Understanding Interactive Behaviour: A Quantitative Approach

Leon Adam Watts

Thesis submitted for the degree of
Doctor of Philosophy

The University of York

Department of Psychology

September 1998

Abstract

This thesis develops an approach to the analysis of behaviour based both on studies of video-mediated communication and current human-computer interaction literature. Human and primate communication studies have shown compelling evidence of the importance of seeing and being seen. Paradoxically, experimental studies have consistently reported little or no benefit for adding a visual channel to audio communications links. Through an analysis of these studies this paradox can be cast as a problem of understanding interactive behaviour. It is argued that interactive behaviour involves far more than can be accounted for by a generalized communication model. Thus video-mediated communication would be better conceived of as video-mediated interaction (VMI). Activity Set Analysis is proposed, building on a whole-interaction model as an analytic unit and focusing on the contemporaneity of activity-system behaviours or states. A software tool, Action Recorder, is described as an implementation of the Activity Set approach to observation. Possible treatments of the resultant observational data are considered and a new statistic, 'S', is derived for comparing the empirical co-occurrence of two behaviours in time against an estimate of their chance level of coincidence. Two dyadic VMI studies are then reported. The first examines gaze and speech behaviour patterns in VMI as a paradigmatic application of Activity Set Analysis. The second goes on to examine similar behaviour in an experiment contrasting three configurations of video link. It pays particular attention to the contextualizing effects of both experimental task and differentiated role in task. Video image characteristics are found to have pronounced effects, but strictly as a function of these contextualizing factors. This thesis not only sheds some light on the VMI paradox but further suggests that Activity Set Analysis has something to offer for students of other kinds of interactive behaviour.

List of Contents

List of Tables

List of Illustrations

Preface: A general approach to interactive behaviour, motivated within a particular investigative context

The work reported in this thesis is a conceptual and analytical approach to interactive behaviour, called Activity Set Analysis. It was developed as a means for addressing a particular question within a particular investigative context. The question was: "does audio-video mediation of interactive behaviour differ from that mediated by audio-only communication facilities". The investigative context was Human-Computer Interaction (HCI) research. This thesis is, however, presented as a work of psychology and it is thus important to indicate the contribution it has to make in that respect; the kind of psychology it purports to represent. It should be noted that, whilst this question and research context frame the work reported here, it is not the purpose of this thesis to provide a definitive answer to the question or to contribute some new knowledge to Human-Computer Interaction *per se*. Activity Set Analysis is proposed as an analytic method with general applicability for a wide range of questions in a wide range of investigative contexts.

The work described is within an empirical tradition of laboratory experimentation, relying on the operationalization of concepts for statistical measurement. The measurements reported refer to *interactive behaviour;* that is, behaviour closely coupled with and reactive to the emerging opportunities for and resources available to mediate action. Whilst fundamentally psychological in its approach and execution, this thesis has as its central theme the investigation of the interaction between people and their functional environment. Studying and arriving at an understanding of interaction in this sense sets HCI apart from academic forms of psychology in several ways, not least of which is the centrality of considerations for the design of new technologies. Psychology has been defined many times, often in conflicting ways. However, to clarify the position of this thesis as a work of psychology, the following definitions are offered. Academic psychology is concerned with building a systematic and general understanding of the components, processes and practices by which animals regulate the various forms of their behaviour. Applied psychology attempts to put existing psychological theory to use for some practical purpose, often evolving unique approaches to make possible such application.

HCI is closely related to the rather better known ergonomics as a form of applied psychology (Shackel, 1996). However, it would be as inaccurate to describe HCI as a part of ergonomics as it would be to describe ergonomics as a part of psychology. Both ergonomics and HCI are multidisciplinary fields. The contributing set of disciplines for each are distinct. Physiology and anthropometrics are vitally important to ergonomics but of passing interest in HCI. Anthropology has a demonstrated role in HCI and yet has little contact with ergonomics. The common thread in ergonomics and HCI that sets them apart from academic psychology is a central concern for how people act in the world with the purpose of the better design of artefacts there encountered. An account of extrinsic factors is of equal importance to the intrinsic capacities of the human organism, the latter being the predominant concern of academic psychology. Thus any investigative technique of use to HCI must be able to lend itself to extrinsic as well as intrinsic analytic dimensions. It is the study of *interactive behaviour*, not individual behaviour, that is of focal importance. Another major distinction to draw between academic psychology and psychology in HCI concerns human adaptability. Often, academic psychology focuses on intrinsic processes considered to be effectively invariant properties of the organism, such as psychophysical reactions or text interpretation processes, and environmental factors are controlled out as confounds. Where environmental factors are admitted, it is in terms of learning, so that they have a cumulative effect as a function of frequency of exposure. Understanding interactive behaviour requires that properties of the organism and the potential of its setting are brought together to interpret action as moment-by-moment adaptation to available opportunities. Settings may be conceived in a number of ways, incorporating place, tool and other people. Person-setting relationships are very likely *vital* and *sensitive*, to the extent that conceptual distinctions blur.

HCI must be able to deal with such dynamics as time dimensions which condition the whole interactive behavioural system. Any investigative technique must be sensitive to effects that may be extremely subtle in relation to combinations of factors making up an activity context. In a limited-term laboratory experiment these may be highly elusive and yet of some significance when translated to the everyday grind of real work. Activity Set

Analysis has been developed with these considerations in mind; taking into account extrinsic factors and sensitivity to the contextualization of interactive behaviour in time.

Acknowledgements

For any work evolving over a six year period, there is inevitably a very long list of those people whose efforts and encouragement have helped to make it what it is. My ideas didn't get to the stage of forming a thesis all by themselves; the effort required to produce this work has been supported by far more than academic advice.

Owen Daly-Jones contributed enormously to the early development of the work I report and to the ideas behind it. He was a vital part of the approach to studying video-mediated interaction developed here at York University. I am greatly in his debt. I'd like to thank Simon Fletcher and Garry Turner for all the technical support and uncomplaining emergency rewiring they've done for me over the years. Technical skills of a different kind were needed to work out an appropriate model for Synchronization Analysis. I am very grateful to Dr Nick Merriam for some useful extended discussion of this matter, for convincing me I wasn't mad or stupid and for providing some very timely advice. Dr Peter Lee, of York University's Mathematics Department, gave me a lot of his time in the final stages of this work. He further reassured me that I wasn't merely seeing shapes in the fog, taking the trouble to understand the parameters of the problem and eventually formulating a workable solution. Also more than worthy of special mention are Dr Peter Davies and Dr Margaret Riseborough, who inspired me to a research career during my studies at Bradford University; the inspirational Jarnail Chudge, MSc, the ever-buoyant Ian Hearnden, MA; the ever-questing Dr Francesmary Modugno and the cheerfully indefatigable Sarah Edwards, MA. Besides the help they have freely offered, I am proud to say that I consider all those named as friends. I have also benefited enormously from an institution known as the Basic Research Symposium, a companion event to the annual ACM CHI conference. Over the years, some extraordinary people have given me the benefit of their wisdom at the BRS, both on specific aspects of my work and also on the general problems motivating it.

I'd also like to acknowledge some of the wonderful people who kept Su and I alive and sane during our epic motorcycle journey through Europe, Africa and North America. Moulay Idris Dowdowa rescued us from an uncertain fate in the sands between Reggane and In Salah, Algeria. Harry Diehl, Connie and Terry Hughes prevented us from being interned or worse during the Iraqi war, Connie and Terry later becoming very special friends to us.

Above all, there are four people without whom it is difficult to imagine this work could ever have been completed: Dr Andrew Monk, Su Watts, MA, Robert John Watts and Patricia Anne Watts, each of whom deserve more than a little space here. Andrew has been a mentor and a friend, beneath whose façade of bumbling amiability lies an intellect of surprising clarity and insight. He has successfully struggled to keep my ideas this side of the unworkable, showing me that statistics are a tool not an answer, and enthusing me for a psychology as a science sternly disciplined by pragmatics. His contribution to the work here is incalculable for the personal and selfless intellectual support he has given.

Su has put up with my ravings for no less than fourteen years. She has consistently demonstrated how boundless is her patience, tolerating her hermit-husband's inconveniences with a Stoicism that would have made Zeno wonder why he bothered. Whilst I grumbled and cursed my way from paragraph to paragraph, Su quietly completed a Masters degree in European Studies at the same time as bearing our son, Rory, earning a living and learning to be a master chef. As if that wasn't enough, she is a skilled copy editor and carefully proof-read the whole of my thesis to the highest of standards. But these are merely details: we have shared much in this time that would have meant nothing, were it not for one another. Indeed, my cup runneth over.

Robert John Watts, my father, and Patricia Anne Watts, my mother, have given unquestioning encouragement and support to me for as long as I can remember. They have held education in high regard from my earliest memories, especially my father. He could not have had even the glimpse of a university education. He was of the wrong class, at the wrong time. He left school for a factory job at the age of fourteen. Thereafter, he took it upon himself to learn, to make a future, building a better horizon for himself and later his wife and two sons. His attitude towards scholarly activity in an age of unchecked materialism is as unusual as it is refreshing. He has often said to me, "Give of your best, boy, nobody can ask for more." It has been a constant reminder that performance and achievement are defined in relative terms, temporing the value of any result by the conduct leading to it. I thank you for that, Dad.

Finally, I dedicate this thesis in three ways: to whosoever has the courage to risk derision by volunteering that they don't understand, in order that they might; to all those who have made this possible, and especially to my son, Rory William Alexander Watts because he's teaching me all over again.

Author's Declaration

Several people have contributed to the work reported in this thesis, most notably Dr Andrew F. Monk and Owen Daly-Jones. It remains substantially and primarily the author's work, in its conception, development and execution. Various parts of this work have been published prior to the preparation of this thesis, including:

aspects of Chapters One, Three and Four in:

> Monk, A. F., McCarthy, J. C., Watts, L. A., & Daly-Jones, O. (1996). Measures of process. In P. Thomas (Eds.), *CSCW Requirements and Evaluation* (pp. 125-139). Berlin: Springer Verlag.

a short description of Action Recorder, presented in Chapter Three, some data from Chapter Four and an early version of Chapter Five's Synchronization arguments:

> Watts, L. A., Monk, A. F., & Daly-Jones, O. (1996). Inter-personal awareness and synchronization: assessing the value of communication technologies. *International Journal of Human Computer Studies*, 44(6), 849 - 875.

some elements of Chapter One and other empirical data from the study reported in Chapter Four:

> Daly-Jones, O., Monk, A. F., & Watts, L. A. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*, 49(1), 21 - 59.

early findings from the experiment reported in Chapter Seven, using a primitive form of the *S* statistic reported in Chapter Five, were published in:

> Monk, A. F., & Watts, L. A. (1995). A poor quality video link affects speech but not gaze. In *CHI'95 Conference on Human Factors in Computing Systems*, Conference Companion (pp. 274 - 275). Denver, Colorado. USA: ACM Press.

Overview

Each chapter is summarized here as a general introduction to this thesis and as a more extensive precis of the work as a whole.

Chapter One. A review of research on interactive behaviour mediated by video communications technologies

The idea of augmenting remote communication devices with video is presented as a paradox: it retains many adherents and attracts considerable investment, in spite of a paucity of evidence to suggest that it is efficacious for communication and the high costs involved in its deployment. On the one hand, scientific investigations of video-mediated interaction have concluded that there is little if any value added by video to communication mediated only by audio technologies. On the other, ethological work on gaze patterns in primate societies and communication studies of gaze patterns in conversation indicate that looking behaviour matters. Seeing others and directing ones visual attention both manages the dynamics of the ongoing conversation and carries something of the social structures within which interactions are embedded. Where people are concerned, looking behaviour is critically conditioned by whether or not speech, the primary engine of human communication, is involved. The apparent contradiction between findings from technologically mediated and face-to-face co-present interaction is taken as good grounds for reconceiving the matter as one of the conception and measurement of interactive behaviour.

Measures of video-mediated communication have tended to exploit a limited conception of performance by concentrating on task-outcome measures. It is argued that, before dismissing video links as uninformative to mediated interaction, some attempt should be made to understand the relationship between gaze and speech behaviour for the required task. Evidence of gaze-related behavioural sensitivities between users of a video link would imply that there is indeed an influence on or of looking behaviour in that communicative context. As gaze and speech vary on a moment-by-moment basis, an approach is required that casts these as mutually responsive behaviours taking place in time. Where interaction behaviours of this kind have been recorded, it has often been as counts. Treating data in this way assumes that the value of each behaviour is discrete and independent of other matters that may have obtained at the time the measured behaviour was displayed. Behaviour so assessed is to do with individual productions, a matter of emission rather than interaction, and is all about the consequences of whatever took place. Deeper conceptions of interaction dynamics have rarely been addressed and little or no account has been taken of the relevance of the provision of a video link for the experimental task to be performed. Video-mediated interaction (VMI) is proposed as a better description of what goes on, based on a broader conception of interactive behaviour.Knowing the result or outcome of a set task implies something about the facility with which the aggregate of available resources, mental and physical, were put to work. It says little about the marshalling of those resources in reaching the outcome, information that may be of crucial importance in application. An examination of such effort and its organization may help to reveal whether in fact the available resources were recruited in carrying out the task concerned.

Chapter Two. Activity Set Analysis: building on state-based temporal characteristics to examine interactive behaviour

Some of the conceptions of action and behaviour evident in the HCI literature are examined in relation to the discussion in Chapter One. An account of interactive behaviour is further developed, with special reference to activity theory and Hutchins' Distributed Cognition account of collective action. Both emphasize how the action of an individual is only meaningful within some kind of context including available resources, to the point where any analysis must explicitly account for these as framing conditions of action. For activity theory, context is crucially related to the tools available for work and to pursue an 'object'. Distributed Cognition asserts that an appropriate unit of analysis for collaborative work comprises, as a minimum meaningful context, the people and artefacts through which coordination is routinely organized.

Identifying relationships between behaviours contributing to the outcome of an experimental task provides an opportunity to expose the important parameters of performance as a whole. These might be taken as evidence of the level of use to which the various resources are put. Furthermore, following the behaviours against variable external conditions as they apply allows something to be said about these as contextual effects. Activity Set Analysis is a general approach devised to examine performance as something occurring over a period of time rather than as an endpoint or final result. It takes its name from looking at combinations of intrinsic and extrinsic conditions as a number of possible transient states, made up of a number of potentially simultaneous activities involving constellations of factors. Each individual activity is defined as a member of a set, $A_1, A_2 \ldots A_n$, where membership is based on a principle of mutual exclusion in time. The occurrence of Activity Set member behaviour, A_1, entails that no other member behaviour of the same Set, *A*, could obtain at that time. In this way, the state of individual and context, made up perhaps of other individuals, tools and environmental variables, is defined by the activity conditions, *1* to *n*, across a collection of Activity Sets, *A* to Z. Activity Sets can describe one or more individuals and the environmental conditions, where appropriate, in which they are operating. Activity Sets would be defined by an investigator to include a range of behaviours of relevance to the research question at issue.

Activity Set Analysis offers two kinds of insight for interactive behaviour. The first is a simple description of the amount of time and frequency with which particular behavioural or environmental states occurred for any of the defined Activity Sets. Statements may thus be made about the relative investment in particular activities on behalf of the actors or the variability of environmental conditions. The second is to expose the contingency of Activity engagement as a function of the conditions given by other Activity Sets over a period of performance.

Chapter 3. Implementing Activity Set Analysis: the Action Recorder software tool

In this chapter, the principles of Activity Set Analysis are applied to the design of a computer logging tool, Action Recorder. Action Recorder is a computer programme formalizing Activity Set Analysis by structuring target behaviours into Activity Sets and automating data organization into the relevant classes through the operation of parsing rules. It thereby prepares data in a form appropriate for statistical analysis of the kind described in Chapter Two. Action Recorder's design and operation are described.

Chapter 4. Experimental case study, Part One: Activity Set Analysis as a methodological approach to video-mediated interaction

In this chapter, video recordings of video-mediated interaction were coded with Action Recorder and subsequently analysed as a paradigmatic demonstration of the Activity Set Analysis. Three basic temporal metrics of Activity are proposed. Inter-observer reliability and statistical validity of these data is demonstrated, the latter by comparing a definitive data set with a data set generated with Action Recorder. Simple statistics, that is for one Activity Set at a time, are reported together with some preliminary data and discussion of contingent statistics, referring to potential Co-activity states. Co-activity states are those comprizing conjunctions of defined Activities. Whereas the former kind of information cannot distinguish between essentially random and organized activity, Co-activity data permit determination of any systematic variation of one Activity against the occurrence of Activities from other Activity Sets.

Chapter 5. Calculating the temporal organization of Co-activity: Synchronization Analysis and the 'S' statistic

Whilst conventional statistical procedures allow a satisfactory analysis of Simple Activity Set data, Co-activity data present conceptual and mathematical difficulties. Duration and frequency data for Co-activities suggest a sense of purpose about Co-activity states that belies their origin, as mathematical derivatives of whole-Activity ratings. Proportional data, unless subject to some transformation, suggest a degree of contingency that may be no more than a mathematical artefact. Proportional data are seen as reflecting an overall investment in Activity, rather than referring to specific instances of Activity, and thus as conceptually the most interesting measure for understanding organization in Co-activity. Some analytic techniques are considered and used to formulate a new approach to Co-activity proportional analysis within the Activity Set approach. Synchronization is presented as a basic estimate of the degree to which the co-occurrence of Activities departs from independence and 'S' translates this estimate into a probabilistic account of Co-activity temporal organization.

Chapter 6. Experimental case study, Part Two: applying Synchronization Analysis to VMI Co-activity data

The statistics developed in Chapter Five are applied to Co-activity standardized time data from the experimental case study. Evidence of organization in Co-activities from between-subjects' Speech and Gaze Activity Sets is found but not for within-subjects' Speech and Gaze states. That there is a gaze relationship between speaker and listener is suggestive of gaze distribution being an intrinsic part of their activity system. Activity Set Analysis has revealed a behavioural contingency improbable without video mediation. It is further taken as evidence that Activity Set Analysis is a viable approach to interactive behaviour.

Chapter 7. Speech and gaze distribution with three configurations of video display and two experimental tasks

Having demonstrated the application of Activity Set Analysis to video-mediated interaction, an experiment is reported in which three different kinds of video image of conversational partners were compared. These were of a relatively large full-motion image, a relatively small full-motion image and a large still image. Few video-mediated interaction studies take into account the effect of the task to be performed on the value of available resources. Typically, tasks are described as 'problem solving', or perhaps as 'negotiation', without indication of how video mediation might lend itself to one kind more or less than another. Two task-phases were included in the experimental design: a general getting-to-know-you discussion was followed by a playing card guessing game. The latter was designed to emphasize the visible aspects of one of the pair. Speech patterns were found to differ between the conditions, though the audio quality was not varied during the experiment, implying a differential sensitivity to the kind of video image on offer in conversation. This is taken as evidence that the video images changed what was required of speech hence implying that: (a) some work was done via the video link; and (b) the amount of work done by the visual component is robust against dramatic differences in the size of the image provided. Activity Set Analysis permits the determination of a relationship between any behaviour and one or more other co-occurring conditions. Chapter Four reported such co-occurring conditions as other kinds of behaviour. Here, a co-occurring condition is taken as the phase of the experimental task. A strong effect of task-phase on gaze and speech patterns was found. Given the artificial nature of the task, no strong conclusions are drawn here but the general notion of task-sensitivity for resource exploitation is demonstrated.

Chapter 8: Discussion and conclusions

The thesis ends with a summary of the Activity Set Analysis approach to interactive behaviour. Consideration is given to how well video-mediated interaction can be assessed by observations of time spent in Activity. The thesis is contends that Co-activity Synchronization demonstrates not merely epiphenomenal but systematic relationships amongst the elements of interactive behaviour. Problems of interpretation and limitations to the meaningfulness of relationships found between the gaze of one person and the speech of their remote counterpart are discussed. Activity Set Analysis is further discussed in a wider context, with time sequence analysis, as a general-purpose analytic method for temporal dimensions of interactive behaviour.

1 Chasing rainbows in fast cars: behaviour, measurement and interaction research

Chapter summary

Research into video-mediated communication is described as a matter of understanding interactive behaviour. The addition of a video image to enhance communications with a remote party has been keenly anticipated for many years. It has been assumed that such an addition would put at least some of the 'gold' in the experience of co-present face-to-face interaction into remote communication. This chapter addresses the substance of this assumption and the way in which this gold has been prospected. A broad consideration of the scientific evidence shows that the expected benefits of video-based technologies over audio-only technologies have been surprisingly illusive. Motives for the technological conveyance of interactive visual behaviours are reconsidered by briefly reviewing evidence for the role of gaze from non-technological studies of interaction. Finally, the approaches adopted by investigators of video-mediated interaction are contrasted, to differentiate bluntly insensitive paradigms from those with promise. Where measures have discriminated audio-video from audio-only mediation, they have tended to focus on the process of interaction. A process analysis of gaze, in combination with some indication of speech activity, should indicate whether or not a video link is exploited in video-mediated interactions. Within the psychological tradition of behaviour studies, a quantitative analysis of empirical data is proposed to complement insights gained from a variety of qualitative approaches to the subject. In order to operationalize such an analysis, a closer consideration of the nature of interactive behaviour is required, an undertaking reserved for the next chapter.

1.1 Envisioning video links: interactive behaviour in a context of technological mediation

> "He said that what makes them alive (...) is their gaze, is their glance. And so that he looking at you and your looking at him was very very important - it meant that you weren't an object. He wasn't at all interested in what's called depicting character in his portraits but he was very interested in the fact that you were a human being sitting there and that there was a human rapport. "
>
> David Sylvester on Giacometti, portrait painter and sculptor, friend of Jean-Paul Sartre [1].

This thesis has the twin aims of carrying some ideas from the interdisciplinary fields of Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW) into a quantitative approach to interactive behaviour and of adding to the debate on video-mediated communication (VMC). The notion of interactive behaviour is intriguing from the viewpoint of experimental psychology. Within its established investigative paradigms, behaviour is very often conceptualized as a discrete matter. Interactive behaviour does not readily conform to such a model. Any particular behaviour is only meaningful with reference to the setting of its production. The notion of interactivity requires that there must be at least two poles between which to coordinate action and against which to judge action. Any individual's behaviour therefore cannot be taken as discrete action: action in this interactive sense requires specification beyond a focal individual to include one or more other entities with or upon which the individual is acting.

Video-mediated communication research is interesting as a matter of interactive behaviour for several reasons. First, it has been subject to some notable psychological research investment, as shall be discussed, without any great reward. Secondly, as a mediator of action, the technologies of communication are not obviously in quite the same relation to their "users" as, say, a pointing device or presentation technique for an object structure. The first-order interactivity of a video link is between those who are in communication with one another. Whereas single-user computer systems involve the device in some sense as a partner in the business of getting some job done, communication technologies are simplistically cast only in the role of conduit for action sourced and directed independently of them. The tool enabling connection is easily marginalized as passive, its proper functioning

[1] In an interview for BBC Radio 4's "Kaleidoscope" programme, broadcast 9th August, 1996.

identified with invisibility to those who use it. However, its invisibility is easily confused with an absence of influence. The devices mediating that communication potentially affect the process of joint action, at least in so far as they constrain or amplify its various aspects. Video-mediated communication, then, provides this thesis with a concrete focus, as a theoretical and applied treatment of interactive behaviour.

In all conversation, one must carry an awareness about the state of one's audience in order that contributions to the joint projects of communication be timely and appropriate (Clark & Schaefer, 1989). Equally, one's audience must be attentive to some degree in order to successfully negotiate the perils of turn taking, described so thoroughly by Sacks, Schegloff and Jefferson (1974), and to build "common ground" in Herbert Clark's terms, as partners in the business of conversation (Clark, 1996).

Seeing and being seen by one's interlocutors involves a special kind of interaction. These activities together bring about an instantaneous mutual awareness. As Sylvester describes above, there is something special about this reciprocal awareness brokered by gaze. One's own identity as a social agent comes into focus, alongside that of the person subject to gaze. To repeat the above: "so that he looking at you and your looking at him was very very important - it meant that you weren't an object".

Audition differs from vision in a way that makes such mutual awareness rather more ephemeral and, should it be sought, demanding. Just as so many aspects of the visual environment pass beneath one's notice, many sounds are ignored as a matter of course. However, to hear something one has no need to turn towards the sound source, a necessary condition for seeing. So, whereas attended sounds are not given evidence by the orientation of the listener, one cannot but conclude that one's self has come to the notice of another person when that other's gaze alights on the self corporeal. When orientations coincide, those looking become aware, in a new way, of their status both as the subject and the object of gaze. Together, those who make such contact are somehow bound by the act as personalities rather than merely as other objects of vision. Visibility and identity seem suddenly intimately related. In some special sense, one becomes individuated by contacts metered in this way. Sylvester went on:

> "It was almost as though he wanted somebody there [whilst he was] painting in order to have someone to be apart from."

Given the subjective force of mutual gaze, it is scarcely surprising that huge efforts and resources have been directed at the provision of technologies to make available some form of mutual visibility for the physically separated. This chapter looks at visibility in remote interaction in three sections. The first summarizes several reviews of existing empirical work on video-mediated interaction, as a statement of the received wisdom on the role of video as an accompaniment to audio communication. The second section considers the non-technological evidence for supposing that there should be some benefit for supporting visibility in interaction. It attempts to draw together findings from studies of animal social conduct, human non-verbal behaviour and cross-modal perception, as they relate to synchronous communication. The third section sets out to reconsider the technological provision of visibility, with a closer examination of some of the key studies in this area. It focuses on the investigators' conception of VMC interactivity and the way in which they approached the business of its assessment.

1.1.1 Mimicking the "gold standard" of face-to-face communication

From the time of the telephone's invention, the addition of a picture of the remote party to a conversation was immediately seen as a natural and relatively unproblematic next step for telecommunication. Its arrival in the context of broadcast communication was forecast in the last century (Mee, 1898) and anticipated in innumerable works of science fiction. The videophone became a technological reality in 1927, around the same time as broadcast television services were becoming established, when a special line was installed between Washington and New York in the United States of America (Williams, 1981). In stark contrast with the steady increase in demand for and availability of television services, it was some time later before such a facility was even within reach of the ordinary household. In 1971, the Bell System attempted unsuccessfully to put a videophone on a commercial footing. It is easy to assume, with running costs at around 250 times greater than an ordinary phone call, that this failure was purely a matter of expense (Noll, 1992; Williams, 1981). However, as Williams observed:

> "there are real doubts, based on psychological evidence, as to whether people want the videophone at any price. It is fairly certain that most of us will not be using videophones by the year 2000" (p. 254, op. cit.).

Despite Williams' bleak forecast, the number of commercial products of this type are testament to the huge commercial interest and investment in the potential of videophones and videoconferencing[2]. In 1994, the Advisory Group on Computer Graphics found over forty different systems available on a commercial basis, covering a huge range of cost and connection options (Butters et al., 1994). It would seem that at least the developers of these systems consider the value of seeing the person with whom one is communicating to be self-evident. As Brittan (1992) put it:

> "By bringing two-way video to desktop computers, multimedia researchers hope to recapture some of the flexibility and human warmth that electronic communication has lacked" (p.59) [although] "there is no underestimating the challenges of getting psychology and technology to mesh" (p. 64).

It is difficult to pin down just what is constituted by 'flexibility and human warmth' but at the same time it is a fundamental matter for the purpose of psychological measurement. This chapter goes on to look at some of the 'psychological evidence' referred to by Williams and from other studies published since she made these comments in 1981.

1.1.2 Reviews of video-mediated communication studies

Alphonse Chapanis, one of the earliest investigators of video as a communications medium, reported a series of experiments comparing a range of contact media, from typewriters to ordinary, or 'face-to-face', interaction (Chapanis, 1975). Anticipating the development of sophisticated computer interfaces, he considered video as a cue-laden resource for providing interactive information as one of four available channels, the remainder made up of voice, typewriting and handwriting. Each of these channels, in Chapanis' terms, was thus an information-bearing "enrichener" of the total

---

[2]Videophones may be contrasted with videoconferencing facilities in the sophistication of the technologies concerned, the formality of the communication they support and the scope of their intended audience. Videophones are conceived of, quite simply, as a telephone or analogous personal audio communication device for real-time communication, augmented by an image of the party with whom one is conversing. Videoconferencing is associated with high-budget group communication, taking place in dedicated suites. Videophones are easily imagined as communications devices for a whole range of situations, just as with the telephone, from a quick chat to a formal enquiry. However, 'videoconferencing' has come to mean any kind of communication mediated by technology providing a view of remote parties. The term for videocommunications devices integrated into personal computers is 'desktop videoconferencing', rather than 'desktop videophoning'. In consequence, the terms will be treated as synonyms for the purpose of this thesis.

communications exchange. This channel-based conception of alternative communications media has been common to several investigators, but is more usually associated directly with sensory apparatus (thus type and handwriting would not be distinguished). It is assumed that each potential channel uniquely conveys certain kinds of information such that the more channels employed, the richer the whole resource to be drawn upon by those in interaction.

Chapanis' studies took place in adjoining soundproofed rooms and were carried out by two parties, or a dyads. The rooms had a window between them which could be screened off as a manipulation, a replaceable opaque panel to allow the passage of sound, and telecommunications links to workstations supporting each of the channels under investigation. In each case, the dyad collaborated to solve an information search task as an 'information seeker' and an 'information source'. The source and seeker each had part of the information required to complete one of the ten tasks, exemplified by: the assembly of a transporter for a waste bin, where the source has the instructions and the seeker the parts; finding a set of newspaper articles on a given topic, where the seeker has the papers and the source an index; and finding the nearest medical practice to a given address, where one has a list of doctors' addresses, and the other maps and street indexes.

For a wide range of 'communication modes', meaning channel combinations, a consistent set of findings emerged regardless of the method of connection between the rooms. These were that all voice-inclusive modes are associated with very similar time requirements for completing the tasks set and, similarly, non-voice modes are not discriminable on this measure. In other words, the anticipated benefits of channel combinations failed to emerge, and specifically, no difference was found between the "communication rich" video plus voice and other non-visual voice modes.

The most concerted investigation of video communication to date was carried out in the early 1970s, under the auspices of the British Post Office Corporation's Communications Studies Group (CSG). A comprehensive series of person-to-person telecommunication studies included contrasts of various audio-video links with audio-only and face-to-face communication, and for a variety of tasks. These studies were later summarized in three distinct works (Short, Williams & Christie, 1976; Williams, 1974; Williams, 1977). Williams' (1974) review of the CSG work indicates that the CSG's motivation was predominantly to evaluate the effectiveness of telecommunications equipment as a substitute for co-present, or face-to-face[3], activities of various kinds. There was no particular commitment to, or expectation of, the value of video devices. Indeed, several innovative audio-only systems received close attention. Video, then, was seen as just another possible technology to contribute to the communications process, in approaching the "gold standard" of face-to-face contact.

However, writing later with her CSG colleagues, Short and Christie (Short et al., 1976) a more sophisticated position was developed. The work reported is, in combination, presenting data and a review. In consequence, much of the anticipated benefit for seeing as well as hearing one's colleagues in some communication task is difficult to distinguish from a general, theoretical position taken on interpersonal communication processes.

> "The research discussed...anticipates the broad consequences of technological change...In studying the way in which people react to and use new telecommunications equipment, we are providing some of the information basic to any clear view of the way in which such systems will or should develop and the impact that they are likely to have upon our society." (Short et al., 1976, pp. 8-9)

Their notion of telecommunications impact was broad indeed, including relatively high-level societal issues, such as economic and environmental change, urbanization, and changes in social organization, as well as the matters that are the focus of this review: human communication and interpersonal relations. Where the latter are concerned, telecommunications were implicated in the maintenance of relationships over a distance and the availability of previously hard-won information or contacts to the individual.

Primarily, the benefits of seeing one's co-participants are not considered to be a consequence of additional, non-redundant visual cues. Rather, a dimension of communication is postulated which may be informed by a number of

---

[3]There are many dimensions to co-presence, or being in the same place at the same time, besides being able to see those with whom one is co-present. Co-presence has been closely identified with visibility to the extent that these other dimensions are often unfairly ignored. The idea of 'face-to-face' communication evokes a restricted notion of co-present interaction, in that only a disembodied notion of the face is implicated. Consider, for example, spatial proximity. Another person's proximity to oneself is: i) a necessary precondition to knowledge about artefacts in their range, and so to the determination of what might serve a joint purpose; ii) signifies social and affective matters; iii) whether or not speech is likely to be audible; iv) whether or not third parties are or should be in range of the exchange.

behaviours, via a number of physical dimensions more or less supported by various telecommunications facilities. This dimension is described as 'Social Presence'. Social Presence is said to be a quality of the media supporting communication, attributed by their users. Crucially, it is not simply a matter of the sum of the information available, as some investigators have implied (Rutter, Stephenson & Dewey, 1981), but is highly sensitive to the purposes of communication for which a medium is selected. Some purposes are intrinsically highly-charged in terms of Social Presence and so extremely sensitive to media choice. Conversely, other tasks, such as uninvolved information exchange, have low social involvement and thus the Social Presence of interlocutors would not be expected to vary very much with contact media. Social presence, then is an attitudinal construct applied to telecommunications technologies as "a cognitive synthesis of all the factors discussed in Chapter 4 as they are perceived by the individual to be present in the medium" (p. 65, op. cit.). This is important for the purposes of the current review, since the Chapter 4 referred to in this quote takes as its focus the role of *visual communication* in social interaction. So, visibility is closely identified with this special dimension of communication, not as a simple cue-additive enrichening of a communications environment, but as a special enrichener for a particular kind of information.

Writing in the late 1980s, Egido (1988; 1990) addressed the poor uptake of videotelephony by carrying out a market-focused review of the available technologies. Three large-scale questionnaire surveys were drawn upon to perform a "needs assessment" study, comparing the characteristics of ordinary meetings with the connection possibilities of video and other media. Criteria for including videoconferencing as a viable alternative were limited to the specification claims of the manufacturers, ignoring the actual technological adequacy of the equipment and the organizational culture of various potential investors for meetings. On this basis, she concluded that even the potential of these technologies in business was considerably overstated.

Egido observed that two periods of interest in VMC coincided with the mid-1970s' oil crisis and the renewed energy threats of the early 1980s. She reports that three major hotel chains in the USA, Marriott, Hilton and Holiday Inn, invested in videoconferencing facilities during this period, to try to counter the anticipated drop-off in business travel. The point is an important one, in the context of the treatment of the topic as a design issue, as pragmatic constraints can be the final arbiter of viability. The easy availability of travel and reduced journey time associated with modern transportation systems compete in a very real sense with telecommunications systems. A threat to travel and hence to face-to-face meeting would inevitably increase the appeal of videoconferencing, regardless of its intrinsic value as a communications medium. Egido's substantive point is that videoconferencing is misunderstood simply as a substitute for face-to-face meetings. It is a new medium for interaction that must be understood in terms of the new opportunities it presents, as well as a range of existing communications functions it might take on. Ford and Wang invested in videoconferencing to train their sales force, capitalizing on the speed of information updating to gain a competitive edge. In 1980, Boeing, again primarily in the interests of speed, cobbled together a two-way TV system as their 757 aircraft development schedule slipped, to link airfield, manufacturing and engineering facilities. The US department store chain, J C Penny, used videoconferencing to allow individual stores' senior buyers to view potential clothing product lines. Egido draws a sharp distinction between a kind of interaction she terms "meeting" and these other applications, and between formal conference suites and other implementation sites. An application-centred approach, coupled with an understanding of organizational priorities and development, she argues, is much more telling than accounts of failure based on the technical quality of available systems. In doing so, she hints at some residual potential for video as a communications device but fails to grasp the nettle: what is it about video that lends itself to some applications rather than others. Furthermore, as an assessment of the market history of videoconferencing, Egido's account can only find value where videoconferencing has demonstrated itself in some application. It does little to inform a proactive investigative strategy.

Whittaker (1995) again argues that little in the way of the kind of benefit expected of video-mediated communication has been forthcoming. He goes on to argue against using video as a substitute for a remote party's face and to recommend that video be thought of in an entirely different way (see Section 3.1.1 for discussion). It is noteworthy that he does not reject out of hand any potential value for video, as he puts it, in a "talking heads" role:

> "the benefits for non-verbal communication are subtle and subjective ... the most effective use of video may not be in a stand-alone application, e.g. videophone, but combining it with other communications applications, including shared workspaces" (Whittaker, 1995, p. 512).

1.1.3 Little evidence for the effectiveness of video

Video as a communications technology has been studied from a number of viewpoints. There has not been a single study to clearly demonstrate a significant benefit for audio-video mediation over audio-only mediation, whereas many have concluded that face-to-face performance surpasses technologically mediated interaction. The majority suggest some non-significant, and hence tenuous, indication of an advantage for video or else no difference at all, when compared to audio mediation: little reason, then, to add a video component to an audio link. It may be that Williams was right and that the commercial efforts to make videoconferencing happen have been misguided by an inconsequential association of faces and talk in everyday conversation. Findings of this kind, quite apart from their incompatibility with the public perception of video links both commercial and in subjective evaluations, are curiously at odds with a substantial literature. The link between a speaker's words and their visibility is not in the least straightforward but a considerable body of work demonstrates a significant role for non-verbal activity in unmediated communication processes. Although the words exchanged between parties in conversation are without doubt the primary engine of human communication, non-verbal behaviour can both modify the sense of an utterance and communicate alongside the spoken word. The majority of non-verbal behaviour is not acoustically based (Argyle, 1988; Short et al., 1976) and it may be that its informative content is the basis of a strong prejudice in favour of seeing those with whom we speak.

There are many forms of visually encoded non-verbal information, including facial expression, figural gesture, posture, dress or even lip movements as 'semi-verbal' communication. Gaze is of particular relevance to this discussion since the face has been the focus of all of the technologies considered. Many of the video images employed by investigators have not included whole body images or even views including hands and arms, limiting what is seen of remote interlocutors to head and shoulders at best. Whittaker describes such systems as "talking-heads video". It is unlikely that postural and gestural information would be conveyed. Proponents of talking-heads video might claim that their devices support facial expression and also gaze behaviour. Gaze has a unique status in the gamut of visual non-verbal behaviour, as both an expressive and an information acquiring act. Perception of another's gaze focus, where that focus is constrained to the others head and shoulders, is thus not at all the same thing as seeing someone looking around in an unmediated context.

It would seem, then, that the case for talking-heads video is weak. However, Whittaker points out that there is a consistent disparity in the literature between the way subjects report their experience of video-mediated communication compared to audio-only and the objective findings reported. There are several grounds for supposing that there is some value in the kind of information conveyed by a view of the face. The next section takes a closer look at some of the ways in which visibility and gaze have been implicated in unmediated communication processes.

1.2 Unmediated interaction, gaze and visibility

Non-verbal communication (NVC) has been of interest to a number of research fields, the longest established of which is ethology. Ethological research takes an evolutionary perspective on the forms of animal behaviour, in so far as they may be seen to be innate or in some sense be subject to evolutionary forces. Of central importance are the interaction between the predisposition of species to behave in particular ways and the environmental opportunities for the expression of those behaviours. Social psychologists and linguists have also carried out non-verbal communication research, looking for 'body language' structures and rules analogous to those of verbal language.

1.2.1 Signalling and communication

On the face of it, speech is easily identified as communicative behaviour. It is uniquely dedicated to the transmission of messages from a speaker to an audience. It is far less straightforward to conceive of the communicative component in non-speech behaviour. In the context of this thesis, it is instructive to consider the problems of distinguishing between behaviour and communication.

Ekman and Friesen (1969) have argued that any sense in which non-verbal communication is used must take into account the intentions of the individual who is the source of the behaviour in question: if the behaviour was not deliberately instituted or modified by that individual for the purpose of conveying some meaning to an identified observer, then that behaviour does not represent a communication. It is merely an attribution of intention by the observer or, in the absence of attributed intention to communicate, an interpretation of the enacted behaviour. For example, an observer may see someone making errors and exerting effort in some physical operation, from which the conclusion may be drawn that the person is struggling. The observer may additionally infer the person to be "asking" for assistance, perhaps identifying these struggles as exaggerated motions. The attribution of exaggerated motion would count as communication if the struggler intended his struggles to be taken as a request for assistance. If the struggler did not intend any such thing, then the evidence of his struggles would simply be a signal of difficulty.

In contrast, Watzlawick (1968) maintained that intention is irrelevant. Behaviour is meaningfully construed as communicative regardless of the intention of the observed, just as long as it is systematically interpretable by an observer. In any case, intention itself is not a terribly useful concept for the researcher of non-verbal communication since it is difficult to determine whether any particular behaviour was explicitly designed by an individual to convey a particular message.

Wiener (1972) pointed out that there is an important distinction to be made between *signs* and *communication*. If an observer assigns a particular meaning to a behaviour, then that assignation is as much a function of the observer as it is of the individual who produced it. In an instance where the individual attached no particular significance to a behaviour, then any inference made by the observer is purely an artefact. For a non-verbal behaviour to be truly communicative, it is necessary that the information encoded in behaviour is decoded without a material change of meaning. So on this account, a communicative act requires that the informative content of the behaviour be both transmitted and received. Whether or not a behaviour is potentially meaningful is a secondary consideration.

Bull (1987) also questions the validity of an intention criterion for defining NVC. He reports that negative affective state information, such as the boredom of an addressee, can be systematically interpreted through posture and gesture when one might imagine the individual concerned may not wish to make public their disinterest. Communication has taken place, in so far as the true state of the listener was transmitted and received, without the positive intention of the listener being established, and possibly the reverse being true. Bull describes this as "non-verbal leakage". Furthermore, the effect of non-verbal cues may be cumulative, such that in isolation a particular behaviour may not give rise to a specific impression, whereas the interaction over time may give the participants strong feelings about the mood of their counterparts.

Communication requires both that information is made available by one individual and that it is picked up by a second individual, so that both encoding and decoding of a non-verbal behaviour takes place. Provided that encoding and decoding can be said to have occurred, the status of the information conveyed need not be included in a conscious account of communication at all. In other words, some communication can be considered to be, in effect, automatic. Execution and reception of the information, via the mediating behaviours, occurs without the need to invoke notions such as planning or formulation.

1.2.2 Comparative studies of primate visual communicative behaviour

Consideration of the meaningfulness of facial expressions has a long-established place in the study of communication. In 1892, Charles Darwin published his "Expression of Emotion in Man and Animals". This work presented a case for natural selection of social as well physiological attributes and the argument that mechanisms of communication are centrally implicated in such a process of selection (Darwin, 1892). Indeed, as Larson puts it Darwin was explicitly concerned to demonstrate that:

> "social units are not just expressions of feelings of sociability. Animals ... became sociable as it became profitable to live together and as those animals who were best able to live in close proximity came to be the most successful at leaving progeny"(Larson, 1976, p. 257)

Broadly, ethologists are concerned with functional accounts of behaviour within a Darwinian framework of natural selection. An account of the adaptation of some behaviour requires that its role is properly articulated with reference to the forces to which it is assumed to be subject. Darwin's assertion that 'man is descended of the apes' invited direct comparison with extant primate species, both from proponents and detractors of his thesis (Chevalier-Skolnikoff, 1973). Various taxonomies of expression were developed to describe facial muscular configurations, in order to categorize and enumerate the development of communicative abilities in mammals. These studies lead on, as sociobiology, to the consideration of animal sociability in evolutionary terms. Whether or not this heritage is a function of the gene pool or behaviours normalized into customs through social mediation, the idea of there being an 'evolution' of expression (Snowdon, 1983) and social behaviour (Bateson, 1987) persists. It should be noted that this point is the converse of a more familiar psycholinguistic debate. The issue here concerns whether human communicative mechanisms (at least in part) can be identified with those of other animals; it is not whether chimpanzees can be taught an abstract symbolic language. Some communicative behaviours, whether by genetic transmission or enculturation, are endemic constituents of the human behavioural repertoire. In this sense, they are fundamental to human communication.

One of the oldest principles of ethology is the inheritance of fixed action patterns (FAPs), stereotypical species-specific behaviours taken to have adaptive significance (Tinbergen, 1963). FAPs have been studied as behavioural units upon which individuals can capitalize in particular circumstances. Facial expressions have been examined in this way by atomic or elementary analysis, such as the systematic catalogues of behaviour known as 'ethograms' (Chevalier-Skolnikoff, 1973). In this application, ethograms comprise 'expressive elements', judged to be representative of units of communication as analogues of verbal statements. Very often, the sense in which 'expressive' has been employed is highly subjective. For example, bared teeth might be interpreted as "Stay away from me" or "That's mine: don't touch it". Non-verbal signals could be interpreted as discrete messages of some kind or classified as indicators of affective state. Such behaviour (e.g. teeth baring) is usually interpreted in terms of emotive content (i.e. being angry or upset) following the Darwinian adoption of an anthropomorphic taxonomy. Primatologists have operationalized terms like 'fear' and 'affection' by way of correlating expressive behaviour with active behaviour, such as approach and proximity. Van Hoof (Russell, 1991) proposed an hierarchical framework for integrating clusters of expressive elements, categorized under five general emotional contexts: play, aggression, affinity, excitement and submission. In any case, the persistence of general emotional categorization of facial expression, and indeed other non-verbal behaviours, has remained a central focus of theoretical and research activity (Chance, 1967; Hinde, 1974).

Two theoretical preoccupations may be discerned from the above. One is concerned with the mechanisms supporting social status and the maintenance and development of groups; the other is for the abstracted meaningfulness of non-verbal behaviours in individual exchanges.

1.2.3 Eye contact, attention and social structures

As discussed above, the earliest interest in non-verbal communication was overwhelmingly in those aspects manifesting as visual phenomena. Of these, eye contact[4] is perhaps the one which most subjectively encapsulates the potential force of non-verbal communication.

Comparative research has produced a coherent set of principles applicable to all primate societies. In particular, directed visual behaviour is seen to be important in regulating primate social interaction, propping up existing power relations, as dominance status, in their societies. Non-human societies commonly employ visual signals as an overt index of social status. Subordination is associated with large amounts of looking at a few members

[4]The term 'eye contact' is used loosely here, referring not specifically to mutual gaze but to a reflexive awareness of the deployment of gaze.

of an animal's community, whilst those to whom they are attending are not simultaneously looking back. Conversely, dominant status is associated with a freely roving gaze, uninhibited by bouts of eye contact and not involving extended scrutiny of any one of its low status fellows (Camras, 1980). A variety of primate societies use the stare as a central part of their behavioural repertoire for assertion: it is identified as a threat gesture that can result in submissive responses from other members of the group (Camras, 1980).

A considerable body of ethological and social research suggests that staring at someone is a dominance gesture and that the breaking of eye contact is a signal of submission (Bull & Frederikson, 1994; Ellyson & Dovidio, 1985). The activity of looking in itself is largely considered to be indicative of attentional focusing. M.R.A. Chance, in his paper "Attention Structures as the Basis of Primate Rank Orders", describes attention distribution as the primary medium of social stratification (Chance, 1967). Human groups are commonly hierarchically structured: social or organizational ranking is all pervasive and dominance is one such stratification process. Chance contends that the attention structure of groups is indeed a direct indication of the social relationships that exist within those groups.

Status structure of primate groups is predominantly controlled through visual attention, in particular the maintenance of persistent attention by subordinates on dominant animals. Ellyson and Dovidio (1985) similarly argue that dominance itself should actually be defined by visual attention and not, as was previously most usual, in terms of aggression or access priority to resources. Dominance is said to be realized through a monopoly of a social group's attention by the dominant animals, rather than having a material focus: it is the basic engine of all hierarchical structures and thus should define any measures of dominance. It is the amount of attention paid to dominant members of a group that determines the structure of interactions between members, not the events that initially gave rise to the relative positions in the group's hierarchy.

Although social dominance is closely related to distribution of attention, the relationship is clearly not one of identity. In attention-based studies of children, for example, correlations are notably higher in some groups than others. The influence of the other correlated factors significantly modifies the relation between attention and dominance (Omark, Strayer & Freedman, 1980) . Attention owes its deployment to at least as many factors as the number of factors in which it is in turn influential. Group members engaging in prosocial behaviour, or deviating from group norms, will all attract disproportionate amounts of attention, regardless of dominance status. Attention, when defined by looking behaviour, manifests itself in consistent patterns, related to various other activities and structures in the group. Imitation, popularity and particularly dominance are all related to looking, when compared with a variety of other factors. Attention deployment as discussed in this section is best considered as a mechanism through which all the hierarchical and other aspects of group cohesion operate. It offers the revelation of several of these mechanisms, with some form of structural analysis of the attentional patterns apparent in the behaviours of group members.

1.2.4 Seeing and hearing 1: functional relationships between speech and gaze

Ellsworth and Langer (1976) hold that the stare is an involving and arousing cue functioning as a platform for dominance signalling. The context of the stare therefore is vital for its interpretation. The meaningfulness of any behaviour is defined by the set of behaviours forming a functional unit for the originator of the communication. In particular, the activity of speaking seems to significantly alter the context for interpretation of gaze deployment.

As we have seen, in primate societies a dominant animal tends to meet the gaze of a subordinate, establishing mutual gaze, and the subordinate tends thereafter to look away, breaking mutual gaze. Investigating human dominance relations in a conversational context, Strongman showed that more dominant members actually broke more glances than did their subordinate partners during floor exchange periods (Strongman, 1970; Strongman & Champness, 1968) . This is precisely the opposite of a prediction that might be made from the non-speech findings for dominance and attention. Kendon (1967) suggested that gaze breaking is used as a signal for floor acceptance. Following Strongman's findings, it may be that the emphasis on cooperation, inferred from the smooth regulation of conversation described by Kendon, is misplaced, or at least contextually limited. Just as the meaning of verbal language units varies depending on the context in which they occur, so it is reasonable to assume the same for non-verbal behavioural units. Where a participant asserts herself in a conversation, it may be that continual appropriation of the 'my turn for the floor' signal functions as an act of dominance, overlaid on the dynamics of turn-taking management.

Several individually meaningful behaviours may be an important part of a simultaneous behavioural production that, in combination, serves the function of conveying a supervenient meaning. This type of interpretation has been variously expounded by a number of mid-twentieth century philosophers, but primarily as Speech Act Theory (Austin, 1962; Searle, 1970). Speech act theory makes a critical distinction between the *forms* of speech and the *function* of speech. The phrase "streams of messages", introduced above, was used deliberately in preference to the more familiar "channels of information". The latter is invariably identified with sensory mode; for example the visual channel. Messages of particular kinds, such as communicating displeasure in some circumstance, may be conveyed simultaneously through many channels. This has led some investigators to suggest that communicative behaviour is highly redundant: if someone says they're angry, that they also look angry is not additionally informative. However, language itself is highly contextualized, so that a particular set of words can stand for a number of different messages. Ambiguity is resolved in a number of ways: by drawing on the immediate history of conversation; by interaction between the individuals concerned; by tone of voice or gesture; or, needless to say, facial expression. It is difficult to dissociate vocal tone from facial expression in such instances. Irony, sarcasm and humour can hinge on the style of the delivery of the words employed, just as much as on the other behaviours accompanying those words. Given the preceding account of the evolutionary age of non-verbal facial expression, one might suggest that however a phrase is constructed, the face of the person concerned frames its interpretation rather than vice versa.

1.2.5 Plain old paying attention

The preceding section takes a particular stance on overt demonstrations of attention, viewing it as a matter of *projection* rather than of *reception*. Of course, paying attention to one's environment in a wider sense is of importance quite independently of these social concerns. All animals must devote some effort to assessing what is going on in a wider sense at any time. Attention deployment is a skilled activity subject to a range of constraints (Neisser, 1976). It may be "grabbed" by some occurrence consequent of passive monitoring (an orienting response) or directed in pursuance of some intentional concern. The social role of attention deployment must be managed with the requirements of attention for planning and acting[5]. Thus accurate

[5]Whilst the reader might have no trouble with the idea of human planning and action, that other animals plan as well as act is somewhat contentious. However, as Bateson (1987) has observed, the metaphoric application of such anthropomorphic concepts to animal behaviour can be highly profitable.

and reliable sampling of environmental conditions is critical to the success of an organism. For that 'most social of animals', the human being, accurate and reliable sampling must be inclusive of an environment defined by peers and variable behavioural imperatives.

Selective attention is an invaluable, not to say vital, mechanism for deriving maximum benefit from the available information sources. There are many potential sources of information within an individual's immediate environment, of varying moment-to-moment utility. Figuratively, knowing which voices to listen to is a great accomplishment (Niemark, 1970). The scope of human communicative capacity is somewhat beyond that of other primate societies and so the number of available voices, or streams of messages, is considerably greater and correspondingly more complex that the foregoing discussion suggests. They threaten to place an intolerable burden on the available attentional resources and yet, in normal interpersonal interaction, no such strain is in evidence. Neisser has suggested that a complex and integrated set of skills is developed to deal efficiently with the raft of information pertinent to an animal's activities. These streams of messages are seamlessly and actively brought together in a form readily digested by an information acquiring organism specially designed for that purpose.

