Moving together: the organisation of non-verbal cues during multiparty conversation

Battersby, Stuart Adam

Moving Together: The organisation of non-verbal cues during multiparty conversation

Stuart Adam Battersby

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2011

Declaration

I declare that the work presented in this thesis is my own work carried out under normal terms of supervision and that the research reported here has been conducted by myself unless otherwise indicated.

Stuart A. Battersby

London, June 3rd 2011

Moving Together: The organisation of non-verbal cues during multiparty conversation

Stuart Adam Battersby

Abstract

Conversation is a collaborative activity. In face-to-face interactions interlocutors have mutual access to a shared space. This thesis aims to explore the shared space as a resource for coordinating conversation. As well demonstrated in studies of two-person conversations, interlocutors can coordinate their speech and non-verbal behaviour in ways that manage the unfolding conversation. However, when scaling up from two people to three people interacting, the coordination challenges that the interlocutors face increase. In particular speakers must manage multiple listeners. This thesis examines the use of interlocutors' bodies in shared space to coordinate their multiparty dialogue.

The approach exploits corpora of motion captured triadic interactions. The thesis first explores how interlocutors coordinate their speech and non-verbal behaviour. Inter-person relationships are examined and compared with artificially created triples who did not interact. Results demonstrate that interlocutors avoid speaking and gesturing over each other, but tend to nod together. Evidence is presented that the two recipients of an utterance have different patterns of head and hand movement, and that some of the regularities of movement are correlated with the task structure.

The empirical section concludes by uncovering a class of coordination events, termed simultaneous engagement events, that are unique to multiparty dialogue. They are constructed using combinations of speaker head orientation and gesture orientation. The events coordinate multiple recipients of the dialogue and potentially arise as a result of the greater coordination challenges that interlocutors face. They are marked in requiring a mutually accessible shared space in order to be considered an effective interactional cue.

The thesis provides quantitative evidence that interlocutors' head and hand movements are organised by their dialogue state and the task responsibilities that the bear. It is argued that a shared interaction space becomes a more important interactional resource when conversations scale up to three people.

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2011

Contents

List of Figures

List of Tables

Related Publications

Conference papers

Battersby, S. A. and Healey, P. G. T. (2010a). Head and Hand Movements in the Orchestration of Dialogue. In Ohlsson, S. and Catrambone, R. (eds) *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, pages 785 – 790. 11th - 14th August, Portland, Oregon

Battersby, S. A. and Healey, P. G. T. (2010b). Using head movement to detect listener responses during multi-party dialogue. In Kipp, M., Martin, J-C., Paggio, P. and Heylen, D. (eds) *Proceedings of LREC Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analysing Multimodality*, pages 11 – 15. 18th May. Valletta, Malta

Healey, P.G.T. and Battersby, S.A. (2009). The Interactional Geometry of a Three-Way Conversation. In Taatgen, N. and van Rijn, H. (eds) *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*, pages 785 – 790. 29th July - 1st August. Amsterdam, Netherlands.

Battersby, S. A., Lavelle, M., Healey, P. G. T., and McCabe, R. (2008). Analysing Interaction: A comparison of 2D and 3D techniques. In Martin, J.C., Paggio, P., Kipp, M. and Heylen, D. (eds) *Proceedings of LREC Workshop on Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*. pages 73 – 76. 27th May, Marrakech, Morocco.

Position papers & abstracts

Battersby, S.A., Healey, P.G.T., Lavelle, M., Eshghi, A. and McCabe, R. (accepted) Non-verbal cues to recipient roles in dialogue. In *Proceedings of the 21st annual meeting of the Society for Text and Discourse*. July 11th - 13th. Poiters, France.

Battersby, S.A., Healey, P.G.T. and Eshghi, A. (2010) Triangulations: Simultaneous engagement in multiparty interaction. In Deppermann, A, Spranz-Fogasy, T and Tanrisever, A. (eds) *Proceedings of the International Conference on Conversation Analysis*, page 225. July 4th - 8th, Mannheim, Germany.

Battersby, S.A. and Frauenberger, C. (2008) Shared 3D interaction spaces with humans and avatars: A summary of motion capture case studies and an introduction to an experimental platform. In *Sussex HCT seminar*. November, Sussex, UK.

Battersby, S.A. and Healey, P.G.T. (2008a) Sharing 3D Spaces: A summary of motion capture case studies showing the use of 3D shared spaces in face to face interaction. In *Shareable Interfaces for Learning Workshop*. September 11th - 12th, Sussex, UK.

Battersby, S.A. and Healey, P.G.T. (2008b) Inhabiting 3D Interaction Spaces. In *Language, Communication Cognition*, page 14. August 4th - 7th, Brighton, UK.

Acknowledgements

Coming soon...

Chapter 1

Introduction

Consider the following everyday scene consisting of three people, Ann, Bob and Claire chatting: Ann and Bob have just walked together from the tube station in the rain to the office that they share with Claire, and are now engaged in informal discussion with her. The chat revolves around the weather in London (not an uncommon topic), in particular the recent journey from the tube in the rain during which Bob was covered in water as a bus drove past him through a puddle. They are stood in the hallway outside their office. In order to hold a conversation, there are many challenges which the three interlocutors must collaboratively address. Before any speech happens, they had to advertise their availability and lack of engagement in another conversation and then collaborate to build their new conversation, ensuring that everyone is mutually aware of each others' involvement. During the dialogue they must maintain this involvement, coordinating their speech and bodies to maintain a sense of continued attention. The three must organise their behaviour such that whoever speaks at that time works with their intended addressee, and vice versa. Unlike a two person conversation, speakers must manage multiple addressees of their utterances. As Bob tells Claire about the journey, he must ensure that Claire understands and accepts that she is being addressed and continually understands what is being said. This is not a one-way relationship, but collaborative; the responsibility also falls upon Claire to demonstrate this and as such the pair must collaborate to achieve this. But what about Ann? She is engaged in the conversation but is neither the speaker nor the direct addressee of the utterance. She should maintain a level of engagement in the discussion. She is privy to the shared experience of the walk from the tube station with Bob, so she should ensure that he tells this story correctly. Indeed, she can be treated as a supporting resource for Bob; he can call on her for details of the journey when he gets stuck. When this happens, Ann must act accordingly. Moreover, Claire must also recognise the shift of the conversational relationship between Ann and Bob. How is it that all three interlocutors maintain the flow and engagement of this unfolding dialogue? In particular, how does Bob as the speaker coordinate the multiple recipients of his utterance?

As well as speech, they have at their disposal a repertoire of non-verbal cues which they can deploy. The interlocutors can use their hands communicatively. For example, Bob can make use of a gesture to both depict the tube station, and also to make it known to Ann that he needs her assistance. The repertoire also contains cues which make use of the face, eye gaze and head orientation. Ann can give Bob his desired feedback by making him the focus of her eye gaze and head orientation. As the three are co-located in a structured shared space, they have mutual access to each other's actions. As Bob retells the bus incident to Claire, gesturing the motion of the bus through the puddle, she catches Ann's eye, causing them both smirk at Bob's unfortunate soaking. Bob notices their shared gaze and exclaims 'Hey! It's not funny'.

These uses of the body come under the definition of non-verbal behaviour. Mehrabian (1972) defines non-verbal behaviour in its narrow sense to 'refer to actions as distinct from speech. It thus includes facial expressions, hand and arm gestures, postures, positions, and various movements of the body or the legs and feet.'

This thesis aims to explore how the shared space supports the coordination of a conversation. It is expected that, as conversations with groups of people are more challenging to manage than conversations with two people, groups will maximise the potential to make space relevant. Thus, the thesis examines the organisation of non-verbal communication which support the smooth unfolding of multiparty dialogue with the shared space. Work over recent decades has identified communicative patterns of non-verbal behaviour, usually making use of dyadic (two person) data. Much of this comes from the field of conversation analysis (see e.g. Streeck (1993, 1994)) and experimental psychology (see e.g. Bavelas et al. (2002a)). More recently there have been efforts from more computational fields, concerned with building complex models of interlocutor movement (see e.g. Ashenfelter et al. (2009)). Whilst the complexities of multiparty data have been addressed to an extent in the verbal domain (see e.g. Eshghi and Healey (2007); Goffman (1981); Clark and Carlson (1982)) there are relatively few non-verbal counterparts (although see Kendon (1973); Özyürek (2002); Jokinen (2010)).

The work presented in this thesis contributes to an understanding of the organisation of interlocutor head and hand movement with respect to the roles and responsibilities that they bear within a multiparty dialogue. A finer granularity in the definition of a shared interaction space is presented, and an argument is made that a shared interaction space becomes a more important interactional resource when conversations scale up to three people.

1.1 Research questions and approach

This thesis will address three main questions concerning the organisation of non-verbal behaviour within the context of multiparty dialogue. They are:

1. Do patterns of detectable inter-person head and hand movement exist within a group of interacting people that are systematically different from the inter-person patterns measured across non-interacting individuals?

2. Are there global patterns of head and hand movements that relate to the unique structure of a multiparty dialogue in terms of its constituent interlocutors' dialogue roles?

3. Are there communicative behaviours that intrinsically rely on their deployment in a mutually accessible shared interaction space and, if so, what do they look like?

These questions will be answered after performing a review of the existing literature. This review will be broken into two stages: literature which addresses the relationship between speech and non-verbal behaviour and literature which examines the communicative properties of non-verbal behaviour, oriented around the spatial organisation of dialogues. With this base the three questions can be addressed empirically. The first question allows for a systematic test of the communicative nature of non-verbal behaviour, accounting for the possibility of statistical artefacts. The following question opens up the multiparty dialogue to expose its organisational structure and explores the patterns of interlocutors' behaviour that relate to their responsibilities within the dialogue. The final question concerns the use of the shared interaction space within the dialogue.

The approaches used to answer these questions sit on a scale which, at one end, features methodologies that are extremely naive and quantitative, whilst at the other end sit methodologies that are interpretive and qualitative. This scale combines novel analysis methods making use of three dimensional motion capture with traditional video based techniques. The varying methodologies require much explanation and hence have been provided with a chapter in their own right (Chapter 4).

1.2 Thesis overview

This thesis is structured around eight chapters. The first chapter, this chapter, introduces the types of problems to be addressed in the coming thesis and states the main research questions.

Chapters two and three constitute the literature review. Chapter two begins the review by examining work which uses the individual as the basic unit of analysis, and hence concerns the relationships between an individual's speech and their non-verbal behaviour. Driven by the content of the literature, this will focus mostly on co-speech gestures and their denotational features. The third chapter moves to literature which examines interlocutor behaviour within dialogue and thus considers non-verbal behaviour as organised by the interaction. The discussion is organised around the spatial structures that delimit an interaction, and particular attention is paid to the dialogue structure and the coordinating behaviours that interlocutors make use of. Collaborative use of head and hand movement within a shared interaction space is examined in detail. The literature studied here allows for a critique of the previous chapter, arguing that the individual approach is insufficient for studying dialogue. Contrasts will be made between dialogues which have access to the shared space and those that don't, and some questions concerning the shared space will be presented.

Chapter four details the common methods used through the thesis. Software has been created as part of this thesis to analyse motion capture and video annotation data. Details of the motion capture system, the analysis methods developed and the corpora of data used through the thesis are documented here.

Chapters five, six and seven constitute the empirical work of the thesis. The starting methodology is particularly naive, making use of machine analysis then moving through to a more interpretive human annotation analysis. The fifth chapter provides a grounding for the rest of the empirical work by testing the coordinated nature of non-verbal behaviour. By comparing the triadic interactions held in the corpus to artificially created interactions in which the constituent members came from different original interactions, coordinated behaviours are effectively factored and tested. The next chapter, Chapter six, opens up the multiparty dialogue to expose its structure. The organisation of the interlocutors' behaviour is analysed with respect to their roles and responsibilities within the dialogue. Chapter seven moves to explore the shared interaction space in more detail. It is split into two studies: The first is rather open and documents intrinsically spatial features of the interactions. The findings from this study are then used to motivate a coding scheme for annotation in the latter study.

Chapter eight summarises the findings from the thesis as a whole and ties the work presented in with the current state of the field and related fields. Implications of the new findings are raised and discussed, along with some open questions that have been created or remain unanswered.

Part I

Literature Review

Chapter 2

Encoding and decoding non-verbal signals

2.1 Introduction

This section opens the literature review by studying work on non-verbal behaviour which makes use of a different analytical unit to that of the main thesis. The work presented here takes the individual, as an encoder or decoder of non-verbal signals, as the basic unit of analysis. Questions about the relationships between a spoken utterance and body movements tend to be the their main focus. This type of work has been segmented from the rest of the literature review as the types of questions asked, the data used and the conclusions made are interpreted differently; the work in the rest of the thesis specifically focuses on the non-verbal interaction with others. However, a thorough review is still important because a) it is a wide field of study related primarily to hand gesture, b) the literature which does have a focus on interaction adopts a large amount of its terminology from here and c) a critique can be performed to highlight what is missing from the approach taken when interaction is not considered. How individuals make use of their bodies, in particular their head and hands, and the space around them when producing an utterance will be the main concerns.

Some of the data used come from individuals in dialogues (e.g. Kendon, 2004). However, the data presented in many of the reported studies come from a narrative monologue paradigm (e.g. McNeill, 1992). Commonly this involves a participant watching a cartoon and, following this, recounting the story to either a camera or a confederate. In this field of psycholinguistics this is a valid approach as the main aim is to study the cognitive aspects of gesture production and its relationship with concurrent speech, in particular the role of gesture in conveying the content of the accompanying speech. The suitability of this approach for the study of dialogue will be questioned at the end of this chapter.

This chapter will mainly be concerned with covering the literature that addresses an individual's deployment of speech and non-verbal behaviour, and the associated relationships between them. As such, this will naturally lead to a focus on factors relating to their content. These will form the first part of the chapter, structured around hand gesture and head movement. Methods of classification and discrimination of behaviours, and the behaviours' functions with respect to speech will be addressed. Following this, the chapter will move from questions of 'what' to questions of 'where' by studing the spaces that are used and exploited around the body. Literature which covers the structure of qualitative space will be reviewed, and the use of a person-centric frame of reference will be highlighted. A link will be drawn to sign language which is intrinsically spatial and includes person-centric spatial structures within the definition of the language.

2.2 Denotation and classification of non-verbal behaviour

An individual's non-verbal behaviours are intimately intertwined with their speech. Together they have been said to form part of a composite signal (Engle, 1998). In this view non-verbal behaviour is said not to be an independent 'channel' to speech; instead a representation is conveyed through a composite signal composed of both verbal and non-verbal actions. Kendon demonstrates this using the following excerpt:

Speaker: He used to go down there and throw

Gesture: Right arm extended, the fingers are curled with the tip of the thumb making contact with them. Moved by the wrist extension, the hand moves outward, rapidly, twice.

(0.3)

Speaker: ground rice over it

He states that it is not the case that the gesture represents the same thing as the speech. In his example, the verb 'to throw' does not describe how the throwing is done. It is only when the verb is considered with the gesture (which was prepared for in advance of the verb), that it becomes apparent that the throwing was actually a type of scattering. A similar argument is made in McNeill's later work (McNeill, 2005).

In other situations, body movements including gesture can be referred to directly in speech (Gullberg, 1995). Consider the passage and example from Kendon (2004):

"She says: 'She used to play the organ then and of course I think there was someone who used to have to stand there doing this (0.9 sec) and that was sort of great fun.' Here, as she begins to say 'I think there was somebody...' she holds both hands in front of her, palms down, formed as loose fists, and she moves them down in unison, also bowing her body forward repeatedly, in an enactment of someone operating an organ pump."(Kendon, 2004)

In this example, the reference of 'this' was to an image created with the hands and movement of the body. Here the gesture is explicitly referenced in the speech, and together they construct the utterance and its intended meaning. The claim here is that language is multimodal. If one were simply to read a transcript of spoken words the full meaning would not be conveyed. However, an argument has been made that gestures do not appear to add anything extra than that specified in the speech. In a study which manipulated the clarity of referents, So et al. (2009) found that if the referent was unclear in speech (for example, using the words 'the man' when describing a scene in which two men are present), instead of clarifying the referents the gestures will be equally unclear. One possible explanation to this apparent contradiction is that the findings in So et al. (2009) are an artefact of the data used. Participants described video scenes to the experimenter, who knew which man was which. Had this been a real dialogue in which the recipient did not know who the referent was, they would have been able to demonstrate their lack of understanding to the speaker, perhaps leading to different gestures.

It is not only in depicting content that the modalities work together. It has been argued that the speaker coordinates the moment-by-moment temporal synchrony of their multimodal utterance. By analysing high speed recordings of a speaker's movements, Condon and Ogston (1966) found that speech and body movements were in prosodic alignment. Body movements are used to mark out tonic stresses in an utterance (Bull and Connelley, 1985) and it has also been claimed that the movements of the body occur more around times in which the speech is disfluent (Christenfeld et al., 1991). They show that gestures are more likely to occur at times when the speaker has trouble encoding their words, resulting in pauses in speech.

2.2.1 Hand gesture

2.2.1.1 Definition

Gesture is defined[1] as ‘a movement of the body or limbs as an expression of feeling’. The first thing to note in this definition is that gesture is not just movements of the hands. Whilst the hands are the most commonly discussed gesturing limbs, one is able to gesture with all body parts including the head. The second part of the definition states that these movements express feeling. There is no mention of gestures as coordination mechanisms, rather that they are a medium used to express content. This definition is in line with the literature to be reviewed in the current section. The remainder of this section will focus predominantly on hand gestures, a reflection of the focus in the literature.

2.2.1.2 Discriminating gestures

In order to be able to identify a gestural unit it is necessary to understand the constituent components of the gesture. Gestures are most commonly triphasic (McNeill, 1992). Their deployment can be broken down into three stages: preparation, stroke and retraction (sometimes called recovery) (McNeill, 1992; Kendon, 2004). Only the stroke is mandatory, the preparation and retractions phases are optional. The preparation phase sees the hands move up from their resting place towards the front of the speaker. The stroke comes next. McNeill states that it occurs at the phonological peak of the accompanying speech where the gesture and speech are semantically matched (McNeill, 1992). Kendon states that it is the expressive phase of the gesture and is often matched with the ‘high information’ word of a phrase (Kendon, 2004). Following the stroke, the hands fall back to their resting place in the retraction phase (see Fig. 2.1).

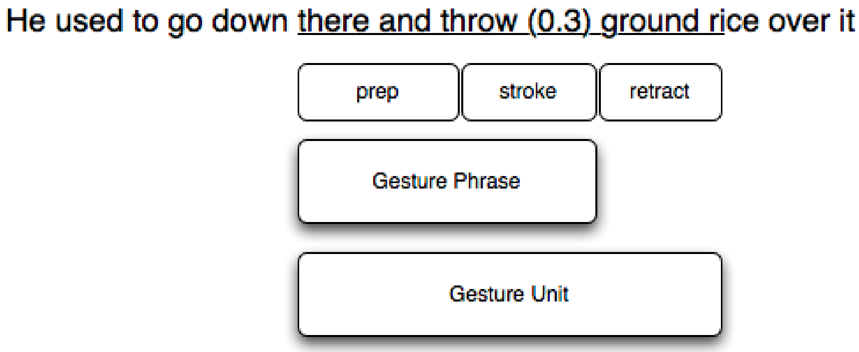

He used to go down there and throw (0.3) ground rice over it
prep
stroke
retract
Gesture Phrase
Gesture Unit

Figure 2.1 – The structure of a gesture as defined by Kendon (2004)

[1] Definition is from the Oxford English Dictionary, second edition

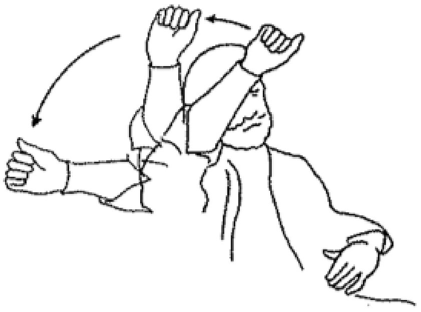
These phases are used to distinguish gestures. A gesture unit is defined as all the movement from the time that the hands move from their resting position to the time they return to their resting position. Each gesture unit may contain one or more gesture phrases (Kendon, 2004). The gesture phrase is made up of the preparation and stroke phases. Each gesture phrase can only have one stroke, and does not contain the retraction. Therefore, in the canonical form of a gesture described above, the whole sequence of preparation, stroke and retraction constitutes the gesture unit, however only the preparation and stroke fall into the gesture phrase (see Fig. 2.1). As the retraction has been kept seperate from the gesture phrase, it is possible for there to be more than one gesture phrase in the gesture unit. In these situations we see multiple preparations and strokes joined together, with a final retraction phase when the hands return to the rest position.

Given that gestures can be individuated in this way, it is possible that a common repertoire of gestures exist which conform to some conventions within a social group. However, unlike sign language which is a well defined language, co-speech gestures are unconventionalised (McNeill, 2000). There is no requirement for a gesturer to depict something in a particular way or with a specific hand shape other than the properties of the subject under depiction. The exceptions to this are emblematic gestures such as the OK sign, or V for victory which hold varying, but culturally specific conventions (see Section 2.2.1.3).

2.2.1.3 Taxonomy

The literature on gesture regularly refers to different types or classes of gesture. There is not a single authoritative taxonomy of gesture; a thorough discussion of many is offered by Kendon (2004). This section will be oriented around the taxonomy documented by McNeill (1992) as this is used most regularly in the co-speech gesture literature, however where possible links will be made to others' work. It is important to reiterate that McNeill's classification uses the individual as the frame of reference and as such is not designed to consider factors pertaining to interaction. Indeed, McNeill (1992) states that gestures 'belong, not to the outside world, but to the inside one of memory, though, and mental images'.

The first class of gesture to be described is the iconic gesture. McNeill states that an iconic gesture displays some concrete reference to the semantic content of the speech. The gesture created offers another representation of the same scene described by the speech. For example, in Fig. 2.2 McNeill is showing that the gesture of gripping something and pulling it back is showing aspects of the same scene described by the speech. The gesture is timed with the speech such that its stroke falls at the same time as the part of the utterance containing the meaning. Similar notions of iconicity are described by Ekman and Friesen (1969) as iconic codes. They assign this code to non-verbal acts which look in some way like what they mean and hence their appearance can help with their decoding.



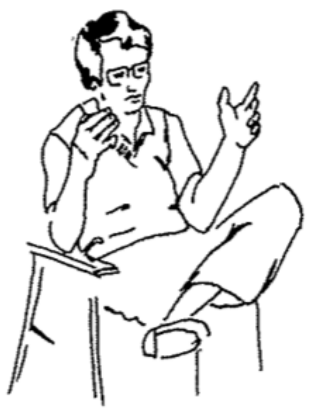
Speaker: and he bends it way back

Gesture: Iconic. Right hand appears to grip something and pull it back from front to own shoulder

Figure 2.2 – An example of an iconic gesture taken from McNeill (1992)

Unlike iconics, the metaphoric gesture does not contain a concrete reference to the speech. However, it does form a pictorial representation of an abstract concept from the speech. The image created is an abstract representation of, for example, some concept, event, or knowledge. In Figure 2.3 McNeill shows a speaker referring to the abstract genre of a cartoon. This is depicted as a bounded object created with both hands. Thus, whist the gesture creates an image, it does so without any direct imagistic reference to the speech. This class of gestures is harder to align with those of other taxonomies, although it does share properties with the arbitrary codes of Ekman and Friesen (1969), which are said to 'bear no resemblance to what they signify'. This more general notion could also encapsulate the next class of gestures that McNeill (1992) defines: the beat gesture.

Beat gestures can mark out segments of discourse and often fall into prosodic alignment with speech. Rather than being triphasic, as the iconic and metaphoric gestures are, beat gestures are usually biphasic. For example, a simple up & down motion (McNeill, 1992). Beat gestures have



Speaker: it was a Sylvester and Tweety cartoon

Gesture: Metaphoric. Hands rise up and offer listener an "object"

Figure 2.3 – An example of a metaphoric gesture taken from McNeill (1992)

also been documented in early gesture work as 'batons' (Efron, 1941) and can serve to emphasise the structure of speech.

A further type of gesture is the pointing, or deictic (McNeill, 1992), gesture. These gesture do not depict any form of imagery. Pointing gestures indicate objects, a location or a direction by projecting a line into space from the most extended body part (it could be that this is not the hands) (Kendon, 2004). McNeill et al. (1993) break down the process of pointing into three elements (see Fig. 2.4):

- *Origo* - The point of origin.
- *Target* - The referent object, which may be concrete or abstract.
- *Trajectory* - This is a line linking the origio to the target.

However, pointing gestures do not only locate existing referents. McNeill et al. (1993) carried out a study based on narrative discourse in which pointing gestures were seen to have a referent in empty space, that is, people were pointing to empty spaces around them. It was reported that the deictic gesture actually creates the referent in the gesture space and, although these referents are abstract, they are salient with respect to the content of the speech.

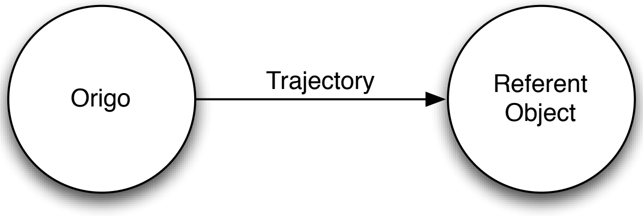


Figure 2.4 – Three elements of a point as defined by McNeill et al. (1993)

McNeill et al. showed that these spaces can then be used referentially throughout the discourse as is usually done with pointing. They highlight this with the following example:

Speaker: and in fact a few minutes later we see the artist

Gesture: Points to left side of space

Speaker: and uh she looks over Frank's shoulder at him

Gesture: Points to the left side of the space again

McNeill et al state that in the first utterance the point creates the referent 'the artist' at the empty space pointed to. In the second utterance the space, which is no longer 'empty' is used referentially to determine who is being looked at. Similar notions are seen in sign language; Bellugi and Kilma (1982) show that when signers are referring to persons not present they will create a referent for them in their signing space and point at this space when they need to make the reference.

One remaining type of gesture, defined by Ekman and Friesen (1969), which does not fit cleanly into the above taxonomy but rather spans a number of categories must be mentioned. It is the emblem gesture. These are gestures which, rather than being deployed along side speech, may completely replace a word. They tend to only be recognised within cultures. An example is the OK symbol.

Holding Gestures When considering the triphasic gesture structure, McNeill (1992) suggests that there are times when a gesture does not move directly through the preparation phase into the stroke, or does not move directly from the stroke to the retraction phase. In these cases holds can be performed both pre-stroke and post-stroke. He states that a pre-stroke hold is when the hands are held still after the preparation phase awaiting the stroke to begin. The post-stroke hold occurs at the final position of the stroke, awaiting the beginning of the retraction phase. He attributes these types of holds to delays in the stroke, perhaps due to delays in the spoken utterance.

Kita et al. (1998) adopt a slightly different approach to holds. Whilst they still suggest that a non-moving hand (during a gesture) is a hold, they propose that there can be a dependent or an independent hold. A dependent hold requires, and is dependent upon, the stroke phase, in line with McNeill's definition. However, with an independent hold, the hold phase of the gesture can replace the stroke phase and form what they refer to as the expressive phase. From this perspective it would seem that a gestural hold is attributed more significance than simply waiting for a delayed stroke.

Despite their differences, both of these definitions of holds refer to the production and deployment of the gesture, and hence are discussed at the level which concerns the temporal structure of the gesture. Other possible functions of holds which facilitate an unfolding dialogue are discussed in Chapter 7.

2.2.2 Head movement

The head is an important locus for non-verbal signals. It is from the head that we speak with our mouths, and we use both head and eye movement for gazing and looking at things. The face also displays information regarding our emotional state, although a full discussion of this is not within the theme of this thesis. Instead, head movement and gaze orientation will be examined. Of particular interest are head gestures and their use to mark out sections of a spoken utterance.

2.2.2.1 Head gesture

As mentioned, it is not only the hands that are used for gesture; many other parts of the body including the head can be used also. Head gesture can occur in a number of forms. Kendon (2004) notes in his discussion of pointing that this can be done with the head. In addition the head can produce gestures that are classed as culturally recognised and hence emblematic. McNeill (1992) reports an example of this in which a head toss gesture holds the meaning of negation within Greek society. A more commonly recognised head gesture is that of the head shake. This is very often equivalent to the word 'no', but may also show or imply negation, or be used by speakers as a self-correction of their own utterance (Kendon, 2002).

A much studied feature of head gestures is their relation to a speaker's utterance. Kendon (2004) documents an example[2] in which non-deictic head gestures accompany hand gestures and speech. In two gesture phrases, the head is seen to perform a lateral shake and then a lowering. The lowering gesture coincides with the ending of the utterance. However, this lowering is by no means consistent across all utterances. If the utterance was a question then it is likely that the speech will rise in intonation towards the end and be accompanied by a upward head movement (Argyle, 1975).

By measuring head movement and speech patterns, it has been show that speakers nod more than non-speakers (Hadar et al., 1983). The pattern of nodding produced can vary within the utterance. As with some hand gestures, head nodding has been seen to occur with stressed words (Dittmann and Llewellyn, 1969; Bull and Connelley, 1985) and can be used by the speaker in order to emphasise sections of their utterances (Argyle, 1975). Dittmann and Llewellyn (1969) comment further on the movement-speech relationship and suggest that body movements, including movements of the head, are caused by trouble with speech encoding. They state that tension causes encoding errors and that these "spill over into the motor sphere".

In Kendon's work, head gestures are transcribed in the same way as hand gestures and are reported as gesture phrases. However, it is not clear if they follow the triphasic structure common to hand gestures.

2.2.2.2 Gaze behaviour

An individual's gaze also relates to the structure of their spoken utterance. Argyle (1975) notes that when speakers reach a grammatical pause, they are likely to shift their gaze upwards. As mentioned in the discussion of head gesture, the head can be used diectically. This is commonly through the use of gaze. In studies which involve monologue descriptions of basic scenes or images, subjects gaze at the object in the image before they verbally identify it (Griffin and Bock, 2000; Meyer et al., 1998). The gap between gazing and speaking is greater if the subject is asked to label the object incorrectly (Griffin and Oppenheimer, 2006). These findings lead the authors to match gaze behaviour with internal word retrieval behaviour. Indeed, gaze is said to aid memory. Subjects often gaze at the space where an object is, was or is expected to be to remember its properties (a phenomenon known as spatial indexing, see Richardson et al. (2007) for an overview).

[2] The example can be found on page 121, example 3

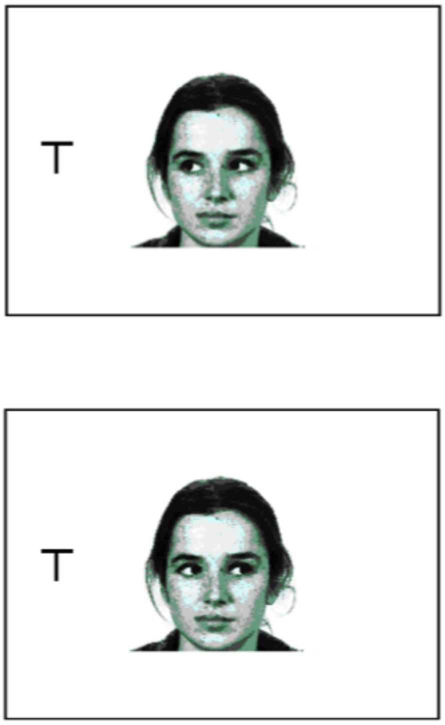


Figure 2.5 – Two conditions of stimulus material used in Driver et al. (1999) showing the eyes oriented towards the letter to be detected and the eyes oriented away from the letter

In an attempt to identify the causes of gaze shifts, Driver et al. (1999) and Langton and Bruce (1999) measured the gaze behaviour of subjects when exposed to various images. They claim that people orient their gaze in the direction of other people's orientation. The stimulus given to participants were pictures of others' faces gazing in a particular direction, an example of which is shown in Figure 2.5. A randomly chosen letter was placed on one side of the pictorial stimulus. Subjects' response times to determine the letter, either in the direction of the stimulus gaze or not, were recorded. Faster responses were found when the direction of gaze in the stimulus image matched the location of the letter. Whilst these studies make this claim based on a 'social' stimulus, the use of photos of people does not represent the true presence of an interlocutor and hence the social claims must be interpreted with caution. Thus the claims should be interpreted in the context of an individual's gaze behaviour.

A discussion of the patterns of gaze behaviour found within dialogue, including those that provide cues to focus of attention and those that demonstrate joint focus of attention, will be explored in Chapter 3.

2.3 Space

The chapter will now turn from the content of non-verbal behaviour to examine where these behaviours take place. When considered from the perspective of an individual, the space around the body is structured such that it can be exploited for these behaviours and a person-centric frame of reference is used to delineate the structure. A parallel can be drawn here to sign language in which a highly structured use of space is seen and is indexed by the body. These topics will be examined because, as the thesis progresses to examine interaction, the importance of structured space within a dialogue will become apparent.

2.3.1 Space around the body

Kendon (2010) discusses a use-space which is described as the space that an organism uses for some activity. His examples include a bird requiring a use-space to build a nest, or a cat requiring a use-space in its sleeping basket. The properties of these use spaces will vary depending upon the use (contrast the space for the bird and the cat). Moving specifically to humans, Vine (1975) offers the concept of an individual space which he states is a mobile, body centred area and typically extends in front of the body. Scheflen (1976) discusses an area called a segment which describes the 'space that the orientation of a body region covers and claims'. Kendon suggests that there is a space which extends in front of a person in which they carry out their activities named the *transactional segment* (Kendon, 1990, 2010). The actual size and boundaries of this space are fuzzy as it is dependent upon the activity currently being performed (for example, it would be small for activities such as reading a book, but larger for an activity such as watching the television). It projects forwards from the orientation of the lower body. Figure 2.6 shows this space. As this transactional segment is defined by the lower body, it is possible for parts of the upper body to be temporarily turned outside of the transactional segment without shifting its placement. This space holds exclusive access rights for the owner; if another individual is to enter or use this space uninvited it would be a marked event and may appear rude.

This space which falls in front of the body can be further subdivided. McNeill's findings, based upon data from six people giving narratives of a cartoon scene, show that there are specific regions in which the different types of gestures specified in his taxonomy fall (McNeill, 1992). The space is divided into the 'center-center' (directly in front of the chest), the 'center' which surrounds this, the 'periphery' around this and finally the 'extreme periphery'. According to

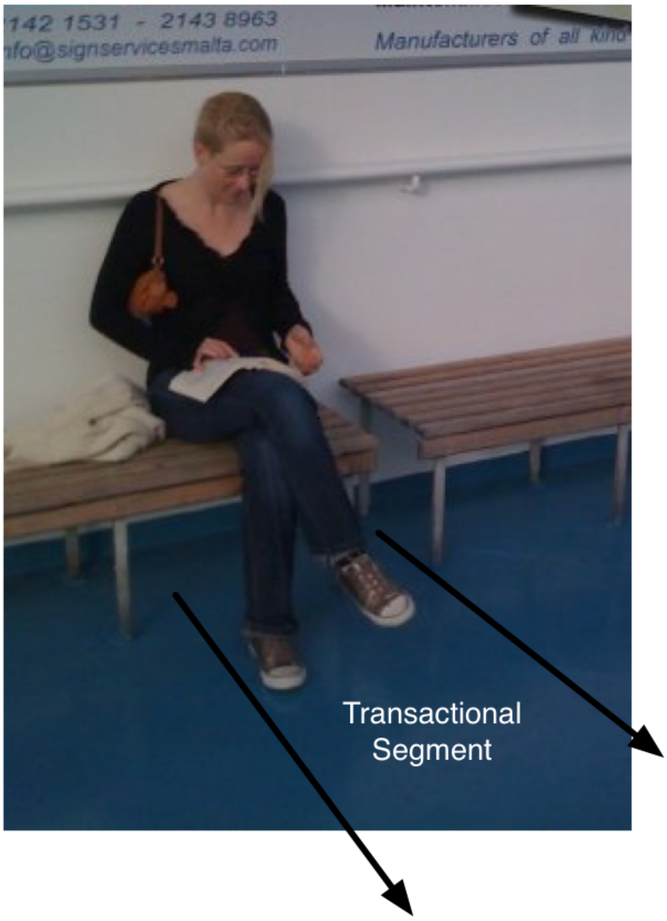


Figure 2.6 – The shape and location of an individual's transactional segment

McNeill, there is a tendency for iconic gestures to fall in the center-center space with metaphoric gestures in the lower center. Pointing falls in the periphery. There is no deployment segment for beat gestures that is common for all individuals, instead each speaker tends to be idiosyncratic in where they cluster their beat gestures. The frequency with which gestures fall into these segments has been shown to vary between cultures (Levinson, 2003).

