77). London: Routledge.

Grice, P. H. (1975). Logic and Conversation. In P. Cole, & J. L. Morgan, *Syntax and Semantics* (pp. 48-58). New york: Academic Press.

Gross, A. G. (2009). Towards a Theory of Verbal-Visual Interaction: The Example of Lavoisier. *Rhetoric Society Quarterly, 39*(2), 147-169.

Gross, J. J., & Levenson, R. W. (1997). Hiding Feelings: The acute Effects of Inhibting Negative and Positive Emotion. *Journal of Abnormal Psychology, 106*(1), 95-103.

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin, Y. S. Lincoln, & (eds), *Handbook of Qualitative Research.* Thousand Oaks, California: Sage.

Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A study of Learners of French and Swedish.* Univesity of Lund Sweden, Department of Linguistics and Phonetics. Lund: Unpublished.

Gumperz, J. J. (1999). On Interactional Sociolinguistic method. In S. Saranghi, & C. Roberts, *Talk, Work and Institutional order. Discourse in Medical, Mediation and Management Settings* (pp. 453-471). Berlin: Mouton de Gruyter.

Gumperz, J. J. (2009). The speech community. *Linguistic anthropology: A reader, 1*(66), 66-73.

Gunes, H., & Hung, H. (2016). Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block . *Image and Vision Computing, 55*(1), 6-8.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*, 29-48.

Gwet, K. L. (2012). *Handbook of Inter-Rater reliability: The definitive guide to Measure the Extent of Agreement Among raters* (Third Edition ed.). Gaithersburg MD: Advanced Analytics, LLC.

Hadjikhani, N., Hoge, R., Snyder, J., & de Gelder, B. (2008). Pointing with the eyes: the role of gaze in communicating danger. *Brain and cognition, 68*(1), 1-8.

Harmer, J. (2007). *The Practice of English Language Teaching. Fourth Edition.* London: Pearson Longman.

Harrison, S. (2014). Methods of Gesture Analysis: Describing gesture forms and Interpreting their meanings. *Intercampus Mobility Grant, UNUK* (pp. 1-42). Nottingham: University of Nottingham, UK.

Hartford, S., Carr , L., & Pope, C. (2010). A Manifesto for Web Science? In M. Rowe , & F. Ciravegna (Ed.), *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, (pp. 1-6). North Carolina: Raleigh.

Hartley, P. (2002). *Interpersonal Communication.* London and New York: Routeledge.

Hayashi, M., Raymond, G., & Sidnell, J. (2013). Conversationa Repair and Human Understanding. In M. Hayashi, G. Raymond, & J. Sidnell, *Conversationa Repair and Human Understanding: an introduction* (pp. 1 - 40). Cambridge: Cambridge University Press.

Hedge, A. (1998). Design of hand-operated devices. In N. A. Stanton, *Human factors in consumer products* (pp. 203-222). Boca Raton: CRC Press, Taylor & Francis Group.

Hendrikse, A. P., Nomdebevana, N., & Allwood, J. (2016). An exploration of the nature, functions and subcategories of the discourse functional category, Interactive, in spoken Xhosa. *South African Journal of African Languages, 36*(1), 93-101.

Henley, C. N. (2010, December 06). *Comment: Motion of the fingers, thumb and wrist-Language of hand and arm surgery series*. Retrieved from Noel Henley website: http://noelhenley.com/532/motion-of-the-fingers-thumb-and-wrist-language-of-hand-and-arm-surgery-series/comment-page-1/#comment-7037

Herring, S. C. (1990). Information structure as a consequence of word order type. *Annual Meeting of the Berkeley Linguistics Society (Vol. 16, No. 1, pp.16 (1)*, pp. 163-174. Berkerely: Berkeley Linguistics Society.

Hess, U., & Kleck, R. E. (1997). Differentiating Emotion Elicited and Deliberate Emotional Facial Expression. In P. Ekman, E. Rosenberg, & (Eds), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (pp. 271-286). Oxford: OUP.

Hess, u., Adams Jr, R. B., & Kleck, R. E. (2009). The face is not an empty canvas: how facial expressions interact with facial appearance. *Philosophical transactions of the Royal Society, 364*, 3497 -3504. doi:10.1098/rstb.2009.0165

Hesse, S., Schulte-Tigges, G., Konrad, M., Bardeleben, A., & Werner, C. (2003). Robot-assisted arm trainer for the passive and active practice of bilateral forearm and wrist movements in hemiparetic subjects. *Archives of physical medicine and rehabilitation, 84*(6), 915-920.

Hilpert, M. (Director). (2013, December 20). *Information Structure* [Motion Picture]. Retrieved from https://www.youtube.com/watch?v=PJecXZp_SYw

Hilpert, M. (2014). *Construction Grammar and its Application to English.* Edinburgh: Edinburgh University Press. Retrieved from https://www.youtube.com/watch?v=QzCZJL5Wk1k

Holler, J., & Wilkin, K. (2011). Co-Speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behaviour, 35*, 133-153.

Hollway, W. (2009). Identity Change and Identification. In S. Bromley, J. Clarke, S. Hitchcliffe, S. Taylor, & (eds), *Exploring Social Lives* (pp. 247-290). Milton Keynes: The Open University.

Hymes, D. (1972a). Models of the Interaction of Language and Social Life. In J. J. Grumperz, D. Hymes, & (eds), *Directions in sociolinguistics: The Ethnography of Communication* (pp. 35-71). New York: Holts Rinehart & Winston.

Hymes, D. (1972b). On communicative competence. In J. B. Pride, & J. Holmes, *Sociolinguistics* (pp. 269-293.). Baltimore, USA: Penguin Education, Penguin Books Ltd.

Hymes, D. (1974). *Foundations in sociolinguistics: An Ethnographic Approach.* Philadelphia: University of Philadelphia Press.

Hymes, D. (1975). Breakthrough into performance. In D. Ben-Amos, K. S. Goldstein, & (eds), *Folklore: Performance and Communication* (pp. 11-74). The Hague: Mouton.

Hymes, D. (1992). The concept of communicative competence revisited. In M. Pütz, *Thirty Years of Linguistic Evolution: Studies in honour of René Dirven on the occasion of his 60th birthday* (pp. 31-58). Amsterdam: John Benjamin BV.

Iyiola, H. I. (2010). *Success In Oral English For Schools and Higher Institutions* (Third Edition ed.). Kaduna: Zenith Prints.

Jakobson, R. (1960). Linguistics and poetics. *Style in language*, 350-377.

Jaworski, A., & Coupland, N. (2006). *The Discourse Reader* (Second Edition ed.). London and New York.

Jennings, N. (2012). *Orchid End Of Year Report, 2012.* Southampton: www.orchid.ac.uk. Retrieved 07 02, 2013, from www.orchid.ac.uk

Jennings, N. R., Moreau, L., Nicholson, D., Ramchurn , S., Roberts, S., Rodden , T., & Rogers, A. (2014). Human-agent collectives. *Communications of the ACM, 57*(12), 80-88.

Jewitt, C. (2012). *An introduction to using video for research.* London: National Centre for Research Methods Working Paper 03/12.

Jewitt, C. (2013). Multimodal methods for researching digital technologies. In S. Price, C. Jewitt, & B. (. Brown, *SAGE Handbook of Digital Technology Research* (p. Chapter 17). SAGE.

Jewitt, C. (2014). An Introduction to Multimodality. In C. Jewitt, *The Routledge Handbook of Multimodal Analysis* (2nd ed.). London: Routledge.

Jiang, J., Awadallah, A. H., Jones, R., Ozertem, U., Zitouni, I., Kulkarni, R. G., & Khan, O. Z. (2015). Automatic Online Evaluation of Intelligent Assistants. *Proceedings of the 24th International Conference on World Wide Web.* Florence, Italy: ACM.

Johnson, D., & Gardner, J. (2005). Effects of team-based computer interaction: the media equation and game design considerations. In F. Kishino, Y. Kitamura, H. Kato, & N. Nagata (Ed.), *Entertainment Computing ICEC 2005* (pp. 468-479). Berlin Heidelberg: Springer.

Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences, 4*(1), 39-103.

Kaun, A. &. (2013). *Research report: Mediatization of culture and everyday life.* Riksbanken Jubileumsfond: Commissioned by the sector committee Mediatization of culture and everyday life of the Riksbanken Jubileumsfond. Retrieved /04/ 16, 2014, from http://www.rj.se/Global

Kaun, A., & Fast, K. (2014). *Research report. Mediatization of culture and everyday life.* Södertörn University . Huddinge: sh.se/publications. Retrieved /01/ 24, 2014, from http://www.diva-portal.org/smash/get/diva2:698718/FULLTEXT02

Kendon, A. (1980). Gesture and Speech:two aspects of the process of utterance. In M. R. Key, & M. R. Key (Ed.), *Nonverbal Communicationand Language* (pp. 207-227). The Hague: Mouton.

Kendon, A. (1988). *Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives.* Cambridge: Cambridge University Press.

Kendon, A. (1990). *Conducting Interaction: Patterns of behaviour in focused Encounters.* Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Kersta, L. G. (1978 (1962)). Voiceprint-Identification Infallibility. *The Journal of the Acoustical Society of America , 34*. doi: 10.1121/1.1936995

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System, 41*(3), 829–840. doi:http://dx.doi.org/10.1016/j.system.2013.08.005

Kipp, M. (2003). *Gesture Generation by Imitation: From Human Behaviour to Computer Character.* der Universit□at des Saarlandes, der Naturwissenschaftlich-Technischen Fakult□aten. Unpublished.

Kipp, M., Neff, M., & Albretch, I. (2007). An Annotation Scheme for conversational Gestures: How to Economically capture timing and form. *Language Resources and Evaluation , 41*(3-4), 325-339.

Kirk, D., Crabtree, A., & Rodden, T. (2005). Ways of the Hands. *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work* (pp. 1-22). Netherlands: Springer.

Kiseleva , J., Williams, K., Jiang, J., Awadallah, A. H., Cook, A. C., Zitouni, I., . . . Anastasakos, T. (2016). Understanding User Satisfaction with Intelligent Assistants. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieva.*

Kita , S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production . *Language and cognitive processes, 22* (8), 1212-1236.

Kita, S. (2013). Architectural issue in the model of Speech-gesture production: Gesture, Action language.

Kita, S., Van Gijn, I., & Van der Hulst, H. (1998). Movement Phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Ed.), *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop Bielefeld, Germanyseptember 17-19, 1997 Proceedings. .1371*, pp. 23-35. Berlin Heidelberg: Springer.

Klein, J., Moon, Y., & Picard, R. (2002). This Computer Responds to User Frustration: Theory, Design, Results, and Implications. *Interacting With Computers, 14*(2), 119-140.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2014). *Nonverbal Communication in Human Interaction* (Eighth ed.). Boston MA, USA: Wadsworth Cengage Learning.

Knight, D. (2009). *A multi-modal corpus approach to the analysis of backchanneling behaviour.* University of Nottingham, School of English. Unpublished PhD Thesis.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development.* New Jersey: FT press.

Komatsu, M., Arai, T., & Sugawara, T. (2004). Perceptual Discrimination of Prosodic Types. *Speech Prosody*, 725-728.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition.* Oxford: Pergamon.

Kress, G. (2011). Partnerships in research': multimodality and ethnography. *Qualitative Research, 11*(3), 239-260.

Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd Edition ed.). London and New York: Routledge.

Krifka, M. (2008). Basic Notions of Information Structure. *Acta Linguistica Hungarica, 55*(3), 243-276. Retrieved /05/ 13, 2014

Krippendorff, K. (2004). Reliability in content analysis: Some Common Misconceptions and Recommendations. *Human Communication Research, 30*(3), 411-433. doi:http://dx.doi.org/10.1111/j.1468-2958.2004.tb00738.

Laban, R. (1975). *Laban's Principles of Dance and Movement Notation (First published 1956).* (R. Lange, Ed.) London: MacDonald and Evans.

Lane, R., Stanton, N. A., & Harris, D. (2006). Applying hierarchical task analysis to medication administration errors. *Applied ergonomics, 37*(5), 669-679.

Latour, B. (2014). subjects of Distance:(Country) Girl Culture Capital. In C. Driscoll, *The Australian Country girl: History, Image, Experience* (pp. 113-136). Routledge.

Laundry, J. S. (2009). Human Computer Interaction in Aerospace. In A. Sears, & A. J. Jacko, *Human-Computer Interacction: Designing for Diverse Users and Domains (Second Edition)* (pp. 198 - 213). New York: CRC Press.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with NEUROGES-ELAN system. *Behaviour Research Methods, Instruments, & Computers, 42*(3), 841-849.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of Reliability and Validity in Ethnographic Research. *Review of Educational research, 52*(1), 31-60.

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender?: an experimental test of gender stereotype. *In CHI'00 extended abstracts on Human factors in computing systems* (pp. 289-290). ACM.

Lee, H., Choi, Y. S., Lee, S., & Park, I. P. (2012). Towards unobtrusive emotion recognition for affective social communication. *Consumer Communications and Networking Conference (CCNC), 2012* (pp. 260-264). IEEE.

Lee, K., & Nass, C. (2003). Designing social presence of social actors in human computer interaction. *CHI Letters*, 289-296.

Leech, G. (2013). Introducing Corpus Annotation. In R. Garside, G. Leech, T. McEnery, & (Eds), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 1-18). London and New York: Routledge.

Lemke, J. (1990). *Talking Science: Language, Learning and Values.* Norwood: Ablex.

Lemke, J. (1997). Cognition, context and learning: A social semiotic perspective. In D. Kirshner, & A. (. Whitson, *Situated cognition theory: Social, neurological and semiotic perspectives* (pp. 37-55). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Levinson, S. C. (2016). Turn-taking in human Communication-Origins and Implications for language Processing. *Trends in Cognitive Sciences*, 6-14.

Lewis, M., Dauphin, Y. N., Parikh, D., & Batra, D. (2017). Deal or No Deal? End-to-End Learning for Negotiation Dialogues. *Facebook Github*.

Lin, H., Guo, F., Wang, F., & Jia, Y. (2015). Picking up a soft 3D object by "feeling" the grip. *The International Journal of Robotics Research, 34*(11), 1361-1384.

Lincoln, Y. S., & Guba, E. (1985). *Naturalistic Inquiry.* Beverly Hills, California: sage.

Liska, A. E. (1974). The Impact of Attitude on Behaviour: Attitude-Social Support Interaction. *The Pacific Sociological review, 17*(1), 83-97. Retrieved 03 18, 2017, from www.jostor.org/stable/1388599

Litman, D. J., & Forbes-Riley, K. (2006). Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. *Speech communication, 48*(5), 559-590.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6-14.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication. *Human Communication Research*, 587-604.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2010). *Practical Resources fo Assessing and Reporting Intercoder Reliability in Content Analysis research Projects.* Retrieved 01 22, 2016, from Matthew Lombard website: http://matthewlombard.com/reliability/#What is intercoder reliability

Luger, E., & Sellen, A. (2016). “Like Having a Really bad PA”: The Gulf The Gulf The Gulf between User Expectations and Experience of Conversational Agents. *CHI 2016* (pp. 5286-5297). San Jose, CA USA: ACM.

Maatman, R. J. (2005). Natural behavior of a listening agent. In J. G. T. Panayiotopoulos (Ed.), *Intelligent Virtual Agents, 5th International Working Conference, Kos, Greece.* Berlin Heideberg: Springer- verlay.

Maclay , H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1), 19-44.

Mai, X. Q., Luo, J., Wu, J. H., & Luo, Y. J. (2004). Aha!" effects in a guessing riddle task: An event-related potential study. *Human brain mapping, 22*(4), 261-270.

Malik, A. A. (2010). *The secretsof body language.* Retrieved from http://www.slideshare.net/arifanees/secrets-of-body-language: http://www.slideshare.net/arifanees/secrets-of-body-language

Martin, G., & Pear, J. J. (2016). *Behavior modification: What it is and how to do it.* New York: Routledge.

Matsumoto, D. (2012, March 21). *How to tell a lie with the naked eye.* Retrieved from Humintell website: http://www.humintell.com/2012/10/how-to-tell-a-lie-with-the-naked-eye/

Matsumoto, D., & Hwang, H. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion, 35*(2), 181-191.

Matsumoto, D., & Hwang, H. C. (2013). Cultural Similarities and Differences in Emblematic Gestures. *Journal of Nonverbal Behaviour, 37*, 1–27. doi:0.1007/s10919-012-0143-8

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of personality and social psychology, 96*(1), 1-10.

May, C., & Finch, T. (2009). Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology, 43*(3), 535-554.

Mazzone, M. (2015). Constructing the context through goals and schemata: top-down processes in comprehension and beyond. *Frontiers in psychology, 6*, 1-13.

McAfee, A. (2011, /11/ 28). The Machine Are coming: Full Speed Ahead. (S. Paikin, Interviewer) Retrieved from https://www.youtube.com/watch?v=-0g8DDsv1MM

McGill, P., & Emerson, E. (1992). Normalisation and Applied Behaviour Analysis: Values and Technology in Human Behaviour. In H. Brown, & H. Smith, *Normalisation: A Reader* (p. Chapter 5). Abingdon Oxon: Routeledge.

McKenna, M. C., & Stahl, K. D. (2015). *Assessment for reading instruction.* New York: Guilford Publications.

McNeil, D. (1979). *The Conceptual Basis of Language.* New Jersey, USA: Erlbaum/John Wiley & Sons Inc.

McNeil, D. (1985). So you Think Gestures Are Nonverbal? *Psychological Review, 92*(3), 350-371.

McNeil, D. (1992). *Hand and Mind: What Gestures Reveal About Thought.* Chicago: The University of Chicago Press.

McNeil, D. (2005). *Gesture and Thought.* Chicago, Illinois, USA: University of Chicago Press.

McNeil, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse processes, 16*, 363-386.

McNeill, D. (1989). A straight path—to where? Reply to Butterworth and Hadar. *Psychological review, 96*(1), 175-179.

McNeill, D. (2006). Gesture: a psycholinguistic approach. *The encyclopedia of language and linguistics*, 58-66.

Meadors, J. D., & Murray, C. B. (2014). Measuring nonverbal bias through body language responses to stereotypes. *Journal of Nonverbal Behavior, 38*(2), 209-229.

Mehrabian, A. (1967). Attitudes Inferred from verbal neutral communications. *Journal of Consulting Psychology, 31*, 414-417.

Mehrabian, A. (2007). *Nonverbal Communication.* New Brunswick(USA) and London (UK): Aldine Transaction.

Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology, 31*(3), 248.

Merriam-Webster Dictionary. (2012). *Definition of AHA MOMENT.* Retrieved May 05, 2015, from Merriam-Webster Dictionary: https://www.merriam-webster.com/dictionary/aha%20moment

Microsoft, C. (2017).