1.2.6 Seeing and hearing 2: modal characteristics of vision and audition

Visual and vocal signal reception differ dynamically in a number of ways, with implications for the interpretation of non-verbal cues. One of the most salient characteristics of vision, when compared with audition, is it's reliance on head orientation. Hearing is omnidirectional and, from a projective viewpoint, it cannot be clear which auditory information source is being attended. Definite limits to the human field of vision entail a close relationship between imputed attention and orientation towards visible information sources. The deictic value of this directionality is added weight by the difference in perceptual quality at the periphery compared to the fovea. Gaze direction is thus vital to the exploitation of the visual information presented by the environment. The detection side of the coin is that whereas visual information may be localized only if the head is directed towards it, auditory information is potentially detectable regardless of the orientation of the head.

Another modal constraint concerns the requirement for some sort of alternating synchronization between those involved in a verbal exchange, i.e. taking turns at talk. Audition is rather more likely to become overwhelmed with multiple sources than vision, such that the content of each source of information is increasingly difficult to establish. It is a relative disadvantage rather than an absolute distinction, as demonstrated by the "cocktail party" and other selective listening effects on the one hand and selective looking on the other (Cherry & Taylor, 1954; Neisser, 1976). Although speech can be discriminated even when audio quality is severely degraded, production and reception of speech are normally closely coupled to retain intelligibility. There are few such constraints on visual communicative behaviours.

In addition to the separate constraints on vision and audition, the two modalities interact in meaningful ways. Stroop (1935) reported subjects struggling inordinately to name colour words presented in colours differing from the colour name spelt out. Of more direct relevance to the current discussion, McGurk and McDonald (1976) found that subjects viewing video recordings of people saying one word, for example "goes", whilst listening to recordings of another phonetically similar word, such as "bows", reported having perceived a hybrid resolution of the two, as "those". Speech reading, i.e. interpreting speech from seeing lip movements and changes in facial musculature, is a tremendously powerful conversational resource operating independently of the emotive components discussed earlier (Campbell et al., 1996). People with significantly impaired hearing are able to function as full members of a conversation, provided they are able to see the faces of their counterparts (Campbell, 1998).

There is some evidence that people who can see one another take less care to speak clearly than those who cannot (Anderson et al., 1998). This work implies that the potential for improved intelligibility in noisy environments might be undermined by a false confidence in the medium. There is no guarantee that the quality of a video image received by an addressee will match the expectations of a speaker. However, where people experience background noise or audio interference, it is quite possible that Anderson et al.'s results, drawn from words excised from clear audio and video-mediated conversation, are not applicable.

These characteristics have obvious consequences for the utility of communicative behaviours demonstrated in environments where the transmission media make the production and/or reception difficult in either sensory mode. Cross-modal redundancy stands to compensate for conditions such as low light or high ambient noise levels, undermining the linguistic capacity of one or other of vision or audition.

1.2.7 Visibility and eye contact as regulators of interaction

As discussed above, the manner in which communication proceeds is highly contextual, such that particular behaviours may be identified with particular meanings only in a contingent and general manner. NVC research in social psychology has struggled with this intractability, that non-verbal signals seem to fit effortlessly into a normal pattern of communication and yet elude categorization. Radley (1991), in his book 'The Body and Social Psychology', identifies NVC research as a prime example of the difficulties faced by social psychology as a whole:

> "To take one example, the field of non-verbal behaviour has long seemed to me a body of work from which significance has slipped further away with the accumulation of findings." (p. vii)

Even so, there is a tradition of considering non-verbal behaviours in interaction as having some generality. Birdwhistell (1970) distinguished two major functional groups of non-verbal signals: *integrational* behaviours, which reference the larger context of conversation and help to manage its progress; and *informational* behaviours, as affording directly meaningful interpretation.

Short et al. (1976) relate these to Argyle's (1969) six-category taxonomic structure for NVC. Argyle's mutual attention and responsiveness, and channel control and feedback were considered to correspond with Birdwhistell's integrational behaviours. These include behaviour such as head nodding and affirmative gesture, non-verbal vocalization, and maintenance of gaze and eye movements. Illustrations, emblems and interpersonal attitudes were said to represent informative behaviours. The latter, informative group, includes illustrative or figural gesture, gesture standing for words or phrases (for example, a shrug), and facial expression for affective reactive information. Short et al. identified Argyle's "feedback" and "interpersonal attitudes" categories of visual cue as of greatest relevance to interaction:

> "These two functions are together by far the most important functions of the visual signalling lacking in audio-only communication" (1976, p. 46).

Visual, non-verbal behaviours said to support the conveyance of interpersonal attitudes include:

- proximity, orientation and eye contact, or mutual gaze, as they relate to conceptions of an intimacy equilibrium and indications of cooperativeness;
- physical appearance, posture and gesture, in terms of attributions of interpersonal attitude and personality;
- facial signals, as a conveyance for speaker state information (e.g. happy and dishonest) and also as a semantic modifier of an utterance (e.g. upward eyebrow movement combined with a statement becomes an interrogative);
- direction of gaze, for monitoring and turn transition (following Kendon - see below), dominance signalling (as discussed above), and indicating emotional intensity.

They go on to propose that the primary distinction between audio-only and audio-video or face-to-face interaction is a composite communicative dimension, known as Social Presence, to be discussed in depth later. Their brief discussion of the role of visual communication and social interaction does not warrant such a strong emphasis on these informational rather than integrational categories. However their major experimental discriminations between audio and face-to-face interaction were found for tasks with a strong interpersonal or interparty component. These were described, with a hint of circularity, as tasks with a high Social Presence requirement. It is perhaps unsurprising that they settled on Argyle's "interpersonal attitudes" class of cue, admitting the value of his "feedback" class only as a secondary matter.

An explanation for this emphasis may be the particular scenario envisaged by Short, Williams and Christie for remote interaction; two or more people engaged in some verbal task, without mention of mediating artefacts. In retrospect this would seem to be a fairly serious omission. Chapanis' studies of 'interactive communication' involved artefacts, both to support and as the focus of interaction (Chapanis, 1975). His experimental tasks included the construction of a dustbin trolley ("trash can transporter", in American English) and a negotiation exercise relying on printed matter. Artefacts are increasingly recognized as important elements in communication at work. Frohlich (1995) has observed that the majority of office tasks are mediated by the introduction of and reference to documents. Whittaker's (1995) 'rethink' of video is based entirely around the role of video in providing a shared context for and conveying artefacts common to ongoing collaborative work.

There has been some significant work undertaken on the integrational aspects of non-verbal behaviour, implicating a variety of visual factors. The most significant social psychological paper on NVC, for the purpose of this review, was published by Kendon (1967) as "Some functions of gaze-direction in social interaction". It was the first time that an investigation was carried out on the role of gaze as a conversational regulator. Kendon's work was inspired by Sartre's account of the personal significance of gaze, as alluded at the beginning of this review, and by Goffman (1964) who described a group in communication as "an eye to eye ecological huddle" for "maximising one another's mutual perceivings". Kendon made sound and film recordings of seven dyadic interactions, the latter at two frames per second. Pictographic transcriptions codes were used for gaze direction, facial expression, and head, trunk, hands and arm positions.

Kendon is often cited for his finding of a consistent relationship between the act of gaze towards and away from participants as a turn exchange signal. Discourse is considered to proceed by taking turns at talk. These turns are often described as if they were the sole resources in communication, to be offered, given and taken in a kind of 'economy' of speech. Looking away was associated with taking up the conversational floor and looking towards the other participant with "a sustained gaze" with yielding the floor. Kendon interpreted this behaviour as both signal presentation and information gathering: the opportunity was signalled and whether or not it would be taken up was determined from the visible reaction of the addressee. Kendon additionally noted that during long passages of fluent speech, speakers look at addressees towards the end of phrases. He interpreted these as attention seeking acts, anticipating Yngve's work on backchannels. Expressive functions carried by gaze were also reported as embedded within a regulatory mode of gaze deployment and to accompany a distinct monitoring function.

Rutter (1984) points out that a number of studies have subsequently examined similar indices of behaviour in interaction without finding the relationship claimed by Kendon for looking behaviour and turn exchange. He observes that Kendon's samples of conversation were few and selective of a particular kind of utterance; fluent speech of over five seconds in duration. Furthermore, each of the utterances serving as the unit of analysis contributed both turn beginning and turn completion data. The behaviour preceding and succeeding each of these was not reported. In a purpose-designed replication study, Rutter, Stephenson, Ayling and White (1978) found that, broadly, looking behaviour followed Kendon's model. However, they noted that the model he proposed was inconsistent with their additional finding that mutual gaze obtained on half to two thirds of these occasions. Unless one is looking at the other, there is unlikely to be any value in the act of looking.

Over a number of years, Duncan and co-workers carried out a series of intensive studies to investigate an integrational aspect of non-verbal behaviour, that of regulating conversation (Duncan, 1972; Duncan, Brunner & Fiske, 1979; Duncan & Fiske, 1977; Duncan & Niederehe, 1974). They identified five classes of regulatory cues: turn-yielding cues; attempt-suppressing cues; back channels; within-turn signals; speaker-state signals.

Turn yielding cues are behaviours associated with the surrender of the floor, in other words a speaker indicating that she is about to stop speaking and thereby provide an opportunity for a listener to reply. These offers of a turn include: i) rise/fall of pitch at the end of a clause; ii) a drawl or stress on the final syllable; iii) the termination of hand gestures; iv) stereotyped expression e.g. "but, ah...", "sort of..."; v) a drop of pitch or loudness associated with stereotypic expression; and vi) the completion of a grammatical clause. These cues share a degree of redundancy, so that some or all of them might be found at any turn termination point. Attempt-suppressing cues are competitive behaviours on behalf of a speaker to prevent a listener from taking the floor. Duncan identified only one of these, hand gesticulation (Duncan, 1972).

Back-channels are short messages from an addressee not constituting a claim for a turn at talk, but providing status information, primarily of continuing attentiveness and degree of agreement (Yngve, 1970). Duncan identified five back-channel behaviours: i) sentence completions; ii) requests for clarification; iii) brief phrases; iv) head nods, and v) head shakes. Brunner (1979) noted that smiles also function as a back-channel, since they invariably accompany the other, acknowledged cues. Within-turn signals (i.e. from the current speaker) allow for the use of back-channels. They are sensitive to a listener's intention to back-channel, providing appropriate points in the conversation for the listener to contribute in this way. Duncan found that back-channels are generally preceded by a shift in head direction towards the listener and the completion of a grammatical clause.

Speaker-state signals are indicators of conversational intent: they denote when a participant has or intends to take the floor. Duncan and Niederehe (1974) found that pre-emptive speaker-state signals, as turn-taking attempts, additionally involve: i) shift away in head direction; ii) starting to hand gesture; iii) an audible inhalation of breath and iv) "paralinguistic overloudness", or abnormally high volume non-speech vocal activity. However, these cues are sensitive to the context of the interaction. Duncan's studies tended to examine few interactions in great detail. Two psychotherapists used all of these cues as distinct from back-channels, whereas a psychotherapist with a client used only (i) and (ii). Duncan and Fiske replicated this latter finding, hence headturning and gesticulation are considered to be the key characteristics of speaker-state.

Duncan, Brunner and Fiske (1979) went on to refine the notion of speaker-state signals to add 'strategy' to the matter of 'organization' described above. Strategy signals overlay organizational signals in relation to their importance or effect from the speaker viewpoint, here allowing a role for other-directed gaze. Rutter (1984) observes that this is the only point at which Duncan's work finds any role for gaze. Yet in making this observation, Rutter misses that many of the other signals identified as conversational regulators by Duncan and co-workers are purely *visible* and hence presuppose gaze deployment (see below).

A note on studies examining general consequences of eye contact is worthwhile here. Argyle and Dean (1965) found an inverse relationship between eye contact and interpersonal distance, leading them to postulate a factor - "intimacy" - maintained in some form of equilibrium between individuals through the manipulation of a set of behaviours. Exline has similarly related looking behaviour to an "affiliation" trait (Exline, 1965). Kleck and Nuessele (1968) artificially manipulated the amount of eye contact during interaction with instructions to a confederate and found that high levels of this behaviour were associated with subjects' rating the confederate as warm and friendly. Champness examined frequency of eye contact in terms of agreement during discussion. Above-chance levels of eye contact were found when subjects were in agreement and less than chance when they disagreed. The role of chance in analysis of behaviour has not always been accounted for. In early studies of posture and gesture, synchrony was regarded as an index of affiliation. Later studies, controlling for the 'background', or expected-by-chance co-occurrence of these behaviours, found that the earlier reports of synchrony for gesture were not confirmed but that postural congruence was robust (Bull, 1987).

1.2.8 Interactive deployment of gaze beyond eye contact

Eye contact tends to be thought of in terms of mutual gaze and yet, as should be clear from the ethological work previously discussed, gaze deployment towards other individuals requires no definite or extended gaze reciprocity. In this literature, 'making visual contact' almost by definition involves fleeting periods of mutual gaze. When a dominant animal looks towards a submissive animal who was already looking towards the dominant, mutual gaze is briefly established before the submissive looks away. Gaze at others has a role within which mutual gaze should be understood. Within a dominance framework, gaze serves the function of establishing who is the subject of whom. A broader framework might cast gaze as a matter of finding out what is the subject of whom, whether or not it is another group member or something physical, perhaps with a bearing on group activity.

Rutter (1984) seeks to distinguish between the act of looking and the business of seeing. It is closely related to Duncan's identification of gaze and of visible behaviours, suggesting that gaze-in-itself has little to contribute to the process of conducting interaction, but that it subserves the operation of other visible behaviour. Rutter questions the focus of prior work on gaze behaviour, especially that of Kendon which set in train the research paradigm, as pertaining to looking towards interlocutors' eyes. The whole body is informative and it is far more productive to consider gaze behaviour in a whole body context. It would be rather simplistic to dismiss Kendon's work as all a matter of eye contact however. Most of Rutter's critique sets Kendon up as a protagonist of eye contact as a necessary and sufficient mechanism for managing turn exchange. Regulation of this kind was certainly a central theme of Kendon's work but only alongside precisely the monitoring function Rutter advocates. Also, Kendon's data necessarily cast gaze in a probabilistic role, as one of the resources available to effect turn exchange, and not as the soul control of the mechanism. In a reply to Rutter, Kendon (1978) makes reference to the exhaustive work of Sacks, Schegloff and Jefferson (1974) on turn exchange to emphasize that it is one of many resources to be drawn upon in an highly contextualized fashion to achieve regulation in a redundant system. He explicitly distinguished between conversation 'introduction' and 'discussion' in selecting samples for his original study, choosing to analyse only the later portion of recorded interactions. Writing later still, he maintains a privileged place for mutual gaze in the gamut of gaze-related activity, but adds:

> "It seems reasonable to suppose that [mutual gaze] will have marked arousing consequences, but what line of action it arouses another to take will depend upon the context in which the look is perceived." (p. 51)

Considered in this light, gaze maybe seen to function not just in a monitoring or interrogating capacity for interlocutor state but also as a regulator depending on the accepted context of talk. Clark's "grounding" sets such monitoring activity as a fundamental prerequisite for any communication and yet is typically expressed in purely linguistic, rather than visual, terms. An additional step is still needed to appreciate the role of gaze in this capacity, and that is to understand that gaze operates in a wider context than interlocutor body and the miscellaneous "away" space necessary to oppose other-directed gaze. In a world where communication is so often about material things referenced through visible media, gaze serves to inform about attention not only on one another but about attention on the materials of conversation. Heath and Luff (1992a) point out that everyday work frequently involves coordination on a wide range of potential activity. As they put it:

> "Mutual visual access provides individuals with the ability to discern, to a limited extent, the ongoing organization and demands of a colleague's activities, and thereby coordinate their interaction with the practical tasks at hand. Moreover, mutual visual access provides individuals with the ability to point at and refer to objects within the shared local milieu." (p. 4)

Thus, the critical consideration for analysis of gaze should be its target defined by conversational resource. Even within the remit of the above discussion, the role of video in providing visual information about speakers and listeners is somewhat restricted by the dyadic nature of so many of the studies. Heath and Luff found that knowing who was speaking in larger mediated interactions was one of the key benefits of video. Another matter that the preceding account makes no reference to is that communication is a process that must begin and end besides being managed once in progress. As Daly-Jones (1998) has discussed, each of these phases is equally nuanced with the additional matter in the current context of making decisions about the most appropriate media choice for the purpose of the would-be communicator.

Following on from this last point, it is worth remembering that non-verbal communication is a part of a whole communication process. Clark's model is of communication as joint action, occurring through an often complex series of steps whereby information is offered and received (Clark, 1996; Clark & Schaefer, 1989). To be able to communicate with another person requires that they be identified as a reachable audience and that they are to some degree individuated. The person who offers some information will typically look for some evidence of its having been understood by the recipient(s) prior to advancing the conversation by offering further information. Clark's account capitalizes on the immediate referents of the speech exchanged as part of a general 'evidence gathering' process. Evidence of this kind presupposes generation of a model of the recipient's level of linguistic, declarative and environmental or physical context knowledge. Gaze is a powerful indicator of attention distribution and, in so far as orientation conditions access to facial expression, interpersonal affective state. It is best appreciated as a mechanism to jointly access and convey both of these independent kinds of evidence in conversation, overlaying the integrative and regulatory functions described by social psychological investigators.

1.2.9 Summary of visual non-verbal communication characteristics

Section 1.2 has discussed some of the evidence concerning the role of non-verbal behaviour in communication, its significance and function. By describing the breadth of interest in the topic, it has attempted to demonstrate some recurrent properties of visual non-verbal communication. These include status as a fundamental component of communication; its close relationship to attention, and consequent implication in attentional social structures; its facility for coordinating group interaction by individual floor-holding and turn transfer, and its function as a mediator of the sharing of objects in cooperative activity.

In sum, it would appear that visual, non-verbal behaviour is widely implicated as a highly developed and fundamental process for mediating the sharing of information about the status of individuals, about the information being exchanged, and about their mutual role in a collective enterprise. Visual non-verbal behaviour plays a part in the coordination of available resources subject to important contextual effects given both by social setting and physical characteristics of the environment in which the interaction is played out. These characteristics inform an account of the characteristics of interactive behaviour in that they show:

- redundancy, suggesting the management of a range of personal resources;
- close sensitivity to the state of systems with which the focal individual is interacting;
- a vital relationship with artefacts shared between interacting parties, setting the context for resource management.

1.3 Understanding video-mediated interaction

Following on from Section 1.1's overview of current opinion on VMC, Section 1.2 found a number of reasons for supposing that seeing one's interlocutors should be beneficial to interaction processes. Three ways of accounting for this failure are first considered and rejected before a re-examination of VMC evidence within a technological-historical framework. The concept of interactive behaviour to be supported by video-inclusive media is exposed as having evolved in tandem with new technologies, yet with little evidence of any parallel evolution in assessment methods. Section 1.2's conclusion, that gaze distribution stands to contribute to many levels of communication subject to a variety of contextual parameters, is found not to have been substantively addressed. Qualitative studies seem to support the mechanism of gaze distribution as an important factor in the success or failure of VMC systems. It is suggested that a complementary quantitative approach would help to substantiate this evidence on the one hand and, given an appropriate theoretical framework, to formulate a workable solution to measuring interactive behaviour on the other.

1.3.1 Implementational and conceptual criticisms of VMC research

Section 1.2's examination of non-technologically mediated interaction suggested that gaze behaviour fulfils a variety of functions and that, in conversation, it has some important relationships with speech behaviour. In a conversational context, the role of gaze becomes implicated in the purpose and referent of talk, alongside its interlocutor attention and regulatory functions. Purpose may be reflected in interpersonal and interparty attitudes evinced by non-verbal expressiveness. Referential implication casts gaze deployment in a gestural role, as evidence of environmental interrogation or a cue to direct the gaze of an interlocutor. It would seem likely that gaze in some of these roles should have a place in mediated interaction.

Three accounts for the failure to find any evidence of this supposed value are now considered. The first suggests that the technologies might not have been adequate, the other two that the role of visibility in VMC has been misunderstood by investigators, such that their implementation of VMC has been seriously flawed.

An aspect of telecommunications technologies so far ignored, and yet of vital importance to their success or failure as practical means of communication, is the very integrity of the technology itself. Whittaker provides a lucid account of this, based in no small part on an earlier study of the spoken aspects of telecommunication as they are affected by bandwidth and lip-synchronization handicaps to work with networked video-image transmission (O'Conaill, Whittaker & Wilbur, 1993; Whittaker, 1995). Similar findings have been obtained in a more recent study of telemedical consultation. Again, the work included many contextual attributes that could be conveyed by video, corresponding closely to Whittaker's video-as-data conception, but suffered from inadequate audio support (Watts & Monk, 1997; Watts & Monk, In Press). Audio support presents many engineering challenges, largely for achieving effective echo cancellation without impeding the fluidity of multiparty exchange, but standards for coding-decoding and transmission protocols are continually yielding factors of performance gain (cf. Schaphorst, 1996). There are three distinct problems here: the basic adequacy of the technology to support transmission of usable quality, however it is realized (what is experienced); the availability of a common high-bandwidth infrastructure for those transmissions (the "wires"); and the basic compatibility of the strategies employed for using that infrastructure once selected (the "video languages spoken"), reflecting the "interoperability" of systems. One commentator summarized these difficulties with the following advice: "let the standards dust continue to settle and ask yourself if your users would be just as happy with a pictureless voice/data-sharing WAN link"(McCarthy, 1995) .

However, the purpose of this thesis is not to account for the technological problems that beset telecommunications devices; at issue is the interactive consequence of video-mediation assuming that the technical difficulties will be overcome. In just this way, the majority of studies have sidestepped the basic engineering difficulties, opting for the control of a laboratory setting. Where others have been out in the real world, special high-budget and temporary systems have been put in place, such as the microwave television link used by Moore et al. (1975).

The other two objections to VMC studies are rather more challenging. They might be described as a 'barking up the wrong tree hypothesis'. They suggest that the way VMC has been conceived has little to do with the way visibility informs people about one another in practice.

In a review entitled "Rethinking video as a technology for interpersonal communications", Whittaker (1995) suggests that a video channel has simply been misunderstood. Rather than as a conduit for the facial, non-verbal content of communication, the real value of VMC is in augmenting speech with additional information about a shared workspace.

This approach draws heavily on an earlier paper, "Turning away from talking heads" (Nardi et al., 1993), describing the use of 'video-as-data' to support interpersonal communication in a neurosurgical unit of a North American hospital. Audio-video links connected a series of operating theatres to a central monitoring suite. A senior neurosurgeon was constantly in "virtual" attendance at all the procedures being carried out. A descriptive account of the work is provided alongside a brief resumé of the very limited evidence of the utility of video to support person-to-person communication. The video images conveyed far more than the "talking heads" view commonly identified with video links. They included a wide view of the entire theatre and all those physically in attendance. It was also noted that the activities and quality of the interaction between surgical team members was carried by the audio part of the link. A more extensive version of this article has (Nardi et al., 1996) further expanded upon the mechanism by which video-as-data is said to function in support of collaborative work:

> "Video-as-data contrasts with the traditional 'talking heads' approach. In applications such as videophone and videoconference, video is used to show the head and upper body of remote interactants. By contrast, we focus here on using the video image to display *shared dynamic work objects* that are critical to the task being carried out by a distributed team." (Nardi et al., 1996, p. 74)

Any case one might make for the effectiveness of VMC for conveying "data" in this way is, in principle, impartial towards talking-heads video. The mechanism described is informed by a theoretical account of group behaviour and coordination. It reflects the indicative role of visibility rather than any of the other matters discussed in Section 1.2: the grounds for supposing that a "data" conception of VMC might be useful are largely orthogonal to those supporting "talking heads" video. Gale has demonstrated that seeing, where someone looks in both a physical and a virtual space, can exert a powerful effect on conversations about spatial stimuli. The video-as-data approach corresponds more closely with an 'information space' rather than a 'social space' account of visible cues. As yet, this work is in its infancy, but whether or not video-as-data is successful, the disparity between evidence for the effects of non-verbal behaviour on interaction and video-mediated interaction is not addressed. It does not make much sense to see these conceptions as competitors and the lack of evidence for talking-heads VMC is not the success of video-as-data.

Much of the evidence discussed in the second part of this chapter, concerning gaze as a social regulator, relies upon a notion of mutuality in visual behaviour. Stare is an instrument of a dominant group member only so long as that member is aware of the subject of staring and that the subject in turn is aware of that stare. Consider again Sylvester's words at the start of this chapter. If the technologies deployed in VMC studies have not captured reciprocity, for example, then the functions that are subserved by mechanisms relying on reciprocity will be disabled. Gaze loses relevance when disjoint from its referent within the space over which it is cast. It is one thing to say that visibility is important in interaction and another to say that providing disembodied moving head-and-shoulders pictures is important in interaction or even that in real life it's people's heads and faces that count. Prosaically, if the "rainbow" is thought of as the projection of gaze and "gold" as its target, there might be gold but to date there has been no rainbow. This is a similar argument to Nardi and Whittaker's "video as data" proposition but requisite of a relationship between gaze and referent across a VMC link as well as within it. It is also firmly based on the idea of gaze distribution and interpersonal space, whereas video-as-data simply suggests that the work and placement of people and things revealed through a video image is where the value of video is to be found.

There is a wealth of evidence to suggest that proximity and its correlates are strong determinants of some basic social processes, and that they interact with gaze and speech activity (Argyle, 1988; Rutter, 1984). Furthermore, consideration of collaborative activity would also implicate some aspects of the distribution of parties. For example, Hutchins and colleagues have shown how small teams of co-workers integrate their activities in a number of ways, including a mutual knowledge of what counts as a shared artefact in the vicinity (Hutchins, 1994; Hutchins & Klausen, 1991; Seifert & Hutchins, 1992).

Matters of this kind can only be resolved empirically. The technologies described have not been designed to support the kind of mobility that underlies much of the group interaction research. What do "approach and retreat" mean when the parties are positioned in fixed chairs perhaps switching between views of remote interlocutors? Or proximity with varying sized monitors and zoom lenses?

These three high-level accounts of VMC failure do not convincingly demonstrate that a view of someone else's face adds nothing to interaction. A further account might be that the research paradigms adopted by investigators have simply missed their mark. One should treat such a proposition with caution, given the variety of evidence on offer. However, the picture for VMC is better described as confused than uniformly negative (Finn, 1997). The remainder of this chapter is given over to an historical-functional consideration of video-mediated communication, as a joint case of assessing technologies and interactive behaviour.

1.3.2 A functional-historical consideration of interactive behaviour in VMC

Studies of mediating technologies, including audio and video transception, are commonly described as investigations of "video-mediated communication", or VMC. Thus far, the use of this abbreviation has passed without comment. This thesis contests that there are some aspects of interactive behaviour that have not been adequately understood by such studies. There is a sense in which the lack of reference to audio in VMC systems, on the one hand, and its categorization as a matter of communication rather than as a fuller notion of interaction on the other, could be read as a full statement of their inadequacy. Interaction denotes not only meaningful exchange between people but also the tools and resources in relation to their activities. Interaction further describes something of the joint effects of the parties brought together by such technology as well as these materials in shaping an integrated activity system. From this point on, the abbreviation "VMI" will be adopted, to denote audio-video mediated interaction and interactive behaviour conducted over audio-video links. It is thus intended to make clear the necessity of conceptualizing video-mediated communications as integral with the purposes, people, resources and tasks they support. The following treatment of VMI research seeks to draw out of something of the implicit model of device use, and hence of interactive behaviour, maintained by investigators.

One of the earliest implementations of a video link immediately defies the suggestion that conceptual sophistication of interactive behaviour has developed with technological opportunity. It was a part of the Englebart & English' "research centre to augment the human intellect" (RCAHI). Twelve terminals were connected to a time-sharing computer as a feasibility study for technologies to support cooperative work. Their express aim was not to examine the benefits of particular technologies but to explore alternative ways of:

> "conceptualising, visualising and organising working material, and in procedures and methods for working together individually and cooperatively." (Englebart & English, 1968, p.396)

In association with personal and group computational workspaces, and accompanied by an audio link, a composite display was provided allowing a remote collaborator to be seen at the same time as the information under joint consideration. The data and remote party images were combined with scan converters such that the face of the remote person was seen through the graphical content of the screen.

Englebart and English claimed that this arrangement promised to "enrich system service." There was no psychological evaluation of this system. On page 409 of the paper, the investigators state rather disarmingly "we have experimented with these techniques and found them to be very effective." Looked at in one way, RCAHI was an innovative and radical approach to reconceptualizing remote communication as fundamentally related to the materials of joint interest. In another way, it was heaping together any and every communications technology for no better reason than that it could be done.

From a functional-historical perspective, this work is worthy of note for two reasons. Firstly, it has proven highly influential, significantly anticipating research during the late 1980s and into the 1990s. The second is that it neatly encapsulates a problem for behavioural research involving high technology: until an artefact is invented its use cannot be investigated by human scientists. For engineers, the challenge is to make the technology work; the so-called 'feasibility study', or proof by demonstration, and behavioural matters are less at issue.

This section continues by examining investigations of video-mediated interaction in sections according to levels of conceptual sophistication. These range from those imagining an audio-visual communications system as a stand-alone facility in its own through to those conceiving of video communication as an integrated part of a modern digital environment.

1.3.3 Televideophones for seeing the person

Moore et al. (1975) examined a potential role for television and telephone to support remote medical consultation. Television, it was felt, should improve the "quality and effectiveness" of consultation, thereby supporting a greater decentralization of manpower than the telephone alone. Nurse practitioners were stationed in three separated clinics in neighbourhoods remote from a hospital. Each clinic was provided with medical supervision by consultants at a hospital via microwave telecommunications links and conventional telephone lines, the microwave transmissions supporting a "television" link. No information is given about the size or scope of the images used or the picture quality of the image itself. However, one might reasonably assume that 1970s' broadcast quality was achieved. Over a seven and a half month period, 1408 encounters at the clinics resulted in 354 remote consultations. Of these, 234 were by television and 120 by telephone (there were twice as many days assigned to television consultation as for telephone, hence these figures represent equivalent usage of both media). Variables included: the length of time taken per remote consultation; the number of consultations requested by nurses for each medium; the kind of consultations carried out in each medium, and the number of referrals for further medical attention following remote consultation. Time was broken down into four phases of activity: initial dealings with the patient; contact and waiting time; remote consultation itself; and nurse-patient contact time following remote consultation.

Irrespective of the kind of medical case, no differences were found in the clinical performance of consultation by mode of contact, other than that television consultations were less likely to be followed by immediate referral to hospital than telephone consultations. However, consultations by television took longer than by telephone (seven minutes rather than five minutes). Additionally, more time was spent with the patient before consultation on 'television days' and there were longer delays before contact was established. This was explained in part by the unwieldiness of the technology and in part by other, unquantified factors:

> "The ability to see and interact with all three parties to the consultation seemed to increase the consultation's social and medical content ... The ability to visualize and interact with all three parties to the consultation may have helped to decrease the sense of isolation created by the geographic separation between the physician and the nurse and patient." (Moore et al., 1975, pp. 731 - 32)

All participants (patients and medics) expressed a high degree of satisfaction for both telephone and television consultations. In sum, they could find little evidence that either effectiveness or quality were enhanced by the television. Interestingly, no change in the pattern of use of the technology was observed over the study period, confounding expectations of a "learning curve" in video-link deployment.

Dunn, Conrath, Bloor and Tranquanda (1977) also looked at video-mediation to support remote consultation. Over six months, approximately 250 encounters were compared for each of four kinds of link: colour, black-and-white and still-frame black-and-white television (all with audio), and a hands-free (loudspeaker and microphone) telephone without a visual component. Data were gathered on referral requests plus additional medical test requests (categorized as either essential or helpful), time taken for diagnosis, and diagnostic accuracy considered against two levels of diagnosis (primary and secondary) and a complicated diagnostic category system. Advantages were conceived of as "effectiveness", defined as diagnostic accuracy, and "efficiency", or time taken to make a diagnosis.

The remote consultations were staged, rather than as a genuine part of patient care: patients who were to see a regular physician at the remote clinic were recruited in the clinic waiting room. They were ushered to a booth by an attendant nurse who was not recorded as playing a part in the consultation. Patients knew that this consultation procedure would not contribute to their personal care. Also, whereas Moore et al. looked at the whole patient contact time, finding substantially and significantly more time spent in the VMI condition, Dunn et al. considered purely diagnostic time, and found no such mode difference. Patient management was recorded by the remote consultant for comparison with the real patient management decided upon by the clinic physician. It should be noted, however, that Dunn et al. did not report the kind of referral requested in each mode, the only favourable difference for VMI recorded by Moore et al., but simply the extent to which it agreed with the clinic physician.

It would seem that Dunn et al. attempted to reduce the process of medical consultation to an information gathering and problem solving procedure:

> "It must be stated emphatically that in this study no attempt was made to measure quality of health care. All results were compared with the actual care received by the patient as standard" (Dunn et al., 1977, p. 23).

Moore et al., intentionally or not, included all aspects of patient care in their study. Dunn et al. found no difference in patient attitude between the four telecommunications media. Overall, no differences in performance were found between communications media for any measure taken, regardless of who the consulting physician was or the type or complexity of the presenting medical case. An exception was for diseases of the skin and subcutaneous tissue, where black-and-white image media were found to be more effective than both colour television and hands-free telephone alone. Again, no "learning curve" was detectable. Although concluding decisively against VMI systems, Dunn et al. nevertheless observe that there was some indication that physicians practice differently when using different media, including both diagnostic method and patient management practices, even within the highly restrictive form of diagnosis examined. It should be noted, however, that for the audio-only condition, Dunn et al. were not using a conventional handset telephone. All those in earshot of the hands-free loudspeaker could benefit from involvement in the interaction, thus audibility was partially confounded with visibility.

The investigative programme of the Communication Studies Group (CSG) is described in a number of reviews (Short et al., 1976; Williams, 1974; Williams, 1977). Given the coherence and relative sophistication of this body of work, and despite over twenty years separating it from other work considered here, it warrants an extended consideration.

Many of the investigations were modelled closely on a white-collar, business conception of work. Williams, in summarising the 29 studies of telecommunication substitutions for co-presence, concluded that the adequacy of the technology in this regard is highly contingent on the nature of the meeting to be substituted (Williams, 1974). In other words, some kinds of meetings require more of certain aspects of communication than others, and that these aspects of communication are not supported by some media. She proposed a four-stage process for determining the typology of the meeting room, based on Reid's Telecommunications Impact Model[6] (Reid, 1971; Short et al., 1976): 1. typing the face-to-face meeting; 2. allocating appropriate media; 3. measuring frequency of meeting types; 4. calculation of proportion of face-to-face meetings which might be transferred, given (2) and (3). The majority of meeting types included in Reid's model were supposed to be quite adequately supported by audio links and no necessary role anticipated for seeing the remote parties.

Several questionnaire-based studies are cited as strong supporters of this explanation of benefit. The earliest of these recruited 72 managerial civil servants in pairs to choose between alternative courses of action, based on their perceived risks, using each of the three media successively (Champness, 1972a) The communications conditions included face-to-face, closed-circuit television (CCTV) or an audio system. After each one, they were asked to complete questionnaire, rating their experience on 24 semantic differential scales[7]. On 20 out of the 24 scales, audio was statistically differentiated from the two visual media. Responses were factor analysed. The first factor included seven scales: colourless-colourful; small-large; constricted-spacious; boring-interesting; ugly-beautiful; unsociable-sociable, and insensitive-sensitive.

---

[6]Reid's model was intended to aid forecasting changes in organizational structures, by predicting which aspects of organizational contact might be altered by the availability of new telecommunications technology. It did not address the effects of technologies on the communications he supposed might be supported by them.

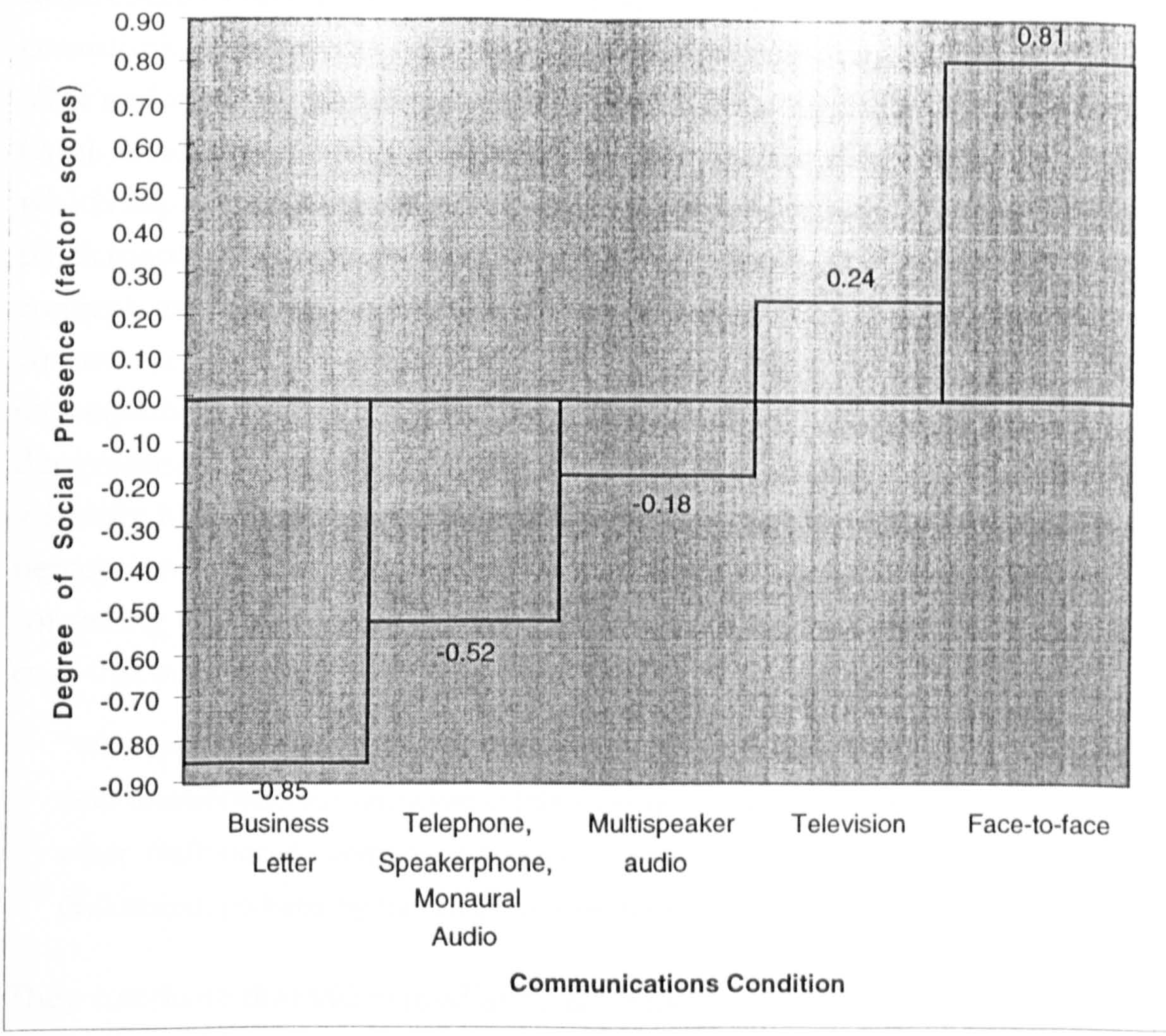
[7]The scales were initially derived from Snyder and Wiggins' 1970 methodological study of "Affective meaning systems: a multivariate approach" (*Multivariate Behavioural Research, 5,* pp. 453-68).

Short, Williams and Christie reinterpreted Champness' findings in terms of their position on Social Presence. They identified the first factor as a composite of Social Presence, implicating it as a major determinant of media effects, and another dimension they described as "Aesthetic Appeal". However, only two of these scales were taken to represent Social Presence (unsociable-sociable, and insensitive-sensitive); thus this factor did not stand out as uniquely identifying the visual media. Furthermore, four of the scales differentiated VMI from face-to-face communication and only one of these was identified with Social Presence. The remaining three were true-false; meaningless-meaningful, and public-private. No particular interpretation is placed on Aesthetic Appeal, in spite of its primacy in the factor analysis.

In the study described above, pairs interacted over closed-circuit television. As a partial replication, 90 civil servants in groups of three performed the same, decision-making risk assessment task with the same equipment (Champness, 1972b). Two people were at one end of the link and one person was at the other. Whereas participants in the first study saw one another from much the same point of view, this time the co-present pair saw a head-and-shoulders view of the lone participant and the lone participant had a wide-angled view of the pair, including the individuals and their seating arrangements. It was expected that the less detailed view would be associated with lower Social Presence than the head-and-shoulders view. Indeed, responses to Social Presence items were significantly higher for those who were co-present than the lone, distant party. However, this evidence may not be as closely associated with the image characteristics as Short et al. suggest: it is unclear to what extent the very co-presence of the pair of participants may have contributed to these responses, independently of the interaction with the remote party and hence the technological mediation. They do not report, one way or another, whether a similar disjunction of response by location/co-presence was obtained in the audio condition.

Three other CSG studies are discussed, in the same, factor-analytic vein. These consistently demonstrate the emergence of a primary factor containing items consistent with the authors' Social Presence concept. More significantly, the Social Presence factor so obtained was associated with an increasing trend in factor score on media conditions of increasing sophistication. These ranged from a business letter, through various monaural audio devices, multi-speaker audio, and VMI to face-to-face.

Figure 1: Degree of Social Presence for several Communications Media, after Short, Williams & Christie(1976, p.71)



Differences between adjacent media in this list rarely reached statistical significance, where adjacency is determined by Social Presence factor score, although the such differences between non-adjacent media were frequent. These are represented in Figure One, as a combination of factor scores obtained in two of the studies referred to above.

The status of the audio-video link on the Social Presence dimension was less stable than the other media and Short et al. note:

> "The value for the video medium is very approximate and is likely to vary markedly depending upon the conditions of testing (e.g. the size of the image on the screen)...Whether or not we would expect differences between the video medium and face-to-face would seem to depend upon the details of the experimental conditions, including, for example, the size of the subjects' image in the television picture." (pp. 71 - 72)

Given the homogeneity of the subjects and tasks used for these studies, it is rather surprising that the results obtained were not more consistent.

An account of Social Presence, as the factor identified in these studies, based purely on visibility was not intended by Short et al. They demonstrate differences in factor scores for Social Presence between non-visual communications media equivalent to, if not greater than, differences between VMI and some audio-mediating communications technologies (see Figure One). A notable differentiation was between multispeaker, spatialized audio, which supported a separate sense of the location for each of the remote participants, and various other kinds of audio mediation. The Social Presence concept, as a synthesis of all the factors associated with visual communications behaviours, is identified with whatever functional consequences visibility has for interaction. Such an approach sounds dangerously empty, as though Social Presence is merely a label for all that visibility has to offer, notwithstanding that it is attributed to individual perceptions as discussed above. Short et al. substantiate the concept as consisting of a number of empirically derived functions (as distinct from the cues that support them) of social regulation.

> "We hypothesize that Social Presence depends upon not only the visual non-verbal cues transmitted, but also more subtle aspects such as the apparent distance of the other (influenced, perhaps, by voice volume) and the 'realness' of the other (influenced, perhaps, by the fidelity of speech reproduction)." (Short et al., 1976, p. 75)

They conclude that video mediation does indeed have relevance to interpersonal communication processes, albeit conditioned by the task concerned. The task will be affected depending on whether or not it "requires" Social Presence and the extent to which the medium concerned supports it. This is somewhat problematic as Social Presence is by definition an attitude towards mediating technology. However, they go on to say:

> "We would predict that in tasks where the emphasis is not on the people involved, and where the outcome does not reflect on the personal qualities of the individuals communicating, the outcome will be unaffected by the degree of Social Presence of the medium. Such tasks would include, for example, information transmission and simple problem-solving tasks." (Short et al., 1976, p. 75)

So emphasis on the people involved and reflecting on the personal qualities of individuals is implicated in lieu of a technology to have an attitude about.

1.3.4 Visual and spoken interaction

Chapanis (1975) summarized a series of laboratory experiments on four types of technology for mediating communication: voice, handwriting, typewriting, and video, with a CCTV link coupled to voice. His aim was to understand how each of these communication channels affected performance, compared to standard performances achieved in face-to-face communication. This standard was described as "communication rich", and VMI was expected to approximate to face-to-face by offering a similarly rich "information bearing" communication for those who use it. Chapanis' (1975) tasks were set up with a conception of visibility being of value in regard of its informative content, and hence in principle to any kind of information, but without any clear notion of a social dimension to communication. This is succinctly demonstrated by a statement that the role of the 'information source' equally could be taken by a computer. The activity was conceived as a truly collaborative exercize, however uninvolved the material might have been in terms of Social Presence. Explicit reference was made to this point:

> "how do communication patterns vary according to the purpose of the communication? All our experiments so far have involved factual problems. The problems all had single answers, and the information needed to solve them was directed toward that one goal." (p. 232)

Although some detailed breakdown of communicative behaviour was carried out, the bottom-line measure reported was the time taken to solve the problem. This measure did not differentiate any of the voice media from one another, including face-to-face, microphone and loudspeaker alone, plus handwriting, and plus audio-video communication. Nine measures of "linguistic output" included: number of sentences generated, number of messages generated, number of words per message, total number of words, percentage of messages as questions, a ratio of different words to total words, and rate of word production. Again, voice media were found to differentiate from others in terms of the performance of production. They were higher on each measure bar the type-token ratio, taken to indicate that voice media are considerably more redundant than non-voice communications. No interpretation is placed on the value of redundancy, one way or another. Interestingly, preventing subjects from interrupting (taken to mean simultaneous speech rather than aggressive floor appropriation) changed this situation, such that voice transmission became closer to teletype on these measures than to face-to-face communication. More words were used per message and overall fewer messages were composed. Although no information is provided regarding the significance of the figures included in this article, there is an indication that "communication rich" conditions involve shorter, more frequent messages than other voice media.

This finding, described as a formalizing of speech, has been repeated in comparisons of telephone with face-to-face conversation (Morley & Stephenson, 1977), although at least one author has suggested that such changes are a consequence of the differing content of messages (Rutter et al., 1981). Given the low level of personal exchange in these tasks, it is difficult to see how communicative style by content could be consequent of content differences. However, Chapanis' measures were insufficiently specific to conclude one way or the other.

So, in common with the CSG indices of telecommunications impact, much of Chapanis' assessment of performance was concerned with the productions of those who carried out these 'factual problems'. Although some insights into the dynamics of mediated communication were provided, and indeed seemed to differentiate the media, no particular emphasis or interpretation was made of the causes or likely consequences of these process differences. Of perhaps greater concern, and also acknowledged, was the limitation of his work to dyadic interaction. A suggestion that all voice media are equivalent based on evidence from two-person studies is considerably less than reliable.

Champness (1971) compared face-to-face and telephone conversation, and found differences of a similar order. Pairs discuss the fate of an elderly production worker, holding up factory output by face-to-face conversation and by telephone. Chapanis' outcome measures, including nature of the solution, satisfaction with solution or agreement between individuals, did not differentiate the groups. On the other hand, process differences were found between face-to-face and audio conditions. The length of utterance of each participant was more similar in face-to-face than in telephone interactions, suggesting interactive turn coordination was operating under substantially different constraints in each case.

Rutter, Stephenson and Dewey (1981) conducted a study to try to isolate the role of physical presence from visibility, as influences of interpersonal and interparty matters in tasks of persuasion. They were interested in the reports of formality in audio-only communication, associated with task-focused, unspontaneous and depersonalized speech, as well as opinion change. Their agenda was not set by designing for remote communication but simply to add to the research on non-verbal communication. The first of two experiments examined communications exchanged between blind people, either one metre apart at 45 degrees or in separate rooms, communicating via microphone-headsets. The 'face-to-face' pairs were video recorded and all pairs audio recorded.

They found no difference in outcome, in terms of opinion change. The audio recordings were analysed for evidence of the kind of formality described above. They coded the fifteen minute discussions with a scheme called Conference Process Analysis (CPA). CPA represents a significant analytic departure from the kinds of measure employed by the investigators of the early and mid-1970s. It is concerned with the *manner and mechanism* by which the communication proceeded. CPA coded for a variety of measures associated with communicative style, such as formality and spontaneity, and content, including self- and other-referring speech, versus task-oriented speech. There was no evidence of the blind subjects in the audio condition using different style or including different content in the things they discussed. The only significant difference obtained was in the 'spontaneity of communication' but in the opposite direction to that hypothesized. More 'speech disturbance', indicated by more frequent and longer bouts of overlapping speech, was found "face-to-face" than in audio. Rutter et al. attributed this to the situational anxiety induced by the knowledge of video recording. The picture is confused, as the data included in the paper also indicate that there were longer mutual silences in the audio condition than face-to-face, but that there were more and shorter utterances in the audio condition. An important consideration here is that the blind people who took part in this experiment were highly practised at communicating in the absence of visual information of any kind. Had sighted people been, for example, blindfolded for the same experiment, the findings would perhaps have been confounded by the social context of a "ridiculous manipulation", as Williams would have it (1977). It nevertheless stands out in the literature as an indication that physical presence in itself adds little or nothing to communication, in the absence of vision and the spatial information that may be derived from the visible characteristics of interlocutor environment.

In the second experiment, sighted subjects conducted the same task in the same conditions as those in the first experiment with two additional manipulations. In one of these, pairs were positioned in adjacent rooms separated by a curtain, as Chapanis and Ochsman had arranged, and in the remaining condition , CCTV cameras relayed head-and-shoulders views of pairs of subjects to one another in an adjacent room. This follow-up experiment did reveal content and style difference between audio-only and face-to-face, but not evidence of differences between the additional and original conditions: the curtain and VMI conditions were said to lie "somewhere *between* the audio and face-to-face".

Rutter et al. took this to mean that they each were superior in some sense to audio but inferior to face-to-face. Visibility was seen as only one of several contributory factors to distinguish face-to-face from audio-only interaction, such that the determinants of the content and style of communicative behaviour were a range of cues with additive effect. It is difficult to interpret this finding in relation to any difference that may have been found between audio-video and audio-only conversations. The purpose of the ANOVA procedure employed to obtain their significant difference is to partition variance in scores obtained against the factors included in the analysis. That the mean scores of VMI and curtain conditions lay between the audio and face-to-face scores did not allow Rutter et al. to reach any kind of conclusion about the relative role of these conditions in mediating communication. It means the variability of the scores was too great to distinguish their influence from chance. They went on to re-examine the CPA scores for a relationship between their content, style and outcome measures:

> "audio outcomes differ from face-to-face outcomes because what the participants *say* is different in the two conditions. That is, cuelessness leads subjects to be task-oriented and depersonalized in the content of their discussion, and it is content which, in turn, produces the different outcomes." (Rutter et al., 1981, p. 50-51)

The pattern of correlation in CPA scores leads them to conclude that a combined effect of a variety of cues determines the style of speech, and thereby alters the balance in speech content and (potentially) the task outcome.

1.3.5 Video as an integrated technology

The advent of computer-to-computer networked communications has led to a new perspective on the role of computational media, as a part of a battery of mechanisms that might be deployed for a variety of purposes. In a networked context, VMI is conceived as an integrated part of a computerized infrastructure for groups at work, much as envisaged by the Englebart and English workstation. That is, VMI is anticipated to be recruited along with a variety of other resources in support of collaboration. The distinction from earlier studies is subtle: latter-day VMI studies are almost inevitably in the "integrated" rather than stand-alone mould, simply by virtue of the advent of the ubiquitous PC in work and internetworked digital communications. The H.324 system for video communication over a range of networks, including analogue telephone lines, also covers real-time audio and data transmission, described as an "international standard for multimedia communication" (Lindbergh, 1996).

The distinction is worth making as the whole idea of communication is radically altered when set into this kind of framework. An interaction might be something it makes sense to conceive of over a period of days or months, or simply two or three words exchanged without the stereotypical opening or closure sequences (Fish et al., 1992; Frohlich, 1995). To borrow from Dix's terminology, the "pace" of interaction is conditioned by a variety of factors crucially governed by systemic environmental constraints (Dix, 1992). In this case, the availability and rate at which communicative exchanges can take place is fundamentally altered by the simultaneous availability of a range of communications media amongst which choices are routinely made (Daly-Jones, 1998).

Gale set out to consider whether there would be a cumulative benefit of adding audio, and audio and video together on a facility for sharing data between computer workstations (Gale, 1990). He measured both the time taken to complete a task and the quality of the work completed, finding no evidence of a gain from either additional communications channel. His subjects additionally completed questionnaires and, although perceptions of productivity increased with the "extra bandwidth" of audio and video, they consistently gave low ratings for usage of the video component. Posthoc, Gale examined recordings of their behaviour and found that, confusingly, subjects had frequently used the video link.