2.3.2 Frame of reference

A key point to make note of is the frame of reference used when describing these spaces. The common factor between all the work on these spaces (and as an extension, the non-verbal behaviour described in this chapter which falls within the spaces) is that they use a person-centric frame of reference. Often called a relative, or viewer-centric frame of reference (see Levinson (2003) for details of the similarities and subtleties of these terms), this frame of reference defines spatial relations with respect to an individual's body. A demonstrative example of a person-

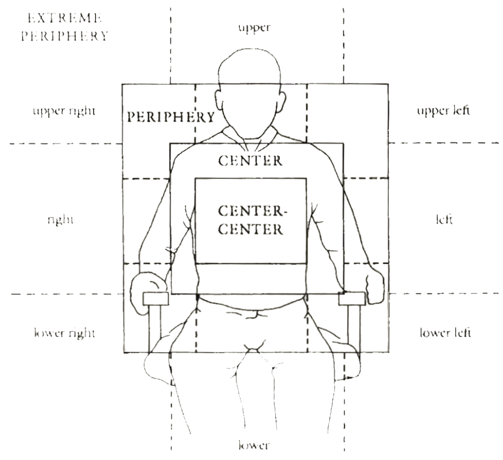


Figure 2.7 – An individual's gesture space, taken from McNeill (1992)

centric frame of reference is found when people give route descriptions as if the person were in the scene. For example, when giving instructions of a route through a building a participant may say 'As you come in, turn right. To your right will be the personal computers room. Continue until you're forced to make a left.' (Taylor and Tversky, 1996). Consider the work covered in Section 2.2.1.3 on abstract referents. As discussed, these make use of space and give it a salient structure with respect to the speaker only. There are other frames of reference that can be used involving other people in the conversation (see Chapter 3) or the world with its cardinal directions (Majid et al., 2004; Taylor and Tversky, 1996). These choices can be influenced by the encapsulating culture and the availability of spatial terms in its language (Levinson, 2003).

2.3.3 Space in sign language

Sign language is interesting to the current discussion because it is intrinsically spatial. The signing space is directly in front of the body at about chest height. The most common frame of reference used in sign language is a person-centric one (Emmorey, 2002). For example, if a speaker of sign language wanted to refer one of two packing boxes in front of themselves (without pointing at the actual box), they would directly map these to their signing space from their own perspective (that is, if the box was on the speaker's left, it would be on the left of the speaker's signing space). This means that anyone trying to interpret these signs must perform a mental rotation (which is done easily and regularly) of the speaker's space.

Structured space is defined as part of the language (at least in American Sign Language) using a person-centric frame of reference. Signers use the space around the body to represent time

(Bellugi and Kilma, 1982). Emmorey (2002) reports on three different timelines that a signer may use, indexed by their body (see Figure 2.8). These are:

- **Deictic** - runs back and forth in the direction that the signer is facing. The future is forwards such that now is the signer's body. Thus if the sign for Friday was placed with an extended arm it would be next Friday, whereas close to the body would be this Friday.
- **Anaphoric** - runs diagonally across the signing space, away from the signer's body. The timeline is anaphoric because the point of reference is not the current time, but a time determined from the content of the utterance. Ahead of this point is located ahead and across from the signer's body.
- **Sequence** - as with the anaphoric timeline the reference point is determined from the utterance, however here the timeline runs parallel to the signer's body (although the direction of time across this line may be determined culturally).

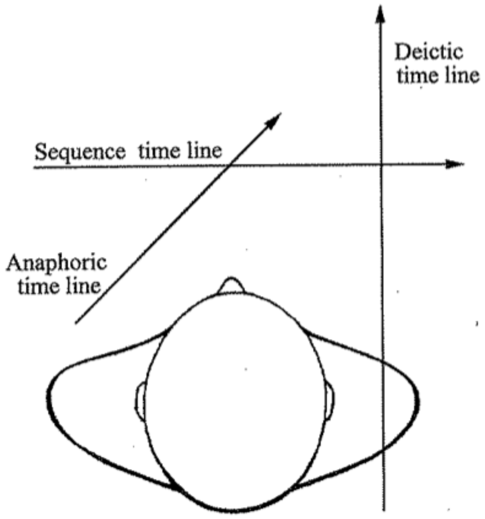


Figure 2.8 – Timelines used in American Sign Language, taken from Emmorey (2002)

Each of these timelines show a clearly structured use of space, which is defined by the language and indexed by the individual.

Instances of abstract spatial structures, which are not part of the language, have been documented also. Emmorey (2002) reports an example of a Chemistry teacher at a deaf school describing the stages of substances changing from solids to liquids to gasses. She signs at different heights to depict a vertical diagram of these stages, loading gas high up in the space and solid low down, moving up and down this diagram as she details the state changes.

2.4 Summary

This chapter has provided a review of the literature which approaches non-verbal behaviour using the individual as the unit of analysis. By following the focus of the literature and studying the denotation of body movements, the multimodal aspects of utterances have been described. In planning, constructing and deploying an utterance speech, hand gesture and head movement are integral constituent components. They are delivered as a composite signal, rather than the gesture and head movement being redundant to, or parasitic upon, speech. The modalities work together to build meaning and are deployed in tight temporal synchrony. One of the implications of this is that, when studying spoken dialogue (in contrast to say, text dialogue), it is essential that the non-verbal actions of the speaker are also considered. There is significant information bundled as part of the utterance that would be missed if only the speech were transcribed.

In the study of space it was shown that individuals exploit the space around their bodies. This may be through claims of access rights to the space, or by gesturing within the space and structuring it with abstract referents. As this exploitation takes place the space is transformed from a measurable space towards a structured qualitative space. This structure goes beyond the physical constraints of the body and is defined by normative characteristics; it is only when someone walks uninvited into the transactional segment that it is marked and the segment becomes manifest. The crucial point here is the use of the person-centric frame of reference in understanding space and the non-verbal behaviours within it. This frame of reference will have naturally emerged from a focus of study on the speech-movement link and whilst these findings give insight into how an individual behaves, by themselves they are inadequate for understanding dialogue. For this the analytical unit must be the interaction, rather than the individual. When this is the case, notions of sharing and collaboration come into play and give rise to frames of reference which encompass multiple individuals. This in turn gives rise to possible functions of non-verbal behaviour that do not just depict content, but manage the coordination of multiple people.

Consider the taxonomy of gesture that was covered in this chapter. If it is the case that interaction allows for gestures which do not just relate to the content of a discussion, then the taxonomies that consider only an individual gesturing are insufficiently comprehensive. This is also the case for functions of head movement. A person's gaze behaviour holds strong social implications; most people can relate to the feeling of being the recipient of an unwanted stare. If the study of non-verbal behaviour was to stop at exploring only isolated individuals, then the use of the body in a conversation of two people would simply be classed as two distinct sequences of activities. As the thesis moves through the next chapter, it will become clear that individuals within an interaction bring about activities which are more than just the sum of the number of individuals present.

Chapter 3

Coordinating a Face-to-Face Interaction

3.1 Introduction

Conversation is a collaborative activity. In contrast to a monologue it involves closely coupled negotiations and coordination between multiple parties. Together they must determine who is currently allowed to speak, who wants to speak next and who is actually part of the conversation rather than merely co-present. Interlocutors must understand what is being said, and demonstrate this to the others in the conversation. They must make clear their continued attention to the conversation, and monitor others for their continued attention (or lack thereof). In group conversations, who the speaker is addressing, and who has a responsibility to respond must be mutually understood and accepted. On top of all of this, interlocutors need to recover the conversation when things go wrong.

As complicated as these challenges sound, conversation is a routine and everyday activity. Without always realising it, people can design with behaviour in such a way that it coordinates not only the content of the discussion, but the processes by which the conversation is regulated. When these conversations occur in a co-located, face-to-face context the body plays a central role in the communication and works hand in hand with the interlocutors' speech.

The body is able to perform actions that would otherwise need to be communicated explicitly in speech. During conversation people produce hand gestures which relate to the other interlocutors (Bavelas et al., 1992) and they move their bodies in synchrony with each other (Condon and Ogston, 1966; Ashenfelter et al., 2009). They stand in certain ways, structuring space to identify them as part of the conversation (Kendon, 1973), and they can show their understanding and continued attention with movements of the head (Yngve, 1970).

This chapter will open by detailing the structured use of space which occurs when participating in interaction, and will introduce the shared interaction space. In contrast to the use of space seen in the previous chapter, this is rooted in an interactive context. This use of space allows for the delimitation of a unit of interaction and as such the discussion will involve details of the interactive structure of a dialogue. Some of the additional challenges involved with a multiparty dialogue will be highlighted. Following this, the chapter will move to explore non-verbal cues which function to coordinate the dialogue within the shared interaction space. This section will include details of the organisation of interaction by turn-taking mechanisms, the use of head and hand movements for regulation and feedback and the use of collaborative gesture to share content within a shared space. The chapter will conclude with an exploration of a conversational situation which does not support a shared interaction space. As with many socially implied norms and features of interaction, the shared space becomes most apparent when it is missing. The scenario that will be used here is that of conversations using video mediated communication (VMC). This medium supports simultaneous bi-directional visual and audio ‘channels’ of communication, but precludes access to the shared space. The discussion will introduce VMC and explore the literature which examines the effectiveness of non-verbal behaviours which attempt to coordinate the interaction within this medium. Some problems will be noted and some findings suggesting compensatory behaviour will be covered.

3.2 Spatial units of interaction

“Everyday terms refer to different aspects of encounters ‘cluster,’ ‘knot,’ ‘conversational circle’ – all highlight the physical aspects, namely a set of persons physically close together and facially oriented to one another, their backs toward those who are not participants.”(Goffman, 1966)

This quotation from Erving Goffman’s study of behaviour in public places shows an informal introduction to the spatial aspects of interaction. He noted that when in focused interaction, people orient themselves with respect to each other. Goffman defines focused interaction as being “concerned with clusters of individuals who extend one another a special communication license and sustain a special type of mutual activity that can exclude others who are present in the situation” (Goffman, 1966). Observations presented by Sommer (1965) suggest a potential difference in the spatial behaviour of people who interact and those who don't. By sketching the seating arrangements of people observed informally in semi-public places such as student canteens and libraries, he saw that interacting pairs tend to sit closer, potentially opposite each other or sharing the corner of a table, whereas non-interacting pairs sit diagonally away from each other, maximising distance. The work that follows in this section considers a frame-of-reference that is rooted in the interaction, rather than the individual. Following the theme of perspective taking in the previous chapter, Schober (1993) notes that when speaking in an interaction, the perspective taken has much higher variation than monologue and requires a negotiation between the interlocutors to come to an agreement which is usually maintained for the rest of the interaction.

3.2.1 Body positioning during interaction

As people interact, their choices of body positioning are not arbitrary. It has been noted by Scheflen (1964) that during group meetings, the members will position their bodies to define or delimit their immediate group from others present. Similar observations were made by Lyman and Scott (1967). They offer a discussion of territoriality in which they define an interactional territory. They suggest that surrounding any form of interaction is an invisible boundary or membrane. It is also noted that notions of maintenance and norms occur during the interaction:

> *"Every interaction territory implicitly makes a claim of boundary maintenance for the duration of the interaction. Thus access and egress are governed by rules understood, though not officially promulgated, by the members."*

These formations can be both stationary and mobile.

In his 1973 paper, Kendon stated that, whilst in focused interaction, the interlocutors:

> *"orient their bodies in relation to one another in such a way that for each one, the angle through which the head would have to rotate from its orientation in the sagittal plane of the body to an orientation in which it would be directly facing another participant is less than ninety degrees"*(Kendon, 1973).

This shows an initial attempt to theorise about what is happening within these spatial formations and an introduction of the notion that interlocutors adopt spatial locations and orientations relative to those of their co-participants. Importantly, this suggests that *interaction* as a unit can be spatially defined.

Kendon further adapted this concept, and gave the more formal notion of an *f-formation* (originally termed ‘Face Formation’) (Kendon, 1990, 1992). When people interact, Kendon suggests that they stand in a formation[1] such that their individual transactional segments[2] overlap with that of each other person. This collaboratively created spatial unit is termed the f-formation.

3.2.2 The dynamic structure of dialogue

When people hold a face-to-face interaction, delimited by the f-formation, the dialogue which they engage in is structured such that the interlocutors adopt dynamic roles. The roles shift as the conversation unfolds. During dyadic dialogue the interlocutors fall into one of two roles. Excluding cases of overlap, if an interlocutor is speaking and holds the floor they are the *speaker*. The notion of a speaker can, in difference situations, have different meanings and can be decomposed further. They may or may not be the author of the words that they speak (for example reading out-loud from a book, or performing from a script), or they may not be speaking from their own position (as when speaking for another) (Goffman, 1981). For current purposes a speaker in a dialogue is assumed to be both speaking for themselves and be the author of their words. During this time in dyadic interaction, the other person is the *listener* (sometimes referred to as a *hearer*, or *auditor* (Duncan and Fiske, 1985)). They have the responsibility to keep track of what the speaker is saying, but also have the right to request clarification or try to take a turn of talk. Goffman notes that in a dyadic dialogue, the listener is also implicitly the addressee; there is nobody else that the speaker could be addressing with their speech. The coordination of turns, that is the exchange of current-speaking rights, is managed only across these two people.

In multiparty dialogue, the roles are more complex than dyads; the basic speaker-listener model is not sufficient as it does not account for the multiple non-speaking interlocutors, who can adopt different roles with distinct patterns of behaviour. In Goffman’s discussions he demonstrates that conversations of more than two people have multiple listeners, not all of which are be-

[1] Kendon defines the difference between a *formation* and an *f-formation* as being dependant upon the presence of an interactional element. For example, a simple formation may occur where two or more persons sustain some spatial orientation, however if there is no interactional element then it is not possible that this could be an f-formation

[2] see Section 2.3 for a discussion of the transactional segment

ing addressed by the speaker. He defines a participant's *participation status* relative to a speaker's utterance to be 'the relation of any one such member to this utterance'. There also exists a distinction between people who are part of the conversation, and people who are not part of the conversation but are able to hear the speaker's words. Goffman states that together, these people's statuses form a *participant framework*. Clark and Carlson (1982) provide relatively clear definitions of the roles that people adopt during multiparty conversations. These definitions can be applied to the example conversation given in the introductory chapter featuring Ann, Bob and Claire. As Bob tells Claire about the bus incident the roles are:

- **Speaker** This is Bob. This is the person who has performed the utterance.
- **Addressees** These are the people that the utterance is directed at. This is Claire in the above example (she may often be called the *primary addressee*). Whilst Ann is involved, she is not a direct addressee of the utterance. She is commonly referred to as the *side participant* in a 3-person dialogue.
- **Overhearers** These are the other people in the corridor who can hear Bob's utterance. Whilst they can listen to what is being said, the speaker does not intend them to take part in the conversation.
- **Participants** These are the people who the speaker identifies to take part in the conversation and include the speaker and addressees, but not the overhearers. In this example Ann, Bob and Claire are included, but the overhearers in the corridor should not be.

Goffman notes that there would be a difference between those in the hallway as eavesdroppers and those in the hallway as overhearers; an eavesdropper attempts to conceal their ability to hear the words, whereas an overhearer, or bystander, may be able to hear by chance. Clark and Shaefer (1992) use the term *ratification* to distinguish between participants and overhearers; participants are ratified members of the conversation, whilst overhearers are not[3]. Clark and Carlson (1982) have more precise definitions and the distinction between ratified and non-ratified participants is based upon the intentions of the speaker. During a face-to-face dialogue, the f-formation itself serves as a ratification device, distinguishing between participants and overhearers (see

[3] This discussion of ratification brings about the need to define what exactly a conversation, or interaction, is and how participants of a dialogue are considered ratified. Goffman defines a focused interaction (the counterpart being unfocused interaction) as being 'concerned with clusters of individuals who extend one another a special communication license and sustain a special type of mutual activity that can exclude others who are present in the situation'(Goffman, 1966).

Kendon (2010) in which he also discusses how an f-formation discriminates important from unimportant things). Those in the formation are ratified participants, those outside are not and may be considered overhearers. This ratification has been seen in an ethnographic study of conversations in a tourist information center (Marshall et al., 2011). They saw that, when a person was not included in the f-formation, they don't participate directly in the interaction. It is clear that, however participants are ratified, they are different to overhearers. Schober and Clark (1989) demonstrated this using a tangram task in which addressees were part of the interaction (and fulfilled the role as defined), however the experimental setup was manipulated to allow for an overhearer to be present. This overhearer, by definition, was not able to request any clarification or show any feedback to the speaker. The overhearers consistently performed significantly worse at the task than the addressees. This demonstrates that even though an overhearer of a dialogue may be able to hear fully what the speaker is saying, they are different from addressees because they lack the ability to *interact* with the speaker. Schober & Clark's results demonstrate that this difference is meaningful.

This thesis will be concerned only with ratified participants of a conversation: the speaker and the addressees, specifically the primary addressee and side participant. The important point to note from these definitions is that non-speakers of multiparty dialogue are not simply able to adopt the role of listener, but must dynamically switch between primary addressee and side-participant. This then raises a higher coordination problem than that found in dyads; the three participants must coordinate with each other to effectively manage these roles and the switches between them. The concerns of this thesis include how non-verbal behaviour is exploited to manage this coordination.

3.2.3 The shared interaction space

A feature of f-formations, and a point which is central to this thesis, is that when the transactional segments of interlocutors overlap, the physical space between the interlocutors is transformed into a shared *interaction* space. The transactional segment is where an individual carries out their activities; the shared space is where interactional activities, organised as a dialogue, are carried out. This space is called the o-space by Scheflen (1964) and Kendon (1973), the positioning of which is shown in Figure 3.1. The participants whose transactional segments constitute the o-space now have joint and exclusive control of, and access to, the space. It is important to note that the exclusivity here is socially implied. It would be entirely possible for an outsider who is not part of the interaction to use the o-space but this would result in appearing odd or even rude. The maintenance of the shared space is dynamic and mutual meaning that people tend to perform a continual 'dance', collaborating with each other as they move to ensure the shared space is sustained (Kendon, 1990).

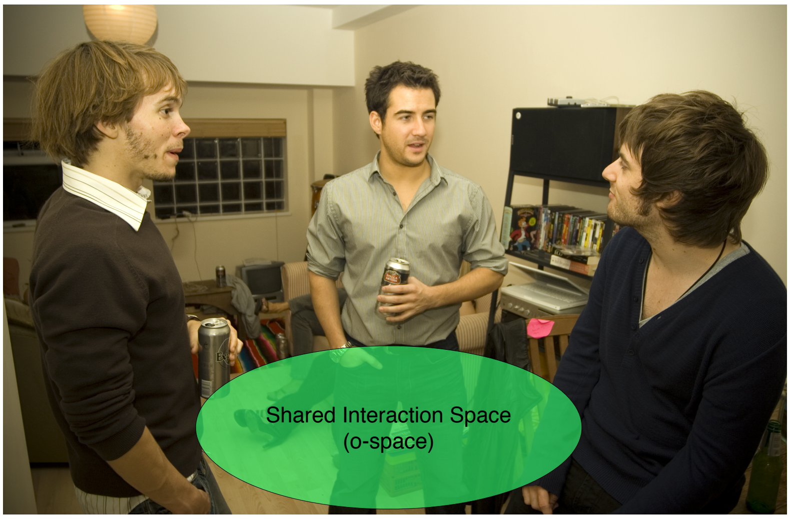


Figure 3.1 – The positioning of an interactional o-space

Gill et al. (2000) have a similar notion to the o-space called the 'space of engagement'. This is said to be an aggregate of individuals' body fields. It allows interlocutors to use their bodies as cues to their willingness for co-operation, and undertake parallel and coordinated communicative moves. As with Kendon's o-space the space of engagement fluctuates over time, adjusting with the interlocutors' level of comfort with each other and requiring reconfigurations for its maintenance.

Kendon notes that it is possible to see people's sensitivity to others' shared spaces occurring regularly; think of when two people are talking in the corridor, if a third person is forced to walk between them due to some spatial restriction and enter the shared space they typically dip down, ducking their head and sometimes even verbally apologise. This demonstrates their recognition of the socially implied exclusivity and that they are attempting to minimise their impact on the space.

3.2.3.1 Instantiating a shared space

When moving from mere co-located people, to those who jointly manage a shared space and are constituent members of an interaction, people coordinate their bodies. Mondada (2009) examined opening sequences that lead up to a social interaction. Her study made use of data in which passers by were approached and asked for directions in a public place. Her findings demonstrate that before speaking, people achieve mutual orientation of their bodies and their gaze. She shows that having decided upon who to approach, the direction seekers would gaze at the passer by and begin to approach them, progressively matching the trajectory of their walking. Commonly a number of gazes are exchanged between the approaching person and the passer by before they begin the process of moving from two independent units to one joint unit of interaction. As the passer by slows down, they deploy mutual gaze with their approacher, slowly turning their upper body such that it is twisted from their lower body. They reach a stable state of interaction with a shared space once they have turned to align their lower bodies with each other. Once this has happened it is common for the space to be reconfigured for the task at hand (in this situation this is the giving of directions). An example from Mondada (2009) demonstrates some of these features:

1. Approacher: excuse me, madam?

   Approacher movements: walks towards passer-by

   Passer-by movements: looks forwards, looks below, looks at approacher, continues walking

2. Approacher: I am looking for the church of saint Roch

   Approacher movements: stops, initiates body torque, changes feet

   Passer-by movements: stops and turns to approacher

3. *Approacher & Passer-by movements: turn face-to-face and exchange mutual gaze*

4. Passer-by: ehm, it's there

   Passer-by movements: turns head and points

Once they have reached 3) both people have now coordinated with each other to create an f-formation and form a shared space. As they move to step four it is reconfigured for the task at hand such that both people are now oriented around the space facing towards the place of interest.

This excerpt shows the initiation, and growing levels, of coordinated movement that leads to the initiation of the shared space.

Once formed, the f-formation and the encapsulated shared space are always under the control of the current interlocutors, however they are not limited to these people; as they are features of interaction rather than of the interlocutors, the constituent people may change (perhaps a complete cycle of people) and the formation and shared space will remain, albeit under new control.

When this process of admitting a new participant to, or a participant wishes to leave from, an already instantiated interaction a set of socially implied norms create a process to follow. Goffman suggests that 'minor ceremonies' are performed to mark the entrance and departure of participants from a formation (Goffman, 1966). In order to analyse these, it is necessary to examine Kendon (1973)'s definitions in finer granularity. Surrounding an o-space is the p-space which is the space in which the participants' bodies are located. Extending further out from this and the interaction is a somewhat vague and undefined r-space.

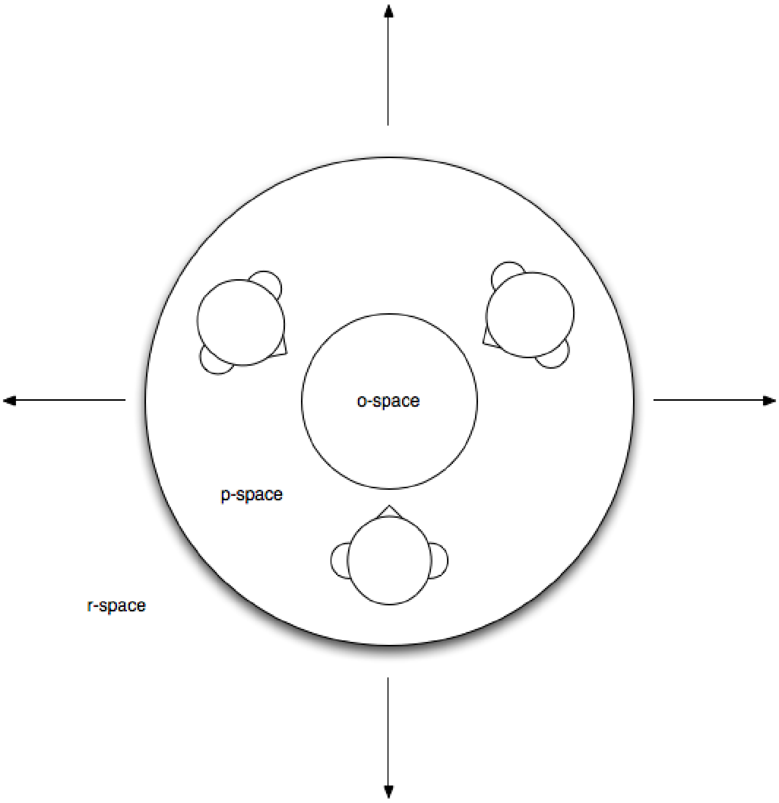


Figure 3.2 – A diagrammatic representation of Kendon's definitions of o, p and r spaces showing a bird's eye view of three people in interaction.

The spaces are shown in Figure 3.2. An approaching person with a desire to join the interaction will initially enter the r-space. As they signal their desire to join the interaction to its current interlocutors, they must wait on the periphery of the p-space until the interlocutors accept them. At this point the existing interlocutors' positions and orientations will adjust to accommodate the new interlocutor and give them equal access to the shared space. It is worth noting that this process of entry only occurs when an external person wishes to join of their own accord; had they been invited to join, they would not need to wait for acceptance by the interlocutors of the interaction, this acceptance was implied by their invitation.

Upon deciding to leave, an interlocutor is likely to step back from the space, but then return in (potentially closer than before) for salutations before walking out, away from the o-space. This may be in the opposite direction to where they are actually going, but the first movement is away from the o-space (without crossing it). Once the formation has been cleared, they can exit in the desired direction (Kendon, 1990).

Whilst the shared space requires continual maintenance, it is tolerant to minor deviations from the interaction at hand. Interlocutors are able to adopt different orientations and postures of the body during conversation. Indeed, it has been shown in a variety of studies that people are able to orient their bodies towards different activities (Scheflen, 1976; Schegloff, 1998; Kendon, 1990). Scheflen discusses how an interlocutor may orient their head to other involvements than their primary one, which is indicated by the lower body. Most of the examples that he uses involve some interaction as one activity, with a non-interactive activity as the second. For example, doing the washing up whilst also talking to people at a table (see Fig 3.3). Similar notions have been discussed by Schegloff (1998) under the name of 'body torque'. He suggests that differing orientations of body regions can show differing levels of priority to differing activities. When the body adopts this kind of posture it is said to be in torque. The primary activity is identified by the orientation of the lower parts of the body, with any secondary activity being identified by the upper body parts' orientation.

Kendon makes note of this in the context of a shared space (Kendon, 1990). He observes that, whilst a twisted posture usually holds both orientations within the bounds of the shared space, it is possible for a posture to be held temporarily such that it orients an upper body part outside of the shared space. However he notes that if this orientation is held for a lengthy period of time, it is likely the the lower body will turn to fall in line with the upper body orientation. This will

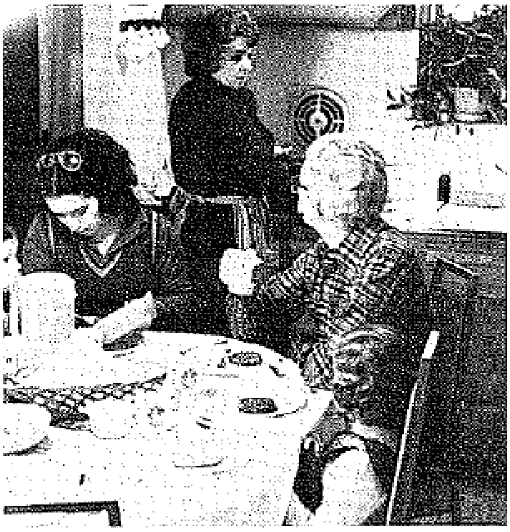


Figure 3.3 – An example of a person attending to two activities, one of which is the people at the table (interactive) the other is the washing up (non-interactive). From Scheflen (1976)

cease involvement with the prior shared space, and will create a new one at the new orientation.

3.2.3.2 The shared space in action

Özyürek (2000, 2002) demonstrated interlocutors' sensitivity to the shared space by examining the effect of specific configurations of the shared space upon gesture formation. She was concerned with how addressee location impacted upon the formation of gestures. Her experiments required a participant to recall a scene from a cartoon in which a cat was thrown out of a window to either one or two addressees. In the one addressee condition, the addressee was to the side of the speaker whilst in the latter they were placed either side of the speaker (see Fig. 3.4), resulting in shared spaces with different configurations. It was found that instead of the gesture which accompanied the motion event 'out' being the same for each condition, it was adjusted for each such that the 'outwards' action was out of the shared interactional space. In the dyadic condition the speakers produced lateral (left/right) gestures which went across and out, whereas in the triadic condition they produced frontal (forward/backward) gestures which went back and out of the space.

Emmorey (2002) notes that phenomena such as these do not occur when conversing with sign language. The gestures seen in Özyürek's study make use of a collaborative reference frame. In sign language, space is always referenced by the individual's body and as such gestures such as the out gestures above would be from the individual's space. Moreover there is variation between cultures. Haviland (1993) reports on speakers of Guugu Yimithirr, an aboriginal language in

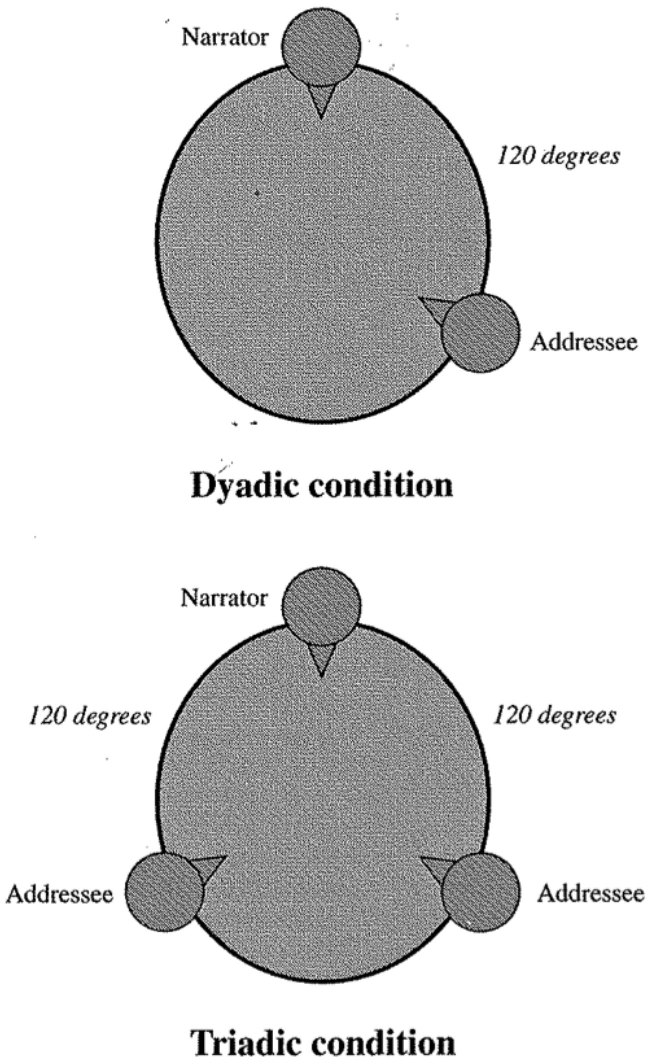


Figure 3.4 – The positioning of addressees in each condition in Özyürek (2000)

Australia, who typically use cardinal directions in their speech and gestures. That is, if the movement were out in a southerly direction, that is the direction they would gesture in. However, the study notes that at certain points speakers will make use of the shared interaction space between the interlocutors, such as when recreating a gesture which has no intrinsic orientation.

3.3 Non-verbal cues for managing an interaction

The chapter so far has shown that an interaction can be delimited spatially, and that the dialogue which takes place within this spatial formation has interactional structures which are more complex in a multiparty dialogue than a dyadic one. This discussion will now move to examine the non-verbal cues used to support the dialogue. During an interaction, the behaviour of an interlocutor is influenced by the presence and actions of others. Marchand (2010) documents a situation where a pair who collaborate on a woodwork construction task interject non-verbally, completing each other's actions. For example, as one member of the pair makes use of a particular holding technique, the other steps in and modifies the method; an action which is accepted by both and thus used further. The modification and acceptance is completed without speech.

At a basic conversational level, Tabensky (2001) offers evidence that the information content of an interlocutor's speech is interpreted and reproduced by the listener. In a discussion of verbal rephrasing she notes that part of an utterance delivered by one interlocutor can be interpreted and reused in another form by the other interlocutor (a common form of this is rephrasing "you haven't" to "I haven't"). Tabensky suggests that this form of rephrasing also happens for gestures. She demonstrates this with an example involving an interlocutor named Daniel. In a description of an apartment he states "a hundred and eighty square meters", meaning a large apartment. His gesture involves both hands moving outwards to show the width and his head pointing upwards to show the height, even though the height aspect was not encoded verbally. Nicola, the second interlocutor replies with "a duplex", with her gesture involving one hand with palm forwards, index finger pointing up and the other on her lap. She has interpreted and rephrased the content of Daniel's head gesture into both her speech and hand gestures. Similar findings are noted by Hayashi (2005) who, in a study involving Japanese speakers, saw that the content offered gesturally but not verbally by one interlocutor was taken up by the other and reproduced verbally. The findings of these studies demonstrate both the communicative nature of the interlocutors' behaviour, and underline the multi-modality of interaction.

The shared space can also be used as a resource that influences the trajectory of the interaction. A good example of a conversation with multiple interlocutors within a shared space is offered in Lerner (1992). Here, four people sat around a table are involved in a story telling situation (see Fig. 3.5 for their positions). In this example, Phyl is explaining a story to Curt, which requires some input from Mike. When she begins, her lower body is oriented towards Curt within the shared space, and her gaze is just off to the side of his head. As she approaches the section of speech which requires input from Mike she turns her head round towards him. However, here she leaves all parts of her body lower than the head oriented towards Curt. Lerner suggests that this is because the requirements from Mike are only side to the main involvement: Curt. Her body position signifies that telling the story to Curt is still the main activity and that it will be returned to after the side activity with Mike has finished. Lerner does not offer a description of the movements of Curt or Mike during this excerpt, or any more specific details of Phyl's body movements. This example demonstrates the importance of the spatial positions of the interlocutors around the shared space as a method of indexing them in the dialogue. This claim is supported by Bekker et al. (1995), who provide further evidence from studies of group design meetings that gestures occur in relation to the spatial arrangements of interlocutors, and Gill et al. (2000) who note that the orientation of interlocutors' bodies can identify their joint focus.

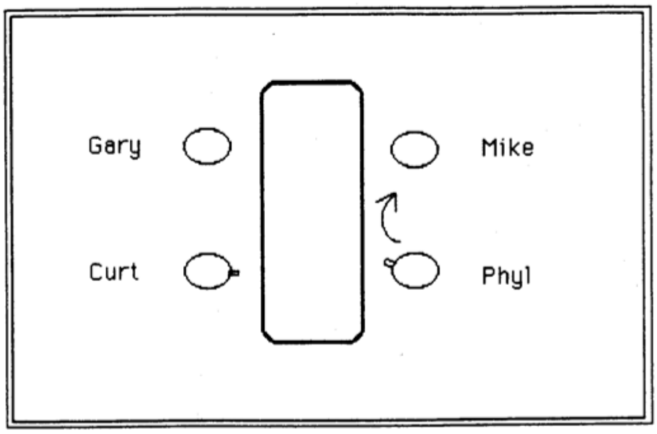


Figure 3.5 – Four people involved in story telling. From Lerner (1992)

Moreover, people can manipulate their positions around the shared space as part of their conversation, potentially based upon rank or authority. For example, LeBaron and Streeck (1997) show how police interrogators can manipulate their position with respect to the suspect during a murder investigation. With two officials and the suspect sat around a table, initially the formation is equally spaced. This is said to attempt to elicit information from the suspect and keep him relaxed during this time (the type of speech used also implies this). As the interrogators move to apply more pressure and authority upon the suspect they shift their positions in the formation such that there is an unequal balance with the focus of attention directly on the suspect.

Kendon (1990) reports on some specific configurations of interlocutors around the shared space. Dyadic interaction often elicits three different orientations: firstly a *vis-a-vis* arrangement which sees the interlocutors facing each other (often associated with intimate or confrontational interaction), secondly an *L-arrangement* and finally a *side-by-side* arrangement which is often used by participants of the opposite sex. In multiparty situations different shapes may form assuming that the transactional segments of the participants overlap. Kendon's views are in line with those of Le Baron & Streeck, noting that it is possible that rank or authority of members may influence their position in the arrangement.

Notions of regulation which require a shared space have been demonstrated in situations which are not exclusively conversational. For example, Gill and Borchers (2003) note that when collaboratively drawing on a whiteboard, people will step back when someone else is drawing and step in to try to take the turn at drawing, and Healey et al. (2005) demonstrate that the shared space is used in collaborative music improvisation in order to regulate who can play when.

These examples demonstrate that the configuration of the shared space, in terms of the positions of the interlocutors (which is directly influenced by the number of interlocutors), is tightly coupled with the non-verbal behaviour of the interlocutors and can influence the trajectory of the conversation.