Miller, J. (Director). (2014). *Top Documentary Films - The Secrets of Body Language* [Motion Picture]. Retrieved from https://www.youtube.com/watch?v=HTb7XQGKp-g

Miller , J., & Schwanenflugel, P. J. (2008). A Longitudinal Study of the Development of Reading Prosody as a Dimension of Oral Reading Fluency in Early Elementary School Children. *Reading Research Quarterly, 43*, 336-354.

Mitchell, T. F. (1957). The language of buying and selling inCyrenaica: as situational statement. *In Hesperis,, 44*, 31-71.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz , M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2*(1), 10-12.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language, 66*, 249-264.

Monroy-Hernández, A. A. (2011). *Operation waffles: expressive tagging on Scratch.* Massachusetts: Microsoft Research. Retrieved /04/ 11, 2014, from http://research.microsoft.com/

Montazeri, M., Hamidi, H., & Hamidi, B. (2015). A Closer Look at Different Aspects of Private Speech in SLA. *Theory and Practice in Language Studies, 5*(3), 478-484.

Moon , A., Panton, B., Van Der Loos, H. F., & Croft, E. A. (2010). Using hesitation gestures for safe and ethical human-robot interaction. *Proceedings of the ICRA* (pp. 11-13)). ICRA.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities?: Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651-674.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Hman-Computer Studies, 49*, 79-94.

Mooney, J. (2012). *Farlex Partner Medical Dictionary/Somatic.* Retrieved june 26, 2017, from Somatic: http://medical-dictionary.thefreedictionary.com/somatic

Moos, D. C., & Marroquin, E. (2010). Multimedia, hypermedia, and hypertext: Motivation considered and reconsidered. *Computers in Human Behavior, 26*(3), 265-276.

Morris, D. (2015). *Bodytalk: A world guide to gestures.* Random House.

Mozdzenski, L. (2013). Verbal-Visual Intertextuality; How do Multisemiotic Texts Dialogue? *Bakhtiniana: Revista de Estudos do Discurso, 8*(2), 176-200.

Muller, M. (2009). *Participation in Web2.0 and e-Participation in Government: Toward a "Third Space" for Deliberation and Government.* www.primodpdf.com. Retrieved from www.mullerkeynoteslidesharefinal

Muller, M. J. (2003). Participatory Design:The third space in HCI. In J. A. Jacko, & A. Sears, *The human-computer interaction handbook: fundamentals, evolving technologies and emmerging applications* (pp. 1051 - 1068). Hillsdale NJ USA: L.Erlbaum Associates Inc.

Murdoch , J., Salter, C., Cross, J., & Poland, F. (2013). Misunderstandings, communicative expectations and resources in illness narratives: Insights from beyond interview transcripts. *Communication & medicine, 10*(2), 153.

Mutlu, B. (2011, fall). *Human-Computer Interaction: Step-By-Step Experimental Design.* Retrieved January 21, 2014, from University of Wisconsin: http://hci.cs.wisc.edu/co.../HCI-Week06-Lecture07.pdf

Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *CHI 2000.* Amsterdam: ACM.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology:Applied, 7*(3), 171-181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5), 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender Stereotypic Responses to Computers with voices. *Journal of Applied Social Psychology, 27*, 864-876.

Nass, C., Reeves, B., & Leshner, G. (1996). Technology and roles: A tale of two TVs. *Journal of communication, 46*(2), 121-128.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors. *CHI Conference* (pp. 72- 78). Boston, Massachusette: ACM.

Neto, F. (2009). Implicit and explicit attitudes among students. *Universitas Psychologica, 8*(3).

Nieto, S. (2013). Language, Literacy, and Culture: Aha! Moments in Personal and Sociopolitical Understanding. *Journal of Language and Literacy Education, 9*(1), 8-20.

Ning, H., Han, T. X., Hu, Y., Zhang, Z., Fu, Y., & Huang, T. S. (2006). A realtime shrug detector. In Automatic Face and Gesture Recognition. *FGR 2006. 7th International Conference* (pp. 505-510). IEEE

Norman, D. (1988). *The Psychology of Everyday Things.* New York: Basic Books.

Nunan, D. (2002). Listening in language learning. In J. C. Richards, , & W. A. Renandya, *Methodology in language teaching: An anthology of current practice* (pp. 238-241). Cambridge, New York: Cambridge University Press.

Ofemile , A. Y. (2015a). A framework for Analysing Discourse between Humans and Software Agents. *MAUTECH Journal of Language and Communication (JLC), 1*(1).

Ofemile, A. Y. (2016). Human-Agent Interaction: When L1 Agents instruct L2 Speakers of English in hybrid contexts. *BAAL Languages in Africa Special Interest Group Annual Meeting (BAAL-LiA SIG, 2016).* University of East London, UK,.

Ofemile, A. Y., Clark, L., Adolphs, S., & Rodden, T. (2016). A Multimodal Assessment of Facial Actions as Listener perceptions in Human-Computer Interaction. *The Inter-varietal Corpus Studies (IVACS) Conference.* Bath Spa University, UK,.

O'Halloran, K. L. (2009). Multimodal analysis and digital technology. *Interdisciplinary Perspectives on Multimodality: Theory and Practice, Proceedings of the Third International Conference on Multimodality.* Campobasso: Palladino.

O'Halloran, K. L., & Smith, B. A. (2012). Multimodal text analysis. In C. A. Chapelle, *The Encyclopedia of Applied Linguistics.* New Jersey: Blackwell Publishing Ltd.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus To Classroom: Language Use and Language Teaching.* Cambridge: Cambridge University Press.

Onwuegbuzie, A. J., & Leech, N. L. (2007). A call for qualitative power analyses. *Quality & Quantity, 41*(1), 105-121.

Oreström, B. (1983). *Turn-taking in English Conversation.* Lund, Sweden: LiberFörlag Ltd.

Osada, N. (2004). Listening Comprehension research: A Brief Review of the Past Thirty Years. *Dialogue, 3*, 53-66.

O'sullivan, M., Ekman , P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit . *Journal of nonverbal behavior, 12*(3), 203-215.

Oxford, R. (1993). Research Update on Teaching L2 Listening. *Systems Journal*, 205 - 211. Retrieved /01/ 01, 2010

Papageorgiou, C. P., Oren, M., & Poggio, T. (1998). A general framework for object detection. *sixth international conference on Computer vision, 1998.* IEEE.

Paulston, C. B. (1971). The sequencing of Structural Pattern Drills. *FifthAnnual TESOL Conventions.* New Orleans.

Paulus, P. B., Kohn, N. W., & Aditti, L. E. (2011). Effects of Quantity and Quality Instructions on Brainstorming. *Journal of Creative behaviour, 45*(1), 38- 46.

Pease, A., & Pease, B. (2004). *The Definitive Book of Body Language: How to read other's thoughts by their gestures.* Buderim, Australia: Pease International.

Petrou, P., Bakker , A. B., & den Heuvel, M. (2016). Weekly job crafting and leisure crafting: Implications for meaning-making and work engagement. *Journal of Occupational and Organizational Psychology*, 1-24.

Pfister, T., Li, X., Zhao, G., & Pietikäinen, M. (2011). Recognising Spontaneous Facial Micro-expressions Pfister. *2011 International Conference on Computer Vision* (pp. 1449-1456). IEEE.

Picard, W. (2015). Resilient and robust human-agent collectives: a network perspective. *Working Conference on Virtual Enterprises* (pp. 79-87). Springer International Publishing.

Pine, K. J., Gurney , D. J., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior, 34*, 169 –178.

Poole , D., & Mackworth , A. (2010). *Artificial Intelligence: Foundations of Computational Agents.* Cambridge University Press.

Porter, S., Ten Brinke , L., & Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity.". *Journal of Nonverbal Behavior, 36*(1), 23-37.

Potts, A. (2016). Women who Kill: Methods to explore the doubly deviant identity in English sentencing remarks. *Language and Identity in law and evidence.* Nottingham, UK: BAAL/Routledge Research Development Workshop, 2016.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. (1989). *Hamburg Notation System for Sign Languages: An Introductory Guide.* Hamburg: International Studies on Sign Language and the Communication of the Deaf 5.

Prodromou, L. (2005). *'You see, it's sort of tricky for the L2-user' the puzzle of idiomaticity in English as a Lingua Franca.* University of Nottingham, UK. Nottingham, UK: Unpublished.

Raban, J. (1991). *Hunting Mister Heartbreak.* London: Pan.

Rafaeli, A., & Sutton, R. I. (1987). Expression of Emotion as Part of the Work Role. *Academy of Management Review,*, 23-37.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*(4), 226-231.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of research. *Psychological Bulletin, 124*(3), 372 - 422.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Linguistics, 33*, 829-856.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places.* New York: CSLI Publications and Cambridge university press.

Reeves, B., & Nass, C. (2014, January 29). The Media Equation. (E. Griffin, Interviewer) https://www.youtube.com/watch?v=26BclMJQUwo.

Ribes, E., & Rodriguez, M. E. (2001). Correspondence Between Instructions, Performance, And Self-Descriptions in A Conditional Discrimination Task: The Effects of Feedback and Type of Matching Response. *The Psychological Record, 51*, 309-333.

Rickenberg, R., & Reeves, B. (2000). The effects of animated characters on anxiety, task performanc, and evaluations of user interfaces. *CHI 2000.* Amsterdam.

Rittle-Johnson , B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587-597.

Robinson, D. A. (1964). The mechanics of human saccadic eye movement. *The Journal of physiology, 174*(2), 245.

Rutherford, J. (1990). The Third Space: Interview with Homi Bhabha. In Ders, *Identity: Community, Culture, Difference* (pp. 207-221.). London: Lawrence and Wishart. Retrieved 02 05, 2016, from http://www.wsp-kultur.uni-bremen.de/summerschool/download%20ss%202006/The%20Third%20Space.%20Interview%20with%20Homi%20Bhabha.pdf

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696-735.

Schegloff, E. A. (1997). Whose Text? Whose context? *Discourse & society, 8*(2), 165-187.

Schegloff, E. A., Jefferson, G., & Harvey, S. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language 361, 53*(2). doi:http://dx.doi.org/10.2307/413107

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social science information, 44*(4), 695-729.

Sekine, K., & Kita, S. (2015). The parallel development of the form and meaning of two-handed gestures and linguistic information packaging within a clause in narrative. *Open Linguistics, 1*(1), 490-502.

Sekine, K., & Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychologica, 179*, 89-95.

Sekine, K., Sowden, H., & Kita , S. (2015). The Development of the Ability to Semantically Integrate Information in Speech and Iconic Gesture in Comprehension. *Cognitive Science, 39*(8), 1855-1880.

Shea, C. T., Davisson, E. K., & Fitzsimons, G. M. (2013). Riding Others' Coattails: Low self-control individuals value self-control in others. *Psychological Science, 24*(6), 1031-1036.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: how people behave differently when interacting with computers and people. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281-288). ACM.

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the 'Aha!' experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology, 107*(2), 281-298.

Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and clinical Psychology*, 13-39.

Shipley, S. D. (2010). Listening: A concept analysis. *Nursing forum, 45*(2), 125-134.

Shriberg, E. (2001). Toerrrr'is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association, 31*(1), 153-164.

Sindoni, M. G. (2013). *Spoken and Written Discourse in Online Interactions. A Multimodal Approach.* New York and London: Routledge Studies in Multimodality.

Sobotta. (n.d.). *Atlas and Text-book of Human Anatomy 1909.* Retrieved November 12, 2015, from https://en.wikipedia.org/wiki/Corrugator_supercilii_muscle: https://en.wikipedia.org/wiki/Corrugator_supercilii_muscle

Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces . *International Conference on Spatial Information Theory* (pp. 37-50). Berlin, Heidelberg.: Springer.

Stack, L. (2012, July). *Repeating Instructions: A Three-Step Process to Avoid Unproductive Communication.* Retrieved (Downloaded 15/02/2015), from http://www.booher.com/booherbanter/repeating-instructions-a-three-step-process-to-avoid-unproductive-communication/.

Standring, S., & Borley, N. R. (2008). *Gray's anatomy : the anatomical basis of clinical practice (40th ed.)* (40th Edition ed.). London: Churchill Livingstone.

Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied ergonomics, 37*(1), 55-79.

Strait, M., Canning , C., & Scheutz , M. (2014). Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 479-486). ACM.

Swales, J. (1990). The Concept of Discourse Community. In *Genre Analysis: English in Academic and Research Settings* (pp. 21-32). Boston: Cambridge University Press.

Tannen, D. (2006). Intertextuality in interaction: Reframing family arguments in public and private. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies, 26*(4-5), 597-617.

Tannen, D., & Wallat, C. (1987). Interactive Frames and Knowledge Schemas in Interaction: Examples from a medical Examination/Interview. *Social Psychology Quarterly Special Issue: Language and Social Interaction, 50*(2), 205 -216.

Taylor, S. (2009). Who do we think we are? Identities in everyday life. In S. Taylor, S. Hinchliffe, J. Clarke, & S. Bromley, *Introducing the Social Sciences: Making Social Lives* (pp. 165-202). Milton Keynes, UK: The Open University.

Thabane , L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., . . . Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC medical research methodology, 10*(1), 1-10.

Tipples, J., Atkinson, A. P., & Young, A. W. (2002). The eyebrow frown: a salient social signal. *Emotion, 2*(3), 288.

Tope, D., Pickett, J. T., Cobb, R. Y., & Dirlam, J. (2014). Othering Obama: Racial Attitudes and Dubious Beliefs about the Nation's First Black President. *Tope, Daniel, et al. "Othering Obama: Racial Attitudes and Dub Sociological Perspectives, 57*(4), 450-469. doi:10.1177/0731121414536140

Topolinski, S., & Reber, R. (2010). Gaining insight into the "Aha" experience. *Current Directions in Psychological Science, 19*(6), 402-405.

Torrey, C., Fussell, S. R., & Kiesler, S. (2013). How a robot should give advice . *Proceedings of the 8th International Conference on Human-Robot Interaction*, (pp. 275–282).

Trandis, H. C., & Lambert, W. W. (1958). A restatement and test of Schlosberg's theory of emotion with two kinds of subjects from Greece. . *The Journal of Abnormal and Social Psychology, 56*(3), 321.

Tsuchiya, K. (2013). *Listenership Behaviours in Intercultural Encounters.* Amsterdam, The Netherlands; Philadelphia, USA: The Benjamins Publishing Co.

Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432-451.

Vallduvi, E. &. (1995). Focus and Kontrast: Finnish and Catalan. *Being a Paper presented at the 18th GLOW Colloquium, University of Tromsoe*. Tromsoe.

Vallduvi, E., & Engdhal, E. (1996). The Linguistic Realisation of Information Packaging. *Linguistics, 34*, 459 - 519. doi:0024-3949/96/0034-0459

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Vandergrift, L. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *The Modern Language Journal, 81*, 494-505.

Vandergrift, L. (2004). 1. Listening to Learn or Learning to Listen? *Annual Review of Applied Linguistics, 24*, 3-25.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and Speech in interaction: An Overview. *Speech Communication, 57*, 209-232.

Wakefield, R., & Wakefield, K. (2016). Social media network behaviour: A study of user passion affect. *Journal of strategic Information System, 25*(2), 140-156.

West, R. L., & Turner, L. H. (2013). *Introducing communication theory: Analysis and application.* McGraw-Hill Humanities/Social Sciences/Languages.

White, S. (Mar, 1989). Backchannels across cultures: A study of Americans and Japanese. *Language in society, 18(1)*, 59-76. Retrieved from http://www.jstor.org/stable/4168001

Williams, M., Mercer, S., & Ryan, S. (2016). *Exploring psychology in language learning and teaching.* Oxford: Oxford University Press.

Wimmer, R., & Boring, S. (2009). HandSense: discriminating different ways of grasping and holding a tangible user interface. *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (pp. 359-362). ACM.

Wolvin, A. D. (2009). Understanding and misunderstanding. In W. F. Eadie, *21st Century Communication* (pp. 137-146). Thousand Oaks CA: Sage Publications.

Wolvin, A. D. (2010a). *Listening: Political and Civil Leadership.* Sage Publishers.

Wolvin, A., & Coakley, C. G. (1996). *Listening.* Dubuque IA: Wm C. Brown Publishers.

Wong, H. R., & San Hu, J. (2011). Why do people hesitate? Perceived risk in workplace spirituality. *International Journal of Business and Management, 6*(11), 57.

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *MBC Medical Research Methodology, 13*(1), 61- 67.

Wooldrige, M. (2009). *An Introduction to Multiagent Sysytems Second Edition.* John Wiilley and Sons Ltd.

Wronski, C., & Daum, M. M. (2014). Spatial orienting following dynamic cues in infancy: grasping hands versus inanimate objects. *Developmental Psychology, 50*(8), 2020 - 2029.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language processed by a common neural system. *Proceedings of the National Academy of Sciences USA*, *106*, pp. 20664-20669. doi:10.1073/pnas.0909197106

Yan, W.-j. J., Wu, Q., Chen, Y.-H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior, 37*(4), 217-230.

Zelkowitz, M. V., & Wallace, D. R. (1998). Experimental models for validating technology. *Computer, 31*(5), 23-31.

Zillner, T. (2000). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Information Technology and Libraries, 19*(1), 54-54.

Zimmerman, D. H. (1998). Identity, Context and Interaction. In C. Antaki, & S. (. Widdicombe, *Identities in talk* (pp. 87-106). London: Sage.

Zwitserlood, I., Ozyurek, A., & Perniss, P. M. (2008). Annotation of sign and gesture cross-linguistically . *6th International Conference on Language Resources and Evaluation (LREC 2008)/3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora (ELDA)*, (pp. 185-190).

Guide to Appendices

Appendices

Appendix I: Lego Assembly Tasks - Study Information Sheet

We are a team of researchers from the Mixed Reality Lab and the School of English, University of Nottingham working on an EPSRC funded project called ORCHID. An area of research we are particularly interested in is the interaction between people and computers. One of the aims of ORCHID is to explore the different ways in which people react to this automated technology that is able to act on their behalf, and also present them with information of interest/benefit.

As part of our research, we have developed a group of tasks in which you will be expected to construct two Lego models by following verbal instructions received from a computer, henceforth referred to as the *instructor*. During each task, you will be reliant on the information received by the instructor as no visual cues will be provided. These tasks will be followed by a short interview. The total length of the study will not exceed 60 minutes in length and you will receive the compensation of a £10 Amazon voucher for your participation.

All of the data we collect will be held in a secure and safe manner in accordance with the Data Protection Act 1998. Access to this data will be restricted to researchers involved in the project, and will be processed confidentially without ever linking it to your name or identity. It is within your rights to refuse the collection of any of the specific types of data identified above.