Where more than two parties are engaged in conversation, the regulatory burden of conversation is considerable increased. Dyads admit only one possible new speaker at any point and attention may be fully focused on a single interlocutor (Daly-Jones, 1998). Cohen (1982) used Bell's "Picturephone Meeting Service", a voice-activated video switching system, to allow eight people to take part in a video-mediated discussion. Picturephone sensed when one of the participants began to talk and changed the view seen by other group members to that given by a camera pointed at the new speaker. She compared Picturephone with co-present groups on certain metrics of the structure of conversation, including turn length, number of turns and simultaneous speech. Picturephone interactions were found to involve less simultaneous speech and fewer turn exchanges than face-to-face interactions. Cohen claimed that these results showed that video-mediation promotes more formal, less spontaneous interaction. Sellen (1992) points out that Cohen's decision to impose a 700ms delay in audio switching time, corresponding to real-world transception latencies, clouds the issue considerably. Such a delay would have had a marked effect on Cohen's measures, regardless of any effect intrinsic to mediated visibility.

O'Conaill, Whittaker and Wilbur made an explicit attempt to differentiate the effect of mediated visibility from transception delay by comparing a high-bandwidth (dedicated fibre-optic network) analogue system with a digital system (O'Conaill et al., 1993). The digital system was designed to run on two 64kb/s Integrated Services Data Network (ISDN) lines and was associated with transception delays of at least 820 milliseconds, a figure comparable to Cohen's VMI condition, and up to 1560ms depending on transmission medium. In addition, the digital system's audio operated on a half-duplex protocol (either transmission or reception at a node at any juncture) whereas the analogue system also had full-duplex audio with negligible delay. Five everyday, real-work meetings were recorded and transcribed for the digital system and also for face-to-face contexts as a control condition, and four similar meetings mediated by an analogue VMI system. The second twenty-minute period of each meeting transcript was subject to a microanalysis for "conversational features", including auditory backchannels, interruptive (unsignalled) and overlapping (signalled) simultaneous speech, turn length in words, formal handovers and floor dominance.

Sample sizes precluded statistical analysis but data suggested that the analogue system allowed similar conversational structure to the control condition on turn length and interruptions, whereas interruptions were very rare and turn length about double the face-to-face and analogue figures. Backchannels with the digital system occurred at approximately a tenth the rate of the face-to-face meetings and a quarter of the rate of the analogue system; analogue involving about half the number of audio backchannels compared to face-to-face. Overlaps and dominance did not distinguish the conditions. Formal handovers were of the same order of magnitude across all conditions, with a suggestion that the video-mediated meetings involved slightly more of these. Overlaps were expected to markedly differentiate the ISDN meetings due to the switching delays. Additionally, the analogue system's video-image quality was reportedly far better than the ISDN video presentation, leading to expectations of a greater need for explicit handover due to speaker identification problems.

As real-work meetings, it is perhaps unsurprising that there was a degree of formality about the conversational structure data. Indeed, the face-to-face control suggests that an appropriate level of formality, in terms of the reported measures, was reflected in the analogue meeting exchanges. Subjective reports were that both VMI systems were more effortful than their face-to-face context. O'Connail et al. note that the physical arrangement of participants in the VMI conditions was such that attending to co-present parties was not as easy as the face-to-face context. Rather than being located around a meeting table, with a variable requirement in physical movement to attend to one another, the VMI conditions involved all participants being seated at 90 degrees to a flat display screen. Interestingly, informants were said to have invested more time in preparing for the VMI meetings than for face-to-face encounters. The implications for the value of real-world performance comparison of mediated and conventional meetings are obvious.

Sellen was also concerned with supporting communication between groups of three or more at separate locations. She carried out two separate experimental comparisons of VMI systems for this purpose (Sellen, 1992; Sellen, 1995). The first extended the metaphor of camera and microphone acting as "surrogate" eyes and ears at the remote location by mounting them into slim vertical racks as "minitowers" with integral display VDUs. The towers then acted as individuated surrogate units from the viewpoint of the person sharing a space with them. Sellen argued that if gaze was normally important in regulating turn transfer in conversation, it ought to be important, in her case, to tell to which of the other three participants a speaker was attending. This "HYDRA" system was compared with a split-screen picture-in-picture system (PIP), as an equivalent system without interpersonal gaze support, and also with same place interaction. Twelve groups of four discussed a separate contentious issue successively in each of the three experimental conditions.

Sellen also adopted measures of speech patterns as indicators of conversational structure and placed a high value on simultaneous and continuous visibility during interaction. Fast rates of exchange transfer, short turns at talk and an incidence of overlapping speech were thought of as demonstrating effective support for highly interactive, multiparty communication. Since HYDRA was designed to convey interpersonal gaze, it was anticipated to be associated with speech patterns more similar to the co-present group than groups using the PIP system. However, PIP and HYDRA were not found to differ from one another on any of Sellen's conversational structure measures. The co-present interactions did not differ from the VMI conditions on number of turns per session, turn duration or a measure of within-group turn distribution (how evenly turns were taken by different group members). However, the co-present interactions were associated with more simultaneous speech and overlapping turn exchange than interactions from either of the VMI conditions. A further comparison of the video-mediated technologies with audio-only mediation failed to differentiate communication on these measures.

That these VMI communications systems supported group interactions with conversational structure similar to face-to-face interaction is noteworthy. However, it is difficult to claim that Sellen's data represent a success for VMI, as to do so would require an argument from the null hypothesis. Furthermore, the method of speech measurement was automated on the basis of a preset trigger for utterance loudness. Turns were defined to exclude all utterances of less than 1.5 seconds, to ignore silences mid-utterance of less than 200 milliseconds and to include all silences prior to the next speaker taking the floor. The adequacy of such a method is at the very least constrained to provide data on longer, fully-voiced utterances and curiously at odds with the intention of delivering insight into interactivity.

1.3.6 Integrated video and object gaze

Drawing in part on Tang and Minneman's studies of collaborative drawing (Tang & Minneman, 1990; Tang & Minneman, 1991), a different slant on the value of video has been developed by Ishii. Ishii sees video as an intuitive attentional cue, not so much for conversational management but as a deictic gesture and thus only making sense when enough of the relevant environment is also included (Ishii, 1990; Ishii & Arita, 1991; Ishii & Kobayashi, 1992). Ishii capitalized on the notion of translucent overlay, taken from Englebart and English's RCAHI, in application to drawing surfaces. As a first development, TeamWorkstation included separate, inset head-and-shoulders participant views alongside a large view of a drawing surface with overlaid hands. Ishii was specifically interested in allowing seamless interactions between views of the workspace and views of the people, with coherent relations between gaze and objects of gaze. ClearBoard achieved this by using combinations of video cameras and projectors to integrate views of participants' faces with the drawing surface to support gaze awareness, i.e. knowing where someone else is looking by treating gaze as a deictic reference. Little in the way of evaluation is reported of these systems.

Barnard, May and Salber (1996) have experimentally evaluated the effects of camera angle arrangement on joint work, where the work involves some visible referents. Their argument was closely related to the assumptions underlying Ishii's work: that resolution of visual referent is possible on the basis of gaze direction and that the process is highly interactive and automatic. They hypothesized that communication effectiveness would be related to the degree of congruence in the spatial relationships between the viewpoint of collaborators and their views of one another in relation to the objects of work. In principle the relative transformations may be learnt, such that the viewpoint of one participant may easily be used to communicate some desired manipulation to a second. Barnard and May asked subjects to perform several tasks together on a computer workstation including a video graphical window positioned below a second window, showing various graphical objects. Three conditions all varied the remote participant view and included a direct video signal from a camera mounted above the remote workstation VDU, a horizontally mirrored signal (left-to-right; right-to-left) and a view from a camera position behind and to one side of the remote person. They found where camera arrangement affords a view that is coherent with respect to the workspace, interactions were associated with less directional speech ("left, right", etc.) and fewer false starts and errors.

Kuzuoka was also much taken with the idea of video conveying data, as a consistent and intuitive way of reciprocally understanding viewpoint and gesture (Kuzuoka, 1992; Kuzuoka & Shoji, 1994). He anticipated a requirement for collaborative systems that would translate into a physical workspace and so be able to support three-dimensional coordination. Kuzuoka (1992) systematically manipulated the viewpoint of a fixed camera, finding progressive degradation in performance again measured by time to completion and also linguistic output (descriptive words of a number of types). Whereas Ishii hoped to allow users to resolve their gestures and viewpoints by mimicking separate presences in a physical workspace, Kuzuoka went on to attempt to place two people at a single presence point in space by using a head mounted camera to show a remote "instructor" where a local "operator" was looking.

The camera was servo driven so that it would turn about the operator's helmet to adjust for eye deviation on a simple principle, derived from observational data relating head and eye deviation to point of focus. This "point of focus" view was displayed to the instructor who was able to interact with the operator via the view as a set of shared referents. The video signal then assumed the role of a shared eye and hence provided, in principle, absolute visual coreference to a space. Meanwhile, a second camera at the instructor's location transmitted a view from behind the instructor and instructor's display to send the instructor's hand movements back to the operator's head mounted display. In this way, absolute view coordination was coupled with absolute gestural reference. Although this work was in the main exploratory and inappropriate for statistical analysis, a microanalysis in much the same mould as Tang and Minneman's (1990) study of collaborative drawing, provided some insights into work with SharedView.

A system using the SharedView principle has been compared with face-to-face operation and audio-only (Kuzuoka & Shoji, 1994). Pairs completed manipulation and assembly tasks, including manipulation of a set of model objects, controlling a machining centre and controlling a milling machine. The model manipulation tasks were crossed with an additional independent variable: completion with and without gesture permitted to assist instruction. Tasks were carried out more quickly where gesture was allowed than where not, with video-mediated and face-to-face conditions not otherwise differing on this measure (there was no audio-only condition for the model tasks).

Transcripts of all interactions were coded for speech, gesture and gaze focus and indicated a consistent pattern of operation for each stage in coordination between instructor and operator. These steps in coordination were taken regardless of available media but required different behaviour as a function of their availability. The milling centre interactions took place over audio-only or audio-video links. Interactions with video were associated with fewer breaks in instructor utterance than audio-only and less operator utterance in general.

On the face of it, finding longer turns and fewer floor exchanges in VMI compared to audio mediation is perplexing and at odds with most other findings (cf. O'Connaill et al., 1993; Cohen, 1982). As O'Malley et al. (1996) have observed, it is not without precedent and may be due to differences in the task demands and technologies employed by different investigators. Indeed, Kuzuoka examined the nature of the interaction involved in his studies in detail to understand why his data had taken this form. The speech pattern change was taken as evidence that the instructor was able to continue to give directions on an operation, and have received confirmation of instruction receipt and understanding by monitoring the remote space. In other words, maintenance of common ground was insured by visibility of operator action via the video images.

A view of interactivity is in evidence in Kuzuoka's work which demonstrates that the temporal context of gaze deployment, in relation here to the task characteristics of instructor-operator interactions, is a fundamental determinant of other contemporaneous activity. It also shows the emptiness of simple, unqualified statements about "formality" of speech from the so-called "conversational structure" metrics so commonly employed. A cycle of activity through various potential objects of attention was observed to impact on communications requirements and the ability of available media to support them. People are required to respond to constraints on their mutual activity, in collaboration and flexibly, depending on stage of work and available resources:

> "A problem in groupware is the handling of irregular human actions. For example, in the current groupware using video communication, it is difficult to respond to unexpected events because cameras and displays are fixed. To support spatial workspace collaboration, it is necessary to analyse and to clarify the details of human behaviour. Groupware design must be centred around such human behaviour so that natural human reactions are accommodated rather than suppressed" (Kuzuoka & Shoji, 1994, p. 67)

One of the issues commonly reported informally from post-experimental debriefings and the like, concerns the achievement of mutual gaze. Video communication systems typically comprise a camera mounted above or to the side of a VDU. When, for example, a listener looks directly at the eyes of a remote speaker presented on the VDU, the listener is necessarily looking below or to one side of the camera lens. As the camera lens is the viewpoint of the speaker, this disparity is apparent and, at least at first, disconcerting. Smith et al. (1991) reported a means of achieving an illusion of eye contact by furnishing participants with a device known as a "video tunnel": a camera is positioned to pick up light reflected from a half-silvered mirror positioned at 45 degrees about the mid-line of a display. The observer is able to see images on the display through the mirror and the camera transmits the reflected image of the observer.

O'Malley et al. (1996) report experiments where a video tunnel was used to support completion of a standard spatial coordination task, involving collaborative navigation from point to point via maps with different but intersecting sets of features (see Anderson et al.,1991). The three experiments they report are highly material to the current discussion in that they relate a surprising disparity between co-present and video-mediated interaction and do so with care for the nature of the task performed. Boyle, Anderson and Newlands (1994) had shown that dyads carrying out the map task required fewer words and turns, and interrupted one another less when they were able to see one another's faces compared to having to rely on speech alone. No differences in accuracy or time to completion were found. It was an interesting result as in neither case could each see the others' map, the obvious candidate for video mediation on a video-as-data account. O'Malley et al. compared audio-only, video-tunnel supporting mutual gaze and a video-tunnel adjusted to preclude mutual gaze. They also found no task performance differences but more interruptions in the VMI conditions and more talk in the video with eye contact conditions. Boyle et al's comparisons were with co-present pairs but it is nevertheless difficult to square their findings with the O'Malley et al. data. A second experiment compared audio-only with an upper torso view and a view of smaller scope, just head-and-neck, presented on the same video tunnel display and hence objectively larger. Findings partially replicated the first experiment, where visibility was associated with no task performance difference compared with audio and with the production of more words when subjects could see one another. Turns and interruptions failed to reach significance but variance was high. The smaller scope/larger image dyads also used more words in discharging the task than the larger scope/smaller image dyads. A third experiment compared audio-only with video-tunnel mediation, as before, but also with a commercial low-bandwidth, small display, long transception delay videophone device in audio-only and audio-video modes. The delayed system conditions resulted in a 36% deficit in solution accuracy, regardless of the presence or absence of video, a performance difference for the first time. Visibility was again associated with greater linguistic output than audio-only, however interactions over the poor audio-video link involved a much greater level of interruption than either of the video-tunnel or audio-only conditions.

O'Mally et al. attribute these differences entirely to delay, reflected in an explosion in the interruption rate (up to 50% of all utterances). Certainly, the delay differences are extreme and the evidence suggests that the style of interaction adopted was incompatible with the ability of the system to convey speech. The delayed system could not cope with the level of interactivity required to coordinate sufficiently on the incremental building of a passage through the map's marked terrain. Some adaptation was in evidence, as a low rate of back-channel speech, but subjects were unable or unwilling to adapt their style of communication sufficiently to perform the task adequately within the constraints of the medium. One should be cautious about interpreting this result as evidence of video being wholly ineffective at compensating for poor audio. The video here was not just delayed but also of marginal quality in a range of other respects, not least of size, given the result of the second experiment.

These three experiments together tell a consistent story of subjects being more verbose when they can see one another at a distance compared to when they cannot. Given such a constrained task, that these findings disagree so markedly from Boyle et al.'s results for the same task is very surprising. O'Malley et al. are firm in their interpretation of the amount of linguistic output as a direct consequence of and in proportion to communication difficulty or failure. Yet how reasonable is such an assertion? Their data show that subjects are performing the task perfectly well. Might it not equally be that subjects in the VMI conditions are able to achieve satisfactory performance and still have sufficient communicative capacity to engage in prosocial, informal communications?

A Conversational Games Analysis failed to differentiate video-mediated and face-to-face dialogue structures but this system involves strictly informational coding. Examination of interpersonal gaze showed far more time (56%) was spent looking towards conversational partners in the VMI conditions than in the face-to-face conditions. O'Malley et al. then suggest that this counts as 'overuse' of the visual channel, contributing to difficulties processing verbal information and increased cognitive load.

They do not consider that their version of the map task was implemented on computer workstations and video monitors in a common vertical plane, whereas Boyle et al. used paper maps on horizontal surfaces with participants sat across a table from one another. The head movements involved in the Boyle study and the retinal detail of visual cues would have been far greater and hence more salient than those in O'Malley et al.'s video-tunnel, let alone videophone, configurations. In the context of the current discussion, their final comments are of more interest:

> "simple measures such as number of turns, pauses an interruptions are not sufficient to explain the relationship between verbal and non-verbal information in different communicative contexts." (O'Malley et al. 1996, p. 190)

In order to make sense of video-mediated interaction, it is not enough to describe communicative behaviour in terms of general distribution of speech but that it must be related to gaze behaviour within the material context of joint activity.

1.3.7 Video as an element in the definition of a collaborative space

The "media space" is another class of video-integrated system placing great value on seeing people with their physical context (Gaver et al., 1992; Gaver et al., 1993; Gaver, Smets & Overbeeke, 1995). A media space involves a series of networked connections to physical locations, often within a particular building, conveying a variety of information about the occupancy and ongoing activities. Typically, video is one of the available media. In a media space application, video again is seen to be of value for its capacity to convey gesture with reference to common artefacts. In addition, video can convey the contact availability of people in a space, a pre-communication issue so often missing from social psychological studies (Daly-Jones, 1998; Frohlich, 1995), concerns. In this way it serves to indicate *what* is going on as much as *how* it is going on. It is difficult to reconcile reports on the use of media spaces with the systematic comparative scheme adopted here. They are practical systems typically introduced in high-technology organizations to solve challenges to collaboration emergent of a wholly new way of working. In this sense they transcend traditional collaborations, affording users a hitherto inconceivable "awareness of their work environment" (Fish, Kraut, Root & Rice, 1992, p. 37).

BellCORE's CRUISER is a possible exception to this, as it was specifically developed to support and promote "informal communication" as a contrast to the formality of scheduled meetings often seen as the bedrock of commercial videoconferencing systems (Fish, Kraut & Chalfonte, 1990; Fish et al., 1992; Fish et al., 1993). This notion is primarily to do with spontaneous and frequent contact, typically of fleeting duration. Video provides a way of 'noticing' whether or not someone is available to establish contact in this way and of maintaining a sense of continuous presence. However, as Heath and Luff have variously reported from usage studies of the XEROX RAVE media space, video mediation can be highly insensitive to attempts to initiate contact (Heath & Luff, 1992a; Heath & Luff, 1992b). The viewer might perceive an opportunity to engage with someone but be unable to attract their attention. From a 'video-as-data' perspective, they also found that, once contact is established, even to gesture as an accompaniment to talk is highly problematic, at least via RAVE's 14" monitors. As Monk and Watts have recently discussed, in environments where there is a degree of personal movement, noticing and contributing to a conversation is also a continuous process with strict dependencies on audio and video reciprocality (Watts & Monk, 1997; Watts & Monk, In Press). In any case, media spaces are what they are by the combination of media they involve. Abstracting a special role for video from them remains a matter of speculation in absence of comparative data.

Following on from Sellen et al.'s HYDRA and the various media space systems, Vertegaal (1997; 1998) has attempted to support interpersonal gaze direction as part of a more general provision of "conversational awareness". He takes the unusual position that the burden on evidence marginally differentiating audio from audio-video mediated interaction is an argument for accepting video as a purely supplementary information source to audio channel. Positioning video as a subsidiary medium calls into question the enormous network resources it requires.

Vertegaal suggests that video information should therefore be conveyed in a parsimonious rather than a realistic manner, modelling a space of interaction within which participants are located (a so called "virtual space"). He has designed systems making use of canned still images as "personifications", both to substitute a video window in a conventional desktop-video arrangement and set into a two-dimensional representation of a three-dimensional space on single or dual VDU monitors. Audio has been mapped to separate speakers to add to the illusion of physical distribution. The stills are animated in response to signals from eye-trackers trained on each of the collaborators in their separate locations.

He is primarily looking for engineering solutions to support groups of people collaborating at a distance, with an understanding of the technological limitations of available resources, in much the same way as Englebart and English, reported at the start of this section. It is worthy of note in that it shows how two of the several functions of gaze distribution have been identified as doing the real work of seeing someone: attentional distribution amongst interlocutors, as a regulatory mechanism, and objects, as a projective deictic gesture.

1.4 Conclusion: an approach for Video-Mediated Interaction

This chapter has examined the literature on experimental studies of video-mediated communication, together with some of the pertinent findings from investigations of non-verbal behaviour. A critique of these experimental studies focused on a shortcoming with account of interactive behaviour presupposed by the measures employed, failing to address video link usage with adequate sensitivity.

Short et al. looked at the literature comparing face-to-face communication with communication mediated by audio-video, audio-only and teletype technologies, for a number of task types and group factors. These included information transmission, problem solving, conflict resolution, bargaining and negotiation, getting to know someone, and persuasion. They concluded that only tasks with a requirement for interpersonal or interparty considerations have found some significant effects of medium.

This conclusion is based on different assumptions that inform the question that sets up this thesis: whether or not a video image contributes anything to spoken communication. Differences were found between co-present or face-to-face interaction and interaction through technology with an explicit commitment to outcome measures, excluding any measures describing "the mechanics of interaction" (Short et al., 1976, p. 80). This commitment was justified by asserting that process effects are only of interest if they impact measurable performance. The question then becomes whether or not the performance measures employed are sufficiently sensitive to detect such impact and whether or not the investigative context for the deployment of these measures affords the exposure of these impacts. Short et al. devote Chapter Nine of their book to some of the real-world factors distancing their laboratory interventions from communications practice. A contention of this thesis is that whilst process change without material consequence is conceivable, it is not safe to assume that measurable changes of interactive behaviour are inconsequential. An assertion to the effect that process factors are only relevant if they impact on task performance is not at all the same thing as demonstrating that process factors are an irrelevance.

A scientific approach to behaviour is built on determinism: all that happens is a consequence of other things. An inability to detect change in task performance when clear differences in process measures are obtained may mean that measured performance is unchanged but material performance may yet be affected. The value of a performance null result is entirely dependent on the value of the measure as a summary of the work performed. An understanding of interactive behaviour through the resources and relationships involved would place one in a strong position to argue for the relevance of process differences, rather than to dismiss them in ignorance. There are two aspects of a criticism of outcome measures. One concerns the adequacy of a measure to address the characteristics of a given task. The other concerns the adequacy of the conception of the work distilled into the identified task. Achieving the latter is a rather ambitious project, although certainly a vital one for HCI.

A persistent problem for the laboratory investigator concerns the generalizability of results obtained for short-term novel tasks to a long-term working environment. Measures designed to throw light on the interactive nature of joint activity can inform this translation by indicating *how* tasks are performed, such that for example the effort required of those performing them might be revealed. This might illustrate how even an adequate conception of the work to be performed might yet fail in generating an appropriate understanding of the task.

An example, drawn from HCI, of this kind is a model devised by Kieras and Polson (1985). Their "Cognitive Complexity Theory" (CCT) provided a means by which to judge the mental effort required of a computer user to carry out particular tasks. It may be that a user could have performed a set task to criterion with two alternative designs by taking on a greater personal load for one than the other such that performance would be indistinguishable in an empirical evaluation. A riposte might be that a well-designed experiment of this kind would include some kind of stressor, such as bursts of white noise or counting backwards, to amplify the usability differential between systems.

Unfortunately, there are any number of levels of disturbance that might be imagined, each with different consequences for the task at hand and the process effects at work. This is particularly true for communication studies. A burst of white noise cannot be considered reasonably as an independent distracter, but an additional influence prejudiced against the auditory channel. In any case, a basic time pressure has failed to differentiate media on outcome measures in a number of studies, even though visual media tend to involve longer interactions when there is no such time pressure. In other words, whatever promotes the extension of interaction when participants can see one another is easily put aside when subject to time stress without affecting the measurable outcome. A CCT for communication is not advocated here, nor indeed is CCT presented as a perfect solution to problems of this kind. It is described merely to illustrate an explanatory principle for real-world behaviour.

Another quite general matter for real-world design of technologies of this kind, indeed of work to incorporate such technologies, is that asymmetries are unpredictable and often invisible. Heath and Luff (1992a) report on attempts to use a ubiquitous VMI facility at Rank Xerox Research Laboratories, Cambridge[8]. They adopted an analytic approach focusing:

> "on the *in situ* or contextual character of human conduct and in particular the sequential and socio-interactional organizations which inform the production and intelligibility of social actions and activities." (p. 3)

---

[8]Now known as Xerox Research Centre Europe.

It is not possible to conclude that the visual aspects of visually mediating technologies are exclusively vested in the gaze behaviour of participants. By accident or design technological mediation may not be reciprocal. Simply knowing that one may, in principle, be subject to another person's gaze, whether or not that other person is a ratified member of one's identified conversational group, must transform conduct as a "Big Brother" effect (Bellotti & Sellen, 1993). However, when visibility is fully reciprocal, it is difficult to argue that such an effect would operate without contribution of active or reactive gaze deployment of the person concerned.

In sum, then, it is unreasonable to assert that a paucity in outcome measures to differentiate audio-only from audio-video interaction means that there really are no consequences for outcome, when there are measurable differences on the progress of interaction. However, it would also be inappropriate to assert that any measurable process differences must inevitably affect the outcome of a specific task in some way. This is the familiar problem of balancing the risks of Type One and Type Two errors. All things being equal, measures that do not differentiate two independent variables indicate that the theoretical function(s) operationalized as those measures is not informed by them. However, process measures differentiating independent variables mean that all things are not equal. It is a basic contention of this thesis that the role of process factors in interaction is fundamental to any approach to interactive technology. The work to be described moves on from this proposition to develop an approach for analysing process factors in interaction. The approach is described against a background of work consistently failing to differentiate audio-video from audio-only mediated interaction on outcome measures.

Where process measures have been obtained, they have indicated a role for VMI but have not effectively discriminated between the behaviours exhibited in these media. A substantial proportion of these have been in the form of subjective evaluations. A problem with measures of these kind concerns their reliance on the conscious availability to the subject of the factors at work. If, as has been argued, much of the power to be derived from non-verbal communication is a deep-seated automatic perception of various aspects of other people's demeanour, intent and conscious preoccupation, then it is likely that the failure of such processes or their successful operation will not be immediately apparent. It is not intended to suggest that no report would be possible, nor indeed that any conceivable report of this kind would be valueless. Rather, it is asserted that the value of such reports is limited and that their limitation could be usefully addressed by a complimentary, behavioural approach to the dynamics of interpersonal interaction.

The next chapter develops an account of interactive behaviour by drawing upon a developing set of theoretical approaches to collaborative work. It then presents Activity Set Analysis as a methodological approach suitable for informing understanding of interactive behaviour as interrelated temporal activity.

2 *Interactive behaviour as integrated activity*

Chapter summary

This chapter turns from addressing the specifics of investigating VMI technologies to a broader consideration of interactive behaviour. Problems stemming from an ergonomic heritage focussed on taxonomies of independent action are described. Developments in studies of Human-Computer Interaction (HCI) and Computer-Supported Cooperative Work (CSCW) are considered for their notions of *embedded interaction* as alternatives to the preponderance of performance-outcome emphasis within the tradition of VMI investigation. Distributed Cognition and Activity Theoretic accounts, as they have been advocated for use in HCI, are then drawn upon to build an understanding of the characteristics of interactive behaviour. It is suggested that such observable behaviour should be considered as both *continuous,* rather than discrete in time, and as visible evidence of *integrated* activity systems, rather than as independent sequences of events. The challenge for the investigator of interactive behaviour is first to identify systems of activity in such a way that systemic behaviour may be tractable to analysis. A simple observational strategy is described for recording quantitative, continuous data as contemporaneous Activity states, comprising measureable conditions of actors, of actors' tools, work context and general environment. Application of this method is proposed to provide additional, quantitative insights into the structure of interactive behaviour to complement other means of exposing the mechanisms of interactivity in relation to the objects of work.

2.1 A methodological reflection on interaction studies

Chapter One discussed a thirty-year history of explorations in technologically mediated interactive communication. Evidence was drawn upon from a number of sources to show how visual information can play varied, powerful and subtle roles in interaction. Although differences between some mediation techniques along some dimensions of measurement have been exposed, the ways in which live video transception are implicated in the interactions of those exposed to it remain problematic. The generally indifferent scientific evidence is at odds with the continuing emphasis placed on video telecommunications products by all the leading technologists, as well as the consistently positive ratings given by video study participants (Bruce, 1996).

How might one account for the lack of evidence to substantiate the expected positive role for a visual channel in mediated interaction? A simple answer might be that merely seeing a disembodied image is so far removed from ordinary, face-to-face experience that there are no savings carried over to its technologically mediated analogue. However, the onus must rest on any who adhere to such a viewpoint to demonstrate that this is so, with the accompanied risks of dicing with the null hypothesis. An alternative account might be to think again about the methods employed by the previous researchers to formulate their enquiries, and the model of interaction driving such formulations.

2.1.1 Work and cognitive processes

Much of the work investigating mediated communication has been carried out within a distinctly work-outcome or production-oriented research paradigm. Chapter One showed how the preponderance of VMI investigations have taken this form and only more recently have aspects of the process of interaction via audio-links been considered material to their evaluation. In part, this may be seen as a natural consequence of any work undertaken in a commercial context and, indeed, much of the work undertaken has been motivated by industrial concerns. In any case, the stance of scientific psychology, in its 'ergonomic' guise, has been very much in the Taylorist tradition. The measures of behaviour taken are predominantly in terms of task outcomes in the immediate timeframe of the study, focusing on the purely individual, sequential, mechanistic fluency of human organism in the act of concern. Performance is operationalized in terms of production rates, so that improvements are seen as increased unit output for a given work period or less time required for a given production volume. The tasks of a worker are broken down into constituent actions, each being optimized on the basis of biomechanics and physiology. Much as Taylor examined in detail the movements comprising the work of loading coke wagons, so task analysts have looked at the human demands of work and, later in HCI through ergonomics and human-factors engineering, conceptual demands of computer systems (Bannon, 1991).

Latter-day task analysis has taken this idea some steps further by anticipating the psychological factors as well as the physical composition of action required for the completion of an identified piece of work. Models of mental processes, drawn from cognitive science, have been recruited to reason about how people marshal their mental resources in approaching a problem. Careful consideration is made of the knowledge required by operatives to carry out their work, and this is related to the effective design of tools to support individuals by modelling the demands of alternatives on their cognitive apparatus (Barnard & Harrison, 1989; Barnard, Wilson & MacLean, 1987; Diaper & Johnson, 1989; Johnson & Johnson, 1991; Kieras & Polson, 1985). By modelling the processes, a deeper insight into the notion of "performance" is obtained, making possible the optimisation of task requirements against putative human action.

Critiques of task analysis turn on the problems of abstracting away from the contextual factors so often influential in providing the additional and crucial real-world constraints on human action. In other words, the action anticipated by cognitive models is so putative that it is barely useful. Still worse, this rational and omniscient view of the world can prove counterproductive when made concrete, by singularly refusing to admit all alternative courses of action. Winograd and Flores implemented a model of how meetings are carried out, the COORDINATOR, by extrapolating from Speech Act Theory (Austin, 1962; Searle, 1969). In some well-reported instances, the COORDINATOR failed spectacularly. Its failure prompted some heated exchanges in the CSCW Journal (Bannon, 1995), revolving around the role-in-context of the rigid category structures, effectively policed by the software. Bowers (1991), Suchman (1993) and Robinson (1993) have variously argued that the highly contingent nature of human action mitigates against the effectiveness of binding control on work practices and that, however well intentioned and objective in abstraction, controls become irreversibly bound up with the politics of an organization in application.

2.1.2 Opportunistic action and available resources

These critiques derive from examinations of work as actually performed rather than in-principle considerations of what it ought to involve. The view of work arising from these studies paints human action not as sequences of independent behaviours, either within the individual or between the individual and his or her tools, but as a highly opportunistic, flexible and joint activity. This is in stark contrast to the view of work underlying traditional ergonomic approaches, as part of a fundamentally rational and compartmentalized organization, with the compartments structurally predetermined to follow one of a finite number of action patterns. Any decision points are well-defined and transfer of control or responsibility between functional units are supposed to be clear. Suchman's (1987) "Plans and Situated Action" revolutionized HCI conceptions of work, within which the business of interacting with computers is understood. She argued that action may be explained more effectively in terms of local conditions than by the sort of planning described by Card, Moran & Newell (1983). Rather than action being driven by purely rational processes, for Suchman action is in a very fundamental way driven by the situation of which the actor is a part. This "situated action" perspective emphasizes how the world reacting to the efforts of individuals provides a set of resources rich in potential for action. The patterns of interaction embedding the individual are consequent of their situational understanding. Local practices, negotiated between people and sensitive to the possibilities offered by their world, evolve and govern by providing a set of resources viewed as such by virtue of their significance and reminder for the communities concerned.

Studies in Human-Computer Interaction have shown how nominally "single-user applications", in the jargon, can and do function seamlessly as multi-user systems. Nardi and Miller (1991) looked at how spreadsheet software was deployed and used in office environments. Rather than particular spreadsheets remaining solely at the disposal of particular individuals, they were contributed to by a whole range of people, recruited through informal communication networks. In short, they were used as mediators of joint work, rather than simply as tools for independent use by individuals, supporting the "sharing of domain knowledge among co-workers". Where work is explicitly collaborative, communication itself seems to owe a great deal to situational dependencies, especially through shared artefacts. John Tang examined the collaborative processes involved in producing a design for an interface to an interactive computer-controlled system (Tang, 1991). His intention was to generate some insights for the design of computer tools to support this kind of work and so his interest was focused on how the paper and drawings were used by his teams to support their own design activity. Protocol analysis, coding for various manipulations and actions with the paper as well as speech, showed that the drawing space functioned as "a key resource for mediating group interaction", as well as providing the conventional role of serving as a record of the group's work.

Hutchins and colleagues (Flor & Hutchins, 1991; Hutchins & Klausen, 1991; Seifert & Hutchins, 1992) have looked at how navigation work is carried out by flight and ship crews, and the process of software development, by examining the way in which team members communicate with one another. They showed how patterns of discourse in the expected multiparty exchanges often relied upon metaknowledge about the work, including the knowledge status of fellow team members. Team members constantly "kept an eye open" for the things going on around each of them, as well as discharging their own particular responsibilities. That meant each would contribute to a continually evolving shared understanding of the project in which they were all engaged, not only as pertinent opportunities arose but also grounded on assumptions of others' understanding.

Hutchins describes this as "distributed cognition", meaning that the elements of the state and the direction of ongoing work is in some sense shared out between participants and also vested in the artefacts of work. Distributed cognition is said to reside in a 'complex cognitive system', comprising collections of individuals and artefacts participating in the performance of a task. A distinction has long been made between declarative and procedural knowledge, the difference between "knowing that" and "knowing how" (Ryle, 1949) , and Hutchins' work can be thought of as making an explicit link between the two. For example a pilot might not "know that" his flight speed is 450 knots at any given moment but he "knows how" to find out at a glance: he or she entrusts the information to the aircraft's instrumentation. Of course, crew may have a rough idea about air speed at any moment but this is only possible by maintaining intermittent contact with the instrumentation. The collaborative dimension to this insight is that the whole flight crew "knows how to know" through the same artefact (air speed indicator) and they each understand this. Similarly, radio communications notionally just for the pilot are heard by the flight crew as a whole. Consequently not only are they all in receipt of this information but understand that they have all been in receipt of it. The information automatically becomes part of the flight crew's "common ground", to borrow Clark's terminology (Clark & Brennan, 1991), and hence subsequent exchanges between them are premised on mutual possession of that knowledge.

Latterly, cognitive models, in application to interactive behaviour, have moved towards incorporating some notion of situation by defining usage scenarios of various kinds (Young & Barnard, 1991) or attempting to model the processes of interleaving system or environmental change alongside human cognitive processes (Howes & Payne, 1990; Young & McNeese, 1993; Young, Howes & Whittington, 1990; Young & Simon, 1987). Howes' *Ayn* model in particular is significant in that it attempts to deal with interaction as a continual recognition-reconstruction cycle. *Ayn* couples two models, one of a computer system and one of the user, to generate an overall model of interaction. It ties system/environmental behaviour and user behaviour at each point of action, practically the antithesis of the early plan-based, state-chaining, problem-solving approaches (Howes, 1994). Monk has recently argued that interactive behaviour is so fundamentally conditioned by person and setting in combination that separate models are unintelligible (Monk, Accepted subject to revision; Monk, In press). He presents a fully integrated interaction model, *SARC*, comprising components from actor and (computer) tool in a continuous interchange of action, recognition, reorganization and reaction.

These properties of interactive behaviour - opportunistic action, integration, coordination through shared artefacts and mutual knowledge - would seem to place a high premium on looking at the manner of expression of component behaviours over the course of interaction, and its relationship to the artefacts through which it is articulated. From a design viewpoint, human interaction is a generic concept, refined in the context of information artefacts to mean something loading heavily on cognitive activity. It also means that whatever assumptions are made about the ubiquitous users, their persistent action in its setting governs their use of available artefact(s). And by persistent, I refer here to the history-context of interactive behaviour. Moment-to-moment changes within and between individuals as well as their environment are fundamental to the expression of individuals' behaviours and the progressive development of a mutual understanding of the work.

An approach to analysing interactive behaviour is required, oriented towards the means by which engaged individuals interact with one another and sensitive in time to whatever differences may result from the details of their environment. Since there is an enormous range of behaviour that might be of interest to an investigator, not to mention the range of descriptive granularity appropriate for the kind of problem to be addressed, the approach should be capable of accommodating any kind of behaviour. There should be opportunity for any observer to define and record elements of the interaction which accord with the concerns of their own practical and theoretical motives.

The rest of this chapter describes such an approach, conceptualizing momentarily visible behaviour as reflecting continuously changing activity states within and between actors and also their environment. Coordination is seen as a time-based concept, given evidence by the contemporanaeity or temporal interleaving of activity states.

2.2 Time, events, activities

The idea of behaviour has many connotations including, from a psychological perspective, something that is in some way observable, concrete and, very often, discrete. Rather than thinking of separate behaviours incrementally constituting action, this chapter argues for thinking in terms of the continuous expression of a variety of influences as "Activity". Activity, as the term is to be used here, is taken as referring to the performance of action over time in relation to resources: it is fundamentally a matter of mutually sensitive, changing temporal states. It owes a good deal to the activity theory of Leont'ev and in turn to Vygotsky's Cultural-Historical psychology (Kaptelinin, 1996; Kuutti, 1996; Zinchenko, 1995).

2.2.1 Activity and activity theory

Activity theory (AT)[9] is not in fact a single canonical theory as its grand title suggests, but a way of looking at the world, a set of assumptions and implications for the study of human behaviour admitting a range of methodological possibilities. AT invests heavily in notions of intent, history, mediation, collaboration and development. These are realized through the social matrix of which individuals are a part; a combination of people and artefacts. Artefacts are tools or mechanisms, physical or conceptual, for getting things done. The "interpenetrated" relationships between people and artefacts are the subject matter of AT. Nardi (1996) argues that HCI has failed to come to grips *systematically* with terms such as "context", "situation" and "practice", and that AT is a likely candidate for an appropriate expansion of HCI by virtue of this very focus. However, AT is not replete with techniques and procedures for investigation. Its value, if any, lies in the explanatory constructs it offers. As Kuutti (1996) puts it:

> "Broadly defined, activity theory is a philosophical and cross-disciplinary framework for studying different forms of human practices as development processes, with both individual and social levels interlinked at the same time." (Kuutti, 1996, p.25)

AT relates to the use of the term Activity, as described in this chapter, both through its commitment to the notion of 'object' as an intrinsic part of behaviour as a governing motive or influence, and also to the mediation of tools in the performance of behaviour. This is not to say that Activity is only in evidence where tools can be seen in use. Tools, in AT, comprise any potential device, literal or figurative, that might be recruited in addressing an object through action. Vygotsky, the originator of AT, was far more interested in the conceptual devices available to people through language to effect

---

[9]Conventionally, "activity theory" is written in lower case but here I have adopted capitalized AT for the sake of making clear and distinct the use of the term Activity as it is developed in this chapter.

transformations of their social and cultural world (Blackler, 1995). Activity, as the term is used here, is conceived as instrumental and sensitive to available resources, whether by way of exploitation or transformation, in pursuit of the individual's goals. The relation to AT's 'object', then, is via instrumentality of action and also AT's recognition of the families of objects framing all human behaviour. Activities range from the lowest level of moment-to-moment decisions about immediate imperatives, Zinchenko argues, through to the societal and cultural mores that influence action. For current purposes, this last point is material in that it admits a range of analytic possibility in terms of the granularity and operational concreteness of recording, within which the current take on Activity might be adopted.

There is an important distinction to be made between someone's intentions and their observable behaviour. Duncan and Fiske (1977) set out on a similar enterprise to that undertaken in this thesis, though their aim was to understand interpersonal communication. They make the point that there is a huge gulf between the observable behaviour and the mental processes assumed to have given rise to it. Their units of analysis were given as 'acts', about which they said simply that they must be discrete, rather than present by degree, that they should require little or no interpretation, and that they should have a presumed relevance.

> "They were seen as samples of activity, not as signs which had some a priori significance. We wanted the meaning or importance of each kind of act to emerge from its relationship and patterns without pre judgement on our part." (Duncan & Fiske, 1977, p. 40)

Perhaps in AT terms, this might be rephrased as a matter of how an individual's family of objects are made manifest in the operations ultimately comprising their activity. Object must be inferred from observable behaviour in order to determine the course of group activity, or "trajectories of action" in Distributed Cognition terminology. However, the categorically distinct 'act' class envisaged by Duncan and Fiske is inevitably muddied by the theoretical position and associated research questions of the observer. This is made abundantly clear in classic forms of task analysis, on the face of it highly operationalized descriptions, where the development of goal-structured hierarchy of behavioural acts, as tasks, is quite endemic. This is also a critical matter from an AT perspective; object is intrinsic to action. As is later discussed, the method to be described here makes no commitment to a particular view of motive or intention. Decisions must be made about the level of description appropriate for an observer to reliably record and to inform the issue behind the study: the investigator's object(s). It is assumed that all observable behaviour, as part of an activity system, is systematically organized at some level but that there is of necessity a distance of interpretation to be bridged.

Susanne Bødker has been instrumental in introducing Activity Theory to the HCI community, in application to requirements capture and system design (Bødker, 1990). She has also specifically addressed issues of behavioural analysis with videotaped recordings, from an activity-theoretic perspective (Bødker, 1996). There is a persistent issue in behaviour analysis for the proponents of contextual approaches concerned with the imposition of the analyst's preconceptions onto the situation under study. It is a difficult matter, as all analysis is to a greater or lesser extent the business of deciding what is and what is not relevant from amongst the huge range of potential information in and around an interaction. Bødker presents Activity Theory as a means of structuring recorded interactions and yet does not prescribe absolutely what to look for. One of the arguments she makes resonates strongly with Hutchins' position on interaction:

> "a better understanding of use [than the limitations of Cognitive Science can provide is] important to the continuing development of methods and theories in HCI ... activity theory seems to provide an interesting alternative framework for developing a more comprehensive unit of analysis for our studies." (Bødker, 1996, p. 146)

Remember that, for Hutchins, the "complex cognitive system" is the unit of analysis, such that all interpretation must be framed in its terms and at its level. This again informs the notion of Activity here. Single behaviours removed from their place in interaction, are thereby stripped of any kind of context for interpretation, and in the process are deprived of any meaning. Taking single behaviours as a unit of analysis is seen as inappropriate for students of interaction. Activity is the unit of analysis proposed here, primarily as a concept of behaviour embedded in the temporal flow of interaction and in the physical modes of its performance.

Bødker suggests that the real potential of AT is to allow for analyses at several levels, depending on the focus of analysis, where level is taken as a metaphor for abstraction from behavioural data, or reductionism from social experience, depending on one's initial viewpoint. Bødker describes a way of building on the supposed tripartite structure of behaviour as activity, action and operation by asking why something takes place, what takes place, and how it is carried out. She discusses primarily the "what" and the "how", the former exemplified by changes in the explicit focus of work from moment to moment, the latter by breakdowns in carrying out that work by a different kind of focus shift, to dealing explicitly with the tools or instruments of work themselves. "The analysis [described] suggests that one analyse relevant objects and subjects of the web of activities at two levels: a contextual level where the purpose is to situate the artefact in the web of activities, and the level of analysing and tracing the actual focus shifts in specific use situations" (p. 154-55). Activity as used here benefits from this notion in allowing for different periods or rates of Activity state change, so that the granularity of description can effectively accommodate nesting of Activity Streams in order that their forms can be examined within one another. For example, the context of carrying out a particular task within work is an Activity State at one level, while the Activities of "assembly", "transport" and "planning" might interchange many times at other levels.

This insight, informing the use of the term Activity here, is elaborated as Bødker writes:

> "Activities never take place in isolation; they are interwoven with other activities that deal with the same or connected objects, or produce the instruments used in the activity in question. In the course of a specific activity, the object change may be viewed as a change of activity or as a change in the purposeful actions or *clusters* of actions." (Bødker, 1990, p. 149)

So, AT describes human behaviour as the manifestation of multiple objectives or motives at a variety of levels, both on a social scale, from the individual to the cultural, and on an temporal scale, from the immediate to long-term transformations. It further acknowledges that the instruments mediating the address of human objects are of central importance to understanding the behavioural operations performed by individuals. These instruments may in principle include not only purpose-built artefacts but any means afforded an individual within their environment. Finally, the relationship between an action and its object through the mediating artefact or instrument is likely to be highly contingent as a consequence of the interwoven nature of activity.

2.2.2 The principle of the "Activity Set"

This thesis proposes a simple framework for integrating observational entities, grounded in the principle that individuals are occupied in various phases of activity over time, some of which can occur in concert with others, and some of which preclude the performance of others. The notion of activity is extended to include the environment of performance, including background conditions, work context and other systems the conditions of which vary, over time and in relation to the behaviour of the "focal" individual. In this way, interactive behaviour in totality is seen as an indefinite number of streams of Activity occurring in time, with those of individuals running seamlessly together with those of their environment, and organized by the overall expression of particular Activities in relation to the (invisible) objects of each person. It is important to understand that the activity concept in AT is intimately bound up with motive or intention, whereas the methodological concept of Activity described here is intended to drive systematic and integrated observation of behaviour. It is assumed that all observable action is ultimately motivated in some way but that motive plays no part in principle to the definition of Activity, as used here. Any inference of motive behind visible behaviour can be relevant to observation if and only if the observer is confident, and can later demonstrate, that their inferences are reliable. Mindful of the theoretical and practical problems of intention-based definitions of non-verbal behaviour, as discussed in Chapter One, and Watzlawick's solution, it would seem that, from an observational viewpoint, an Activity concept tied to an *explicit* 'object' is unnecessary and undesirable. In fairness, there is no necessity built into the AT tradition that object should be articulable by the person concerned. Indeed, as Wertsch relates, following the schism between Vygotsky and Leont'ev, and also Luria, much of AT development occurred in a period when it was politically inexpedient to make any reference to consciousness. It is not a contention that some would agree with: Nardi's account is premised on the relationship between AT and conscious experience. Blackler, discussing the cultural-historical AT tradition, asserts that:

> "activity theory conceptualizes working as an active, purposive endeavour. Expertise is distributed within a community of practitioners but activity systems achieve their coherency by virtue of the shared object of activity that unites participants. Neither social cooperation nor workplace technologies can sensibly be understood in isolation from the purposive nature of the activity system of which they are a part." (Blackler, 1995, p. 240)

Nevertheless, it is clear that AT has more to say about the necessary connection between individual, object and medium for action than the clarity of the object notion for these participants. Relating Engestrom's (1993) AT analysis of a Finnish medical centre, Blacker notes that medical staff had widely differing notions of their work objects, including "biomedical", "administrative-economic", "psychiatric" and so on. Furthermore, one of AT's attractions to contemporary theorists of technological design is the centrality of automatization of once-cognitive activities, as operations and tacit knowledge. Blackler observes that activity in AT sidesteps many of the dichotomies that have had a central place in Western intellectual traditions, not least those between thought and action, between actors and their tools, and between individuals and their setting. Blackler is also at pains to point out that it is not so much a focus on isolated behaviour in absence of context that contrasts with AT or even a preponderance of interest in the context of action at the expense of individual considerations. At the heart of AT is the indivisibility of actor and context for understanding action. Indeed, it is all of these elements together that comprise "activity". It may seem strange that environmental states are included as active in the approach advocated here, in the same way as the ongoing behaviour of individuals who are, on the face of it, focusing the study. This is precisely from a recognition of the indivisibility of individual, action and setting: environments are dynamic, and from an interactive behaviour perspective they must be at least reactive.

Observational Activities provide the materials from which one might begin to build a picture of the components of activity systems. It would perhaps be far more appropriate within an AT tradition to describe these as actions rather than as Activities. I have chosen not to do so with the intention of not losing sight of the purpose of the approach; to build an understanding of systems of interactive behaviour in relation to one another and ultimately the objectives of those involved. However, it cannot be overemphasized that the identification of observational Activities to figure in a potential activity system is not presented here as a carte blanche for recording anything and everything. As later chapters serve to demonstrate, candidacy for observational Activities should follow from some logic for inclusion. Of course, just what that logic is must depend heavily on the purpose of investigation. Activities (capital 'A') are not activities (small 'a'); Activities (capital 'A') serve to present some additional material to an analyst from which something of the underlying activity system might be revealed.

Physical modes of operation place certain restrictions on expressions of behaviour. For this reason, some Activities must be ordered: to imbibe one must first reach for one's glass, then bring it to one's lips and then sip. It would be difficult, if this example were in a public house, to also wave over an acquaintance seen to enter our imaginary hostelry. The act of reaching and drinking would have to be interrupted or modified to include the gesture. Whilst the gestural codes possible within a given manual act can be complex (Bull, 1987), other modes of operation are less flexible. Looking behaviour is closely restricted to a single fixation target at a time, given the directionality of gaze and focal distance of target, however quickly one might shift from one to the next. Modal restrictions of this kind have useful implications for observation: once an expressive mode has been identified, it is recordable on a system of mutually exclusive behavioural states, marked by transition points. Mutual exclusivity amongst Activities means that something is being said about all those precluded by the occurrence of the just-recorded Activity and also that they can be much more efficiently recorded.

To take another example, it is possible to ride a bicycle and have a conversation at the same time. It would be wholly impractical to ride a bicycle and travel on a bus at the same time: travel by buses and bicycle is to all intents and purposes mutually exclusive. The use of words is independent of whether one rides on buses or bicycles and so one might say that conversation is not mutually exclusive with respect to mode of transport. Bicycling is an activity which, at least for the practised cyclist, does not require the full and complete employment of his or her physical and mental resources. Equally, conversation usually does not preclude other forms of Activity. Note that this is not at all the same thing as asserting that the use of buses and bicycles does not *interact* with the use of words. The Activity concept developed here is, in the first instance, purely observational. In fact, the systematic observation of the use of buses and bicycles along with the recording of the use of words might allow one to begin to explore the underlying, functional relationships that do exist between words and forms of transport. The observational Activity then becomes a basis from which to infer the underlying Activity system. Activity descriptors are part of the theoretical total state of an individual or, following Vygotsky, groups, societies or perhaps even culture. Other state information about an individual's context of action, for example location, might usefully contribute to the generation of Activity classes. Being at a bus stop is location information but it is also heavily implicated in catching a bus[10]. Location varies with time and constrains action, and in this

---

[10]Since the deregulation of public transport systems, this relationship has become rather more muddied.

sense it stands to be a full part of the Activity system. Being at traffic lights is again simply location information but with concrete relevance for the road traveller, whether by bicycle or bus.

In the case of the above buses and bicycles example, the use of buses and the use of bicycles are two Activities in which an individual might engage. The use of words is a third Activity. The first two are in some sense equivalent alternative Activities, in this case to do with locomotion. Their exclusive equivalence is the relationship which binds them as observational entities. Although the purpose of buses and bicycles is that they are both forms of transport, the important relationship between them from an observational viewpoint is that they both tie-up the individual's resources at a level that precludes contemporanaeity. Any purpose associated with an observation is a theoretical inference, arising from the particular explanatory constructs of the observer. The observation by Activity category is intended to help make clear the elements of behaviour which constitute some situation, and more particularly the relationships that exist between them. There is no list of permissible or required Activities to be recorded by all observers. Activities refer to the questions addressed by the observer and the relevances of their theoretical origin.

The terms used to refer to this approach to interactive behaviour are:

Activity Set Analysis- the general name describing this approach.

An Activity - any aspect of human behaviour or environmental state varying over time and identifiable by some observational criteria.

An Activity Set - a set of entity states comprising Activities mutually exclusive in time and collectively forming an exhaustive description of the entity to which they apply on the specified observational dimension.