3.3.1 Who goes when?

In order to understand how non-verbal behaviour assists in the regulation of an interaction, literature which covers a core feature of interaction, the turn-taking system, must be addressed. Conversations are structured by who is allowed to talk, and norms exist which determine who is going to talk next. This is known as the turn-taking system (Sacks et al., 1974; Duncan and Fiske, 1985). The turn taking system leads to the collaborative creation of 'units of interaction' which are said to be the building blocks of conversation (Duncan and Fiske, 1985). The norms are socially implied, but can be recognised when a person breaks the norm and interrupts a speaker mid-turn; it appears rude. Conversely, if nobody takes the turn and speaks then the situation becomes embarrassing. There is, however, no formal rule system which states how interlocutors are to negotiate the turn exchanges and so some form of coordination must exist within the dialogue to facilitate it (Wiemann and Knapp, 1975). Various discussions in the literature exist on how the exchange of turns is handled (Sacks et al., 1974; Wiemann and Knapp, 1975; Duncan and Fiske, 1985), but there is general consensus that:

- Turns can be yielded by the speaker at appropriate points

- Turn can be requested at appropriate points

- Continued attention to the current turn must be demonstrated

The management of these activities can be, but is rarely, done explicitly (at least in natural conversation). Cues to the upcoming end of a turn can be found in the spoken utterance, such as its lexical and syntactic structure (de Ruiter et al., 2006) including 'uhs' and 'umms' (Clark and Fox Tree, 2002), and movements of the body are regularly used as coordinating signals (Koutsombogera et al., 2011). When considering turn yielding, the speaker who holds the turn must coordinate with their listener to signal that they have the opportunity to take over. This, however, is only an opportunity and is not mandatory; the listener does not have to take the turn if they do not want it. Conversely, when requesting a turn, the listener must coordinate with the speaker to demonstrate that they would like to take the next available turn. These activities should occur at a point in the dialogue where an exchange would be appropriate.

Whilst a speaker holds the turn, the listeners are not silent receivers of information, but must provide feedback in order to fulfil the requirements of showing continued attention and understanding (Yngve, 1970; Kendon, 1977; Enfield, 2005). Indeed, Yngve (1970) states that in conversation, 'messages flow simultaneously in both directions'. This enables both the speakers and the listeners to acknowledge that what has been said has been contributed to their shared common ground (Clark and Schaefer, 1989). Common ground consists of any prior common knowledge and beliefs that an interaction starts with. As conversations progress, the interlocutors contribute to and build upon their prior common ground with their dialogue (Eshghi, 2009). Listeners must provide feedback in order to show understanding and hence acknowledge the contribution to common ground.

When listeners provide feedback it is commonly verbal in the form of short phrases such as "uh huh" or "O.K.", or through head and hand gestures (Yngve, 1970; Allwood and Cerrato, 2003). Bavelas et al. (2000) (and reported further in Bavelas and Gerwing (2011)) provide an additional level of resolution in the types of feedback listeners can give. They state that it can be generic, which is not tied to the meaning of the speaker's speech ("uh huh" and "O.K." would fit generic feedback), or that it can be specific to particular points in the speaker's speech (such as wincing at their pain). Feedback means that holding the turn or not, is not the same as being a speaker or a listener. There are times when an interlocutor is able to briefly be a speaker without having claimed the turn (Yngve, 1970). The need for feedback becomes particularly apparent when it is absent. In these situations a listener may often be told to 'pay attention'.

Within a multiparty dialogue, it is reported that there are different responsibilities between the primary addressee and the side participant; the primary addressee must audit the speaker's speech and give evidence to the speaker of this as in a dyadic dialogue. However the side participant, whilst part of the conversation, has a lower responsibility to demonstrate understanding and as such should collaborate with the speaker less than the primary addressee does. This is as stated by Clark and Shaefer (1992)'s principle of collaboration: 'Speakers collaborate directly with addressees and only indirectly with side participants'.

3.3.2 Regulation with head movement

Movements of the head communicate a lot during interaction. They overtly display focus of attention and are able to identify intended addressees. A speaker's gaze can work together with their gesture to guide the addressee to salient information. Very often an addressee will focus their attention upon a speaker's gesture only after the speaker has looked at their gesture themselves (Streeck, 1993; Enfield, 2005). Gullberg (2003) confirmed these findings in a quantitative study. She provides a quantitative estimate of the relative importance of a speaker's face and hands by measuring the eye-gaze patterns of addressees. Her face to face ('live') condition consisted of two people, one of which had watched a cartoon. This person then retold the cartoon in narrative form to an addressee who had been configured with eye tracking equipment. The gaze patterns of this addressee were recorded. Only 7% of the speaker's gestures were fixated by the addressee. 96% of the time the addressee looked at the speaker's face; only 0.5% of the time was spent on their gestures with the remaining time spent looking at other objects in the room. This data demonstrates the relative importance of the head, when compared to the hands, for addressees of dyadic dialogue.

The specific regulatory functions of the head are particularly manifest around turn exchanges (Kendon, 1977). When approaching the end of a turn, Kendon states that a speaker is able to signal to their addressee that they are about to give up the floor by looking up and at the addressee. Wiemann and Knapp (1975) also suggest that gaze behaviour corresponds with turn-yielding and turn-requesting activities. They show, in line with Kendon, that when yielding a turn the speaker will look up at the listener to form mutual gaze (although see de Ruiter (2007) for a discussion of the reducing effect of a visual artefact on person oriented gaze). Knapp (1978) defines mutual gaze to occur in a "situation where the two interactants are looking at each other - usually in the region of the face" and state that the amount of time spent in mutual gaze can vary with the intimacy of the pair or their relative statuses. Wiemann and Knapp (1975) also show that when a listener attempts to request a turn, they shall look at the speaker (attempting to form mutual gaze) and dramatically increase the amount of head nods that they produce (Kendon (1973) likens this to a musician beating time just before he starts to play his instrument) . It is possible that these behaviours are detectable by an observer to an interaction. Foulsham et al. (2010) show that observers of a recorded interaction whose gaze targets are measured, gaze at the speakers of the dialogue. When measuring the timing of the observer's gaze towards a person, it lands on the target 150ms prior to their turn starting. Whilst this must be interpreted with caution because the data is from an observer rather than a participant of the dialogue, it suggests that something is happening to the behaviour of the interlocutor prior to gaining the floor which is different from their normal behaviour.

When considering dyadic dialogue there is an imbalance between the speaker's and the listener's gaze behaviour (Kendon, 1977; Bavelas et al., 2002b). In these dialogues it has been shown that the listener is much more likely to be looking at the speaker than the speaker is to be looking at the listener. The gaze of the speaker becomes more important when examining multiparty dialogues. Interlocutors are able to advertise to whom their utterances are addressed by orienting their heads (and thus faces) to their intended recipient. Kendon (1990) names this a *face address system*. This has been used computationally by Frampton et al. (2009). This study used computational models which attempt to predict who the intended addressee of the spoken referring expression 'you' was. By varying the inputs to the model, using both features of the spoken utterance and gaze features, the accuracy of the prediction with the various features can be tested. The accuracy was improved by using the gaze of the speaker. The study noted that whoever the speaker was in mutual gaze with was most likely to be the addressee.

Similar notions are seen in Goodwin (1979), who demonstrates that the collaborative gaze behaviour of the interlocutors *within* a turn actually influences its construction. From the speaker's perspective Goodwin states that, whilst speaking in a group, the speaker's gaze is able to identify the intended recipient of the utterance. This is in line with Kendon's findings. Goodwin also demonstrates that the recipients must return gaze when a speaker gazes at them (and, although Goodwin doesn't state this explicitly, show continued attention to the speaker). This, of course, does not always happen and may result in the speaker reformulating their utterance mid-turn to either make it more appropriate for the current addressee, or to locate an addressee that it is appropriate for and is willing to return gaze. Bavelas et al. (2002b) see similar activity within dyads. They propose that a speaker creates a *gaze window* starting when they gaze at the listener. This should elicit feedback from the listener, either verbally or non-verbally. They provide evidence to show that the window is terminated once the listener has provided feedback.

Interlocutors are able to use each other's gaze direction as cues to their focus of attention, leading the interactants to share a joint focus of attention. Hanna and Brennan (2007) performed a director/matcher task in which they show that matchers make use of the director's gaze direction to identify a target object before it is specified in their speech. Moreover, by manipulating the alignment of the director and matcher's objects, they demonstrated that matchers are able to understand and remap the director's gaze depending on the scenario.

When considering the literature concerning gaze reviewed in the previous chapter, it should now be clear that an interlocutor's gaze behaviour is *not* automatic and stimulus driven but is instead part of a complex system of collaborative and communicative behaviour (a point which made clear by Hanna and Brennan (2007)). This is a point picked up on specifically by Kingstone (2009) who makes the claim that the method of presenting pictorial based stimuli to participants does not account for the *social* factors and suggests that the findings may be the same if the participants were presented with images of the feet, or pointing arrows. Kingstone's point is that if one excludes real people and a social situation from the study, no claims can be made about the *social* nature of eye gaze.

3.3.2.1 The effectiveness of gaze

When considering multiparty dialogue a problem arises with treating eye-gaze as an interactional cue. Interlocutors are only able to monitor the eye-gaze of one person at a time. Loomis et al. (2008) demonstrate that in situations such as small group conversations, people are only able to reliably judge another's eye gaze within a 4° rotation of their own head or eyes. Head orientation can be effectively judged up to rotations of 90°. Thus, given the reduction in effectiveness of eye-gaze in group conversations, it might be expected that head orientation should become a more suitable cue than eye-gaze. Indeed, Rienks et al. (2010) measure patterns of exclusively head orientation in multiparty dialogue and show that speakers tend to orient to listeners more than listeners orient to speakers. They do not offer a break down by listener role however.

This argument is in line with findings by Jokinen et al. (2010) in which a small corpus of multiparty dialogues are used. Head orientation and eye-gaze data are collected and used as features in a support vector machine which predicts turn exchange events. The results, which must be considered tentative due to the small sample, demonstrate that the relationship between turn taking events within the dialogue and eye-gaze is ranked lower than that with head orientation. This is calculated by comparing the accuracy of the prediction when the eye-gaze and head orientation features are systematically excluded.

3.3.3 Posture and synchrony

An apparent feature of interaction is that interlocutors coordinate their overall posture and fall into synchrony of movement with each other. Kendon (1970) demonstrated that in informal discussions it is possible for interlocutors to match their postures (Scheflen (1976) documents similar examples). Changes in posture are likely to happen around changes in communicative phenomena. For example, whilst shifts in interlocutor postures correlate with topic shifts, they are much more likely to occur around a turn exchange than within a turn. Cassell et al. (2001) demonstrate this by comparing the frequency of interlocutor posture shifts within a dialogue that occur at a turn exchange or within a turn. Condon and Ogston (1966) provide early quantitative findings that speakers and addressees fall into temporal synchrony. By manually annotating the micro-movements of interlocutors' bodies, they show that an addressee's body movements are in prosodic synchrony with the speaker's speech.

Recent work in Ashenfelter et al. (2009) has attempted to use more complex mathematical techniques, such as cross-correlation analysis, to determine if there is any motion-level synchrony between interlocutors of dyadic dialogue. The study looked at the how correlated interlocutors' head movements were and if these correlations occurred with any delay. Results showed levels of coordination around offsets of two seconds that were not present at greater offsets. In similar quantitative work, Shockley et al. (2003) provide data based upon interlocutors' hip positions, suggesting that pairs who interact with each other adopt more similar postures than those who interact with an experimental confederate. However, it is unclear whether the findings show coordination of body movement, or coordination by proxy of speech. This is because the coordination existed during a condition when the interlocutors could not see each other. These two studies use novel methods, and provide evidence that levels of coordination of body movements exist. The specifics of how long these persist for and what is actually being coordinated are not fully understood.

When considering the relationship between interlocutors of multiparty dialogues, Kendon (1970) notes that the speaker and 'direct-addressee' fall into synchrony with their bodies and that this relationship is distinct from the other interlocutors. The data for the study are annotations of interlocutors' movements, who are in group conversations in a bar. The conclusions drawn from this study allow for the introduction to the notion that there are varying levels of synchrony within a group interaction, however the methodology does not allow for finer grained conclusions to be drawn. Firstly, the level to which these other interlocutors were engaged in the conversation is not clear. Therefore whilst the behaviour of the direct addressee was distinct from others present, it is not clear if it was distinct from other participants or overhearers. The latter case would be demonstrating similar findings to those mentioned previously by Schober and Clark (1989). A more constrained conversational environment may make this clearer. Secondly it is not clear what the relative contributions of interlocutors' non-verbal cues are. It is clear from Kendon's annotations that the whole body has the potential to contribute, but the specific levels of coordination are not known. Thirdly, the amount of data used needs to be greater to validate the findings. The hand annotation of video data employed at the time of the study gave great detail, but is extremely time consuming. Current technologies are able to reduce the amount of time required for indexing interlocutor movements.

3.3.4 Using gesture to manage the interaction

The literature has documented numerous examples of gestural activities that are unique to dialogue. For example, recent observations presented in Jokinen (2009) suggest that interlocutors attempt to avoid gesturing at the same time as each other. In an early study of non-verbal behaviour Ekman and Friesen (1969) define 'interactive' behaviours to be acts by one person in an interaction which clearly modify or influence the interactive behaviour of the other person(s). Whilst this definition is somewhat cyclic, the view of gesture that is adopted is distinct from those seen in the previous chapter as it is clearly defined by the act's effect on *another* interlocutor. Ekman & Friesen continue to define a 'regulator' as a category of non-verbal behaviour which 'regulate the back-and-forth nature of speaking and listening between two or more interactants'.

Bavelas et al. (2011) take this notion further and provide evidence that the content depicted in a speaker's gesture contributes to the growing common ground between the speaker and listener. This common ground requires that understanding is evidenced, either explicitly or implicitly. The follow is an example of content depicted in the gesture contributing to the interacting pair's common ground:

Speaker: So we could have, like, you come in.

Gesture: places two index fingers together on the table

Addressee: Yeah.

Speaker: There's a kitchen.

Gesture: places left hand slightly to the left of prior gesture location

Here the speaker places their hands on the table as they say 'you come in' to mark the location of the entrance. This is explicitly acknowledged as understood by the listener, at which point the speaker continues to discuss the location of the kitchen with respect to the entrance. Throughout the study, 95.5% of transcribed addressee responses showed that they had understood the information that the speaker presented (Bavelas et al., 2011).

The influence of another's actions can be seen in another conversational situation as demonstrated in work by Jurgen Streeck (Streeck, 1993, 1994). He demonstrates this with a word searching example. When a speaker has forgotten a word for something, they will often produce a gesture instead. This gesture is then continually revised depending upon the actions of the addressee. In the example, found in Streeck (1994), the speaker has forgotten the name of a herb (speaking in the Filipino language, Ilokano). She uses pointing gestures to attempt to provoke the listeners into aiding her. As she is offered the incorrect answer, she continually modifies her gestures until she reaches the correct herb.

Bavelas et al. (1992) examine gestures which relate directly to the interaction. They document a class of gestures called *interactive* gestures which they say do not directly refer to the topic of conversation, but are instead involved with the process of conversing with another person. They state that the function of these gestures is to aid the maintenance of conversation as a social system and involve the orientation of a hand to another participant, perhaps using a pointing gesture or an open palmed offering gesture. They offer a demonstrative example of a discussion concerning a library:

Participant 1: then you look up the author or the title
Gesture: mimics skimming through cards

As this interlocutor says ‘look up’ they produce a gesture which mimics skimming through cards. This is a topic gesture. The other person in the dialogue later summarises this by saying:

Participant 2: “then look up under the appropriate thing”
Gesture: flicks hand to participant 1

As they say ‘appropriate’ they produce a quick movement towards the other interlocutor. This would be classed as an interactive gesture because it is in reference to something previously mentioned by the interlocutor who was oriented to with the gesture. Interactive gestures have been reliably found in an independent corpus presented in Holler (2010). In her work, Holler demonstrates that a subset of interactive gestures known as ‘shared information gestures’ directly relate to the shared common ground between two interlocutors. By manipulating common ground and measuring occurrences of these gestures there was a significantly higher number of them when common ground existed compared to when it didn’t (independent of gesture and speech rate). A similar class of gestures to interactive gestures has been noted by Jokinen and Vanhasalo (2009) called ‘Stand Up Gestures’. These are said to coordinate the conversation in order to highlight the important part of an utterance. Both interactive gestures and stand up gestures bear similarities to the more general notion of ‘body moves’ introduced by Gill et al. (2000). These are meta-communicative acts, embodied by body movements, which share information about the communication structure rather than the content (for example, using gaze as a checking mechanism). In examining a corpus of multiparty conversations, Jokinen (2010) provides evidence that pointing gestures are not simply used in order to identify or even create referents, but instead can aid in the management of the dialogue. The examples provided show speakers making use of points to manage the dialogue and coordinate information flow (such as identifying common ground). Cassell et al. (1994) single out beats, as defined by McNeill, as gestures which aid activities such as yielding a turn of talk.

In a discussion of gesture placement during interaction, Sweetser and Sizemore (2008) give examples of dyads deploying gestures in different spatial locations during their conversation. They highlight ‘regulatory’ gestures as unique in being deployed in the shared space (what they refer to as the ‘interpersonal gesture space’, i.e. the space between the interlocutors). Their regulatory gestures are said to serve functions such as floor and topic management, and when they are deployed the interlocutors reach out of their own, personal space into the shared space. Once the regulatory gesture is complete the interlocutors return to gesturing in their personal spaces.

The implication so far from the literature is that, even when considering a dialogue context, it is only speakers who gesture. This is not the case. Furuyama (1993) (cited in Furuyama, 2000) shows that listeners of an instructional dialogue produced gestures, often without concurrent speech of their own. Whilst Furuyama does not provide details on this, it is possible that these non-speaker gestures demonstrate understanding to the speaker and allow the interlocutors to acknowledge their shared common ground.

Whilst these studies all approach conversation and body movements in different ways, the commonality between them all is that body movements, in particular head and hand gesture, contribute to the coordination *between* the interlocutors and facilitate the successful management of the dialogue.

3.4 Sharing content gesturally

When studying individual, or personal, spaces it was demonstrated that speakers load spaces around their bodies with content, creating abstract referents in space. Within the shared space, not only are the locations of the interlocutors important, but the shared space can be loaded with content which becomes interactionally salient. This is commonly done using gestures and the shared space is transformed from a space which can be used for interactional activities, to a space which also contains salient content.

Bekker et al. (1995) show that gestures can be used to create and reference abstract referents (which they term ‘imaginary objects’), and that these referents last the length of the conversation[4]. They note that new gestures can refer to gestures in the past by using the space in which they were previously located. Murphy (2005) reports similar findings when studying architects

[4] In the reported study these were meetings of designers

conversing about a design over a drawing. He noted that, using the plan as a base, architects would gesturally build upwards to show structures and movement through these structures, and that these gestural constructions remained in place in subsequent sections of the dialogue. Gestural 'diagrams' have also been noted without a base as a plan. In a study of kinship representations, Enfield (2005) demonstrates that speakers will orient their bodies towards their intended recipients and then make use of the shared space to deploy complex hierarchies of family trees.

An example of the *interactive* use of abstract referents comes from Olson and Olson (2000) which details a study of engineers' meetings. The meetings involve groups of engineers who are co-located (i.e. in the same room, with a shared space) discussing new ideas for engine parts and some possible modifications to them. It is reported that the engineers would describe a complex idea by drawing it with their hands in the air. Later in the discussion, other people would refer to "that idea" by pointing at the space in which the idea had originally been placed. Furthermore, McNeill (2005) offers an analysis of what are termed 'shared gesture spaces', using a dyadic context. Here, one of the two interlocutors is being rather evasive in their speech, leaving the second interlocutor unclear as to the true meaning of the utterances. However, the two interlocutors exploit the space between them with pointing gestures, and this serves to disambiguate the evasive speech. It is shown that, when the interlocutors point to different areas of space in accompaniment to their verbal utterances, they assign meaning to them. As the conversation progresses these meanings shift and it is only through use of pointing to these spaces and explicit reference to their meanings by the co-participant that the ambiguity in the dialogue is cleared up.

There are also times during dialogue during which the interlocutors directly exploit each other's hands and the spaces around them to co-construct a gesture. Furuyama (2000) documents examples of gesture events, distinguish by their collaborative nature. The common sequence of a collaborative gesture begins with one interlocutor deploying their gesture. This gesture may then be repeatedly modified by both the original gesturer and the other interlocutor. This can be done using pointing gestures with their target as either an area on the hands of the original gesturer, or the space around the hands. It was also noted that further gestures may be created based on the original gesture. In Furuyama's work he examined an origami tuition situation. Instructors were asked to teach learners an origami technique in dyads without using an actual piece of paper. What resulted was the use of the hands to make iconic gestures representing the paper and the folding actions that must be applied to it. This lead to the learners invoking collaborative gestures by pointing directly to the instructor's hands, and manipulating their representation of the paper in line with the semantics of the discussion.

These examples serve to demonstrate uses of gesture to collaboratively depict content. The formation of these gestures are contingent upon the shared space as they exploit the locations of either prior gestures or concurrent gestures, and underline line the fact that the space is constitutively shared with equal access. It is the gestures' deployment in the shared space which makes them meaningful. They also make clear the fact that the space is rarely empty; as conversations unfold, the shared space is exploited and a dynamic layout of interactionally salient content is collaboratively constructed.

3.5 Organising the dialogue at a higher level

A note must be made which concerns the organisation of the dialogue. There are few immediately apparent links to non-verbal behaviour, and as such this section has been made distinct from the previous. The concerns are with the units used to discuss the organisation of dialogue. This chapter has presented a level of organisation which makes use of the individual interlocutors within the dialogue. They are organised by their dialogue role with the canonical form of a three-person conversation structured into the speaker, primary addressee and side participant. An alternative is to consider the unit to be a group of individuals (Lerner, 1993), or parties. Schegloff (1995) puts forward the notion of a party such that within an interaction 'there are fewer parties than there are persons'. This means that there are occasions when a party may contain only one individual, and there may be times when it contains more than one individual. Parties are often made up of people who have a shared experience (such as Ann and Bob's shared experience of walking in the rain, detailed in the Introduction) or are in a relationship. These structures become important within a multiparty dialogue when considering turn-taking. Turn taking, as discussed previously, requires negotiations between two interlocutors. When considering multiparty conversations there are now more people to consider in the coordination of the next turn. One suggestion is that multiparty turn-taking is actually organised at the party level rather than at the individual level seen in dyadic conversations (Lerner, 1993; Schegloff, 1995).

Party structures also influence the responsibilities that interlocutors bear within the dialogue. The people involved in a party share some form of common ground, perhaps a shared experience, relevant to the current conversation. Within the context of the dialogue, one member of the party is able to respond on behalf of the whole party. It is only if the speaker has trouble (for example hesitating or pausing) that the other members have the responsibility to provide a correction (see Eshghi and Healey, 2007; Eshghi, 2009). This means that common ground can be acknowledged for and on behalf of a party.

3.6 Precluding access to the shared space

Perhaps the most effective way to demonstrate the necessity of a shared space for a conversation is to examine the behaviour of interlocutors when access to the shared space is precluded. To do this a medium which allows interlocutors real time visual and auditory access to each other, but does not allow them access to a shared space is needed. Video mediated communication (or VMC) is a form of interaction which fulfils these requirements.

The type of video mediated communication studied here will be the most basic form; there have been attempts to solve some of the problems that will be noted with novel interfaces, but this is not of concern here. The aim is to demonstrate the impact on dialogue when the interlocutors are precluded access to the shared space, not to investigate the evolution or design of all video mediated communication systems.

3.6.1 What is video mediated communication?

Video mediated communication is a form of technologically mediated communication which adds a video channel to a standard audio connection. Whilst there are no formal specifications upon the hardware used, the system usually consists of a video camera with audio which give a 'head and shoulders' view of each of the interlocutors; this view is then displayed on some form of screen. An example of a three way video mediated conversation can be seen in Figure 3.6.

This form of communication differs from face-to-face interaction because, whilst it preserves visual access to each interlocutor, it precludes mutual access to the shared space. Consider diagrams 3.7 and 3.8. In a conversation in a shared space, A and B are able to form mutual gaze. In addition to this, C is able to see this mutual gaze and A and B know that C can see this. In the conversation over a video mediated link there is no longer a shared space, but a series of peer-to-peer video channels. A and B are not able to form direct mutual gaze because of the separation of the camera from the image of the interlocutor on the screen. Even if they were, C would not be able to interpret their mutual gaze.

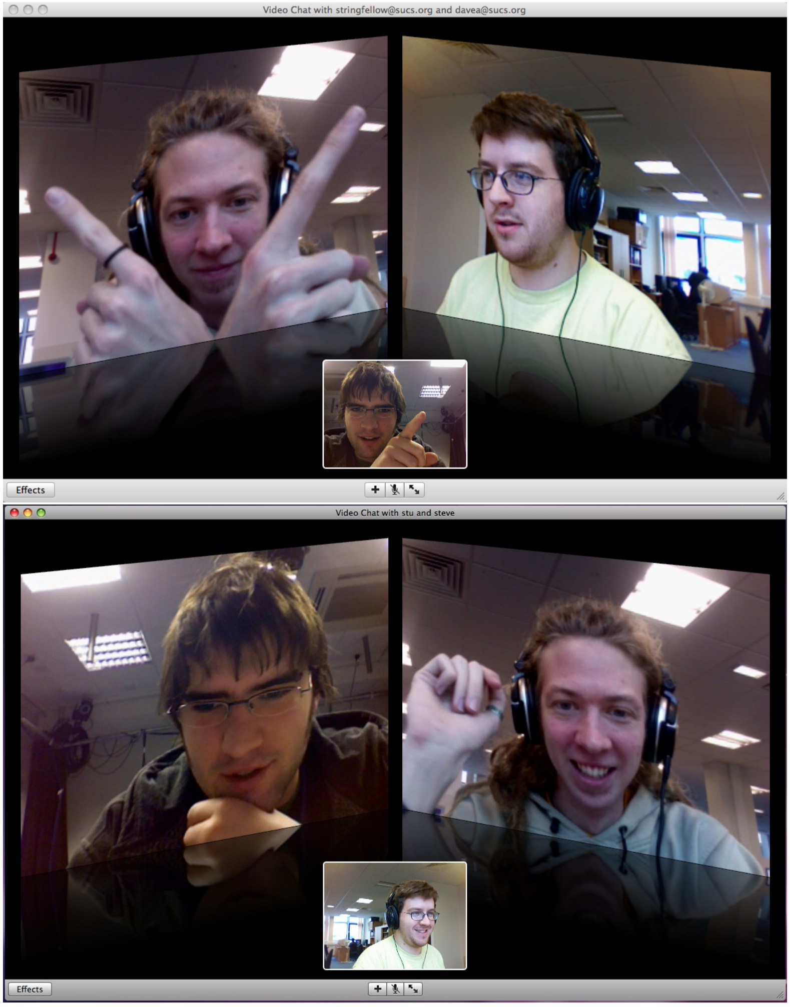

Video Chat with stringfellow@sucs.org and davea@sucs.org
Effects
Video Chat with stu and steve
Effects

Figure 3.6 – A traditional video mediated conversation from two interlocutors' perspectives

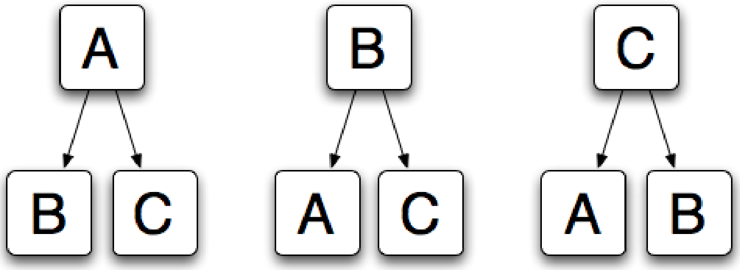


Figure 3.7 – A diagrammatic representation of a multiparty video mediated conversation showing peer-to-peer access provided by the visual channels.

When comparing VMC to both face to face conversations and audio only, such as the telephone, VMC appears to be more similar to audio-only than face to face (Sellen, 1995; Hauber et al., 2006) (although video which provides a shared workspace has been shown to be advantageous, see for example Mondada (2007)) . In observing users of an in-house VMC system,

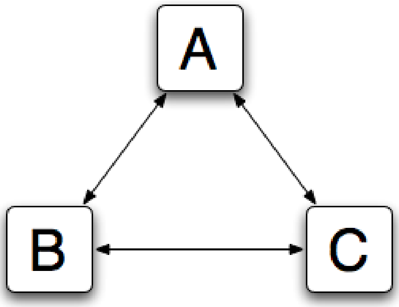


Figure 3.8 – A diagrammatic representation of interaction in a shared space (e.g. face to face interaction).

Heath and Luff (1997) suggest that video mediating technology provides an environment which is still markedly different to that of face to face.

It could be argued that the cause of the problems is based in the technology and that the difference from face-to-face is only because of factors such as poor quality video or delay. To examine examine this, Whittaker and O'Conaill (1993) studied three forms of interaction: mediated over a high quality and high speed network with no delay, mediated over a low speed ISDN network with delay and unmediated face to face. The study made note of many interactional problems, however, importantly for the current discussion, the problems which occurred in the low speed condition also occurred in the high quality and speed condition, but not in the face to face condition. This demonstrates that it is not the technical aspects of the communication channel *per se* which affect the conversation.

3.6.2 The effect on communicative behaviour

Whilst the adverse effects of VMC might be minimal in terms of the denotational aspects of non-verbal behaviour, it is their regulatory functions which appear to be most masked or ineffective. As was demonstrated in this chapter, these are critical to maintaining coordinated conversation.

The most overt problem is that gaze is disrupted. Heath and Luff (1997) studied an in-house VMC system. They found that people attempted to use gaze to initiate a conversation (as was demonstrated in real world work by Mondada (2009)), but that this failed to be noticed. Moreover when in a conversation they found that the lack of mutual gaze and the inability to provide feedback by gaze caused problems in the spoken utterance. The inability to provide backchanneling feedback causes problems for demonstrating continued attention (Whittaker and O'Conaill, 1993). Vertegaal (1997) noted that the lack of effective gaze causes interlocutors to misinterpret who an utterances was addressed to. Gestures also appear to be treated differently to those deployed in face-to-face. Heath and Luff (1997) found that gestures become 'mutated' from their original form and that they don't elicit the expected response (perhaps an attempt to take the floor) from the listener.

The striking finding from studies that examine conversational processes within VMC is that these interactions become more monologic and lecture-like, with lower levels of interactivity than face to face. Turns in VMC are of greater length (van der Kleij et al., 2009; Whittaker, 2003) and there is less simultaneous speech in VMC[5](van der Kleij et al., 2009; Sellen, 1995). Indeed, Heath and Luff (1997) state that there is not any compelling evidence that VMC supports non-verbal interactive cues.

Interlocutors increase their behaviour of other aspects of the interaction, which may be compensatory behaviours. When examining number of words per turn, Whittaker and O'Conaill (1993) saw that interlocutors of video mediated conversations used more words per turn than in face to face. Interlocutors resort to managing turn taking using formal control, such as explicit hand overs by naming the next speaker or using tag questions such as "isn't it" (Whittaker, 2003; Sellen, 1995)

The findings suggest that when access to the shared space is removed, conversations become less interactive. Interlocutors not only have one less resource, the shared space, at their disposal but many interactive cues which depend on it are mutated.

3.7 Summary

This chapter has reviewed literature which provides evidence that, when people are engaged in a face-to-face conversation, they are not simply co-located individuals but a coordinated group. The group works together to maintain the flow of the unfolding conversation and this is done using both speech and the body. Interlocutors' body orientations can be used to show their membership of an interacting group, and their heads and their hands can be used to collaboratively manage the conversation. It should be clear by now that to adopt a person-centric frame of reference in the study of non-verbal behaviour, which may provide insight into the relationship

[5] The definitions of simultaneous speech vary across the literature, some of which appear to include speech that would normally be classed as backchannels.

between speech and gesture, would simply be insufficient to understand dialogue because it is precisely the behaviours which coordinate the group which would be masked. A similar argument in this vein which discusses the communicative properties of gestures is found in de Ruiter (2006).

Throughout this chapter, where possible it has been noted that the communicative features described may not be the same in multiparty as they are in dyadic interaction. In particular the structure of the conversation around turns at talk and the exchanges between these turns require much higher coordination within multiparty conversations. The literature has suggested, although slightly vaguely, that there is a difference between a primary addressee and a side participant. It seems that there is a stronger relationship between the speaker and the primary addressee, than between the speaker and the side participant. However, there is a gap in the literature when trying to explain exactly what it is that differs between the primary addressee and the side participant, particularly with respect to their non-verbal behaviour. Indeed, when exploring interlocutors' non-verbal behaviour it has sometimes been challenging to make direct claims about their behaviour in a multiparty conversation due to the relative shortage of literature in this niche.

The shared interaction space, which is encapsulated by the f-formation, has been identified as a space in which conversations take place. It is unique to co-located face-to-face interaction, although attempts have been made to recreate a sense of co-location through virtual environments and telepresence (see, for example, Slater and Usoh, 1994). The shared space is a resource that can be exploited for communicative purposes and is used in almost all conversations, at the most basic level as a ratification device. It was seen that, not only are interlocutors sensitive to its presence, but other communicative resources such as mutual gaze are dependent upon it; when the shared space is removed there is an apparent shift in the behaviour of interlocutors. The types of activities that are seen as regulatory in face to face dialogue, such as turn taking, are less efficient and interlocutors start to adopt different compensatory behaviours which shift the interaction more towards a monologue. This suggests that the efficacy of non-verbal communication is not due to visual and auditory access *per se*. Without their deployment within a mutually accessible shared space, they do not have the same interactional effects.

The notion that the shared space facilitates coordination behaviours within a dialogue also needs further exploration. Whilst a number of studies presented here have shown that there is an effect on the interactiveness of a conversation when the shared space is manipulated, it is still not entirely clear how the shared space is used in co-located face-to-face interaction. What role does an interlocutor's body play in exploiting the shared space? This brings about a further question: What happens when dialogues shift from dyads to multiparty? Is there a different use of the shared space? Özyürek showed that the interlocutors behave differently with different configurations of the space, but there exist no studies which attempt to address the comparative importance of a shared space with respect to the differential dynamics that come about with scaling up to multiparty from dyadic dialogue. At the current level of granularity, we simply know that the shared space is an important resource for interaction.

Part II

Methods

Chapter 4

Analysing Multiparty Interaction

4.1 Introduction

This thesis is concerned with the analysis of shared space during multiparty interaction. In particular, with the ways in which people coordinate non-verbally within it. This gives rise to the need for data, tools and methods which enable relatively free-form dialogue to be used as tractable data for analysis.

In contrast to some of the studies reviewed in Chapter 2, the types of data used for this thesis must necessarily be real interactions. For these to generate a useful dataset for current purposes, several criteria need to be satisfied. The interactions should be, as far as possible, spontaneous (not scripted), engaging and should elicit the active involvement of at least three naive[1] people. Given the concern with the use of shared space, any task should avoid topics that are intrinsically spatial which may result in spatial patterns relating more to properties of the content under discussion than due to interactive features.

Due to the interest in the relative positions of multiple heads and hands it is important that the motions of all the bodies in a space can be captured. For this reason 3D motion capture technology is used in conjunction with traditional video annotation methods. This chapter will begin with a discussion of motion capture technology and what analytic problems it addresses, and will the move on to detail the corpora used for analysis. The chapter will finish by detailing

[1] Some studies employ a confederate or experimenter to sit in on interactions and take part in, or attempt to manipulate, them. However, it is likely that the confederate will not behave naturally or systematically across all trials and as such this method is avoided.

the type of data that is collected and the post-processing that is applied to it.

4.2 Motion capture as an analytic tool

4.2.1 The motion capture system

The Augmented Human Interaction lab at Queen Mary houses an optical motion capture system (Vicon MX controlled by Vicon iQ). The system tracks movement in three dimensional space. To do this it makes use of an array of 12 infra-red or near infra-red cameras placed around the lab. These cameras track reflective markers which are attached to the clothing of subjects at a rate of 60 frames per second. The cameras are calibrated before each data capture and as such the system knows the position of each camera, relative to all the others. On each frame of data, each individual camera records a two dimensional image of the markers in its scene. The system combines all 12 of these images to build an accurate, three dimensional scene with coordinates for each marker.

The makers attached to subjects are not active, that is, all work is done by the cameras and the computer system. This means that there are no electronics in the markers themselves (they are rubber balls covered in high visibility tape) and no cables are attached to the participants. The system is able to record multiple people at the same time. Once each capture is complete the data points must be labelled. At the start of this process the system is only aware that there exist a number of markers moving in three dimensional space. The labelling processes informs the system where the markers are on the body (for example, the front right of the head on participant 1). The resulting data exported from the system is time series data with the position of each labelled marker in three dimensional space.

One apparent potential weakness of this system for studying human behaviour is the need for subjects to wear markers. It is challenging to determine the impact of this on their dialogue, although anecdotally it has been observed that subjects tend not to focus on the markers but instead the task at hand. The main justification for using this technology is the volume of accurate spatial data that can be obtained using it.

4.2.2 Advantages over traditional techniques

Whilst motion capture is not a panacea and does bring with it problems itself, it addresses some of the key troubles associated with traditional, video based techniques. The first issue that it addresses is that of that quantity and granularity of the data available. Consider the study by Condon and Ogston (1966). This highly labour intensive approach required frames of data to be manually annotated, and resulted in a study of only three dyads. The motion capture system can gather data at up to 120 frames per second. With 3 data points for each marker on each person and the automated nature of the capture this results in vast datasets (up to 583200 data points per person per minute using a standard upper body marker set) ready for analysis.