The results from the study will be used for publication in academic conferences and journals. You are free to withdraw at any point during, or after, the study and any data collected will be erased from our records. To withdraw, simply inform the researcher during the study or use the following details to contact us with any queries you might have:

Leigh Clark
Mixed Reality Lab (MRL)
School of Computer Science
University of Nottingham
Jubilee Campus
Nottingham
NG8 1BB

Leigh.Clark@nottingham.ac.uk

Abdulmalik Ofemile
School Of English
The University of Nottingham
University Park
Nottingham
NG7 2RD
aexacof@nottingham.ac.uk

Additional contact information

Supervisors:

Professor Svenja Adolphs – svenja.adolphs@nottingham.ac.uk

Professor Tom Rodden - tom.rodden@nottingham.ac.uk

School of English Ethics officer

Dr Gareth Carrol – gareth.carrol@nottingham.ac.uk

Appendix IIa: Consent Form (HHI)

Lego Assembly Tasks Consent Form

	Yes	No
I confirm that I have agreed to take part in this study, have read the information sheet provided and understand what is involved.	☐	☐
I understand that the study will gather recordings of my participation, and I agree to the use of this data in an anonymised form for research and analysis.	☐	☐
I understand that I can withdraw at any time by informing the researcher conducting this study, and my personal data will be erased from the records.	☐	☐
I confirm that I am over the age of 18.	☐	☐
In addition to the data analysis, I give permission for data that could identify me (e.g. photos, video) to be used in publications, conferences, presentations and future research	☐	☐

This is to confirm that I have agreed to take part in this research study on the date:

………………..

Signed ………………………………………..

Name …………………………….……………

Department / School ………………………………………..

Contact Tel No (optional): …………………………………

Email: …………………………………………...

Appendix IIb: Consent Form (HAI)

Lego Assembly Tasks Consent Form

I confirm that I have agreed to take part in this study, have read the information sheet provided and understand what is involved.

I understand that the study will gather recordings of my participation, and I agree to the use of this data in an anonymised form for research and analysis.

I understand that I can withdraw at any time by informing the researcher conducting this study, and my personal data will be erased from the records.

I confirm that I am over the age of 18.

This is to confirm that I have agreed to take part in this research study on the date:

………………..

Signed ………………………………………..

Name ……………………………….……………

Department / School ………………………………………..

Contact Tel No (optional): …………………………………

Email: …………………………………………...

In addition to the data analysis, I give permission for data that could identify me (e.g. photos, video) to be used in publications, conferences, presentations and future research □

Signed ………………………………………..

Appendix III: Assessing Inter-Rater Reliability

Percentage agreement is computed using decision Tables (Table1 &2) containing agreement ratios and percentages. These will be matched with benchmark scales in Landis and Koch’s Table, and Altman’s benchmark (Gwet, 2012).

To calculate values on the decision Table for facial actions, the number of AUs responsible for each action is sufficient and no averages are calculated as shown in decision Table1 below.

Table 1 Decision Table Facial Actions

Facial Action	**Annotator 1 Action Units**	**Annotator 2 Action units**	**Agreement**
Neutral	1	1	1
Neutral Intense concentration	2	2	1
Neutral face down	1	1	1
Tight-lipped smile	2	2	1
Swallowed/ compressed lips	1	1	1
Micro smile	2	2	1
Agreement Ratio			6/6
Percentage Agreement			**100%**

Gesture phases are calculated in averages by adding the number of phases in each member of a family of gestures then divided by the number of times the gesture occurs. For example, if three knowing hands occur that is 'N'. If the first knowing hand has 4 phases, the second has 3 and the third has 3 these will be summed up as 10 phases divided by 3. If the value is more or less than the next whole number by 5 it is rounded up or down to the next whole number respectively. The formula is given below:

Equation-1:Calculating Average gesture phases

$$\text{Average Gesture phases (AGP)} = \frac{\text{Number of phases } g1 + g2 + g3 + gn}{N}$$

Rounding up AGP is given as ± 5

To calculate agreement on the decision Tables, compare values for annotator 1 with values from annotator 2 if there is no difference write 1 for agreement however, where there is a difference, write zero (0) indicating no agreement(Table2).

Table.2 Decision Table Gestures

GESTURE FAMILY	GESTURE PHRASE	Phrases Annotator 1	Phrases Annotator 2	Agreement
Aligning Hand	Basic Aligning Hand	4	4	1
	Micro-aligning Hand	3	3	1
Picking Hands	knowing hand	3	3	1
	Basic picking Hand	3	3	1
	Picking with a hold	3	3	1
	Two-finger picking	3	2	0
	Two-Hand setting Aside	2	1	0
Presentation	Basic presentation/ displaying hand	4	4	1
	Deictic Hands	3	3	1
Joining Hand	Basic joining hand	3	3	1
	Multiple Joining hand	5	5	1
	Failed Joining Hand	6	6	1
	False starts	4	4	1
	Twisting hands	4	4	1
Agreement Ratio				13/15
Percentage Agreement				**86.67%**

sing the formula given below, the percentage agreement is calculated for each variable. This begins with adding up similar values to form agreement ratio of the whole as shown in decision Tables. The ratio is multiplied by 100 divided by 1 to get a percentage agreement shown below.

$$\text{Percentage agreement} = \frac{x}{n} * \frac{100}{1}$$

Where: x= ∑Agreement

n = ∑Total variables sampled

Equation 2: Calculating Percentage agreement(IRR)

Percentage agreement (Gestures) $\frac{13}{15} * \frac{100}{1}$ = 86.67%(Table 2 above)

Percentage agreement (facial actions) $\frac{6}{6} * \frac{100}{1}$ = 100%(Table 1)

Appendix IV: Second coder Annotations

Participant IH1- Task1

Instruction1

Figure 1 Neutral Face Down

The facial muscles responsible for Neutral Face Down is Action Units AU0.

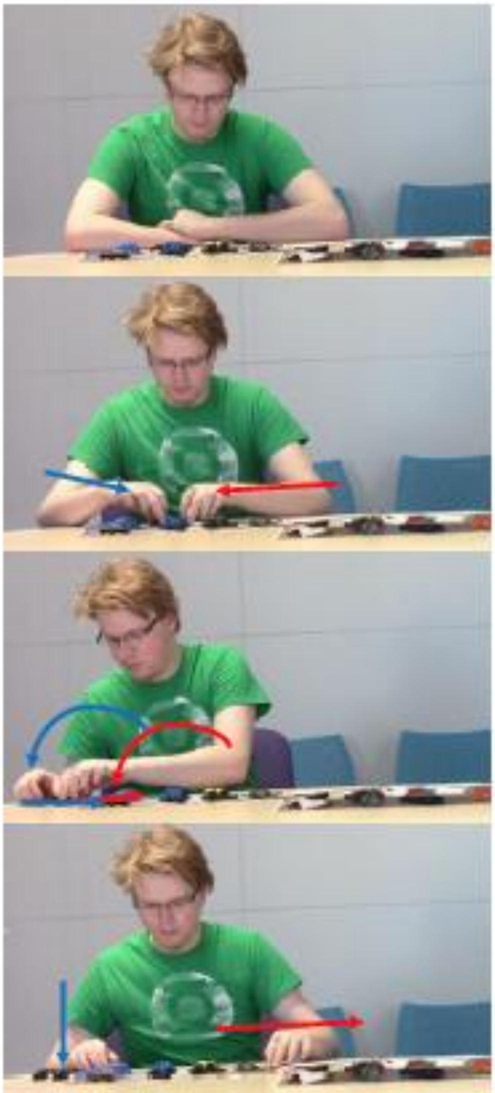


Figure 2 Dual Picking Hands Gestures

Preparation:
Participant IH1 has both hands in hold before executing a stroke. Both hands are palm down, as one is on top the other.

Stroke: As participant IH1 executes a stroke, by using both fore and middle fingers of both hands in a coupler-shaped form to move down and grasp the black claws.

Retraction1:
Retraction1 begins as participant IH1 drops and slightly moves them to the left whilst the middle and fore fingers of the left hand are stretched.

Retraction2:
Retraction2 begins as participant IH1 palms are placed down which slightly move the assembly bits forward.

Instruction 2

Figure 3-Multiple Joining Hands

Prepartion1: ParticipantIH1 has a False Start, whereby he has the wrong bit but then drops and picks the right bit with his left hand.

Stroke1: Participant by using multiple joining hands to attach the bits together. As the left hand moved to the right hand and attach the bits together.

Prepartion2: Participant IH1 then grasp a bit up with his left hand using his thumb, fore and middle fingers in a coupler-shape.

Stroke2: The participantIH1 then moves his left hand towards his right hand and attach the bits together.

Retraction1: Participant IH1 puts down the assembled bits by both hands.

Retraction2: The participant IH1 finally rest and clasps his hands together on the table

Figure 4 Neutral Face

The facial muscles responsible for Neutral Face is Action Units AU0.

Introduction 3

Preparation: ParticipantIH1 initiates gestures with hands under the table and has a neutral concentration on task.

Stroke1: Participant IH1 grasps the yellow bit up using his middle, fore and thumb fingers in a coupler-shaped.

After participant IH1 pulls up bits with his left hand, he then moves the yellow bit right to his right hand.

Stroke2: Participant IH1 the grasps another yellow bit and moves it upwards.

Post stroke hold: Participant IH1 retracts both hands to a lateral position still holding the yellow bit.

Figure 5-Picking with a hold

Preparation: Participant IH1 initiates the gesture with a neutral face down staring at the assembly bits and hands underneath the table

Preparation 1: Participant IH1 pulls up an assembly bit using his middle finger and thumb in a two-finger picking upward movement by his right hand.

Failed Stroke Phrase: After participant IH1 has picked up the assembly bits, he tries to put the assembly bits together with the right hand attaching the bit to the black piece in the left hand. These movements co-occur with neutral concentration on the task. participant IH1 is unable to attach the assemble bits down.

Post Stroke Hold: This image shows participant IH1 retracting both hands down towards the sagittal axis as he listens to repeated instructions with a neutral face

Figure 6-Failed Stroke Hold

Figure 7-Neutral Face

Participant IH1- Task 2

Stroke Phase1- Participant IH1 performs a basic joining hands together using the thumb and middle finger of left hands with a crooked aim whilst turning and twisting the piece in an up and down movement

Stroke Phase 2- Participant IH1 performs a decisive point movement by using twisting hands as he sorts out the assembly bit.

Stroke Phase 3 Part1- Participant IH1 performs two different actions in this phase. Firstly, he displays a gesture which include two finger picking (illustrated) his right hand using his thumb and fore-finger, while doing this he is also holding the object. Thus, he is picking with a hold whilst performs a two-finger picking gesture.

Stroke Phase 3 Part2- After the picking the bit, he then performs a checking hand gesture and a basic joining hands gesture using both hands.

Stroke Phase 4- Participant IH1 performs a circular presentation with the object whilst conducting self-repairs that are self-initiated. Whilst displaying these set of gestures, participant IH1 has a neutral concentration on task.

Failed Stroke Phase- Participant tries to perform a basic joining using his thumb, fore-finger and middle finger of his right hand to assembly bit together but he made a failed joining hand. Due to a failed joining hand, participant IH1 performs a recoil with the object.
Stroke Phase 5- Participant IH1 performs multiple joining hand gesture after recoil with a neutral face down.

Retraction- After the bits have been perfectly assembled together, participeIH1 places the object in a vertical position

P19_T1_HRAgent One (1)

Figure 9-Facial Actions

Pre-Stroke Phase- During this phase, the participant quickly brings both hands together to measure their fit (micro-aligning hand gesture) as she looks at the assembly bits with a neutral faced down.

Failed Stroke Phase- The participant brings both hand together as she tries to attach the assembly bits unsuccessfully (failed joining hand). This gesture co-occurs with a micro-tight lipped smile

Post Stroke Phase- The participant drops the piece in the right-hand palm down and picks up another (basic picking) and these co-occur with compressed lips as she brings both hands together again to align them.

Stroke Phase- The participant moves the RH leftwards and LH rightwards to first assess the fit of assembly bits (aligning hand) then attaches them (joining hand) along with compressed lips.

Retraction- The participant simply clicks on for more instructions after the bits are joined.

Figure 10-Neutral Face Down

Figure 11-Compressed Lips & Tight Lipped Smile

Agent Two (2)

Figure 17-Basic Picking

Preparation- The participant conducts a basic picking hand gesture with one hand as the one hand.

Failed Stroke Phase- The participant has a failed joining hand with jaw drop.

Post Stroke Phase- During this phase, the participant conduct a concatenated picking hand whereby she had to identify, isolate and pick up the assembly bits.

Failed Stroke Phase- The participant had a slight and angry disgust as she conducted a failed joining hand.

Post Stroke Phase- This participant during this phase has a searching hand which has her picking an assembly bit with a eureka aha moment assuming she found the right piece.

Stroke Phase- The participant at this phase uses three fingers of both hands also known as deitics hand gesture as a multiple joining hand. These gestures are applied with work face and a neutral concentration task

Retraction- The participant sets the item as she prepares for more instruction

Figure 13-Jaw Drop

Figure 14-Nervous Smile

Figure 15-Angry and Slight Disgust

Figure 16- Eureka (Aha) Moment

P19_T2_CL

Figure 18-Picking with a hold

Preparation Part1 –The participant conducts a checking hand as she holds the assemble bits on her left hand. This is done with a neutral face.

Preparation Part 2- This is another aspect of preparation but it includes the participant having a neutral hard face.

Preparation Part 3- With her neutral hard face due to not understanding the instruction she performs a double hand picking gesture. With both an assembly piece and the item itself. This leads to false start as she has an angry dis gust expression.

Stroke Phase Part1- The participant has a neutral face as well as instruction whilst picking with a hold.

Stroke Phase Part2- Assembly pieces are brought together and basic joining hand is conducted with a neutral concentration on task.

Figure 19-Neutral Face

Figure 20-Neutral Hand

Figure 21-Angry Disgust

Figure 22-Neutral Concentration on Task

Figure 23-Neutral Face

Appendix V: Lego Instructions – Nex Vague (Amended for HHI)

1. To start with locate the two orange feet and place them on the desk.
2. Now take two of the black pieces with two ball joints.
3. Then just attach the joints to the sockets of the oranges pieces. The black joint should just be pointing vertically in the same way.
4. So, keep these black pieces vertical and just give each one a little twist 90 degrees or so to the right.
5. So now take two of the grey pieces with ball joints.
6. Just attach them to the sockets of the black pieces. The sockets should be pointing vertically in pretty much the same way.
7. Now pick up the two small identical white pieces that look something like armour.
8. These should attach to the front joint of each black piece. The ends with the holes should be closest to the grey piece.
9. Now find two of the orange pieces that again are the same size.
10. Just attach them to the outside of each grey piece. The ends with the holes should be closest to the sockets.
11. Now locate the largest black piece with seven ball joints. This is the body.
12. Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.
13. Now pick up the two remaining grey pieces with ball joints.
14. Just attach the sockets onto the top two black joints of the body.
15. Then just take the two remaining black pieces with ball joints.
16. The same should be done with the sockets of the grey pieces so they are more or less in line with each one.
17. Now locate the large white armour piece.
18. This should attach to the middle ball joint on the front of the body and the narrow end should point downwards.
19. Now just take the white piece that has like an 'H' in the middle.
20. Just attach this to the holes on the top of the white piece.
21. Now pick up the small green and the large orange pieces that look like heads.
22. So keep the eyes of both pieces facing forwards and basically just connect the orange crossed connector inside the large piece to the green crossed hole of the smaller piece.
23. Now just attach the socket inside the green piece to the top black joint of the body. The eyes again should more or less be facing forwards.
24. Now locate the small, thinnest black piece and a small orange armour piece.
25. Just attach the orange piece to the grey socket on the right arm. The end with the holes should be closest to the head.
26. The black piece has tothen be placed just inside the orange hole.
27. Now to make the right arm weapon just pick up the two crossed red connectors, the black fist and the long orange piece.
28. Basically, connect the red piece to the crossed connector of the orange piece. The other half should stick out at the other end.
29. Just connect the black fist to the other end of this connector so the socket is just pointing away from the orange piece.
30. Then just connect the other red piece to the crossed hole of the fist.
31. Now pick up the large grey piece that is more narrow than the other and does not have, like, a small black attachment.

32. Just connect the crossed hole of the grey piece to the remaining half of the red connector so it more or less lines up with the orange piece.
33. Just connect the black fist to the black ball joint on the right arm and now pick up the gr.
34. Now locate the remaining orange armour piece and the smaller grey armour piece.
35. Just attach the orange piece to the grey ball joint on the left arm. The end with the holes should be closest to the head.
36. Now just attach the grey piece to the orange holes. The end that looks a bit wider should be closest to the head.
37. Now to make the left arm weapon just pick up the large grey piece and smaller green piece that looks a bit like a tube.
38. The widest end of the tube should connect to the smallest end of the black attachment just on the grey piece.
39. The socket of the grey piece should be attached to the black ball joint just on the left arm.
40. So now the arms should be turned so they are more or less pointing forwards.
41. Now pick up the long yellow piece that looks like a tube.
42. One end should be placed in the remaining hole of the black attachment.
43. The other should be placed in the bottom left hole around the back of the large white piece.
44. Now take the small grey piece and the longer one that looks sort of like a chain.
45. So then place the grey piece in the top hole of the orange armour piece that is just on the right leg.
46. One wide end of the chain thing should be attached to this small grey piece.
47. Now just rotate this wide end a little so it is pretty much pointing vertically.
48. The model is complete. Please inform the researcher.

Appendix VI: Lego Instructions – Aquagon Vague

1. To start with locate the two black feet and place them claws down on the desk.
2. Now take two of the small black pieces that are sort of a Y shape and have a crossed hole at the bottom.
3. Just put these in the rear gap of each foot and attach them to the cross-shaped connector.
4. These should be pushed towards the front of the feet until they are more or less firmly in place. The round socketshould point vertically towards the ceiling.
5. So now take two of the yellow pieces with ball joints.
6. Just attach the ball joints to the sockets of the black pieces so the yellow socket is pointing vertically in pretty much the same way.
7. Now pick up the two black pieces with ball joints.
8. Just attach them to the yellow sockets so they are again pointing vertically the same way.
9. So keep these black pieces vertical and just twist each one a littlebit 90 degrees or so to the right. These are the legs.
10. Now find the two smalldark grey pieces that look something like armour.
11. These should attach to the remaining black ball joints around the outside of each leg. The widest end should be closest to the feet.
12. Now locate the largest black piece that has seven ball joints. This is the body.
13. Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.
14. Now take a black cylinder, a grey cylinder and a small light grey piece with a curved thing that looks a bit like a fin.
15. Just place the black cylinder into the bottom round hole at the back of the body piece.
16. Now just attach the grey piece with the fin to this cylinder. The end that looks a bit wider should be closest to the body.
17. Just give this piece a little bit of a twist so the fin is more or less pointing towards the desk.
18. The grey cylinder should go in the rear hole of the grey finned piece slitted end first.
19. Now locate the largest dark grey piece that looks like a big piece of armour.
20. Just attach this to the middle ball joint on the front of the body using the round socket. The end that's a bit more narrow than the other should point downwards.
21. Now to make the arms just take two grey pieces with ball joints, two yellow pieces with ball joints and the two small black pieces that look like fists.
22. Just connect the yellow joints to the socket of each fist.
23. The same should be done with the end grey joints in the yellow sockets.
24. The grey socket of each arm should be attached onto the top two black joints of the body.
25. Now locate the two small, identical transparent and the fourcurved yellow pieces that sort of look like spikes.
26. Just place the big spikes into the holes that are closest to the edge of each piece.
27. Then just place the small spikes into the other outside hole so they are just below the big ones.