2.2.3 Observing a modal stream of behaviour on a time base

An ethological scenario

Imagine watching a group of gazelle by a water hole[11]. There are four of them but you are concentrating on just one, a doe. She is grazing. Occasionally she looks over to where a lion is resting beneath the shade of a tree on the other

[11] The example is not based on real gazelle feeding behaviour data. It is included for illustrative purposes only and in recognition of the ethological descriptive paradigm from which it stems.

side of the water hole. Sometimes she takes a few steps to the water's edge to drink before returning to her grazing.

All things happen in time. Some acts may be of such a short duration that they might be taken as instantaneous. Our gazelle's attention on the lion may be maintained by very brief glances. The Sttaw gazelle are highly adapted to a marshy savannah so that they are intermittently able to keep an eye on predators whilst grazing or drinking. Other behaviours are more time-consuming: grazing involves finding suitable plant life and tugging at this thick and wiry vegetation, a Sttaw gazelle delicacy. Drinking requires careful positioning along the steep-sided gullies of the raging torrents that course their way through the Sttaw plateau, besides the act of imbibing itself.

Approaches to Recording Behaviour

Categorization and enumeration of behaviour has a long history in ethological and non-verbal behaviour research (Argyle, 1988; Martin & Bateson, 1993). The stream of observational subjects' activity is divided up into categorical units amenable for identification, depending on available recording equipment and the kind of question framed by the investigator's research agenda. Behavioural categories are chosen to be:

- relevant to the research question.
- precisely defined in terms of observable content rather than imputed purpose
- homogeneous (all behaviours counted as instances of the category share the same relevant properties)
- independent (categories are fully exclusive, admitting no overlap of inclusion criteria).

The most fundamental measure deployed by researchers of either tradition has been the coding occurrences of each behavioural category. These codes might then be used to determine rate of performance given a time base, as intra-event latency or behaviour duration. Once again, the dominant approach has been to compile summary statistics with the intention of forming statements about behavioural production under certain conditions.

An observer must first decide how to go about collecting observations by choosing a sampling method for the subjects concerned (which ones to observe and when to do it) and a way to record what they have seen.

Four sampling regimes have evolved for ethological observation:

- *ad libitum*, or free of systematic constraint.
- *focal*, or concentrating on only one individual in a group within a defined time period.
- *scan*, or observing a set of individuals in several sessions, broken by intervals.
- *behaviour*, or concentrating on one kind of behaviour evident in a group of individuals within a defined period.

In principle, decisions about recording methods are independent of the adopted sampling technique. In a field setting, these would be whether to collect data continuously or periodically. Periodic collection, not to be confused with scan sampling, involves noting whether or not a category is in evidence at the instant terminating a within-session sampling period (instantaneous recording) or else whether or not a category has occurred during the within-session sampling period (one-zero recording). Periodic recording is rather crude by comparison with continuous recording, generating inherently biased estimates of both behaviour frequency and duration[12]. The advent of inexpensive and reliable recording, observation and cataloguing video equipment has meant that practical constraints on the use of continuous recording have relaxed[13]. Formerly, in many cases periodic recording was the only practicable means of live data collection and so preferable in many cases to continuous recording just by default. Social psychological research on the other hand has been more closely concerned with observation in controlled environments, where film and sound recording has been a costly but realistic investment since the early '70s. Cheap and reliable video equipment has been available for some years and so interactive sessions are frequently recorded for review *ad infinitum*. One may play, pause, review and take breaks to aid concentration. Continuous observation recordings may be compiled over a number of such viewings, limited only by the investment of effort deemed worthwhile.

---

[12]There are mathematical techniques available to correct such biases but these inevitably add an additional degree of uncertainty (as an additional level of estimation) to the resultant frequency and duration data, compared to those given by continuous recording.

[13]In a field setting, especially where there are practical constraints on deployment of equipment (prevailing environmental or ambient conditions, socio-political acceptability, physical portability, tapes, power etc.), video recording can fail to produce the kind of re-usable resource implied here. These are the exception rather than the rule. Generally, if it can be observed by eye, it can be observed by camera.

In any case, these techniques are not necessarily to be used in an exclusive sense. Coding must still take place at some time and it has been argued that film recordings are always second best to being on the spot (Martin & Bateson, 1993). Some behavioural categories of relevance to a research question may be typified by characteristics (short, rare) that make them realistically amenable to one approach whereas others of equal *a priori* relevance to the issue may only be amenable to another. A host of other factors are necessary for deciding on such matters, such as behavioural conspicuity, *a priori* significance of behavioural category against expected incidence, observer concentration and load etc. For the purposes of this discussion, it is more important to consider in the round the nature of the information returned by the analytic investment. A generalized account is given below, based on the above "ethological scenario". It assumes a succession of focal samplings to compile a single data set for all individuals under observation.

As a simple event-oriented method, an observer might mark off instances of target behaviours over the course of a defined time period. Event-based coding of behaviour takes the view that duration does not meaningfully contribute to the research issue at stake. Figure 2.1 thus records that the gazelle in question drank twice and looked at the lion on seven occasions, but takes no account of the 'bout lengths' of either category.

Drinks	Observes Lion
II	~~IIII~~ II

Figure 2.1: Focal gazelle frequency recording

As a refinement, data of this kind may be recorded with the time at which it occurred. The data may be divided subsequently into fixed intervals or "bins" and analysed by unit time. Thus our gazelle's behaviour is represented with defined markers on a time base in Figure 2.2. Tall lines stand for "looks at lion", short lines for "drinks".

Figure 2.2: Sample data on a time base, within five-minute intervals.

Interval length would typically be the shortest possible without sacrificing observation reliability. For the sake of argument, staying with an event conception for the moment, imagine five-minute intervals from a total observation period of twenty minutes. Analysis on the five-minute sampling interval might allow one to frame statements about typical gazelle gaze frequency. These data then preserve that in the first five-minute interval the gazelle drank once and glanced at the lion three times, did nothing in the second interval , looked at the lion twice in the third interval and drank once, and finally looked at lion twice more in the final interval. Summary frequencies could only reflect what happened overall. Yet within this description, there is a notion of progression or sequential dependency. For example, one might ask whether drinking leads on looking at the lion: an issue of continuity from one behaviour to the next..

Continuity in this sense may be examined with time-series analysis, typically testing for "Markovian order", provided the actual sequence of observed behaviours is maintained in the observer's record. Markov models are mathematical descriptions of some *process*, in this case meaning the effects of systematic influences at work to shape the occurrence of pre-defined behavioural categories in the ethological record. They are of particular interest here in so far as their application serves to expose process relationships between behaviours in the flow of their execution. As previously discussed, there are good reasons to suppose that process analyses stand to be far more revealing for the student of interactive behaviour than outcomes analyses.

Markovian modelling is premised on a stochastic view of interaction, where consequent interaction events are not fully determined by antecedent events. Earlier events are seen as altering the probabilities of subsequent events (Attneave, 1959; Haccou, 1987). In this way, an event may be predicted from the occurrence of some other event, in terms of number of intervening events. Predictions are expressed as state-space 'transition probabilities': the likelihood of event A in a space of possible events being followed by event B, B being followed by A, and so on. Such an analysis computes actual transition probabilities as a function of the sequential occurrences of behavioural categories in a recorded set, in order to compare with them with a random transition model. More concretely, if "looks at lion" consistently and immediately followed an event, "swishes tail", swishes tail would be said to give a first-order prediction of looks at lion. Thomas, Bull & Roger (1983) applied this method to informal dyadic conversations, using transcripts coded for up to seven categories of speech, and showed that the occurrence of one allowed a reliable prediction of the next.

Discrete Markovian models treat coded behaviours as sequence elements with a degree of hybridity between event and state characteristics. They are without duration and yet exhaustively describe the activity of the individual(s) observed, so that each behaviour corresponds to occupancy of a position in a finite state-space. The "process" modelled in this way is always to be considered as in some state and yet state residence time is irrelevant to the analysis. Discrete Markovian processes are thus applicable only to sequences of instantaneous and discrete observations. Events are instantaneous and thus there is no inherent provision for dealing with temporal relationships such as *whilst* or *during*. It is furthermore necessary to make some strong assumptions about the "stationarity", or uniformity, of transition probabilities over the an observation period. Just in terms of the duration of a behaviour alone, this assumption is very difficult to defend as time spent engaged in behaviour is very likely to alter transition probability given the time-sensitivity of most behaviours of ethological significance (Haccou, 1987).

As discussed in Chapter One, the duration of certain kinds of signals, such as mutual gaze, have been recorded by many investigators. For such behavioural categories, squarely as states of the focal person or animal, recording bout length or duration is essential. Duration data would typically be aggregated to compute the total time over the session.

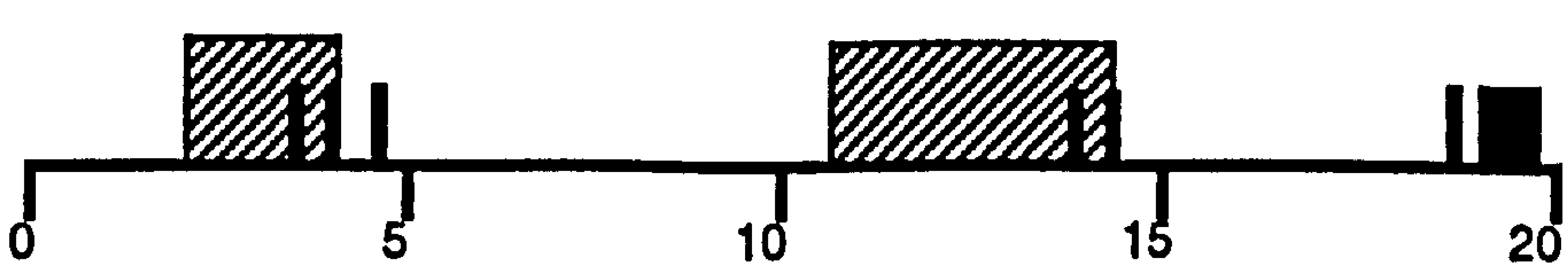


Figure 2.3: Frequency and duration data representation for gazelle feeding behaviour (hatched) and gaze behaviour (filled)

Figure 2.3 thus adds some duration information to our representation, recording that our gazelle drank for two minutes and for four minutes, that six of the looks at the lion were brief and one extended for a minute. Recording of behaviour duration may be within the sampling paradigm or a part of continuous observation. However, these data also fail to address the matter of chronology in action implied by Figure 2.3. Each observation becomes associated with a summary statistic, independently of the other acts. Issues of process contingency between gaze and feeding behaviour cannot be resolved by these data alone. Duncan and Fiske summed up their efforts to develop a research programme for face-to-face interaction on the basis of much this type of data by suggesting that this is an important omission:

> "In the research reported in Parts II and III acts in the stream of interaction were not located in chronological time. In Part II the duration of various acts was noted, but the respective locations of acts within the five-minute conversations were not recorded. In Part III the accurate location of acts was a central concern, but this location was accomplished in terms of an act's sequential position with respect to other acts or to the pauses between acts.
>
> We believe, however, that the consideration of temporal factors would be a valuable addition to the study of interaction. There is clearly much to be learned concerning the temporal aspects of action sequences. One can imagine a kind of interaction chronograph study that took into account a wide variety of speech and body-motion actions, rather than focusing exclusively on temporal parameters of speech and silence. One might ask, for example, are there minimum and-or maximum time intervals associated with the appropriate activation of, and response to, given signals? Further information on this sort of issue is of potential interest, not only to investigators of interactions, but also to investigators of reaction time and human information processing. It appears, however, that considerable attention has to be given to problems of analysing the temporal aspect of face-to-face interaction." (Duncan and Fiske, 1977, p. 311)

Haccou (1987) has discussed the application of continuous time, rather than discrete, Markov modelling for ethological analysis. As discussed above, discrete Markovian models take no account of bout duration or "residence time", as Haccou puts it, yet residence time can condition transition probability. Provided that statistical independence between behavioural categories is preserved, residence time may be factored in to the calculation of transition probability. Such techniques are considerably more promising than their discrete counterparts, having a much clearer relationship with the time-dependency of interactive processes. As Haccou observes, sequence stationarity is highly unlikely in coded records of animal behaviour. That is, behaviour is typically *inhomogeneous*. Indeed, issues such as arousal, aggression and feeding have all been of central concern in ethology. One might test for stationarity in an ethological record to determine homogenous periods and then cross compare behaviour. Unfortunately, dividing sequences up in this way is far from precise and very likely practicable only for longer time periods. In ethology, this does not pose too much of a problem. Animal observations may take place over weeks or months. For shorter spells of interactive behaviour, such as for particular interactive encounters, it could be critical.

Notwithstanding problems of observational data meeting criteria for Markov modelling, there is a rather more difficult issue to contend with: contemporaneity. Discrete Markov models allow for no contemporaneity in principle or practice by allowing sequence elements no duration. As such, it is impossible for two elements to occur at the same time. Continuous time Markov models put in place the potential for contemporaneity, by allowing for duration of states, but cannot allow it on principle since it is vital for all the elements in a recorded sequence to be mutually exclusive. Behavioural categories may be composite of the activity of several individuals, as a derivative of their separate observation. In this way, *while* conditions of a limited kind may be considered, again in terms of transition probabilities. However, the exclusivity requirement for analysis (as distinct from recording) can be problematic when, for example, some behavioural category is inherently a joint action. Haccou provides the example of shaking hands. The category might be defined as the occurrence of hand contact and thus belong in some sense to both participants in the action. The only way continuous time Markovity may be preserved in such cases is by subdividing the category in such a way as to deny or circumvent its joint nature. Haccou suggests for example assigning a truly distinct notion of initiation and response, whilst admitting that this is not always possible for an observer to determine.

This thesis is concerned with interactive behaviour. The earlier part of this chapter was at pains to make clear that much contemporary thinking suggests that interactive behaviour is not divisible into initiation and response in any coherent way. This statement might be re-phrased in statistical terms as that the unit of analysis can never be as small as the individual, but must reside as some level of the joint evidence of action in an activity system. Remember: the conception of interaction is not in any case about independent behaviours recorded in a sequence but explicitly tied together by some higher-level intention. Activity extends so that it is concurrent not only with lower-level activity for an individual but so that it is concurrent with other individuals' activities. This theoretical consideration would seem to leave continuous time Markovian modelling with seriously flaws in application to interactive behaviour.

Contemporaneity explicitly links elemental activities at any given time as an integrated display of a functioning system. Figure 2.4 shows the gazelle's behaviour as two separate streams in time, but indexed to one another explicitly on a time base. The streams represent the gazelle's feeding and gaze direction.

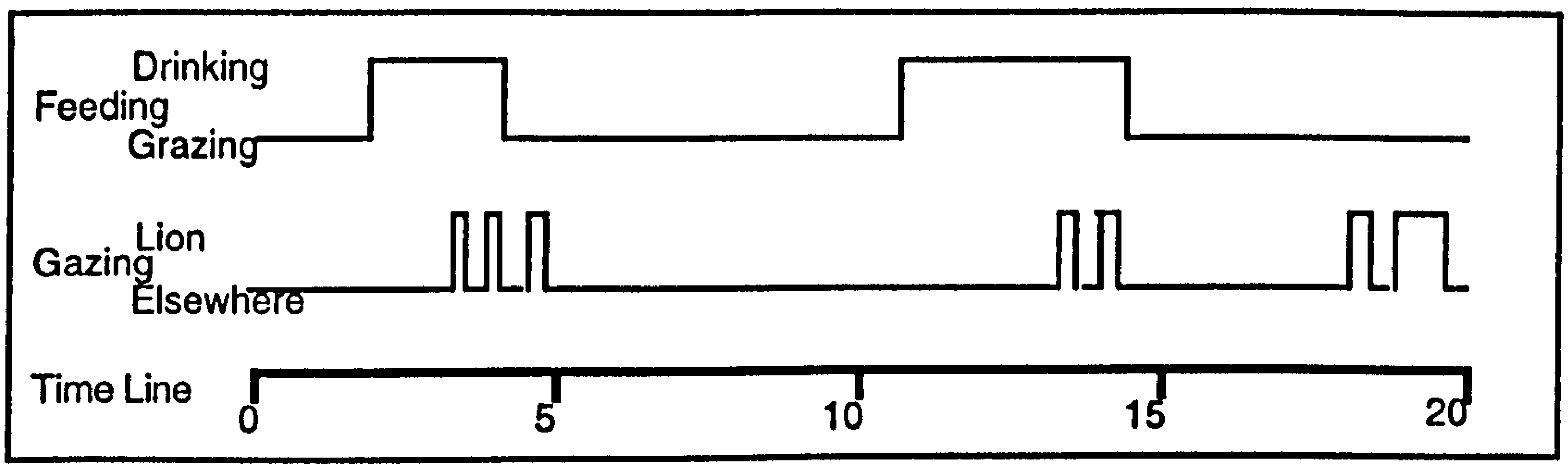


Figure 2.4: Continuous trace, representing two binary Activity Sets over a twenty minute period

This scheme allows one to describe the gazelle's feeding and gaze behaviour both independently, in terms of frequency and timing, and with respect to one another. Thus, the gazelle was grazing at the outset of the period of observation. After two minutes, she moved to the water's edge and drank for two minutes before returning to grazing for a further seven minutes, and so on. Meanwhile, she glanced at the lion twice initially and then looked for a little longer, and so on. We can further say that all of her longer looks were whilst she was at the water rather than whilst grazing. The ethological activity system includes both lion and gazelle: the activities of each are unintelligible without being referenced against one another. A consequence of asserting that individuals are always engaged in some kind of Activity at all analytical levels is that something must always be ongoing. This means that, for multiple-Set analyses, something definite can always be said about the connected involvement of defined Activities in the interactive setting and, by implication, about the functional states of the underlying activity system.

Simple characteristics

Taking each Activity Set on its own allows one to plot the temporal distributions of its constituent Activities and thereby say something about their relative prevalence and actor investment in their own right. These data are here described as "simple statistics", since they deal with one Activity Set at a time.

So it would be said that:

Stream 1(Feeding)	Stream 2 (Gazing)
Activity Drink	*Activity Look at lion*
number = 2	number = 7
mean duration = 180 secs.	mean duration = 6.5 secs.
total duration = 360 secs.	total duration = 36 secs.
Activity Graze	*Activity Look elsewhere*
number = 3	number = 8
mean duration = 280 secs.	mean duration = 145 secs.
total duration = 840 secs.	total duration = 1164 secs.

It can be seen that data on the non-target Activities is forthcoming as a by-product of the convention for exhaustive mutual exclusivity. It is thus possible to make statements of the kind that the gazelle was grazing for more than twice the time she drank and that, although she looked at the lion on seven occasions, she spent only a fraction of the time doing so compared to looking at other things. Where Activity Sets comprise only two Activities, the frequency data is rather less interesting (for every 'on' of a state, there must be an 'off') but similar comparative statements within Activity Sets might otherwise be just as revealing.

Compound or "contingent" characteristics

Data on the joint presentation of behaviours across two or more Activity Sets are referred to as 'contingent statistics'. Examining how Activities from different Activity Sets co-occur, or equally never co-occur, allows one to ask questions about the dependencies between behaviours in time. Whereas the dependency implicit amongst behaviours within Activity Sets is purely to do with the mechanics of performance, potential dependencies amongst Activities from different Activity Sets are all to do with functional sensitivity.

Data of this kind would be of the form:

<u>Stream 1 - Feeding by Stream 2 - Looking</u>
Activity Drink with Activity Look at lion
number = 4
mean duration = 1 secs.
total duration = 4 secs.
Activity Drink with Activity Look elsewhere
number = 4
mean duration = 44 secs.
total duration = 176 secs.

Activity Graze with Activity Look lion
number = 4
mean duration = 8 secs.
total duration = 33 secs.
Activity Graze with Activity Look elsewhere
number = 8
mean duration = 103 secs.
total duration = 829 secs.

From this kind of analysis, one might conclude that the gazelle invests most of its effort in keeping tabs on the lion over the period when she is grazing and barely any time when she is drinking. To this extent, were this observation to be repeated for other gazelles, one might begin to establish a relationship between feeding, drinking and predation.

The early part of this chapter developed a case for rejecting analyses of interactive behaviour as collections of essentially independent actions, related to one another only in terms of procedure. It was argued that levels of contextualization form a set of fully integrated behaviours as an Activity system. The gazelle example, in a very simple way, illustrates how one might use the idea of an Activity system within an observational framework. One might conceive of the water as environmental mediation of the gazelle's survival repertoire. That the gazelle's observation of lions was found to be so much less evident during drinking behaviour might be used to frame an hypothesis to the effect that gazelles are less threatened by lions when they are within reach of water. The observational characteristics of Activity Sets stand to reveal the behaviour membership of an Activity system. One might imagine a host of other Sttaw gazelle behaviours that may or may not fall into an Activity system of predation avoidance. In this simple example, the lion was assumed always to be dozing in the sun. Other lion Activity should also have been recorded in the confident expectation that the lion's engagement in other behaviours would have changed the gazelle's context of observation and behaviour, and in turn likely to have changed lion behaviour until they both settled into a new stable enactment of the system. One might further conceive of higher-paced Activity fluctuations than drinking and grazing examples. As the "signal-to-noise ratio" increases with variability and frequency of Activity state change, so such relationships must become increasingly difficult to determine. A detailed recording and analysis method, such as that proposed here, has the potential to see through the noise to the signals, if any, giving evidence of Activity system involvement.

In the following three chapters, this strategy is fully illustrated in application to data from a VMI case study.

2.3 Summary

Chapter One's discussion of VMI studies prompted a broad consideration of the nature of interactive behaviour. The process of interaction was discussed in terms of continuity of actions and their temporal contiguity. A position was developed for considering all behaviour as continuous, as ongoing Activity, rather than as fragmented sequences of instantaneous events. At psychological or overt behavioural levels, mutually adaptive but separate streams of activity contribute to and express the overall state of the organism. Activity Theory and Distributed Cognition accounts suggest that a meaningful description of interactive behaviour requires inclusion of those environmental states mediating effort and coordinating joint action.

At a psychological level, Activity streams might be interpreted as a variety of motives and processes, specifying goals and conditions that ultimately drive behaviour in relation to environmental opportunity. At a behavioural level, the streams are consequent of physical changes in state, such as positioning in space, orienting, looking, reaching, holding, manipulating, speaking, etc. These acts can be constrained by their physical mode of expression, such that performing one act precludes the performance of certain others: an operational relationship of 'mutual exclusivity'. Observations may be organized on this basis and grouped into Activity Sets. Observing interactive behaviour on the Activity Set principle gives descriptive data tied to the time-base of the interaction. This time base allows any special conjunctions of Activities between Sets to be identified, thereby showing where and how one Activity is related to another Activity. The dependencies of interactive behavioural systems might then be exposed to show which of a set of available resources are implicated in the grand activity and which are not.

3 Action Recorder: an implementation of Activity Set Analysis

Chapter summary

Previous Chapters have proposed a temporal examination of interaction processes. It was argued that an understanding of interactive behaviour must be cast within a system of related factors,together comprising an Activity system. The time and frequency with which available resources are recruited, and evidence of temporal relationships between separate activities, were argued to be indicative of such an Activity system. Behavioural observations to capture dynamic interaction are often costly to obtain and difficult to interpret. A software tool for recording interactive behaviour, Action Recorder, is reported. Action Recorder was designed to make use of the Activity Set Analytic principle described in Chapter Two, to structure observations and to prepare behavioural logs suitable for statistical analysis.

3.1 Methods for dealing with rich data

There are many members of the HCI community who have to build an understanding of how people manage their environment, including the computer systems focusing their work. As discussed in Chapter One, in the context of video-mediated interaction, this type of activity is not necessarily amenable to assessment or understanding from general measures of task performance alone. It is not uncommon to make detailed or recordings of people interacting with computer systems or other information artefacts to provide analysts the "rich data" from which such an understanding might be derived.

Data may be described as "rich" if they afford a range of analyses. Transcripts may be generated of speech and of action taken whilst carrying out a specific task and later examined for a variety of purposes (c.f. Tang & Minneman, 1991). They may additionally be retained after the conclusion of a study to be consulted in pursuance of some later investigation. For example, the Human Communication Research Centre has compiled a corpus of this kind of data, for a set of interactions with the "map task" (Anderson et al., 1991).

One of the richest resources for exploring the dynamics of interaction is offered by filmed or videotaped recordings of activity. These video data[14] are commonly used in the social sciences, having superceded the one-time contemporaneous recording of interaction from observation suites. As Chapter One described, many techniques have evolved for its assessment. Duncan and Fiske's work (Duncan & Fiske, 1977) in particular relied on an extensive analysis of a range of behaviours recorded from films of individuals in interaction. Indeed, Kendon's original report on gaze in interaction was of this form. Industrial usability practitioners consider the video recording of interactions between users and test systems as a vital but unstructured resource. They exploit video data to exemplify and explain prototype interface problems to software engineers (Nayak, Mrazek & Smith, 1995).

Video data is a mixed blessing for analysts: without the inherent complexity of these data, the dynamic characteristics of interest are difficult to obtain, and yet with it the nature of interaction is at risk of being swamped by detail. Although valuable insights may be gained from examining video data, the analysis can be a costly, time-consuming business. The advent of the cheap and reliable desktop computer has the potential to transform analytic investment time. Indeed, the 1990s has seen a marked growth of interest in computer tools as analytic aids for video data (Harrison & Baecker, 1991; Olson & Storrosten, 1990).

There are three general forms to video analysis tools: annotation and indexing systems, event recorders and state recorders. Event recording has a long history in ethological research, as described in Chapter Two. Automated means for observers to note the frequency and moment of behaviours are thus well established. State recorders are less common in behavioural analysis, as facilities for making note of the extent or persistance of certain kinds of behaviour. Noldus' Observer[15] allows some state information to be recorded as a special case of event with a property of duration. Noldus' EthoVision[16] allows the encoding of state information as an implicit characteristic of an automatic tracking facility. The latter allows, for example, the position of rats in a horizontal plane to be recorded by image recognition software. Some of these offer highly sophisticated video-manipulation and control techniques

[14]Video data should not to be confused with the "video-as-data" view of VMC.

[15]Information available at: http://www.noldus.com/products/observer/obs_index.htm

[16]Information available at: http://www.noldus.com/products/ethovision/ev_general.htm

(Owen, Baecker & Harrison, 1994; Weber & Poon, 1994) and involved data analysis procedures (Noldus, 1991; Sanderson et al., 1994). Harrison and Baecker have observed that the design space of video analysis tools is enormously variable and yet revolves around an inherant design tension between needs for open-ended recording and structured methods: typically, the price of flexibility is a lack of rigour. For example, Weber and Poon's Marquee (1994) was designed for unstructured video-tape annotation and would struggle to cope with the demands of other approaches to video analysis. Equally, highly structured tools tend to be inappropriate for analyses other than those from within the theoretical framework that gave rise to them, such as Losada and Markovitch's (1990) 'GroupAnalyzer' tool, based on Bales'(1983) SYMLOG group dynamics coding strategy. GroupAnalyzer is a highly specialized tool built to meet explicit methodological and theoretical objectives.

MacSHAPA[17] is similarly committed to a set of analytic objectives given by a manifesto for "exploratory sequential data analysis" (Sanderson et al., 1994). MacSHAPA is particularly noteworthy, in the context of this discussion, as Penny Sanderson, the primary force behind it, recognized two vital matters relating interactive behaviour to video analysis. The first is that interactive behaviour involves many dependencies, the shape of which are fundamental to its progress. MacSHAPA seeks to expose these by looking for relationships in the train of events recorded by an observer. The second is that these are likely to be highly complex, the product of an indeterminate number of contextual effects and factor interactions, and so an approach is required that is capable of looking at the set of recordings in relation to one another. MacSHAPA seeks to achieve this by allowing observational data to be cast in various ways through the definition of codes, the relationships between which may be reorganized independently of the recordings themselves. A variety of options for data visualization are also included. Temporal relations are an important organizing principle in MacSHAPA. MacSHAPA data, including comments, annotations and events, are associated with a particular point in time. The time-point principle renders these data particularly amenable to sequential and linear analysis but is less useful for analysing non-linear aspects of observations(Sanderson, McNeese & Zaff, 1994). Chapter Two

---

[17]Up-to-date information and software available from a website at the following address: http://www.aviation.uiuc.edu/institute/acadprog/epjp/MacShapa.html

made a case for considering behaviours not as point events but as ongoing Activity. An examination was proposed of both a notion of "behavioural investment" in activity and a "whilst" relation between Activities and non-exclusive Activities.

3.2 Qualitative insight and quantitative reliability

Paying heed to what is said in the course of interaction has a more recent history than the majority of the NVC research reported in Chapter One. Pragmatic linguistics concerns itself with language in use, contrasted with the theoretical abstractions of the likes of Chomsky, relying heavily on the analysis of conversations in various settings (Levinson, 1983). Much of interest may be gained from the scrupulous care of microanalysis, that is, subjecting fragments of an interaction to detailed and intensive examination. Conversation Analysis (CA) has been conspicuous in the provision of many insights into the mechanisms of discourse, including not only the spoken exchanges of individuals but also the way in which they organize their interactions around available resources (Greatbatch et al., 1995).

For all the benefits they may ultimately offer, the costs of microanalytic approaches are acknowledged to be heavy by even their most ardent protagonists. The investment of analytic time against real conversational time typically suffers disparities of orders of magnitude. In no small part, this very cost has motivated the development of many of the available computer-based recording tools. Furthermore, the tradition from which CA has sprung, Pragmatic Linguistics, does not operate within a methodological framework amenable to quantification. Observations are rarely operationalized in such a way as to allow meaningful numerical summary. They are more often concerned to expose the interaction structures enabling the generation and maintenance of shared understanding by the subjects of their observations. The intensive nature of CA and allied techniques invariably concentrate on a great amount of information extracted from a small number of sources - far too few to constitute a statistical sample suitable for the framing of general statements about the interactive behaviour concerned.

Some who stand to benefit from research in the CA mould are reluctant to ignore its lack of statistical foundation. Even proponents of a fundamentally qualitative approach suggest that some quantification would provide useful support to an otherwise discursive analysis (Silverman, 1985). Conversely, studies of interaction within the methodological traditions of the behavioural sciences typically employ performance-based metrics of interaction. These are subject to criticism on the part of qualitative investigators, as theoretically uninteresting and of doubtful utility. The pragmatic limitations to obtaining appropriate data for a manageable statistical sample has been seen by those working in the Linguistic tradition to be a loss of theoretical validity, on the basis of superficiality of the measures. Indeed, the very notion of the "representativeness" of a sample of conversations is often challenged by those who have adopted the CA approach. Nevertheless statistical viability brings in principle a sound and common basis for comparing findings from a variety of investigations.

Coding schemes have been developed for the systematic observation of various aspects of interpersonal interaction in association with the content of talk (Bull, 1987; Bull, 1994; Thomas, Bull & Roger, 1982). Conversational Games Analysis attempts to lend some generality to analyses of interaction by quantifying instances of various communicative actions revealed through speech (Newlands, 1998). However, the costs of microanalysis, when applied to an appropriate statistical sample, rapidly escalate.

It should be made clear at this juncture that this thesis does not set out to challenge the methodological validity or even the pragmatic value of qualitative research. Activity Set Analysis is set within a tradition of behaviour research on the basis of quantification and so promotes a particular view of observational and analytic rigour. This thesis asserts the value of quantification to bring additional insights to the understanding of interactive behaviour. Quantification of the great detail of interaction stands to allow the discovery of relationships between recorded behaviours that might otherwise be very difficult to discern. It further sets out quite specific findings on the basis of repeatable method so that other investigators might attempt to disconfirm or replicate the reported results. These are matters that simply do not enter into the discourse of qualitative research and yet have been vital to the building of the various bodies of scientific knowledge. The work reported here is framed within the concerns of Human Computer Interaction research. As such, methodological purity takes second place to value in application. The value of quantification as described here is thus to be demonstrated by the light it throws on the use of video links, to be described in subsequent chapters.

Chapter Two described an approach to recording interactive behaviour, to exploit a variety of dimensions of observation, using a simple process-oriented behavioural classification method. For the Activity Set model to afford any practical value, an implementation is required. Action Recorder was developed for just this purpose and is described over the rest of this chapter.

3.3 Action Recorder

A general purpose event recorder, Action Recorder, was developed to implement Activity Set Analysis, as described in Chapter Two. Action Recorder was written in the HyperTalk programming language of Apple's HyperCard environment. It was intended to be a low cost and more expedient means of recording interactive behaviours than frame-by-frame microanalysis, although it can also support this type of work. Action Recorder's main purpose is to allow an observer of video recordings to record state information about people, the things they work with and the conditions of their work, but logging at a real-time pace. Action Recorder, complete with documentation, is available by file transfer protocol from the following location:

ftp://ftp.york.ac.uk/pub/users/psyc11/law4/ActionRecorder.sea.hqx

This section gives an overview of the principles and operation of Action Recorder. A full user manual is available as Appendix 1a. Some implementation details are included in Appendix 1b, with HyperTalk code for some of Action Recorder's more important functions.

There are four stages to data collection with Action Recorder: i. defining behaviours to record; ii, recording Activity states; iii, checking the recorded Activity state logs, and iv, compiling the logs into a table of states that obtained over the course of the interaction, the "Time Slice Table". An overview of each of these stages is followed by a full user manual for Action Recorder.

3.3.1 Defining Activities

The analyst decides which aspects of the interaction are of interest and defines keyboard keys to represent them. Action Recorder belongs to the "event recorder" software genre. Observers note events as they occur, the events being some salient behaviour. However, Action Recorder, following the Activity Set approach to interactive behaviour, takes observer action as signals of state change or state assertion in a particular observational dimension. In other words, the observers note of particular 'Activity Set' is taken as a statement about the condition of the Activity Set to which it belongs. As discussed in Chapter Two, membership of an Activity Set is determined by a simple rule of mutual exclusivity: if an Activity can occur concurrently with another activity, then those two activities belong on different dimensions (i.e. to different Activity Sets); if the occurrence of one precludes the occurrence of another, they may be members of the same Activity Set. Mutual-exclusivity should be effectively transitive for all Set members[18]. Typically, Activity Sets describe aspects of a particular individual's behaviour (*B* and *A* in the figure) however they might equally apply to environmental states. Although the effect of pressing a key in Action Recorder is to assert that some Activity is occurring, typically a key press would be made when an observer notes some change to a particular Activity Set state (i.e. a different Activity has been engaged by the individual concerned).

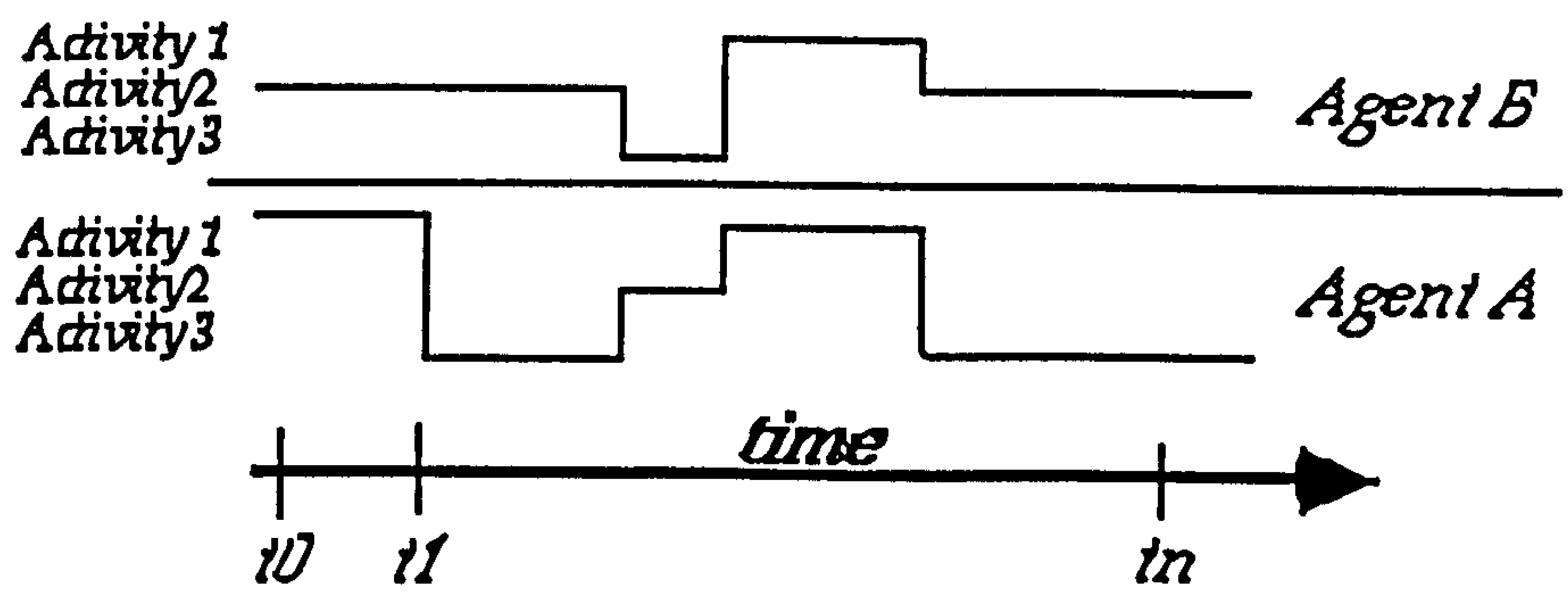


Figure 3.1 A polygraph schematic of the activities of Agents A and B

Having defined the Activity Sets, the observer is required to decided which of them to record in the coming recording session. Action Recorder automatically parses the definitions each time the observer goes through this decision process. The use of check boxes allows observers to record Activities from one, a subset, or all of the defined Activity Sets.

3.3.2 Recording Activity States

Action Recorder is now ready to begin logging observer keystrokes. This approach creates a digital equivalent of the kind of record produced by a polygraph: each Activity Set comprises all possible values as positions or values to mark out a particular trace (see Figure 3.1). The observer is required to cue a videotape at the starting point of the session to be analysed. Cueing

[18]Care should be taken to ensure that Activity Set membership is *fully* exclusive: having confidence that one particular behaviour precludes performance of another particular behaviour is not sufficient to ensure that is precludes all others in a candidate Set.

means finding the right point in the video recording and setting the VTR to "pause".

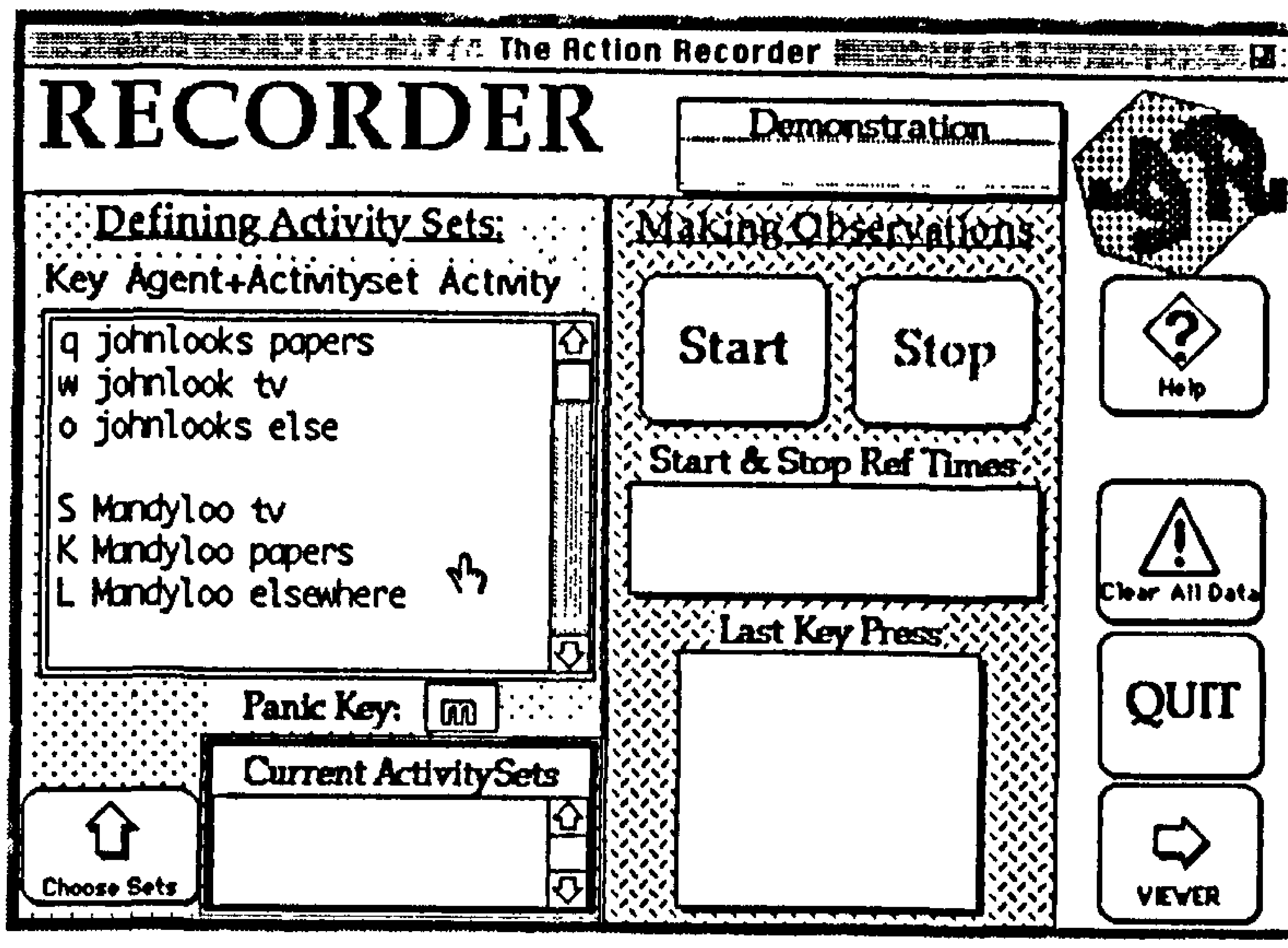


Figure 3.2 Action Recorder's definition and recording screen

The observer clicks a button, Start, and Action Recorder presents him or her with one or more dialogue boxes to set the initial state of the interaction log, depending on the number of Activity Sets selected for recording (see Figure 3.2). Each dialogue includes the list of characters defined for that particular Activity Set and checks the observer's entered code accordingly. The observer is then required to type in a video reference time for the start of the session (from the VTR counter or some other form of time code associated with the videotape). Action Recorder is now set for recording and presents a final dialogue, confirming the start time and presenting a "Begin" button. "Begin" may be clicked, or return or enter pressed, at the same time as the VTR "pause" is released. In this way, Action Recorder's clock and the VTR are set to run in parallel. The observer proceeds to press the appropriate keys as Activities are engaged. Where the observer believes a 'm'istake has been made, it is possible to mark the log with a key "m", predefined for this purpose.

At the end of the recorded interaction, the observer clicks the "Stop" button (see Figure 3.2). Action Recorder then converts timing information from its internal clock into seconds. A list of all the keystrokes made by the observer is compiled into an editable field, with their respective time codes (see Appendix 1c for an example). Where the "m" key has been used, the time code is listed in a separate field for the observer to check back against the video recording and edit the log accordingly. The observer may then add further Activity Set records to this log by selecting a new combination and starting the process afresh.

3.3.3 Collating time-stamped keystrokes

Action Recorder organizes the keystroke Activity codes by sorting into temporal order and then sorting into a "Time Slice Table", according to Activity Set membership and time stamp, as time-stamped state descriptions. A "time-slice table" is a table where the columns comprise all defined Activity Sets plus time, duration and error information (see Table 3.1 and Appendix 1d).

johnlook	lizlok	TIME	Duration	Exception
q	u	035.00	1.33	
q	t	036.33	1.67	
w	t	038.00	0.57	
w	u	038.57	0.75	
q	u	039.32	0.36	
w	u	039.68	0.74	
o	u	040.42	1.88	

Table 3.1: The organization of an Action Recorder "Time Slice Table"

Each row then consists of a state vector made up of the characters, one for each set, standing for the Activity that obtained at each juncture in the recorded interaction, the time at which it became true, the duration for which it obtained and undefined characters, if any. Activity Set membership is checked actively against the observer's definitions each time to determine the grand Activity State of the individuals, systems and environments defined as of interest by the observer.

3.3.4 Analysing Time Slice Tables

Action Recorder saves Time Slice Tables together with SPSS syntax files (see Appendix 1d). SPSS includes a function "aggregate" that collapses cases (state vectors) together on the basis of specified variables (in this case, Activity Sets). Aggregation in SPSS allows the duration variable to be summed over cases where there is no change in Activity State within the specified Activity Sets. In this way, the Action Recorder syntax files for SPSS are capable of providing summary statistics for the frequency, mean duration and overall amount of time spent in each of the defined Activities (see Appendix 1e). Later chapters use Action Recorder to operationalize Activity Set Analysis as an approach to interactive behaviour. A full account of inter-Activity Set Analysis working from Action Recorder data is developed in Chapter Five.

3.4 A short user manual for Action Recorder

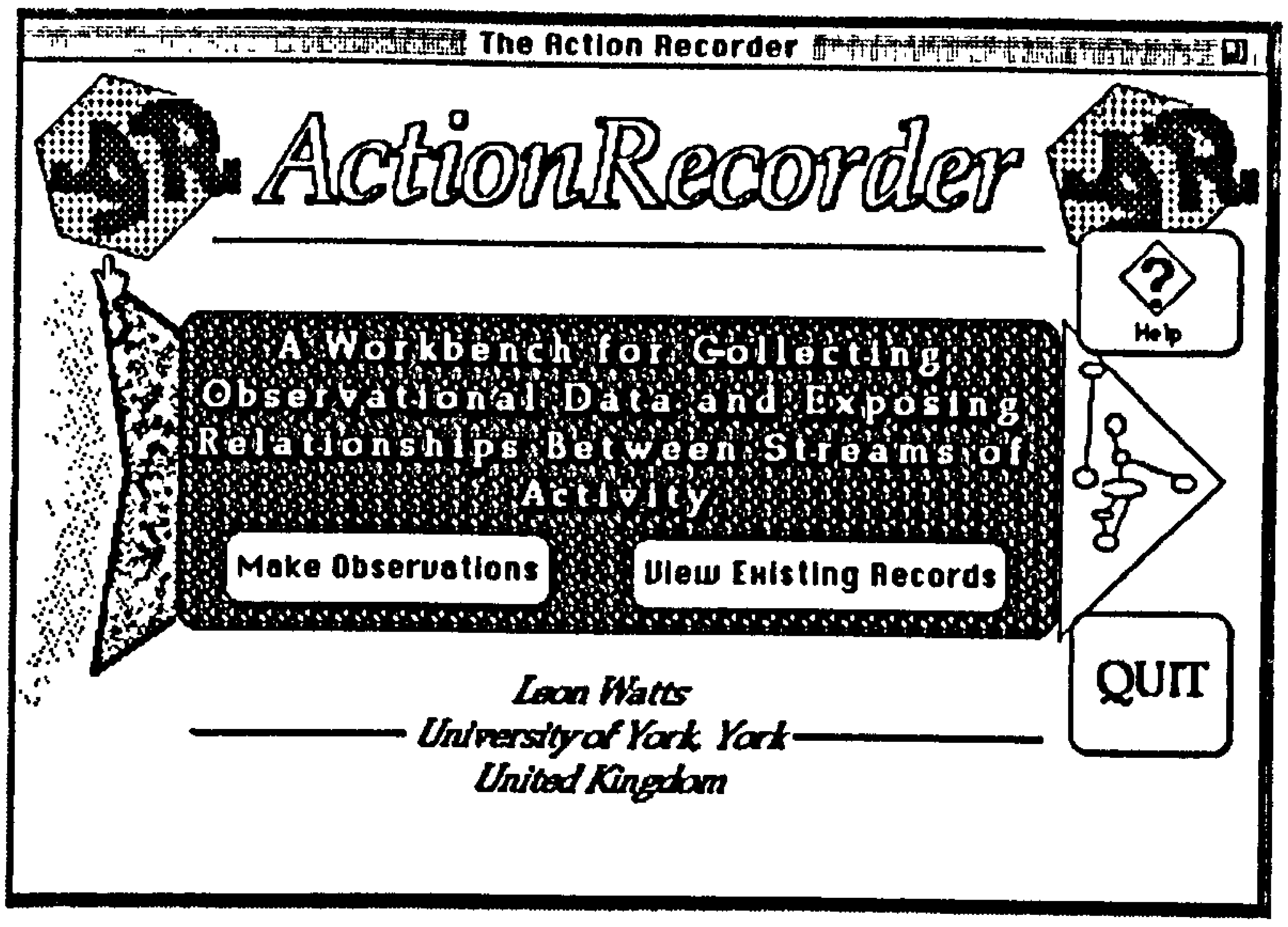


Figure 3.3: Action Recorder welcome screen

Action Recorder is a computer program that lets you record things as they happen, and then to inspect and analyse the log of your observations. It is made up of three parts: the RECORDER, the VIEWER and the TIMESLICER.

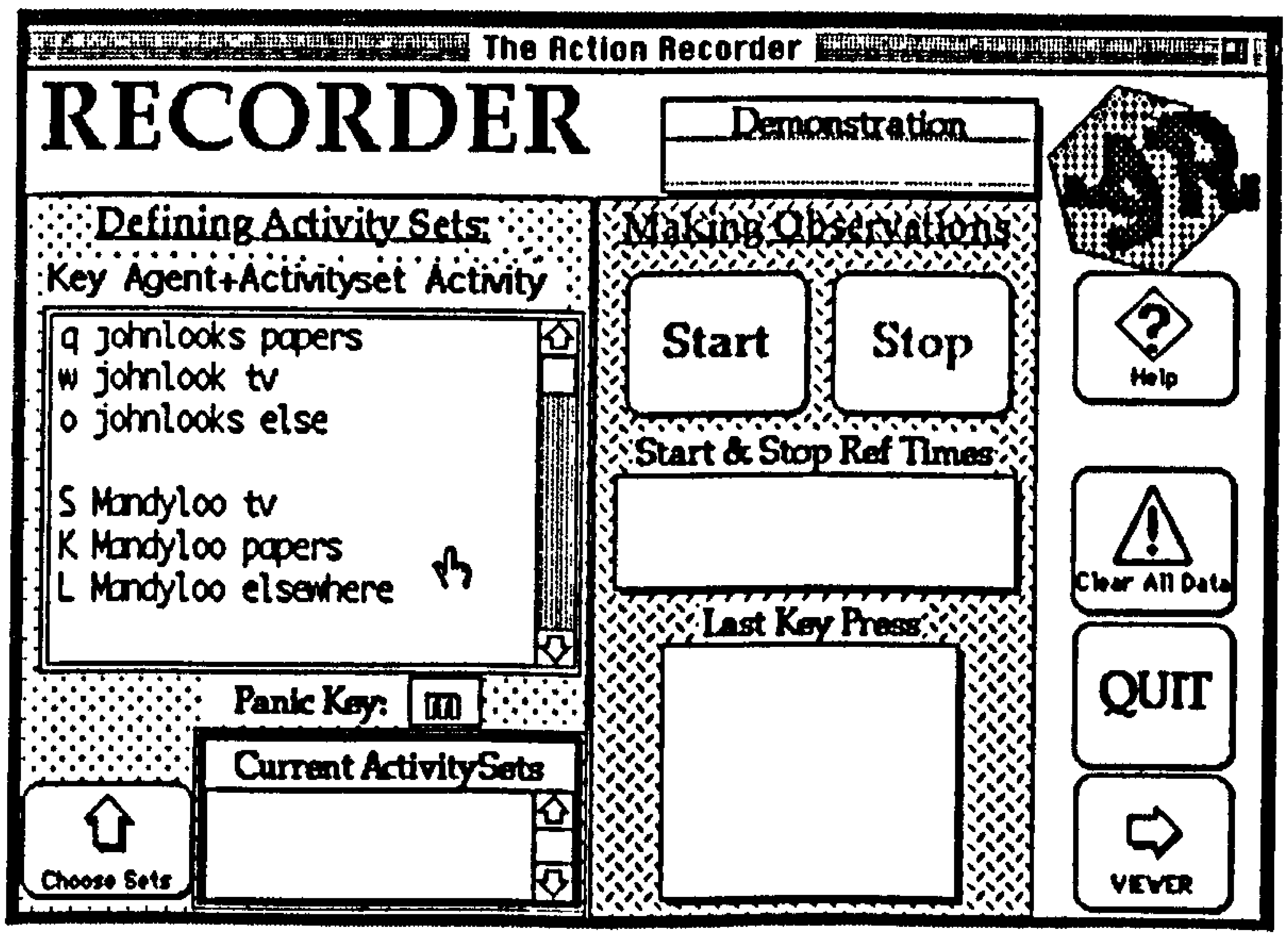


Figure 3.4: Action Recorder RECORDER screen

The RECORDER allows the collection of observational data in real-time, by coordinating observations taken with a reference time from the videotape itself. Observational recordings are organized by defining behaviours of interest as Activities in mutually exclusive groups - effectively different streams of behaviour.