Perhaps the most important characteristic of the motion capture system is that it preserves the three dimensional nature of the interaction. As demonstrated in the literature review, face to face interaction takes place in a co-located, three dimensional world. A two dimensional video recording compresses the three dimensional scene down into the two dimensions that it can represent. Whilst information from the 3D scene can be preserved to some extent using multiple video cameras, the motion capture data is inherently 3D as it is only concerned with the 3D position of each marker. In addition, the system enables us to extract the relative position and movement of any two points in 3D space. These features mean that, for the study of factors relating to space and body movements in interaction, motion capture technology is very well suited as it yields an accurate representation of how people's body movements are deployed with respect to each other.

Motion capture also allows for systematic analyses of the data which will, unlike human coders, never vary. When analysing data, features of the interaction can be indexed using a set of parameters, and these parameters will remain consistent. Likewise, the data can be manipulated and combined in a repeatable and systematic fashion (see Section 4.6 for combinations of different interlocutors' data into control groups).

4.2.3 Challenges of motion capture analysis

Although the rigidity and non-interpretive nature of parameterised analysis is an advantage, it also presents a challenge. Human coders are able to understand and interpret the interaction and identify unexpected features or patterns, this is not possible with the motion capture. For this reason it is necessary to be able to integrate the motion capture with other sources of data. In this thesis these sources are speech and human coding data from ELAN (see Section 4.3 for details). Unlike a human coder, the system is not able to understand what it is that people are doing. Consider nodding, a human is able to watch a video and determine that a subject is nodding even if they would find it hard to characterise exactly what it was about the head movement that made them class it as a nod. In motion capture terms, the detection works in reverse and any detection is must be explicitly defined. For this reason we do not try to identify features or events in the interaction, but provide a parameter based index of them (see Section 4.4).

Another issue is that motion capture is susceptible to drop outs in the data. This is where the system is unable to locate a marker for a period of time. This does not affect the other makers in the scene, but it means that throughout the session there will be gaps in the data for some markers. This may happen because of occlusion or light fluctuations in the lab. When the marker is from a rigid structure such as the head, the Python software (described in Section 4.4) will attempt to fill the gap. However, with other structures this is not possible and the data will be missing.

4.3 Corpora

Two three-person corpora are used in this thesis. Groups of three were used as they offer the most tractable form of multiparty conversations. The main corpus, collected in the course of this research and known as the Tuition Task corpus, will be used throughout. The alternative corpus, known as the Balloon Task corpus, will be used in order to assess the possible effects of task characteristics on the kinds of non-verbal coordination observed in Chapters 5 and 6. The alternative corpus was collected by Mary Lavelle[2].

4.3.1 The main corpus: Tuition Task

4.3.1.1 Pilot studies

Pilot studies were carried out in order to test the appropriateness of the task and equipment setup. In a series of studies, triads carried out different tasks including acting out a scene from a pub, describing the layout of their houses and performing tuition tasks. The acting task was rejected as the acted interaction did not meet the requirements of being spontaneous and unscripted. The task which required a description of the layout of the house was also rejected as this elicited spatial patterns of gesture which were intrinsic to the house; participants could potentially locate the lounge on their left hand side because of its location in the house. The tuition tasks proved to be successful, requiring only modifications to the motion capture setup and the delivery of the

[2] PhD student at Queen Mary University Of London

tuition material to the subjects.

4.3.1.2 Participants

33 participants took part in the study in groups of 3 meaning the corpus consists of 11 triads. The participants were aged between 18 and 34, and included 14 males and 19 females. They were recruited from the student population at Queen Mary, and studied either English, Computer Science or Psychology. All participants were fluent English speakers, some of whom were native, others spoke English as a second language. Each participant was paid 10 pounds for their time and potentially received module credits. The total length of the corpus is 2 hours and 54 minutes.

4.3.1.3 Task & procedure

The corpus was collected in the Augmented Human Interaction (AHI) lab at Queen Mary. This lab houses an optical motion capture system and video recording equipment. Six tuition tasks were developed that consisted of either a short computer program or a description of a system of government. The material was all text based with no graphical representations to attempt to avoid potential bias on any gestural representations.

Each group performed three rounds of tuition. On the first round one member of the triad was assigned the role of 'learner' and the remaining two members were assigned the roles of 'instructors'. On the subsequent rounds these roles were rotated so that each person was a learner once. On each round the instructors were given printed tuition material which they were asked to collaboratively teach to the learner. They were allowed to familiarise themselves with the material prior to tuition and then returned it to the experimenter. During this time the learner was removed to another room. The learner and the instructors then sat on stools in the AHI lab and the tuition commenced. The stools were placed to create an approximation of the f-formation involving the three subjects, with mutual access to a shared interaction space (see Figure 4.1). The mobility of this f-formation was restricted as the subjects were seated which was a necessity to ensure that the participants were located within a space with good coverage by the motion capture system. There was no time limit and were no restrictions other than they were not allowed to use pen and paper. To motivate the participants to teach and learn, a post-completion test was used. By the nature of the task, all three subjects were involved in the dialogues and as such can be considered conversational participants. There were no overhearers[3]. This remained constant

[3] The exception to this could be the experimenter who could not be seen, as they remained behind a curtain, but could be spoken to in order to signify completion of the task

through out all data capture sessions.

Computer Code Task The Computer Scientists' material was computer code, written in Java. Computer code was selected as the subject material in an attempt to avoid any artefacts in the participants' use of space that could result from an intrinsically spatial domain (for example, this may occur when using tasks that involved giving route descriptions or layouts of houses). On each round the two instructors were given one printed copy of a Java application with its associated class hierarchy. Each Java class was printed on a separate sheet (the order of which was randomised). Three different Java applications were used: 'Student', 'MP3Player' and 'Retailer'. Each application was designed such that it avoided any spatial features.

Government Task The remaining participants' material was descriptions of government structures. This material included an overview of the government, including which bodies are authoritative over whom, and the roles and responsibilities of certain departments. Three different governments were used: Jersey, Indonesia and Saudi Arabia; the information was doctored to attempt to reduce any spatial features. On each round the two instructors were given one printed copy of the government description, consisting of an overview on one sheet of paper, and a series of departments, each on individual sheets of paper. The order of these departments was randomised.

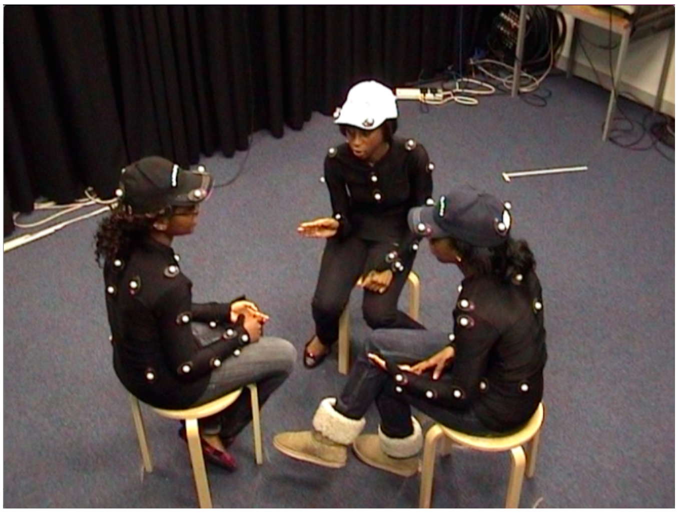


Figure 4.1 – An example scene showing the arrangement of participants in the Augmented Human Interaction lab

4.3.2 The alternative corpus: Balloon Task

This corpus, collected by Mary Lavelle, is used in a study of social interaction in schizophrenia. For current purposes we make use of only the control groups, which total 1 hour and 49 minutes of interactions.

4.3.2.1 Participants

The corpus consists of 45 people, 23 male and 22 female, aged 18-51 who were paid 15 pounds for their time. They were recruited from non-academic staff at Queen Mary and via local community websites.

4.3.2.2 Task & procedure

The corpus was collected in the Augmented Human Interaction lab using the same equipment setup as the tuition task corpus. Each group performed multiple conversation tasks, however only the balloon task conversations are used here. These were always the first conversation to take place. Participants were asked to sit on the prepositioned stools in the lab, and were given instructions for the task verbally. In this task participants are presented with a moral problem, in which they must decide who is to be thrown out of a falling hot air balloon in order to save the lives of the rest of its passengers. Each passenger is described, including their name and their occupation. Participants are given the instructions as a group and there are no rules about what they can say or how they can tackle the task.

4.4 Analysing the interaction

4.4.1 Data collected

Motion data from the head and hands are used to encapsulate participants' gesturing, head nodding and head orientation. These features were selected on the basis of the literature review; it was seen that both head and hand movements have the potential to serve as coordination mechanisms within dialogue. To capture address cues such as those detailed by the face address system (Kendon, 1990), head orientation is measured. As eye gaze can't always be effectively judged at wide rotations (Loomis et al., 2008), it is predicted that head orientation will be a meaningful cue as it is mutually manifest to all participants in the multiparty interaction. In order to make effective use of the motion data, simple methods to index the features are required.

4.4.2 Speech annotation and manual coding

ELAN was used for all hand coding of events and for transcribing participants' speech. To transcribe the speech, ELAN's waveform viewer was used to ensure timing accuracy (see Fig. 4.2). This transcription was done by the author by hand and was very verbose. All verbal contributions were classed as speech, including feedback speech such as "mm hmm". For annotation of subjective events, a subset of randomly selected events was annotated by a second coder and reliability measures (Cohen's Kappa) calculated. The specifics of hand annotations will be presented as part of the methodology in Chapter 7.

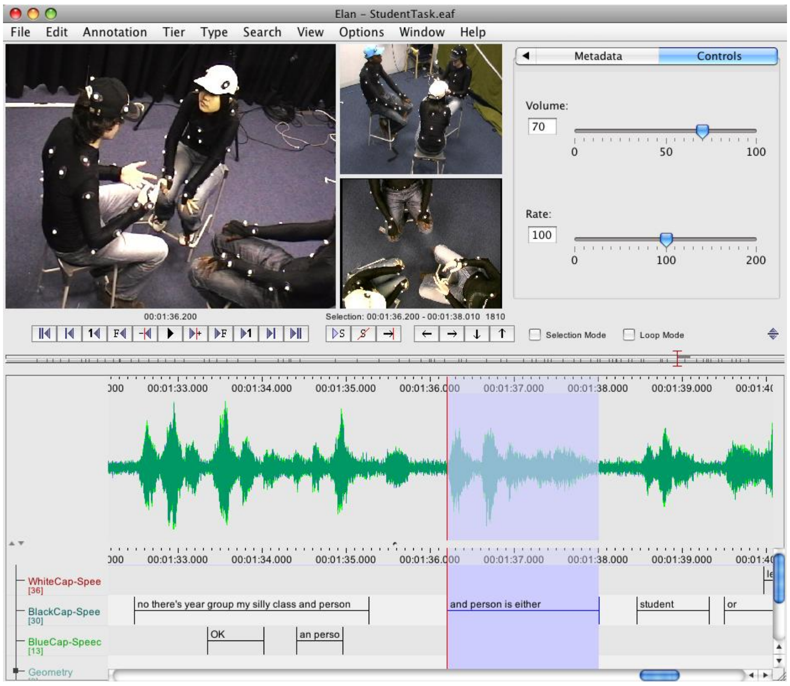


Figure 4.2 – A screen capture showing the speech coding of videos in ELAN

4.4.3 Post-processing

For each session of data there exists a labelled motion capture file, which is exported from Vicon iQ, and an ELAN XML file which contains timing information for speech and potentially hand coded events. Analysis software has been written using Python[4] that reads the motion and annotation data files for an entire corpus. The timings from each data source are aligned such that the speech and motion capture data match. Following this, various measures are taken of the interactions, and quantitative analyses are performed (these will be detailed in subsequent

[4] http://www.python.org/

chapters). SPSS (versions 18 and 19) or Microsoft Excel were used for all statistical tests.

4.4.3.1 Assigning recipient roles

The speaker is determined from the hand annotated speech. Primary addressee and side participant roles are approximated to an operational definition of primary and secondary recipients. This definition of recipiency is based upon the head orientation of the speaker, allowing for the hypothesis that head orientation is a significance cue in multiparty interaction to be tested. For each frame of data, the head orientation is calculated from four markers on the speaker's baseball cap. This orientation is then compared to a centre line which bisects the interaction space. This is calculated based upon the two recipients' positions and is indexed to the rear of the speaker's head. If the current head orientation falls within two degrees of the centre line, this data is excluded. Otherwise, whichever recipient inhabits the same side of the line that the speaker's head orientation falls into is assigned the role of primary recipient, the other becomes the secondary recipient (see Fig. 4.3). Therefore for each time point, assuming complete motion capture data exists or can be accounted for, it is known who the speaker (based on manual speech annotation), the primary recipient and the secondary recipient are (based upon the head orientation of the speaker). These definitions of recipiency are used throughout the thesis.

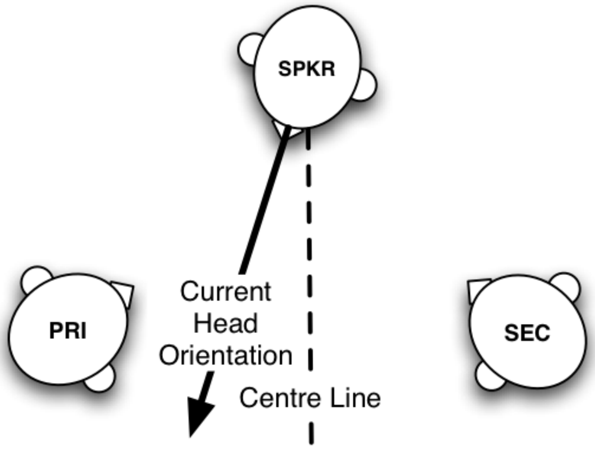


Figure 4.3 – Using speaker head orientation to assign recipient roles

4.4.3.2 Indexing head movements

As measures of interlocutor head movements, two indexes are created: rotations of the head (an approximation of changes in orientation) and movements in the vertical plane (an approximation of head nodding).

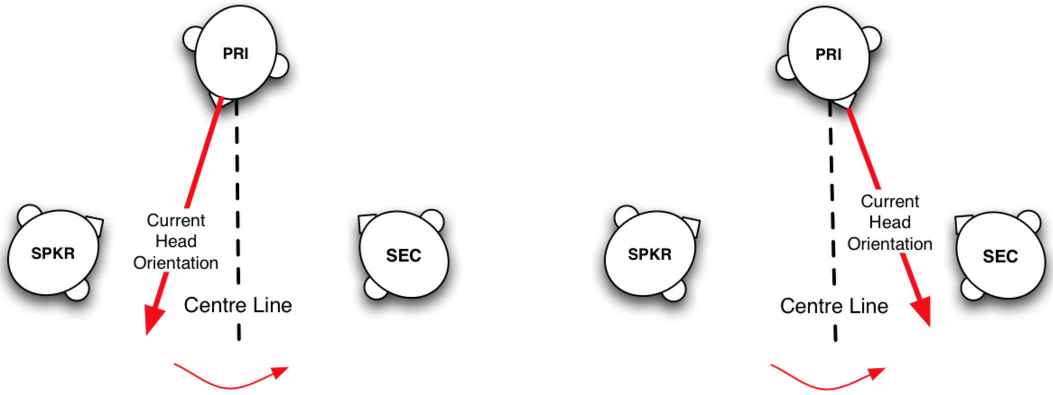


Figure 4.4 – Indexing re-orientations of the head, here those of the primary recipient

Head Rotation Following a similar process as described for assigning recipiency, the current head orientation is extracted for each person in the interaction. The next step looks through the individual's head movements for the interaction and indexes all the times at which their head orientation crosses the centre line (see Fig. 4.4), switching orientation from one interlocutor to another. These moments are listed as changes in orientation.

Head Nodding The aim of this measure is to index all time points at which a person is producing vertical head movement, an approximation of head nodding[5]. The data used for this measure come from a front head marker in the vertical axis. This coordinate data can be interpreted as a signal (over time). As this signal is global to the whole 3D scene, along with head movement it will contain a variety of movement including prosodic body sways, gross shifts in posture, unintentional body movements, and more (see Figure 4.5). Signal processing techniques are applied in order to filter out some of the unwanted information. The signal is first shifted such that zero is the mean position of the head marker. This gives values that show movement above and below this mean position. Next, any low frequency movements, potentially caused by body sways, are excluded by applying a 1Hz high-pass filter. High frequency movements, potentially caused by body shakes or camera error, are removed next by applying a 4Hz low-pass filter. The result is a range of frequencies which narrow down the data to movement of the head. These frequencies are in accordance with those described as the parameters of normal head movement

[5] The term 'head nodding' is used with caution here as this requires a definition of what exactly a head nod is. For the current purposes the analysis is restricted to one based purely upon motion data

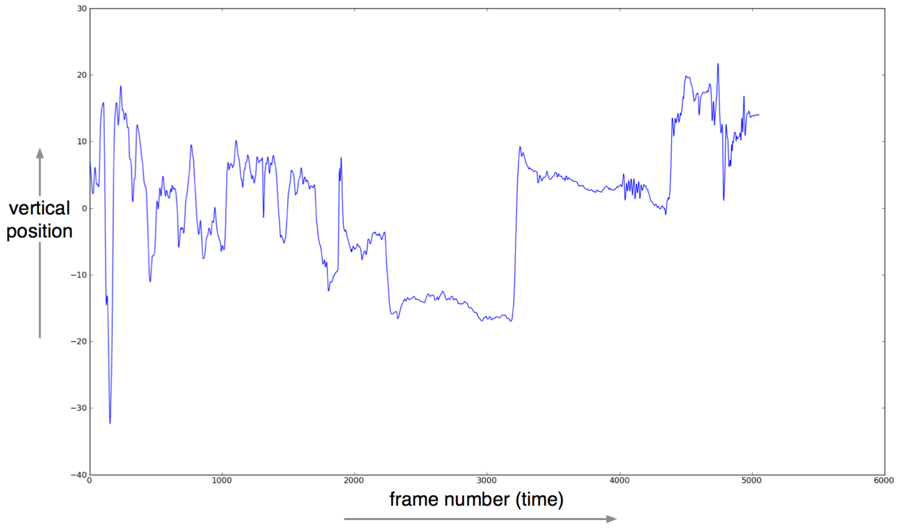


Figure 4.5 – Raw motion capture data of one participant's vertical head movement over one interaction. The vertical axis is the position, the horizontal axis is frame number (or time).

in the British Journal of Ophthalmology (Gresty et al., 1976), and are inline with findings from Hadar et al. (1983) who observe a peak frequency of head movement in speakers at 2.51Hz which falls into their range of 'ordinary movements' which are between 1.8 and 3.7 Hz.

Peak and trough detection is applied to the filtered data to identify the times at which the movement direction changes (i.e. the top and bottom of a nod). Only movements with greater than 7 frames of data between a peak and a trough are considered to possibly be a nod. If the rate of change is greater than 0.3mm/frame the participant is considered to be nodding at this time point. These two values are used to index potential head nodding within the head movement signal and are based on findings from a study of Swedish speakers' head movement (Cerrato and Svanfeldt, 2005).

4.4.3.3 Indexing hand movements

In order to create a measure of interlocutor hand movement activity, the speed of each hand is calculated. The data for this measure come from a single marker placed on the back of each of the hands. The speed of each hand is calculated using the distance moved in three dimensional space between frames of data. This results in a mm/frame unit (where there are 60 frames per second). A check is in place to ensure that there have been no data drop outs, which would result in erroneously inflated hand speeds for one frame. If there has been data drop out, then data for the affected times are excluded.

To provide a single measure of hand speed, at each time point the maximum speed of the interlocutor's two hands is reported. This is done to encapsulate any gesturing performed by the interlocutor in one measure. When there is the need to provide a binary decision of whether an interlocutor is gesturing or not, the mean and standard deviation of hand speed for each individual is calculated across each interaction. At any point in time, if an individual's hand speed is greater than one standard deviation from the mean they are classed as gesturing.

4.4.3.4 Creating control groups

As the motion capture data is held as discrete, labelled position data for each person, it is possible to manipulate the structure of the interactions. The motivation for doing this is to allow for the investigation of spurious correlations which have been attributed to interaction (for example, the Condon and Ogston (1966) study may suffer from this), and will be detailed further in Chapter 5. One possible manipulation is to break up the triads of interlocutors and match them up into new interactions such that the constituent members of the new interactions did not originally interact with each other. These are termed control group interactions (see Fig. 4.6).

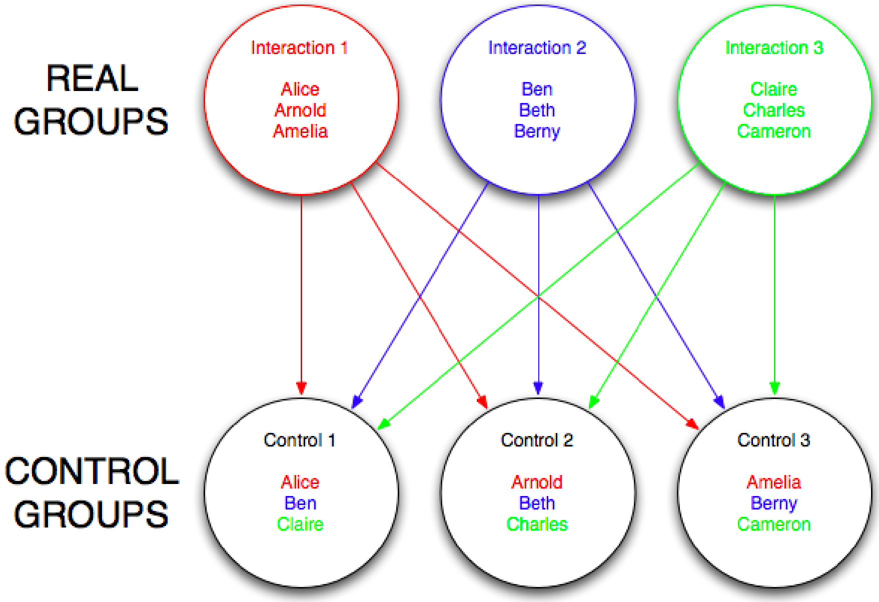


Figure 4.6 – Creating three control group interactions from three real interactions

Control groups are constructed by first breaking the triadic grouping of the interactions resulting in a list of interlocutors. As each individual took part in three interactions in the main corpus, that individual will appear three times in this list. The order of this list in then randomised. The resulting list is then used to construct the new triads. Each consecutive three people are assigned to a control group. Any interactions which contain interlocutors who were from the same original interaction are excluded, along with any interactions that contain the same interlocutor twice. The remaining interactions will now contain interlocutors whose original interactions were of different lengths. To normalise them, the shortest original interaction is identified and the remaining two interlocutors' data is removed from this point on. As each original interaction has its data indexed by the time of each frame of data (determined by when the motion capture system took the frame of data), the control group interactions will have different time indexes across its three members. One member of the triad's timing indexes are chosen, and the remaining two members' data points are re-indexed to match these. The resulting control group interactions can now be treated and analysed in exactly the same way as real interactions.

For the control group interactions the ratio of instructors to learners is not consistent across the dialogues. In the real conversations there are always two instructors and one learner. In the control group dialogues there may be any combination of instructors and learners within the triple. This is done to preserve the quantity of data in the control group corpus.

4.5 Statistical significance

This thesis will make use of a criterion level of 0.05 throughout. Values greater than this will not be considered significant. Values which fall between 0.05 and 0.1 whilst not considered significant will be considered to show a trend. This is in line with the exploratory theme of this thesis.

Part III

Exploring Multiparty Interactions

Chapter 5

Testing coordinated behaviours

5.1 Introduction

The concerns of this chapter are focused upon the behavioural relationships that may exist between interlocutors. The position that has been built up over the literature review is that during an interaction, the presence and actions of one person in a dialogue have an influence over the behaviours of the other people in the dialogue. However, potential problems exist with some of the methodologies used to make these claims. Consider the study by Condon and Ogston (1966). Their coordination findings, which were attributed to interaction, may exist due to the presence of statistical artefacts and not to inter-person coordination in conversation[1]. The results could be explained by non-interactional sources of correlation such as collective fatigue, the temporal structure of the task at hand or constraints due to seating arrangements. McDowall (1978) expresses similar concerns, suggesting that coding errors may exist in the manual methods used and that the correlations found may not be due to the conversation the people are engaged in.

In this chapter the availability of motion capture data is exploited to question the interactive patterns of behaviour within dialogue more fully. It asks whether or not people coordinate and structure their movements with the people they interact with. Put another way, this chapter aims to investigate whether patterns of detectable inter-person head and hand movement exist within a group of interacting people that are systematically different to the inter-person patterns measured

[1] Another concern, albeit less intrinsic to their approach, is that a particularly small sample size was used, likely due to the time consuming methodology.

across non-interacting individuals. The most effective way to do this is to factor out the inter-person relationships that are being analysed and see if this has any measurable effect on the statistical correlation of the interlocutors' behaviour. This chapter applies two different methods to do this, both will be detailed below.

As mentioned, it is necessary to consider the possible sources of coordinated movement that are not due to the interaction. Thus it is also questioned whether the interactions within the corpus share a common temporal structure. For example, the task itself may have a temporal organisation which gives rise to certain interactional phenomena at common time points. These could include a heightened necessity to show understanding or more turn exchanges, regardless of the actual inter-person relationships. This chapter makes an approximation of these key points by employing a relatively simple method which divides the interactions up temporally and analyses the frequencies of behaviours within these segments.

At this initial stage of the thesis all tests will be applied independently to head movement, hand movement and speech (see Chapter 6 for analyses which consider the modalities together). The tests make use of binary variables based upon the motion capture and speech annotation data described in Chapter 4. This chapter will first apply the tests which factor out interaction and, following this, will test if there are common temporal structures across the corpus. All tests will be applied to both the tuition task corpus and the balloon task corpus.

5.2 Method

5.2.1 Testing inter-person relationships

This section performs four tests: 1) identifies the inter-person relationships that exist in the interactions, 2) factors out the inter-person relationships by making use of artificially created control groups, 3) factors out the inter-person relationships and the temporal structure of the interactions using randomisation and 4) performs a direct statistical comparison of the relationships present in real interactions with the control groups and randomised data from 2) and 3). By performing the two step factoring of structures, the tests identify the extent to which relationships observed in the corpus can be attributed to inter-person coordination and the extent to which they can be attributed to non-interactive sources of coordination that follow a common temporal structure. These tests do not detail what the temporal structures may be, this is addressed by subsequent tests, they merely identify their presence (or lack of).

5.2.1.1 Calculating relationships for real interactions

In order to understand the inter-interlocutor relationships of speech, head movement and hand movement three independent tests are performed, one for each of these modalities. To index when people are speaking, a variable called isSpeaking is generated (along with counterparts for nodding and gesturing). To index whether anyone other than this person is speaking in the interaction a variable called otherSpeaking is generated, with counterparts. The tests therefore make use of two binary variables, one representing an individual's behaviour and one representing the combined behaviour for the remaining two interlocutors. The variables for analysis are:

- isSpeaking together with otherSpeaking
- isNodding together with otherNodding
- isGesturing together with otherGesturing

Each of these variables can be either 1 or 0. For example, with an interaction consisting of Ann, Bob and Claire, when we analyse Ann at a given time point her isNodding variable will be 1 if she is nodding, 0 otherwise. If either Bob or Claire are nodding at this time then Ann's otherNodding variable will be 1, or 0 if neither of them are nodding. These measurements are taken at each frame of data and for each person in the interaction. These data across the whole corpus are entered in a binary logistic regression analysis (a binary logistic regression is appropriate because the dependent variable will always be binary). To account for the fact that individuals tend to move in self-similar ways the regression accounts for the ID of the individual under analysis and a unique pair ID of the remaining two interlocutors. This is in order to account for within individual variation. Three regressions are performed across the whole corpus, one for each modality. Each regression tests whether the data for an interlocutor in one of the modalities can be predicted by the behaviour of the other interlocutors in the same modality. Hence, when testing head movement the dependent variable will be isNodding and the independent will be otherNodding. Each regression produces an Exp(B) value that describes the relationship between the dependent and independent variables. If this is 1.0 there is no relationship between the independent and the dependent variables. If it is greater than 1.0 the relationship is a positive one, if it is less than one the relationship is a negative one. The Exp(B) value also has a corresponding significance value.

5.2.1.2 Factoring out the inter-person relationships with control group interactions

To create a comparison level of inter-person behaviour the same regression tests are applied to control groups. These are constructed in line with the method described in Section 4.6 and preserve the temporal structure of the interactions but are made up of interlocutors from different original interactions.

To ensure that the results are not distorted by artefacts in the construction of the artificially created control groups, such as chance pairings of correlated interlocutors, the process is repeated multiple times. The randomised creation of control groups and the subsequent regression analysis are performed 100 times. The resulting Exp(B) and significance values are averaged to produce a mean Exp(B) and mean significance level.

5.2.1.3 Factoring out the inter-person relationships and temporal structure with randomisation

The second method to factor out the inter-person relationships differs from the previous method by also factoring out the temporal structure of the interactions. The data for this test are constructed in the same way as the data for the initial regression tests applied to the interactions in the corpus. An additional step is applied before the regression is performed. The data file representing the entire corpus for the regression consists of pairs of variables (such as isSpeaking and otherSpeaking) indexed by the time that each pair was recorded in the interaction. This index is broken by randomising the order of the measure of the other interlocutors whilst leaving the order of the individual's measure constant. Thus the isSpeaking variable will be paired with a value of otherSpeaking which was taken at a different time point in a different interaction. The regression that is then performed tests whether the data for an individual in one of the modalities can be predicted by the combined behaviour of two other individuals taken from a different interaction and a different time point in the interaction, in the same modality.

As with the previous test, the randomisation and regression analysis are performed 100 times to avoid statistical artefacts. The resulting values are averaged.

5.2.1.4 Comparison of real and manipulated interactions

To perform a direct comparison between real and control group interactions, a statistical test is required that can include both sets of data. The regressions do not do this, as they perform two independent tests for each condition. A generalized linear model is appropriate as this functions similarly to a binary logistic regression (allowing binary dependent variables), but also allows for the inclusion of a Condition (real v.s. control and real v.s. randomised) factor. Statistical interactions between the Condition factor and the independents can be calculated. For example, when examining nodding this will enable us to determine directly if the relationship between otherNodding and isNodding varies at the different levels of Condition (real v.s. control). Some of the results of this test will be redundant to the regressions, however it is the ability to directly compare conditions that makes it necessary.

5.2.2 Exploring the structure across interactions

To test if there is a common temporal structure of behaviour across the interactions within the corpus, a simple method is used. Each interaction is broken into quarters. For each quarter the percentage of time spent nodding, gesturing and speaking is calculated for all interlocutors across all interactions. This calculation is made by testing each person on each frame of data for their isNodding, isGesturing and isSpeaking variables and reports percentage values for each modality and quarter. If there is no common temporal structure across the interactions then the percentages reported should not be reliably different.

This test is also applied to control group interactions. As these control interactions are trimmed to the length of the shortest original interaction of its constituent members, the quarter cut-offs here will not be correctly aligned to two of the interlocutors' original interactions. This misalignment will get progressively worse through the quarters. That is, the first quarter will be the most accurate, the final quarter will be most inaccurate. The differences between quarters are tested using multiple non-parametric pairwise comparisons. The test used is a Wilcoxon signed rank test because the data are not assumed to be normally distributed.

Applying this test to both the tuition task corpus and the balloon task corpus, which differ in their task, will help understand to what extent the task is responsible for any observed temporal structures.

Table 5.1 – Results of binary logistic regression analyses for each modality in the tuition task corpus

	Real Groups		Control Groups		Randomised Data	
Modality	Exp(B)	Sig	mean Exp(B)	mean Sig	mean Exp(B)	mean Sig
Speech	0.080	0.000	0.937	0.053	1.001	0.502
Head	1.160	0.000	1.018	0.095	1.000	0.521
Hands	0.890	0.000	1.055	0.064	1.001	0.495

Table 5.2 – Results of binary logistic regression analyses for each modality in the balloon task corpus

	Real Groups		Control Groups		Randomised Data	
Modality	Exp(B)	Sig	mean Exp(B)	mean Sig	mean Exp(B)	mean Sig
Speech	0.126	0.000	0.974	0.042	1.001	0.519
Head	1.151	0.000	1.027	0.041	0.999	0.483
Hands	0.787	0.000	0.995	0.079	0.999	0.507

5.3 Results

5.3.1 Inter-person relationships in real, control group and randomised interactions

5.3.1.1 Calculating relationships for real interactions

The results indicate that apparent inter-interlocutor relationships exist in all modalities, in both corpora. The results for the tuition task are presented in Table 5.1 and the results for the balloon task are shown in Table 5.2. As the reliable Exp(B) values are lower than 1.0 for hand movement and speech, they suggest that interlocutors are less likely to gesture if another interlocutor is gesturing and less likely to speak if another interlocutor is speaking. As the reliable Exp(B) values are greater than 1.0 for head movement, they suggest that an interlocutor is more likely to nod if another interlocutor is nodding.

5.3.1.2 Factoring out the inter-person relationships with control group interactions

The results of testing with control groups indicate different patterns in all three modalities from the real groups. In both corpora, for speech, head and hand movement the value of the relationships are closer to 1.0 (the no effect value). In the tuition task corpus these relationships are insignificant, but still reach trend significance; in the balloon task corpus the relationships for speech and head movement reach significance, and hand movement reaches trend significance. This indicates that the inter-person statistical relationships are less strong in control groups, but the remaining significance, and trend significance suggests that there may exist some residual statistical relationship.

5.3.1.3 *Factoring out the inter-person relationships and temporal structure with randomisation*

When both the temporal structure and the interaction are removed there is a marked effect on the statistical relationship between interlocutors. Whilst in real interactions there is a significant relationship between interlocutors for head movement, hand movement and speech, this is removed entirely in the randomised interactions. There is no significant relationship between the head movement, hand movement or speech of interlocutors who didn't interact when there exists no common temporal structure.

5.3.1.4 *Direct comparison of conditions*

There exist reliably different inter-person relationships between real interactions and control groups, and real interactions and randomised data in both corpora. When using data from the tuition task corpus, there is an effect of Condition x otherSpeaking ($Wald\ \chi^2 = 219194,\ p = 0.00$), Condition x otherNodding ($Wald\ \chi^2 = 686,\ p = 0.00$) and Condition x otherGesturing ($Wald\ \chi^2 = 129,\ p = 0.00$) when comparing real and randomised data, and an effect of Condition x otherSpeaking ($Wald\ \chi^2 = 146925,\ p = 0.00$), Condition x otherNodding ($Wald\ \chi^2 = 259,\ p = 0.00$) and Condition x otherGesturing ($Wald\ \chi^2 = 60,\ p = 0.00$) when comparing real and control groups. The marginal means for each of these tests are shown in Tables 5.3 and 5.4.

When using data from the balloon task corpus, there is an effect of Condition x otherSpeaking ($Wald\ \chi^2 = 94801,\ p = 0.00$), Condition x otherNodding ($Wald\ \chi^2 = 306,\ p = 0.00$) and Condition x otherGesturing ($Wald\ \chi^2 = 347,\ p = 0.00$) when comparing real and randomised data, and an effect of Condition x otherSpeaking ($Wald\ \chi^2 = 75437,\ p = 0.00$), Condition x otherNodding ($Wald\ \chi^2 = 656,\ p = 0.00$) and Condition x otherGesturing ($Wald\ \chi^2 = 31,\ p = 0.00$) when comparing real and control groups. The marginal means for each of these tests are shown in Tables 5.5 and 5.6.

Table 5.3 – Estimated Marginal Means for real vs control group data in the tuition task corpus

Modality	Condition	Marginal Means		Sig
		No	Yes	
Speech	Control	0.276	0.300	0.000
	Real	0.580	0.104	0.000
Head	Control	0.193	0.198	0.000
	Real	0.193	0.216	0.000
Hand	Control	0.128	0.125	0.001
	Real	0.133	0.120	0.000

Table 5.4 – Estimated Marginal Means for real vs randomised data in the tuition task corpus

Modality	Condition	Marginal Means		Sig
		No	Yes	
Speech	Randomised	0.307	0.306	0.202
	Real	0.574	0.106	0.000
Head	Randomised	0.201	0.202	0.541
	Real	0.192	0.214	0.000
Hand	Randomised	0.130	0.129	0.201
	Real	0.133	0.120	0.000

5.3.2 Structure of interactions

5.3.2.1 Tuition task corpus

Within the interactions used in the tuition task corpus there is no reliable common temporal structure of the interlocutors' speech. Whilst differences exist in the values for each quarter (see Figure 5.1), these fail to reach significance (see Table 5.7). The same lack of patterns occurs in control group interactions (Table 5.8 and Figure 5.1).