28. Now just attach the whole thing to the available grey joints on the arms. The end with the big spikes should be closest to the top of the body.
29. Then just give each spike a little twist so they point towards the body.
30. Now take the remaining yellow piece with a ball joint.
31. Just attach the yellow socket to the top black joint of the body. The yellow piece should point forwards so it's more or less in line with the feet.
32. Now take the remaining blue piece that looks a bit transparent. This is the head.
33. Basically attach the socket of the head to the joint of the yellow piece. The longer part of the head should point forwards. One crossed hole should point forwards and one should be pointing vertically towards the ceiling a bit like the other pieces.
34. So now locate the large blue and red piece.
35. Basically, attach the crossed red connector to the crossed blue hole on top of the head. The eyes should face forward.
36. Now take the tail end and place it inside the back hole of the grey finat the back of the body.
37. So now locate the yellow face piece.
38. Just attach it to the remaining hole on the front of the head.
39. Now take the remaining grey finned pieces and two blue cylinders.
40. These cylinders should be placed into the back of each fist using the crossed connector.
41. Just attach the finned pieces to the cylinders and then just twist them a bit so the fins point to the sides.
42. Now take the black cylinder and blue sword pieces.
43. Just place the cylinder into the remaining hole in the left fin.
44. Now just take each sword and just attach it to the front hole of each fist. They should both be basically pointing outwards to the sides.
45. The arms then have to be turned so they are more or less pointing forwards.
46. The two finned pieces should be connected together using the black cylinder.
47. Finally just twist one of the fins a bit so it faces the body.
48. The model is complete. Please inform the researcher.

Appendix VII: Lego Instructions – Nex Vague

1. To start with locate the two orange feet and place them on the desk.
2. Now take two of the black pieces with two ball joints.
3. Then just attach the joints to the sockets of the oranges pieces. The black joint should just be pointing vertically in the same way.
4. So keep these black pieces vertical and just give each one a little twist 90 degrees or so to the right.
5. So now take two of the grey pieces with ball joints.
6. Just attach them to the sockets of the black pieces. The sockets should be pointing vertically in pretty much the same way.
7. Now pick up the two small identical white pieces that look something like armour.
8. These should attach to the front joint of each black piece. The ends with the holes should be closest to the grey piece.
9. Now find two of the orange pieces that again are the same size.
10. Just attach them to the outside of each grey piece. The ends with the holes should be closest to the sockets.
11. Now locate the largest black piece with seven ball joints. This is the body.
12. Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.
13. Now pick up the two remaining grey pieces with ball joints.
14. Just attach the sockets onto the top two black joints of the body.
15. Then just take the two remaining black pieces with ball joints.
16. The same should be done with the sockets of the grey pieces so they are more or less in line with each one.
17. Now locate the large white armour piece.
18. This should attach to the middle ball joint on the front of the body and the narrow end should point downwards.
19. Now just take the white piece that has like an 'H' in the middle.
20. Just attach this to the holes on the top of the white piece.
21. Now pick up the small green and the large orange pieces that look like heads.
22. So keep the eyes of both pieces facing forwards and basically just connect the orange crossed connector inside the large piece to the green crossed hole of the smaller piece.
23. Now just attach the socket inside the green piece to the top black joint of the body. The eyes again should more or less be facing forwards.
24. Now locate the small, thinnest black piece and a small orange armour piece.
25. Just attach the orange piece to the grey socket on the right arm. The end with the holes should be closest to the head.
26. The black piece has tothen be placed just inside the orange hole.
27. Now to make the right arm weapon just pick up the two crossed red connectors, the black fist and the long orange piece.
28. Basically, connect the red piece to the crossed connector of the orange piece. The other half should stick out at the other end.
29. Just connect the black fist to the other end of this connector so the socket is just pointing away from the orange piece.
30. Then just connect the other red piece to the crossed hole of the fist.

31. Now pick up the large grey piece that is more narrow than the other and does not have, like, a small black attachment.
32. Just connect the crossed hole of the grey piece to the remaining half of the red connector so it more or less lines up with the orange piece.
33. Now pick up the green ball.
34. This should be placed in between the grey and orange pieces.
35. Now locate the remaining orange armour piece and the smaller grey armour piece.
36. Just attach the orange piece to the grey ball joint on the left arm. The end with the holes should be closest to the head.
37. Now just attach the grey piece to the orange holes. The end that looks a bit wider should be closest to the head.
38. Now to make the left arm weapon just pick up the large grey piece and smaller green piece that looks a bit like a tube.
39. The widest end of the tube should connect to the smallest end of the black attachment just on the grey piece.
40. The socket of the grey piece should be attached to the black ball joint just on the left arm.
41. So now the arms should be turned so they are more or less pointing forwards.
42. Now pick up the long yellow piece that looks like a tube.
43. One end should be placed in the remaining hole of the black attachment.
44. The other should be placed in the bottom left hole around the back of the large white piece.
45. Now take the small grey piece and the longer one that looks sort of like a chain.
46. So then place the grey piece in the top hole of the orange armour piece that is just on the right leg.
47. One wide end of the chain thing should be attached to this small grey piece.
48. Now just rotate this wide end a little so it is pretty much pointing vertically.
49. The model is complete. Please inform the researcher.

Appendix VIII: Lego models Set-up Layout

Nex Model Layout

Aquagon Model Layout

APPENDIX VIV: LISTENER FACIAL ACTIONS AND GESTURES IN HAI (CHAPTER 4)

1. Facial Actions- Neutral Face

2 The second type of neutral face co-occurs with listeners leaning forward

Neutral Face- Head bending forward

In HHI people cock their heads or move closer to speakers or objects of speech when they want to have a better listening position or look. Similarly, listeners bend forward,given by AU 54 or the head down (Ekman, 1988), with eyes focused at some point in space with hands on the Table doing nothing. This suggests that listenersare paying close attention as instructionsare repeated for better comprehension.

3 Another sibling is the neutral face with a hard expression. Here the face is predominantly expressionless yet showing traces of negative emotions such as tightly compressed lips shown in Figure 4.8 below. Ekman (1997) suggests that tight lips may indicate a trace of anger; however, this may also reflect a mood or way of concentrating on the task.

Neutral Face-Hard Expression

4 Another variation of the neutral face shows listeners' faces expressionless, but their hands are engaged in various tasks.

4a-Neutral Face-concentrating on task

This suggests that participants are multitasking by assembling or trying to pick an assembly piece while listening to the repeated instruction (Figure 4.9) above. Sometimes it combines pushing up the lower lip as the upper lip presses down (AU8). Ekman (2007) suggests that it is a sign that listeners are concentrating on the task at hand.

The next variation is the intense neutral face shown (4b) below. It is characterised by the partially furrowed forehead or horseshoe-like mark produced by Darwin's muscle of difficulty or eyebrow gatherer(Ekman, 2007), which is a separate strand of AU 44 (Table 3.7)while other facial muscles are relaxed (AU 0).

4b-Neutral Face-Intense

Gathering the eyebrow may indicate agony, perplexity, confusion or determination depending on the context in which the emotion is expressed.

Appendix VIV.6 Surprise

Surprise (Figure4.12) overleaf is often triggered by something unexpected such as a sudden loud noise or sighting an unexpected phenomenon and Ekman (2003) says that it is very brief probably lasting a few microseconds. For example, surprise is probably triggered by CP's voice when giving instruction to the participant for the very first time. The following muscles (Figure 3.33; Table 3.7) make the expression possible: AU1 raised the inner brow, AU2 raised the outer brow, AU5B slightly raised the upper lip while AU26 made the jaw to drop down (Ekman, 1997).

6a-Surprise

Research suggests that, direction of gaze may indicate where the source of stimulus is located (Hadjikhani et al., 2008; Ambadaret al., 2009). In Figure 6a above, the participant displayed fixed gaze towards the probable source of surprise in the task layout.

Another variation is the slight surprise (6b) that occurs with a lower intensity shown in Figure 4.13 below and given by AU1B AU2, AU5B and AU26 (see Table 3.7). The inner brow is raised slightly, the upper lip is raised, and the jaw drops down (Ekman, 1982b).

6b-Slight Surprise

The emotion is triggered by a sudden realisation that the previous assembly steps were wrongly executed of after repeating instruction 12. This makes the participant to carry out self-correction which begins with dismantling the model and re-assembling preceding stages before executing the current instruction.

Appendix VIV.7 *Nervous Smile*

The nervous smile (Ambadar et al., 2009) earlier called the miserable smile (Ekman, &Friesen 1982) is produced deliberately in a very short period. It may be superimposed on a negative expression or come after the expression with the negative expression persisting (Ekman, 1997).

The participant (Figure7) below displays a mixed feeling of smiling and embarrassment elicited by a negative physical and tactile stimulation (Ekman &Friesen 1982), alarm, and discomfort due to failure to execute assembly tasks. The morphological characteristics (Figure 3.33; Table 3.5) responsible for this emotion include AU 6 that raises the cheeks, AU12 pulls down lip corners and AU 54 does the head downward movement.

7-Nervous Smile

The participant is probably embarrassed but smiling just as participants in Ekman et al., (1980) were unhappy and smilingin adverse circumstances. The nervous smile is not as intense as the felt smile but, it may have a longer duration, greater amplitude, more head downward movement and is more likely to include open mouth than tight-lipped smiles (Ambadaret al., 2009).

Appendix VIV.8 *Tight-lipped Smile*

The tight-lipped smile (Figure8) overleaf is signalled by lips stretched tight across the face in a straight line with the teeth unexposed (Ekman, 1988). The muscles responsible (Figure 3.33; Table 3.7) include AU 6B that slightly raises the cheeks without visibly reducing the eyes' apertures or creating very visible crow's feet around the eyes. AU12B slightly pulls lip corners thus creating some wrinkles.

8- Tight-lipped Smile

In HHI, tight-lipped smile has been found to be used by people that do not want to reveal their true feelings (Malik, 2010) and may be interpreted by others as a sign of rejectionamong women in Western societies(Pease & Pease, 2004).

Appendix VIV.9 *G.W Bush Grin*

This facial action is named after trademark smirk on the face of President George W. Bush(Pease & Pease, 2004). The participant displays a smile-like appearance (Ekman, 2007) in the Bush grin (Figure 9) below. The sad element becomes clear when one covers the upper part of the face and looks at the lower part.

9- George W Bush Grin

The Bush grin (Figure 9) is characterised by the archetypal oblique eyebrows, lowered lip corners, slightly stretched mouth and raised cheeks. The muscles responsible (Figure 3.33; Table 3.7) include AU 1 the inner brow raiser, AU 15 the lip corner depressor, AU 27B the mouth stretcher and AU 6 the cheeks raiser. Here the eyelids are lowered by AU 5 helping the eyes focus on the piece held up while listening to the instruction again.

Appendix VIV. 10 Slight disgust

Appendix VIV.11 Frown

Appendix VIV.12 Compressed lips

Appendix VIV. 13 Biting the Lip

Another form of the tense mouth is biting lips shown (Figure 4.27) below. The muscle basis (Table 3.7) for this is AU12 which pulls the corner of the lower lip over teeth in the lower jaw then AU32 executes the lip bite (Ekman, 1988).

13 Lip biting

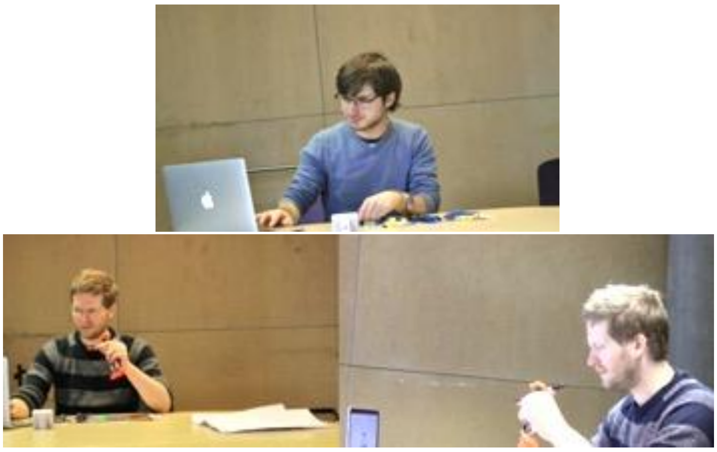
Morris (2015) suggests that the lip bite coincides with anger in the angry gesturer. While this may be tenable in HHI, biting the lip here suggest uncertainty as participants are thinking and interpreting the agent's instruction just as Kita's (2013) co-thought gestures indicate cognitive processing of information.

Appendix VIV.14 Micro Expressions-anger

Appendix VIV.15 Micro Expression –sadness

Appendix VIV.16 Micro-smile

Just as in high intensity emotions, enjoyment experienced in assembly tasks is sensory pleasure (Ekman, 2007) derived from appraisal such asexperienced by listeners in assembly tasks appraising instructionsand tasks using the listener's automatic processing strategy (Chapter 2). This process may trigger micro-smiles (Figure 4.32) below.Morphologic features responsible for these smiles include the slight contraction of AU6B which slightly raises the cheeks and AU12B which pulls the lip corner.



16- Micro-smile

A quick drawing up of the eyebrows with lips spreading suggest that participantsare experiencing a slight enjoyment or may be pleased, amused, or feel good with the task at hand. On the other hand, it may just be that participants are having a tough time and smiling to make the best of a bad situation.

Appendix VIV.17 Micro-tight-lipped Smile

The tight-lipped smile (Figure 4.33 overleaf) is indicated by lips that are stretched tight across the face in a straight line with the teeth unexposed (Ekman et al., 1980). The muscles responsible (Table 3.7) include AU6B that slightly raises the cheeks while AU5B lowers the eyelids thus visibly reducing their apertures. AU12B slightly pullslip corners and creates wrinkles.

17- Micro-Tight-lipped smile

And there is an indication of the slight action of AU27B-the mouth stretcher giving the smile a Bush-like grin suggesting that, the participant may be managing emotions with this smile (Ekman, 2007).

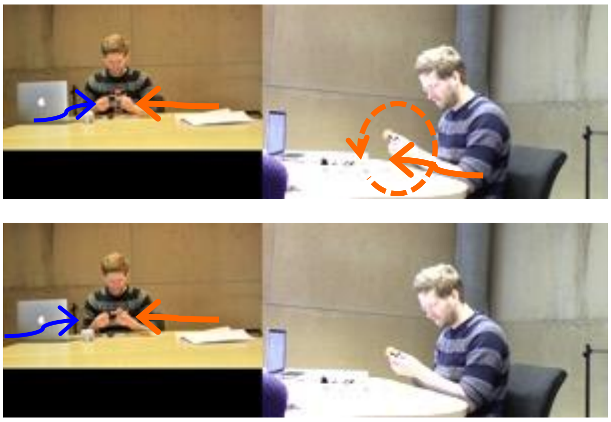
Appendix VIV. 18 Assembly Gestures- Aligning hands

In the vignette (Aligning hands –a) below, the process is same as in 4.34 except that there are two strokes suggesting two attempts at joining the assembly bits.

Preparation: P43 initiates this gesture with both hands in a brief hold. Preparation continues with P43's RH moving leftwards to present an assembly piece for examination and visual comparison.

Stroke 1: P43's LHgoes rightwards, palm lateral to grasp and pulls up the assembled kit, and turn it diagonal.

Stroke 2: P43's RHnow brings the assembly piece closer to the kit in the LHto visually measure their fit without them touching.



18a-Aligning Hands

In addition, presentation here is done with the participant's hands resting on the Table instead of in mid-air.

Another form of the aligning hand gesture has three gesture phases like Kendon's (2004) gesture phrases and the participant executes the gesture with a retraction that is more of a post-stroke hold as illustrated in the vignettes (Aligning hands –b) below.

Preparation P34 initiates the gesture with the **LH** holding the receiving piece while the **RH** is holding the piece to be joined to the Lego kit.

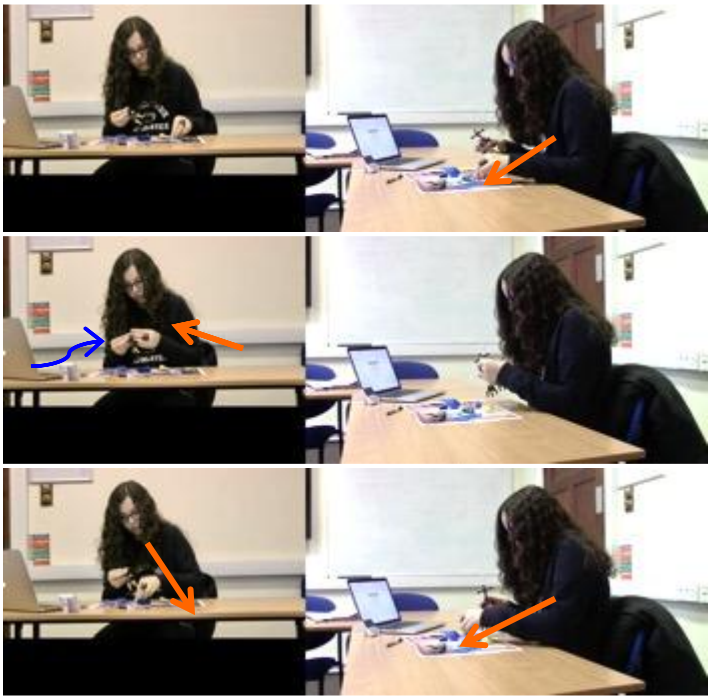
Stroke P34 performs the phase by bringing the hands together by taking the piece in the **RH** (giving hand) close to the Kit in the **LH** (receiving hand) and sizing them up.

Retraction: occurs as P34's hands move apart without fixing the pieces together. The**LH** moves leftwards (M1-L) and down (M1-D) to the middle of the sagittal axis (SCC) while the **RH** moves right and upwards to the right-hand side of the sagittal axis. Both hands are in a very brief hold in preparation for another gesture

18b-Aligning Hands

Participant 34 (Aligning hands –b) above measures fit and size of assembly bits visually and physically by touching the assembly bits together without joining them.