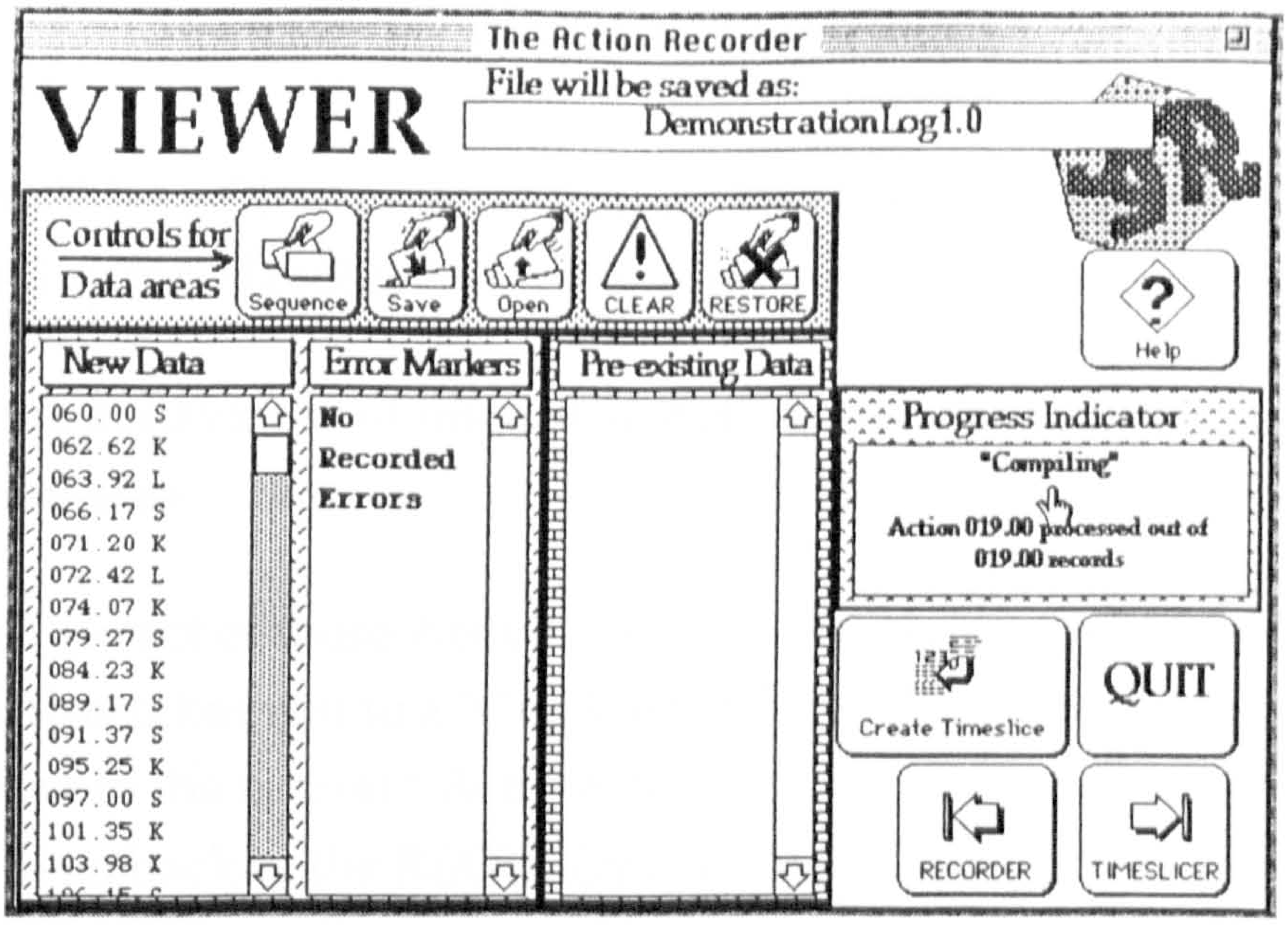


Figure 3.5: Action Recorder VIEWER screen

Data is examined and, if necessary, edited in the VIEWER. Previously recorded data may also be examined and merged with new data with this component of ***Action Recorder.*** At this point, all data is of the form *'time keystroke'*.

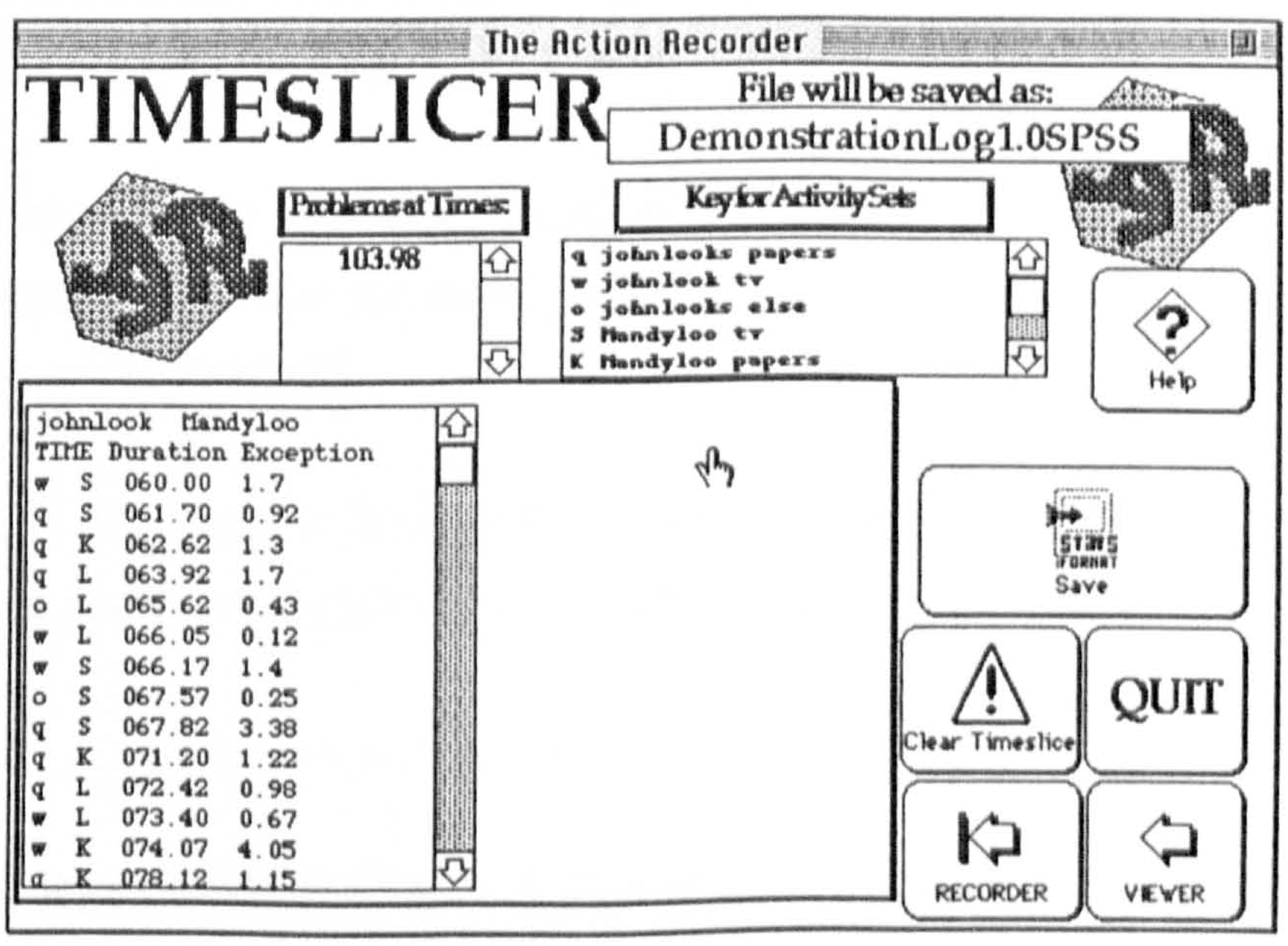


Figure 3.6: Action Recorder TIMESLICER screen

The TIMESLICER is an extension of the VIEWER. It is a state-based representation of data collected with the RECORDER, including checks for 'valid' keystrokes in the data and calculation of the duration for which each 'state', as defined by the status of each Activity Set at any point in the interaction, was true. Saving TIMESLICER data allows analysis with SPSS™.

TO RECORD NEW DATA

1. Click the "Make Observations" button. This takes you to the RECORDER, after asking for a label to index this observational recording session.

2. Define the behaviours of interest as Activity Sets in the Activity Set Definition area.

3. Choose a subset of these Activity Sets by clicking on the "Choose Sets" button. This takes you to a "CHOOSE ACTIVITY SETS" screen. Check the boxes next to the relevant Activity Set names. Click the "RECORDER" button to go back to the RECORDER screen.

4i. Click the "Start" button.

4ii Find the place in your videotape from which you wish to make your analysis.

4iii Enter the current state of behaviour in the Activity Sets, as visible on your video monitor, as Action Recorder prompts you for them.

4iv Enter a time code in seconds (not hours, minutes and seconds) for the place where you have paused your video. Preferably this is a time code in the video picture itself.

(4v. OPTIONAL enter a time code in seconds for the point at which you wish to end the section of videotape for analysis.)

---

5. ACTION RECORDER IS NOW PRIMED AND READY TO GO. You will see a dialogue box confirming your video reference times. Make sure you know which keys are defined to which behaviours. Turn to face the video monitor and, at the same moment, press the pause release on your video player and click the "Begin" button or hit the Macintosh keyboard ENTER key. As you see things relevant to the current Activity Sets, hit keys appropriately.

ADDING MORE DATA TO YOUR RECORDINGS

1. Go to the RECORDER by pressing the "RECORDER" button.

2. Click the "Choose Sets" button to select Activity Sets for observation this time around.

3. Click the "Start" button. This time, after answering Action Recorder's prompts about the current collection of Activity Sets for recording, you will be asked if you wish to "Add" more data to that which you previously recorded, or to "Delete" the previously recorded data. Choose "Add".

4. You will then be asked if you wish to carry on from the point you left off in the videotape, or if you wish to make your observations from some other point than this. Respond accordingly.

5. Type in the video reference (if adding from another point, for the start of the video segment) the end of the video segment.

6. If happy with the video references shown by Action Recorder, click the "Begin" button or hit the Macintosh keyboard ENTER key simultaneously with activating the video player.

CORRECTING ERRORS WHILST RECORDING

1. If an incorrect keystroke is made, and the correct keystroke is known, hit the correction key (defined as the Panic Key on the RECORDER screen) twice and the correct key once. The incorrect keystroke will be substituted for the correct keystroke.

2. If some kind of problem is detected with the observational recording process, hit the correction key once. This will act as a "marker" for later reference and correction.

CORRECTING ERRORS AFTER RECORDING

1. On the VIEWER screen, scroll through the New Data area until one of the video reference times shown in the "Error Markers" area appears. Next to this video reference time should be the letter defined as your correction key.

2. Find the equivalent reference in your videotape, and work out which key, if any should have been pressed.

3. Manually edit (delete and type in) the New Data area where the error marker was found, to bring Action Recorder's data into line with the videotape.

SAVING RECORDED DATA

1. Go to the VIEWER screen. Click the "Save" button in the "Controls for Data Areas". If you have data in both data areas, you will be asked which one to save.

2. A standard Macintosh file dialogue asks you to name and save your data, using the label for the recording session you provided at the start, appended with "Log#", where # is an integer.

3. Action Recorder confirms that the data has been saved by showing the file name with a full path.

RETRIEVING SAVED DATA

1. Go to the VIEWER screen. Click the "Open" button in the "Controls for Data Areas".

2. You will be asked if you wish to add the Activity Set definitions for the retrieved file's data to the Activity Set definitions area on the RECORDER. If you intend to go on to create a Time Slice Table (for statistical analysis), it is ESSENTIAL that you choose to "Add" rather than "Ignore".

3. If there is already some data in the "Pre-existing Data" area, you will be asked whether to add the retrieved data to it, or to delete these data and replace them with the data from the file.

Note: if you inadvertently click "Delete" here, the "RETRIEVE" button in the "Controls for Data Areas" with allow recovery of these data.

CREATING A TIME SLICE TABLE

1. Go to the VIEWER screen. Click the large "Create Timeslice" button .

2. If you have data in both Data areas, indicate which one to use.

3. Action Recorder will flash the TIMESLICER screen, and then work for some while to produce the Time Slice Table, but lets you know how it's doing via the "Progress Indicator" on the VIEWER screen. When done, Action Recorder chimes and shows a completion dialogue. Clicking "OK" shows the TIMESLICER screen, containing the Timeslice Table; a list of problem video reference times (if any were found), and an area defining keystrokes to activities, to aid interpretation.

4. If any video reference times are shown in the "Problems" area, scroll through the Time Slice Table until each video reference time in turn is found. Here, the offending keystrokes are displayed in the extreme right-hand column. As with the New Data (see above), use your videotape to work out what should have been recorded in each case.

Note: unlike the error correction in the New Data area, at this stage Action Recorder checks each recorded keystroke against the Activity Set definitions you have provided. So errors, where they occur, are instances of completely undefined keystrokes.

SAVING A TIMESLICE TABLE

1. Go to the TIMESLICER screen. Click the large "Stats Format Save" button.

2. A standard Macintosh file dialogue asks you to name and save your Time Slice Table data file, using the label for the recording session you provided at the start, or file name if the Timeslice Table has been created with retrieved data, appended with "SPSS".

3. Action Recorder will confirm the successful saving with the name of the Time Slice Table *data* file and full path, and immediately ask you if you wish to create an automatic analysis file for **SPSS**. You are strongly recommended to do this.

4. A standard Macintosh file dialogue asks you to name and save your Time Slice Table*analysis* file, using the label for the recording session you provided at the start, or file name if the Time Slice Table has been created with retrieved data, appended with "**SPSS.bat**".

5. Action Recorder will work for a little while, letting you know what it is doing via the "Progress Indicator" on the VIEWER. It will then chime and confirm the creation of the analysis, or command file, for SPSS™.

ANALYSING A TIME SLICE TABLE

1. Start SPSS™ for the Macintosh, version 4.0.

2. Choose "Open" from the "File" menu. A standard Macintosh® file dialogue asks you to find a file. Select the analysis, or command file (i.e. the "**.SPSS.bat**" NOT the "**..SPSS**" file).

3. Click in the "Input Window" of SPSS.

4. Select All.

5. Choose "Run Selection" from the "Run" menu, or type "R" with the command key, or simply hit the Macintosh "ENTER" key.

6. Inspect the statistical data in the SPSS "Output Window".

4 *Activity Set Analysis as a methodological approach to interactive behaviour*

Chapter summary

Activity Set Analysis is reported in application to video recordings of eight dyads carrying out role-playing discussion tasks over a VMI system. Observations of gaze and speech behaviour were operationalized with Activity Set Analytic principles, using the Action Recorder computer logging tool described in Chapter Three. Analysis of the resultant data is used as a paradigmatic demonstration of this approach to interactive behaviour. Gaze and Speech Simple Activity Set Analysis shows that about half of participants' time was spent looking at the screen of a shared computer link, a third looking at their own handwritten notes and a sixth looking towards one another, via the video link. However, these interactive resources were "referred to" by glance just as frequently as one another. Consideration of Activity Set complexes, as Co-activity states, suggested that participants tended to look at the shared link more when they were in the state SpeechSilence than when in Utterance, but GazeNotes and GazeVideo Activities were independent of Speech state. A 'standardized time' measure is recommended as the most meaningful Co-activity measure but some special problems relating to its measurement preclude any firm conclusions.

4.1 A case study of Activity Set Analysis

Chapter One discussed VMI research and identified gaze and speech Activity as a potentially revealing of VMI system usage. In Chapter Two, Activity Set Analysis was described as a way of thinking about and analysing interactive behaviour. Chapter Three presented a computer tool, Action Recorder, designed to take advantage of Activity Set analytic principles for recording behavioural data from live or recorded interactions. This chapter brings these ideas together in application to some real data. Four sections describe: (i) the rationale and design of a VMI case study, (ii) preliminary analysis, and (iii) a problem for analysing continuous rather than discrete data, as represented by temporal data.

An investigator interested in video-mediated communication might be informed by a great range of behavioural data, depending on the issue under consideration. Here, the VMI system is cast as a resource in an integrated system of people, objectives and resources. Its information potential is set against all the other resources for the purposes of those who use it. The better the contribution of the VMI system to the Activity system as a whole, the better able participants should be to do their work. In this way, one might assume that the information potential of a mediating environment will be expressed through the overall facility with which people are able to carry out specified pieces of work. However, as discussed in Chapter One, performance measures taken in experimental contexts have proven insensitive to VMI mediation, regardless of the actual metric adopted. Rather than looking at these "secondary" measures of mediation on interactive behaviour, the primary behavioural data, drawn from interactions as they occur, might provide some insight into the role of available resources in discharging the experimental task.

Reviewing research literature on video-mediation and visual non-verbal communication, Chapter One suggested that a direct approach to interactive behaviour, examining gaze itself in association with spoken communication, should expose any resource-based variations in communications media. Chapter Two described an approach to interactive behaviour, Activity Set Analysis, intended to show how observations of such behaviour may be organized in terms of the temporal characteristics of transient states, or activities. Establishing relationships between these observational activities might open a window on the underlying Activity system giving rise to such behaviour.

When people interact with one another, they do so in a space that may constrain their behaviour in a variety of ways. If there are high levels of ambient noise, they have to raise their voices. In a public space, conventions on acceptable forms of behaviour differ from those appropriate in a private space. Were such a space to be a library, conventions would be of a substantially different form to those applicable at, for example, a funfair or a shopping arcade. One might reasonably expect both gaze and speech behaviour to fall into the class of behaviours constrained by such conventions. The environment might almost be viewed as a full partner to interactive behaviour. Where people are working together, they invariably do so in amongst a rich set of resources. These might be tools in a workshop, flora in a garden or documents in an office. As discussed in Chapter Two, the status of artefacts in an interactive context may be conceived as a function of the "distributed cognition" of a work group.

Assuming that behaviour is not intrinsically wasteful or needlessly redundant, interactive behaviour is conditioned to exploit only those resources, visible and otherwise, that pertain to an agent's purposes. The visual environment in which people operate is tremendously rich with places to look. Only a subset of these, however, are of relevance at any moment. Within the context of an interaction, assumptions must be made about the physical materials of joint significance for those involved. Speakers and listeners must be just as sensitive to their role in the economy of communication as they are to the words they trade.

Effective exploitation of the information in the visual domain, coordination of turns at talk, and social forcefulness of gaze mean that, during interaction, participants must in some sense manage their gaze behaviour. Gaze distribution must be organized in terms of when and for how long peers should be looked at in order to hold their attention or to monitor their understanding; consult or add to visual information resources, such as text or graphics; organize such resources, perhaps by manipulation; scan for visual developments; and how much to attend to any additional visual information detected.

In this way, the case study described here is intended to show how Activity Set Analysis might be used as a vehicle to motivate questions about gaze distribution between individuals and their "interaction environment". It further raises questions about the visibility of a remote physical space as an information resource, as part of a larger work system.

4.2 Analysis of gaze deployment in an audio-video remote communications task: a case study of Activity Set Analysis

Some aspects of interactive behaviour were examined in a technologically-mediated communication task. The task was devised as a dyadic, negotiative, role-playing exercise, intended to exploit the "interpersonal and interparty" factors implicated in "visible" communications by, for example, Morely and Stephenson (1969) and Short (1974). The exercise required pairs of participants to discuss fictional canditates for hardship grants. They were asked to try to reach agreement on which three out of a total of ten student applicants for financial assistance should receive the help they requested. The situation was chosen to be relevant to subjects drawn from a student population and to which they might bring personal experience. The VMI environment, described in detail below, connected two physically separated rooms with a live synchronous audio-video link, a working surface and a computer link with shared text views. Each of these resources served as physically separate and perceptually discriminable objects for analytic purposes as well as to support the video-mediated negotiation task.

Measures of interaction focused on the looking and utterance behaviour of the individuals concerned, operationalized with reference to the principles of the Activity Set Analysis.

Gaze Activity Set	Speech Activity Set
GazeVideo: Looking at TV monitor, comprising head and torso image of distal[19] person and their immediate surroundings.	SpeechUtterance: Making a vocal contribution to the discussion.
GazeComputer: Looking towards the computer monitor, displaying screen-share of applicant information.	SpeechSilence: Not making any sort of sound that might be construed as a contribution to the discourse.
GazeNotes: Looking at paper-based notes taken in preparatory phase of experiment.	n/a
GazeElsewhere: Looking somewhere other than the first three locations.	n/a

Table 4.1: Defintions of Gaze and Speech Activity Sets

Looking or "Gaze" behaviour was divided into four Activities, including a default category, and all speech behaviour was generalized into two "Speech"

[19]'proximal' observations are those taken of a given physical location, and 'distal' observations are those relating to cross-link activity.

Activities as either Utterance or Silence. These are laid out in Table 4.1. Whenever Speech and Gaze Activities are referred to in the text, these operational definitions should be consulted.

4.2.1 Rating videotaped recordings

During the experimental sessions, each of the cameras providing images for the TV monitors at each end of the communications link, and also to two VCRs in an observation room. In this way, the audio and video transmissions from each workstation were recorded onto separate VHS videotapes. These tapes were subsequently rated one at a time, using the Activity Set operationalizations of gaze and speech behaviour described above. Each observation of a relevant Activity was to be recorded using an Apple macintosh IIvx personal computer, running a custom-built recording tool, Action Recorder' (see Chapter 3). The beginning of each session was indicated in the recordings by listening out for the point at which an experimenter said to the participants "when I leave the room, introduce yourselves in your 'professional capacities', to start the meeting." The raters found this point in the videotaped record and used it to benchmark Action Recorder's time codes. When the rater was ready to begin rating, s/he simultaneously pressed the pause button on a remote control device for the VCR and the 'enter' button on the computer keyboard. In this manner, the internal computer clock was aligned with the pre-recorded time of the VCR. The end point was given by the point at which the experimenter re-entered the room. Taking one Activity Set at a time (i.e. either Gaze or Speech, see Table 4.1), the observer watched the recorded session on a video player at normal running speed.

For gaze behaviour, the rater looked for changes in eye, head and posture, indicating a shift in visual focus, within the parameters described by the gaze Activity Set. Although the Speech Activity set was defined as a simple, binary state, it is marked by rather less clear-cut transitions than for the Gaze Activity Set (see Table 4.1). Onset of utterance was relatively straightforward as any vocalization issued by the observed, with a limited number of exceptions. Coughing, laughter and sighing were excluded as contributions, but all other vocalizations, such as 'uh huh' and 'um', were allowed as legitimate contributions in the form of backchannel speech (Yngve, 1970) or pause-filling. Raters were required, in so far as they were able, to log every instance where vocalization started and then when, in their judgement, the contribution came to an end. Several consecutive turns at talk may be taken by a given individual in conversation (Sacks et al., 1974). The cues for recognition of turn transition are various, including paralinguistic (e.g. changes of pitch or rhythm) and semantic attributes of talk (Duncan, 1972; Duncan & Niederehe, 1974). So raters were instructed to indicate the end of a 'bout' of speech by calling on their 'everyday knowledge' of conversational turn taking. A potential source of confusion is of a form which might be described as 'suspended speaking.' It is a frequent and natural feature of ordinary discourse that participants behave with uncertainty: as opinions are formed and consolidated, new directions for the discussion are explored, tentatively rejected, and so on. In these circumstances, it may be seen that a 'neat end' to an utterance is rare indeed. Offset of utterance was defined as a silence on the part of the speaker of such a length as to constitute an opportunity for the observer, should s/he have been the speaker's conversational partner, to have taken a turn at talk. Non-breaking silence would not be picked up by instructions to only record when people were actually vocalizing. Even if raters had been instructed to record vocal activity in this way, it is not at all clear that they would have been able to disengage their ordinary apparatus of conversation to the extent that they would have been able to ignore non-breaking silences. Spoken contributions were thus reflected in the data in a way that a speech-energy analysis, such as employed by Sellen (1992) and Vertegaal (1997; 1998), would not.

Since there were two individuals per session (i.e., dyadic interactions), the primary rater watched recordings of each session a total of four times, recording one Activity Set on each pass (two Activity Sets defined per participant).

4.2.2 Anticipated findings

This study has the dual purpose of (i) clarifying and demonstrating the Activity Set concept, and (ii) providing some descriptive data on gaze distribution in mediated communication. It is not a comparative evaluation of the effects of some experimental manipulation and so there are no hypotheses to test per se. The videotaped recordings were taken from a comparative study of audio-only and audio-video mediated communication. The audio-only condition did not lend itself to a meaningful comparison with the audio-video condition on measures of gaze deployment. However, something can be said about the expected outcome of applying Activity Set Analysis in this way. The foregoing discussion has outlined several grounds for supposing that the use of gaze in interaction is an active and effective communicative process. If an individual looks somewhere, they do so for some purpose. As such, some clear patterns of "visual resource exploitation" should become apparent.

The task chosen for this study relies upon paper- and electronically-based textual information to fuel collaborative decision making. The three gaze activities defined above identify "paper-based information" and "computer-based information" as task-relevant resources. The "looking at video" Activity differs from the others in two important ways: it is seen as a transmission as well as a reception device, and it offers subjective, interpersonal information about the people concerned. Adopting Yule's (1985) terminology, it offers both *interactional* and *transactional* information to the participants. In the first instance, it was expected that participants would consult all three visual resources repeatedly over the course of their interaction.

As discussed in Chapter One, gaze behaviour was expected to depend at least to some extent on speech behaviour. If it were the case that each visual information resource represented an equivalent utility and accessibility for the purposes of the participants, equivalent attention to each might be expected. However, the information given in the paper notes and computer resources was written and thus required a considerably different investment of the participants than for the video image. Furthermore, the video link was the means for the reception of status information about the distal partner and also transmission of the same to the distal partner. It was anticipated that this category of looking behaviour would demonstrate a markedly different contingency with Speech, compared to the paper GazeNotes and GazeComputer Activities[20]. Correspondingly, the GazeNotes and GazeComputer categories should be associated with similar patterns of Activity. GazeComputer and GazeNotes activities were not expected to absorb the same amount of effort, since these differ on a number of dimensions from familiarity to opportunity for deictic expression.

4.3 Method

4.3.1 Participants

Sixteen individuals were each paid £6 to attend two, one hour experimental sessions to carry out mock negotiations in same-sex dyads. Approximately half of these were recruited via a local email network. Dyads were composed of twelve male and four female members of York University, from a wide

[20]Refer to Table 4.1 for interpretation of these terms.

range of degree course backgrounds. Age range was approximately 20 to 30 years of age.

4.3.2 Design

Dyads negotiated over fictitious financial hardship claims in each of two communication conditions, audio-only and audio-video combined. Videotaped recordings of interactions in the audio-video combined condition were used as the subject matter for the case-study reported in this chapter, the audio-only condition not lending itself to meaningful gaze deployment analysis[21]. Each interaction was between unique same-sex dyads, who reported that they had not met prior to the experiment. Two sets of materials were devised to generate discussion, together with instructions for two alternative sets of decision priorities for the task.

4.3.3 Materials

Information

Two electronic versions of ten sets of 'hardship application forms', representing requests for financial assistance by (fictitious) students at a fictitious institute of higher education (Appendix 2b).

Blank paper forms, to contextualize the electronic information provided by the 'Aspects' software (Appendix 2a).

One set of general instructions for 'assessors' of forms, outlining the general nature of the task (Appendix 2c).

Two sets of specific instructions for assessors of forms, directing the assessor to make a particular effort to promote the interests of one of two hardship groups, namely: being disadvantaged as a result of family circumstance or disability. (Appendix 2d).

---

[21]See Daly-Jones, Monk and Watts (1998) for a report on the comparative aspects of this experiment, use different dependent measures.

Equipment in each room:

A Finlux 'Black Design' 26" colour television, type 3028E.

An Apple Macintosh IIvx personal computer with large, colour display, running Group Technologies' "Aspects" screen-sharing/joint-editing software.[22]

A chair, fixed desk and writing materials.

A 'Panasonic' VHS-C portable cam-corder, type NV-G1, with 'Sigma' 'Wide Converter' x 0.5 teleconverter.

A 'Realistic PZM' microphone.

A 'Panasonic' radio-frequency (RF) adapter, type VW-RF8.

Telecommunication connections

Combined audio and video in PAL format, transmitted as radio-frequency signals to the link monitors over coaxial cables.

'Aspects' software (Group Technologies, Inc., 1990) linking the two computers over an 'Appletalk' local area network.

Equipment in the Observation Suite:

Two 'Panasonic' VCRs, type NV-F 76 HQ, with a 'Panasonic Editing Controller', type VW-EC310.

A 'Panasonic' digital video mixing desk, type WJ-AVE5.

A 'Panasonic' character generator, type VW-CG2E.

4.3.4 Procedure

Each session began with the introduction of participants (Ss) to one another and a resumé of the experimental aim. Ss were informed that the study was to examine how people communicate with one-another via mediating technologies. It was explained that, after having prepared some arguments to support particular applicants, Ss would be expected to role-play the university assessors for a series of applications for financial support, in a meeting lasting about half an hour.

Although it was stressed that assessing applications for financial support on grounds of some hardship is a real-world task, equally, it was made clear that the university in which the task was to be set, the applicants whose cases the forms represented, and the forms themselves, were fictitious. The meeting was described as an opportunity to resolve differences between the assessors

[22]In one room, an 'Apple Macintosh IIvx' personal computer with an 'Apple Macintosh 16" Colour Display', in the other an 'Apple Macintosh II' personal computer with a 21" 'SuperMac' colour screen, running 'SmallScreen' emulate a 16" colour screen.

over a shortlist of candidates, subject to a requirement to encourage a particular type of candidate. Each assessor was said to have been seconded to the task from a particular university college. Each college had a particular viewpoint on the type of candidate it wished to encourage, and this was to be expressed through their appointed assessor. Three candidates from the shortlist of ten were to be offered funding. No provision was to be made for division of the available awards. Full, in-role instructions for Ss can be found in Appendices 2c and 2d.

Ss were given copies of the general information for assessors, including their particular decision priority, and a blank form. Questions were encouraged by the experimenters (Es) and answered within the general framework outlined above. Since the study was focused on the process of discussion, it was suggested that the discussion need not necessarily lead to any final conclusion within the allotted time, although they should try to work towards consensus.

Ss were taken to their respective ends of the communication link (two offices at opposite ends of a building). Seating positions were established 24" away from, and directly facing, the computer monitors and 26" away from, and at right angles to, the televisions (see Figure 4.1). In the V condition, the view of each participant was from the waist up, the image occupying roughly 30% of the 26" TV screen area. The monophonic audio afforded good speech quality, permitting fluent conversation at normal volumes, allowing for overlapping speech and interruption without distortion or break-up.

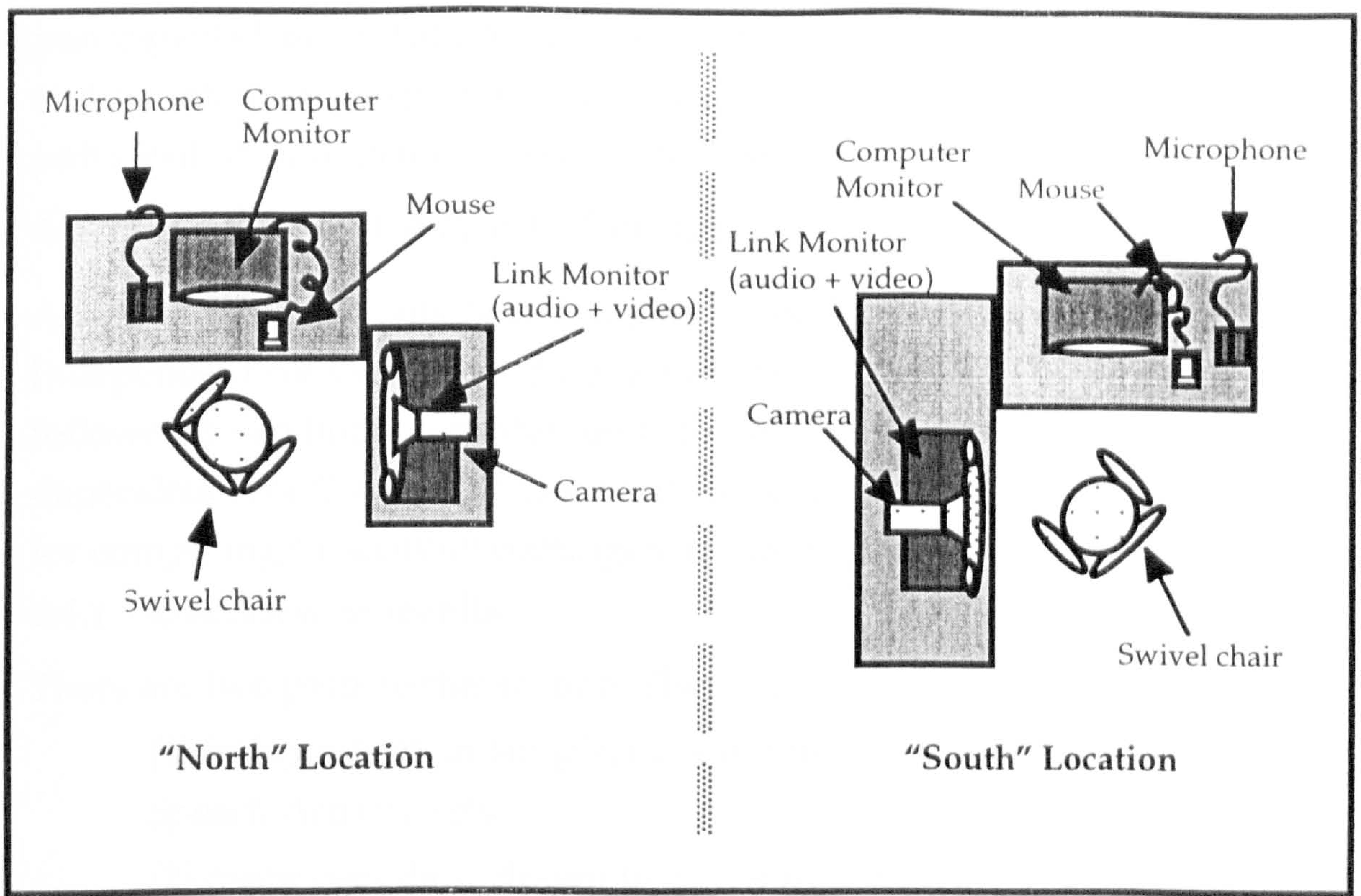


Figure 4.1: Schematic diagram of audio-video and linked computer workstations used in case study

Experimenters gave basic instruction on how to use the computer-displayed documents, by scrolling with a mouse input device. No keyboards were fitted. Ss were allotted fifteen to twenty minutes to overview electronic copies of the ten application forms in the set (Appendices 2a & 2b). Ss were instructed to identify the candidates corresponding to their particular selection criteria (as indicated in their specific guidelines document) and to take notes to support arguments in favour of these candidates. They were asked not to begin their discussion of the forms until their review period was terminated by the experimenters. Experimenters left the experimental rooms. The document navigation mechanism (scrolling) with the 'Aspects' software (Group Technologies, Inc., 1990) was not linked at this stage.

After the allotted review time, or until it was apparent from the observation room that participants had completed the note-taking Activity, experimenters re-entered the experimental rooms. The linked scrolling facility of the Aspects software was enabled, so that the view of the electronic forms was identical from both machines. Participants were instructed to introduce themselves to one another in their 'professional capacities' (i.e. in role) and told that Es would re-enter their conferencing room after around twenty minutes. The VCRs, located in the briefing/observation room, were started.

Experimenters monitored dyads progress throughout the experiment from a video observation suite. After either twenty-five minutes had elapsed, or participants had agreed on three of the applications, experimenters knocked and re-entered the experimental rooms. Participants were thanked for their participation and, in the case of the final session, paid.

4.4 Analysis of Activity Set data

Analysis of two Activity Sets, Gaze and Speech, are reported (see Table 4.1). Independent consideration of Gaze and Speech Activity Sets is given first, followed by an initial consideration of inter-Activity Set temporal dependency, or Co-activity, is set out. Issues in dealing with proportional data for computing Co-activity contingencies are raised.

4.4.1 Overview of results

There are two parts to this section. These are:

(1) independent, or simple, data summaries for the Gaze and the Speech Activity sets.

(2) contingent data, drawn from the **two-way, within-participants conjunctions** of the Speech and Gaze Activity Sets.

Each includes tabular and graphical data summaries, with statistical analyses where appropriate. All data are given to two decimal places.[23] For clarity of presentation, later subsections include some tabulated information repeating the findings of earlier tables, although this will not be referred to in the text unless to disambiguate some marginal effects. Data demonstrating statistical validity and inter-rater reliability follow each section.Two levels of significance are set for evidence of an effect. These are: $p < 0.05$, for 'significant', and $p < 0.01$ for 'highly significant'. Where the determined 'p' values fail to reach these levels, following the exploratory nature of this study, the calculated value is presented. Standard deviations (SD) of the mean for each measure are tabulated in parentheses. Column and line graphs plotting means of the various measures also plot standard deviations of the mean as error bars, rather than plotting the more conventional standard error in this way.

As discussed above, the goals of this analysis are: (a) to illustrate the Activity Set Analysis approach to interactive behaviour and (b) to provide some

---

[23]The data presented appear to reflect accuracy of up to 100th of a second. However, the Apple Macintosh computer on which the recordings were made implements time in units of a sixtieth of a second. Two decimal places are required to express this. The meaningfulness of all results is reserved for the Discussion.

baseline information as a quantitative natural history of gaze distribution and utterance dynamics in video-mediated communication. Some attempt is thus made to explain where Activity Set Analysis has discriminated relationships between Activity Sets and to interpret any effects in terms of video-mediated communication.

4.5 Simple Activity Set data: analysis of separate Activity Sets

The tables in this section comprise various measures which examine the occurrence of each Activity in a given Activity Set, without regard for their concurrency with other Activities. They are described as "non-contingent" summaries of the data, since any relationships or contingencies between simultaneously occurring Activities are ignored at this stage. Data on Gaze are followed by data on Speech Activity. In each case, the data are taken from 16 individuals.

4.5.1 Gaze Activity Set

Taking a straight average of all the Gaze data from the 16 participants (see Table 4.2), overall, most time appears to be spent in the GazeComputer Activity. The time spent in GazeVideo is just a third of this figure and GazeNotes appears to involve a temporal investment somewhere between the two. The default, GazeElsewhere occupied participants for neglible time over the course of their discussions.

Gaze Activity	Total Time in Each Activity (SD)	Activity Frequency (SD)	Mean Duration of Each Activity (SD)	Maximum Activity Duration	Minimum Activity Duration
Video	155.55 (77.31)	76.94 (31.84)	2.09 (0.81)	12.43	0.23
Computer	431.14 (230.53)	77.44 (38.20)	5.58 (1.86)	35.88	0.21
Notes	290.70 (190.39)	71.56 (28.23)	3.92 (1.99)	29.20	0.23
Elsewhere	18.03 (19.45)	18.88 (16.46)	1.00 (0.50)	3.68	0.25

Table 4.2: Gaze Activity Set: Temporal (in seconds) and Frequency Summary Statistics

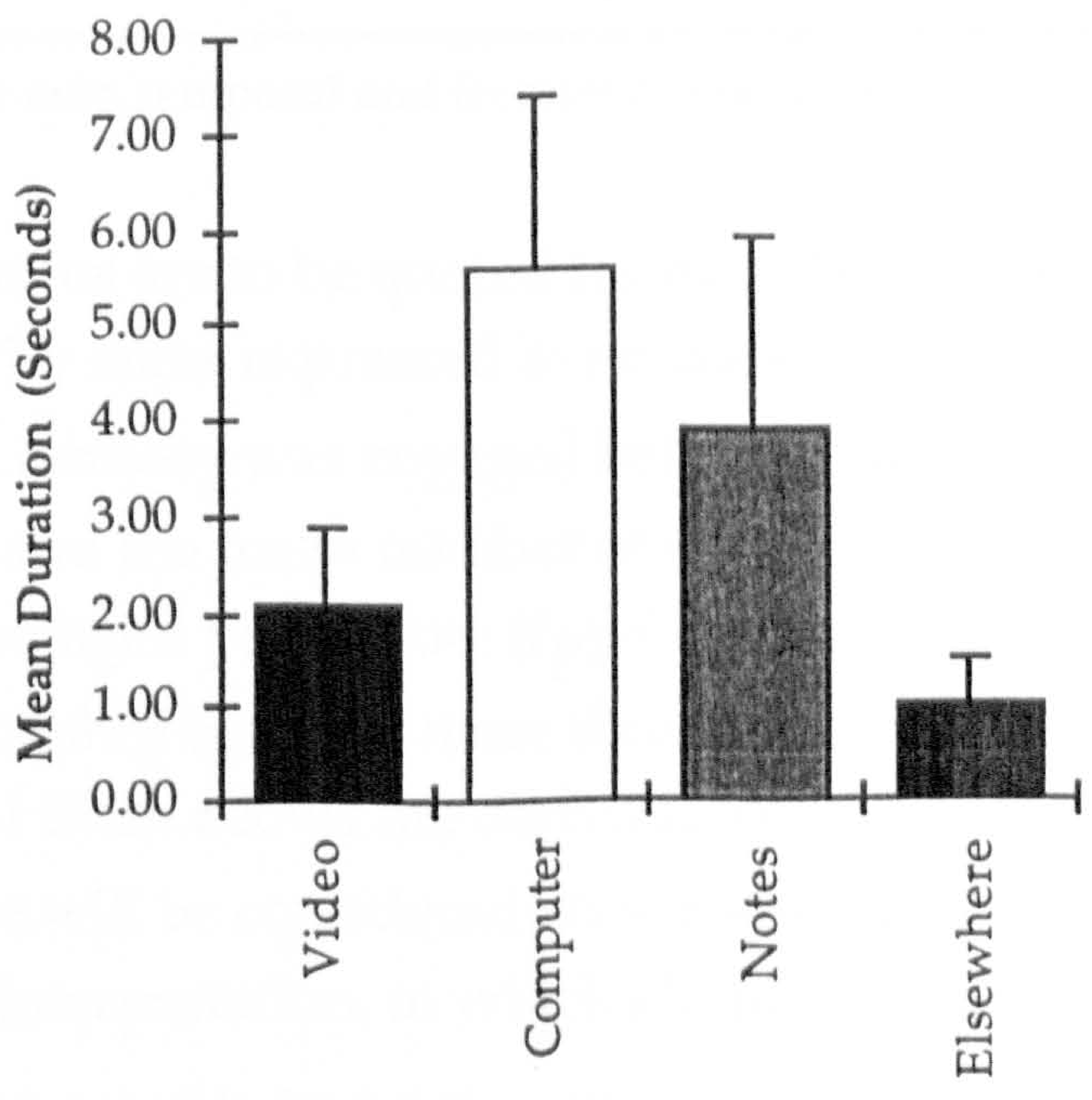
Frequency data, when treated in a similarly direct way, do not discriminate between GazeVideo, GazeNotes or GazeComputer. The default Gaze Activity, GazeElsewhere, again appears to contribute little, as a comparatively infrequent state.

These crude measures of amount of time in and frequency of Activity state occurrence do not allow for the variability in length of experimental discussion (mean length=923secs., median length=768 secs., standard deviation=350 secs.). Consequently, frequency and total time summaries disproportionately reflect the behaviour of dyads in sessions of longer total duration. Comparisons of 'mean duration of Activity' are not subject to this uncontrolled variability between dyads. This measure refers to the average length of each instance of a given Activity state, for each experimental session.

The default GazeElsewhere Activity state is comparable to the other three Gaze Activities by mean duration of occurrence. However, the very low frequency and overall time spent in GazeElsewhere indicate that it is, as expected, to be of vanishing importance. Henceforward, the GazeElsewhere Activity state will not be included in the data or statistical analyses presented. It is noteworthy that the Gaze Activities defined for these analyses were sufficiently inclusive to allow dismissal of the GazeElsewhere Activity state.

As Table 4.2 shows, the mean duration of the Activity state GazeComputer was just over 5.5 seconds, longer than either GazeNotes or GazeVideo. At nearly 4 seconds, participants engaged in GazeNotes for about two thirds of the Computer figure. The mean duration of GazeVideo Activity was comparatively brief, at around 2 seconds long. Figure 4.2 presents these data as a histogram. The data were subject to an analysis of variance, treating GazeActivity as a variable with three levels, and these differences were found to be highly significant ($F(2, 30) = 16.88$, ; $p < 0.01$, $MSe = 2.89$. $Q = 3.49$, Tukey's $HSD = 1.48$).

Figure 4.2 - Mean duration of Gaze Activities



4.5.2 Three base measures of Activity

As discussed above, simply averaging across all dyads to determine the time in each Activity, or of the frequency with which individuals engaged in particular activities, does not account for differences in overall time taken. To compensate for session length variability, some standardized metrics are required. For total-time-in-Activity, a simple proportion of the session length would be appropriate. However, since the data here are time based, there are some advantages to standardizing on a temporal unit. By employing a standardized period, as well as dealing with biasing, it becomes possible to interpret frequency and time-in-Activity together in an intuitive manner. Here, a standard period of 'one minute' is adopted. Also, once the 'number of seconds in Activity state per minute' and 'number of occurrences of Activity state per minute' are known, the duration of Activity is readily determinable as the quotient of standardized time and frequency. In Table 4.3, standard time-in-state is described as 'seconds per minute', and the standard frequency data as 'occurrences per minute.'

Gaze Activity	Standard time in seconds/min (SD)		Frequency of looks in occurrences/min (SD)	
GazeVideo	11.02	(5.87)	5.27	(1.85)
GazeComputer	28.67	(10.76)	5.30	(1.89)
GazeNotes	19.17	(10.47)	4.84	(1.16)
elsewhere	1.14	(0.88)	1.24	(0.88)

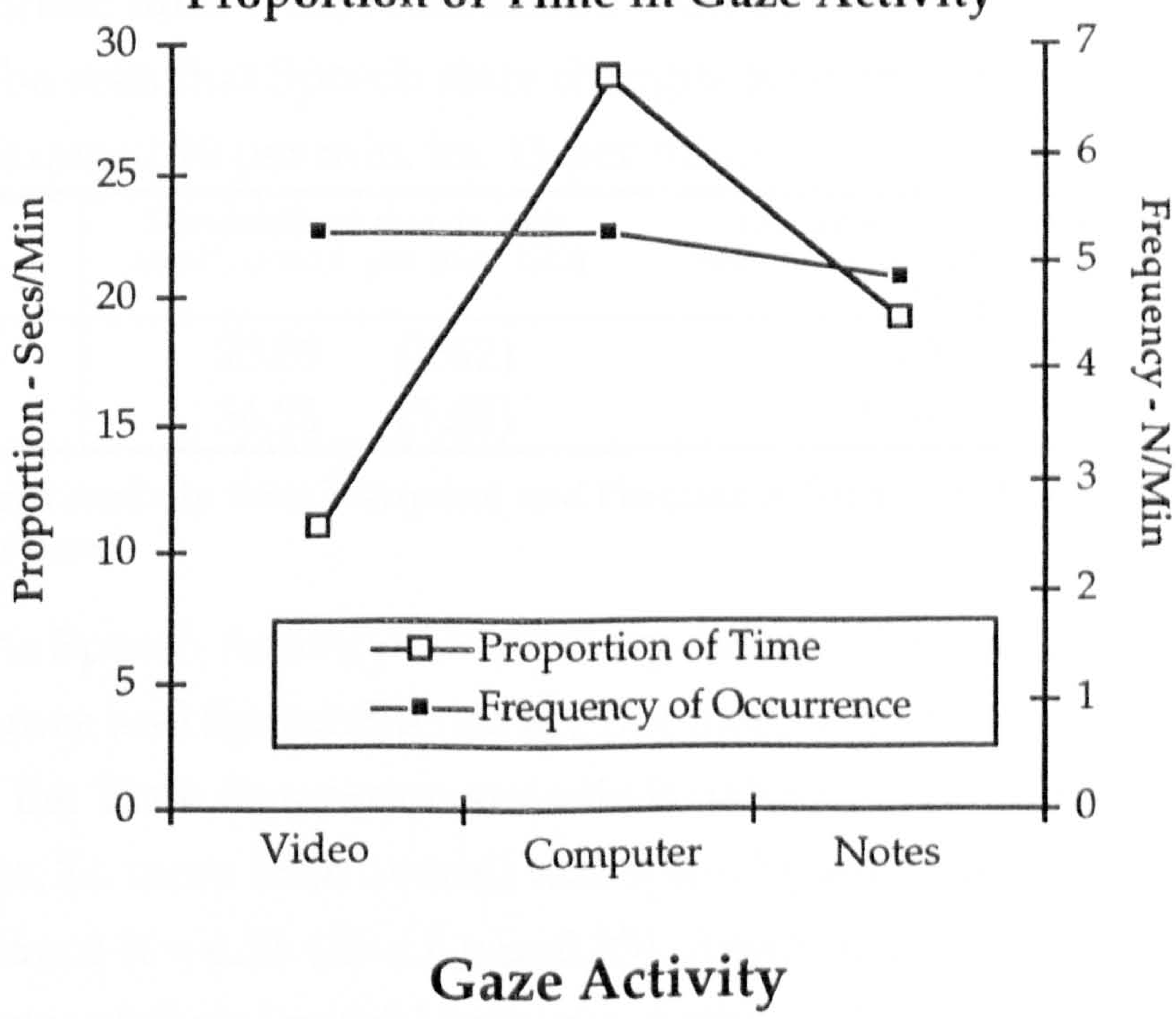
Table 4.3: Gaze Activity sets, temporal and frequency statistics, standardized to a common time base

So, three basic measures are to be quoted for each Activity: the overall time invested in an Activity state, expressed in seconds per minute (t_p); the mean time for which each Activity was engaged before some other Activity from the same Set began (d), and the mean number of occasions upon which the Activity state was engaged per minute (f_p). Clearly, there is a simple mathematical relationship between these three measures, such that the knowledge of two of them allows the derivation of the third: ($t_p = d \times f_p$). For the moment all three will be considered, as some of the later analyses require some non-intuitive interpretation, in which all three assist clarity.

Duration data are intended to be read as evidence of the length of time some Activity needs to be engaged in such that some value is obtained by the actor from carrying it out, assumed to be motivated by some facet of the ongoing interaction. Frequency data are an expression of the number of occasions a motive for initiating an Activity reached a point of execution. Standard time in Activity state can be read as reflecting the overall investment in each of the observed activities, for the period of their recording. Figure 4.1 presents the duration data for Gaze Activities and Figure 4.2 presents the corresponding standard time and frequency data.

Overall, there is no evidence that the three Gaze Activities occurred at different frequencies ($F(2, 30) = 0.55$; $p = 0.581$). However, the time proportional data show that participants spent less time in GazeNotes than GazeComputer, and still less time in GazeVideo ($F(2, 30) = 9.62$; $p < 0.05$, mse=129.8). Following up with Tukey's HSD (Tukey's HSD=9.94, Q=3.49), it can be seen that although participants spent significantly longer in GazeComputer than GazeVideo (difference between means = 17.65), neither the differences between the GazeNotes and GazeComputer (9.50) nor those between GazeNotes and GazeVideo (8.15) quite reach significance.

Figure 4.3 - Contrasting Frequency with Proportion of Time in Gaze Activity



The Figure 4.3 contrasts the similarity of Gaze Activity frequency with differences in the standard time spent in each Gaze Activity. So, although almost half of the standard minute was spent in GazeComputer, compared with 20 seconds in GazeNotes and 10 seconds engaged in GazeVideo link, participants looked in each of these three directions just as often. Figure 4.2

reflects this difference, showing the mean duration of each Gaze Activity in histogram form.

4.5.3 Speech Activity Set

Measures of vocal Activity show that bouts of SpeechUtterance (i.e. mean duration of occasion of Activity engagement) are shorter than bouts of SpeechSilence ($t = 3.44$ (15 d.f.); $p < 0.05$), at around 2.5 seconds compared to around 4 seconds (see Table 4.4 and Figures 4.4 & 4.5).

Speech Activity	Total Time in Each Activity (SD)	Number of Occurrences[24] (SD)	Mean Duration * of Activity (SD)	Maximum Duration	Minimum Duration
Utterance	341.28 (128.25)	141.88 (61.89)	2.57 (0.82)	17.96	0.18
Silence	554.13 (210.81)	142.06 (61.65)	4.27 (1.89)	32.01	0.20

Table 4.4: Speech Activity Sets: Temporal (in seconds) and Frequency Summary Statistics

As discussed above, raw data on total time and frequency of Activity occurrence does not lend itself to meaningful analysis. Hence, Table 4.5 presents the standard time in state and frequency of state shift[25], standardized to the one minute unit. When considered with the Gaze Activity frequency data, it may be seen that Speech state changes occurred less frequently than Gaze shifts (around 10 per min. vs. 15 per min.).