For non-verbal behaviour the findings are different. Interlocutors nod more at the beginning of the interaction than towards the end. There is significantly more nodding in the first quarter of

Table 5.5 – Estimated Marginal Means for real vs control group data in the balloon task corpus

Modality	Condition	Marginal Means		Sig
		No	Yes	
Speech	Control	0.288	0.274	0.000
	Real	0.484	0.109	0.000
Head	Control	0.126	0.115	0.000
	Real	0.122	0.138	0.000
Hand	Control	0.099	0.085	0.000
	Real	0.101	0.081	0.000

Table 5.6 – Estimated Marginal Means for real vs randomised data in the balloon task corpus

Modality	Condition	Marginal Means		Sig
		No	Yes	
Speech	Randomised	0.277	0.277	0.646
	Real	0.483	0.110	0.000
Head	Randomised	0.127	0.126	0.127
	Real	0.122	0.138	0.000
Hand	Randomised	0.096	0.096	0.940
	Real	0.100	0.080	0.000

interactions, with the remaining three quarters equal. This pattern is preserved in control group interactions. Hand movement occurs more at the beginning and end of the interactions than in the middle. Quarters 1 and 4 are the same, but have significantly higher levels than quarters 2 and 3, which are equal. In control group dialogues the pattern across the first quarter is kept, but the final quarter is lost, most likely because of the shortening of interactions.

5.3.2.2 Balloon task corpus

In the balloon task corpus the structures are different (see Figure 5.2). As can be seen in Table 5.9, there are no reliably detectable common temporal structures of speech or hand movement

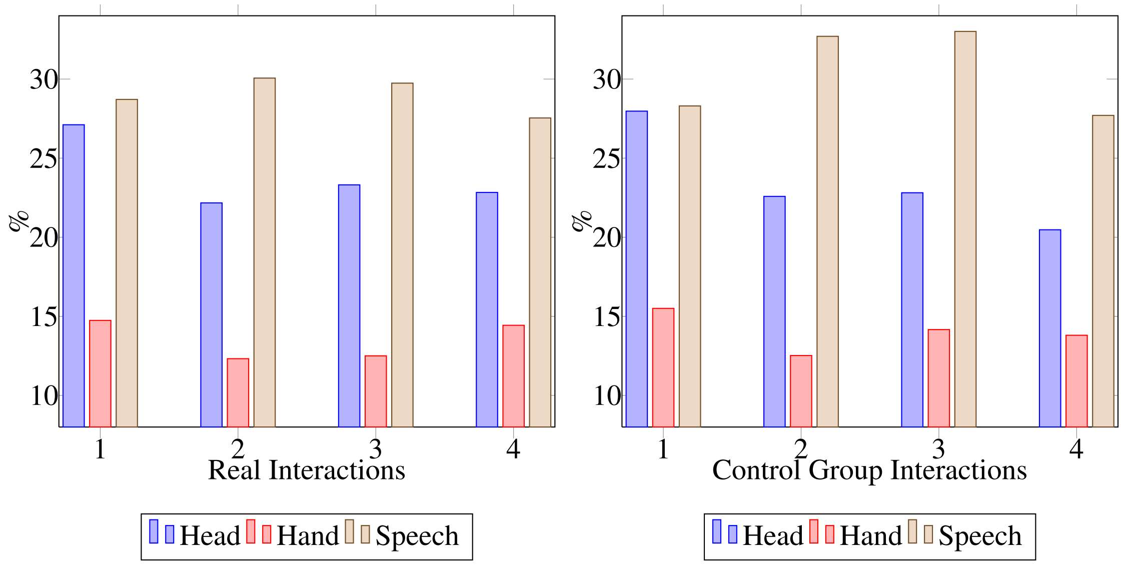


Figure 5.1 – Quarterly breakdown of head, hand and speech across interactions in the tuition task corpus

Table 5.7 – Significance tests between quarters in real interactions in the tuition task corpus

	Speech		Head		Hands	
Quarters	Z	Sig	Z	Sig	Z	Sig
Q1-Q2	-0.764	0.445	-3.037	0.000	-2.307	0.021
Q1-Q3	-0.963	0.336	-3.644	0.000	-2.318	0.020
Q1-Q4	-0.416	0.677	-2.351	0.019	-0.063	0.950
Q2-Q3	-0.014	0.989	-0.607	0.544	-0.038	0.969
Q2-Q4	-0.688	0.492	-1.312	0.189	-2.464	0.014
Q3-Q4	-1.110	0.267	-1.183	0.237	-2.712	0.007

Table 5.8 – Significance tests between quarters in control group interactions in the tuition task corpus

	Speech		Head		Hands	
Quarters	Z	Sig	Z	Sig	Z	Sig
Q1-Q2	-1.352	0.176	-3.080	0.002	-2.584	0.010
Q1-Q3	-1.083	0.279	-2.501	0.012	-2.670	0.008
Q1-Q4	-0.728	0.467	-3.398	0.001	-2.162	0.031
Q2-Q3	-0.287	0.774	-0.368	0.713	-0.076	0.940
Q2-Q4	-0.017	0.987	-0.058	0.953	-0.248	0.805
Q3-Q4	-0.187	0.851	-0.666	0.505	-0.509	0.610

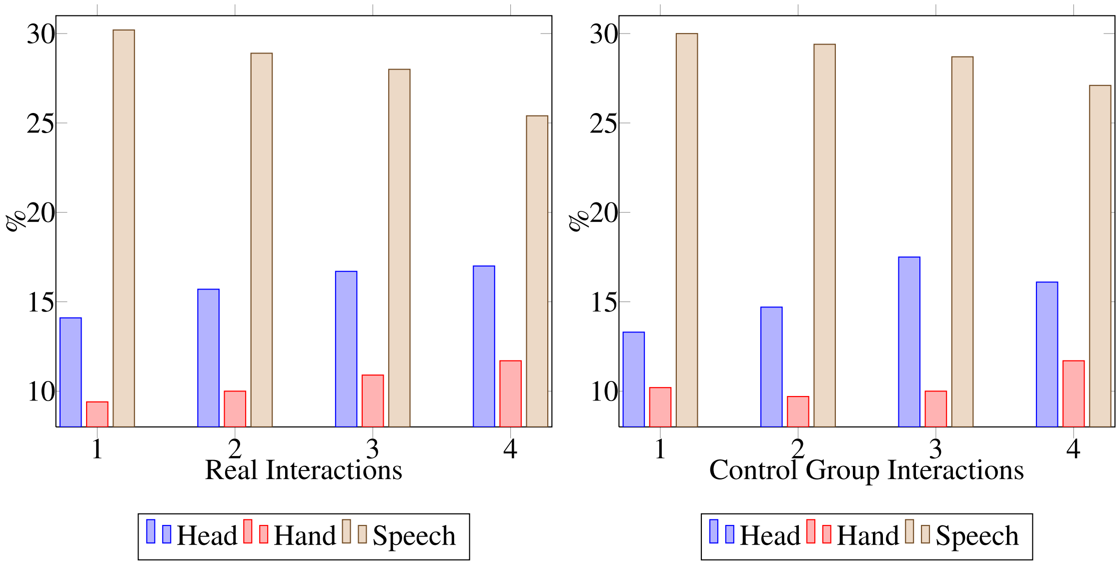


Figure 5.2 – Quarterly breakdown of head, hand and speech across interactions in the balloon task corpus

Table 5.9 – Significance tests between quarters in real interactions in the balloon task corpus

	Speech		Head		Hands	
Quarters	Z	Sig	Z	Sig	Z	Sig
Q1-Q2	-0.615	0.538	-1.462	0.144	-0.751	0.453
Q1-Q3	-1.292	0.196	-1.800	0.072	-0.642	0.521
Q1-Q4	-1.647	0.100	-2.791	0.005	-1.354	0.176
Q2-Q3	-0.976	0.329	-1.394	0.163	-0.408	0.683
Q2-Q4	-1.005	0.315	-1.800	0.072	-0.502	0.615
Q3-Q4	-1.507	0.132	-0.237	0.812	-0.070	0.944

Table 5.10 – Significance tests between quarters in control group interactions in the balloon task corpus

	Speech		Head		Hands	
Quarters	Z	Sig	Z	Sig	Z	Sig
Q1-Q2	-0.506	0.613	-1.157	0.247	-0.227	0.821
Q1-Q3	-0.503	0.615	-1.697	0.090	-0.079	0.937
Q1-Q4	-0.627	0.530	-1.224	0.221	-1.306	0.192
Q2-Q3	-0.299	0.765	-1.838	0.066	-0.283	0.777
Q2-Q4	-0.627	0.530	-1.114	0.265	-1.409	0.159
Q3-Q4	-0.051	0.959	-0.401	0.688	-1.717	0.086

across the interactions. Head movement increases through the interaction, with reliably more nodding in the final quarter when compared to the first quarter. When comparing the first quarter to the penultimate quarter, and the second quarter to the final quarter the difference reaches trend significance. In the control group interactions there are no significant common temporal structures in any modality, although as can be seen in Table 5.10 there exist some trend patterns in head and hand movement.

5.4 Discussion

This chapter has demonstrated that, along with speech, interlocutors coordinate their non-verbal behaviour. By using methods that were unavailable to Condon and Ogston, it has been shown that some of this apparent co-ordination can be attributed to non-interactive factors associated with a common temporal structure. When potential inter-person effects are factored out using control groups, there is a marked effect on the statistical relationship between people's behaviour (as shown by the results of the generalised linear model), however weak correlations still exist. It is only when the temporal structure of the interactions is broken down as well, that the relationships are completely removed. These findings add greater validity and detail to those of Condon and Ogston (1966). By testing high volumes of data, across two independent corpora, it has been shown that an interlocutor's non-verbal behaviour is influenced by the other interlocutors' behaviours, and that it follows a common temporal structure which may result from non-interactive causes such as the task at hand. That the differences in effect sizes for all modalities are greater between real and control groups, than between control groups and randomised data implies that most of the coordination can be attributed to interactive features of the conversations, with a smaller amount attributed to their non-interactive temporal structure.

By examining the temporal structure across two corpora, it has been possible to account for some of the remaining statistical relationships that are present in control groups. In the tuition task corpus patterns were seen in head and hand movement, whilst in the balloon task corpus patterns were reliably found only in head movement. The crude method used to get a handle on the temporal structures showed that the those identified are corpus specific; that is, the patterns seen in the tuition task corpus differ from those of the balloon task corpus. One interpretation of this is that the common temporal structure arises as a result of the task. In the tuition task corpus the interactions commonly begin with the delivery of the tuition material from the instructors to the learners. If head movement is used as a feedback mechanism to demonstrate understanding, this would explain the higher rate of nodding in the first quarter; interlocutors are commonly demonstrating understanding within the interaction at times approximated to the first quarter. Likewise the rates of gesturing could be interpreted at the beginning of the interaction by the instructors teaching the material and hence depicting it gesturally. At the end of the interaction the learners usually recap this and hence will gesture more again. In the balloon task corpus there is an increasing level of head movement towards the end of the interactions. This could be because, unlike the tuition task corpus, there is no immediate requirement to demonstrate understanding but the task does require that the interlocutors converge on an agreement of who to remove from the balloon. It is possible that this increasing agreement is manifest in their head movement.

Whilst the amount of time spent speaking appears generally similar (in the range of 25%-35%) across corpora, the levels of head and hand movement appear to differ. This may imply that the requirements of the task are differentially dependent upon non-verbal behaviour. The balloon task corpus has apparently lower amounts of gesturing, potentially a reflection of the topic under discussion or potentially that coordinating gestures, such as interactive gestures, are carried out more with the tuition task. The tuition task corpus also features apparently higher amounts of nodding. This is in line with the notion that head movement serves as a feedback mechanism, suggesting that in the tuition scenario there is a generally higher need to acknowledge contributions to common ground and demonstrate understanding than there is in the balloon task.

A note must be made on the lack of reliable patterns in the amount of speech in the tuition task corpus, and speech and hand movement in the balloon task corpus. The approach used to detect these is crude and thus it does not accurately detect the precise patterns, only highlights them. It is possible that, at a level of finer granularity than is accessible with this method, temporal structures exist where they have remained undetected. When examining inter-person relationships, it would be desirable to make use of a statistical test which accounts for any effect of temporal structure on the apparent coordination. However, as the current method is unable to precisely identify the structure, it is not possible to accurately account for it statistically.

As an aside, it is also interesting that there are different patterns between modalities. This hints, although does not prove conclusively, that non-verbal behaviour is not redundant to speech; if it were, it is possible that the structures of head and hand movement would be the same as speech.

That they are different both from speech and each other, suggests that there are some aspects of the task which elicit more emphasis on a particular modality.

5.5 Conclusion

This chapter has demonstrated that people influence each other's behaviour, at least on a per-modality basis, within dialogue. Concerns about potential artefacts in prior work, resulting from non-interactive factors, were expressed and investigated. One possible source of artefacts, a common temporal structure across the interactions, was identified and quantified. By comparing two corpora with different tasks, the possibility that this temporal structure was the result of the task was given weight. Even when this common temporal structure remained in place, as the inter-person relationships were broken down using control groups the statistical relationships between the behaviour of people (that people tend to nod together but avoid speaking or gesturing together) were reliably different to those in real groups. This suggests that, accounting for artefacts arising from the common temporal structure, interlocutors coordinate their head movement, hand movement and speech.

However, it is still not clear what, if any, coordination exists between modalities; does the fact that one interlocutor is speaking influence the other interlocutors' likelihood to gesture or nod? This leads into the next point: at this coarse level of analysis it is not possible to say if there are any within group structures that exist with respect to the unfolding dialogue. Within a multiparty dialogue, is the organisation of interlocutors' non-verbal behaviour influenced by their roles in the dialogue? These questions will be examined next.

Chapter 6

Non-verbal cues to dialogue roles

6.1 Introduction

Through the literature review it was demonstrated that interlocutors take turns at speaking, and make use of their bodies as part of production and comprehension. It was shown that movements of both the head and hands can depict content. Moreover, it was shown that interlocutors use head and hand movements to coordinate their dialogue and as this dialogue unfolds, they adopt various different roles. These are dynamic and switch as the dialogue progresses. In dyadic dialogue these roles simply consist of a speaker and a listener. However, it was demonstrated that multiparty dialogue has additional complexities. The introduction of the third person means that, minimally, roles are now considered to be speaker, primary addressee and side participant. Interlocutors are now faced with additional coordination problems that do not exist in the dyadic setting.

The previous chapter demonstrated that interaction has a marked effect on the inter-person statistical relationships that are present in verbal and non-verbal behaviour. The current chapter builds upon this, opening up the interaction to expose the dynamic participant roles that it encapsulates and examining their relationships with interlocutors' non-verbal behaviour.

When considering the recipients in a multiparty dialogue, Kendon (1970) suggested that the speaker and the primary addressee should fall into synchrony based upon observations of group interactions in a bar. However, it is not clear a) which body parts synchronise and b) what the predictions for other participants are. The second question of this thesis aims, in part, to address this and provide a general understanding of the interlocutors' movements within a conversation, including the relative movements of the primary and secondary recipients. As such, the second question that stands to be addressed in this thesis is: 'Are there global patterns of head and hand movements that relate to the unique structure of a multiparty dialogue in terms of its constituent interlocutors' dialogue roles?'. The literature review demonstrated the use of co-speech gestures and highlighted patterns of head movement. For example, it was seen that there exist differences in gaze patterns between speakers and listeners (Kendon, 1977; Bavelas et al., 2002b). If these findings scale up from dyads to multiparty it could be expected that the measured head and hand movement of a speaker is different to that of a recipient. Thus, the first hypothesis is:

Hypothesis 1: Speakers are distinct from recipients in their head and hand movement.

Moving next to consider the two recipients, if their behaviour is defined by them being non-speakers, there should not be any observable differences between them as neither are speakers. However, if their behaviour is a product of their level of collaboration with the speaker, then they may appear different from each other. The latter is in line with Clark and Shaefer (1992)'s principle of collaboration that speakers 'collaborate directly with addressees and only indirectly with side-participants'. If Kendon (1970)'s findings map to a 3-person interaction, it could be expected to see this difference in the interlocutor's non-verbal behaviour. It is possible that any difference would be manifest in head movement, as it has been suggested that addressees should return gaze (Goodwin, 1979; Wiemann and Knapp, 1975) and show feedback with their head (Allwood and Cerrato, 2003), and in their hands, as it has been suggested that speakers and addressees collaborate with gesture (Furuyama, 2000) and that gestures can contribute to common ground (Holler, 2010; Bavelas et al., 2011). Therefore, the second hypothesis is:

Hypothesis 2: Primary recipients' patterns of head and hand movement are different from those of secondary recipients, with more movement expected from the primary.

The previous chapter demonstrated that the task may be responsible for some of the observed interlocutor movement. It is therefore possible that, where interlocutors have different roles as defined by the task, they may differ in their behaviour. Thus the final hypothesis is:

Hypothesis 3: Interlocutor roles, as defined by the task at hand, will be manifest in their head and hand movement.

The approach taken is exploratory, and the question will be addressed by applying statistical analyses to the participants' motion. Data will be drawn from both of the available corpora (detailed in Section 4.3). In both corpora, interlocutors may be assigned a recipient role as either primary or secondary; in only the tuition task may interlocutors be assigned a task role, either learner or instructor. Thus, tests relating to task role are only applied to the tuition task corpus. Head orientation will be studied first and will begin with an analysis of the recipients' orientation with respect to the speaker and, for the tuition task corpus, task role. Following this, analyses which include all the interlocutors will be applied to their head and hand movements. The final tests will look exclusively at the recipients again to directly analyse the interplay of head and hand movement between them. The aim is to give an understanding of how these movements are used in multiparty dialogue.

6.2 Method

6.2.1 Filtering the data

It is not possible to fairly apply statistical analyses to the entire corpus. This is because, using the current operational definitions of recipiency, the assumption that each interlocutor will hold only one dialogue role is somewhat hopeful. It does not account for times of overlapping speech or no speech. These times will influence all members of the conversation, as the roles must be defined in a triad as part of a *multiparty* conversation, rather than in pairs of speaker-recipient. Consider a dialogue in the canonical situation of one speaker (Figure 6.1). Here Bob is speaking to, and

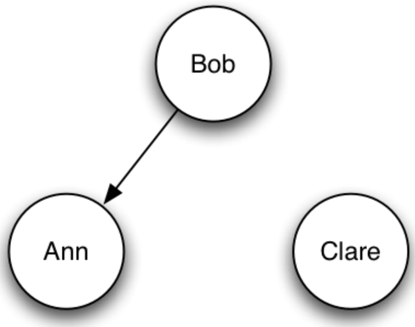


Figure 6.1 – A conversation with one speaker, Bob, orienting to their primary recipient

orienting his head towards, Ann. As such, Bob is the speaker, Ann is the primary recipient and Clare is the secondary recipient. This is mutually understood and accepted by everyone. However, what if Ann was also speaking and oriented towards Bob (Figure 6.2)?

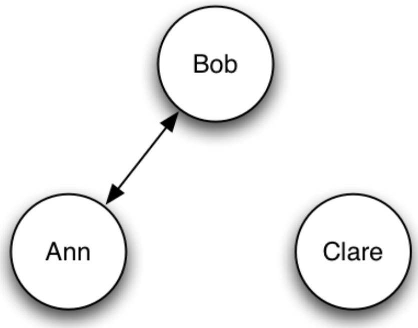


Figure 6.2 – A conversation with two speakers, Bob and Ann, orienting to each other

Here, Ann and Bob are both speaking and both oriented towards each other. As such they are both speakers and primary recipients, and Clare is a secondary recipient. Their behaviours in this situation will likely be different to the first, for example Ann is a primary recipient by definition in both situations however she is also speaking in the second. The nature of the overlapping speech may vary; Ann and Bob may be collaborating or competing, for example. Clare will also likely behave differently because she no longer only holds a recipient-recipient relationship with Ann. There are other constructions possible of overlapping speech, for example both Ann and Bob may be speaking, but Ann may be speaking to Bob and Bob may be speaking to Clare. There may also be three speakers all at once. Discussions of overlap are found by Schegloff (1995) in which he describes cases of overlapping speech similar to those described above. The current concerns of differing behaviour at these times are substantiated by Schegloff's work, as he demonstrates differing patterns of speech and also notes that gaze behaviour will be different at these times of potential competition for the floor.

The approach taken is to filter out times where there is anything but one speaker, meaning that only the canonical multiparty conversation (accounting for roughly 70% of the conversations) data is used. This is because the current operational definitions of recipiency are not able to effectively index recipiency when there are multiple speakers.

6.2.2 Frequency analysis of head orientation

To examine patterns of recipient head orientation the operational definitions of recipiency detailed in Chapter 4 are used. On each frame of data a measurement is made of the head orientation of the primary recipient and of the secondary recipient. Using this data, two frequency count variables for both recipients are produced:

- **Oriented to speaker** - Number of frames of data oriented to the speaker
- **Oriented to other** - Number of frames of data oriented to the non-speaker (either a primary or secondary recipient)

A non-parametric binomial significance test[1] is applied to ensure that these proportions do not occur by chance and a chi squared test is used to determine any difference in these measures between the recipients. A comparison of the recipients across corpora will be made.

In order to explore the effect of an interlocutor's task in their orientation behaviour, a second test takes advantage of the fact that the tuition task corpus has identifiable task roles. In this corpus there are always two instructors and one learner. The aim is to find out whether interlocutors who have the opportunity to orient to either someone who shares their task role or to orient to someone who does not, exploit this opportunity. To do this the orientation of the non-speaking instructor is analysed. This finer grained analysis of head orientation therefore determines, at each frame of data, whether the speaker is a learner or an instructor. Who the secondary instructor is oriented to at that time is then detected and is used to calculate frequency counts for the above two variables for the times that the speaker is a learner or an instructor.

6.2.3 Regression analysis

In order to understand the relative contributions of various features of the interaction, regression analyses will be performed on the motion capture data. The regressions allows us to tease apart the interactions between the variables so that we can understand their individual contributions independently of each other. Data points within the regression are frames of motion capture data (measured at 60 frames per second). At each frame there is a measurement for each person in the interaction. The variables available for analysis are as follows:

- **isNodding** - a binary variable (Yes if this person is nodding, No if they are not)

[1] The data for this test are comprised of binary variables at each time point representing the two options (oriented to speaker/oriented to other). The significance reported is a z approximation

- **isSpeaking** - a binary variable (Yes if this person is speaking, No if they are not)
- **Recipient Role** - a binary variable (either ‘Primary’ or ‘Secondary’ as determined by the speaker’s head orientation)
- **Task Role** - a binary variable (either ‘Learner’ or ‘Instructor’)
- **Hand Speed** - a scale variable of the maximum hand speed of either of the person’s hands

As these descriptions suggest, the variable Recipient Role is a judgement made based upon the actions of the speaker at that moment in time, and as such are a product of the interaction. In the statistical tests where the dependent variables are binary, the type of regression used will be a binary logistic regression; where they are scale, a linear regression will be used.

6.3 Results

This results section will report the results from both corpora where applicable, and are organised by the type of analysis. The tests of recipient head orientation will address, in part, hypotheses 2 and 3. The subsequent inferential statistics will address all three hypotheses.

6.3.1 Recipient head orientation

6.3.1.1 Head orientation by dialogue role

Tuition Task Corpus The results for recipient orientation in the tuition task corpus show a contrast in the behaviour of the recipients. A primary recipient is more likely to be looking at the speaker than look at the secondary recipient. However, when measuring the orientation of the secondary recipient, the preference to look at the speaker is weaker (see Table 6.1). These figures are significant with $\chi^2 = 19639$ (p < 0.01).

Table 6.1 – Recipient head orientation in the tuition task corpus

Recipient Role	Oriented To Speaker	Oriented To Other	Significance
Primary Recipient	70.6%	29.5%	0.00
Secondary Recipient	55.5%	44.5%	0.00

Balloon Task Corpus The results (shown in Table 6.2, ($\chi^2 = 660$ ($p < 0.01$))) differ from that of the tuition task corpus. Whilst still significantly different, the difference between recipients observed in the tuition task corpus has been reduced. Both recipients appear more similar and share a preference to orient towards the speaker rather than each other.

Table 6.2 – Recipient head orientation in the balloon task corpus

Recipient Role	Oriented To Speaker	Oriented To Other	Significance
Primary Recipient	63.6%	36.4%	0.00
Secondary Recipient	59.9%	40.1%	0.00

Comparing corpora When directly comparing the same recipient role between corpora using the data presented in Tables 6.1 and 6.2, the results show that the orientation behaviour of a primary recipient in the tuition task corpus is reliably differently to that of a primary recipient in the balloon task corpus ($\chi^2 = 3360$ ($p < 0.01$)), and that the orientation behaviour of a secondary recipient is reliably differently in the tuition task corpus to the balloon task corpus (($\chi^2 = 1188$ ($p < 0.01$)).

6.3.1.2 Head orientation by task role

Table 6.3 – Head orientation of an instructor when they are a secondary recipient

	Oriented to Speaker	Oriented to Other	Significance
Speaking Learner	83.9%	16.1%	0.00
Speaking Instructor	39.6%	60.45%	0.00

These results are significant ($\chi^2 = 63571$ ($p < 0.01$)). They demonstrate that when the speaker is a learner, the non-speaking instructor has a preference to orient towards them. However, when the speaker is the other instructor, the non-speaking instructor has a preference to orient towards the non-speaking learner. This means that they have a preference to orient towards the learner, irrespective of whether the learner is speaking or not.

6.3.2 Inferential statistics

6.3.2.1 Exploratory correlations

A one-tailed correlation analysis was performed for each corpus prior to the regression analysis to determine which variables are correlated with head and hand movement. This analysis demonstrates which variables have a significant correlation (and hence need further exploration), but does not demonstrate the unique contributions of each variable. The correlation coefficients for the tuition task corpus are shown in Table 6.4, and the balloon task corpus in Table 6.5.

Table 6.4 – Correlation data (Pearson Correlation) of the tuition task corpus. All reported correlations are significant at p <0.001

	isSpeaking	isNodding	TaskRole (Learner)	RecipientRole (Primary)	Hand Speed
isSpeaking	1	0.084	-0.030	–	0.184
isNodding		1	-0.45	0.044	0.211
TaskRole (Learner)			1	0.506	-0.042
RecipientRole (Primary)				1	0.020
Hand Speed					1

Table 6.5 – Correlation data of the balloon task corpus (Pearson Correlation). All reported correlations are significant at p <0.001

	isSpeaking	isNodding	RecipientRole (Primary)	Hand Speed
isSpeaking	1	0.178	–	0.261
isNodding		1	0.059	0.209
RecipientRole (Primary)			1	0.029
Hand Speed				1

These results show no coefficient for isSpeaking x RecipientRole as there are no data points for this combination (only dialogues with only one speaker at a time are used). When examining the relationship between isNodding and the other variables, the correlations suggest that speakers nod more than non-speakers (due to the positive correlation), learners nod less than instructors (due to the negative correlation), primary recipients nod more than secondary recipients and interlocutors move their hands more when they nod than when they don't. Whilst the directions of the correlations are the same between corpora, there are some differences in the values. Both head and hand movement have higher correlation coefficients in the balloon task corpus with speech and recipient role, implying that interlocutors' behaviours are more closely related to their dialogue role.

In addition to those mentioned, there are also the following correlations which are significant and will likely influence each other:

- isSpeaking x TaskRole (restricted to the tuition task corpus)
- TaskRole x RecipientRole (restricted to the tuition task corpus)
- TaskRole x Hand Speed (restricted to the tuition task corpus)
- isSpeaking x Hand Speed
- RecipientRole x Hand Speed

6.3.2.2 Predicting head movement

The binary logistic regression analysis takes all of these inter-correlations into account, and includes a participant ID to account for participants' individual variability. Due to the size of these tables, the participant IDs are not reported here but are reported in full in Appendix A.1. As there is a logical constraint between isSpeaking and RecipientRole (if isSpeaking is 1, RecipientRole will always be missing) two separate regressions are performed to examine these variables. The results of these regressions must be interpreted together. The tuition task corpus results are shown in Tables 6.6 & 6.7 and the balloon task corpus results are shown in Tables 6.8 & 6.9.

The results of the tuition task corpus show that, when the interactions of all variables are taken into account, an individual's role within the dialogue influences how likely they are to nod. If they are speaking they are more likely to nod than if they are not. There is also a contrast between recipient roles, showing that a primary recipient is more likely to be nodding than a secondary recipient is. The task role of the person also has a small but significant influence upon their nodding behaviour, with learners less likely to nod than instructors.

Table 6.6 – Binary logistic regression predicting nodding in the tuition task corpus with isSpeaking (model significant at $p < 0.001$)

Variable	B	S.E.	EXP(B)	Sig
isSpeaking (Yes)	0.253	0.004	1.288	0.000
Task Role (Learner)	-0.164	0.004	0.848	0.000
Hand Speed	0.126	0.001	1.134	0.000

Table 6.7 – Binary logistic regression predicting nodding in the tuition task corpus with Recipient Role (model significant at $p < 0.001$)

Variable	B	S.E.	EXP(B)	Sig
RecipientRole (Primary)	0.233	0.007	1.263	0.000
Hand Speed	0.155	0.001	1.167	0.000
Task Role (Learner)	-0.124	0.008	0.884	0.000

The results for head movement in the balloon task corpus confirm those of the tuition task corpus. They demonstrate that there is a relationship between nodding and speech, and that primary recipients are more likely than secondary recipients to be nodding. The positive relationship between hand and head movement is also present.

Table 6.8 – Binary logistic regression predicting nodding in the balloon task corpus with isSpeaking (model significant at $p < 0.001$)

Variable	B	S.E.	EXP(B)	Sig
isSpeaking (Yes)	0.725	0.006	2.064	0.000
Hand Speed	0.106	0.001	1.111	0.000

Table 6.9 – Binary logistic regression predicting nodding in the balloon task corpus with RecipientRole (model significant at $p < 0.001$)

Variable	B	S.E.	EXP(B)	Sig
Hand Speed	0.164	0.002	1.178	0.000
RecipientRole (Primary)	0.061	0.013	1.063	0.000

6.3.2.3 Predicting hand movement

As with the analysis of head movement, two separate regressions are performed to independently analyse isSpeaking and Recipient Role.

Table 6.10 – Linear regression predicting hand speed in the tuition task corpus with isSpeaking

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.485	0.006	0.198	0.000
isSpeaking (Yes)	1.232	0.006	0.167	0.000
Task Role (Learner)	-0.215	0.005	-0.031	0.000

Table 6.11 – Linear regression predicting hand speed in the tuition task corpus with RecipientRole

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.266	0.007	0.201	0.000
Task Role (Learner)	-0.256	0.007	-0.048	0.000
RecipientRole (Primary)	0.169	0.007	0.033	0.000

The results of the regression predicting hand movement in the tuition task corpus (shown in Tables 6.10 & 6.11) demonstrate that, as with head movement, an interlocutor's hand behaviour varies depending upon their state in the dialogue. Speakers are more likely than non-speakers to be moving their hands. When looking at the difference between primary and secondary recipients the difference is much smaller, although significant; primary recipients are slightly more likely to be moving their hands than secondary recipients. There is a greater difference observed between learners and instructors, with instructors more likely to be moving their hands than learners. As observed when regressing onto head movement, there is a positive relationship between nodding and hand movement.

The results for hand movement in the balloon task corpus (Tables 6.12 and 6.13) are similar to those of the tuition task corpus. Speakers are more likely to be moving their hands, and primary recipients are more likely than secondary recipients to be moving their hands. The positive relationship between head and hand movement is confirmed.

Table 6.12 – Linear regression predicting hand speed in the balloon task corpus with isSpeaking

Variable	B	S.E.	Standardized B	Sig
isSpeaking (Yes)	1.412	0.006	0.232	0.000
isNodding (Yes)	1.229	0.007	0.169	0.000

Table 6.13 – Linear regression predicting hand speed in the balloon task corpus with RecipientRole

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.044	0.009	0.175	0.000
RecipientRole (Primary)	0.069	0.005	0.019	0.000

6.3.2.4 Recipient role

Predicting recipient role The direct analysis of recipient role for the tuition task corpus is shown in Table 6.14.

Table 6.14 – Binary logistic regression predicting Recipient Role (Primary) in the tuition task corpus(model significant at $p < 0.001$)

Variable	B	S.E.	EXP(B)	Sig
Task Role (Learner)	2.684	0.007	14.640	0.000
isNodding (Yes)	0.217	0.007	1.242	0.000
Hand Speed	0.013	0.001	1.013	0.000
isSpeaking (Yes)	*Removed as constant*			

The figures demonstrate that task role has the highest impact on recipient role, a likely result of instructors teaching material to learners. Nodding is the next significant predictor of recipient role, confirming the finding that primary recipients are more likely to nod. Movement of the hands also has a significant impact (again, albeit a small one) on recipient role with primary recipients more likely to move their hands than secondary recipients. As expected, the isSpeaker variable was removed as this is constant 0 when Recipient Role has a value.

The results for regressing onto recipient role in the balloon task corpus are shown in Table 6.15. As with the tuition task corpus, there is a small but significant effect of hand behaviour with a greater hand speed predicting primary recipiency. If a participant is nodding there is a greater likelihood of them being a primary recipient.

Table 6.15 – Binary logistic regression predicting Recipient Role (Primary) in the Balloon Task Corpus

Variable	B	S.E.	EXP(B)	Sig
isNodding (Yes)	0.055	0.013	1.057	0.000
Hand Speed	0.017	0.002	1.017	0.000
isSpeaking (Yes)	*Removed as constant*			

6.4 Discussion

6.4.1 Discriminating dialogue roles

The results above, confirmed across two independent corpora, go some way towards exposing the relationship between body movements and the moment-by-moment organisational structure of the dialogue. Within the restrictions placed upon the analysis, at any point in time an interlocutor may either be a speaker, a primary recipient or a secondary recipient. Confirming hypothesis 1, speakers are distinct from recipients in their behaviours and are more likely to nod and to gesture. This is in line with findings from dyads (Hadar et al., 1983). An initial attempt to interpret the speaker's nodding behaviour may be to link this with the physical properties of speaking, however this is unlikely as movements such as these are removed from the signal during the filtering stage. An interpretation is that these movements hold interactional significance, and fit into a feedback cycle of head movement similar to that usually found from listeners. This suggests that feedback head movement is not unidirectional, from the listener to the speaker, but may form more of a collaborative process between speakers and recipients (see Boholm and Allwood (2010) for similar speaker feedback data from Swedish speakers). Speakers also have a propensity to gesture more than non-speakers, most likely due to depictions of content of their speech. Note that this does not mean that non-speakers do not gesture, simply that speakers gesture more.

A key contribution of this chapter is confirming hypothesis 2; when examining patterns of head and hand movement, the primary and secondary recipients are different. A primary recipient is more likely to be gesturing than a secondary recipient. Furuyama (1993) noted that listeners within a dialogue are able to gesture along with a speaker's speech, without speaking themselves. The findings here confirm these observations and add an additional level of detail, present only in multiparty interaction, by distinguishing between the recipients. The results of head movement show that, all other variables being equal, primary recipients are more likely to be nodding than secondary recipients are. When the findings of head and hand movement are considered in tandem, they contribute to the understanding of recipient behaviour in multiparty dialogue and their relationships with the speaker. They suggest, as alluded to in the literature, that the primary recipient holds a different relationship, and a higher level of coordination, with the speaker than the secondary does with the speaker. The findings confirm, and add to, those of Kendon (1970). When considering Clark and Shaefer (1992)'s suggestion that primary addressees have more of a responsibility to show continued attention to the speaker than secondary recipients, these results confirm this, and add that this is manifest non-verbally in both head and hand movement.

The findings for head orientation prove more of a challenge to interpret in the context of interlocutors' dialogue roles. That patterns do exist, even when using the relatively naive judgements of recipiency and orientation, confirms the prediction made in Chapter 4 that head orientation is used as an interactional cue within a multiparty dialogue, although no claims can be made about eye-gaze given the dataset used. The challenge in interpreting head orientation with an interlocutor's dialogue role exists due to the fact that a primary recipient's patterns of orientation are different in the balloon task corpus to those found for a primary in the tuition task corpus. Likewise, the secondary recipient's orientation patterns are different between corpora. These differences may be caused by the different task structures, and will be addressed next.

6.4.2 The effect of task

It is not possible to fully answer the original exploratory question of 'Are there global patterns of head and hand movements that relate to the unique structure of a multiparty dialogue in terms of its constituent interlocutors' dialogue roles?' without considering the task at hand and what these implications have for the members of the conversation. Indeed, the previous chapter showed, at a crude level, that the task structure is influential for the interlocutors' behaviour. The tuition task corpus is distinct from the balloon task corpus in that there are two different task roles; interlocutors are either instructors or learners. In the balloon task corpus this is not the case. In this corpus, all interlocutors share the same task role in deciding who must be thrown from the balloon. When looking at the relationship between task role and recipient role (see Table 6.14) it was found that primary recipients are more likely to be learners rather than instructors. This is to be expected, given that the instructors are tasked with teaching the material to the learners. The regression analyses go some way to factoring out the effect of task role when measuring other variables, however even when this statistical factoring is applied the results are only interpretable within the context of the interaction. That is, within the tuition task corpus, although the primary and secondary recipients are defined as per current operational definitions, their behaviour is still framed by the tuition scenario. By examining the behaviour of the recipients in the balloon task corpus, the behaviour of recipients who are framed by a different scenario was tested.