Although, the participant (Aligning hands –c)overleaf also uses three steps but instead of a post-stroke hold, she drops the piece instead of attaching it. The form and function of the gesture areillustratedin the vignette overleaf



Preparation P27 initiates the gesture with the **RH** holding the receiving piece while the LH moves towards the kit layout, palm down with digits held coupler-shaped to pull up the piece to be joined to the Lego kit.

Stroke P27 performs the phase by bringing the hands together by taking the piece in the **LH** (giving hand) close to the kit in the **RH** (receiving hand) and sizing them up. The **RH** spins the kit around with the feet towards her body or sagittal axis and the head facing the kit layout thereby positioning it to receive the bit from the **LH**

Retraction: occurs as P27's **LH** goes down to drop the piece without fixing it to the one in the **RH.** The**RH** is in a hold. Retraction here

18c-Aligning Hand

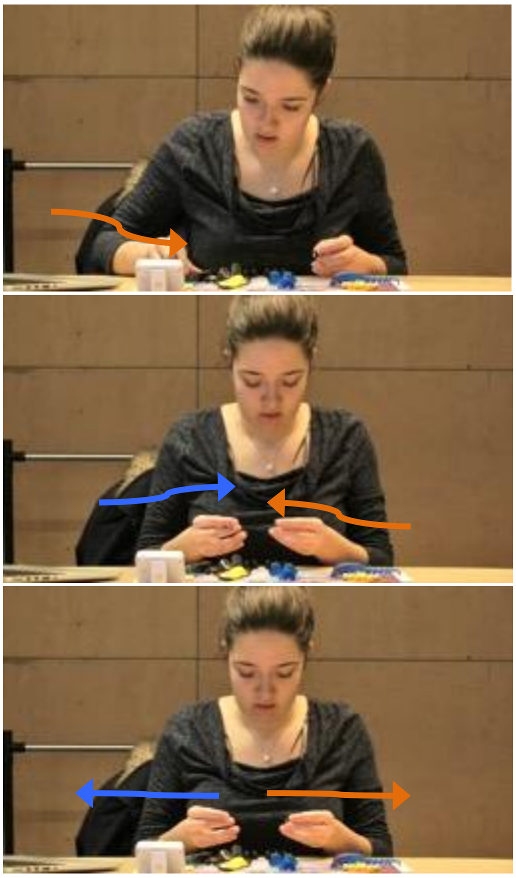
The structure of the gesture illustrating the basic aligning hand is shown below

<~~~~ ~~~~~~~/***********************************/############ >
P S R
… grey piece / ……with the fin to this cylinder… …/........ The

The structure of the gestures shows that the stroke co-occurs with aspects of the instruction. This suggests that the participant combines listening with task execution.

Appendix VIV. 19 Aligning Hands

Micro-gestures are as quick as microexpressions (Figure 4.30) because they both last microseconds. The gesture begins at 00:00:37.890, ends at 00:00:39.220 and the duration is 00:00:01.330. The form and function of the micro-aligning hand are given in the vignettes (Micro-Aligning hands –b) overleaf.



Preparation P17 initiates the gesture by using the **RH** palm down with digits coupler-shaped to pick a piece while the **LH** is holding another piece of the Lego kit in preparation for the stroke.

Stroke P12 performs the phase by bringing the hands together quickly thus physically taking the pieces in both hands close to one another and sizing them up

Retraction: occurs as P17's hands go backwards to a post-stroke hold position still holding the pieces without fixing them together.

19a-Micro-aligning hand

The gesture phrase lasts only 1.33 microseconds.

The multiple or extended aligning hand gesture is executed with four strokes during a long pause andmay also externalise participants' mental strategies in processing instructions. The form and function of the gesture are shown in the vignette (Multiple Aligning hands) overleaf.

19b-Multiple aligning Hands

Preparation: P21 initiates the gesture by picking up a black piece with the LH while the RH is holding the yellow piece and resting on the Table.

Stroke 1: The **LH** brings the black piece from the left while the **RH** brings the yellow piece from the right to measure their fit and size them up.

Stroke 2: The **LH** holds the black piece while the **RH** turns the yellow piece left to measure their fit and size them up.

Stroke 3: The **LH** holds the black piece while the **RH** turns the yellow piece right to measure their fit and size them up.

Stroke 4: The **LH** twists the black piece clockwise while the **RH** twists the yellow piece anticlockwise to measure their fit and size them up.

Retraction: The **LH** retracts to a hold position still holding the black piece while the **RH** retracts holding the yellow piece towards the agent. The pieces are not joined together.

The structure of the gesture phrase during the pause is shown below

<~~~ ~~/*********/************/**************/***********/-.-.-.-##### >
P S1 S2 S3 S4 Post stroke hold and R
00:06:56:720/ 00:06:58:060/ 00:06:59:320/00:07:00:600/00:07:01:960/ 00:08:14:020

The structure of the gesture suggests that the participant is repositioning the pieces as he is thinking through the process.

The aligning hand (Aligning hands with twists) comes with twists. The form and function of this gesture are given in the vignette below

Preparation: In this gesture, the **RH** is the joining hand while the **LH** is the receiving hand. P19 initiates this gesture with the **RH** holding a piece to be aligned while the **LH** holds the main body ready.

Stroke 1: Both hands move closer. The **RH** holds the joining piece while the **LH** positions the body ready for joining

Stroke 2: P19's **RH** remains stable while the **LH** turns in a cyclic twist clockwise then anticlockwise adjusting the position and measuring the fit of the piece held with the body.

Stroke 3: The **LH** repositions the body in clockwise and anticlockwise cyclic twists while the stable**RH** tries to join the piece assessing the fit with the body.

Retraction: Both hands go sideways in retraction to a twin hold still holding the joining parts without fixing them.

19c-Aligning Hands with Twists

The twisting movements suggest that the participant is measuring the fit of the assembly pieces by twisting them within a three-dimensional space (henceforth 3-D space).

Another form of basic gesture begins with the picking hand gesture done with one hand but the two hands perform presentation for the comparison. The form and function of the gesture are shown in the vignette (Basic presentation after picking) below

Preparation: P12 prepares for this gesture with the **LH** in a hold while the **RH** asks for a repeat of instruction before taking the projected path for the stroke indicated by the broken arrow.

Stroke 1: The first stroke occurs when P12' **RH** bends, palm down with digits held coupler-shaped and pulls out a piece (picking action).

Stroke 2: The **RH** then retracts (indicated by the upwards arrow) turning clockwise (indicated by the cyclic arrow) to a post stroke hold position for comparison with what the **LH** is holding for P12 to assess for fit. It is like the aligning hands without bringing the hands together.

20a-Basic Presentation after Picking

Here the right hand is very active; it pulls up the assembly bit then ends in a long post-stroke hold while the left hand is in a hold position with the palm lateral. Both holds form the presentation gesture.

Another variation of basic presentation hand is shown in presentation with a Right hand below. This time the right-hand wrist flexes ulnarly about 180^0 rightwards to make the presentation.

Preparation: P38 initiates this gesture in two stages. In the first stage, P38's **RH** is requesting instructions.

Pre-stroke Hold: In the second stage, P38's **RH** held palm down moves backward towards the sagittal axis to a brief pre-stroke hold with the palm lateral.

Stroke 1: P38's **RH** then moves forward, palm down towards the kit layout with digits held coupler-shaped to pull up the yellow assembly piece. The **LH** is held palm lateral and in a hold.

Stroke 2: P38's **RH** moves up from the Table then spins rightwards as the wrist flexes ulnarly thus turning the hand palm lateral. The fingers are coupler-shaped holding the piece to make a presentation for examination

20b-Presentation with the Right Hand

Another form of presentation is done with the left hand. The form and functions of the displaying hand is shown in the vignette (Figure20c) below

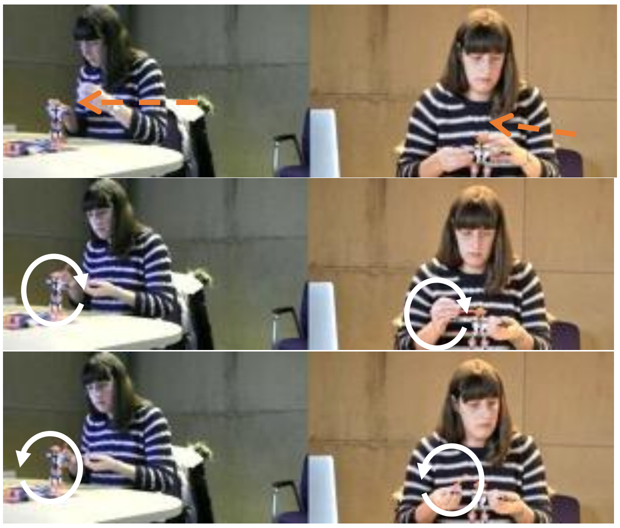
Preparation: P17 prepares for the stroke palm down by picking the pieces at once in a fist (fingers curled around the pieces) with the **LH** palm down.

Stroke: P17 now executes the stroke. The **LH** turns leftwards then palm up with the digits stretched out/opened and the pieces on top for confirmation as P17 listens to the repeated instruction. The **LH** is held midway between the Table and participant's face.

Retraction: Thistakes place when P17 puts the **LH** down palm up or supine on the Table.

20c-*Presentation with Left Hand*

The circular displaying hand gesture in assembly tasks is like beat gestures (McNeil, 1999) but distinct because of the full circular or 360^0movements involved. The form and function of the gesture are shown in the vignettes (circular presentation) below.



Preparation: P3 exchanges the piece from the **LH**to the **RH** in preparation for the multiple strokes

Stroke 1: Turning the piece clockwise inwards towards the body with the **RH**

Stroke 2: Turning the piece anti-clockwise outwards away from the body with the **RH**

20d-Circular Presentation

The participant in above may have used circular presentation to assessher execution of that stage of the assembly process in much the same way that Kendon (2004) observes people presenting things to others for viewing and assessment

During assembly tasks, participants examine assembly pieces using presentation and alignment gestures to measure the fit or size up the assembly bits. The form and function of this gesture are shown in the vignette (Figure 20e) below.

Preparation: P29 initiates the gesture by passing the assembly piece from the RH to the LH.

Stroke: P29's LH turns clockwise about 90^0 to the left, holding up the piece to present it before his eyes for examination. The RH is in a hold.

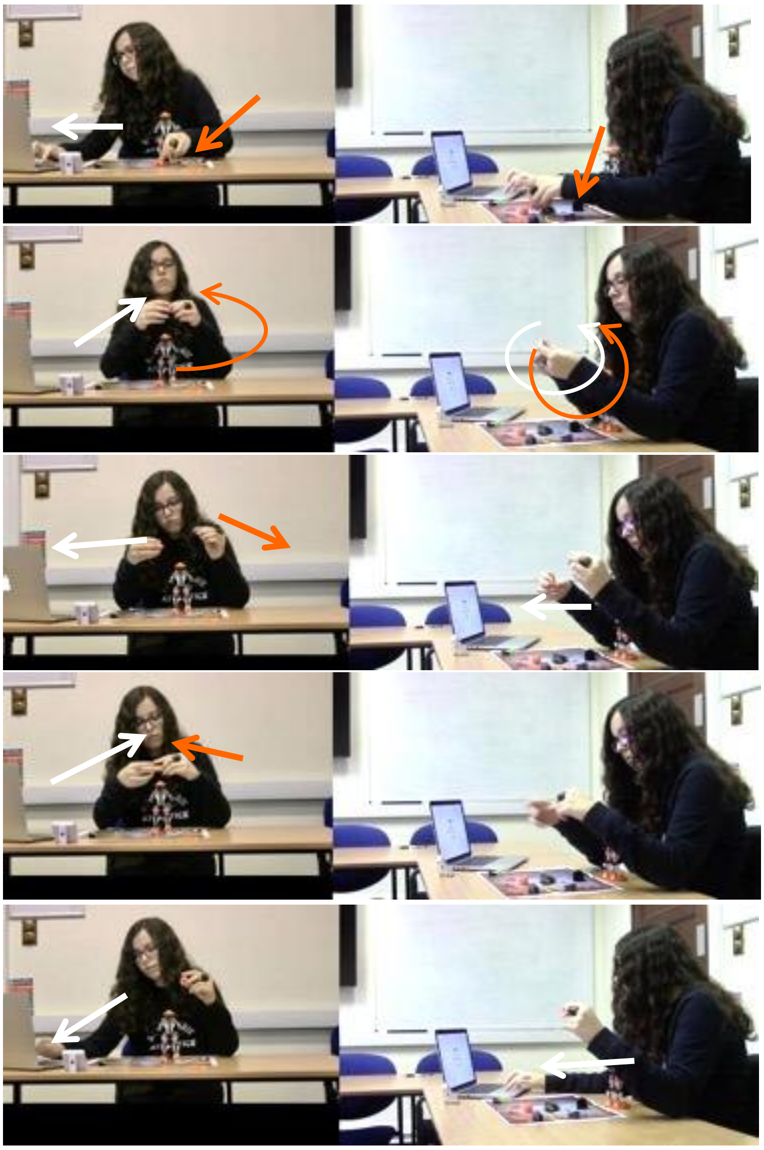
Retraction: P29's LH turns anticlockwise about 90^0 to the right for further examination.

20e-Presentation with Aligning Hand

The structure of the gesture is shown below

<~~~~~~~~~~~~~/*************/.-.-.-.-####>
P S R
00:02:47:828 / 00:02:49:443 /00:02:49:723

Another type of presentation is followed by multiple retractions. Retraction begins with the separation of the hands then the exchanging hand gesture takes place where the participant drops the piece from the right hand into the left hand in preparation for retraction. The form and function are shown in the vignette (Figure 20f) overleaf.



Preparation: P27 initiates the gesture by asking for a repeat of instructions with the RH while the LH picks the assembly pieces.

Stroke: Both hands move up palms lateral then turn clockwise to present the pieces before the participant's eyes for examination.

Retraction 1: P27 begins to retract by first separating the hands. The RH moves rightwards still holding a piece while the LH moves up and leftwards palm down still holding another piece.

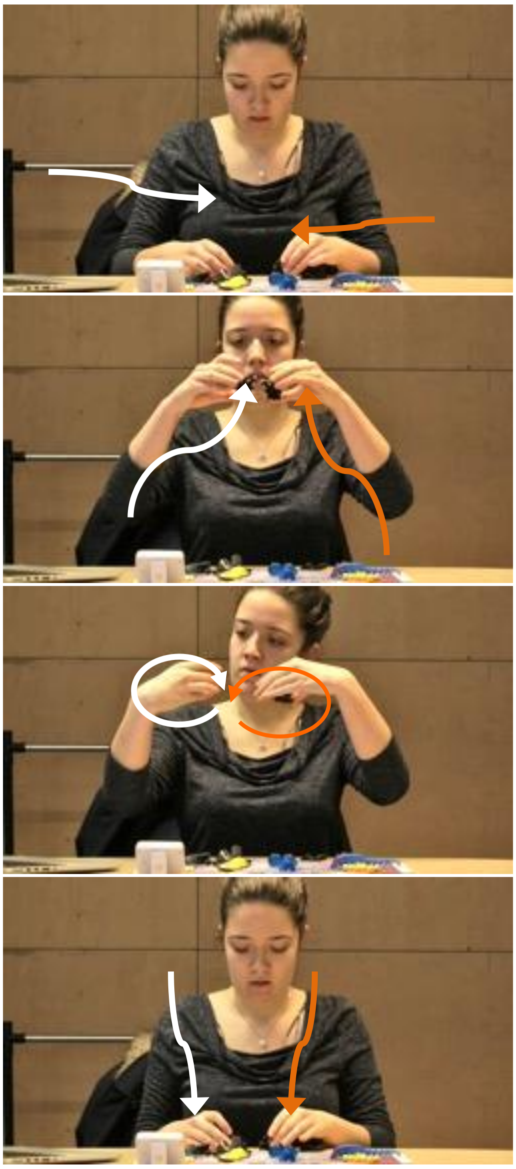
Retraction 2: The second stage is the exchanging hands where the RH drops a piece into the LH

Retraction 3: The RH moves to the left palm down with digits stretched out. The LH is in a hold.

20f-Presentation with Multiple Retractions

The participant (Figure 20f) above presents the pieces for examination as the instruction is repeated suggesting that the repetition is to confirm if the participant picked up the correct assembly bits

Presentation may also be enacted with a combination of picking and examining gestures done with both hands. The form and functions of the gesture are shown in the vignette (Figure 20g)overleaf.



Preparation P17 initiates the gesture by picking the assembly piece in the **RH** and **LH** from the workspace

Stroke 1: P17 performs the first gesture by using both hands to align and present the pieces for examination from one angle and sizing them up

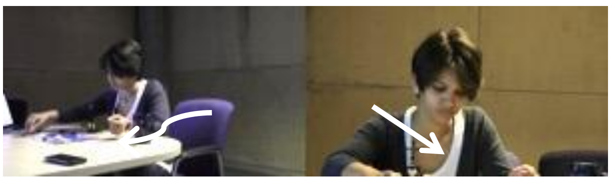
Stroke 2: P17 performs the second gesture by turning (indicated by the cyclic arrows) the pieces picked up inwards towards the body using both hands to align and present the pieces for examination from another angle and sizing them up

Retraction:occurs as P 17's hands go back to the preparatory position on the workspace still holding the pieces.

20g-Presentation combined with Dual picking

The participant the vignette above twists assembly pieces around while holding the pieces for examination which suggest physical measurement of fit.

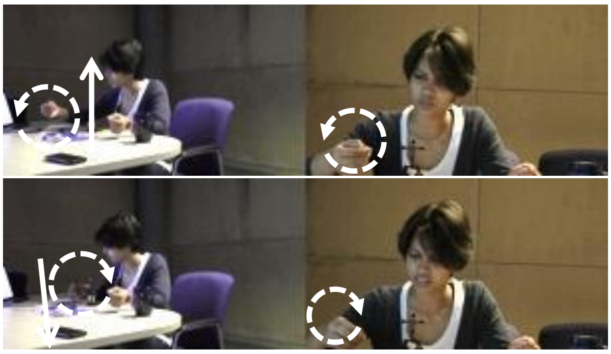
Another form of presentation is done with the tips of the fingers instead of the palm for examination. The forms and functions of the gesture are shown in the vignette (Figure 20h) below.



Preparation: P9 prepares for the gesture by stretching her **RH**, palm down with digits held coupler-shaped.

Stroke: P9 performs the gesture by retracting the **RH** (indicated by the upward arrow) then a brief hold, turning the picking hand round and leftwards (indicated by the cyclic arrow) bringing the palm and the piece held up for examination.