	Standardized time in each state*, in secs. per min. (SD)	Frequency of each state, in number of occurrences per min (SD)
Utterance	23.09 (5.62)	9.38 (2.17)
Silence	36.91 (5.62)	9.40 (2.18)

Table 4.5: Speech Activity Sets: Temporal and Frequency Statistics, Scaled to a Common Time Base

Given that the Speech Activity Set is binary, relative frequencies of SpeechUtterance and SpeechSilence are not meaningful and the information provided by the Time Proportion statistic is naturally congruent with the Duration data. So more time overall was spent by each participant in Silence than in Utterance ($t = 4.56$ (15 d.f.); $p<0.05$). Assuming that participants broadly alternated their SpeechUtterance Activity, the dyad spent most of

24Note that a rounding error has introduced a small disparity between the mean frequency of Utterance and Silence.

25The 'Frequency' figures here serve to indicate the overall rate of change of utterance. Since there are only two states (i.e. on or off) they must alternate and hence there is no contrast between the Utterance and Silence on this measure.

the discussion time in conversation with each other (the sum of individual SpeechUtterance standard time = 23.09 + 23.09 = 47 seconds per minute).

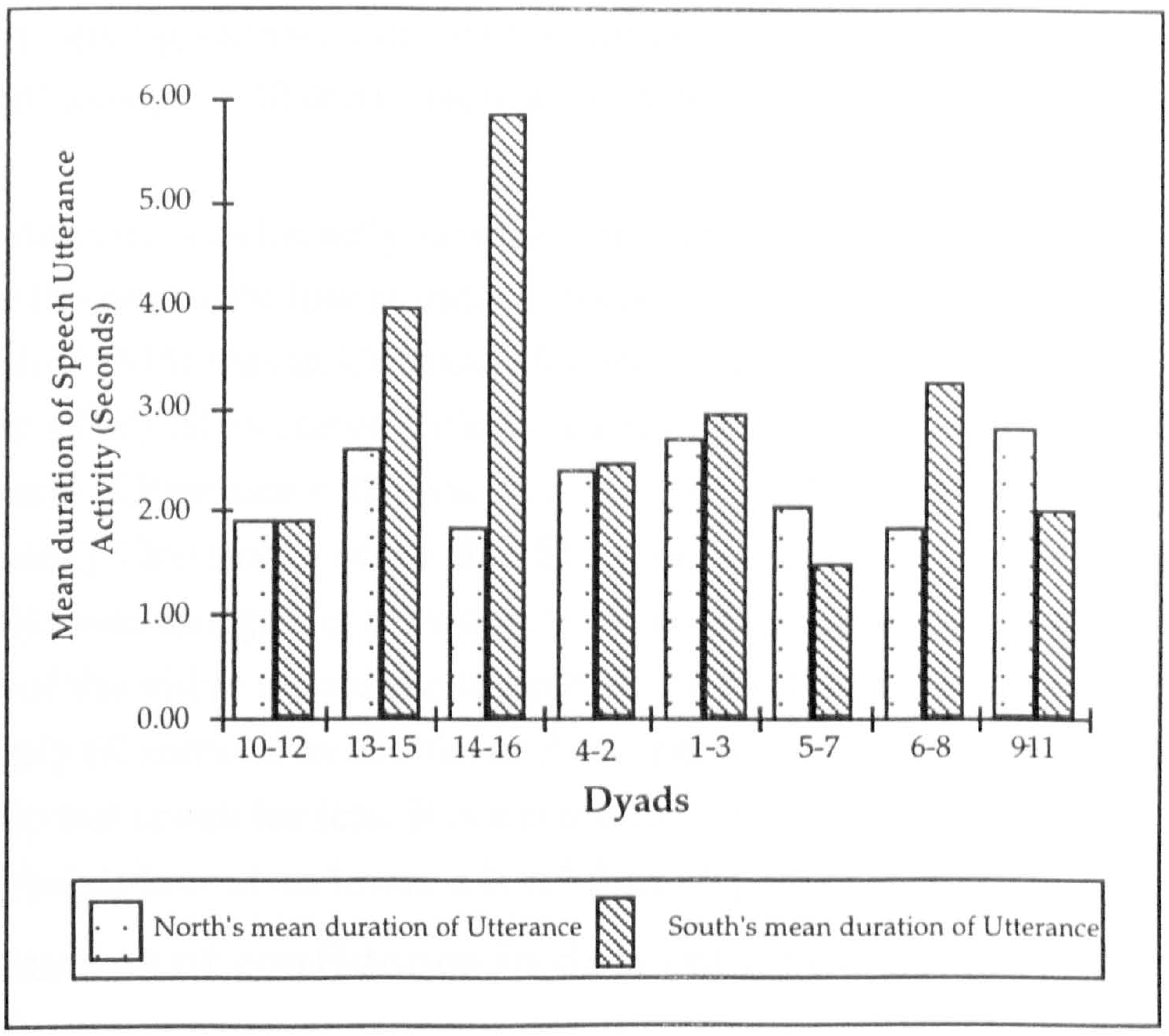


Figure 4.4: Duration of Speech Activities, by Participant (seconds)

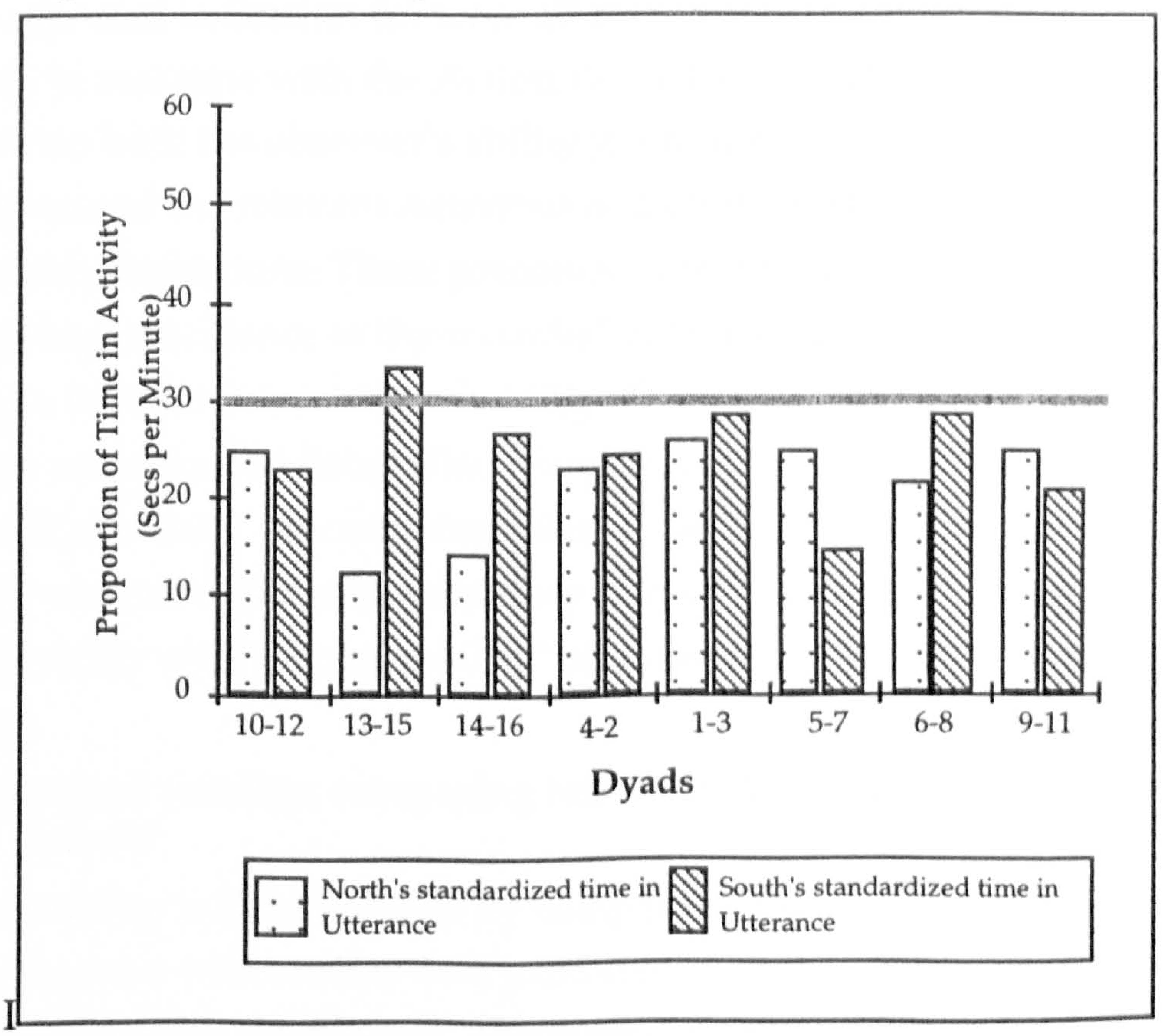


Figure 4.5: Standard time spent in Speech Activities, by Participant (seconds per minute)

Figures 4.4 and 4.5 show how individual Speech Activity data relate to dyad membership. They are histograms of the duration and standard time data respectively, giving scores from dyad members in adjacent positions on the "Participant" axis (i.e. S10 & S12 were a dyad, S13 & S15 were another dyad, and so on).

Utterance Activity was broadly similar across individuals, also statistically evinced by the relatively low standard deviations for these measures. Only one individual (S15) was in Utterance for more time overall than in Silence (35 seconds out of 60). S15's conversational partner (S13) was unusually Silent on both measures (Utterance ≈ 12 secs. min^{-1}; 3.5 seconds, vs. Silence ≈ 48 secs. min^{-1}; 11 secs.). One might argue that S15 was compensating for the reticence of S13, or S13 was struggling to compete for an opportunity to speak. Inspection of the video recording suggested the former to be more likely. If there are only 60 seconds available, and one person talks for most of them, the other can do but speak for less. It is a reminder of the importance of treating interactive behaviour at no lower a level than *all* parties involved.

4.6 Measures of confidence in data collected with Action Recorder

The data reported in Section 4.5 were all collected by a single observer rating videotapes in real time with the Action Recorder tool. Their value is thus dependent on both the observer's ability to use the tool in real time to accurately record the relevant Activities and on the observer's interpretation of the Activity definitions. These premises translate into two requirements for demonstrating confidence in the recorded data, as statistical validity and the rather more familiar inter-rater reliability. The idea of statistical validity is whether or not recorded data reflect the phenomena they are intended to represent. It primarily concerns the value of the tool for making observations on-the-fly, whereas inter-rater reliability is mainly to do with the general communicability and "recognizability" of the Activity Set definitions themselves.

4.6.1 Statistical validity: comparing real-time Action Recorder data with a veridical data set

Statistical validity is determinable by comparing the data produced by an uncertain method with similar data produced by a method in which confidence is already established. The video recorded experimental discussions were re-examined by the original rater using a different method for logging observations.

A benchmark data set was compiled, using a frame-by-frame analytic approach to the video data used in the test case. Equipment limitations precluded the use of this approach for the Speech Activity Sets[26] and so only Gaze Activity Sets were re-examined. As previously reported, Gaze was divided into three active levels and one default level of Activity. It was recorded according to the observers' perceptions of the participants' visual focus. The quality of video recordings under analysis was such that this information was readily detectable from both head and eye Activity.

A three-minute sample of video data was taken at random from each of the sixteen subjects, i.e. one videotape per subject. Using a frame advance facility on a video editing suite[27], an observer carefully progressed through each three-minute sample, noting each occurrence of a change of visual focus on behalf of the subject under observation according to the four defined Gaze Activities.

Each instance of a change from one Gaze Activity into another was recorded as a single letter code, with the time at which it occurred, in a data field in the Action Recorder tool. This frame-by-frame equivalent involved typing event codes and time stamps manually into a data field in the Action Recorder tool, rather than logging in real time. The time stamps were obtained directly from the video image itself, rather than from the VCR's built-in timer, to remove a potential source of inaccuracy. As reported previously, this time code was generated during the experimental sessions and mixed into the video signal at the recorder. Once again, each Activity Set was logged individually (one pass of the videotape per Activity Set). All sorting and preparation for statistical analysis was done using Action Recorder's data categorization and organization routines with SPSS, as detailed in Chapter Three. The statistical validity refers here to process of real-time data collection, as the most likely source of error, rather than to the statistical analyses performed on those data once recorded.

Three-minute segments were taken from the original, real-time ratings to match the frame-by-frame versions. Both were subject to the same SPSS treatments, generated by Action Recorder. Similarity of frame-by-frame and real-time data was then determined by correlation with Pearsons r.

---

[26]On the Panasonic VCRs used, the audio track only functions in 'play' mode, not in 'frame advance' mode.

[27]Panasonic VCR, type NV-F 76 HQ, controlled by a 'Panasonic Editing Controller', type VW-EC310.

4.6.2 Statistical validity - results

In each case, the mean and standard deviation of the three base Activity measures (mean duration in Activity, mean standard time and mean standard frequency) are tabulated for each Gaze state. Scatter plots of the scores for the sixteen participants are also presented to clarify the relationship between data collected by each technique.

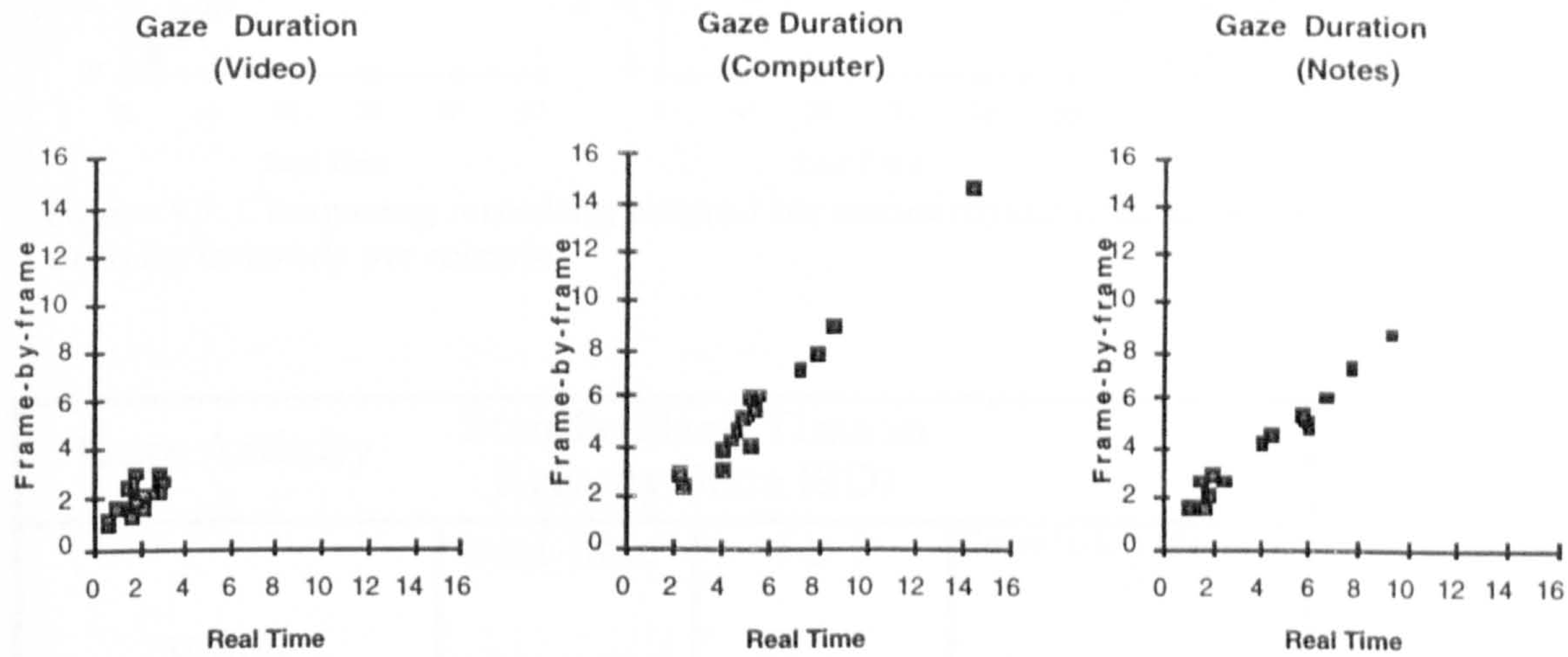


Figure 4.6: "Mean Duration of State" Data on Gaze deployment obtained from real-time Action Recorder and frame-by-frame ratings

Gaze Activity	Duration in Seconds (SD)		
	real-time	f-b-f	Correlation (Pearson' r)
Video	1.92 (0.75)	1.96 (0.63)	0.684
Computer	5.77 (2.88)	5.68 (2.99)	0.985
Notes	4.02 (2.60)	3.89 (2.18)	0.987

Table 4.6: Duration of Gaze Activities, correlating real-time and frame-by-frame (f-b-f) recording

Gaze Activity durations deriving from f-b-f and real-time recording methods were almost identical for GazeNotes and GazeComputer. Agreement for GazeVideo was also very good, although the Pearson r value was somewhat lower than for the other Gaze Activities. Inspection of Figure 4.6 shows that this may have been a consequence of the uniformly small values associated with this measure for the GazeVideo Activity.

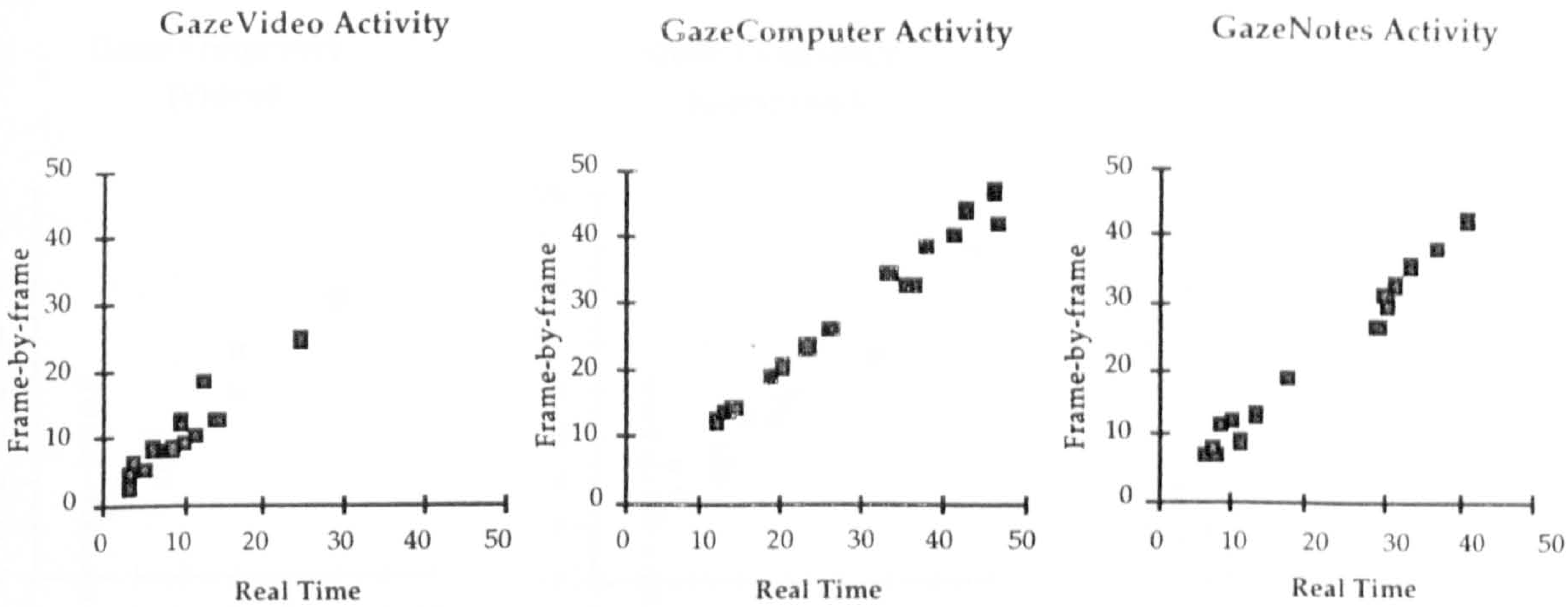


Figure 4.7: Comparing recording method for standardized time measures of Gaze Activity (seconds per minute)

Gaze Activity	Standardized Time in Activity State (SD)		
	real-time	f-b-f	Correlation (Pearson' r)
Video	9.22 (5.18)	10.00 (5.53)	**0.926**
Computer	29.82 (11.85)	29.00 (11.85)	**0.989**
Notes	20.17 (12.91)	19.75 (12.64)	**0.992**

Table 4.7: Time Proportion (Seconds per Minute) in three Gaze Activities, obtained from two data gathering techniques, real-time and frame-by-frame (f-b-f) recording

Extremely good agreement was found for the standard time in Activity state measure.

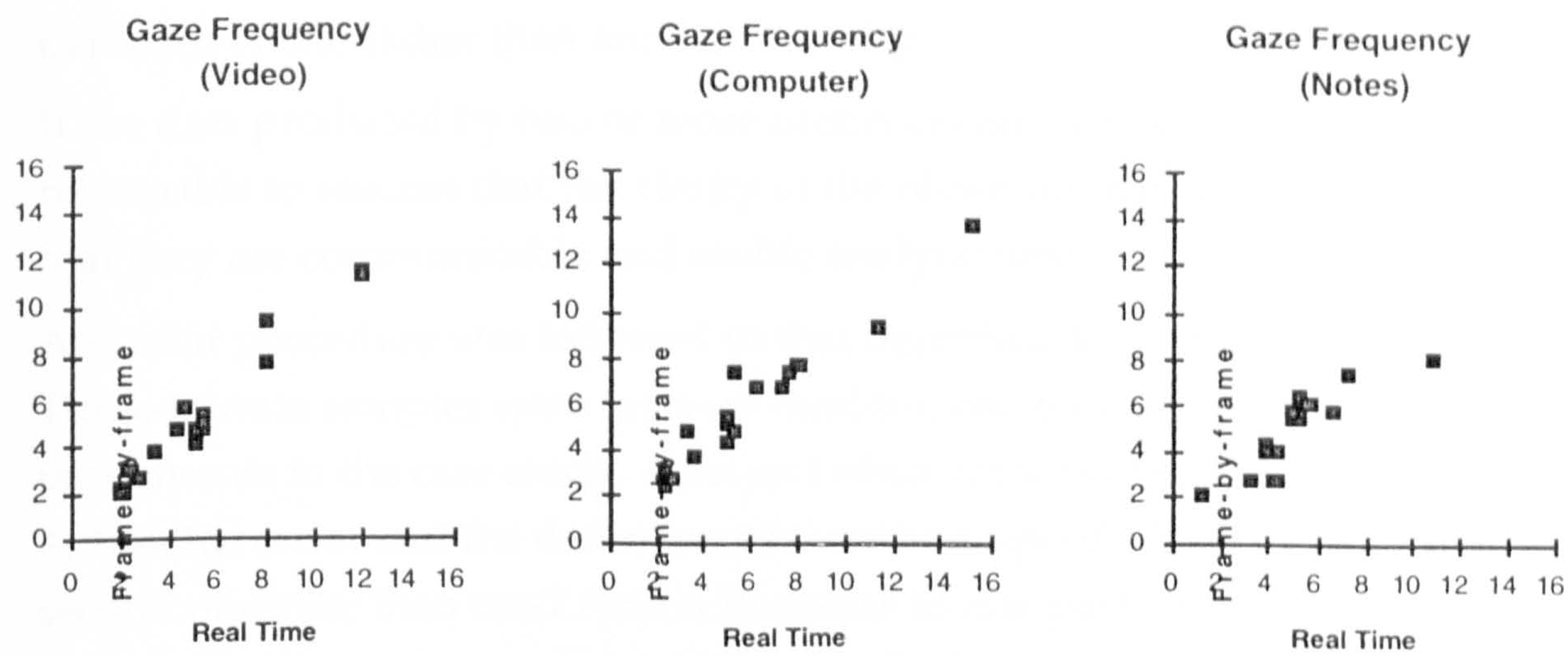


Figure 4.8: Frequency of Gaze Activity occurrence, comparing 'Real Time' Action Recorder and 'Frame-by-frame' ratings

Gaze Activity	Frequency -N per Minute (SD)		
	real-time	f-b-f	Correlation (Pearson' r)
Video	5.06 (2.52)	5.13 (2.44)	0.969
Computer	6.00 (3.45)	5.90 (3.45)	0.968
Notes	5.12 (2.08)	4.94 (1.67)	0.879

Table 4.8: Frequency of three Gaze Activities, , obtained from two data gathering techniques, real-time and frame-by-frame (f-b-f) recording

Very good agreement was found between the two recording methods for this frequency measure. Since the frequency and standard time measures are in a mathematical relationship with the mean duration of Activity measure, the rather lower R for GazeVideo mean duration is probably more of a reflection of the low and uniform scores. Taken together, these comparisons represent an acceptable demonstration of the statistical validity of real-time recording for the defined Gaze Activities with Action Recorder.

4.6.3 Inter-rater reliability

Inter-rater reliability looks at the correspondence of data produced by different observers working to a common coding strategy. In this case, it involves a combination of the ability of other observers to use the Action Recorder tool in real time, how well the defined Activity Sets can be explained, how discriminable each Activity is to its within-set competitors and whether or not different observers can agree on some behaviour as evidence of one rather than another Activity.

If the data produced by two or more observers are very similar, it is reasonable to assume that the clarity of the observational categories is good, in that they are communicable and usable analytic devices.

A similar procedure was followed to that described for statistical validity. Three-minute samples were taken at random, one for each of the sixteen participants in the case study. A second observer was trained in the use of Action Recorder and the definition of Gaze and Speech Activity Sets. The second observer then used Action Recorder to rate participant Gaze and Speech Activity, one pass per Activity Set. Ratings were compared with appropriate extracts from the original observer's ratings by generating summary data with Action Recorder's automatically produced 'SPSS syntax' file.

Once again, comparisons of the three base Activity measures are reported in turn and similarity assessed with Pearson's r correlation coefficient.

As Table 4.9 and Figure 4.9 demonstrate, observers ratings of Gaze Activity show very close agreement when expressed as mean duration of Activity. Observers' ratings of Speech Activity also show good agreement, expressed as mean duration of SpeechUtterance state, and very good agreement for SpeechSilence (see Table 4.10 and Figure 4.10). Extremely good agreement was obtained by extracting the standardized time in Activity state measure from observers' ratings (see Table 4.11 and Figure 4.11).

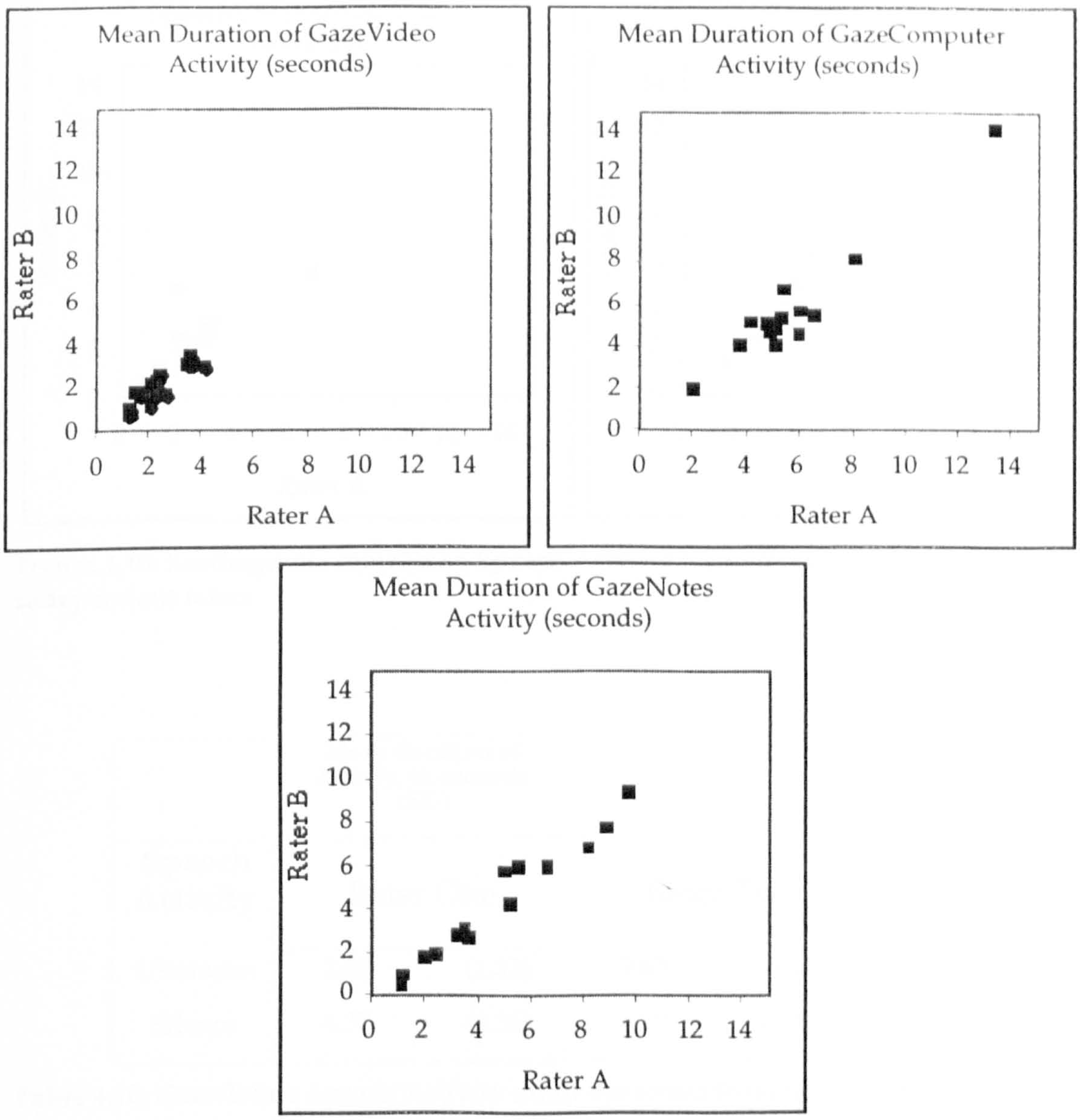


Figure 4.9: Scattergrams comparing Gaze Activity mean durations from independent raters

Mean duration of Activity, in seconds (SD)					
Gaze Activity	Rater One		Rater Two		**Correlation (Pearson' r)**
Video	2.40	(0.88)	2.06	(0.81)	0.864
Audio	5.69	(2.43)	5.54	(2.64)	0.957
Notes	4.46	(2.75)	3.93	(2.65)	0.980

Table 4.9: Correlating Gaze Activity mean durations from independent raters

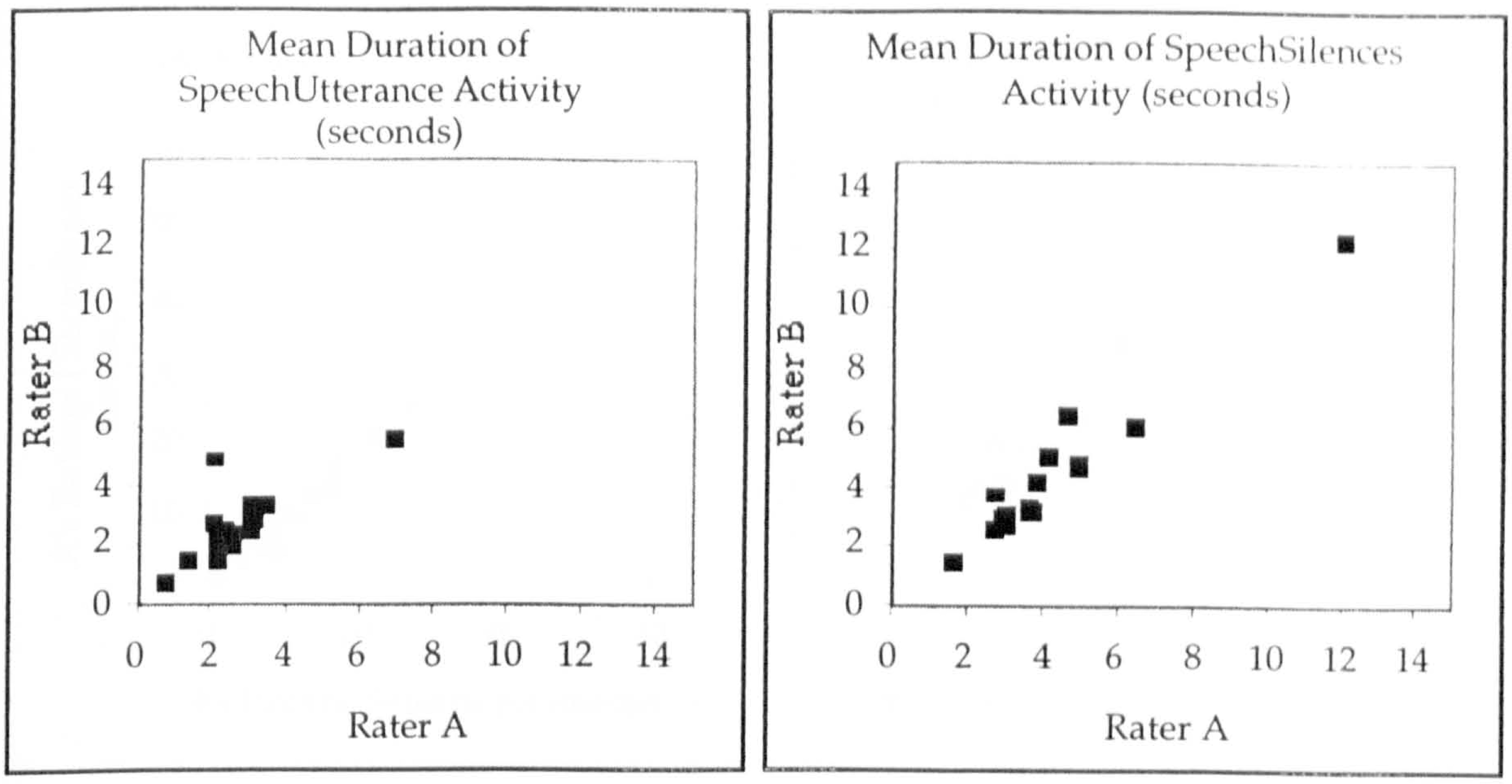


Figure 4.10: Scattergrams comparing Speech Activity mean durations from independent raters

Speech Activity	Mean duration of Activty, in seconds (SD)				
	Rater One		Rater Two		Correlation (Pearson' r)
Utterance	2.69	(1.32)	2.63	(1.22)	0.764
Silence	4.37	(2.36)	4.50	(2.50)	0.970

Table 4.10: Correlating Speech Activity mean durations from independent raters

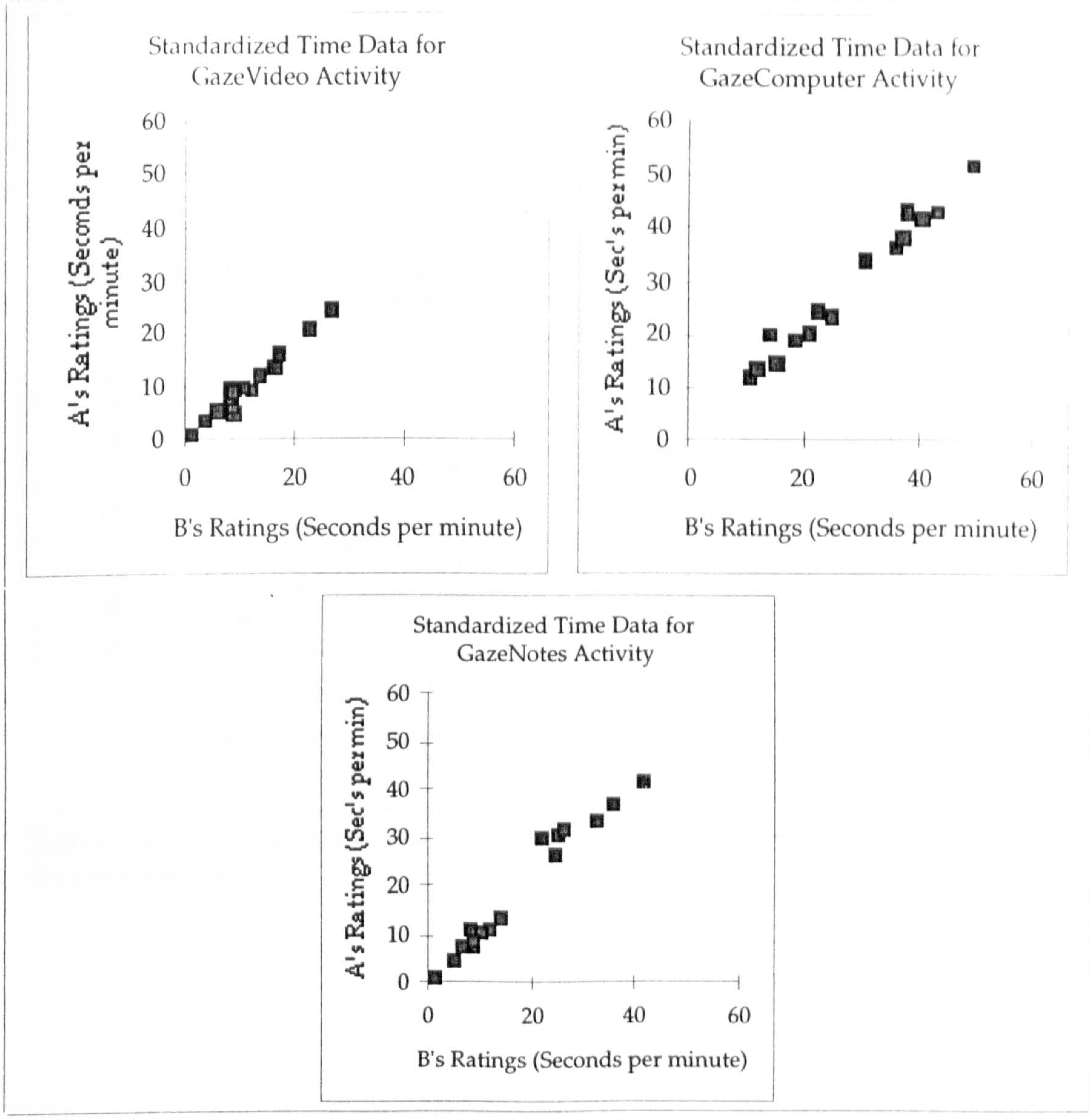


Figure 4.11: Scattergrams comparing Gaze Activity standardized time data from independent raters

Standard Time, in seconds per minute (SD)					
Gaze Activity	Rater One		Rater Two		Correlation (Pearson' r)
Video	11.58	(6.65)	10.27	(6.27)	0.977
Audio	28.08	(12.35)	29.83	(12.77)	0.986
Notes	17.89	(12.07)	19.20	(13.27)	0.981

Table 4.11: Correlating Gaze Activity standardized time data from independent raters

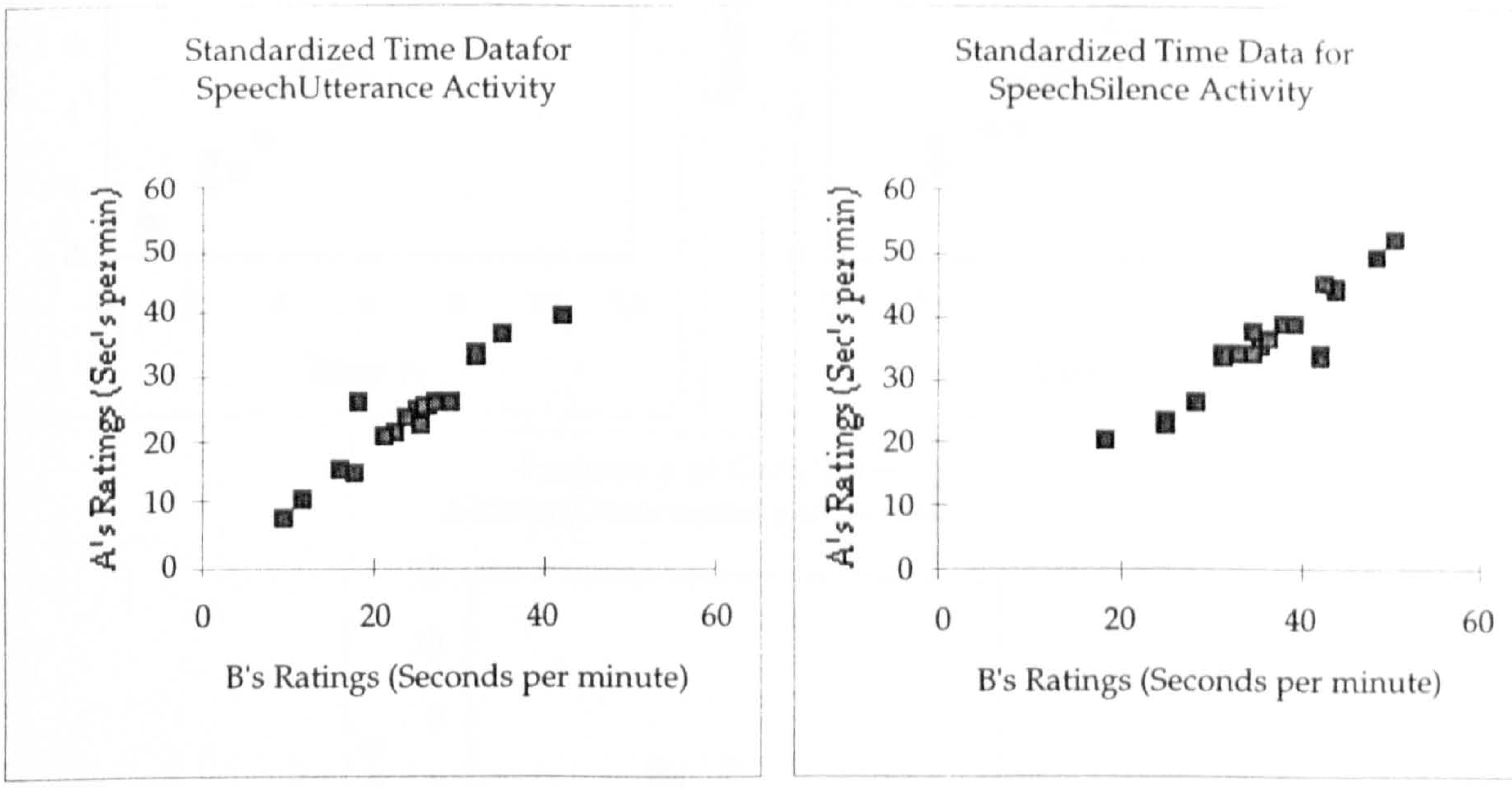


Figure 4.12: Scattergrams comparing Speech Activity standardized time data from independent raters

Speech Activity	Standard Time, in seconds per minute (SD) Rater One		Rater Two		Correlation (Pearson' r)
Utterance	23.23	(8.33)	22.72	(8.73)	0.953
Silence	36.81	(8.32)	37.29	(8.73)	0.952

Table 4.12: Correlating Speech Activity standardized time data from independent raters

Observers ratings of Speech Activity standardized time data were also very similar.

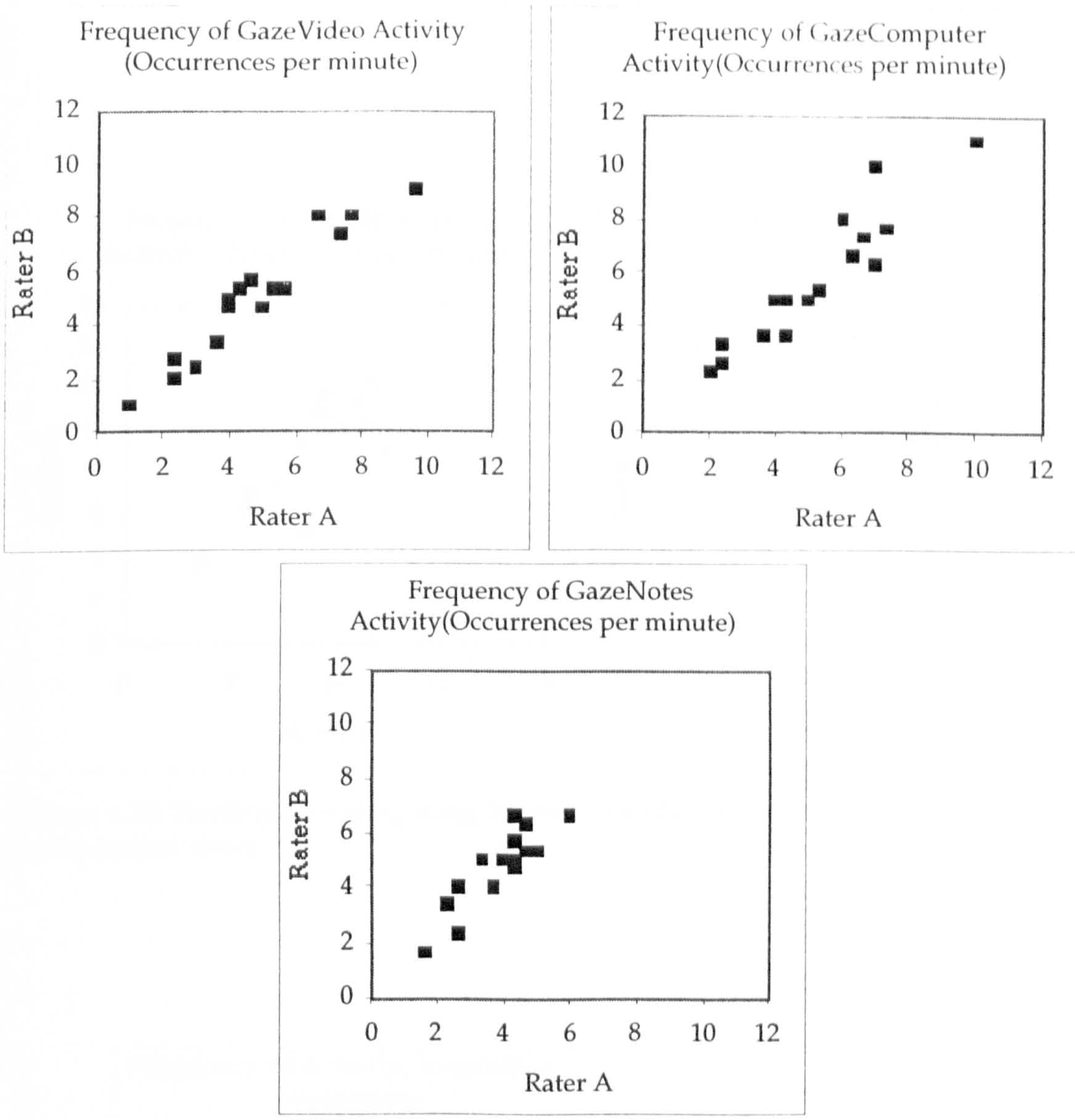


Figure 4.13: Scattergrams comparing Gaze Activity frequency data from independent raters

Frequency of Activity, in number per minute (SD)					
Gaze Activity	Rater One		Rater Two		Correlation (Pearson' r)
Video	4.79	(2.25)	4.98	(2.32)	0.961
Audio	5.23	(2.16)	5.81	(2.52)	0.934
Notes	3.92	(1.13)	4.77	(1.42)	0.879

Table 4.13: Correlating Gaze Activity frequency data from independent raters

Frequency data, as number of occurrences per minute, from the two observers' ratings showed excellent agreement.

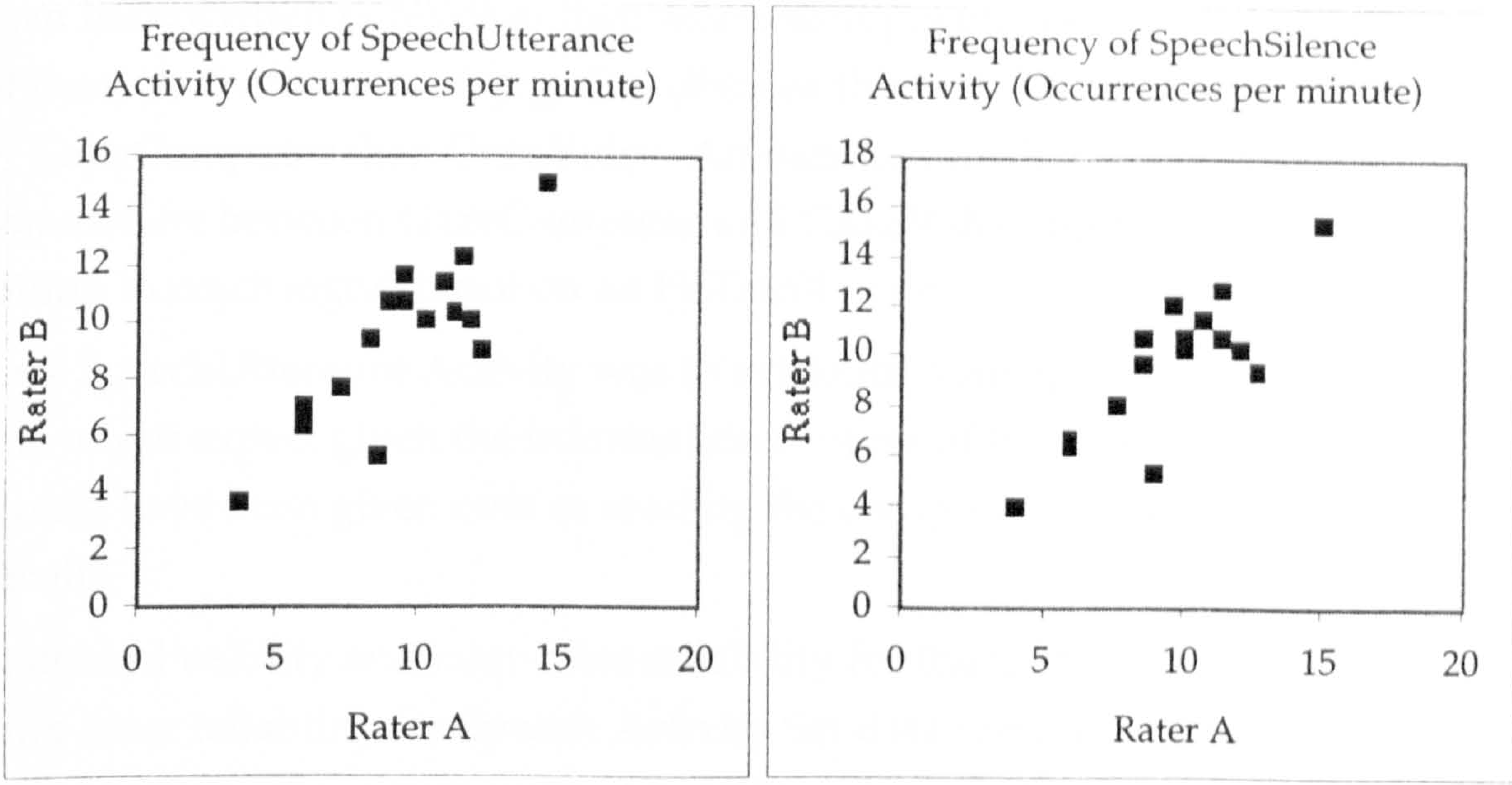


Figure 4.14: Scattergrams comparing Speech Activity frequency data from independent raters

Frequency of Activity, in number per minute (SD)					
Speech Activity	Rater One		Rater Two		Correlation (Pearson' r)
Utterance	9.42	(2.80)	9.36	(2.84)	0.841
Silence	9.51	(2.77)	9.53	(2.91)	0.828

Table 4.14: Correlating Speech Activity frequency data from independent raters

Finally, inter-rater agreement for the number of occurrences of Speech Activities per minute was also high.

4.7 Summary of results for Simple Activity Set Analysis

When participants in the experimental discussion task looked towards their computer monitors, they did so for considerably longer than at either their own handwritten notes or at their discussion partner on the video link. Each of these Activities was engaged as often as the next. More time was invested in GazeComputer than GazeVideo. Amount of time in GazeNotes was somewhere between GazeComputer and GazeVideo; both contrasts just failing to reach significance on an HSD followup.

Less SpeechUtterance Activity was in evidence than SpeechSilence, much as one might expect given the information content of the task, i.e. some time would have been given over to reading the computer-based application details.