A link can be drawn here to the literature on the dynamics of dialogue, in particular that of parties (see Section 3.5). Members of a party share common knowledge and are jointly responsible for answering questions. The task role structure imposed in the tuition task corpus results in the formation of an implied party: the instructors. They are privy to joint knowledge that the learner is not, and they are jointly responsible for teaching the material to the learner. This is not to say that other parties will not form within the dialogue, however the instructor party is well defined and clearly measurable[2]. This implied party construct does not exist, at least at a level of resolution finer than that of the interaction as a whole, in the balloon task corpus; all interlocutors are given the same instructions and the same prior knowledge.

Looking first at the regression analyses of head movement, it is evident that a learner is less likely to nod than an instructor. One possible explanation for this is that, during the times that one instructor is teaching material to the learner, the other instructor nods in confirmation of the material, when the learner is retelling the material to the instructors they both give positive

[2] In ongoing work by Christine Howes using the tuition task corpus, additional evidence is emerging that the instructors form a party (Howes, in prep). This initial evidence suggests that more split utterances, utterances that are started by one interlocutor but completed by another, occur above chance between instructor pairs than do between instructor-learner pairs. Furthermore, an instructor is more likely to ratify the completion when the antecedent to this split utterance comes from an instructor and is completed by a learner (26% ratified), when compared with the third party in the other two possible constructions (instructor-instructor: 17% by the learner, learner-instructor: 18% by the other instructor). These findings are important because, as Lerner (1993) suggests, split-utterances are one way that interlocutors can form a party.

feedback in the form of nodding. When contrasting the effect of primary recipiency in the context of the tuition task corpus' tuition scenario with that of the balloon task corpus it appears that there is a weaker effect of primary recipiency in the latter. This could be interpreted as a covert effect of the task role.

When looking at the regression analyses of hand movement the results follow a more expected pattern. Instructors are more likely to gesture than learners. This could be accounted for with two reasons: firstly, they must depict the material and as such they will deploy gestures more often that the learners who do not have this requirement, secondly the instructors are members of a party and may use gestures to enable coordination between themselves within the party. That instructors gesture more than learners, that learners are most often primary recipients and that primary recipients gesture more than secondary recipients may appear to be contradictory, but this is not the case. As the regressions account for the interactions between variables, each of the findings are statistically independent of each other. That said, the size of the effect of primary recipiency in the balloon task corpus is smaller than that of the tuition task corpus, suggesting that some of the coordination seen between the primary and the speaker could again be a covert effect of task. These findings confirm hypothesis 3 and add granularity to the findings concerning the task in the previous chapter.

The analysis in which the effect of task is most clearly manifest is the analysis of head orientation. The results of the orientation analysis on the tuition task corpus show that a primary recipient will orient to the speaker 71% of the time, whereas for the secondary recipients this preference is rather weak. However, when the option for an interlocutor to orient to an interlocutor within their party (an instructor orienting to an instructor) or to orient to an interlocutor outside of their party (an instructor orienting to a learner) arises there is a preference to orient towards the non-party member. That is, a non-speaking instructor has a preference to look at the learner rather than their co-instructor, even if their co-instructor is speaking. When this choice of party membership, based on task role, is removed, as is the case with the balloon task corpus, the observed patterns of head orientation shift, with both recipients being near equal in their mild orientation preference towards the speaker. The results of the orientation analysis in the balloon task corpus allow the results of the tuition task corpus to be understood more clearly. The unequal behaviour of the recipients can be explained, not only by factors relating to their dialogue role, but also because they are members of different parties within the dialogue which have differing task based responsibilities. In the tuition task corpus it is possible that the instructors' head orientation is influenced by the need to monitor the learner for understanding, rather than demonstrate continued attention to the speaker. Because of this, when measuring head orientation based on recipiency alone, the recipients are likely to behave similarly and orient mainly towards the speaker. However, party membership and the imposed responsibilities take precedence over this and are likely to change the observed behaviour. These findings evidence that the higher-level organisation of the dialogue into parties also indexes interlocutor non-verbal behaviour.

These findings suggest that, whilst an interlocutor's role within the dialogue influences their behaviour, the patterns of non-verbal behaviour present in the dialogue could be organised across parties rather than individuals. This also raises questions about the analytical methods that are used when studying multiparty interaction. Rather than focusing on individuals within the context of an interaction, it may be more suitable to measure the behaviour of parties within the context of an interaction.

6.4.3 The close coupling of head and hand movements

In addition to the intended goal of exploring head and hand movement in relation to dialogue state, the analyses have also demonstrated that there is significant interaction between the movements of interlocutors' heads and hands; if an interlocutor is gesturing it's likely that they will be nodding and vice versa. This is aligned with the holistic views of Hayashi (2005), and the composite signal theories (Engle, 1998). This finding raises the question of what it is that the union of head and the hands can do within communicatively within the dialogue. This will be addressed in the next chapter.

6.5 Conclusion

This chapter tested non-verbal behaviour in relation to interlocutors' dialogue roles (either speaker, primary recipient or secondary recipient) and task roles (either instructor or learner). It was found that, when measuring gesture and nodding, interlocutors' dialogue roles can index their behaviour. This is most interesting when considering recipient roles as there is a clear distinction between them: primary recipients nod and gesture more than the secondary recipients suggesting a higher level of coordination between the speaker and the primary recipient than is present between the speaker and the secondary recipient. Task role, and hence party membership, was also shown to be a valid index of interlocutor behaviour. This was most manifest in the recipients' orientation behaviour where there was evidence for interlocutors monitoring each other for purposes of the task (i.e. instructors monitor learners), as well as monitoring and feeding back to the speaker.

Chapter 7

Exploiting the shared space

7.1 Introduction

The thesis so far has demonstrated that interlocutors make use of their head and hands as part of their conversation. Through the previous two empirical chapters, patterns of movement which can be attributed to a) the interaction and the unique structure of a multiparty dialogue and b) the task structure with its associated responsibilities have been identified. The methodology employed for these chapters has been strictly quantitative, using parameter based analyses. This chapter makes use of a methodology which differs somewhat from those seen so far. A human based, qualitative analysis is used in order to get a more interpretive view of the data. This allows us to go some way towards understanding what the interlocutors are doing with the behaviours that have been identified so far. Without this interpretive step, the thesis would be incomplete.

A remaining question that must be addressed is concerned with the shared space. The findings so far allow for a clearer understanding of the interactional processes involved with non-verbal behaviour, but no investigation has been made into their deployment in the shared space. The literature review demonstrated the importance of a shared space for interaction (Kendon, 1970), and how through manipulations of its construction and constituent members it has an effect on the behaviour of interlocutors (Özyürek, 2002). Excerpts of interactions within a shared space were shown, such as that from Lerner (1992) in which differing orientations of a speaker's head and body were documented. The question that must be answered is: 'Are there communicative behaviours that intrinsically rely on their deployment in a mutually accessible shared interaction space and, if so, what do they look like?'. By using a qualitative approach to understand the interlocutors' movements it will be possible to observe any uses of the shared interaction space. As expressed in Chapter 3, of particular interest is how the shared space is used as a resource to manage the interaction, what role the interlocutors' bodies play in this and if multiparty dialogue is different to dyadic dialogue in its need of the shared space.

The question will be addressed in a two stage process consisting of two studies. The first study will be an observational study with the aim of producing qualitative results which demonstrate spatial features of the interaction. The second study will examine the communicative effect of a key feature from study one using a hybrid of human annotation and machine analysis, producing quantitative results.

7.2 Study One

7.2.1 Aim

This study will attempt to find examples of interlocutor behaviours which directly make use of the shared space. Rather than looking for patterns using the naive parameter based approach so far, this section will try to interpret the movements of interlocutors. The aim is to determine a) if these types of behaviours exist, and if so b) what they may look like.

7.2.2 Method

The data used for this study are multi-angle video data based on the main, tuition task corpus (details of which can be found in Section 4.3). These data consist of one above, and two either side views of the interactions. ELAN[1] is used to play the videos synchronously side by side. The methodology employed is qualitative not quantitative and there is no initial hypothesis. Instead, the data are used to allow patterns and features to emerge. The second study will firm up the qualitative findings and provide a quantitative analysis of them.

Rather than a computer analysing the videos, they are watched by the author, with multiple passes, in combination with the speech annotations described in Chapter 4. Sections of the interactions during which interlocutors appeared to be using their head or hands collaboratively and making use of the shared space were noted and examined further. On further observation the

[1] ELAN annotation software: http://www.lat-mpi.eu/tools/elan/

key features of the sections were identified and will be reported using observations and excerpts from the interactions.

7.2.3 Observational results

7.2.3.1 Known phenomena

The data showed similarities to those commonly reported in the literature. At the most basic level participants oriented their bodies towards each other as they sat on the stools, as suggested by Kendon (1973) and Goffman (1966). Some other examples include instances of gestural mimicry such as that shown in Fig 7.1. Here the learner is wearing the blue cap, the two instructors are wearing the white and black caps. As they introduce one of the elements of the application, a 'still camera', the learner expresses that he does not know what this means. The instructor wearing the black cap clarifies and also produces a gesture that shows the rectangular shape of a camera with his right index finger pressing the shutter button (Fig. 7.1.a). Following this the learner then mimics him, creating a similar rectangular gesture (Fig. 7.1.b). This is in line with discussions of mimicry, such as those reported by McNeill (2005). Examples of gesture that could be classed as *interactive* under the terminology proposed by Bavelas et al. (1992) and observed in Holler (2010) are present in the data also. For example, in Fig 7.2 the participant in the white cap is trying to recall some of the departments in the government and has just suggested the department of trade. After she has said this the participant in the blue cap repeats 'trade' and produces a gesture with an extended arm towards the white capped participant with the palm up. This could be glossed as meaning 'and the department, trade, which you have just

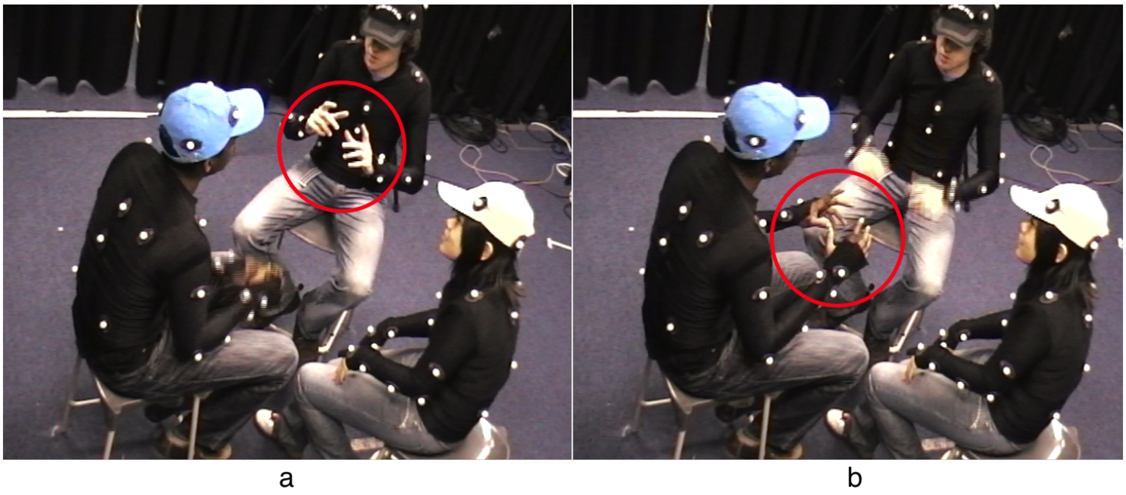


Figure 7.1 – An example of gestural mimicry, involving the description of a *still camera*

suggested'. There are examples of non-speaker gestures, akin to listener gestures (Furuyama, 1993) (which are in line with the quantitative findings in the previous chapter that recipients gesture). A common scene involved a person recalling some departments or classes and listing them gesturally whilst at the same time the other members of the interaction would also gesturally list them (without speech).

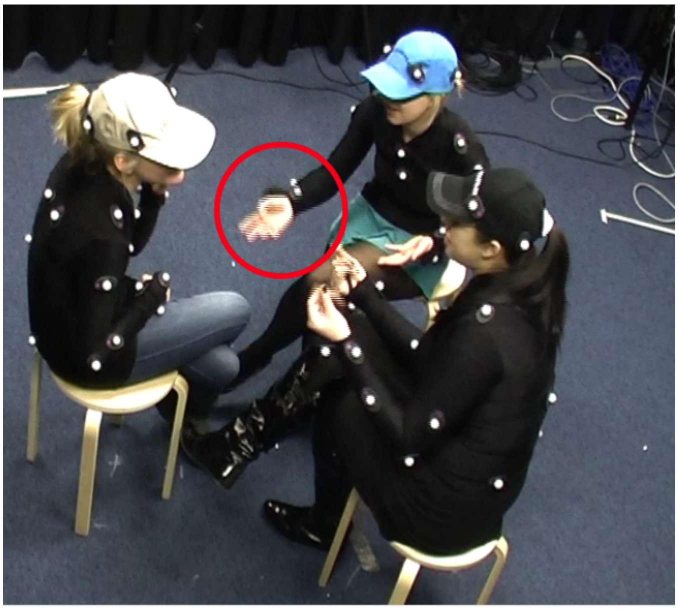


Figure 7.2 – An interactive gesture deployed by the interlocutor with a blue cap

7.2.3.2 *Shared frames of reference*

There were two prominent gestural techniques that were used whilst describing either the government material or the Java code material. The first was to list sections off on the hand in a counting style. These are similar to the *list sequences* described by Bekker et al. (1995). The second technique saw interlocutors gesturally build a class[2] hierarchy in front of their bodies (during the Java task) to represent the Java class hierarchy. It is important to note that the stimulus material given to them was designed to avoid intrinsically spatial domains and did not contain a drawn version of the hierarchy, the instructors were only able to build up the hierarchy by understanding the printed code (which had one Java class per sheet). Thus whilst the stimulus material defines the vertical layout of the hierarchy (i.e. what is above or below each other), it did not define the horizontal layout; this was a choice made during the tuition.

[2] a 'class' is programming terminology and, for current purposes, can be treated as an element in the hierarchy

These hierarchies often formed persistent spaces. That is, the gesturers did not need to keep their hands in place to maintain the hierarchy; they could deploy a number of classes on one tier, then move their hands away but refer back to the spaces where the classes had previously been deployed to describe the elements underneath them. These appear to be similar to the gestural diagrams of kinship noted by Enfield (2005).

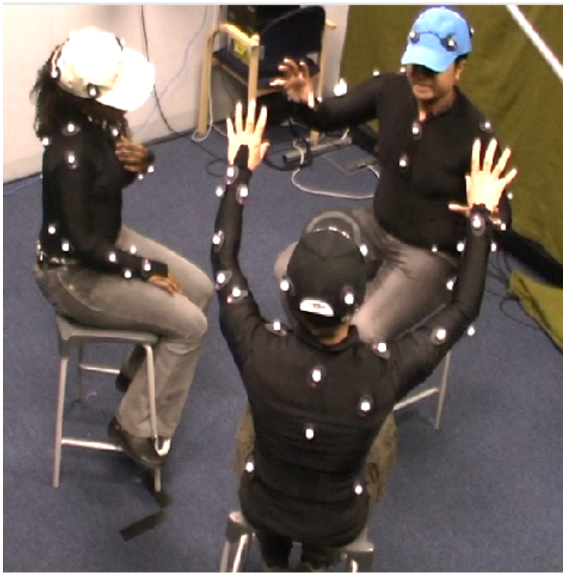


Figure 7.3 – An example showing the sharing of a gestural hierarchy

The following example demonstrates how these types of gestural hierarchies can be used by the learner. In Fig 7.3 the two instructors are wearing white and blacks caps; the learner is wearing a blue cap. The instructor in the black cap introduces the 'Person' class. He says 'person here' and stretches his right arm forwards and to the right of his body with his palm vertical and towards the learner. The learner extends his left hand forwards, deploying the same gesture as the instructor, but slightly lower than his hand. The instructor continues with 'year group here' placing his left hand out stretched and to the left of his body. The learner then places his right hand directly in front of the instructor's left hand.

Firstly we see here that both participants have extended out of their individual spaces towards a mutually accessible area of the shared space. Secondly, the learner is also recreating the hierarchy, not using his own, person-centric frame of reference which would result in a left-right flip of the hierarchy, but using a shared frame of reference with the instructor[3]. The result of this is

[3] The accompanying speech was used to determine which elements of the hierarchy were currently being gestured

that both the instructor and the learner have created, or at least acknowledged the existence of, interactionally shared objects; they, and as a ratified participant the co-instructor, all mutually know that the Person class and YearGroup class are interactionally salient objects and can be referenced spatially.

In the example, the participants are not pointing to exactly the same spaces. However, because they both mutually know where each other are positioned around the shared space, they are able to create these shared objects gesturally by adopting the shared frame of reference. The sharing of loaded space has been seen in Olson and Olson (2000), however what has remained unobserved is the shared frame of reference used to gesturally refer to the spaces.

7.2.3.3 Simultaneous engagement

The next feature of the interaction to be described is the use of a speaker's head and hands to simultaneously engage both recipients at the same time. This is examined because the actions of the speaker directly coordinate with the recipients and, at least on first examination, appear to exploit the shared space. The sequence is as follows:

White Cap (instructor): add play playlist (Fig. 7.4)

Gaze: Towards Black Cap (learner), angled down slightly

Gesture: Left hand is between herself and Black Cap (learner) with fingers extended. Right hand is counting along the left hand's fingers with a pointing gesture

(0.1)

White Cap (instructor) : add to track

(0.1)

White Cap (instructor): is$_1$ it add track$_2$ (Fig. 7.5)

Gaze$_{1\&2}$*: Turns from Black Cap (learner) to face Blue Cap (instructor)*

Gesture$_2$*: Left hand stays stationary between herself and Black Cap (learner), right hand moves to be placed between herself and Blue Cap (instructor), pointing towards Blue Cap (instructor)*

(0.3)

White Cap (instructor): I think$_1$ add track$_2$ (Fig. 7.6)

Gaze$_1$*: Turns from Blue Cap (instructor) to face Black Cap (learner)*

Gesture$_1$*: Right handed point turns in an arced motion around her body to be placed between herself and Black Cap (learner), now pointing at Black Cap (learner). Left hand remains in* *place.*

Gesture$_2$*: Left handed gesture ends and rests on the leg.*

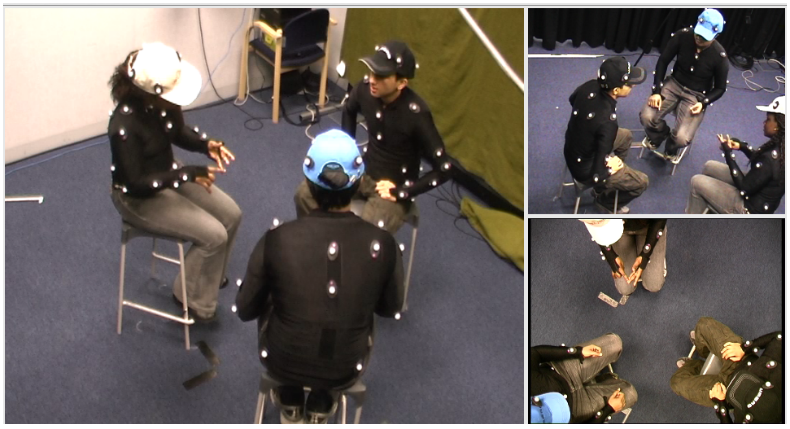


Figure 7.4 – White Cap listing methods to the learner

This example demonstrates a speaker's ability to manage both recipients at the same time and hence simultaneously engage them. In the excerpt, the instructor wearing the white cap (henceforth 'White Cap') is explaining some of the methods (sections the computer code) which are part of the Playlist class to the learner who is wearing the black cap (henceforth 'Black Cap'). As she does this she is gesturing with both hands; the left hand is held out between her and Black Cap, the learner, with her fingers extended, the right hand is counting along the fingers (by pointing at them) as she mentions each method (see Fig 7.4). She is also looking towards Black Cap. As she speaks Black Cap has his hands on his legs and is back channelling verbally and with head nods towards her. The second instructor who is wearing the blue cap (henceforth 'Blue Cap') is looking towards Black Cap during this time and is neither gesturing or speaking.

She continues and erroneously describes the 'add to track' method (it should simply be 'add track'). After carrying on with her list, she comes back to this and again says 'add to track', still gesturing and gazing towards the Black Cap. She poses the question 'is it add track' to Blue Cap, her co-instructor. As she says this her gesture and gaze configuration changes; her gaze turns towards Blue Cap, and shortly after her right hand moves from being a counter on the left hand and becomes a point in the direction of Blue Cap. Her left hand however does not move and stays in its location between herself and Black Cap. She has now turned her attention from Black

Cap to seek help from her co-instructor, however her left hand remains clearly separate to this second, help seeking, task and appears to be meant only for herself and Black Cap. It is as if she is using it as an interactional hold[4] to signify the that the conversation between herself and Black Cap is not over, just temporarily on hold (this could also be being signified to Blue Cap. It may be a sign that he does not have the right to take the floor of the whole 3 way interaction, simply the side dialogue between himself and White Cap).

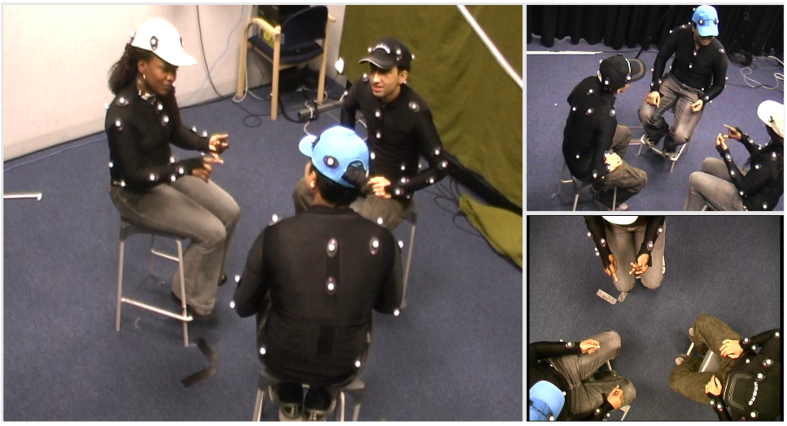


Figure 7.5 – White Cap seeking help from her co-instructor, whilst holding the interaction with the learner.

With her combined right hand and gaze movement she is now eliciting the involvement of the other instructor in the development of the ongoing turn. As she does this combined movement towards Blue Cap, he then turns his gaze to form mutual gaze with hers. Black Cap also turns to face Blue Cap (see Fig 7.5).

Before Blue Cap responds verbally, White Cap answers her own question with ‘I think add track’. It is only after this that Blue Cap managed ‘yeah add track yeah’. Now that she has the answer to her question, White Cap needs to deliver this information to the learner, Black Cap. At this moment her left hand is still left in a hold between herself and Black Cap; her right hand is pointing towards Blue Cap which is the direction of her gaze also. She now changes her gesture and gaze configuration again. She synchronously turns her gaze and point (in a slightly arced trajectory) around herself to now point towards Black Cap. As her right handed point comes into the space between herself and Black Cap her left hand drops from its holding position, Black Cap turns his gaze back towards her and instructional dialogue continues with the new information

[4] Note, the use of the word *hold* here is used to signify that a section of the *interaction* is on hold. This is in contrast to the use of the word by McNeill (1992); Kita et al. (1998) who use it in reference to sections of the *gesture* as being on hold

acquired from the consultation with Blue Cap (see Fig.7.6).

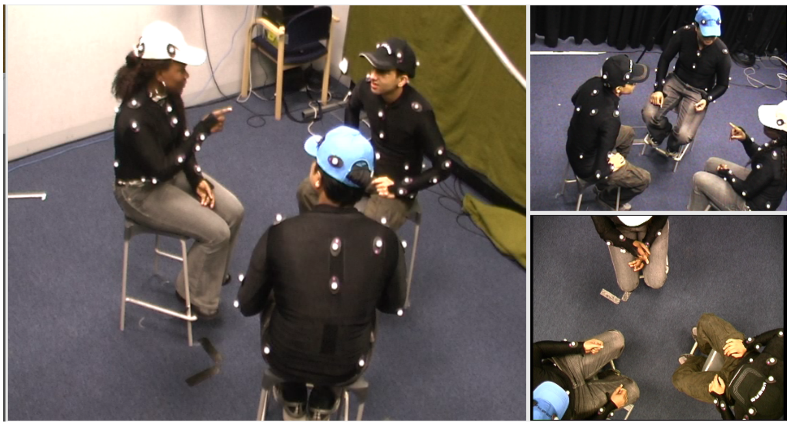


Figure 7.6 – White Cap continuing the tuition with the learner.

Blue Cap does not speak again until significantly later in the tuition. It is also worth emphasizing the fact that White Cap turns her right handed point towards Black Cap. This not only holds interactional implications for Black Cap, but also the removal of this gesture from between the two instructors is marked. It implies that White Cap is no longer enlisting Blue Cap in the construction of the turn.

Of importance here, is that White Cap did not engage in three, serial conversations. She did not have a conversation with Black Cap, cease this, move to Blue Cap, cease this and then move back to Black cap. She continually engages with Black cap and temporarily she also engages with Blue cap. Her use of head orientation and gesture help to facilitate this simultaneous engagement.

Common Forms The case study identified and discussed above is significant as it demonstrates an essential use of shared space. The chapter will soon turn to examine how common this type of event may be. To address this, the common characteristics of simultaneous engagement events are identified. The three main forms are demonstrated below (see Appendix B.1 for the transcripts) and form the basis of the coding scheme for quantitative analysis in section 7.3.3.3.

- **Head Moves**

  In this scene (Fig. 7.7) the instructor in the black cap is gesturing and orienting her head towards the learner in the white cap. She utters "ok now part of the states assembly are two err things we have the executive". She then says 'is that right' and turns her head towards her blue capped fellow instructor but leaves the gesture in place between herself and the learner.

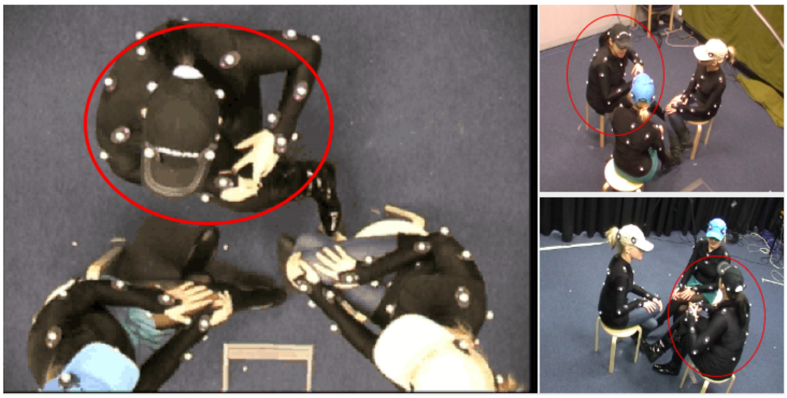


Figure 7.7 – Orienting to the addressee with a gesture and orienting to the third party with the head

- **Hand Moves** In this example (Fig. 7.8) the blue capped instructor is explaining some details which had previously been introduced by his white capped colleague. As he explains them to the black capped learner he says "the l-lower hierarchy classes like the masters and undergraduate". As he reaches 'the masters' he shifts the orientation of this left hand towards the instructor.

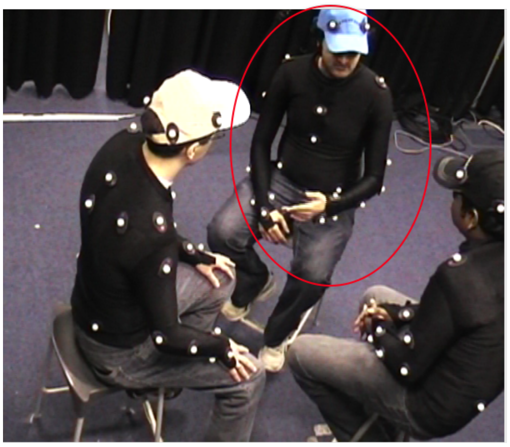


Figure 7.8 – Orienting to the third party with a gesture while continuing to orient to the addressee with the head

- **Both Moves**

  In this scene (Fig. 7.9) the instructors are wearing the blue and white caps. As the white capped instructor adds "and the king chooses the ministers" to blue's descriptions, she first points and looks at the learner in the black cap then shifts both of these in a coordinated movement to the blue capped instructor.

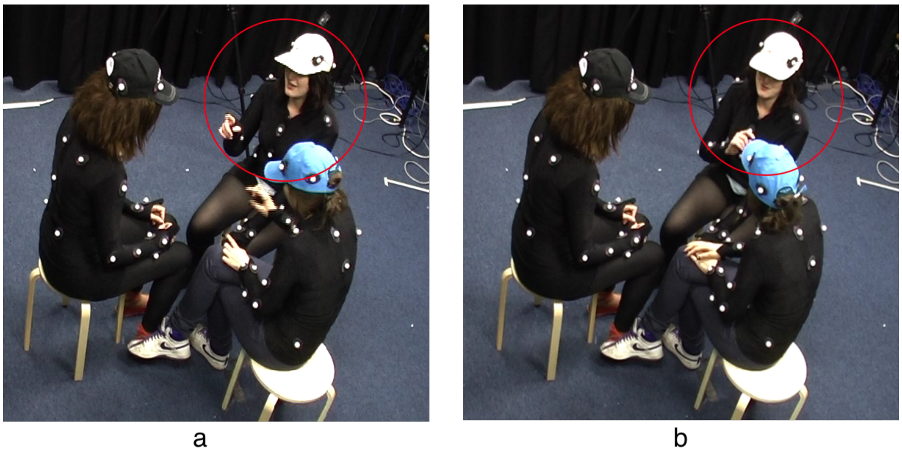


Figure 7.9 – A combined movement of head and gesture by the white capped instructor

7.3 Study Two

7.3.1 Aim

The aim of the second study is to explore more deeply the key finding from the first study. Simultaneous engagement events have been chosen because they are unique to multiparty interaction and require a shared space for deployment. This study will question the contribution of these events to the unfolding interaction. The hypothesis is that these events are interactionally significant coordination events, and as such should elicit different behaviour from their recipients than at other times during the interaction. To get an understanding of the relative contributions of head and hand movement to the dialogue during simultaneous engagement events, a comparison will be made between the frequency of responses by the class of event. These questions will be addressed by:

1. measuring the regularity of simultaneous engagement events across the corpus using hand coded data
2. measuring their effect on recipient behaviour, with a comparison between class of event using a hybrid analysis of hand codings and machine analysis.

7.3.2 Method

7.3.2.1 Annotation of simultaneous engagement events

In order to understand the frequency of simultaneous engagement events throughout the corpus, an annotation based approach was taken using ELAN. An annotation was made for every visible change in *speaker* head or gesture orientation relative to the recipients whilst a gesture was in progress (see Appendix B.2 for the coding scheme decision tree). Using the qualitative findings described above, each event was sub-coded into the one of the following three categories:

- **Head Moves:** Here the head orientation changes but the gesture remains stationary. For example, the speaker may be gesturing towards the primary recipient and glance (by turning their head) towards the other, secondary, recipient.
- **Hand Moves:** Here the gesture moves, but the head orientation remains stationary. For example, the speaker could be gesturing with a palm up gesture towards the primary recipient and whilst continuing to look at them, turn their gesture so that it is oriented to the secondary recipient.
- **Both Move:** Here both the gesture and the head shift orientation. For example, the speaker could be pointing towards the primary recipient then turn their point along with their head orientation towards the secondary participant.

All events were coded by the author. To ensure reliability of annotations, a second pass of 25 randomly selected events was made by a second coder.

In order to explore the relationship between the production of simultaneous engagement events and task role, for each interaction the number of events performed by either an instructor or a learner create two frequency counts. To account for the fact that there are twice as many instructors than learners, the frequency instructor counts are normalised by taking the average between them per session. The values are compared using a non-parametric, related samples test.

7.3.2.2 Machine analysis of recipient behaviour

Measurements To analyse the frequency of responses to simultaneous engagement events, this method makes use of the head nodding and head re-orientation (i.e. crossing the centre line) measures defined in Chapter 4, combined with the ELAN based event annotation data. A window after each event is created and, for each recipient, a score is made for whether a head re-orientation and whether a head nod occurs in that window. If the situation arises where two simultaneous engagement events occur within each other's windows, these are excluded as a direct link between the event and the responses cannot be made. In order to provide a measure of response latency the first change of head orientation or nod that occurs after the target event and before another target event occurs is recorded. Latency data is used to motivate the size of the window used for frequency response calculations. The window is set as the mean response time plus one standard deviation. This is done in order to justify the window size through the data.

Baseline activity To interpret the measures of behaviour after the simultaneous engagement events, it is important to know what the baseline likelihood of a recipient nodding or changing orientation during sequences that do not contain a target event is. To provide this, a baseline comparison sample was created by randomly selecting points in the interaction where someone was speaking but not producing a simultaneous engagement event. Recipient behaviour after these baseline points was then analysed in the same way as it is for the simultaneous engagement events.

Comparisons To identify any differences in responses between recipients, response rates will be compared according to recipients' roles (primary v.s. secondary). The tests will then turn to determine if there exist any differences in recipient behaviour between target (simultaneous engagement) events and control events. This will be done by measuring the frequency of responses and the latency of responses. Comparisons will be made to compare the behaviour between the classes of events try to to determine their relative impact on the interaction.

7.3.3 Results

7.3.3.1 Frequency of simultaneous engagement events

The total duration of dialogues in the corpus was 2 hours and 54 minutes, and each task took on average 5 minutes and 16 seconds. There was a total of 308 simultaneous engagement events, occurring on average every 34 seconds. The breakdown by class of event is shown in Table 7.1. The inter-rater reliability was good with Kappa = 0.78,($p <$ 0.001).

Table 7.1 – Number of simultaneous engagement events broken down by type of event

Event Class	Count
Head Moves	184
Both Move	91
Hand Moves	33

7.3.3.2 Relationship with task role

The mean number of simultaneous engagement events and standard deviation produced per interaction for instructors and learners are shown in Table 7.2. Whilst instructors have a higher mean, using a Wilcoxon signed rank test shows that this difference is not significant (*Z = -1.050, p = 0.294*). This means that the differences observed between learners and instructors in the previous chapter (learners nod and gesture less), is not present above chance for the production of simultaneous engagement events.

Table 7.2 – Frequency of simultaneous engagement events by task role (production of event)

Task Role	Mean	Standard Deviation
Learner	2.64	2.702
Instructor	3.35	2.816

7.3.3.3 Frequency of recipient responses

Using the latency data, the response window was calculated to be 5.39 seconds.

When comparing the response rates of both vertical head movement and head re-orientations there is no significant difference (tested using Chi Squared) between primary and secondary recipients (see Table 7.3). Because of this, the following results will merge primary and secondary recipients together and compare their response rates to the the baseline response rate.

Table 7.3 – Response rates to target events for each recipient

Response Type	Primary Recipient Rate	Secondary Recipient Rate	Sig (χ^2)
Head Nods	76%	82.2%	Not Significant (p=0.13)
Head Re-Orientations	54.8%	54.3%	Not Significant (p=0.92)

Table 7.4 – Response rates by type of event, measured by recipient re-orientations

Event Class	Response Rate	Baseline Response Rate	Sig
Head Moves	50.87%	45.04%	Not Significant (p=0.14)
Both Move	58.82%	45.04%	$\chi^2 = 6.59,$ $p < 0.05$
Hand Moves	65.0%	45.04%	$\chi^2 = 5.99,$ $p < 0.05$

For changes in head orientation the recipients' baseline response rate is 45.04% and their response rate to target events is 54.57%; a small but reliable difference ($\chi^2 = 8.16, p < 0.01$). To examine the effect of the different types of simultaneous engagement, enabling a comparison of head and hand movements, a breakdown according to the classification scheme provided in Section 7.2.3.3 is shown in Table 7.4.

Switching to using nods as a measure, a slightly different pattern is observed. Recipients respond 79.03% of the time compared to a background response rate of 72.32% ($\chi^2 = 5.37, p < 0.05$). The breakdown by type is shown in Table 7.5.

In order to provide a direct comparison of the recipients' relative sensitivity to changes in the speaker's head and hand orientation responses to 'Head Moves' events and 'Hand

Table 7.5 – Response rates by type of event, measured by recipient nodding

Event Class	Response Rate	Baseline Response Rate	Sig
Head Moves	76.96%	72.32%	Not Significant (p=0.18)
Both Move	78.43%	72.32%	Not Significant (p=0.2)
Hand Moves	92.5%	72.32%	$\chi^2 = 7.84,$ $p < 0.01$

Moves' events can be compared. This shows a significant difference between the groups using the values for head nods as a measure of response shown above ($\chi^2 = 5.0, p < 0.05$) and a trend when using head re-orientations ($\chi^2 = 2.73, p = 0.098$).