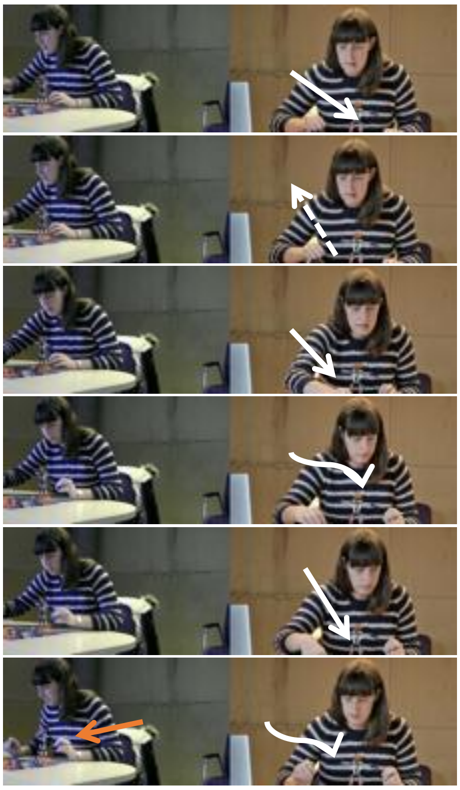
Retraction: P9 retracts by bringing the **RH** palm down to the Table (indicated by the downward arrow) while turning the hand rightwards (indicated by the cyclic arrow), palm down and dropping the piece.



20h- Presentation with Fingertips

Appendix VIV.21 Picking Gestures

Another form of searching hands is the concatenatedsearching hands. It so named because like the concatenated gesture (Kendon, 2004), it is made up of multiple phases (Figure 3.8)as attempts to pick up a piece, hesitations and retractions occurring one after the other with the participant picking up nothing. The functions and structure of the gesture are illustrated in the vignette (21a) overleaf.



a. Preparation 1

RH

b. Retraction
RH

c. Preparation 2
RH

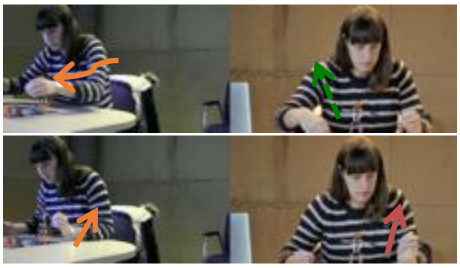
d. Stroke 1 **RH**

e. Preparation3
RH

f. Stroke2 **RH**/
Preparation **LH**

g. Stroke **LH** /
Retraction **RH**

h. Retraction LH



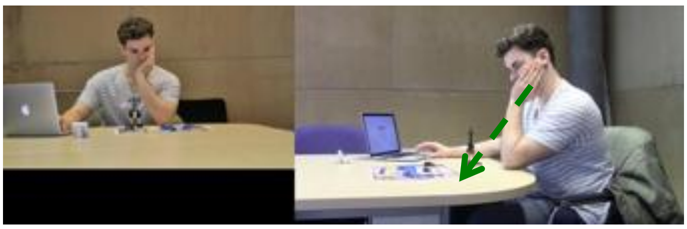
21.a-Concatenated Searching Hand

Preparation begins with an outstretched hand and palm down with digits starting to form the coupler shape. The outstretched handpalm downwith coupler-shaped digits movesand bendstowards the workspace to selectassembly pieces but, there is a quick retraction which, is the backwards movement of the hand to the position of preparation in a, b, c, d, e. In f stroke 2 for the Right Hand (**RH**) co-occurs with the preparation for the Left Hand (**LH**), in g, while the **RH** is retracting, the **LH** is executing a stroke and in h the LH comes to rest.

These gestures externalise the internal comprehension and decision-making process going on within the participant. Retraction with an empty hand may indicate hesitation equated with a “retraced false start” (Maclay & Osgood, 1959, p.21) probably due to the participant’s uncertainty about what to do, incomprehension of instructions or indecisiveness. These become pertinent because the pieces are well laid out (Appendix VIII) yet the listener’s hands come to rest without picking up any piece.

Picking to Hold

This picking hand gesture is used to remove assembly pieces from the kit layout with one hand but the piece is held with both hands after retraction. The form and function of the gesture are shown in the vignette (Figure 21b) below.



Preparation: P21 initiates the gesture from a pensive pose (LH on cheek) and **RH** communicating with the agent. Intended path for **RH** indicated by the broken green arrow.

Stroke: P21’s **RH** moves to the kit space, palm down with digits held coupler-shaped and pulls up a piece while the **LH** is on the Table clenched in a fist

Retraction: The **RH** retracts backwards with the piece and it is now held jointly with the LH in a double post stroke

21b-Picking Hand To hold in Both Hands

The structure of the gesture phrase during the pause is shown below

<~~~ ~~~~~~~~~~~/*******/.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.########## >
P S Post stroke hold and R
-er and a small light grey piece with a curved thing that looks a bit

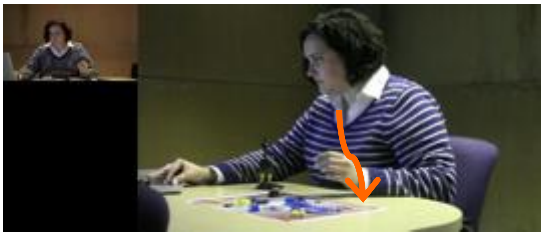
Picking Hand with a Hold

This is another picking hand gesture done with one hand. The hand that picks up assembly pieces also retains them in a palm down post-stroke hold. The form and function of the gesture are shown in the vignette (Figure 21c) overleaf.

Preparation: P4 prepares for the gesture as a continuation of a hold.

Stroke: P4 performs the stroke by moving the **LH**, leftwards towards the piece. The **LH** is palm down, bending forward using a combination of the thumb, fore and mid fingers to pull up one piece

Post-stroke Hold: P4 retracts the **LH** to a brief standstill finally the **LH** comes to rest holding the piece palm down.



21c- Picking Hand with a Palm Down Hold

Another picking hand gesture is done with one hand like a beat gesture, retracts and ends with a supine post-stroke hold. The form and function of the gesture are shown in the vignette (Figure21d) overleaf.

Preparation: P12 prepares for the gesture with the **LH** at rest. The broken green arrow indicates the projected path of the picking hand

Stroke: P12 performs the stroke by moving the **LH**, straight towards the piece. The **LH** is palm down, bending forward with coupler-shaped digits then using a combination of thumb, fore and mid fingers to pull up the piece

Retraction: P12 begins to retractthe **LH** in preparation for a post stroke hold

Post stroke Hold: P12 retracts the **LH** rotating it into a palm up or supine position, holding the piece in a post-stroke hold.

21d-Picking Hand with a Supine Hold

The Concatenated Picking Hand

The concatenated picking hand is executed with multiple gesture phrases as outlined in Figure 3.8. It is used to confirm if the instruction was correctly executed and aid the next operative gesture. The form and structure of the gesture are shown in the vignette (Figure 21e) overleaf.

Preparation 1

P2 prepares for the gesture by stretching his **LH** palm down forward with cupped digits.

Stroke 1 P2 uses the **LH** palm down to do the one-finger flick gesture to select and eliminate pieces that do not fit the description in the instruction.

Stroke2 P2 uses the **LH** to do a two-finger flick gesture to eliminate more pieces that do not fit the description in the instruction

Preparation 2 P2 continues by stretching the **RH** palm down with coupler-shaped digits then using a combination of the thumb, forefinger and middle finger, to pull-up the piece.

Retraction: P2's **LH**rotates 90^0 leftwards into palm up position with the digits holding the piece to make a

21e- Concatenated Picking Hand with a Presentation

The concatenated picking hand comprises the first preparation, two strokes as one-finger flick followed by the two-hand flick (Figure 3.12) to select a piece, then another preparation, before picking up the desired piece. It continues with retraction and the hand turns radially (Henley, 2010) to make a presentation at the end.

Picking and Aligning Hand

The assembly process requires that participants pick then join pieces together. However, the picking action may lead to joining or concatenated gestures combining picking and aligning hands. The form and function of the gesture are illustrated in the vignette (Figure21f) below.

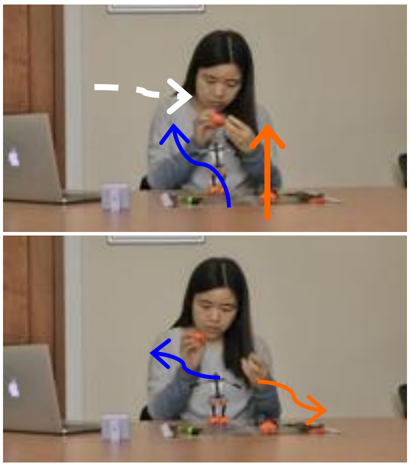
Preparation: P34 initiates this gesture with the **RH** asking for instructions while the **LH** is in a hold.

Stroke 1: P34's **LH** moves forward, palm down with coupler-shaped digits to pull up the green piece

Stroke 2: As the **LH** moves palm lateral backwards towards the sagittal axis, the **RH** moves diagonally forward towards the kit layout, palm down using a combination of the thumb, forefinger and middle finger to pull up the green piece.

Stroke 3: The **RH** moves up palm down while the **LH**rises. The **RH** then goes down with the piece while the **LH** comes up from beneath to execute a top-down aligning hand.

Retraction: Both hands move apart in opposite directions. The **RH** enacts a presentation while the **LH** ends in a post-stroke hold.



21f-Picking and Aligning Hand

This is also a natural progression of the assembly process that participants useto measure the fit of assembly pieces.

Dual Picking Hands

This picking gesture is done with two hands at the same time. It is the opposite of the alternate picking hands (Figure 21h) where both hands are used one after the other simultaneously. The form and function of the gesture are shown in the vignette (Figure 21g) below.

Preparation: P25 initiates the gesture with the **LH** in a hold while the **RH** asks for the current instruction.The broken green arrow shows the intended path of the **LH**

Stroke 1: P25 moves both hands forward, palm down with coupler-shaped digits to pull up the pieces.

Stroke 2: The **RH** moves diagonally to drop the piece on the left while the **LH** drops the piece in front of the participant. Both hands form an incomplete 'X'.

Retraction: Both hands retract backwards disentangling the 'X' to rest on the Table.

21g-Dual Picking Hands

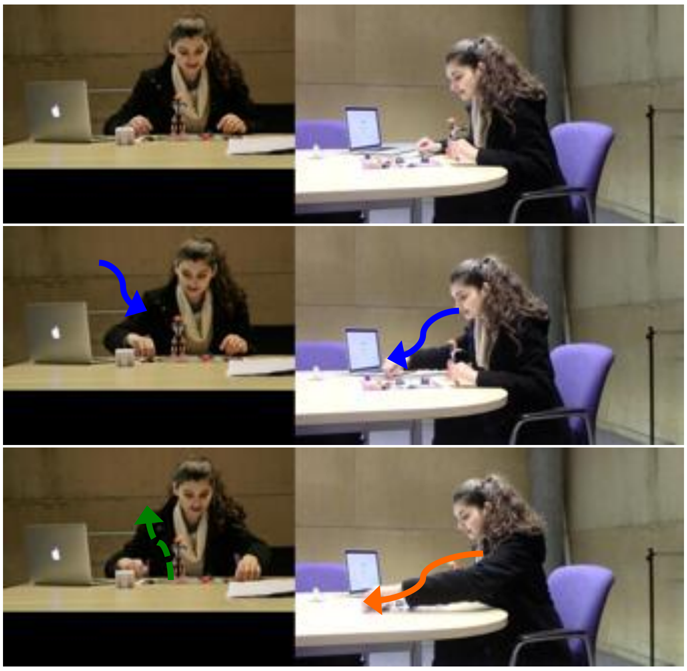
The structure of the gesture phrase is as shown below

```
                         …black cylinder, a…grey cylinder
<~~~~ ~~~~/******************/*************/######>
              P                    S1              S2              R
```

The gesture phrase comprises one preparation, one dual hand picking; X-shaped drops and cross-retraction

Alternate Picking Hands

The alternate picking hands gesture is done with both hands in a back and forth simultaneous motions where one hand picksup an assembly bit at a time from different positions. The form and function of the gesture are shown in the vignette (Figure 21h) overleaf.



Preparation: P20 initiates gesture with the **LH** and **RH** at rest on the workspace

Stroke 1: The first stroke occurs when P20's **RH** b moves forward, palm down coupler-shaped digits and pulls out a piece (picking action) while the **LH** is in a brief hold.

Stroke 2&Retraction the **RH** then retracts backwards (the first green broken arrow) to drop the piece picked up while the **LH** moves forward, palm down, with coupler-shaped digits, bending down picks up another piece (orange arrow) and retracts (the second green and orange arrows) to a post-stroke hold position.

21h-Alternate picking Hands-a

The gesture phrase structure is shown below with timings in microseconds

<~~~ ~~~~~/ ***********.-.-.-/≈******-.-.-.- /-.-.-.-.-.-#####>
P S1 S2 Post stroke hold and R
00:05:48:000/00:05:48:610 /00:05:50:250/ 00:05:53:330

P20 enacts the gesture in Figure 4.60 above where S1 (Table 3.3) is a multi-segment phase (Kita et al., 1998). The first stage picks the piece while in the second a retraction co-occurring with S2 is used by the participant to draw assembly pieces closer.

Another variation of alternate picking hands is just like Figure 4.61 above with a clearer face view of the interaction. The form and function of the gesture are shown in the vignette (Figure 21i) overleaf.

Preparation: P28 initiates the gesture from holding the kit with both hands. The **RH** is palms lateral holding and twisting the upper part while the **LH** is palm down holding down the lower part of the kit.

Stroke 1: The **RH** goes forward palm down with digits held coupler-shaped to pull up the piece with the thumb-forefinger combination.

Stroke 2: The **LH** moves straight to the left of the kit layout, palm down, digits coupler-shaped to grasp the piece with a thumb-forefinger combination.

Retraction/Stroke3: As the **LH** retracts to a hold, the **RH** moves backwards towards the sagittal axis to pull up another piece with a thumb-forefinger combination

Retraction 2: The **RH** moves to the right, palm down, digits curled in a half fist with the forefinger stretched to ask for a repeat of the instruction. The **LH** is still in a post-stroke hold.

21 i- Alternate Picking Hands-b

The structure of the gesture is shown below

<~~~~~~~~~~~/*******************/**********/###********/***####>
P S1 S2 R & S3 R
…grey cylinder and a/ small light grey piece/00:03:11:266/ 00:03:13:106/ 00:03:17:146

Another alternate picking hand gesture is executed with radial movements. The form and function of the gesture are shown in the vignette (Figure 21j) below.

Preparation: P40 initiates this gesture with the **LH**under the Table while the **RH**is in a brief pre-stroke hold.

Stroke 1: P40 moves the **LH** forward, palm down towards the kit layout with the last two digits curled while using a combination of the thumb, forefinger and ring finger to pull up the assembly piece.

Stroke 2: The **LH** withdraws towards the sagittal axis as the **RH** is held in preparation.

Stroke 3: As the**LH**reaches the sagittal axis, the **RH** moves palm down towards the kit layout with coupler-shaped digits to pull up another assembly piece.

Stroke 4: The **RH** then rotates rightwards making the palm lateral with the thumb and the middle finger holding the piece to make a presentation. The **LH** is in a post-stroke hold and palm down.

21j- Alternate picking Hands - Radial Movement

The first stroke is executed with a radial movement of the left hand while, in the last stroke the right-hand rotates in ulnar pronation to make a presentation of the picked piece. The last move (S3) suggests that P40 uses the gesture to confirm or assess the task executed.

Another sibling of alternate picking hand gestures is distinct because it ends with both hands holding the pieces up for either visual alignment or comparison. The form and function of the gesture is shown in the vignette (Figure 21k) below.

Preparation: P24 initiatesthe gesture with the **LH** on the chin and **RH** at rest on the workspace

Stroke 1: The first stroke occurs when P20's **LH** moves forward, palm down coupler-shaped digits and pulls out a piece while the **RH** is in a brief hold.

Stroke 2&AlternateRetraction the **LH** then retracts backwards (broken green arrow) to a post-stroke hold while the **RH** moves forward, palm down, coupler-shaped digits, bending down picks another piece (orange arrow) and retracts (blue and orange arrows) to a post-stroke hold position.

Stroke 3: The **LH** turns anti-clockwise as the **RH** retracts and turns clockwise to display the pieces picked up for P24 to examine.

21k-Alternate picking Hand -Dual Holds

The gesture occurs towards the tail end of the repeated instruction and the structure of the gesture phrase is shown below.

<~~~ ~~~~~/*********/*************/******/##### >
P S1 S2 S3 R
. ange armour piece/ 00:04:33:073/ 00:04:34:390/00:04:33:150 /

The gesture structure indicates that the first stroke co-occurs with part of the instruction while the rest of the strokes occur during the pause with both hands retracting to a post-stroke hold. These moves suggest that, while S1 may be considered a co-interpretation move, S2 and S3 are not.

Picking Hand –Ulnarly

This is another picking hand gesture that occurs with movements of the wrist that does all the work because the piece is very close. The form and function of the gesture are illustrated in the vignette (21L) below.

Presentation: P35 initiates the gesture with the **LH** on the Table while the **RH** is requesting for instructions

Stroke: P35's Left forearm is resting on the Table but the wrist rotates45^0rightwards (half pronation) with the thumb and forefinger stretched forward to pull up the assembly piece. The **RH** moves leftwards, palm down using the thumb's retropulsion and the forefinger to pull up another piece.

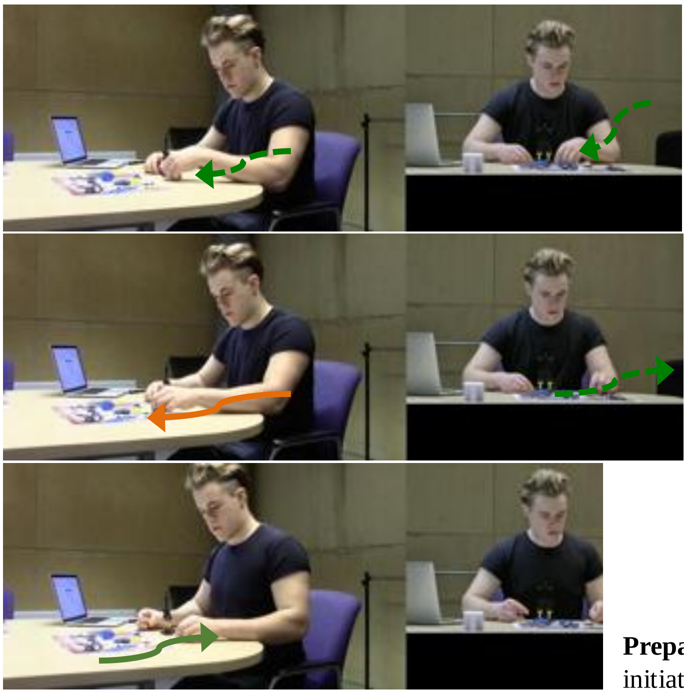
Retraction: P35's **LH**rotates about 45^0 leftward in a half-supination to make a presentation. The **RH** is at rest on the Table.