Statistical validity and inter-rater reliability for the Gaze Activity Set data and inter-rater reliability for Speech Activity Set data were all shown to be in good agreement. As these 'measures of confidence' were all obtained from randomly selected three-minute segments of the experimental discussions, the close agreement is particularly valuable. Since each measure is a mean calculated over all instances in an interval, greater similarity follows from longer intervals. Mean duration of time in Activity state data was somewhat less reliable for short Activities, including SpeechUtterance and GazeVideo, although still perfectly adequate to support the inferences drawn from these data.

The implication of these comparisons is both that Action Recorder can be used effectively in real time and that the Activity Sets defined for this case study are operationally effective as Simple analytic constructs.

4.8 Contingent Activity Set Analysis: data from multiples of Activity Sets as Co-activity states

The data reported in this section comprise various measures of Activity concurrency by examining a pair of Activity Sets. The term Contingent Analysis is adopted since it seeks to establish where engagement in some Activity is contingent on other Activity states then in progress. The Activity Sets described in this chapter are all attributable to each individual in the dyad thus there are four possible Activity Set pairings: the only possible *within-participants* consideration (i.e. A's Speech by A's Gaze), and three *between-participants* analyses (i.e. A's Gaze by B's Gaze, A's Gaze by B's Speech, and A's Speech by B's Speech). Consideration of the within-participants Co-activity data begins by looking at duration and standardized frequency data and then at the standardized time data.

It should be understood that Co-activities are derived from the Activity ratings summarized in the previous section. Each of the summaries described as "Simple Statistics" directly reflects Activities rated as complete behavioural entities by observers. For example, a GazeNotes rating would commence with observation of a glance towards paper material on a table and persist until the observer notes a glance somewhere else. In contrast, the data reported here as Contingent Statistics are fractions of the specified Activities computed to be in temporal conjunction. For example, a GazeNotes Activity state might have been in progress when an observer noted two successive changes in Speech Activity state, all while there has been no change in the on-going Gaze Activity. Three Co-activity states would be derived from the state table compiled by Action Recorder for an analysis of these two Activity Sets: GazeNotes;SpeechSilence -> GazeNotes;SpeechUtterance -> GazeNotes;SpeechSilence. The purpose of the Contingent Statistics reported here is to determine whether or not there is evidence of contingency in the expression of Activities between Activity Sets; in other words, to estimate the extent to which engaging in one Activity might depend on other current Activity states, individual or environmental.

4.8.1 Within-participants, Gaze and Speech Activity Sets

Consideration of Co-activity contingency commences with an examination of Speech and Gaze Activity Set data at the level of the individual. These provide the only "within-participants" Activity Set Contingent Statistics available from the case study data. For example, questions may be asked of the form: is it true that an individual tends to be in the Speech Activity 'Silent' when they are in the Gaze Activity 'Computer'? Or is it that there is no consistent relationship between any of the Gaze Activities and Speech state? Since there are two Speech Activities and three Gaze Activities (excluding the default), there are six possible "Co-activities" to be considered for evidence of a relationship of this kind.

Co-activity states, described by Activity conjunctions	Mean duration of State in Seconds (SD)		Standardized time in State, in Secs./min (SD)		Frequency in number/min (SD)	
Gaze Video, while in SpeechUtterance	0.97	(0.27)	4.49	(2.98)	4.32	(2.19)
Gaze Computer,while in SpeechUtterance	1.56	(0.53)	10.53	(6.17)	6.44	(2.62)
Gaze Notes, while in SpeechUtterance	1.34	(0.37)	7.54	(4.19)	5.34	(2.32)
Gaze Elsewhere with SpeechUtterance	0.74	(0.39)	0.53	(0.40)	0.71	(0.52)
Gaze Video, with SpeechSilence	1.41	(0.46)	6.53	(3.38)	4.59	(1.93)
Gaze Computer, with SpeechSilence	2.67	(0.85)	18.14	(6.36)	7.08	(2.66)
Gaze Notes, with SpeechSilence	1.82	(0.70)	11.63	(7.44)	5.94	(1.96)
Gaze Elsewhere with SpeechSilence	0.76	(0.47)	0.61	(0.65)	0.79	(0.58)

Table 4.15:Within Participants, Gaze and Speech Co-activity data.

4.8.2 Co-activity duration

The data reported here are derived from the raw ratings of separate Activities from videotapes. In consequence, Co-activity duration data reflects the mean period for which separate Activities coincided. Analysis of Co-activity durations serves to determine whether these coincidences are random or somehow organized with respect to one another.

Table 4.15 and Figure 4.15 suggests that there is a within-participant constraint governing the expression of some Speech and Gaze Activities.

Durations of each of the Gaze Activities for the 'Utterance' state of Speech are all shorter than 'Silence', to varying degrees.

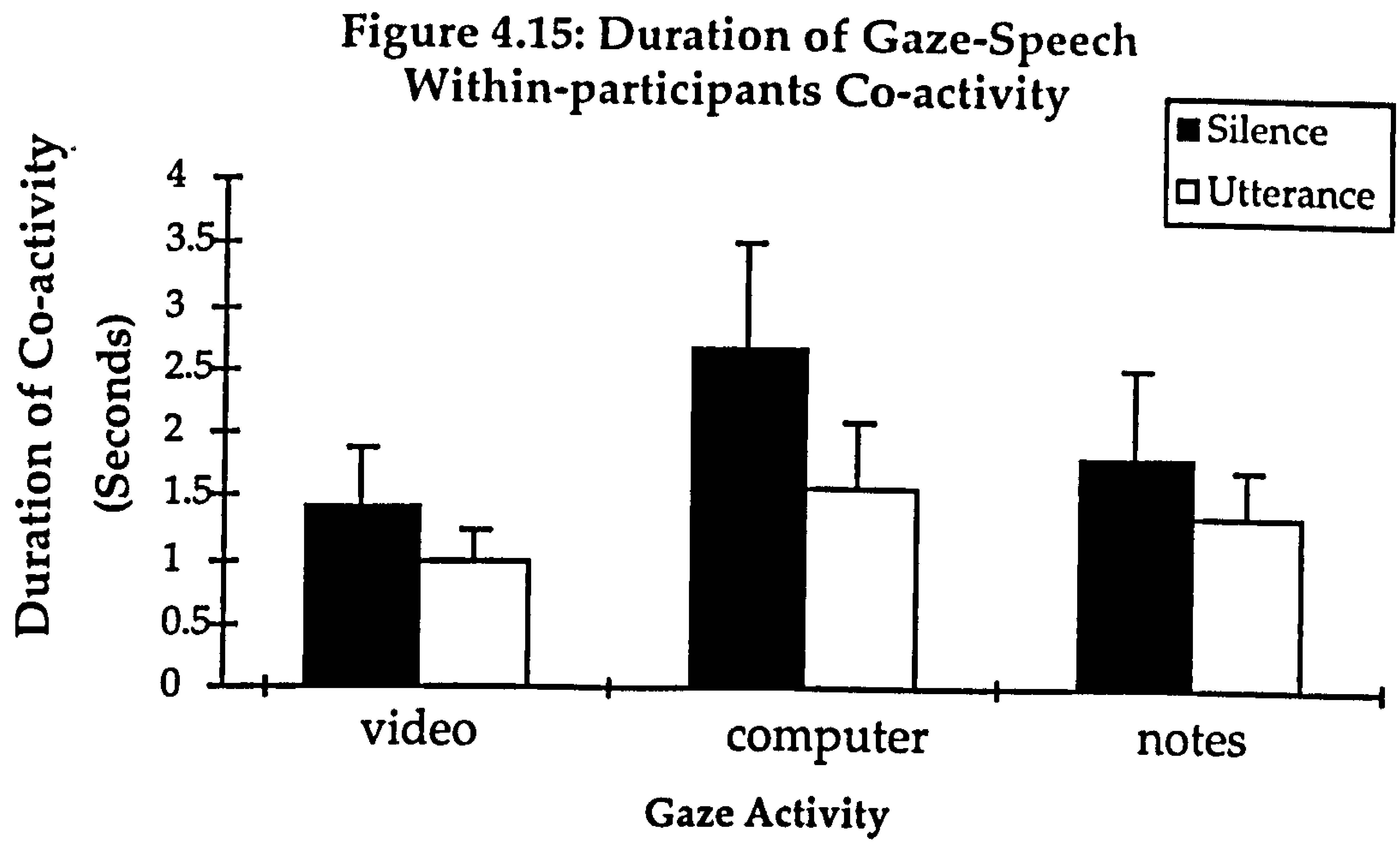


Figure 4.15 suggests that the largest difference in Co-activity duration contrasts GazeComputer by Speech Activity state, such that GazeComputer glances were approximately half as long while participants were in Utterance as when they were in Silence. GazeNotes and GazeVideo were also shorter in Utterance than Silence but not to the same degree as GazeComputer. Table 4.16 presents results of an ANOVA, confirming the presence of a highly significant interaction. Following up with a Simple Main Effects procedure (Kirk, 1968), all three Gaze Activities were found to differ significantly by Speech Activity state (follow-up results are given by the italicized entries).

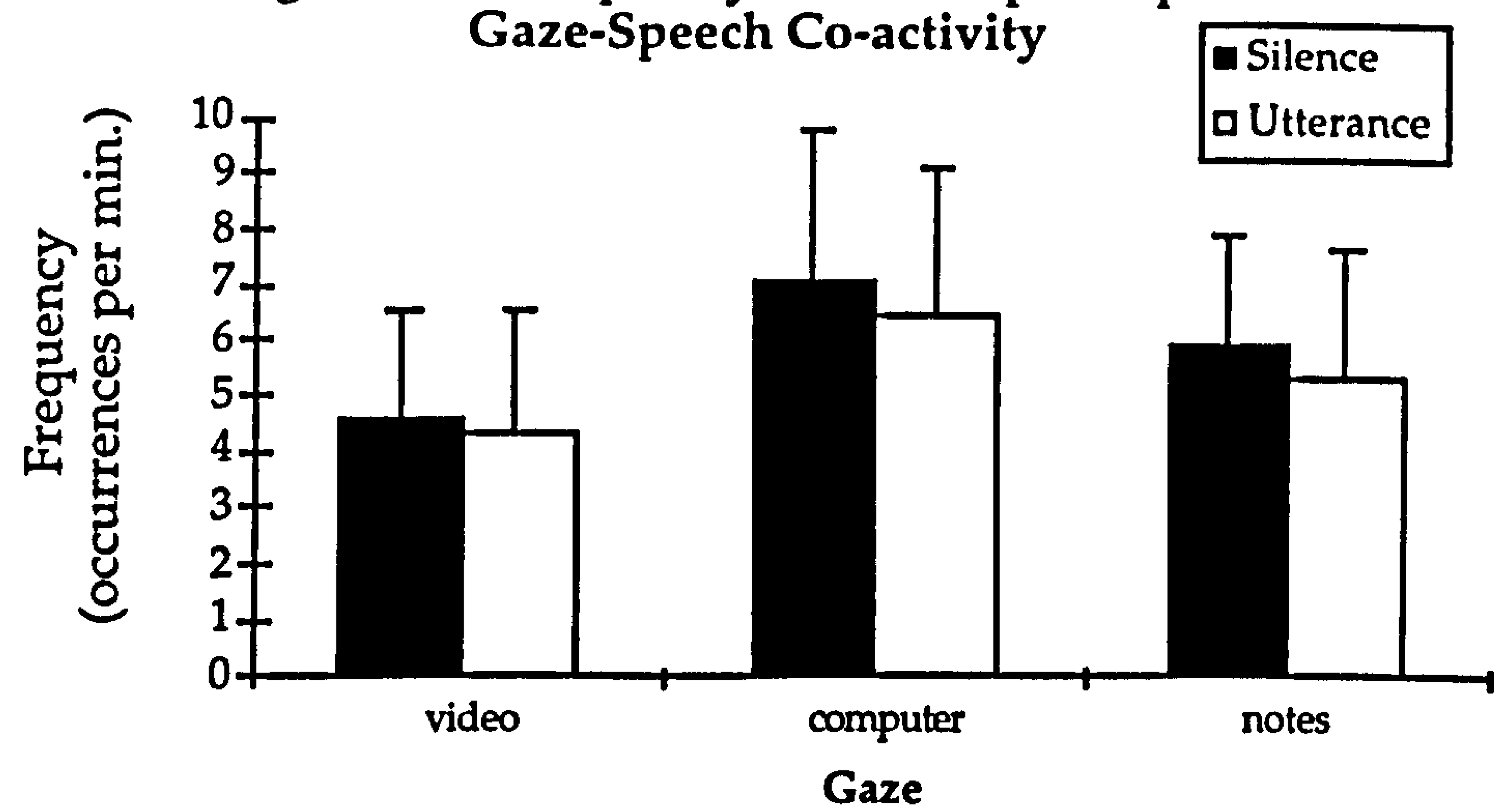
Source of Variation	SS	DF	MS	F	p
SPEECH	10.86	1	10.86	37.03	<0.001
GAZE	13.80	2	6.90	17.29	<0.001
SpV	*1.51*	*1*	*1.51*	*8.41*	*< 0.01*
SpC	*9.77*	*1*	*9.77*	*54.27*	*< 0.01*
SpN	*1.83*	*1*	*1.83*	*10.15*	*< 0.01*
SPEECH BY GAZE	2.25	2	1.12	6.08	0.006
Speech Error	4.40	15	0.29		
Gaze Error	11.98	30	0.40		
Interaction Error	5.55	30	0.18		

Table 4.16: F-table for ANOVA of mean duration of Activity, within-participants Gaze and Speech Activity Sets (simple main effects follow-up in italics)

4.8.3 Co-activity frequency - occurrence per minute

A consideration of Co-activity duration asks whether or not the Activity states randomly coincided by looking at differences between mean per-instance periods of Activity combinations. A consideration of Co-activity frequency instead examines the rate at which Activity combinations are expressed.

Figure 4.16: Frequency of within-participant Gaze-Speech Co-activity



Here, a within-participants analysis of Gaze and Speech Co-activity frequency indicated no differences[28] for Gaze Activity against category of Speech Activity (Table 4.17 & Figure 4.16). In other words, the rate of Co-activity occurrence was not a contingency of their constituent Activity states. Participants most frequently engaged in Gaze Computer and least often in Gaze Video, regardless of their simultaneous Speech Activity state.

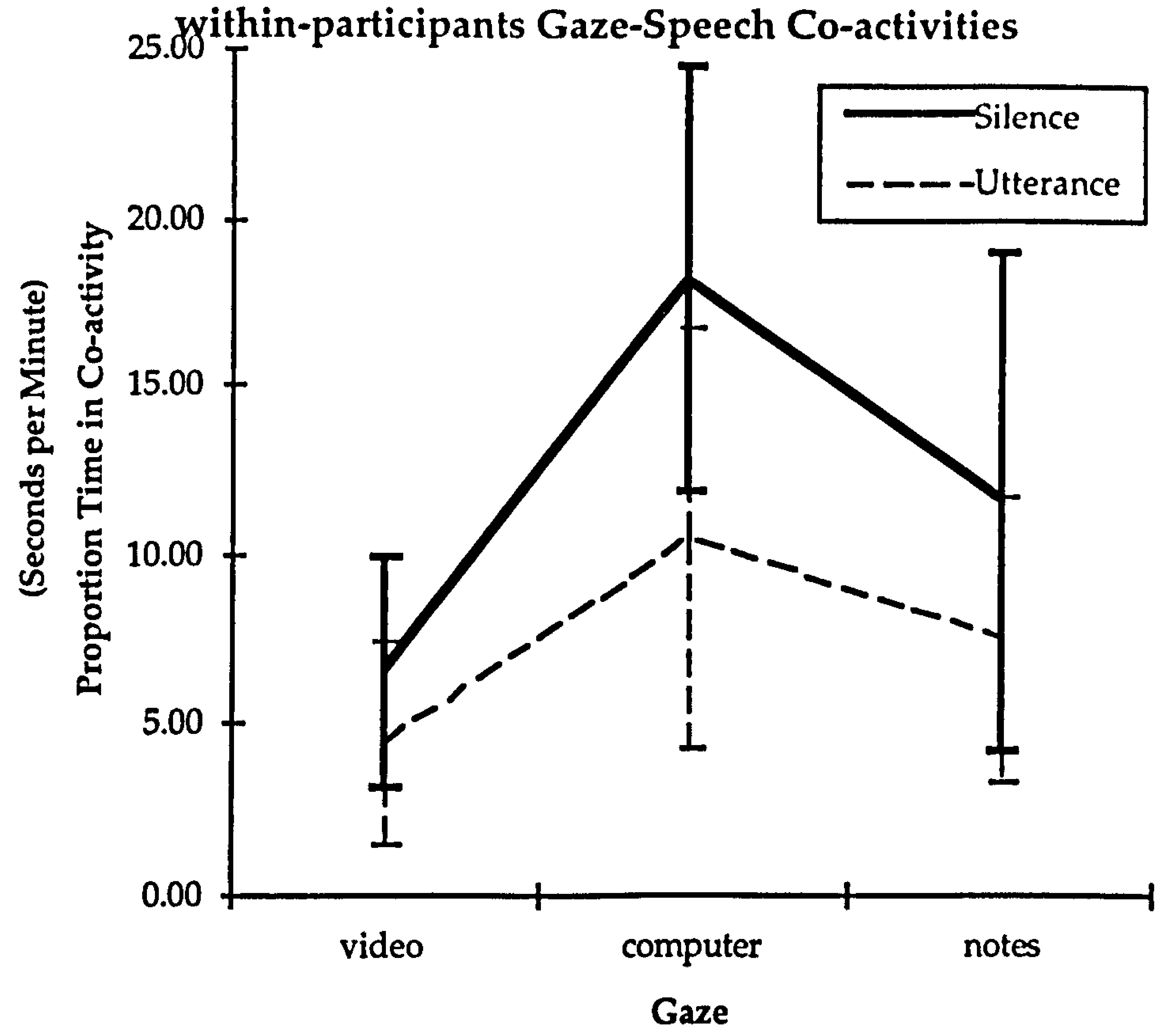
Source of Variation	SS	DF	MS	F	p
SPEECH	6.15	1	6.15	3.11	0.098
GAZE	85.33	2	42.66	3.54	0.042
SPEECH BY GAZE	0.64	2	0.32	0.94	0.401
Speech Error	29.65	15	1.98		
Gaze Error	361.1	30	12.04		
Interaction Error	10.14	30	0.34		

Table 4.17: F-table for within-participants Gaze and Speech Activity Sets standardized frequency data

[28]On the frequency measure, no main effect is found for the Speech Activity Set (as expected from the foregoing analysis and comments - see Table 4.5). However, Gaze Activity Set just reaches significance at the 0.05 level. This does not agree with the result of the Simple Gaze Activity analysis of frequency data (see Table 4.3). There is no contradiction here as Co-activity data are not given by a linear combination of Simple Activity terms.

4.8.4 Standard time in Co-Activity State - seconds per minute

Figure 4.17: Standardized time in within-participants Gaze-Speech Co-activities



Source of Variation	SS	DF	MS	F	p
SPEECH	503.02	1	503.02	25.00	<0.01
SpV	*33.42*	*1*	*33.42*	*3.09*	*n.s.*
SpC	*463.05*	*1*	*463.05*	*42.80*	*<0.01*
SpN	*133.37*	*1*	*133.37*	*12.33*	*<0.01*
GAZE	1248.45	2	624.22	9.62	<0.01
SPEECH BY GAZE	126.80	2	63.40	5.86	<0.01
Speech Error	301.58	15	20.11		
Gaze Error	1946.90	30	64.90		
Interaction Error	324.61	30	10.82		

Table 4.18: F-table for within-participants standardized time data Gaze and Speech Co-activities (simple main-effects follow up italicized)

Rather than looking at the mean duration of Co-activity occurrence, the standard time in Co-activity state considers the relationship between constituent Activities of the state over the whole period of analysis. It reflects the total investment in particular Activities, as a function of one another.

Speech and Gaze Co-activity standardized time data are shown in Figure 4.17. Whilst in SpeechUtterance, participants tended to gaze towards the video monitor for less time than when in SpeechSilence (in a ratio of 1:1.4). This relationship is of the same order for GazeNotes in Utterance and Silence (1:1.4) and for GazeComputer in Utterance and Silence (1:1.7). An analysis of variance (see Table 4.18) produced a statistical interaction between Gaze and Speech Activity Sets. A 'simple main-effects' test was performed, looking at the three Gaze Activities (as levels of the variable 'Gaze'), within Speech (italicized in Table 4.18). The SpV row shows the contribution to the variance made by the 'Video' level of Gaze with the Speech Activity Set. Similarly, SpC is the 'Computer' and SpN, the 'Notes' levels of Gaze with Speech Activity. Highly significant interactions are given between Speech Activity and both GazeComputer and GazeNotes. However, with an F ratio of 3.09 (1, 30), GazeVideo interaction with level of Speech Activity narrowly failed to reach significance (at the $p<0.05$ level, $F=4.17$ (1,30), and at the $p<0.10$ level, $F=2.88$ (1,30).

Given that no differences were evident in the frequency-based analyses, and that there is a simple mathematical relationship between the duration, standardized time and frequency measures, it is somewhat surprising that GazeVideo by Speech Activity interaction in duration measures are not also evident in the standardized time measures.

4.9 Co-activity data confidence

Statistical validity and inter-rater reliability must also be demonstrated for Co-activity data. Although they are also derivative of the observers' ratings, they cannot be arrived at by a linear combination of terms from the Simple Activity data, as each Co-activity state is generated from the raw Action Recorder log. In this section, validity and reliability correlational data are presented for all possible Co-activities, anticipating treatment of between-participants Co-activities to be reported in Chapter Six. Validity data is given first, followed by inter-rater reliability data.

4.9.1 Validity of Gaze-Gaze Co-activities

Tables 4.19 - 4.21 give Pearson's *r* correlation coefficients for mean duration, standardized frequency and time measures for GazeGaze Co-activities. All show very close levels of agreement.

Duration (seconds)					
Co-activity States	Real Time		fbf		Pearson's *r*
GazeVideo GazeVideo	1.25	(0.91)	1.09	(0.49)	0.81
GazeVideo GazeComputer	2.31	(0.79)	2.47	(0.72)	0.67
GazeVideo GazeNotes	1.94	(0.96)	2.25	(1.27)	0.83
GazeComputer GazeComputer	3.03	(1.17)	3.12	(1.22)	0.89
GazeComputer GazeNotes	3.47	(1.49)	3.48	(0.86)	0.89
GazeNotesGaze Notes	2.36	(1.92)	2.28	(1.62)	0.96

Table 4.19: Correlations between frame-by-frame (fbf) and real-time recording methods for Gaze Gaze Co-activity duration data

Standardized Frequency (occurrences per minute)					
Co-activity States	Real Time		fbf		Pearson's *r*
GazeVideo GazeVideo	1.83	(1.38)	2.12	(1.53)	0.81
GazeVideo GazeComputer	6.96	(3.89)	6.71	(3.75)	0.67
GazeVideo GazeNotes	4.04	(2.25)	4.21	(2.09)	0.83
GazeComputer GazeComputer	6.62	(4.72)	5.54	(2.88)	0.89
GazeComputer GazeNotes	8.13	(3.37)	7.67	(2.95)	0.89
GazeNotesGaze Notes	3.92	(1.87)	3.75	(1.85)	0.96

Table 4.20: Correlations between frame-by-frame (fbf) and real-time recording methods for Gaze Gaze Co-activity standardized frequency data

Standardized Time (seconds per minute) Co-activity States	Real Time		fbf		Pearson's r
GazeVideo GazeVideo	2.31	(2.23)	2.52	(2.04)	0.87
GazeVideo GazeComputer	8.28	(5.42)	8.85	(5.54)	0.95
GazeVideo GazeNotes	5.12	(3.68)	5.89	(4.64)	0.87
GazeComputer GazeComputer	18.23	(9.67)	17.10	(9.15)	0.98
GazeComputer GazeNotes	13.96	(5.50)	13.63	(5.66)	0.98
GazeNotesGaze Notes	10.53	(10.30)	9.95	(10.18)	0.99

Table 4.21: Correlations between frame-by-frame (fbf) and real-time recording methods for Gaze Gaze Co-activity standardized time data

As equipment limitations precluded validation of Speech Activity data in this way, it is particularly important to demonstrate good inter-rater reliability for Speech-related Co-activities. Scattergrams are presented in Appendix Three with Co-activity scores derived from each rater's Action Recorder observations. These permit a closer inspection of the relationships between Gaze Co-activity these data.

4.9.2 Inter-rater reliability of Co-activity measures

Within-subjects: Gaze-Speech Co-activities

Co-activity		Mean Co-activity duration	Standardized Co-activity frequency	Standardized Co-activity time
GazeVideo	SpeechUtterance	**0.26**	0.87	0.96
GazeVideo	SpeechSilence	0.81	0.86	0.95
GazeComputer	SpeechUtterance	0.83	0.97	0.97
GazeComputer	SpeechSilence	0.84	0.91	0.96
GazeNotes	SpeechUtterance	**0.46**	0.95	0.87
GazeNotes	SpeechSilence	0.88	0.85	0.99

Table 4.22: Pearson's *r* correlation coefficients for measures of within-participants Speech Gaze Co-activity

Inter-rater agreement for within-participants Gaze-Speech Co-activity measures was very good, with the exception of mean duration data for GazeVideoSpeechUtterance and GazeNotesSpeechUtterance (emboldened in Table 4.22). In both cases, the disparity in ratings could be isolated to data from a single dyad. In Figure 4.18, these Co-activity duration data are represented as a crosses amongst the data from other dyads represented as boxes. With the errant data points removed from the comparison, Pearson's *r* values rose to 0.85 in both cases. As Figure 4.18 shows, the general magnitude of these data was near floor, and so small absolute differences impinge on Pearson estimates of agreement to a greater degree.

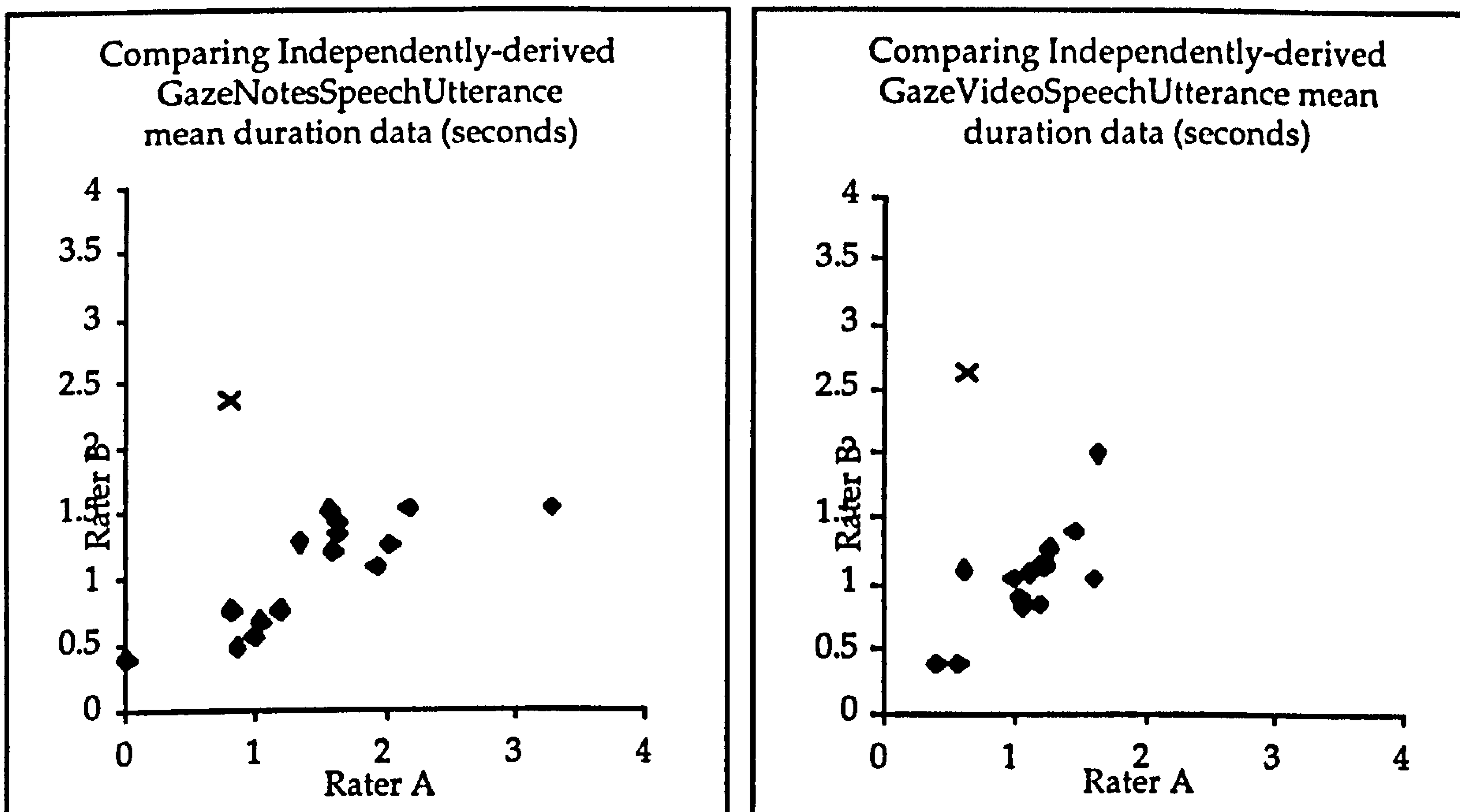


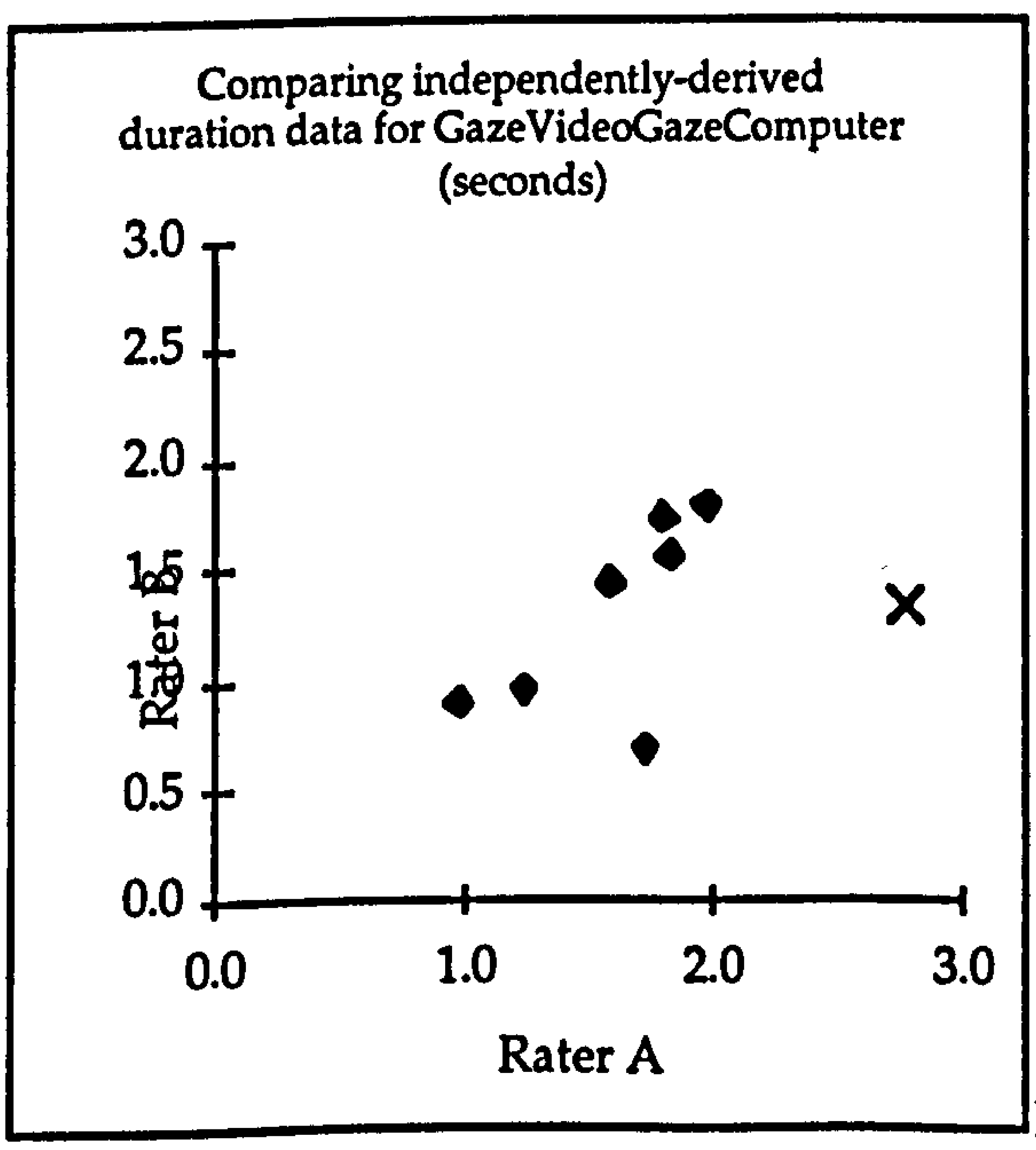
Figure 4.18: Comparing duration data from two observers, for two within-participants Co-activities

Between Subjects: Gaze-Gaze Co-activities

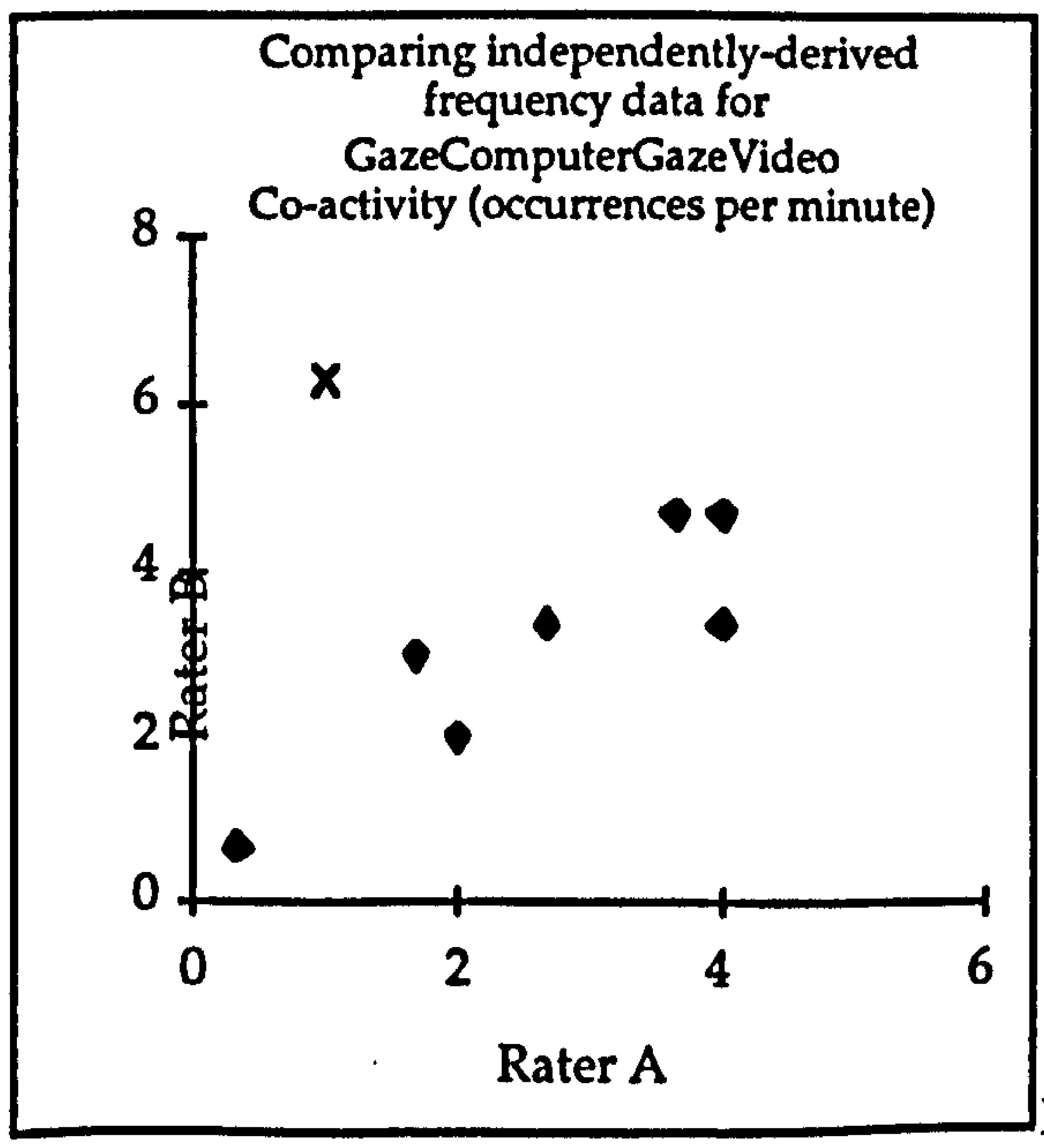
Co-activity (ParticipantA:ParticipantB)		Mean Co-activity duration	Standardized Co-activity frequency	Standardized Co-activity time
GazeVideo	GazeVideo	0.77	0.98	0.92
GazeVideo	GazeComputer	**0.42**	0.63	0.86
GazeComputer	GazeVideo	0.64	**0.34**	0.77
GazeComputer	GazeComputer	0.86	0.94	0.98
GazeComputer	GazeNotes	0.58	0.87	0.95
GazeNotes	GazeComputer	0.78	0.59	0.91
GazeNotes	GazeNotes	0.92	0.84	0.98
GazeVideo	GazeNotes	0.71	0.81	0.96
GazeNotes	GazeVideo	0.96	0.95	0.98

Table 4.23: Pearson's *r* correlation coefficients for independent raters' within-participants GazeGaze Co-activity data

Agreement between raters for GazeGaze Co-activities is good for all Co-activities on the standardized time measures and for most Co-activities on the duration and standardized frequency measures. Two Co-activities have conspicuously low correlation coefficients (emboldened in Table 4.23). These are two comparisons of two forms of GazeVideoGazeComputer Co-activity. There are two data sets for this Co-activity as each Activity is distinguished by participant location within dyads. Conceptually, these data are identical.



igure 4.19: Comparing GazeVideo GazeComputer Co-activity duration data derived from independent raters



igure 4.20: Comparing GazeComputer-GazeVideo Co-activity standardized frequency data derived from independent raters

The scattergrams presented as Figures 4.19 and 4.20 show that there are some clear outliers, from the dyad made up of subjects 11 and 9 and from the dyad of subjects 2 and 4 respectively. Removing these data points improves the correlation for GazeVideoGazeComputer mean duration measure from 0.42 to 0.87 and for the GazeComputerGazeVideo frequency measure from 0.34 to 0.93. The duration measure is associated with another less extreme outlier, which may have been removed, and yet there seems to be a general issue about the reliability of this measure of Co-activity state.

Between Subjects: Gaze-Speech Co-activities

Co-activity (ParticipantA:ParticipantB)		Mean Co-activity duration	Standardized Co-activity frequency	Standardized Co-activity time
GazeVideo	SpeechUtterance	0.86	0.94	0.96
GazeVideo	SpeechSilence	0.70	0.86	0.95
GazeComputer	SpeechUtterance	0.82	0.94	0.94
GazeComputer	SpeechSilence	0.94	0.86	0.95
GazeNotes	SpeechUtterance	0.96	0.95	0.95
GazeNotes	SpeechSilence	0.91	0.86	0.95

Table 4.24: Pearson *r* correlation coefficients for independent raters' between-participants GazeSpeech Co-activity data

Table 4.24 shows that correlations were high for all measures of between-participants SpeechGaze Co-activities.

Between Subjects: Speech-Speech Co-activities

Co-activity (ParticipantA:ParticipantB)		Mean Co-activity duration	Standardized Co-activity frequency	Standardized Co-activity time
SpeechUtterance	SpeechUtterance	0.22	0.90	0.13
SpeechUtterance	SpeechSilence	0.97	0.79	1.00
SpeechSilence	SpeechUtterance	0.89	0.95	0.93
SpeechSilence	SpeechSilence	0.98	0.84	0.95

Table 4.25: Correlation coefficients for independent raters' between-participants SpeechSpeech Co-activity data

There was very poor agreement in Speech Speech standardized time and mean duration data derived from the two raters. Figure 4.21 shows that these low Pearson r values can be traced to a particular data point, in both cases for the dyad including Subject ID 3. These are shown as a crosses amongst the other data points, represented by boxes. Elimination of this dyad brought agreement up to acceptable levels (Pearson's $r = 0.81$ for the duration measure and 0.82 for the standardized time measure).

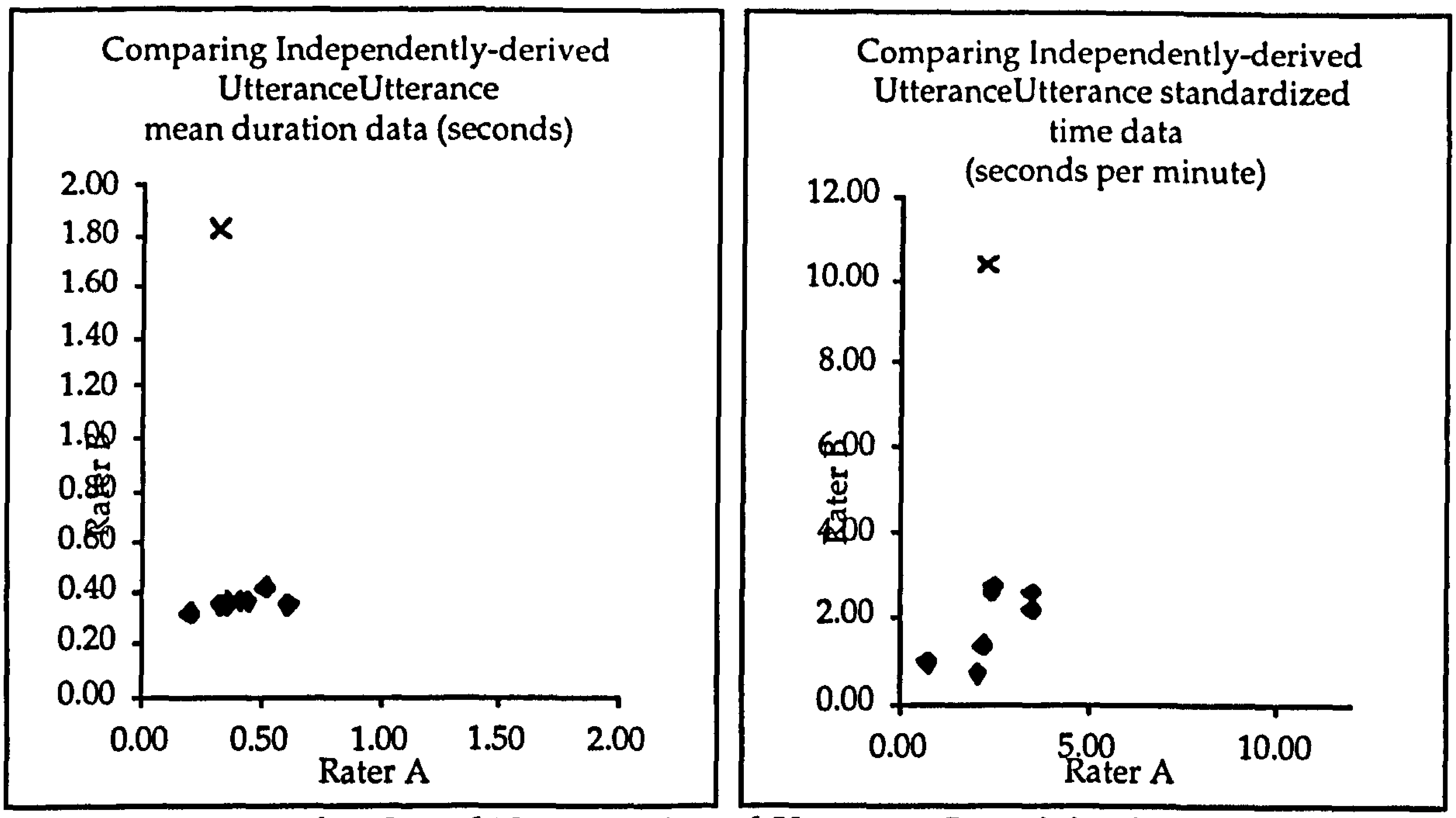


Figure 4.21: Comparing SpeechUtterance:SpeechUtterance Co-activity duration and standardized time data, derived from independent raters

4.9.3 Summary of Co-activity reliability and validation

In general, Co-activity measures appear to be statistically valid, in the case of Gaze, and reliable between observers, for other Co-activity combinations. Duration measures did not fare quite as well as standardized frequency and time measures. One particular subject was treated rather differently by the two observers. As all data must be analysed at the level of the dyad rather than the individual, that subject and their dyad partner were removed from later analyses.

4.10 Interpreting measures of Co-activity

At this point, it may be useful to consider the nature of the information provided by the three measures here employed, in terms of how the results might be interpreted, the potential for redundancy among the measures and the appropriateness of the treatments applied.

Duration data are conceptually quite problematic. While mean duration, standardized frequency and time data are all derived from the single Activity Set ratings of observers, the former is disadvantaged by focusing on the single expression of a Co-activity state. Standardized frequency and time data are both general expressions for Co-activities over the whole period under analysis. As such, they represent to some extent the total effort expended by individuals within the forms of Activity identified. In this way, no great claims need be made about the status of individual Co-activities, it is simply that they emerge from the flow of behaviour as expressions of effort expended in various ways.

In this way, mean duration data might be read erroneously as independent acts, expressed by an individual regardless of all else that they are doing. For example, the GazeVideoSpeechUtterance Co-activity might be thought of as a single intentional behaviour invoked by an individual, for whatever reason. It will be recalled that duration data are supposed to be indicative of the effort required of an individual in order to derive some benefit from the Activity concerned[29]. However, these Co-activity data are merely derivative of the stream of an individual's mutually-exclusive Activities. GazeVideoSpeechUtterance cannot reasonably be argued about as if it were initiated as a new and uniquely motivated behaviour. Although each state is presented as if it were somehow a uniquely specified and performed act, there is a danger of literally interpreting this as equivalent to an assertion that each state is fully and uniquely specified, on the fly, by those taking part in the interaction. Rather, each is a consequence of a change in some particulars of an Activity Set contributing to the prior general state of the interactive system. The Gaze at Video link might have begun before this episode of Utterance and may continue beyond it.

Duration data additionally seem to have faired least well in estimates of inter-rater reliability. As discussed above, the three measures deployed are not mathematicallly independent. Knowledge of any two of the three base measures can provide a full description of dyadic engagement in the defined Activities. From both conceptual and operational viewpoints, considerations of Co-activity seem to be poorly informed by mean duration data. Henceforward, only the standardized measures shall be considered in this connection.

Standardized frequency data might be criticized on similar conceptual grounds. Implicitly, frequencies model the value of any Activity expression as equivalent i.e. one glance is much the same as another. It would seem inappropriate to model, for example, the investment in GazeComputer Activity in the same way when one instance might lasting thirty seconds and

---

[29]This is a simplification as there are a number of determinants of behaviour, including plans, situational opportunity, environmental imperatives and so on.

another only half a second. Since frequency has been 'standardized' to a one-minute period, some of this criticism might be deflected: the investment in each Activity by number is given *on average*, so that an intermediate time in Activity is reflected by these data. Nevertheless, frequency data for Co-activities should be treated with caution. Standardized time data for a Co-activity family, that is, a combination of Activity Sets, has a direct relationship to Simple Activity data. They must always sum to 60. There is no clear relation between Co-activity and Simple frequency data in this way.

Co-activity standardized time measures present the least conceptual difficulties. Very small periods spent in some Co-activity state contribute little to the overall score and long periods contribute a great deal. As a general measure of Activity investment it stands to work very well. Unfortunately, unlike the duration and frequency measures, the relationships between Activities exposed by standardized time data are multiplicative rather than additive. Standardized time, as seconds-per-minute, is a measure of the proportion of time for which an Activity was engaged. Since the ANOVA procedure models contributions to variance in an additive manner, it is not clear that these results are readily interpretable.

Consider the following hypothetical data. Should there have been a total of 10 seconds of Utterance per minute, and 50 seconds of Silence, any more than 10 seconds of a GazeComputer with SpeechUtterance Co-activity could not have been measured - yet it would have been perfectly possible to have found 20 seconds of GazeComputer with SpeechSilence. These figures, should they have obtained, would be far more telling of a relationship between GazeComputer and SpeechUtterance than the, nominally twice as large, GazeComputer whilst SpeechSilence. The figures in Table 4.15 thus make any relationship between these Activity Sets difficult to interpret and the ANOVA reported to be unreliable. Looking again at the bald figures for time proportion in Table 4.15, the number of 'seconds per minute' in each Gaze Activity state differ by very similar orders in combination with SpeechUtterance and SpeechSilence. It is thus not possible to understand the contingencies of expression between Activities with this procedure.

An obvious step to take would be to log transform these proportional data. However, so doing would undermine the intelligibility of any findings: a relationship of some kind would have been demonstrated but the extent of value of the relationship would need some further exploration. The next chapter considers alternative procedures for estimating the degree of Co-activity organization, using the standardized time measure.

5 Synchronization: analysing Co-activities for evidence of temporal organization

Chapter summary

Chapter Four showed how Activity Set Analysis may be operationalized to provide statistical summaries of timed behavioural data. Simple breakdowns into interaction Activity duration, frequency and proportion of time in state were described. Analyses of Co-activity states were carried out for evidence of contemporaneous state contingency in duration and frequency data. Duration data were rejected as a measure of Co-activity due to the difficulty of interpreting a derivative of the original Simple ratings, and frequency data were considered subject to similar limits of interpretation. Analysis of proportion-of-time-in-Activity-state presented difficulties. This chapter considers alternative approaches to gauging departure from temporal independence. A novel approach is described for measuring temporal organization amongst recorded Activities. A new statistic, *S* for extent of synchronization, is presented as a conceptually sound and empirically usable estimate of the extent to which engagement of one Activity is contingent on the state of another.

5.1 Activity Set multivariate analysis of variance

The previous chapter dealt with Simple (single Activity Set) analyses and began to address the problem of analysing Contingent data or determining inter-Activity Set contingency. In other words, a preliminary attempt was made to expose the extent to which Activity engagement varied as a function of other potentially simultaneous Activity states. Pairings of Activities from different Activity Sets were described as potential Co-activities, where Co-activity implies some degree of organization between the constituent Activities. Duration and frequency Co-activity data were made subject to analyses of variance, with the two Activity Sets concerned treated as independent variables, their respective Activities serving as levels of the variable. Contingent duration data were reported with reservations about their intelligibility, given that the Co-activity states were specified purely in observational terms. Contingent frequency data were felt to meaningfully inform questions about Co-activity organization nevertheless to suffer a similar disadvantage to the duration data. This chapter concentrates on proportional data as the most illuminating measure for estimating temporal dependency.

Where the ANOVA procedure indicated a main effect on either of the Activity Sets, this was taken as grounds for supposing that there was a difference in the degree of expression of the Activity Set's component Activities. The main effect thus indicated intra-Activity Set differences. For example, an ANOVA main effect on proportion of time in state for the Gaze Activity Set implies that the amount of time spent in GazeComputer, GazeNotes, and GazeVideo was not equivalent. Tukey's HSD follow-up tests were used to determine whether the three Activities were all differentiated from one another, as levels of the variable.

Dependencies between Activities from different Activity Sets were examined by looking at evidence in statistical interactions, or where some of the variance could be accounted for by a combined effect of both Activity Sets. This was initially assumed to indicate that one or more of the constituent Activities in one Activity Set was conditioned by the occurrence of one or more of the Activities in the other set. In the example worked through, within-subjects Gaze and Speech, an interaction was indeed found. The Simple statistics for Speech showed that, overall, more time was spent in Silence than Utterance so, overall, more time in each of the Gaze states in combination with SpeechSilence would be expected just on this basis. This would then be expressed as a main effect in the multivariate ANOVA. Following up with a simple main effects test, it appears that Utterance Silence is associated with time in the GazeComputer Activity state than in GazeVideo or GazeNotes. One would normally interpret such a difference within the level of one variable as evidence of a statistical interaction. However, as these data are proportions, the dependent variables do not differ in combination by a fixed difference. They effectively "mislead" the ANOVA by combining in a multiplicative manner, violating its additive model of linear contribution to variance. One might get around this problem by log transforming the proportional data. A disadvantage with so doing is then interpreting any differences that are found, concerning the nature of the temporal dependence between Activity Sets. ANOVAs cannot estimate the magnitude or, more importantly for these time-based data, the sense of any association. By sense of temporal association, the difference between *conjoint* and *disjoint* organization is intended, i.e. two behaviours can be coordinated to consistently co-occur or consistently not co-occur.