7.3.3.4 Latency of recipient responses

The time elapsed between a simultaneous event until the first response (movement or change of head orientation) for each recipient was analysed in order to provide another comparison of the behaviour of the recipients to simultaneous and baseline events. A mixed model linear analysis was used with Recipients and Task as random factors and 'Condition' (Simultaneous Engagement Event vs. Baseline) and Task Role (Learner vs Instructor) as within subjects factors. This showed a reliable main effect of Condition ($F_{(1,1064)}$ = 5.07, p = 0.03) but no main effect of Task Role ($F_{(1,30.6)}$ = 0.14, p = 0.71) and no Task Role × Condition interaction ($F_{(1,1060)}$ = 0.89, p = 0.37).

As Table 7.6 shows, recipients' responses to target events were approximately half a second faster than the baseline responses.

7.4 Discussion

The qualitative evidence detailed in this chapter adds to the strictly quantitative findings from the previous two chapters. It has added information about what the interlocutors are

Table 7.6 – Marginal means for recipient response times

Condition	Marginal Mean	Standard Error
Simultaneous Engagement Event	2.3 seconds	0.39
Baseline Event	2.9 seconds	0.38

doing in the dialogues with some of the patterns that have been identified earlier. This section will discuss the implications of some of the current findings.

7.4.1 Simultaneous engagement in dialogue

Simultaneous engagement events allow for the coordinated management of multiple recipients in a multiparty dialogue. Using excerpts from the dialogue it was seen that when speakers need to engage with a second recipient they are not forced to cease their engagement with the first recipient and follow a serial pattern of engagement. Instead, by deploying a combination of head and gesture orientation they are able to engage with both recipients simultaneously. The types of composite orientations seen here resemble examples documented in Schegloff (1998) and Lerner (1992). However, they differ from these findings in their combination of specifically the head and the hands, and their deployment within, and effect on, a dialogue. Moreover, the fact that these events simultaneously engage multiple people is not their defining feature *per se*. Other forms of simultaneous engagement can occur though other means, for example by addressing people with the plural form of 'you'. However, here the coordination is done explicitly via the body. Simultaneous engagement events occur on average every 34 seconds in this task. This appears relatively frequent as Bekker et al. (1995) found that any form of gesture occurred on average every 9 seconds in their corpus of face-to-face interactions.

By measuring recipient behaviour at times after a simultaneous engagement event, and comparing this to their behaviour at baseline times when a simultaneous engagement event has not occurred, it was shown that recipients respond faster, and more often, to simultaneous engagement events than baseline events. This allows us to confirm the hypothesis that they are interactionally significant events.

Whilst it is not possible to pinpoint exactly the communicative function of the events seen here, it is likely that they go some way towards allowing a speaker to tackle the added managerial complexities that arise with multiparty dialogue and multiple recipients. The examples described show (but are not limited to) interlocutors seeking information, performing checks with other interlocutors and referencing items that they had mentioned previously. It is possible that these events, and the responses from recipients, contribute to the incremental construction of utterances. This is in line with early findings by Goodwin (1979), whose work demonstrates how recipient feedback contributes to the construction of an utterance. Simultaneous engagement events extend this to account for the feedback of multiple recipients at the same time.

The lack of any significant distinction between the task roles of the simultaneously engaging interlocutor suggests that any coordination that these events allow for is not restricted to coordination with the party of instructors, but occurs equally, irrespective of task role and party membership.

7.4.2 Interplay of head and hands

Chapter 6 showed through regression analyses that there was significant interplay between an interlocutor's head movements and their hand movements. Discovering the presence and regularity of simultaneous engagement events goes some way towards an understanding of this. When considering the class of simultaneous engagement events in which both head and hands move, this interplay is clear; it it the movement of the head and hands together which define this class. On first glance it appears hard to understand how the other classes of simultaneous engagement events demonstrate this interplay, as only one of them moves. However, there is a difference between the measures used in each chapter which means that this interplay can still be present. The analysis in Chapter 6 made use of definitions based purely on motion data whereas in this chapter it is specifically changes in orientation that are of interest, whilst a gesture is in progress. This means that it is likely that, under the earlier definitions, the hands will be classed as moving during a simultaneous engagement event, even if their orientation does not change (i.e. a Head Moves event). The person would be gesturing simply without changing the orientation of that gesture. Therefore, all classes of simultaneous engagement events have the potential to contribute to the observed interplay.

The current analyses establish the relative interactional significance of changes in orientation of the speaker's head and hands. These situations are marked as the recipients are given the choice of either attending to the orientation of the head or the orientation of the hands. To the extent that the current measures of recipient behaviour can expose, changes in orientation of the hands hold a higher significance for the dialogue than changes in head orientation. This is in contrast to the patterns of addressee attention described in the literature. Gullberg (2003) shows that listeners in a dyadic dialogue gaze mainly at the speaker's face, with little attention shown to the hands. The present analyses demonstrate that these findings do not necessarily directly scale up to the multiparty dialogue; there are times in the dialogue in which the hands, not the face, take precedence.

Recipient response rates also allow us understand gaze behaviour more clearly. When focussing on events in which only the head changes orientation, there is no reliable difference in recipient orientation change rate to that of the baseline measures. The fact that the orientation change in the speaker does not elicit an orientation change in the recipients beyond the baseline measure confirms findings in the literature that recipient gaze does not automatically follow that of the speaker.

7.4.3 Uniqueness of multiparty dialogue

Simultaneous engagement events are one manifestation of the relative complexity of groups in conversation when compared to dyads in conversation. The specific coordination problem, that of coordinating more than one recipient of an utterance, is unique to multiparty interactions. As such it would not be possible to find a simultaneous engagement event, at least in the form described here, without more than one recipient.

7.4.4 Dependence upon a shared space

The features documented in this chapter are distinguished by their intrinsic dependence on physical co-presence. In the cases involving the deployment of shared hierarchies, it was seen that interlocutors adopted a shared frame of reference. Whilst the ability to adopt a shared frame of reference is possible in principle without a shared space, the mutual awareness offered by the shared space underpins the adoption of the reference frame. This frame of reference contrasts with the person-centric ones seen in the literature review. Schober (1993) showed that the frame of reference adopted during dialogue requires negotiation between the interlocutors. Whilst in Schober's study the interlocutors could not see each other, here they can and the observations suggest that part of the negotiation of the reference frame to be used is done by the deployment of the interlocutor's gestures in the shared space. It is possible that if access to the shared space was restricted, the ease with which interlocutors were able to agree on the shared frame of reference would be reduced.

In the cases involving simultaneous engagement events, access to the shared space is essential. Consider the comparison of a multiparty interaction in a shared space with the same interaction over video mediated communication (see Section 3.6). The shared space condition allows the speaker to monitor both recipients' behaviour and likewise the recipients are mutually aware of their engagement. It is essential that a recipient knows that their counterpart is also being engaged so that they can respond appropriately. The peer-to-peer video channels of video mediated communication offer two immediate problems with this situation: 1) the speaker is unable to effectively deploy the combination of head and hands such that it identifies both recipients for engagement 2) the recipients do not have access to the mutual awareness offered by the shared space which is necessary to collaborate in the engagement. At the times of simultaneous engagement, real time audio and visual access to interlocutors over peer-to-peer channels will not adequately support the interaction. It is likely that the compensatory behaviours documented in Whittaker (2003) may help, in part, to compensate for this.

The necessity for a shared space, and the uniqueness of these events to multiparty dialogue also allow for a finer grained understanding of how a shared space is used in the course of an interaction. When discussed by Kendon (1990), the shared space is shown to be delimited by the interlocutors' bodies and used for collaborative activities, however there is no clear distinction in its use between dyads and multiparty conversations. Özyürek (2002)'s work shows that the form of a gesture (specifically an in-out gesture) changes when shifting between 2 and 3 people talking. What is demonstrated here however, is a use of the shared space which is unique to multiparty dialogue. Whilst dyadic dialogue may make use of the shared space, the coordination problems addressed by a simultaneous engagement event do not arise and as such the value of the shared space as a resource for interaction will be lower. Space becomes more important in multiparty dialogues as a resource for coordination.

7.5 Conclusion

This chapter employed a qualitative methodology and subsequently combined this with a quantitative analysis to identify interlocutor behaviours which make direct use of the shared interaction space. Simultaneous engagement events were identified and examined as they directly coordinate multiple recipients of the dialogue and, to be an effective interactive cue, require a mutually accessible shared space. The impact of these speaker events on the recipients was tested, showing a higher level of activity following a simultaneous engagement event than exists in dialogue sequences without an event.

As the events coordinate multiple recipients they are unique to multiparty dialogue. In combination with their dependence on a shared space, they suggest a use of the shared space that is unique to multiparty dialogue; the type of coordination problem that they address would not arise in a dyadic conversation. This adds to the definition of the shared space provided by Kendon (1990). It was also noted that simultaneous engagement events which involve movements of the speaker's hands rather than their head elicit more responses above baseline activity. This implies that, at these times, the hands are a more significant interactional cue than the head.

Part IV

Discussion

Chapter 8

Concluding discussion

8.1 Overview

This thesis questioned how shared space is used as a resource for coordinating conversation. Multiparty (not dyadic) dialogues were studied because of their greater interactional complexity, which maximises the potential to make space relevant. In a co-located space the body plays a key role in coordination, meaning the study of shared space and the organisation of non-verbal cues are intrinsically linked. Thus, this thesis first questioned what the organisation of interlocutors' non-verbal behaviour is within multiparty dialogue.

The key finding is that the organisation of non-verbal behaviour in multiparty dialogue is not the same as the organisation in dyadic dialogue. In multiparty, the shared space is a more important resource, supporting interactionally significant coordination events. The multiple recipients of an utterance are not defined by their lack of speech *per se*, but can be distinguished from each other, with a primary recipient (an approximation of the primary addressee) nodding and gesturing more than a secondary recipient (an approximation of the side participant). Membership of a party (Lerner, 1993) is manifest in interlocutors' non-verbal behaviour, most clearly in their patterns of head orientation. This thesis contributes to a model of multiparty dialogue, particularly concerning non-verbal behaviour, and details the expected head and hand movements of the interlocutors, which are supported by their mutual access to the shared space.

8.2 Summary of thesis

Dialogue is different from monologue; along side depicting content, it brings about the need for interlocutors to coordinate their behaviour in order to regulate the conversation. In face-to-face dialogue the body is said to play a key role in supporting this coordination (Kendon, 1990). For example, gestures can hold managerial functions (Bavelas et al., 1992) and head movements can serve as feedback mechanisms (Boholm and Allwood, 2010). However, existing quantitative claims of coordination (such as those by Condon and Ogston, 1966) are problematic due to the possibility of statistical artefacts (a problem that has been echoed in the literature (e.g. McDowall, 1978)). These artefacts could be non-interactive sources of correlation such as collective fatigue.

Face-to-face interaction takes place in a co-located, three dimensional space. This allows for the creation of a shared interaction space, which is encapsulated by the f-formation system, created using the orientation of interlocutors' bodies (Kendon, 1990). In dialogues which preclude access to the shared space (such as video mediated communication), there is a marked effect on the interlocutors' behaviour (Whittaker, 2003). However, it is not clear how the shared space supports the coordination that is necessary for a dialogue to unfold smoothly.

When conversations scale up from dyadic to multiparty dialogue, the interlocutors are tasked with an added level of coordination complexity. The basic speaker-listener model that encompassed dyadic dialogue is insufficient because it does not account for the multiple non-speaking interlocutors. In this situation the dialogue can be structured by the participation framework (Goffman, 1981) in which people are divided into ratified participants and overhearers. Ratified participants are further divided into speaker, primary addressee and side participant. Clark and Shaefer (1992) provide the principle of collaboration such that 'speakers collaborate directly with addressees but only indirectly with side participants'. However, it is not clear what this means for the organisation of non-verbal behaviour. Kendon (1970) suggests that a speaker and direct addressee can be observed to synchronise their bodies. What is not clear from Kendon's work is how the bodies coordinate, and what the predictions are for the other members of the conversation. In addition, multiple interlocutors give rise to the possible organisation of people into sub-groups that share responsibility and common ground, known as parties (Lerner, 1993). It has been shown that these party structures can influence the organisation of the dialogue, such as turn taking (Lerner, 1993; Schegloff, 1995), and that members of the party share a distinct level of common ground to non-members (Eshghi, 2009). Does this higher-level structure influence the organisation of non-verbal behaviour?

The thesis presented three research questions, aimed at filling some of the gaps identified in the literature above:

1. Do patterns of detectable inter-person head and hand movement exist within a group of interacting people that are systematically different from the inter-person patterns measured across non-interacting individuals?
2. Are there global patterns of head and hand movements that relate to the unique structure of a multiparty dialogue in terms of its constituent interlocutors' dialogue roles?
3. Are there communicative behaviours that intrinsically rely on their deployment in a mutually accessible shared interaction space and, if so, what do they look like?

These questions were addressed across three empirical chapters, exploiting video and motion capture data held in two corpora (one of which was collected as part of this research). Data analysis methods and software were developed in order to answer the research questions.

The first empirical chapter (Chapter 5) tested the statistical relationship between interlocutors' speech, head movement and hand movement. To account for possible statistical artefacts such as collective fatigue, these relationships were compared to a) control groups, which were artificially constructed such that the constituent members were from different original conversations, and b) randomised data in which the interlocutors didn't interact and any temporal structure was disrupted. A test was applied to the corpora to determine if a common temporal structure existed, possibly accounting for the statistical artefacts. It was shown that interlocutors influence each other's verbal and non-verbal behaviour. They tend not to speak or gesture over each other, but do nod together. This is important as it validates the argument made in the second chapter of the literature review: that the behaviour of interlocutors in dialogue is different to individuals performing monologues as, in dialogue, multiple people must coordinate their actions. The new methodology identified a common temporal structure across the interactions in each corpus (see Section 5.3.2 for details of these structures). This structure would cause statistical artefacts in prior work, and may have been incorrectly identified as interactional coordination. This temporal structure may be caused by the task at hand, as it differs between corpora with different tasks.

The second empirical chapter (Chapter 6) opened up the dialogue to expose the dynamic participation framework. It performed descriptive and inferential statistical tests to compare the head orientation, nodding and gesture behaviour of speakers and non-speakers, primary and secondary recipients, and learners and instructors. The results were compared across corpora to identify any effect of task. It was found that, confirming suggestions in the literature, speakers are distinct from non-speakers in their head and hand movement. Importantly, recipients are distinct from each other; primary recipients are more likely to nod and gesture than secondary recipients are. This extends Clark and Shaefer (1992)'s principle of collaboration to the non-verbal domain. There are particularly interesting findings regarding party structures. When there is no measurable party present, as is the case with the balloon task corpus, both recipients prefer to orient their heads towards the speaker. However, in the cases were the task imposes a party, as is the case in the tuition task corpus, interlocutors prefer to orient towards a non-member of the party irrespective of who is speaking at that time. This is important because it shows that party structures are manifest in interlocutors' non-verbal behaviour.

The third empirical chapter (Chapter 7) performed a hybrid qualitative and quantitative analysis to interpret the behaviour of the interlocutors within their shared space. Simultaneous engagement events, which are performed by speakers, were uncovered and identified as unique to a multiparty dialogue because they coordinate multiple recipients at the same time. Mutual access to a shared space is required for the event to have communicative significance. The chapter compared the behaviour of recipients after simultaneous engagement events to baseline measures of their behaviour during the interaction, and compared their response rates to different classes of simultaneous engagement events (defined by which part of the speaker's body changes orientation: Head Moves, Hand Moves, Both Move). It was shown that the events hold interactional significance, eliciting more frequent and faster responses from recipients compared to their baseline measures. It was found that, in contrast to patterns from dyadic dialogue (Gullberg, 2003), events which involve changes in orientation of the hands elicit more movement in the recipients than changes in orientation of the head. This chapter concluded that, as simultaneous engagement events are unique to multiparty dialogue, the shared space may be a more important communicative resource than it is for dyads.

Taken together, these findings provide a number of contributions:

- It has been demonstrated, using a contemporary methodology, that interactions elicit different non-verbal behaviour than monologues. Interlocutors coordinate their verbal and non-verbal behaviour
- A finer level of granularity in the definition of the shared space, specifically its use as a coordination mechanism in multiparty dialogue, has been shown.
- A model of non-verbal behaviour in multiparty dialogue which shows the expected head orientation, nodding and gesture behaviour of the interlocutors has been created
- Evidence for the organisation of non-verbal behaviour by higher-level structures, such as parties, has been provided.
- Evidence that patterns of non-verbal behaviour in dyadic dialogue do not scale up directly to the multiparty dialogue has been presented.
- Methods and software for the quantitative analysis of human interaction data have been developed.

These contributions are important because, taken together, they make crucial steps towards modelling the communicative body movements used during multiparty dialogue. This thesis contributes (but is not limited) to the fields of human interaction, psychology and linguistics as it exposes the relationships between body movements and the moment-by-moment organisation of dialogue that underpin everyday group conversations, making clear that non-verbal behaviour in multiparty dialogues does not scale up directly from dyads. This can be utilised by designers of technologies that both support and analyse group conversations (such as video mediated communication and meeting analysis systems respectively) as the model of interlocutor behaviour and use of space can be integrated into the technologies, and researchers and practitioners of mental health as understanding the mechanisms that drive 'normal' group conversations will enable pathologies that are manifest in social interaction to be detected (see e.g. Lavelle et al., submitted).

8.3 Reflections

This section reflects on the findings of this thesis and considers the implications for both the immediate and related fields. It will be partitioned in four sections: 1) considers how this thesis extends the existing literature that concerns the structure and expected patterns of interlocutor behaviour in multiparty conversations, 2) discusses the revised and new methodologies used, and the potential for exploiting the software developed in this thesis for further human interaction research, 3) assesses how the model of non-verbal behaviour in multiparty interaction provided by this thesis impacts upon the development of social technologies and 4) presents some open questions that remain when considering conversations that take place, not in the lab, but in everyday settings.

8.3.1 Understanding multiparty dialogue

8.3.1.1 Organising dialogue

There is more than one way to organise body movements in a multiparty dialogue. The first is by the roles that the interlocutors adopt during the dialogue, or the participation framework (Goffman, 1981). The current findings provide evidence that this framework is manifest in the interlocutors' non-verbal behaviour, in particular for the recipients. It is not the case that recipients are defined simply by their lack of speech; instead, one must consider the relationship that they hold at any given point in time with the speaker. It has been shown that, in extension to Clark and Shaefer (1992)'s principle of collaboration, the primary recipient (an approximation of the primary addressee) has a level of coordination with the speaker, at least in their head and hand movement, that is higher than the level of coordination that the secondary recipient holds with the speaker. This suggests that the same utterance, produced by the speaker, has different consequences for each recipient. If speakers were simply senders of denotational information, this should be received and, errors aside, comprehended in the same way by both recipients. That it is not, and this difference is visible, suggests that the speaker and the primary recipient collaborate in the construction of the utterance.

The second method by which the movements can be organised is at a higher level, into groups of individuals, or parties. As discussed by Lerner (1993) and Schegloff (1995), it was shown that party membership can influence the organisation of turns within the dialogue. Using text based dialogue experiments, Eshghi (2009) further demonstrated that parties can be indexed semantically, sharing a distinct level of grounding from the others in the dialogue. A key finding from this thesis, and one which adds to existing studies, is that party membership is manifest in an interlocutor's movements. This is interesting because it adds weight to the existing argument that dialogue can be structured by higher-level mechanisms such as parties. The current findings state that it is possible to see, or at least for a computer to see, parties in action.

This organisational duality has implications for the participation framework. Goffman (1981) defined a participant's status relative to a speaker's utterance to be 'the relation of any one such member to this utterance'. However, the fact that, in certain situations, the non-verbal behaviour of an interlocutor is indexed, not just by their relationship to the speaker's utterance but by their party membership raises concerns over using just their status in the dialogue as their distinguishing feature. Consider, by the definition in the participation framework, the secondary recipient. Using data acquired by comparing two corpora, this thesis has demonstrated that a secondary recipient in one situation has different patterns of behaviour to a secondary recipient in another situation, meaning they cannot be treated as equal. This was most manifest in the head orientation of a secondary recipient who, when there was no party present, would orient towards the speaker, but when part of a measurable party chose to orient outside of the party, irrespective of who was speaking. These findings suggest that future attempts to index an interlocutor within the dialogue may need to consider aspects of the participation framework in tandem with any party structures that exist. One approach, similar to Levinson (1988)'s in which he identifies members of a conversation using 'yes or no' parameters such as 'participant' and 'recipient', could be adding a parameter representing party membership to the participation framework. This raises the question (an echo of a question raised by Eshghi (2009)) how do parties arise when the task structure, that was imposed on the interlocutors in the tuition task corpus, is not present? Whilst not possible to answer fully, times of simultaneous engagement may project a transient party. Consider the time that a speaker is speaking to the ostensible primary addressee and, by turning his head orientation, performs a check of what is being said with the side participant. This resembles 'conferring' as discussed by Lerner (1993), which is said to make relevant a party for the interaction. It is possible that simultaneous engagement events, and hence the shared space, allow for a process of conferring non-verbally, thus creating a temporary party.

A link can be drawn here to the incremental construction of utterances. Goodwin (1979) discussed the structure of dialogue in which the incremental construction of an utterance at the conversational level was seen. In the dialogue, a speaker reformulates their utterances depending upon whether they secure an addressee and the subsequent feedback that they are given (in Goodwin's case this is receiving their gaze, and hence attention). The coordination between the primary recipient and the speaker observed in this thesis gives weight to the argument that addressee feedback influences the construction of the utterance. The primary recipient tends to gaze back at the speaker most of the time and nod when the speaker speaks. However, as noted, the behaviour of the secondary recipient is variable depending upon the presence of the party. This does not appear disruptive, suggesting that these gaze patterns are mutually accepted within the system of the dialogue. An interpretation of this is that the secondary recipient's attention holds less significance for the speaker's dynamic construction of the utterance.

Exceptions to this are the marked occasions where the speaker displays a need for the secondary recipient to engage with the construction of their utterance. These occur during times that have been termed simultaneous engagement events. At these times, the simultaneous engagement event contributes to the incremental construction of the speaker's utterance by allowing them to receive feedback from multiple recipients at the same time. The actions of both of these recipients will influence what is said next by the speaker. A further point to note is that, as both recipients are being engaged and providing feedback to the speaker, the distinction, at this point in time, between primary and secondary recipients is blurred. This again raises questions about the use of the participation framework alone to define interlocutor behaviour.

An interesting point to note here is how this discussion is unique to multiparty dialogue. These problems would not present themselves to dyads. Speakers and recipients may only be organised with these roles as the there is no possibility of organisation in groups at a resolution lower than the dyad itself.

8.3.1.2 Orchestrating dialogue

A parallel accumulation of findings relate, not to how the interlocutors of the dialogue are organised, but how their bodies orchestrate the dialogue. A running prediction through the thesis was that head orientation, rather than eye-gaze, should be a significant cue in the interaction. That head orientation is a significant cue was proved correct; it serves as an address cue, differentiating primary and secondary recipients. The salience of this cue is demonstrated as patterns emerge, even with the somewhat naive use of head orientation employed in this methodology. This method assigns recipiency based upon all speaker orientations (except those straight down the middle of the interaction), which are likely to include orientations that are potentially oriented outside of the interaction space. Even with this added risk of simply detecting noise, significant patterns emerge hinting at the potential strength of this cue in the interaction. This confirms initial findings by Jokinen et al. (2010) of the significance of head orientation, and extends the low-level findings of Loomis et al. (2008) to the conversational level. It is not straightforward to make any direct claims about the use of eye-gaze as an interactional cue. It could be expected, given the effective use of head orientation and the Loomis et al. (2008) findings, that eye-gaze is only considered a weak cue. However, the data do not exist to back this up.

Adding to this, directly comparing changes in head orientation to changes in hand orientation during simultaneous engagement, shows that gesture is a more significant interactional cue than head orientation when measuring recipient behaviour (presented in Battersby and Healey, 2010a). It appears then, that multiparty dialogue is orchestrated by cues which sit on a continuum, with eye-gaze as rather ineffectual at one end, moving through head orientation and ending with highly effective gesture. Whilst challenging to account for this, it could be explained again by the Loomis et al. (2008) results, that is, it is easier to monitor large movements in one's peripheral vision than small movements.

These observed patterns, once again, present differences between dyadic dialogue and multiparty dialogue. As covered in the literature review (in particular Chapter 3) eye-gaze is considered an essential cue for interaction in dyads, and gesture is secondary to cues in the face. The reverse has been observed during this thesis using multiparty interaction data.

Open questions remain as to how the rest of the body orchestrates the interaction. Whilst the decision to focus on head and hand movements was motivated by the literature, the role of the rest of the body has remained unvisited. The analyses used in this thesis took data from, maximally, 6 markers per subject whilst the dataset holds 27 makers per subject. This means that, within the existing rich data, there exists the potential to examine the effect of upper body orientation and posture as potential orchestration mechanisms. Kendon (1973) showed that the shared space is indexed by the orientation of the interlocutors' lower bodies. It would be an interesting extension to remove the seating constraints and allow participants to naturally form an f-formation, capturing data on the coordinated orientations of their lower bodies also.

8.3.1.3 Shared space and interactive topologies

The shared space has been shown to be an essential resource for multiparty, face-to-face interaction (see Chapter 7). It could be argued that, during a face-to-face conversation, the shared space supports the construction an interactive topology. Similar notion of topologies have been documented in Healey and Peters (2007) and Heath and Healey (2011) that describe collaborative drawing and how this activity structures space. The interactive topology is borne out of the interactional cues deployed during an unfolding dialogue, such as address and turn taking cues, salient topics and simultaneous engagement - and other coordination - events. In the face-to-face setting seen in our data, the interactive topology is contingent upon mutual access to a shared interaction space. It is the mutual knowledge that the shared space offers that allows interlocutors to collaborate and develop the topology. The frame of reference used to define the interactive topology is centred in the interaction, rather than individuals.

Dyads' topologies will differ to those found in a multiparty dialogue. Those aspects of the topology that are contingent upon access to the shared interaction space are more important for multiparty dialogue. The topology may also change depending upon the context; a performer will share with her audience, who also mutually share, a different interactional topology to that seen in the local context of a dialogue, in which there are different constraints and implications upon its constituent members (see e.g. Healey et al. (2009) for findings based on data from a lecture, and Gardair et al. (2009) for work on the interactional processes involved with street performers and audiences). The topology is important because it contains the interactive cues that need to be supported in media that do not give access to a shared space.

A question that remains is whether the shared space can be decomposed further, beyond its current level where it is delimited by the interlocutors' bodies. Are there different regions in the space that lend support to different cues, perhaps for different members of the interaction. It is possible that party relationships are manifest in the topology, exploiting the shared space to support coordinating cues within the party differently to those that coordinate outside the party. It has not been possible to answer this question during this thesis, however during the course of examining methodologies for the presented studies one technique was developed that may provide a basis to start from. It uses clustering techniques, applied to hand motion data, in order to assign priorities to areas of the shared space. Frequently used areas will develop into clusters, infrequently used space will remain empty. A graph representing a work in progress attempt at this is shown in Appendix C.1, with larger areas representing more frequently visited zones. Tests would need to be developed to determine if these zones correlate with other features of the dialogue. Furthermore, techniques are required to effectively summarise multiple graphs and compare them between conditions.

8.3.2 Analysing dialogue

This thesis has helped with a common problem faced by those that study human interaction; they are limited by the technology available. For example, video analysis reduces the spatial data available in a scene into two dimensions. The motion capture data used has provided full three dimensional data of the multiparty interactions. However, even though this generates high volumes of spatial data, extracting useful statistics from this data is a difficult problem. Techniques and software were needed to make this data useful. In order to answer the research questions of this thesis, analysis methods have been created and coded into Python software packages to do this (presented, in part, in Battersby et al., 2008; Battersby and Healey, 2010b).

This software has been used to interrogate an existing methodology, in which claims of correlation may have been spurious (Chapter 5). By refining the existing methodology, new results hold greater validity. By creating and using indexes to interactional phenomena, the software has been able to derive meaningful statistics that represent the interaction and the collaborative behaviours of interlocutors. Whilst the software packages have proved useful for current purposes, they are able to form the base of a generic tool for analysing interaction data from automated sources, allowing the research community to benefit from the software made as part of this thesis. The software in its current state has been exploited during independent research, examining social interaction in patients with a diagnosis of schizophrenia (Lavelle et al., submitted).

8.3.3 Implications for technologies

The contributions that this thesis brings concerning shared space and the collaborative use of bodies in multi-party dialogues will have implications for the design of social technologies. As Nijholt et al. (2009) describe, social technologies require knowledge from studies of multimodal corpora to be design into the systems. One type of social technology, video mediated communication, was discussed in the literature review. The findings concerning shared space have particular relevance to this technology. There are also other contemporary and emerging technologies that will benefit from the current findings. These have been split into two categories: the first are those technologies that support, or engage in, conversation (of which video mediated communication is one), second are those technologies that attempt to automatically analyse conversation (such as automated meeting analysis systems).

8.3.3.1 Supporting conversations

Video mediated communication is distinct from face-to-face communication in its lack of mutual access to a shared space. In this transformed space, features of the interactive topology need to be supported by the technology rather than the shared interaction space. At a basic level this may be that address cues using head orientation need to be supported in the interface (such as the 'addressee assistant' described in Gatica-Perez and Odobez (2010)), a more complex problem would be to support the ability to coordinate multiple recipients as the shared space allows for in face-to-face dialogue. It is not immediately apparent how this would be done, but this thesis give designers of these technologies the social theory required to help create appropriate systems.

Technology not only mediates communication, but can augment it. Consider the design of interactive table tops, such as the Diamond Touch (Dietz and Leigh, 2001) or the Microsoft Surface[1]. These devices sit in a shared space, and augment co-located conversation with artefacts that can be shared and collaboratively manipulated. The usability of these surfaces can be improved with the current findings by adapting to the current conversational context. Assuming the device has the relevant input data, it could, for example, tailor the information presented depending upon an interlocutor's state within the dialogue or task based role they adopt, or present relevant information to multiple recipients when they are simultaneously engaged.

Further still, the computer can attempt to engage in conversations with people itself. Traditionally this has been done using text or audio based dialogue systems, however new technologies are emerging that make use of some form of embodiment, usually as either a robot or a computer animated avatar (commonly referred to as an Embodied Conversational Agent, or ECA). One of the main concerns that this field currently faces is not how to create an aesthetically pleasing avatar or robot, but to design into the ECA the ability to hold successful social interaction with humans. Significant efforts are under way with this. For example, Vilhjálmsson (2009) reports on a community wide effort to create a structured approach to modelling avatar behaviour, separating intention and behaviour levels, and providing a dedicated behaviour modelling language. The work here can contribute to the theory underpinning this modelling, although translating the findings from dialogues presented here to a state driven avatar is non-trivial.

8.3.3.2 Automated conversation analysis

Research effort in this domain is concerned with building machines that are able to understand and interpret an interaction scene (for an overview see e.g. Tur et al. (2010)). This could be done, for example, to automatically generate meeting minutes. The field is broad, and systems can use audio or textual data as input, or take advantage of intelligent meeting rooms with coordinated sensors covering multiple modalities. What ever the input data, the systems require a model of the dialogue to be able to effectively interpret it. As Gatica-Perez and Odobez (2010) state, non-verbal cues can be essential for these systems to infer important facets of group conversations. The findings presented in this thesis

[1] http://www.microsoft.com/surface

can assist with this, providing a statistical model of how interlocutors bodies would be moving within the dialogue. A basic example, and a problem noted by Gatica-Perez and Odobez (2010), is determining who the current speaker of a conversation is, as often the data from microphones are inaccurate. Using the current findings, adding the nodding and gesture behaviour of interlocutors to audio in a predictive model underpinning the system could improve its accuracy. Similarly, if one were to assume that the speaker may be the person to whom the rest of the members of the conversation are oriented towards (as would be expected if one were to scale up from orientation patterns in dyadic dialogue), this would be incorrect. We have shown that there are times when orientation is governed by party membership, not speaking status. Using our current findings would help to improve methods of speaker identification.

Another challenging problem in this field is segmenting and identifying the topic under discussion (see e.g. Purver et al. (2006) who use transcripts of dialogue). This thesis has seen that interlocutors load the shared space with meaning, such as with shared gestural hierarchies (Section 7.2.3), using space to collaboratively structure and refer to salient topics under discussion. These spatial features of the interaction could assist with identifying the mutually accepted topics of discussion. A further motivation for understanding the structure and topics of co-located conversations in real time is to enable a remote participant joint more easily over a video link.

8.3.4 Conversations in the wild

The claims made throughout this thesis are based upon the corpora used, and hence it is challenging to make broad claims beyond them. The findings of this thesis demonstrate the processes that can happen within conversations, they cannot, and are not meant to, account for all conversations. Some questions that arise when moving from the constraints of the lab to conversations in the wild are:

- How do the complex relationships, prior knowledge and personal history influence the formation of parties? If the dialogue consists of a couple, and a third person do the couple constitute a party? If one member of the couple and the third person shared an experience that is being retold, how does this potential party's requirements conflict with those of the couple? Will these situations be manifest in the interlocutors' non-verbal behaviour and if so how?

- How do artefacts play a role in conversations? Very often people collaborate on tasks such as drawing with pen and paper, using computers and technology, make use of machinery and more. How do people collaborate to structure space with the added constraints of working around a physical object? See Heath and Healey (2011) for initial work on this question.
- What happens when conversations grow and fragment? What are the normative processes that occur at the inter-group level? Although not in the strict dialogic sense, these patterns are organised across groups of people with joint responsibility and may resemble parties.
- What happens if the constraint of only quantitatively analysing the canonical multiparty conversation with one speaker is removed, and the analysis is opened up to explore overlapping speech? The socio-dynamic processes must change drastically when there is competition. Indeed, it will be necessary to identify the nature of the overlapping speech to determine if it is competitive at all (Schegloff (1995) has described some of the possible variations of overlapping speech). It may be that the changes in behaviour seen with co-telling overlapping speech are different to competitive overlapping speech.
- How does culture influence the organisation of dialogue? The corpora used in this thesis were developed in London, England. If they had been constructed in a location with different cultural norms would this change the observed organisation and, if so, how?

To generate valid answers to these questions, it may be necessary to adopt different methodologies to those used in this work. The predominantly quantitative, corpus based approach has been suitable for the types of questions asked in this thesis. However, in the wild it may not be; instead a methodology such as ethnography may be more appropriate.

8.4 Final words

The work presented in this thesis has made inroads into uncovering how the shared space can be used as a resource for coordinating a multiparty dialogue, and the associated non-verbal behaviour that takes place within it. This is a broad and complex field, with findings accumulated from various disciplines with varying methodologies. It is hoped that the findings offered here give the reader a better understanding of the complexity of multiparty conversation, conversations that we take part in almost every day, and some of the ways in which we use our bodies in shared space to address this complexity.

To conclude, it has been shown that the patterns of non-verbal behaviour in dyadic dialogue do not directly scale up to multiparty dialogues. A, currently unanswerable, question that is raised by this finding is: which form of dialogue is the norm and which is the special case? One may assume that the addition of a third party to a dyadic dialogue creates the special case: multiparty dialogue. Instead, it is possible that multiparty dialogue allows us to see fully the underlying norms that are present in *all* dialogue, the reduction to the special case of dyads simply masks them.

Bibliography

Allwood, J. and Cerrato, L. (2003). A study of gestural feedback expressions. In Piaggio, P., Jokinen, K., and Jönsson, A., editors, *First Nordic Symposium on Multimodal Communication*, pages 7–22, Copenhagen.

Argyle, M. (1975). *Bodily Communication*. Methuen & Co. Ltd, Bristol.

Ashenfelter, K. T., Boker, S. M., Waddell, J. R., and Vitanov, N. (2009). Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of experimental psychology. Human perception and performance*, 35(4):1072–91.

Battersby, S. A. and Healey, P. G. T. (2010a). Head and Hand Movements in the Orchestration of Dialogue. In Ohlsson, S. and Catrambone, R., editors, *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 1998–2003, Portland, USA.

Battersby, S. A. and Healey, P. G. T. (2010b). Using head movement to detect listener responses during multi-party dialogue. In Kipp, M., Martin, J.-C., Paggio, P., and Heylen, D., editors, *Proceedings of LREC Workshop on Multi-Modal Corpora: Advances in Capturing, Coding and Analysing Multimodality*, pages 11–15, Valletta, Malta.

Battersby, S. A., Lavelle, M., Healey, P. G. T., and McCabe, R. (2008). Analysing Interaction: A comparison of 2D and 3D techniques. In Martin, J., Paggio, P., Kipp, M., and Heylen, D., editors, *Conference on MultiModal Corpora*, pages 73 —- 76, Marrakech, Morocco.

Bavelas, J., Kenwood, C., Johnson, T., and Phillips, B. (2002a). An experimental study of when and how speakers use gestures to communicate. *Gesture*, 2(1):1–17.

Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive Gestures. *Discourse Processes*, 15(4):469–489.

Bavelas, J. B., Coates, L., and Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6):941–952.

Bavelas, J. B., Coates, L., and Johnson, T. (2002b). Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52:566–580.

Bavelas, J. B. and Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*.

Bavelas, J. B., Gerwing, J., Allison, M., and Sutton, C. (2011). *Dyadic evidence for grounding with abstract deictic gestures*, pages 49–60. John Benjamins.