21l-Picking Hand- Ulnarly

The picking hand-ulnarly gesture (Figure 4.65) combines ulnar movements (left hand's half pronation) with retropulsion (right hand only) characterised by the pincer-like shape of the forefinger and thumb to pull up assembly pieces.

Picking: Two-finger Hand

In the assembly task bits and pieces come in all sizes and this may to an extent determine how the participant picks them up. Heavier and broader objects require the use of more digits than smaller and lighter ones. P14 uses the two-finger hand gesture to pick up small pieces before assembling them. The form and function of the gesture are shown in the vignette (21M below



21m- Picking with Two Fingers

Preparation: P14 initiates the move from a rest position with both hands on the Table. The broken green arrows indicate the intended paths.

Stroke: P14 moves the **LH** palm down with the forefinger, midfinger and thumb extended towards the piece while the ring and small fingers are curved inwards. The stroke occurs when the extended fingers pick up the piece.

Retraction: P14 retracts the **LH** in a straight line backwards to its rest position where the piece is dropped.

Picking Hand - Double

The double picking gesture is done with one hand picking two or more assembly pieces sequentially as distinct from the dual picking hand (Figure4.59). P34 concludes this gesture with a retraction that turns into a presentation with the radial movement of the left hand. The form and function of the gesture are shown in the vignette (21N) below.

Preparation: P34 initiates this gesture with the **RH** asking for instruction while the **LH** with the palm lateralis in a hold.

Stroke 1: P34's **LH** moves forward to the kit layout and palm down. The forefinger, middle finger and the thumb are stretched out to pull up the assembly piece while the rest of the digits are curled inwards.

Retraction 1: The **LH** moves backwards towards the sagittal axis briefly palm down still holding the piece.

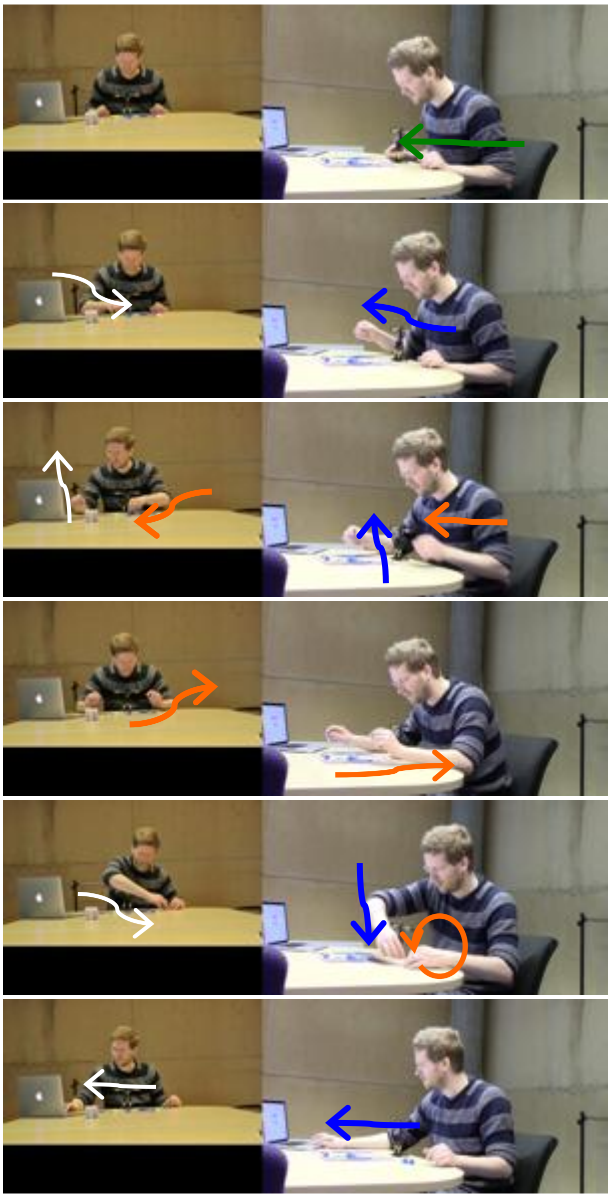
Stroke 2: This continues from the retraction. The **LH** moves forward again, palm down and using a combination of the thumb and forefinger to pull up the second assembly piece.

Retraction 2: The **LH** moves towards the sagittal axis palm down then it turns palm up to make a presentation.

21n-Double Picking Hand

Appendix VIV. 22 Segmented Repositioning - Picking Hand

The first part of a segmented repositioning gesture is the alternate picking hand. Although, it is seemingly done independently but its link to the second segment becomes clearer when the final gesture is enacted when executing instructions. The form and function of the gesture is shown in the vignette (22) below.



Preparation: P43 initiates this gesture with both hands at rest on the Table while the bending forward suggests that the participant is paying attention.

Stroke 1: P43's **RH** moves palm down with coupler-shaped digits towards the kit layout to use a combination of the thumb and forefinger to pull up the assembly bit.

Stroke 2: As the **RH** retracts towards the sagittal axis, the **LH** moves forward, palm down with coupler-shaped digits to pull up another assembly bit.

Retraction1: The **LH** goes backwards towards the sagittal axis to rest in a post-stroke hold with the **RH.**

Stroke 3: P43's **RH** moves leftwards, palm down to drop the piece it is holding. At the same time, the **LH** turns upside down to become palm down and drop the piece it is holding on the Table too.

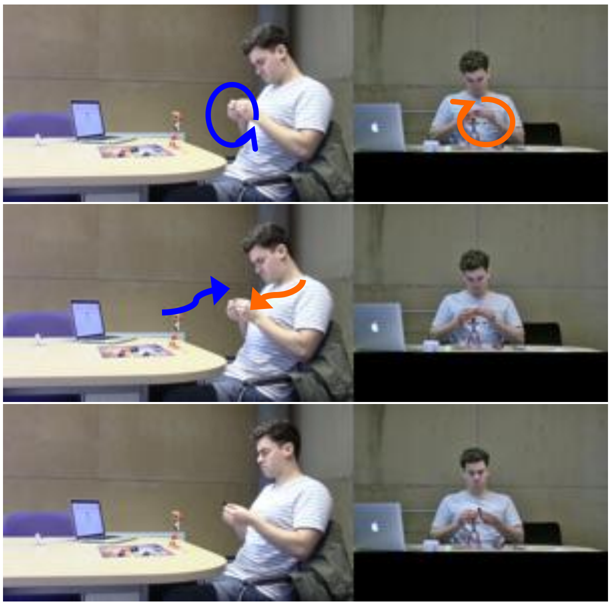
Retraction 2: The **RH** moves rightwards towards the agent while the **LH** rests on the Table

22. Segmented Repositioning

The second part of segmented repositioning is a deliberate dropping hand by the participant (S3) to set aside pieces picked up while waiting for the next instruction. Note that the side view is always showing the progression while the face view being ahead shows the final execution. This staggered presentation enables observers to have two views of the same action in different execution stages.

Appendix VIV. 23 Joining Hands

One variation of joining hand is used for carefully attaching tiny and fragile assembly pieces. The form and function of the gesture is shown in the vignette (23a) below.



Preparation: P21 initiates the gesture by twisting both hands in opposite directions, as indicated by the cyclic arrows

Stroke: The **LH** brings the black piece from the top while the **RH** holds the yellow piece for the two to become attached together.

Retraction: This is more of a display of the finished product for inspection. P21 repeats the instruction while assessing the assembled piece.

23a-Joining Hands for Tiny Bits

The structure is as follows

<~~~ ~~~~~/**************************** /-.-.-.->

P S Post stroke hold and R

00:08:09:630/ 00:08:12:067 / 00:08:14:020

Multiple Joining Hands

The multiple joining hands gesture may indicate the listener's adjustment of various assembly techniques to successfully execute instructions. This may be aided by listener's continuous adjusted interpretation of instructionsto achieve the same meaning with the agent. The form and functions of the gesture are illustrated in the vignette (23b) below.

Preparation: P12 initiates the gesture from an aligning position and joining two pieces together unsuccessfully, then readjusting the hands grip. The **LH** holds up the extended arm of the Lego kit in preparation for the stroke by the **RH**

Stroke 2: P12's **RH** (the giving hand) moves towards the **LH**(holding the receiving piece) to attach the piece to the extended arm in the prescribed location.

P12 adjusts his**LH**'s grip by deliberately dropping the piece held following the intended path (indicated by the broken green arrow in preceding vignette).

P12's **LH** regains the hold-up position (indicated by the upwards broken arrow) in preparation for another stroke.

Stroke 3: P12 successfully joins the pieces together by moving the **RH**, palms vertical to stick the piece laterally to the bodywith a slight twist.

Retraction: P12 retracts the **RH** towards the interface to a request a new instruction while still holding the Lego kit with the **LH**.

23b-Multiple Joining Hands -a

Another variation of the multiple joining hand is used to assemble kits that are paired such as feet and arms as well as bits that are similar in structure. The form and function of the gesture is illustrated in the vignette (23c) below.

Preparation 1:
P2initiates the gesture with the LHpalm down and pulling up one assembly piece. The RH is palm lateral holding an assembly piece.

Stroke 1: The LH moves upwards palm up to a post-stroke hold as the RH goes upwards to meet the LH in a bottom-up co-joining movement that fits the two pieces together.

Post-stroke hold/preparation:
The LH holds up the joined pieces while the RH goes down palm lateral in preparation for another move.
Stroke 2: TheRH palm lateral grasps the upright kit while the LH is still holding up the assembled bit.
Stroke 3: The RH moves up with the assembly piece to meet the LHstill in a post-stroke hold to execute another joining hands
Retraction Both hands come to rest on the Table holding the assembled kit

23c *Multiple joining Hand*

Another form of failed joining hands gesture is enacted with the repositioning of assembly bits after the initial trial before repeating instructions. The form and function of this failed joining hands gesture are shown in the vignette (23d) below.

Preparation: P27 initiates the gesture by from the retraction in the previous aligning hands gesture. The **RH** is in a hold while the **LH** is picking up a piece

Stroke 1: P27's **RH** is holding the body of the Lego kit and the **LH** brings the grey piece picked up from the side to attach to the cylinder. **Failure (?).** The grey piece does not fit as expected.

Reposition 1: P27 uses this move to turn the kit around and bring it down with the **RH** holding the body while the **LH** holds the grey piece.

Reposition 2: Both hands move upwards raising the kit up while the **LH** twirls the grey piece to get a sense of its fit.

Reposition 3: P27 brings the kit down and uses the **LH** to turn the grey piece round in the same manner one turns screws.

Reposition 4: P27's **RH** holds the body up while the **LH** turns the grey piece then pulls it out

Retraction: The **RH** retracts rightwards to ask for another repeat while the **LH** holds the body.

23d Failed Joining Hand with repositioning

The structure of the gesture phrase is shown below

<~~~~/**********??????/*******************************/**************/************/**********/-.-.-.-.-.-.-##>

P S1 (Failed) / S2: Repositioning(R) 1 S3: R 2 S4: R3 S5: R 4 Retraction

… end that looks / … a bit wider should be closest to the body. /00:09:31:125 / 00:09:35:245 /00:09:37:205 /00:09:38:085

After initial failure, multiple repositioning follows, indicative of a continuous change in thelistener's assembly behaviour. Repositioning is a listener coping strategy externalised in four movementsthat visualise the listener's intrapersonal negotiation, amendment and interpretation of instruction during task.

Appendix VIV 24: Self-Initiated Self-Repairs

One other form of self-initiated repairs during assembly tasks is the extended self-repair process that occurs in two stages to correct listener's picking up the wrong assembly piece as outlined in the vignette (24) below.

Trouble Source (TS)

Preparation: P41 initiates this gesture with the **RH** at rest while the **LH**is palm down pulling up an assembly bit.

Stroke 1: Proceeds from preparation as P41's **LH**uses the thumb and forefinger to pull up one assembly piece.

Retraction 1: P41's **LH** moves backwards towards the sagittal axis palm down. However, midway the **LH** turns leftwards making the palm lateral.

Stroke 2: P41's **LH** then turns rightwards to become palm down and then drops the assembly piece on the Table.

Retraction 2: P41's **LH** moves leftwards to rest on the Table.

Repair initiation

Stroke 1: P41's **LH** rises up then moves to the left, palm down and using the combination of the thumb, forefinger and middle finger to pull up the wrong assembly piece. This is the beginning of actual repair

Actual repair execution

Stroke 2: P41 carries out the actual repair here. P41's **LH** moves forward, palm down with digits clutching the piece to drop it in its original position in the kit layout.

Retraction: The **LH** retreats towards the sagittal axis.

24 Extended Self-initiated Self-repairs

The first stage begins with a picking hand that ends with the participant dropping the piece on the Table; this is where the trouble spot is located. The second stage is used to return an assembly piece back to its original position when the listener realises that it was picked up in error. The extended self-repair process, unlike the normal repair process, comprises multiple gesture phrases and each can stand on its own as an independent gesture; however, they are linked by the second retraction in stage one which serves as the beginning of the repair initiation.

Appendix VIV.25: Deictic Gestures

Deictic Co-interpretational

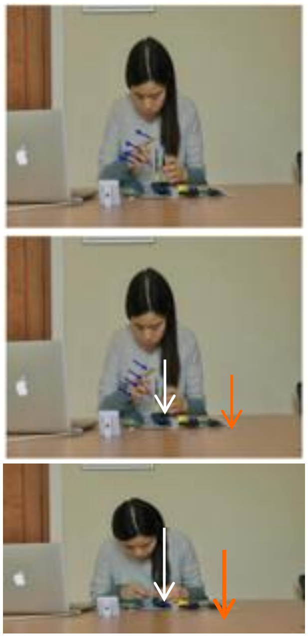
This deictic gesture begins with a representational gesture (Kita 2013; Kendon 2004) and ends with pointing such that it performs two functions at once. The form and function of the gesture are shown in the vignette (Figure 25a) below

Preparation: P34 begins this gesture with both hands in a brief hold with palms lateral.

Stroke 1: This is done in two quick movements. P34's **RH** moves up palms lateral with the digits spreading out back and forth to measure and mimic the *'the widest end'* twice. The **LH** moves in the opposite direction

Stroke 2: The second movement is a repetition of the palm measurement. This time the **RH** is palm down.

Stroke 3: P34's **RH** moves downwards, palm down with digits pointing to the 'feet' of the kit. The **LH** grasps the kit



25a Deictic Co-Interpretational Gesture

<~~~~~~~/**********/*************/*********>

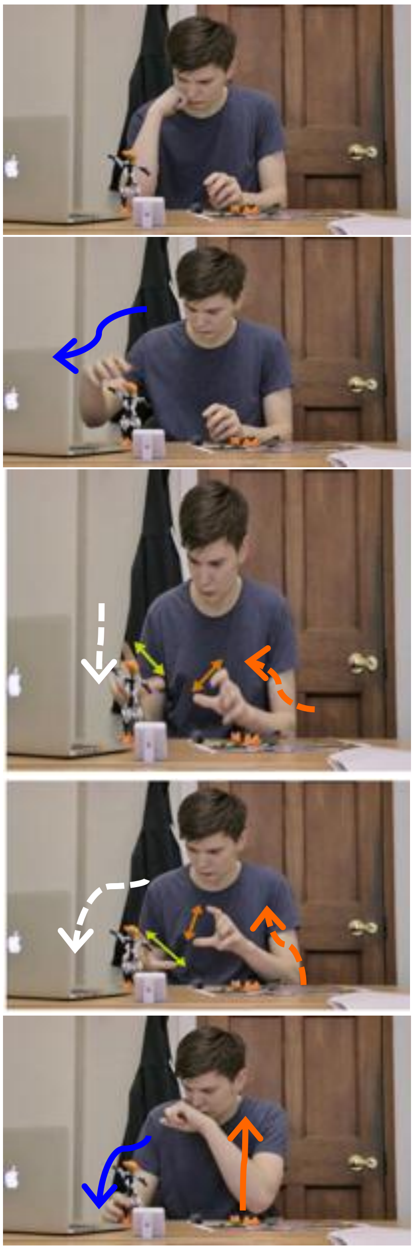
P S1 S2 S3

… the widest end should be closest to the feet

The structure of the gesture (Fig 25a) above shows that S1 and S2 involves the fingers using a horizontal beat-like movement to make spans in quick succession to measure and give an iconic interpretation of the word *"wide end"* in the instruction. S3 is executed with the right hand pointing at the '*feet*' of the kit. Here, listener gesture co-occurs with words in instructions giving rise toco-interpretational gestures like Chu and Kita's (2016) co-thought gestures and McNeil's (1992) co-verbal gestures.

Deictic Representation

This is wholly a representational gesture enacted using many quick successive moves that end with the listener grasping the upright kit. The form and function of the gesture are shown in the vignette (Figure25b) below.



Preparation: P37 initiates the gesture with the **RH**on the chin and the **LH**on the Table as the instruction is repeated.

Stroke 1: P37's **RH**goes forward palm down with digits spread out to quickly hover above the upright kit.

Stroke 2: P37's **RH**moves downward, palm down with the forefinger flexing opposite the thumb's abduction movement. In a simultaneous movement, the **LH**moves up, palm down and when it levels upto the **RH**with the forefinger flexing opposite the thumb's abduction and adduction movements. These two movements create the representational gesture of sizing up

Stroke 3: P37's **RH**now goes downward while the **LH**is raised slightly higher with the forefingers flexing opposite the abduction and adduction thumbs movements.

Retraction: The gesture ends with P37 grasping the upright kit with the **RH**while the **LH**is raised to cover the mouth.

25b *Deictic Representational Gesture*

Portrayal of measurement is enacted by flexing of the forefinger opposite the thumb's abduction and adduction as movements out of and into the palm (Henley, 2010)respectively. The bidirectional arrows indicate both movements used to measure the assembly kit.

Deictic Enumerator

The ticking hands gesture is like the counting fingers gesture that is "…used to enumerate things by means of what can be construed as an iconic representation of a tally sheet"(Lucking et al. 2013, p.7). The enumerating hand is different because the assembly task requires identification of pieces from among many and the hand acts like a ticking pen that checks off each item in a list. The form and function is shown in the vignette (Figure25c) below

Preparation: P19 initiates the gesture with hands in the hamburger position

Stroke 1: P19's **RH**takes the piece from the **LH** and moves palm down with coupler-shaped digits and drops it in the workspace

Retraction 1: P19's **RH** retracts backwards briefly. Indicated by broken blue arrow

Stroke 2: P19's **RH** moves leftwards from the retraction, palm down with forefinger pointing downwards and tapping the first piece as a way of iconically ticking it of a list

Retraction2: The **RH** retracts backwards briefly

Stroke 3: P19's **RH** moves to the extreme left, palm down with the forefinger pointing downwards and tapping the second piece. This stage continues seamlessly to **Stroke 4**: P19's **RH** moves to the right immediately, palm down with forefinger pointing downwards and tapping the third piece. S3 weaves seamlessly into S4 yet each is a distinct move.

Retraction 3: The **RH** retracts finally to request a repeat of the instruction. Throughout the stages the **LH** remains in a hold position with the palm up.

25c *Deictic Enumerating Hand*

The "hamburger" position is taken from the Canadian American Sign Language (ASL) (Bailey et al., 2002, p.298) and is done with the right hand on top of the left picking something from it instead of covering it completely. The participant uses the deictic enumerating hand gesture to locate, select and set aside assembly pieces by iconically ticking-off every piece as mentioned in the instruction.

Appendix X: Listener facial actions and gestures in HHI (Chapter 5)

Appendix X.1 Neutral Face



Appendix X. 2 Felt Smile



Appendix X.3 Disgust

Appendix X.4 Compressed or Swallowed lips


Appendix X. 5 Slight Disgust

Slight Disgust

As outlined in Chapter 4.4.6, disgust is a feeling of aversion towards someone or something.

5- Slight Disgust

Slight disgust (Figure4) is given by the slight contraction of the muscle that wrinkles the nose AU9 and AU44 a separate strand of AU 4, narrows the eyes.

Appendix X.6 Picking Hands

Picking Hand –Ulnarly

This gesture is used for selecting assembly pieces that are very near listeners. The gesture combineswrists-based radial flexion and ulnar supination movements. The form and function of the gesture are shown in the vignette (Figure 5a) below.

Preparation: IH3 initiates this gesture after asking for repeat with both hands on the Table

Stroke: IH3 raises the back of the handwrist (ulnar extension) of the **RH.** Hethen moves the hand about 20^0 radiallyto the left (radial flexion), palm down, and uses a combination of the thumb and forefinger to pull up the piece.

Retraction: IH3's right wrist rotates about 45^0in a half-supination movement rightwards to make a presentation.

6a- Picking Hand - Ulnarly

The wrist and forearm are capable of three pairs of movements namely, pronation and supination, flexion and extension, and ulnar deviation or ulnar flexion and radial deviation or radial flexion(Allen, 2007; Hesse et al., 2003). However, the ones applicable to the picking hand-ulnarly gesture (Figure 5.15) are briefly outlined as follows, extension describes the movement of raising the back of the hand. Radial deviation or flexion describes the movement of bending the wrist to the thumb and supination is the movement rotating the forearm into a palm up position(Allen, 2007; Hesse et al., 2003).

Double Picking Hand

The listener uses one hand to pick up two assembly bits from different locations using this gesture. It is partially dialogic as the listener externalises the private speech going on within her as a nonverbal hesitation before making the right move. The form and function of the gesture are shown in the vignette (Figure6b) below.

Preparation: IH2 initiates the gesture with the **RH** in a hold while listening to the instruction.

Stroke Hesitation: IH2's **RH** moves palm down towards the extreme right of the kit layout but hovers over the green piece. The broken green arrow indicates the projected path.

Stroke 1: IH2's **RH** then moves palm down, leftwards with digits held coupler-shaped to pull up one of the pieces described in the instruction.

Retraction 1: IH2's **RH** immediately moves to the right, palm down with digits clasping the first piece. This is also a preparation for the next stroke.

Stroke 2: IH2's **RH** goes palm down and uses a combination of the thumb and forefinger to grasp and pull up the green piece.

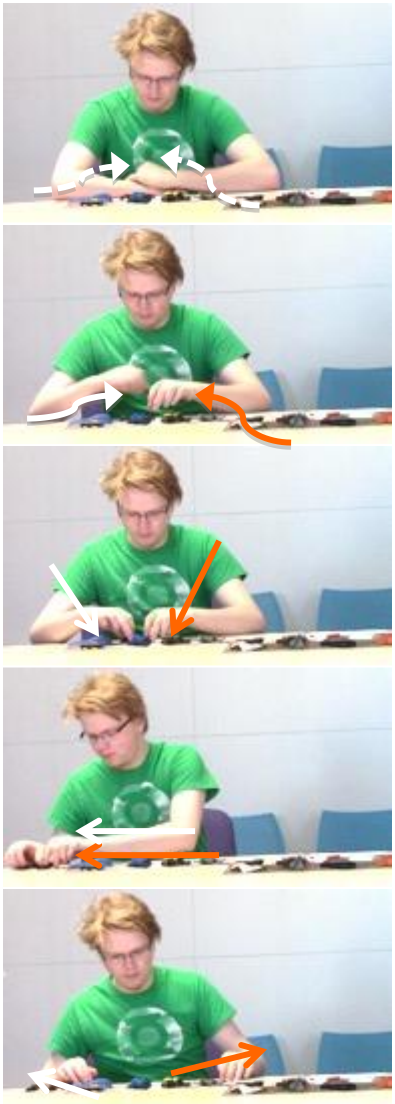
Retraction 2: IH2's **RH** moves back towards the sagittal axis to drop one of the pieces into the **LH**which rises and moves rightwards, palm lateral to receive the piece from the **RH**.

Retraction 3: Thereafter, both hands go halfway towards the workspace to execute a presentation of the pieces for examination.

5b- Double Picking Hand

Dual Picking Hands

This is a picking gesture done with two hands at the same time. It is the opposite of the alternate picking hands where both hands are used one after the other in simultaneous turns, as illustrated in the vignette (Figure6c) below.



Preparation: IH1 initiates the gesture from a rest position with both hands folded on the Table and the **LH** on top of the **RH**. This continues as IH1 moves both hands, palm down towards the kit layout. Broken arrows indicate projected paths

Stroke 1: The first stroke occurs when IH1 moves both hands forward, palms down with coupler-shaped digits and pulls up the pieces at once.

Stroke 2: The second stroke occurs when IH1's hands move rightwards to drop the pieces in the front right position.

Retraction: Both hands retract backwards to rest on the right and left side of the task area.

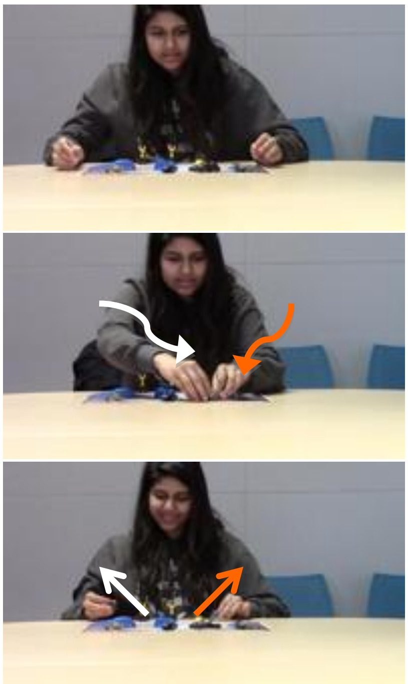
6c- Dual Picking Hand -a

The structure of the gesture phrase is shown below

<~~~~~~~~~/************/********/#####>

P S1 S2 R
………claws down on the desk……….

Another example of the dual picking hand is done by IH2 overleaf. The difference here is that while IH1 (Figure 5c) above drops the picked pieces on the right-hand side then retracts only one hand, IH2keeps the pieces and retracts both hands to their original positions. The form and functions of the gesture are outlined (Figure6d) below.



Preparation: IH2 initiates this gesture with both hands resting on the Table.

Stroke: IH2 moves both hands forward, palms down with coupler-shaped digits and in a combination of thumb, forefinger and middle finger to pull up the two pieces at once.

Retraction: IH2 retracts both hands backwards to rest on the Table holding the selected pieces.

6d- Dual Picking Hand-b

The gesture occurs within the last word in the instruction and the pause that comes thereafter before the next instruction. The structure is as shown below

<~~~ ~~~~~/***************/.-.-.-.-######>
P S R
...armour /

Picking Hand with Repairs

This gesture is executed with one hand. It has repairs embedded before the participant picks the correct pieces with retractions in between the strokes. The form and functionof the gesture are shown in the vignette (Figure6e) overleaf.

Preparation: IH1 initiates the gesture from a rest position and **LH** moves straight towards the kit layout as indicated by the broken green arrow and the yellow solid arrow.

Stroke 1: IH1's **LH** moves to the kit space, palm down with coupler-shaped digits to pull up a piece then realises that it is the wrong one

Stroke 2: IH1's **LH** moves slightly to the right and drops the wrong piece in self-correction

Stroke 3: IH1's **LH** moves further right palm down using in a thumb-forefinger combination to pull up the first correct piece

Retraction 1: IH1'**sLH** retracts briefly in preparation for the next stroke.

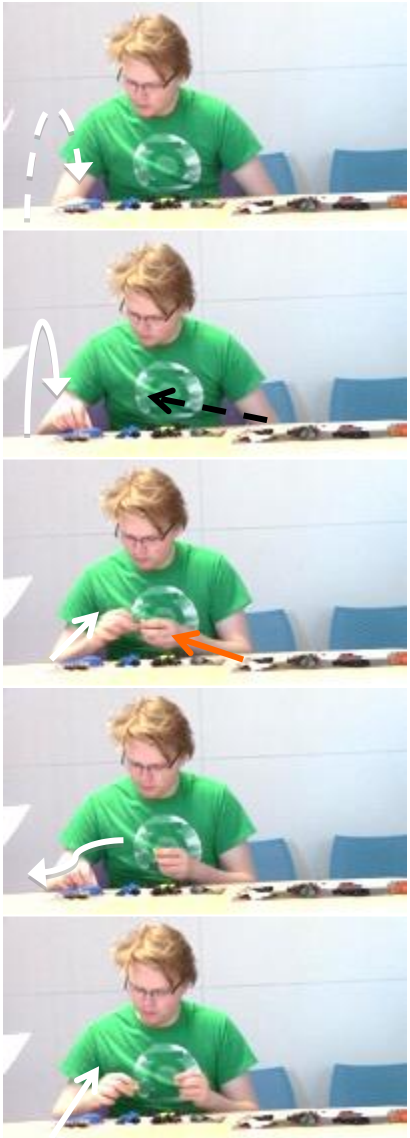
Stroke 4: This continues from retraction 1 and occurs when IH1's **LH** moves forward palm down using a thumb-forefinger combination to pull up the second piece.

Retraction2: The **LH** retracts backwards towards the sagittal axis.

6e- Picking Hand with Repairs

Picking Keeping and Holding Hand

This gesture combines keeping hand and holding hand gestures made by participants when selecting Lego pieces described in instructions with one hand. The form and function of the gesture are described in the vignette (Figure6f) below



Preparation: IH1 initiates the gesture with the two hands underneath the Table. The broken black arrow indicates the projected path of the **RH.**

Stroke 1: The Stroke occurs when IH1's **RH** comes up and moves forward towards the kit layout palm down with coupler-shaped digits to pull up the first yellow Lego piece. The broken black arrow indicates the projected path of the **LH**.

Retraction 1: IH1 retracts the **RH** upwards then leftwards and drops the piece in the **LH**.

Strike 2: The second stroke occurs when IH1's **RH** goes towards the kit layout, palm down with coupler-shaped digits to pull up the second yellow Lego piece. The **LH** is in a post-stroke hold.

Retraction 2: IH1's **RH** moves up in retraction while keeping the second yellow Lego piece and it comes to rest briefly

6f- Picking, Keeping and Holding Hand

The gesture occurs within a pause and the structure is as follows

<~~~ ~~/**************/.-.-.-.-######/**************/.-.-.-.-.-.##### >

P S1 R1 S2 R2

/00:02:01:259/ 00:02:01:899/ 00:02:02:579/ 00:02:03:179//00:02:03

Appendix X.7 Joining hand

Joining hand gesture is used to assemble part of the feet and it begins with picking up the connector with the right-hand then joining it to the feet of the Lego model with the left hand.The procedure is outlined in the vignette (Figure 7a) below.

Preparation: IH1 initiates the gesture from the **RH** dropping one connector while the **LH** is holding another connector.

Stroke 1: Preparation continues with IH1's **RH** picking up a foot with the **LH** still palm up grappolo-shaped and holding the connector.

Preparation 2: The **RH**(the receiving hand) moves up, palm grappolo-shaped laterally holding the foot in preparation.

Stroke 2: The **LH**(the joining hand) moves towards the **RH**, palm lateral, uses grappolo-shaped digits to attach the connector to the foot held by the receiving hand.

Retraction: The **RH** retracts to the front-right of sagittal axis while the **LH** retracts to a brief hold in preparation for the next move.

7a- Joining Hand - with LH

Another joining hands gesture is used to assemble the second foot enacted the same way with Figure6a, but the right hand gives while the left receives. The form and function of the gesture is shown in the vignette (Figure7b) overleaf.

Preparation: IH1 initiates the gesture from the retraction in the preceding gesture. The **RH** goes palm down with coupler-shaped digits to pull up the second connector while the **LH** is holding the foot. Preparation continues with IH1's **RH** moving the connector up and leftwards while the **LH** holds still palm up grappolo-shaped and holding the connector.

Stroke 1: The **RH** moves up, palm lateral with grappolo-shapeddigits holding and inserting the connector into the foot held in readiness by the **LH** or the receiving hand.

Stroke 2: IH1 uses both hands to put the assembled piece down.

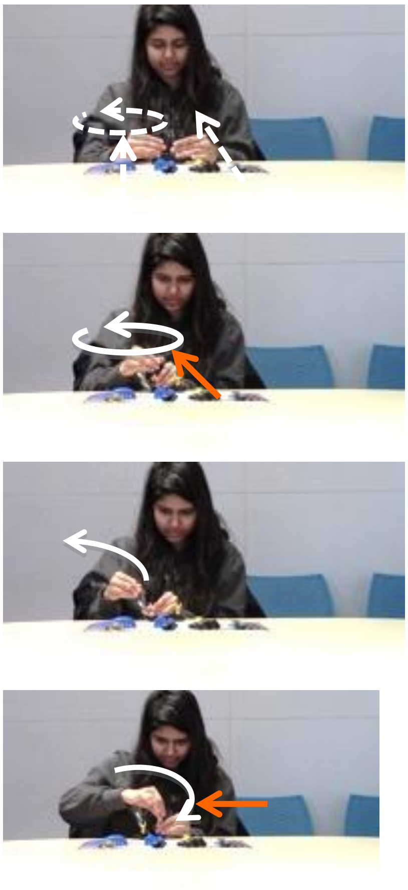
Retraction: This is done in two stages. The first stage occurs with both hands going up and towards the sagittal axis. The second stage concludes the retraction when both hands simultaneously come to rest and are clasped together on the Table.

7b- Joining Hand_ with RH

The participant (Figure7b) above uses the gesture to execute an instruction that has not been issued. IH1 has thus positively violated turn-taking protocol because he carried out the instruction correctly.

Joining Hand with Twists

This form of joining hand is circular in nature like Bressem's (2014) arched movement, measured in degrees for accuracy and is used to attach assembly bits from the side. The form and function of the gesture is shown in the vignette (Figure7c) below



Preparation: IH2 initiates the gesture with both hands holding the assembled kit. The broken green arrows indicate the projected paths

Stroke 1: The **LH** grasps and holds down the assembled kit while the **RH** moves up to grasp and twists the bit clockwise about 90^0 to get the desired orientation.

Retraction 1: IH2 withdraws the **RH** rightwards while the **LH** holds down the assembled piece as IH2 assesseswork done.

Preparation 2: IH2's **RH** moves leftwards to grasp the top of the kit in preparation for another stroke.

Stroke 2: IH2 executes stroke 2 by twisting the assembled kit clockwise about 90^0 with the **RH** as the **LH**holds the base up.

7c- Joining Hands with Twists

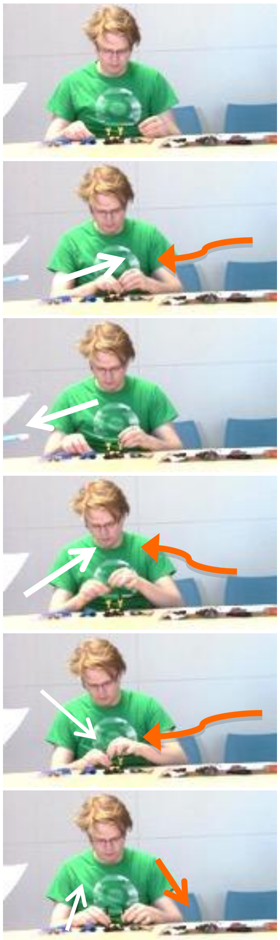
This gesture spills over to the next instruction but the participant also does not ask for a repeat. The structure of the gesture is shown below<~~~~/***/~~~~~~~#####/*******~~~~#####>

P S1 R to P1 S2 as R

…Twist……each one a little bit 90 degrees or so to the right. These are the legs

Dual Joining Hands

This gesture is used to assemble the thighs by attaching the ball joints of the black pieces to the yellow sockets of the Lego kit. The form and function of the gesture is shown in the vignette (Figure7d) below.



Preparation: IH1's initiates the gesture with the **RH** picking one piece while the **LH** is in a hold.

Stroke 1: IH1's **RH** moves towards the left, palm down using coupler-shaped digits in a thumb, middle and ring finger combination to grasp the receiving bit. The **LH** moves rightwards, palm down holding the joining piece to attach it in a top-down joining action into the one held down by the **RH**.

Retraction 1: This is a retraction to pick up another piece with the **RH** while the **LH** is holding the assembled piece in preparation for the next move.

Preparation 2: IH1 uses the **LH** to collect the piece from the **RH**

Stroke 2: IH1's **RH** goes down to grasp and hold down the receiving bit while the **LH** moves rightwards, palm down holding the piece to perform a top-down joining move

Retraction 2: The **LH** retracts downwards to hold one leg while the **RH** holds the other.

7d-Dual Joining Hand

This gesture spills over to the next instruction but the participant never asks for a repeat. The structure of the gesture is shown below and overleaf.

```
<~~~ ~~~~~~~~~~~~~~~~~~~~.-.-.-.-######>
Just attach them to the yellow socket./
P/************************************************/
S1
..so they are again pointing vertically the same way
/####**********~~**************************/
R to pick pause – Exchange of bit } semi-multi-segment phase
002:02:24:201 - 00:02:24:800 } Pause
/*************>
S2
So keep these…
00:02:26:561
```

The dual joining hands gesture contains semi-multi-segment phases (Kita et al., 1997) used to pick up and exchange assembly bitswithin the short pause when executing the main gesture.