Bekker, M. M., Olson, J. S., and Olson, G. M. (1995). Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *DIS '95: Proceedings of the 1st conference on Designing interactive systems*, pages 157–166. ACM.

Bellugi, U. and Kilma, E. S. (1982). *From gesture to sign: deixis in a visual-gestural language*. John Wiley and Sons, Ltd.

Boholm, M. and Allwood, J. (2010). Repeated head movements, their function and relation to speech. In Kipp, M., Martin, J.-C., Paggio, P., and Heylen, D., editors, *Proceedings of LREC Workshop on Multi-Modal Corpora: Advances in Capturing, Coding and Analysing Multimodality*, pages 6–10, Valetta, Malta.

Bull, P. and Connelley, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behaviour*, 9(3):167–187.

Cassell, J., Nakano, Y. I., Bickmore, T. W., Sidner, C. L., and Rich, C. (2001). Non-Verbal Cues for Discourse Structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 114–123, Toulouse, France.

Cassell, J., Stone, M., Douville, B., Prevost, S., Achorn, B., Steedman, M., Badler, N., and Pelachaud, C. (1994). Modeling the Interaction between Speech and Gesture. In *Proceedings of Cognitive Science Society Annual Meeting*.

Cerrato, L. and Svanfeldt, G. (2005). A method for the detection of communicative head nods in expressive speech. In *Gothenburg papers in Theoretical Linguistics 92: Pro-*

ceedings from The Second Nordic Conference on Multimodal Communication, pages 153–165, Göteborg.

Christenfeld, N., Schachter, S., and Bilous, F. (1991). Filled Pauses and Gestures: It's not a coincidence. *Journal of Psycholinguistic Research*, 20(1):1–10.

Clark, H. and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Clark, H. H. and Carlson, T. B. (1982). Hearers and Speech Acts. *Language*, 58(2):332–373.

Clark, H. H. and Schaefer, E. (1989). Contributing to Discourse. *Cognitive Science*, 13(2):259–294.

Clark, H. H. and Shaefer, E. F. (1992). Dealing with Overhearers. In Clark, H. H., editor, *Arenas of Language Use*. University of Chicago Press.

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

de Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *International Journal of Speech-Language Pathology*, 8(2):124–127.

de Ruiter, J. P. (2007). Some Multimodal Signals in Humans. In Van Der Sluis, I., Theune, M., Reiter, E., and Krahmer, E., editors, *Proceedings of the Workshop on Multimodal Output Generation MOG 2007*, pages 141–148.

de Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535.

Dietz, P. and Leigh, D. (2001). DiamondTouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, New York, New York, USA. ACM Press.

Dittmann, A. T. and Llewellyn, L. G. (1969). Body movement and speech rhythm in social conversation. *Journal of personality and social psychology*, 11(2):98–106.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visiospatial Orienting. *Visual Cognition*, 6(5):509–540.

Duncan, S. and Fiske, D. W. (1985). *Interaction Structure and Strategy*. Cambridge University Press.

Efron, D. (1941). *Gesture and Environment*. King's Crown Press, New York.

Ekman, P. and Friesen, W. V. (1969). The Repertoire Of Nonverbal Behavior: Categories, Origins, Usage and Coding. *Semiotica*, 1(1):49–98.

Emmorey, K. (2002). *Language, Cognition, and the Brain*. Lawrence Erlbaum Associates.

Enfield, N. J. (2005). The Body as a Cognitive Artifact in Kinship Representations. *Current Anthropology*, 46(1):51–81.

Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams and objcet demonstrations are integrated in multimodal explanations. In Gernsbacher, M. A. and Derry, S. J., editors, *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, pages 321–326.

Eshghi, A. (2009). *Uncommon ground : the distribution of dialogue contexts*. Phd, Queen Mary University Of London.

Eshghi, A. and Healey, P. G. T. (2007). Collective States of Understanding. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, number September, pages 2–9, Antwerp.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition*, 117(3):319–31.

Frampton, M., Fernández, R., Ehlen, P., Christoudias, M., Darrell, T., and Peters, S. (2009). Who is "You"? Combining Linguistic and Gaze Features to Resolve Second-Person References in Dialogue. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 273–281, Athens, Greece. Association for Computational Linguistics.

Furuyama, N. (1993). Gesture and speech of a hearer as what makes the field of promoting action. In *Poster at Seventh International Conference on Event Perception and Action*, Vancouver, Canada.

Furuyama, N. (2000). Gestural Interaction between the instructor and learner in origami instruction. In *Language and Gesture*. Cambridge University Press.

Gardair, C., Healey, P. G. T., and Welton, M. (2009). I would like you to clap your hands . In *CHI Crowd Computer Interaction Workshop*, Boston, USA.

Gatica-Perez, D. and Odobez, J.-M. (2010). *Visual attention, speaking activity, and group conversational analysis in multi-sensor environments*, pages 1–29. Springer.

Gill, S. P. and Borchers, J. (2003). Knowledge in co-action: social intelligence in collaborative design activity. *AI & Society*, 17(3-4):322–339.

Gill, S. P., Kawamori, M., Katagiri, Y., and Shimojima, A. (2000). The role of body moves in dialogue. *RASK*, 12:89–114.

Goffman, E. (1966). *Behavior in Public Places: Notes on the Social Organization of Gatherings*. The Free Press.

Goffman, E. (1981). *Forms of Talk*. University Of Pennsylvania Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In Psathas, G., editor, *Everyday language: Studies in ethnomethodology*, pages 97–121. Irvington Publishers.

Gresty, M., Leech, J., Sanders, M., and Eggars, H. (1976). A study of head and in spasmus nutans. *British Journal of Ophthalmology*, 60(9):652–654.

Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4):274–279.

Griffin, Z. M. and Oppenheimer, D. M. (2006). Speakers Gaze at Objects While Preparing Intentionally Inaccurate Labels for Them. *Journal of experimental psychology: learning, memory and cognition*, 32(4):943–948.

Gullberg, M. (1995). Giving language a hand: gesture as a cue based communicative strategy. *Working Papers*, 44:41–60.

Gullberg, M. (2003). Eye movements and gesture in human face-to-face interaction. In Hyönä, J., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movements*, pages 685–703. Oxford: Elsevier.

Hadar, U., Steiner, T., Grant, E., and Clifford Rose, F. (1983). Kinematics of Head Movements Accompanying Speech During Conversation. *Human Movement Science*, 2:35–46.

Hanna, J. and Brennan, S. (2007). Speakers eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4):596–615.

Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. (2006). Spatiality in videoconferencing. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work - CSCW '06*, page 413, New York, New York, USA. ACM Press.

Haviland, J. B. (1993). Anchoring, Iconicity, and Orientation in Guugu Yimithirr Pointing Gestures. *Journal of Linguistic Anthropology*, 3(1):3–45.

Hayashi, M. (2005). Joint turn construction through language and the body: Notes on embodiment in coordinated participation in situated activities. *Semiotica*, 156:21–53.

Healey, P. G. T., Frauenberger, C., Battersby, S. A., Oxley, R., Schober, M., and Welton, M. (2009). Engaging Audiences. In *CHI Crowd Computer Interaction Workshop*, Boston, USA.

Healey, P. G. T., Leach, J., and Bryan-Kinns, N. (2005). Inter-play: Understanding Group Music Improvisation as a Form of Everyday Interaction. In *Proceedings of Less is More - Simple Computing in an Age of Complexity*, Cambridge.

Healey, P. G. T. and Peters, C. R. (2007). The Conversational Organisation of Drawing. *First International Workshop on Pen-Based Learning Technologies (PLT 2007)*, pages 1–6.

Heath, C. and Luff, P. (1997). *Reconfiguring Media Space: Supporting Collaborative Work*, pages 323–347. Lawrence Erlbaum Associates.

Heath, C. P. R. and Healey, P. G. T. (2011). Making Space for Interaction : Architects Design Dialogues. In *Gesture Workshop*, Athens, Greece.

Holler, J. (2010). Speakers Use of Interactive Gestures as Markers of Common Ground. In Kopp, S. and Wachsmuth, I., editors, *Gesture in Embodied Communication and Human-Computer Interaction. 8th International Gesture Workshop. Revised Selected Papers*, pages 11–22, Bielefeld, Germany. Springer.

Howes, C. (in prep). *Coordinating in dialogue: Using compound contributions to join a party*. PhD thesis.

Jokinen, K. (2009). Gestures in Alignment and Conversational Activity. In *Proceedings of the PACLING Conference.*, pages 141–146, Sapporo, Japan.

Jokinen, K. (2010). *Gestures and Synchronous Communication Management*. Springer.

Jokinen, K., Nishida, M., and Yamamoto, S. (2010). On Eye-gaze and Turn-taking. In Andre, E. and Chai, J. Y., editors, *International Workshop on Eye Gaze in Intelligent Human Machine Interaction*, Hong Kong, China.

Jokinen, K. and Vanhasalo, M. (2009). Stand-up Gestures Annotation for Communication Management. In Navarretta, C., Paggio, P., Allwood, J., Ahlsén, E., and Katagiri., Y., editors, *NODALIDA 2009 workshop Multimodal Communication: from Human Behaviour to Computational Models*, pages 15–20.

Kendon, A. (1970). Movement Coordination in Social Interaction: Some Examples Described. *Acta Psychologica*, 32:100–125.

Kendon, A. (1973). The Role Of Visible Behaviour in the Organization of Social Interaction. In *Social Communication and Movement*. Academic Press.

Kendon, A. (1977). *Studies in the behavior of social interaction*, chapter 1, pages 13–51. Indiana University , Bloomington and The Peter De Ridder Press, Lisse.

Kendon, A. (1990). *Conducting Interaction: patterns of behavior in focused encounters*. University of Cambridge.

Kendon, A. (1992). The negotiation of context in face-to-face interaction. In *Rethinking Context: Language as in interactive phenomenon*. Cambridge University Press.

Kendon, A. (2002). Some uses of the head shake. *Gesture*, 2(2):147–182.

Kendon, A. (2004). *Gesture: Visible Actions as Utterance*. Cambridge University Press.

Kendon, A. (2010). Spacing and Orientation in Co-present Interaction. *Development of Multimodal Interfaces: Active Listening and Synchrony*, 5967:1–15.

Kingstone, A. (2009). Taking a real look at social attention. *Current opinion in neurobiology*, 19(1):52–6.

Kita, S., van Gijn, I., and van Der Hulst, H. (1998). Movement Phases in Signs and Co-Speech Gestures, and their Transcription by Human Coders. In *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop*, volume 1371. Springer Berlin / Heidelberg, Bielefeld, Germany.

Knapp, M. L. (1978). *Nonverbal Communication in Human Interaction*. Holt, Rinehard and Winston, 2 edition.

Koutsombogera, M., Ammendrup, S. M., Vilhjálmsson, H. H., and Papageorgiou, H. (2011). *Nonverbal Expressions of Turn Management in TV Interviews : A Cross-Cultural Study between Greek and Icelandic*, pages 207–213. Springer-Verlag.

Langton, S. R. H. and Bruce, V. (1999). Reflexive Visual Orienting in Reponse to the Social Attention of Others. *Visual Cognition*, 6(5):541–567.

Lavelle, M., Battersby, S. A., Healey, P. G., and McCabe, R. (submitted). Nonverbal behaviour and conversational role: A 3D Motion Capture Study of Triadic Interaction in Schizophrenia. *American Journal Of Psychiatry*.

LeBaron, C. D. and Streeck, J. (1997). Built Space and the Interactional Framing of Experience During a Murder Interrogation. *Human Studies*, 20:1–25.

Lerner, G. H. (1992). Assisted Storytelling: Deploying Shared Knowledge as a Practical Matter. *Qualitative Sociology*, 15:247–271.

Lerner, G. H. (1993). Collectivities in action: Establishing the relevance of conjoined participation in conversation. *Text - Interdisciplinary Journal for the Study of Discourse*, 13(2):213–246.

Levinson, S. (1988). *Putting linguistics on a proper footing: Explorations in Goffman's Concepts of Participation*, pages 161–227. University Of Pennsylvania Press.

Levinson, S. C. (2003). *Space in Language and Cognition: Exploration in Cognitive Diversity*. Cambridge University Press.

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., and Beall, A. C. (2008). Psychophysics of perceiving eye and head direction with peripheral vision: Implications for the dynamics of eye gaze behaviour. *Perception*, 37:1443–1457.

Lyman, S. M. and Scott, M. B. (1967). Territoriality: A Neglected Sociological Dimension. *Social Problems*, 15:236–249.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in cognitive sciences*, 8(3):108–14.

Marchand, T. H. (2010). Embodied cognition and communication: studies with British fine woodworkers. *Journal of the Royal Anthropological Institute*, 16:S100–S120.

Marshall, P., Rogers, Y., and Pantidi, N. (2011). Using F-formations to Analyse Spatial Patterns of Interaction in Physical Environments. In *Proceedings of CSCW 2011*, pages 445–454, Hangzhou, China.

McDowall, J. J. (1978). Interactional Synchrony: A Reappraisal. *Journal of Personality and Social Psychology*, 36(9):963–975.

McNeill, D. (1992). *Hand and mind: What Gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2000). *Language and Gesture*. Cambridge University Press.

McNeill, D. (2005). *Gesture & Thought*. University of Chicago Press.

McNeill, D., Cassel, J., and Levy, E. T. (1993). Abstract Deixis.

Mehrabian, A. (1972). *Nonverbal Communication*. Aldine Atherton.

Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, 66(2):B25–33.

Mondada, L. (2007). *Operating Together through Videoconference: Members' Procedures for Accomplishing a Common Space of Action*, pages 51–67. Ashgate.

Mondada, L. (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41(10):1977–1997.

Murphy, K. M. (2005). Collaborative imagining: The interactive use of gestures, talk, and graphic representation in architectural practise. *Semiotica*, 156:113–145.

Nijholt, A., Heylen, D., and Rienks, R. (2009). *Creating Social Technologies to Assist and Understand Social Interaction*, volume I, pages 416–428. IGI Global.

Olson, G. M. and Olson, J. S. (2000). Distance Matters. *Human-Computer Interaction*, 15:139–178.

Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In *Language and Gesture*. Cambridge University Press.

Özyürek, A. (2002). Do Speakers Design Ther Cospeech Gestures for Their Addressees? The Effects of Addressee Location on Represetational Gestures. *Journal of Memory and Language*, 46:688–704.

Purver, M., Griffiths, T. L., Körding, K. P., and Tenenbaum, J. B. (2006). Unsupervised topic modelling for multi-party spoken discourse. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06*, pages 17–24.

Richardson, D., Dale, R., and Spivey, M. (2007). *Eye movements in language and cognition: A brief introduction*.

Rienks, R., Poppe, R., and Heylen, D. (2010). Differences in head orientation behavior for speakers and listeners. *ACM Transactions on Applied Perception*, 7(1):1–13.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735.

Scheflen, A. E. (1964). The significance of Posture in Communication Systems. *Psychiatry*, 27:316–331.

Scheflen, A. E. (1976). *Human Territories: how we behave in space-time*. Prentice-Hall.

Schegloff, E. A. (1995). *Parties and Talking Together : Two Ways in Which Numbers Are Significant for Ta1k-in-Interaction*, pages 31–42. University Press of America.

Schegloff, E. A. (1998). Body Torque. *Social Research*, 65:535–596.

Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition*, 43(1):1–24.

Schober, M. and Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology*, 21:211–232.

Sellen, A. J. (1995). Remote Conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10:401–444.

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

Slater, M. and Usoh, M. (1994). *Body Centred Interaction in Immersive Virtual Environments*. John Wiley and Sons.

So, W. C., Kita, S., and Goldin-Meadow, S. (2009). Using the Hands to Identify Who Does What to Whom: Gesture and Speech Go Hand-in-Hand. *Cognitive science*, 33(1):115.

Sommer, R. (1965). Further studies of small group ecology. *Sociometry*, 28(4):337–348.

Streeck, J. (1993). Gesture as Communication I: Its Coordination with Gaze and Speech. *Communication Monographs*, 60(275-299).

Streeck, J. (1994). Gesture as Communication II: The Audience as Co-Author. *Research on Language and Social Interaction*, 27(3):239–267.

Sweetser, E. and Sizemore, M. (2008). *Personal and interpersonal gesture spaces: Functional contrasts in language and gesture*, pages 25–51. Mouton De Gruyter.

Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture*, 1:2:213–235.

Taylor, H. A. and Tversky, B. (1996). Perspective in Spatial Descriptions. *Journal of Memory and Language*, 35:371–391.

Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tur, D., Dowding, J., Favre, B., Fernandez, R., Frampton, M., Frandsen, M., Frederickson, C., Graciarena, M., Kintzing, D., Leveque, K., Mason, S., Niekrasz, J., Purver, M., Riedhammer, K., Shriberg, E., and Vergyri, D. (2010). The CALO Meeting Assistant System. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1601–1611.

van der Kleij, R., Maarten Schraagen, J., Werkhoven, P., and De Dreu, C. K. W. (2009). How Conversations Change Over Time in Face-to-Face and Video-Mediated Communication. *Small Group Research*, 40(4):355–381.

Vertegaal, R. (1997). Conversational awareness in multiparty VMC. *CHI '97 extended abstracts on Human factors in computing systems looking to the future - CHI '97*, page 6.

Vilhjálmsson, H. H. (2009). *Representing Communicative Function and Behavior in Multimodal Communication*, pages 47–59. Springer-Verlag Berlin Heidelberg.

Vine, I. (1975). Territoriality and the Spatial Regulation of Interaction. In Kendon, A., Harris, R. M., and Key, M. R., editors, *Organization of Behaviour in Face-to-Face Interaction*. Mouton.

Whittaker, S. (2003). Theories and Methods in Mediated Communication. In Graesser, A. C., Gernsbacher, M. A., and Goldman, S. R., editors, *Handbook of Discourse Processes*, pages 243–286. Lawrence Erlbaum Associates Inc.

Whittaker, S. and O'Conaill, B. (1993). An evaluation of video mediated communication. In *INTERACT '93 and CHI '93 conference companion on Human factors in computing systems - CHI '93*, pages 73–74, New York, New York, USA. ACM Press.

Wiemann, J. M. and Knapp, M. L. (1975). Turn-taking in Conversations. *Journal Of Communication*, 2:75–92.

Yngve, V. H. (1970). On Getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567 – 577.

Appendix A

Full tables for statistical analyses

A.1 Regression analyses

Table A.1 – Table 6.6 including PersonID

Variable	B	S.E.	EXP(B)	Sig
isSpeaking	0.253	0.004	1.288	0.000
Task Role (Learner)	-0.164	0.004	0.848	0.000
Hand Speed	0.126	0.001	1.134	0.000
Person(1)	-0.195	0.027	0.823	0.000
Person(2)	0.047	0.026	1.048	0.070
Person(3)	-0.394	0.026	0.674	0.000
Person(4)	-0.303	0.025	0.739	0.000
Person(5)	-0.269	0.025	0.764	0.000
Person(6)	-0.139	0.024	0.870	0.000
Person(7)	-0.271	0.024	0.763	0.000
Person(8)	-0.391	0.024	0.676	0.000
Person(9)	0.923	0.026	2.516	0.000
Person(10)	0.866	0.026	2.377	0.000
Person(11)	-1.200	0.040	0.301	0.000
Person(12)	-0.189	0.025	0.828	0.000

Person(13)	0.181	0.023	1.199	0.000
Person(14)	0.006	0.024	1.006	0.815
Person(15)	0.214	0.021	1.239	0.000
Person(16)	0.649	0.021	1.914	0.000
Person(17)	-0.439	0.022	0.645	0.000
Person(18)	0.312	0.022	1.366	0.000
Person(19)	0.789	0.022	2.201	0.000
Person(20)	-0.476	0.024	0.621	0.000
Person(21)	1.145	0.020	3.142	0.000
Person(22)	0.339	0.021	1.403	0.000
Person(23)	0.876	0.020	2.402	0.000
Person(24)	1.682	0.020	5.377	0.000
Person(25)	0.714	0.020	2.042	0.000
Person(26)	0.426	0.020	1.531	0.000
Person(27)	-0.287	0.021	0.750	0.000
Person(28)	-0.988	0.023	0.372	0.000
Person(29)	0.003	0.021	1.003	0.887
Person(30)	0.887	0.020	2.427	0.000
Person(31)	0.437	0.020	1.548	0.000
Person(32)	0.621	0.020	1.861	0.000

Table A.2 – Table 6.7 including PersonID

Variable	B	S.E.	EXP(B)	Sig
RecipientRole (Primary)	0.223	0.007	1.263	0.000
Hand Speed	0.155	0.001	1.167	0.000
Task Role (Learner)	-0.124	0.008	0.884	0.000
Person(1)	-0.299	0.037	0.741	0.000
Person(2)	0.190	0.037	1.209	0.000
Person(3)	-0.616	0.038	0.540	0.000

Person(4)	-0.140	0.036	0.869	0.000
Person(5)	-0.453	0.036	0.636	0.000
Person(6)	-0.485	0.035	0.615	0.000
Person(7)	-0.572	0.034	0.564	0.000
Person(8)	-0.651	0.035	0.522	0.000
Person(9)	0.528	0.036	1.696	0.000
Person(10)	0.394	0.037	1.483	0.000
Person(11)	-1.469	0.063	0.230	0.000
Person(12)	-0.695	0.036	0.499	0.000
Person(13)	-0.173	0.032	0.841	0.000
Person(14)	0.054	0.035	1.056	0.117
Person(15)	0.050	0.029	1.051	0.083
Person(16)	0.507	0.028	1.660	0.000
Person(17)	-0.558	0.032	0.572	0.000
Person(18)	0.204	0.030	1.227	0.000
Person(19)	0.630	0.029	1.878	0.000
Person(20)	-0.624	0.032	0.536	0.000
Person(21)	0.524	0.028	1.689	0.000
Person(22)	0.007	0.028	1.007	0.814
Person(23)	0.369	0.027	1.447	0.000
Person(24)	1.222	0.028	3.393	0.000
Person(25)	0.599	0.027	1.821	0.000
Person(26)	0.105	0.028	1.111	0.000
Person(27)	-0.334	0.029	0.689	0.000
Person(28)	-1.826	0.037	0.161	0.000
Person(29)	-0.372	0.029	0.689	0.000
Person(30)	0.670	0.027	1.954	0.000
Person(31)	0.051	0.028	1.053	0.067
Person(32)	0.400	0.028	1.492	0.000

Table A.3 – Table 6.8 including PersonID

Variable	B	S.E.	EXP(B)	Sig
isSpeaking (Yes)	0.725	0.006	2.064	0.000
Hand Speed	0.106	0.001	1.111	0.000
Person(1)	1.219	0.032	3.384	0.000
Person(2)	0.710	0.035	2.034	0.000
Person(3)	0.325	0.036	1.384	0.000
Person(4)	-0.005	0.039	0.995	0.895
Person(5)	-0.249	0.040	0.780	0.000
Person(6)	0.830	0.033	2.292	0.000
Person(7)	0.740	0.035	2.096	0.000
Person(8)	1.223	0.032	3.398	0.000
Person(9)	0.498	0.035	1.645	0.000
Person(10)	1.024	0.033	2.784	0.000
Person(11)	0.532	0.035	1.702	0.000
Person(12)	1.380	0.031	3.974	0.000
Person(13)	1.930	0.030	6.892	0.000
Person(14)	-0.271	0.040	0.763	0.000
Person(15)	1.123	0.031	3.073	0.000
Person(16)	1.154	0.031	3.171	0.000
Person(17)	-0.512	0.044	0.599	0.000
Person(18)	1.297	0.030	3.659	0.000
Person(19)	-0.790	0.048	0.454	0.000
Person(20)	1.561	0.030	4.764	0.000
Person(21)	0.651	0.034	1.918	0.000
Person(22)	1.520	0.030	4.572	0.000
Person(23)	0.603	0.033	1.828	0.000
Person(24)	0.192	0.036	1.211	0.000
Person(25)	1.508	0.030	4.518	0.000
Person(26)	2.501	0.030	12.199	0.000
Person(27)	0.680	0.032	1.974	0.000

Person(28)	0.308	0.048	1.361	0.000
Person(29)	2.054	0.030	7.802	0.000
Person(30)	0.620	0.035	1.858	0.000
Person(31)	1.020	0.033	2.773	0.000
Person(32)	-3.257	0.174	0.039	0.000
Person(33)	1.296	0.034	3.653	0.000
Person(34)	0.875	0.037	2.400	0.000
Person(35)	1.623	0.033	5.067	0.000
Person(36)	0.434	0.034	1.543	0.000
Person(37)	0.013	0.039	1.013	0.746
Person(38)	0.700	0.033	2.013	0.000
Person(39)	1.234	0.031	3.434	0.000
Person(40)	0.964	0.032	2.508	0.000
Person(41)	0.584	0.034	1.793	0.000
Person(42)	0.794	0.033	2.213	0.000
Person(43)	0.920	0.032	2.508	0.000
Person(44)	1.238	0.031	3.449	0.000

Table A.4 – Table 6.9 including PersonID

Variable	B	S.E.	EXP(B)	Sig
Hand Speed	0.164	0.002	1.178	0.000
RecipientRole (Primary)	0.061	0.013	1.063	0.000
Person(1)	0.824	0.073	2.279	0.000
Person(2)	0.590	0.075	1.804	0.000
Person(3)	0.579	0.077	1.784	0.000
Person(4)	0.695	0.078	2.003	0.000
Person(5)	0.411	0.078	1.509	0.000
Person(6)	1.180	0.073	3.253	0.000
Person(7)	1.030	0.076	2.802	0.000

Person(8)	1.043	0.074	2.839	0.000
Person(9)	0.019	0.088	1.019	0.830
Person(10)	0.880	0.073	2.412	0.000
Person(11)	0.618	0.076	1.855	0.000
Person(12)	1.765	0.069	5.844	0.000
Person(13)	2.348	0.071	10.466	0.000
Person(14)	-0.381	0.087	0.683	0.000
Person(15)	1.418	0.071	4.130	0.000
Person(16)	0.729	0.078	2.073	0.000
Person(17)	-0.466	0.089	0.628	0.000
Person(18)	1.934	0.070	6.920	0.000
Person(19)	-0.799	0.091	0.450	0.000
Person(20)	1.646	0.070	5.187	0.000
Person(21)	0.868	0.073	2.383	0.000
Person(22)	1.701	0.070	5.482	0.000
Person(23)	0.846	0.078	2.331	0.000
Person(24)	0.540	0.074	1.715	0.000
Person(25)	1.779	0.069	5.922	0.000
Person(26)	4.255	0.069	70.430	0.000
Person(27)	0.736	0.078	2.087	0.000
Person(28)	0.374	0.088	1.454	0.000
Person(29)	2.549	0.068	12.799	0.000
Person(30)	0.637	0.076	1.891	0.000
Person(31)	1.705	0.074	5.503	0.000
Person(32)	-3.002	0.296	0.050	0.000
Person(33)	1.460	0.079	4.305	0.000
Person(34)	1.677	0.075	5.347	0.000
Person(35)	1.936	0.075	6.930	0.000
Person(36)	-0.509	0.100	0.601	0.000
Person(37)	0.155	0.082	1.168	0.058
Person(38)	0.197	0.082	1.218	0.016

Person(39)	1.019	0.071	2.771	0.000
Person(40)	1.468	0.073	4.342	0.000
Person(41)	0.228	0.078	1.256	0.003
Person(42)	0.734	0.074	2.084	0.000
Person(43)	0.542	0.078	1.720	0.000
Person(44)	1.340	0.071	3.817	0.000

Table A.5 – Table 6.10 including PersonID

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.485	0.006	0.198	0.000
isSpeaking (Yes)	1.232	0.006	0.167	0.000
Task Role (Learner)	-0.215	0.005	-0.031	0.000
PersonID	0.002	0.000	0.005	0.000

Table A.6 – Table 6.11 including PersonID

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.266	0.007	0.201	0.000
Task Role (Learner)	-0.256	0.007	-0.048	0.000
RecipientRole (Primary)	0.169	0.007	0.033	0.000
PersonID	0.009	0.000	0.030	0.000

Table A.7 – Table 6.12 including PersonID

Variable	B	S.E.	Standardized B	Sig
isSpeaking (Yes)	1.412	0.006	0.232	0.000

isNodding (Yes)	1.229	0.007	0.169	0.000
PersonID	0.005	0.000	0.023	0.000

Table A.8 – Table 6.13 including PersonID

Variable	B	S.E.	Standardized B	Sig
isNodding (Yes)	1.044	0.009	0.175	0.000
RecipientRole (Primary)	0.069	0.005	0.019	0.000
PersonID	0.002	0.000	0.016	0.000

Table A.9 – Table 6.14 including PersonID

Variable	B	S.E.	EXP(B)	Sig
Task Role (Learner)	2.684	0.007	14.640	0.000
isNodding (Yes)	0.217	0.007	1.242	0.000
Hand Speed	0.013	0.001	1.013	0.000
isSpeaking (Yes)	*Removed as constant*			
Person(1)	0.921	0.031	2.513	0.000
Person(2)	0.558	0.035	1.747	0.000
Person(3)	1.075	0.033	2.930	0.000
Person(4)	0.915	0.030	2.497	0.000
Person(5)	0.174	0.031	1.190	0.000
Person(6)	0.636	0.030	1.889	0.000
Person(7)	1.070	0.028	2.916	0.000
Person(8)	0.358	0.029	1.430	0.000
Person(9)	1.594	0.035	4.924	0.000
Person(10)	1.539	0.035	3.893	0.000
Person(11)	-0.535	0.039	0.586	0.000
Person(12)	1.798	0.029	6.038	0.000

Person(13)	0.223	0.029	1.250	0.000
Person(14)	-0.129	0.033	0.879	0.000
Person(15)	-0.674	0.027	0.510	0.000
Person(16)	1.383	0.026	3.985	0.000
Person(17)	1.252	0.028	3.497	0.000
Person(18)	2.459	0.029	11.692	0.000
Person(19)	-0.723	0.029	0.485	0.000
Person(20)	0.284	0.027	1.329	0.000
Person(21)	0.328	0.027	1.388	0.000
Person(22)	0.374	0.025	1.453	0.000
Person(23)	0.858	0.026	2.357	0.000
Person(24)	0.774	0.027	2.166	0.000
Person(25)	0.954	0.026	2.596	0.000
Person(26)	0.767	0.026	2.153	0.000
Person(27)	1.764	0.026	5.838	0.000
Person(28)	-0.530	0.026	0.589	0.000
Person(29)	1.005	0.026	2.731	0.000
Person(30)	2.607	0.029	13.557	0.000
Person(31)	0.780	0.025	2.180	0.000
Person(32)	0.164	0.027	1.013	0.000

Appendix B

Simultaneous Engagement

B.1 Transcripts for simultaneous engagement events

- Head Moves

  Black Cap (Instructor): ok now part of the states assembly are two err things we have the executive

  Head: Oriented towards White Cap (learner)

  Gesture: Towards White Cap

  (0.4)

  Black Cap (Instructor): is that right

  Head: Turns towards Blue Cap (instructor)

  Gesture: Remains towards White Cap

  Blue Cap (Instructor): executive

  Head: Turns to towards Black cap and nods

- Hand Moves

  Blue Cap (Instructor): the l-lower hierarchy classes like <u>the masters</u> and [undergraduate]

  Head: Oriented towards Black Cap (learner)

  Gesture: Left hand gestures palm up towards White Cap (instructor)

White Cap (Instructor): [yeah]

Head: Nods

- Both Move

Blue Cap (Instructor): and there's ministers who vote on policies for

Head: Oriented towards White Cap (instructor)

Gesture: Gesturing with right hand

(0.3)

White Cap (Instructor): and the king chooses the ministers

Head: Orients to Black Cap then turns to Blue Cap

Gesture: Points with right hand towards Black Cap then turns to Blue Cap

Blue Cap (Instructor): yeh king chooses the ministers

Head: Turns to Black Cap

Gesture: Gestures with right hand

B.2 Coding scheme decision tree

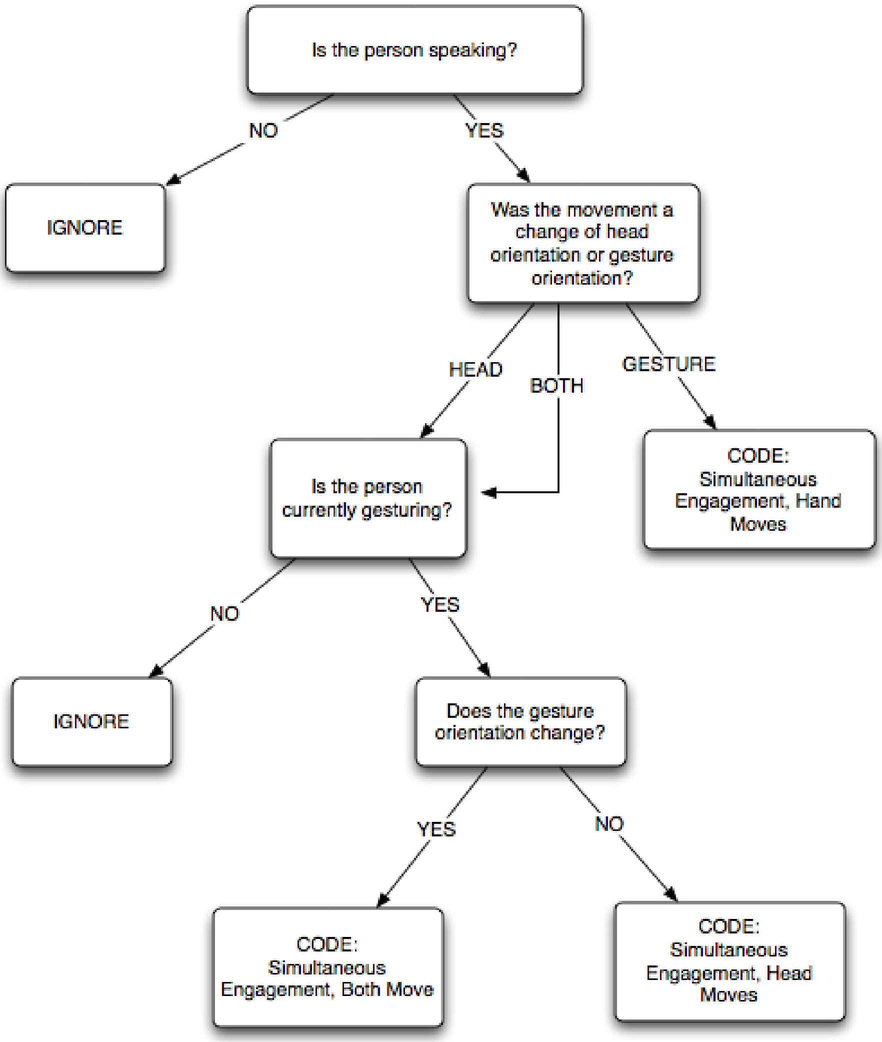

Is the person speaking?
NO
YES
IGNORE
Was the movement a change of head orientation or gesture orientation?
HEAD
BOTH
GESTURE
Is the person currently gesturing?
CODE: Simultaneous Engagement, Hand Moves
NO
YES
IGNORE
Does the gesture orientation change?
YES
NO
CODE: Simultaneous Engagement, Both Move
CODE: Simultaneous Engagement, Head Moves

Figure B.1 – Coding scheme decision tree for Simultaneous Engagement events

Appendix C

Discussion

C.1 Hand movement clustering

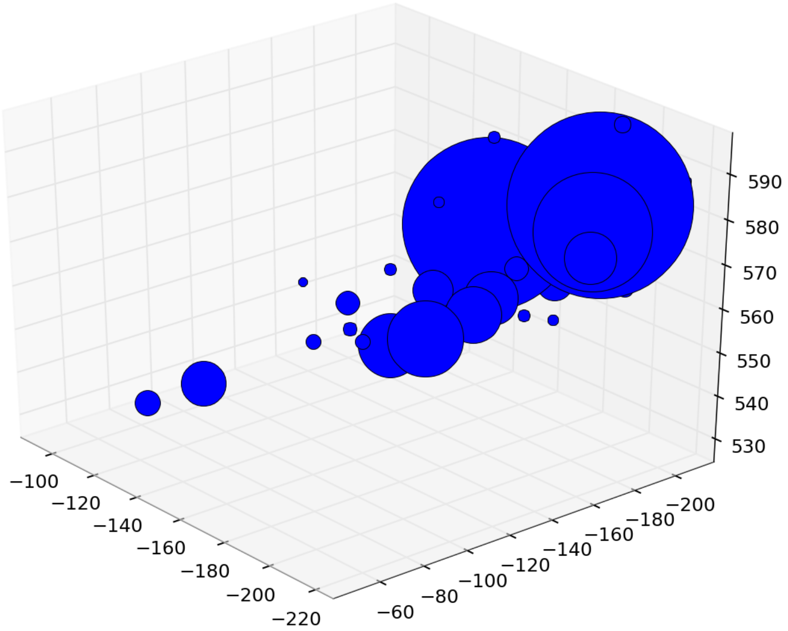


Figure C.1 – An work in progress graph showing clusters of hand movement. Larger circles represent more frequently visited areas.