One might form an opinion on this matter either by looking at tabulated means or graphical plots. Unfortunately, even in the simplest case of pairs of Activity Sets, the data do not readily lend themselves to "by-eye" interpretation. Again this is a consequence of the baseline ratio of time spent in Activity states. In the example of the previous chapter, from Graph 4.6 it is clear that all Gaze Activities make up a greater portion of interaction time in combination with Silence than with Utterance. It is far from clear whether the greater differences between each Gaze Activity within SpeechSilence compared to each within SpeechUtterance are anything more than a simple amplification of the within-Activity Set differences, expected from baseline Speech Activity. Meaningful analyses of several orders of Activity Set synchronization are conceivable. For example, one might wish to determine conjoint synchronization of three collaborators on a document set implicated in a particular task context. This would require five Activity Sets which, in an ANOVA situation, would require prediction of a five-way interaction of amongst n-level log-transformed variables. Such an approach would be intractable and difficult to interpret.

5.2 Correlating behavioural indices

Duncan and Fiske (1977) used correlations to look for evidence of relationships amongst interaction behaviours. They examined a wide range of coded behaviours as 'acts', from 'self adapters' (touching) to smiling, and included speech and gaze. Statistics were generated for each act, variously including frequency, duration and proportion of time in state. These measures were taken as indices of interaction traits and used for the computation of act intercorrelations. Matrices were computed separately for the four possible combinations of gender in a dyad (male-female and female-male distinctions were meaningful for some act correlations) and also between acts of the same participant.

Calculating correlations between such indices stands to estimate the degree of association between coded behaviours *over the whole course* of a coded session. As coefficients range from -1 to 1, they also give an indication of the sense of the association. However, this "sense" differs in an important way from the kind of discrimination advocated in Chapter Two. Activity Set Analysis requires some indication of the contemporanaeity of activity over time, as an understanding of the dynamics of interactive behaviour. A negative correlation implies that the behaviours concerned tended not to co-occur at all. A positive correlation suggests that the behaviours were exhibited at about the same rate or for a similar extent across the sample of observations, not that they tended to co-occur. Whether they occurred within the session but not at the same time is a completely different and more interesting question.

Duncan and Fiske looked at indices coding certain composite behaviours, mapping onto the Co-activity concept developed in this thesis. Rather than deriving them computationally from rater recordings of component Activities, they were explicitly coded as such. So a correlation coefficient would be calculated for the extent indices "smiling while speaking" and "smiling", and for "smiling while speaking" and "speaking", and so on. They intended to determine broad associations between subgroups of behavioural acts as evidence of the organization of social interaction.

Interpretation of association discovered in this way is again subject to the confounding influence of baseline extent of behaviour. Inevitably, someone who spends a lot of time smiling and who also spends a lot of time looking towards their conversational partner will spend a fair amount of time in a state described by these behaviours in combination. Duncan and Fiske struggle to make sense of these associations, observing: "one might expect, in principle, that indices for rates during speaking turns would be artefactually correlated with each other (but) we found no evidence for such misleading effects"(p. 61). Unfortunately, finding no evidence might mean that there is no relationship or that a relationship is masked by conflicts between mathematical and behavioural effects. They conclude by admitting: "Our substantive findings from our studies of correlates of acts in interaction do not impress us ... the proportion of large correlations was not greatly above the proportion of which would be expected by chance" (p.123).

5.3 Information Theory: organization by separating signals from noise

Information Theory provides a mechanism for determining the organization or patterning of behaviour in sequences of events. The problem addressed in this chapter is concerned not with sequences of events but with temporal organization by co-occurrence. Information statistics deal with empirically derived and/or predicted differences in proportional terms. This basis would seem to lend itself to a consideration of temporal synchronization by proportion of time in state. However, it should be clear that in so doing assumptions must be made about the equivalence of proportions of time and proportions of discrete events.

5.3.1 Shannon and Weaver's concept of information as degree of uncertainty

The development of electrical communications media gave rise to a pressing need for some principled means for their assessment. The problem required a new way of considering technologies, not purely as physical devices but as conveyors of "information", a conceptual shift instigated by Wiener's publication of "Cybernetics" (1948). Shannon and Weaver (1949) devised Information Theory as a means for assessing the content of transmitted signals by inventing some rules for its quantification. They observed that any signal exists within a spectrum of transception possibility. The concept of a signal was set up as a function of the probability of its chance occurrence, by relating the signal to the number of "units of uncertainty" in the signal medium. Information was then defined as the number of binary digits (bits), as the power to which 2 must be raised to give the number of alternatives from the set of elements over which the signal was to operate. In the simplest case, where all alternatives are equally likely to occur, uncertainty H is simply a function of the number of these alternatives *m*, as the log to the base two of the probability of any one occurrence, *p*:

$$H = \log_2 (^1/_m) = \log (p)$$

As the log of numbers less than 1 is negative, and it makes sense to describe the number of bits as a positive number, conventionally the log of the reciprocal probability, $^1/_p$, is used. So where the set contains two elements of equal a priori likelihood, the probability of one of them happening is 0.5 and the amount of information measured in this way would be 1 bit. That is, there are two alternatives, expressed by the binary digit '1'. The bit value is variously described as the average information (in something that has happened), the uncertainty (of something that may happen) or, more generally, as the entropy of a set of events (Attneave, 1959; Golomb, Peile & Scholtz, 1994).

The determination of event probabilities is dependent on how the signal itself is conceived. A signal might be constrained to select from among a fixed set of alternative prefabricated targets. The targets might be complex on any number of dimensions. All that is required of the signal is to identify one of them to be invoked. The early telegraph operated in much this way. The *selective information content* of a telegraphy signal is given by the chance of specifying any particular character, say one from a set of thirty seven alphanumeric characters. From the formula above, if all characters are equally likely to be selected, the selective information content would be $\log(1/37)$=5.21 bits.

5.3.2 Organization as uncertainty reduction

The telegraph example assumed that all thirty seven elements are a priori equally likely to be sent. In general, and especially for linguistic elements, an assumption of equiprobability is unsafe: entropy is not monotonically related to the number of elements in the set. It is possible to factor in the differences in likelihood of occurrence as a weighting to the formula for equiprobable event sets. This is formalized as:

$$H(p_1, \ldots, p_m) = \sum_{i=1}^{m} p_i \log_2(1/p_i)$$

where H is the entropy of the set, p are the probabilities of elements in the set, and m is the number of events in the set. Suppose the written language in our hypothetical telegraph comprised two subsets by probability, one of 30 equally likely characters and a second subset of 7, each of double the likelihood of the first subset. The entropy of this thirty seven character telegraph would be 5.14 bits, compared with 5.21 for 37 equally likely characters. That some elements of the alphanumeric set are more likely than others reduces the overall entropy of the set.

The above account deals purely with a priori probabilities to describe information in a signal. Observation gives additional information on the likelihood of elements in the set of possible signals. Comparing the information given a priori with that given by a posteriori probabilities allows determination of the information gained. In information theoretic terms, the relationship between sets is a matter of their probability of sequential association in event series. The separate entropies of each set, from independent observations, are compared to the entropy of their observed association. Attneave illustrates this with a hypothetical experiment, where one set of events comprises X stimuli and a second comprises Y responses.

The empirical entropy of the set of stimuli, responses and observed responses to stimuli are given as:

$$H_{(x)} = \sum^{i} p_i \log\left(\frac{1}{p_i}\right)$$ [Stimulus set, *X*]

$$\hat{H}_{(y)} = \sum^{i} \hat{p}_i \log\left(\frac{1}{\hat{p}_i}\right)$$ [response set, *Y*]

$$\hat{H}_{(x,y)} = \sum^{i} \hat{p}_{ij} \log\left(\frac{1}{\hat{p}_{ij}}\right)$$ [observed stimuli-response set, *XY*]

where *i* is instance of *X* and *j* is any instance of *Y*. Note that the ^ symbol is used to denote empirical rather than *a priori* probability. In an experiment, the stimuli are controlled and hence the probabilities of set *X* are the actual probabilities. Information transmitted, *T*, is then given by:

$$\hat{T}_{(x,y)} = H_{(x)} + \hat{H}_{(y)} - \hat{H}_{(x,y)}$$

If there is no pattern evident in subjects' responses, i.e. there is no systematic relationship between sets X and Y, then there will be no information gain. In terms of the formula above, *T* is zero if the sum of the entropies of sets X and Y elements is equal to the entropy of their coincident observation, XY. If, on the other hand, the entropy of joint element observation (i.e. stimulus-response combinations) is found to have a smaller bit value than the sums of separate entropies, this is the amount of information transmitted as the uncertainty reduced from the independent set entropies alone. Perfect transmission, where there is a consistent and unique response for every stimulus in the example, would be where $\hat{T}_{(x,y)} = \hat{H}_{(x)} = \hat{H}_{(y)} = \hat{H}_{(x,y)}$

Signals are not always organized to select from a finite set of alternative targets. For example, it would not normally be meaningful to measure the information content of a television picture by its raster elements. Measures might be in terms of its *structural information content*, defining the received picture along several dimensions. Alternatively, its *metrical information content* might be assessed, given by considering the content of the received picture in terms of the 'weight of evidence' it contains for the discriminability of elements. Whether structural or metrical information content is to be assessed, both depend on a definition of the elements against which to set these treatments, *logons* as logical information units in the former case or *metrons* as the base unit of statistical evidence in the latter. The translation of signal properties into logons or metrons then gives one the same basis for assessing information content as the structural approach described above.

In principle, Information Theory would appear to have much to offer to the problem of Contingent analysis in an Activity Set approach to interactive behaviour. Activity Sets might be viewed as dimensions, and the proportional data of their constituent Activities as probabilities from which to compute the entropy of sets. Determination of empirical Co-activity probability might then be contrasted with an a priori Co-activity probability, given by the entropies of each Simple statistic. Taking Contingent data on Gaze and Speech Activity from Chapter Four, entropies and information transmitted would be as follows[30]:

Activity	Seconds per minute	Proportion
GazeVideo	11.00	0.183
SpeechUtterance	23.00	0.383
GazeVideo:SpeechUtterance (Observed)	4.49	0.075
GazeVideo:SpeechUtterance (Expected)	4.22	0.070

Table 5.1: Standardize time data for within-participants Gaze and Speech Activity, expressed as proportions

$\hat{H}$ (GazeVideo)= .183(log .183) = 0.449

$\hat{H}$ (SpeechUtt) = .383(log .383) = 0.530

$\hat{H}$ (GazeVideo:SpeechUtt) = (.075) log (.075) = 0.280

$T = 0.449 + 0.530 - 0.280 = 0.699$

Contrast this calculation of T with one based on the a priori entropy for the set GazeVideo:SpeechUtterance (the product of Simple proportions):

$\hat{H}$ (GazeVideo:SpeechUtt) = (.070) log (.070) = 0.709

$T = 0.449 + 0.550 - 0.709 = 0.010$

So T suggests that the occurrence of SpeechUtterance and GazeVideo Activities are not entirely independent, to the extent that these two Activities share 0.7 bits of information.

[30]These data are for illustrative purposes. The H statistic is properly applied to sets of frequency observations, summing over all values of plogp in the set. The assumption here is that H and thereby T would be calculated for each dyad.

It is instructive to pre-empt the discussion of Between Subjects Activity Set Analysis, to be reported in the next chapter, by considering Contingent data on Speech.

Activity	Seconds per minute	Proportion
SpeechUtteranceA	21.37	0.356
SpeechUtteranceB	24.82	0.414
SpeechUtteranceA: SpeechUtteranceB (Observed)	3.12	0.052
SpeechUtteranceA: SpeechUtteranceB (Expected)	8.84	0.147

Table 5.2: Standardize time data for between-participants Speech:Speech Co-activity, expressed as proportions

Here, Speech Activity for dyads is differentiated by a Location variable with two levels, A and B[31]:

$\hat{H}$ (SpeechUttA) = .356(log .356) = 0.531

$\hat{H}$ (SpeechUttB) = .414(log .414) = 0.527

H(SpeechUttA:SpeechUttB) = (.052) log (.052) = 0.222

T = 0.531 + 0.527 - 0.221 = 0.836

Again, *T* shows that there is some organization of engagement in SpeechUtterance on the part of the dyads, to the extent that they share 0.836 bits of information. Clearly, from the difference between observed and expected Co-Activity proportions, the sense of organization this time is the opposite of within-subjects GazeVideo:SpeechUtterance.

5.3.3 Entropy in Activity and Co-activity proportions of time in state

The entropy-based metrics, as described above, originate in frequency probabilities. The *T* statistic sets up contrasts of empirically derived and/or predicted differences in proportional terms. Given that the temporal synchronization data in question here are expressed as proportions, both for Simple Activities and for Co-activities, calculation of entropies for each

[31]That is, the Activities of participants in this experiment were differentiated only by their communications workstations.

Activity Set and for their contemporaneous combination appear to be straightforward. Co-activity data require no special transformation to render them amenable to producing a T estimate of information transmitted for each recorded dyad. These T data, in bits, might then be subject to a parametric comparison of means.

However, such an approach has a number of shortcomings in this application. First, information statistics are established to deal with whole sets of possible combinations and produce bit estimates of their organization. In this application, dealing with continuous data rather than the discrete frequencies normally associated with Information Theory, it is not very clear how one should properly conceive of an event set. Activity Sets themselves, through their mutual exclusivity and exhaustiveness properties, are rather more analogous to event sets. Entropies for the Gaze and Speech Activity Sets would be separately calculated, and an entropy for the set of Co-activities made up by their combination:

$$\hat{H}_{(Gaze)} = p_{iv} \log(1/p_{iv}) + p_{in} \log(1/p_{in}) + p_{ic} \log(1/p_{ic}) + p_{ie} \log(1/p_{ie})$$

$$\hat{H}_{(Speech)} = p_{iu} \log(1/p_{iu}) + p_{is} \log(1/p_{is})$$

$$\hat{H}_{(GazeSpeech)} = \left(\begin{array}{l} p_{ivu} \log(1/p_{ivu}) + p_{inu} \log(1/p_{inu}) + \\ p_{icu} \log(1/p_{icu}) + p_{ivs} \log(1/p_{ivs}) + \\ p_{ins} \log(1/p_{ins}) + p_{ics} \log(1/p_{ics}) \end{array} \right)$$

where i is the entity under observation, v is Video, c is Computer, n is Notes, e is Elsewhere, u is Utterance and s is Silence. Information transmitted would then be:

$$\hat{T}_{(Gaze:Speech)} = \hat{H}_{(Gaze)} + \hat{H}_{(Speech)} - \hat{H}_{(GazeSpeech)}$$

The Activity Set is an *observational* construct, the membership of which has no necessary conceptual significance. It is important to be able to ask questions about specific Co-activities, defined across Activity Sets. This is analogous to a requirement for identifying links between specific stimuli, in Attneave's example, and specific responses.

A second matter concerns the idea of temporal organization itself, expressed as synchronization. The ordinary use of the term means happening at the same time. One form of organization in time, then, is when things are coordinated such that they occur at the same moment or within the same period, rather than sometimes together and sometimes not. In other words, it may be expressed as co-incidence at greater than chance levels. Another way of thinking about temporal organization is the polar opposite: consistently not occurring at the same moment or in the same period. This would be avoidance at greater than chance levels. Synchronization as described here is intended to expose temporal organization and so it makes sense to account for the former *positive Synchronization* as well as the latter form, *negative Synchronization.* Information transmitted is given as a magnitude in bits, without preserving the *sense* of association. Again, this is not a trivial problem given the difficulties of working "by eye" to interpret a single bit estimate of organization. This problem rapidly escalates with as the order of Activity Set Analysis increases.

5.4 Synchronization: a consideration of chance co-occurrence and consistent temporal organization

The analytic problem described in Chapter Four was that the Simple proportion of time spent in each Co-activity component made any estimate of the organization of Co-activity states very difficult. For an analysis of mutual gaze (that is, when two individuals make eye contact) Argyle and Ingham (1972) observed that mutual gaze is likely to occur just by chance on the basis of independent looking levels. The Simple proportions for each of the component Activities of a Co-activity state provide a posteriori grounds for estimating an a priori probability for Co-activity state incidence. Consideration of the difference between observed values and those that would be expected on the basis of chance alone suggests a χ^2 approach. However, the χ^2 model deals with frequency data for a sample group, including an N term, and Simple Activity Set Analysis provides proportional data for the time spent in each of the defined Activity states, with no clear analogue of N.

5.4.1 Estimating expected co-occurrence of Activities

Simple statistics for proportion of time in state estimate the likelihood an Activity state would have obtained at any given moment during the period of observation. Where a Co-activity comprises a pair of Activities, the expected probability of that Co-activity state at any moment over the period, p_{ce}, is the product of the two observed Activities' Simple proportions of time in state, a and b:

$$p_{ce} = a\,b \qquad (1)$$

Activity Set Analysis' Contingent statistics summarize the observed proportion of time a dyad spent in the possible Co-activity states. The difference between these empirically derived probabilities for Co-activity states, p_c, and their expected occurrence, p_{ce}, thus indicates how far the Co-activity's constituent Activities depart from independence, as the extent of temporal organization between them. This is presented as a measure of 'Synchronization':

$$\text{Synchronization}(c) = p_c - p_{ce} \qquad (2)$$

Consider the data in Table 5.1, drawing from the study reported in the last chapter. The proportion of time spent in GazeVideo Activity as 11 seconds per minute, or 0.183 as a proportion. The proportion of time spent in the SpeechUtterance Activity state is given as 23 seconds per minute, or 0.303. Thus, the a priori probability of the GazeVideo:SpeechUtterance Co-activity state is 0.183 x 0.383 = 0.070.

Table 5.1 shows that on average the dyads in this study spent 4.5 seconds per minute both looking at the TV monitor and talking at the same time, or 0.075 as a proportion of their discussion. The Synchronization score is thus 0.075 - 0.070 = 0.005. The same process can be applied to simultaneous SpeechActivity, data for which are shown in Table 5.2.

There are two problems with this example. The first is that, for explanatory convenience, these figures are means drawn from the case study sample. Synchronization involves a non-linear combination of terms and thus it should be determined separately for each Co-activity. Analysis should then be carried out with individually obtained Synchronization scores as the units of analysis. The second problem concerns the interpretation of the score itself. The χ^2 statistic simply expresses the square of the observed-expected difference as a proportion of the expected frequency. The analogue for Synchronization would then be expressed as a fraction of the expected value:

$$\text{Synch}(C) = \frac{\sum_{i=1}^{m} \frac{p_{ci} - p_{ce.i}}{p_{ce.i}}}{m} \qquad (3)$$

The idea of expressing the degree of departure from independence in terms of some other quantity brings this discussion to a limitation of the Synchronization measure.

5.4.2 A restrictive relationship between Simple and Contingent proportions (1): maximum Synchronization

The logic of Synchronization as described above has been used to estimate the extent to which mutual gaze is non-random (Rutter et al. 1977; Argyle and Ingham, 1972). Observed 'individual looking levels' in an experimental discussion have been used to calculate the amount of time both would be expected to look at each other at the same time, and the difference between this figure and the observed amount of eye contact taken as an indication of mutual gaze coordination. In a critique of this technique, Wagner, Clark and Ellgring (1983) noted that there is a restrictive mathematical relationship between 'individual looking levels', or Simple Activity proportions to use the general scheme adopted here, and mutual gaze, or Co-activity state. Any Co-activity state cannot extend for longer than the observed extent of the smaller of its constituent Activities. To put it another way, some critical information is lost when departure from independence is expressed as the difference between observed Contingent proportions and product of Simple proportions.

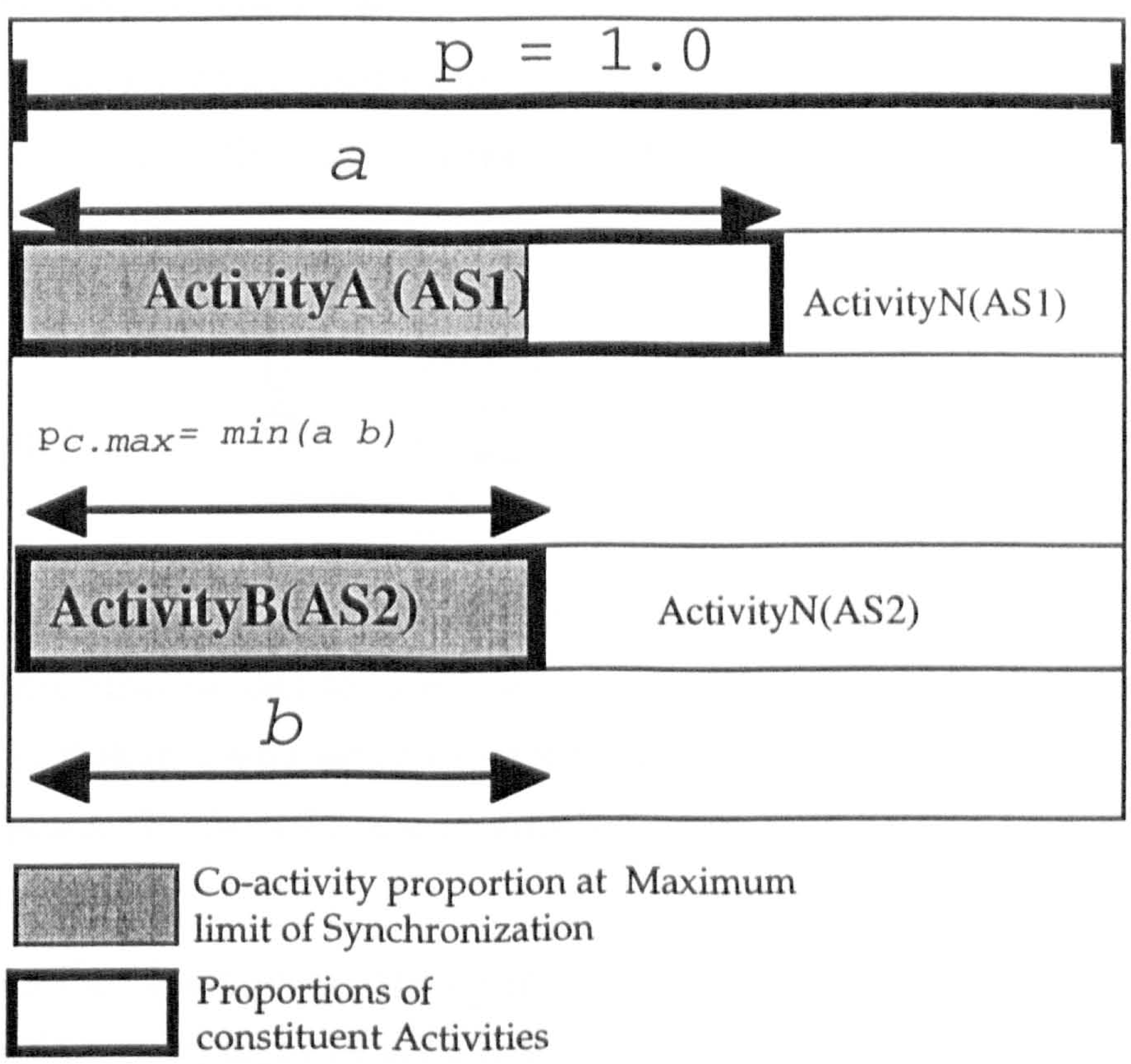


Figure 5.1: Representing positive Synchronization as maximum alignment by relative proportion of available time

The maximum proportion of time in a Contingent Activity state is limited by the smaller proportion of its Simple constituents. In the case of the example above, GazeVideo:SpeechUtterance, the smaller proportion of time in state is for GazeVideo. Working from the observed Simple proportions, it is not possible for the Synchronization score of GazeVideo:SpeechUtterance to exceed 0.183 of the discussion time, i.e. the proportion of time spent in GazeVideo. Figure 5.1 represents this limiting case schematically, showing the proportional extent and conjunction of two Activities, A and B, from different Activity Sets, AS1 and AS2 respectively. Figure 5.1 is intended to convey this idea both as a matter of relative proportions of time in state and also as a matter of mean probability of conjunction in a period of observation. The case shown here is where maximum coincidence obtains between two Activities observed to have extended individually for different proportions of time. The shaded area within Simple Activity A shows that a potential Co-activity, AS1.A:AS2.B, cannot extend for more time than given by the proportion of time in state known for Simple Activity B.

It may thus be seen that Synchronization is conditioned by the Simple proportions from which a maximum possible Co-activity proportion, $p_{c.max}$, may be derived:

$$p_{c.max} = \min(a\ \ b) \\ \Rightarrow \\ \text{Synchronization}_{max} = p_{cmax} - p_{ce} \tag{4}$$

The extent to which Co-activities are organized to co-occur would be better expressed in terms of this limit as the *maximum* Synchronization that could obtain, given the observed Simple statistics of Co-activity components. For the set of Co-activity observations, C, this is expressed as:

$$S{+}ve(C) = \frac{\sum_{i=1}^{m} \dfrac{\text{Synch}_i}{p_{ci.max} - p_{ci.e}}}{m} \tag{5}$$

This quantity is described as S^{+ve}, in reference to the denominators' status as the maximum limit on departure from independence, rather than the χ^2 zero departure, and that the sense of departure is towards greatest synchronization. So S+ve for the example here would be 0.005/0.183-(0.183 x 0.383) = 4.44%.

There is then the other sense of departure from independence, of organized temporal disjunction, against which to moderate the Synchronization score.

5.4.3 A restrictive relationship between Simple and Contingent proportions (2): minimum Synchronization

Where the sense of temporal organization is not of contemporaneity but avoidance, the Synchronization score would be negative. This score should similarly be expressed as a percentage, this time of the *minimum* possible co-occurrence. This minimum limit depends on the sum of Co-activity constituents' time in state in relation to the total time available.

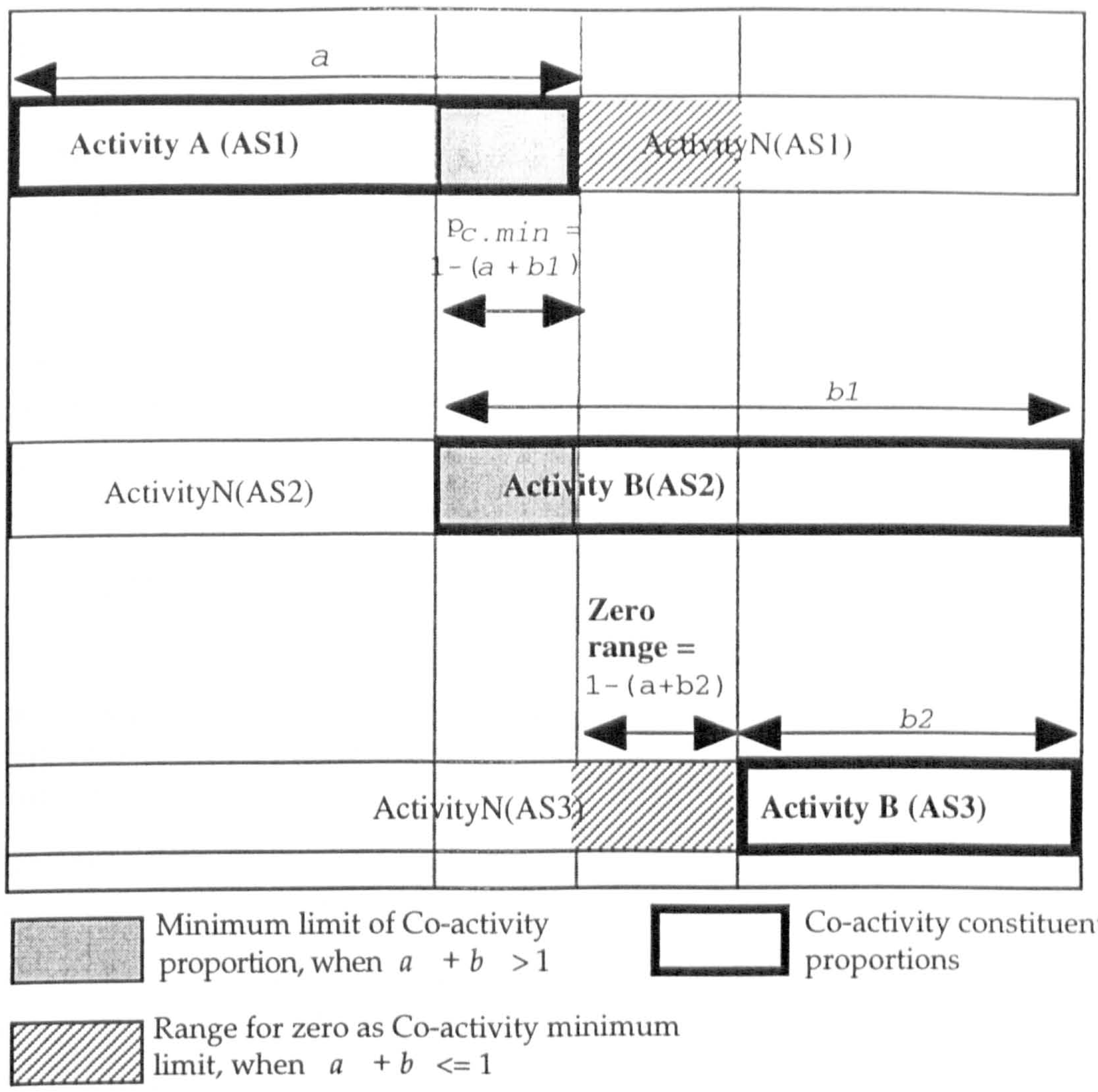


Figure 5.2: Representing negative Synchronization as maximum avoidance by relative proportion of available time

The top and centre Activity Sets represented in Figure 5.2. again show a potential Co-activity *AS1.A:AS2.B* as a shaded portion of the relevant Simple Activities. Minimum Synchronization represents the case where the potential Co-activity constituents are as disjoint as their Simple proportions allow. Here, the critical consideration is what opportunity there was, based on the a posteriori Simple proportions, *a* and *b1* , for these Activity states to have "avoided" one another. That is, how negatively Synchronized might have been the Activities *AS1.A* and *AS2.B*. In Figure 5.2, the minimum Co-activity proportion, $p_{c.min}$, for the Co-activity *AS1.A:AS2.B* is represented as an overlap in constituent Activity proportions, *a* and *b1*. The relationship governing this opportunity is given by the time remaining, if any, after allowing for the eventuality of Co-activity constituents occurring at maximally divergent intervals. In proportional terms, this is the difference between the sum of the constituents, *a* and *b*, and 1.0 (representing total time available). Nominated Activities in the top and bottom Activity Sets, *AS1* and *AS3*, show a different relationship. The *AS1.A:AS3.B* Co-activity can be specified without any overlap at all. In fact, there is a probability range for zero overlap, indicated by the hatched portions of *AS1* and *AS3*, given the Simple proportions $\hat{}p_a$ and $\hat{}p_{b2}$ represented here. So where $a + b <= 1$, the minimum limit is zero, otherwise it is given by the remainder of 1- $(a + b)$:

$$p_{c.min} = \min((1-(a + b)), 0)$$
$$=>$$
$$\text{Synchronization}_{min} = p_{c.min} - p_{ce} \quad (6)$$

Synchronization$_{min}$ might then be used as the denominator to determine how close the observed Co-activity proportion approaches it minimum limit, as the maximum departure from independence for negative Synchronization, S^{-ve}:

$$S^{-ve}(C) = \frac{\sum_{i=1}^{m} \frac{(p_{ci} - p_{cie})}{p_{ci.min} - p_{cie}}}{m} \quad (7)$$

5.4.4 A restrictive relationship between Simple and Contingent proportions (3): Synchronization, limit ranges and expected distribution of coincident states

The preceding discussion took the product of Simple Activity proportions as an appropriate model for the random coincidence of Activities in a fixed period. If both Simple proportions are small, it follows that the expected Co-activity proportion, the relative amount of time spent in a state composite of both Activities, evaluates to a correspondingly small proportion. Similarly, if both Simple proportions are large, the expected probability of coincidence is also relatively large. There are limits to both conjunction and disjunction, the maximum and minimum possible synchronization, against which this expected coincidence must be set. This is where Watts, Monk and Daly-Jones (1996) formalized Wagner, Clark and Ellgring's (1983) observations on this matter, adopting measures of temporal coordination between the Simple Activities, S^+ and S^-, as the difference between observed and expected coincidence standardized on these maximum or minimum possibilities.

There is an important and indeed severe limitation to this treatment of synchronization. Conjunction, or disjunction, exists on a continuum between the maximum and minimum synchronization, conditioned by the magnitude of each Simple proportion in relation to the other and to the total available time. Consider negative synchronization. One can define a set of cases falling within limits where $p_{c.min} = 0$ (i.e. $a + b < 1$), with the same value for expected coincidence and yet, due to differences in relative magnitude of a and b within the period, with widely differing potential for precisely aligning to this minimum. That is, the smaller the proportion of time remaining after allowing for a complete disjunction between *A* and *B*, the less likely it should be that such a disjunction might be observed. There is no provision in the preceding formulations for taking this matter into account.

Thus far, only one Co-activity state has explicitly entered this discussion, AB. There are in total four states that may be reasoned about on the basis of Activities A and B: AB, A not-B (A~B), not-A B (~AB), and not-A not-B (~A~B). The zero range is the proportion of time spent in the state ~A~B. These states allow one to consider what might be termed the 'scope of opportunity' for alignment, positive or negative, between potential Co-activity constituents. Figure 5.3 represents the proportion of an interaction period remaining after allowing for complete disjunction of the two Activities A and B, as a 'zero range'. That is, a range of conjunction cases on a uniform distribution, satisfying a condition of absolute minimum Synchronization ($p_{c.min}$=0).

Consider the following hypothetical data, in association with Figure 5.3. On the preceding account, an expected conjunction, p_{ce}, of two Activities from different Activity Sets, with proportions *a* and *b*, separately observed to take up 0.4 parts of the period (i.e. $a = b = 0.4$), would be 0.16. Exactly the same p_{ce} obtains for Activities occupying 0.2 and 0.8 of the available time or even 1.0 and 0.16. As Figures 5.2 and 5.3 illustrate, the minimum possible Synchronization between Activities depends on their proportions and may not be zero. The crude estimate of p_{ce} does not allow one to distinguish between cases where Simple Activities must be minutely coordinated to "achieve" minimum Synchronization ($a + b >= 1$, the zero range =0), and cases where Simple Activities have a range of opportunities for mutual avoidance ($a + b < 1$, the zero range is large).

Figure 5.3 contrasts Co-activities with identical p_{ce} of 0.16, $p_{c.min}$ is zero in both cases, yet in the top example, $p_{\sim a \sim b}$= (1 - 0.8 - 0.2) = 0 allowing only one possible "position" of *ab*. In the bottom example, $p_{\sim a \sim b}$ = (1 - 0.4 - 0.4) = 0.2, there is a range of possible positions. It is convenient to think of *ab* in terms of *a* "moving" with respect to *b*. This is really a shorthand for reasoning about sets of conjunctions that might obtain over a number of observations, or for counting up possible *ab* states from short divisions of the full interaction period. In this way, the 0.2 of space represents a range of movement possible for *a* whilst still completely avoiding *b*, minimum Synchronization remaining at zero. Hence, the "zero range" in Figures 5.2 and 5.3 describes the time making up the remainder of a period, after allowing for *a*and *b* at minimum synchronization. So, describing these two examples as having the same expected Co-activity proportion is self-evidently false.

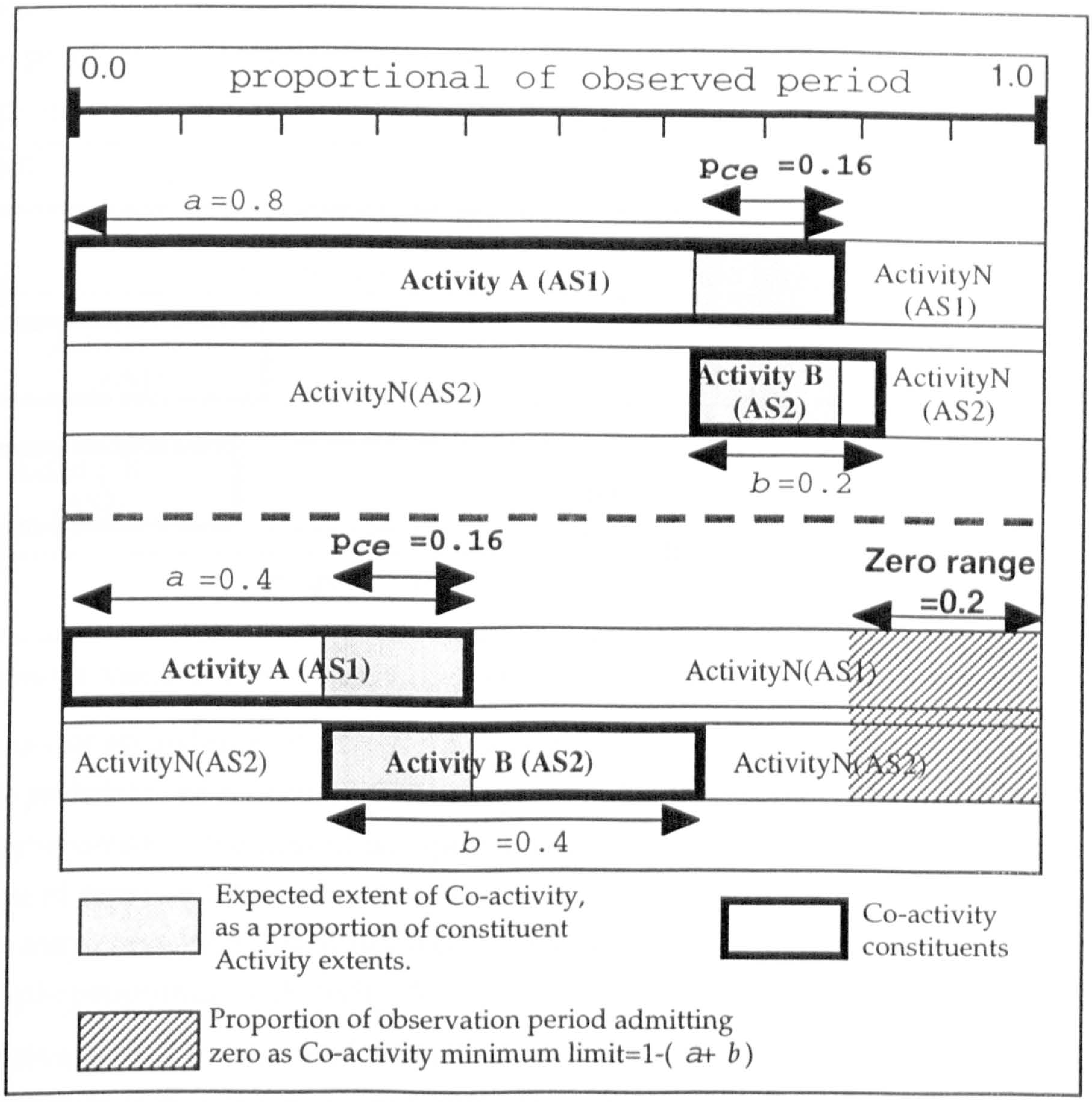


Figure 5.3: Representing cases with equivalent p_{ce} while $p_{c.min}$ = zero

Similarly, achieving maximum Synchronization would require minute coordination where Simple Activity proportions are very similar. Where one Activity is ongoing for the majority of the period and the other is known to obtain for a small fraction of it, maximum Synchronization would require somewhat less coordination.

This "zero range" notion is important since it brings into focus the idea of variance in alignment of each Activity at intervals within the interaction period, conveniently described as the motion of A with respect to B. Overlap covers a range of Co-activity state possibilities with possible conjunctions as a function of their Simple proportions *a* and *b*. That is, in some instances A will occur at a time entirely disjoint from B, in others partially disjoint partially conjoint, or whole conjoint with B. There are circumstances, given by the empirical Simple proportions of time in A and B, where some Co-activity states might be excluded from consideration immediately. As we have seen, when $a + b >= 1$, ~A~B cannot obtain. Where $a = 0$, then only the state ~AB can obtain. The earlier p_{ce} estimate might be refined by examining more closely the possible p_c components given by a random distribution of Activity A and B proportions within the scope of the interaction (a and b limited by 1)[32].

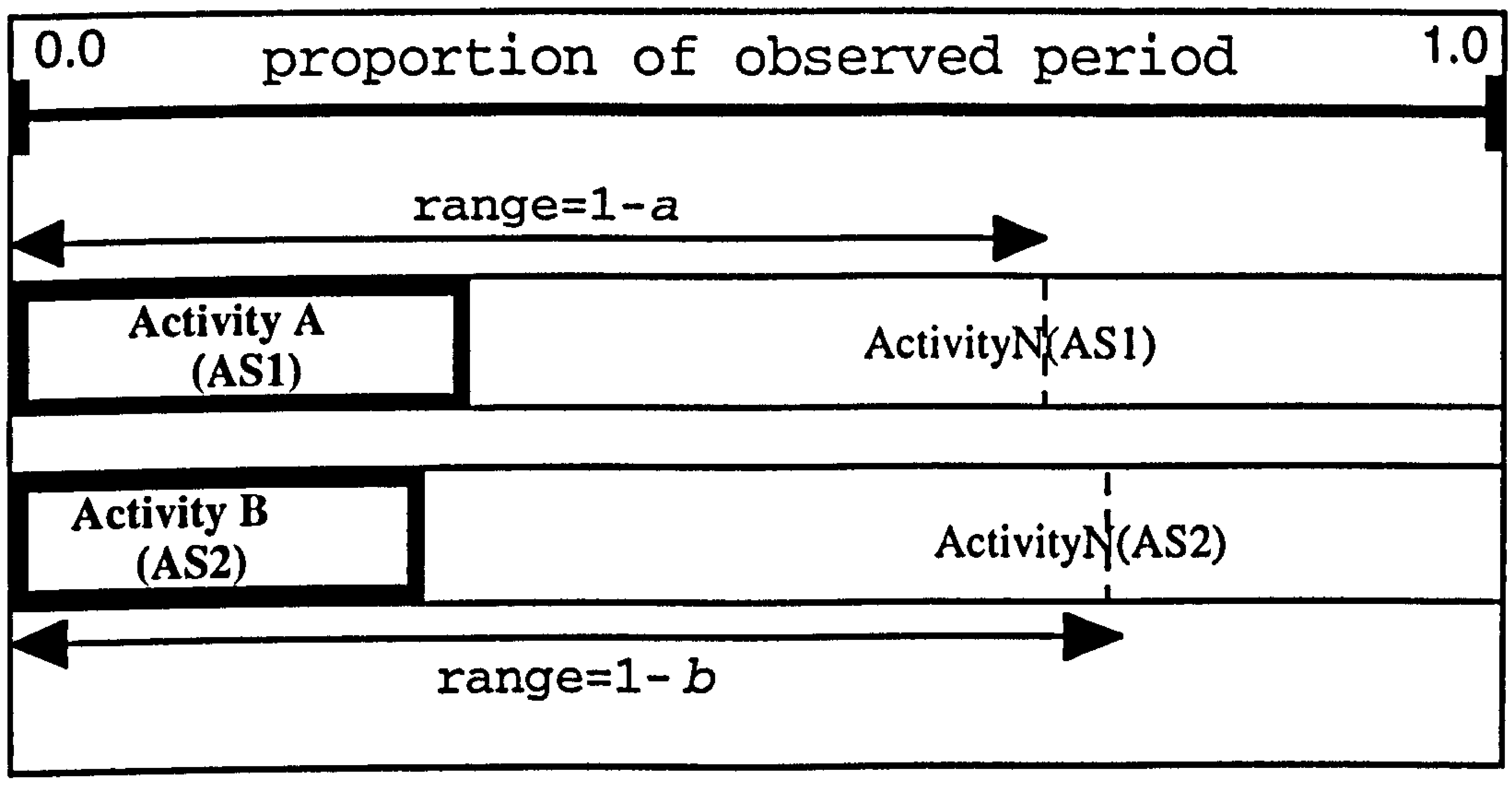


Figure 5.4: Variability of a Pair of Activities

Consider an instance of Activity A, *ai*, the earliest occurrence of the Activity. The *probability* of *ai* occurring at any point in the interaction period is given by the *proportion* of the period occupied by Activity A, *a*. Continuing with the form of representation introduced earlier in this section, Figure 5.4 shows that one may consider a distribution of *ai* as limited by the magnitude of the Simple proportion of Activity A.

Given all other instances of Activity A, *am*, in the known proportion *a*, *ai* could occur only within the restricted range from zero to 1-*a*. The remaining proportion of the period (i.e. between 1-*a* and 1) is taken up by the other instances of A, *am*. It should be noted that for the purposes of this argument, all instances of Activity A have been conceptually grouped into a contiguous unit, thus the scale in Figure 5.4 should not be confused with linear time. The range "zero to 1-*a* " represents the limits of a distribution of Activity A as a proportion of the whole interaction period and not a sequential representation of time in seconds from a time zero to time 1-*a* of an interaction.

---

[32] I am much indebted to Dr Nick Merriam, York University Computer Science Dept., for advice on this problem and still more to Dr Peter Lee of York University's Mathematics Department for steering me away from unproductive avenues in my quest to improve on S^+ and S^-, suggesting this approach and his subsequent work on the proof.

Grouping in this way allows consideration of how much of A and how much of B falls into each of the possible states on a random distribution of the separate Activities A and B. One might thereby estimate the variance of the p_{ce} (the expected proportion of interaction taken up by the state AB). A value for the expected variance based on the known proportions, *a* and *b*, would allow a full standardization of the original Synchronization formula (2), rather than the partial standardization given by prior formulae relying just on $p_{c.min}$ and $p_{c.max}$, (5) and (7).

All else being equal, *ai* would be equally likely at any point, in a range X, within this proportional range and so the a priori distribution of A is uniform. By the same token, *bi* could occur at any point, in a range Y, between zero and 1-p*b* but for the time being, assume that p*b* is fixed to begin at some point, y, in this range. As previously discussed, the maximum possible coincidence is limited to the smaller of the two Simple proportions. Let Activity A exist but take up less time than Activity B, so $0 < a < b < 1$ and $p_{c.max} = a$. The assumption of *a* falling on a uniform distribution being uniformly distributed within its 0 to 1-*a* range leads on to a consideration of how *a* might overlap with *b*, for all the possible relative positions of *a* and *b*. The degree of coincidence may be calculated by refining the prior considerations of this chapter into four key cases of Simple proportion magnitude and relative timing.

These cases are represented in Figure 5.5. Each case describes a set of possible values of *a* and *b*, given the assumptions described above, and a range of earliest first instance of Activity B (the fixed 'y'). The earliest instance of Activity A, *ai*, is considered to "move" from left to right up to its range limit. As these assumptions translate into Co-activity states, assuming a uniform distribution, the degree of p_{ce} is calculable by integration between limits defined by transitions into state combinations. Case I describes a set of relative conjunctions of *a* and *b* where:

(i) *a* could occur entirely before *b*, giving absolute disjunction as the state A~B (*ai* in the range 0 -> $y - a$).
(ii) partial overlap so that *ai* is the last instance to coincide with *b*, allowing for some disjunction, as A~B, and some conjunction, AB, for *ai* within limits $y - a$ -> y
(iii) total coincidence, state AB, for *ai* within limits y -> $y + b - a$.
(iv) partial coincidence, AB, and partial disjunction, A~B, within limits $y + b - a$ -> $y + b$.
(v) total disjunction again as *ai* moves up to its furthest position, y+*b* -> 1 - *a*.

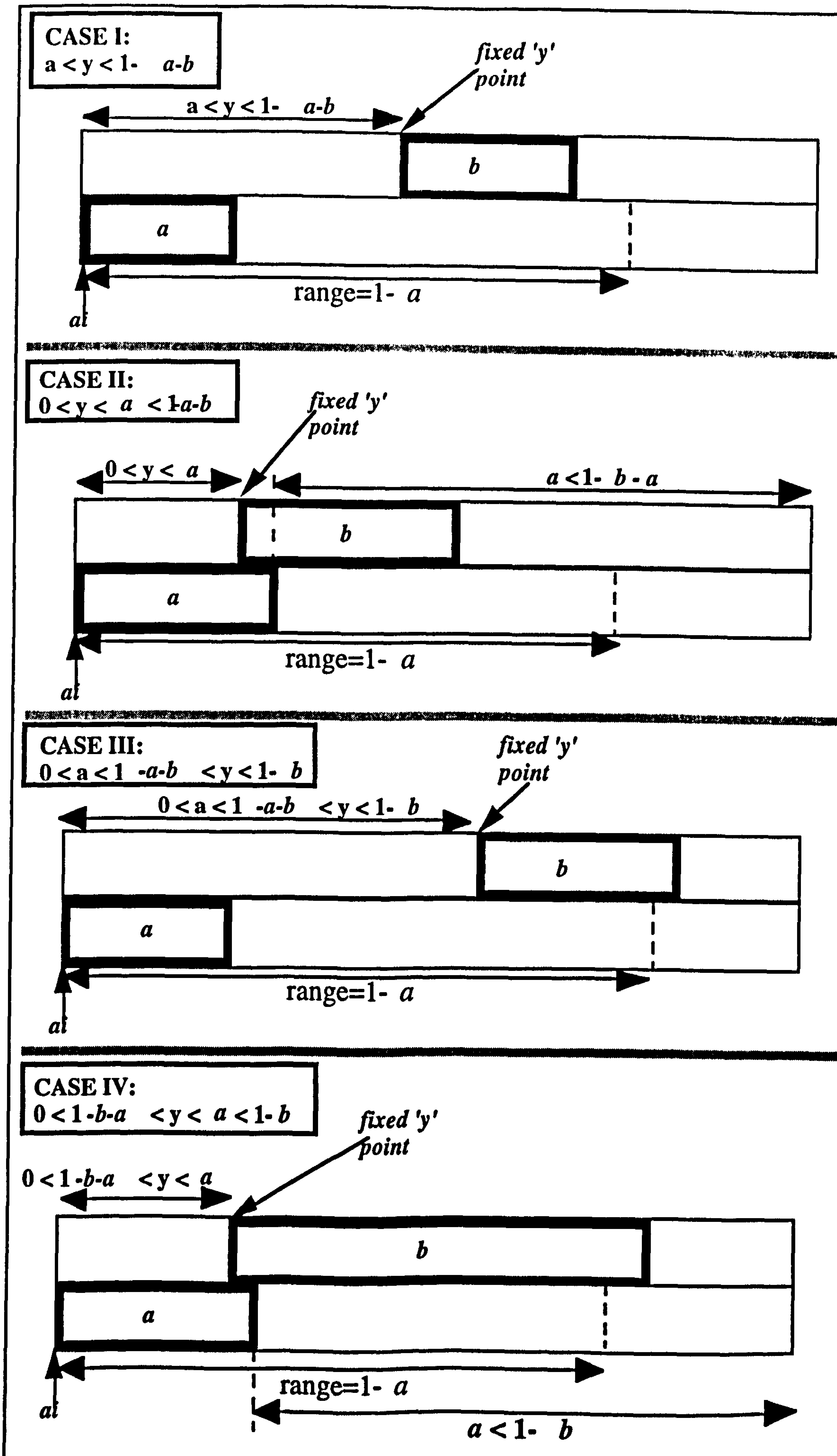


Figure 5.5: Cases of A in relation to an Activity B, known to be longer, at different moments of occurrence in an interaction period

Clearly, relative conjunctions (i) and (v) do not contribute to p_{ce}, as they are defined by complete disjunction, (iii) gives the proportion of time for full conjunction, (ii) and (iv) the proportion of time for partial conjunction. By integrating functions for each of these components between the relevant limits one may derive an expression for expected coincidence, taking into account cases of partial as well as full overlap (see Appendix 4):

$$p_{ce} = \frac{a\,b}{1-a} \tag{8}$$

Case II describes the set of conjunctions where the earliest occurrence of Activity B, y, precludes absolute disjunction before *b*. The first integral contributing to p_{ce} thus operates within more restrictive limits, from 0 to y, y given as less than *a*. The new integral gives the expression:

$$p_{ce} = \frac{a\,b - {}^{1}/_{2}\,(a - \mathrm{y})^2}{1-a} \tag{9}$$

Case III represents the set where y is too late to allow absolute disjunction after *b*. In this case, the final component contributing to p_{ce} is restricted by the limit of *ai* movement, 1-*a*. Integrating for Case III gives:

$$p_{ce} = \frac{a\,b - {}^{1}/_{2}\,(1 - b - a - \mathrm{y})^2}{1-a} \tag{10}$$

Case IV represents the set where y is too early for absolute disjunction before *b* and also where *b* is of such a magnitude that there is also no possibility of absolute disjunction afterwards. To put it another way, there isn't room for all of *a* to fit in either before or after *b*. In this case, the integral of the three components evaluates to:

$$p_{ce} = \frac{a\,b - {}^{1}/_{2}\,(a - \mathrm{y})^2 - {}^{1}/_{2}\,(1 - b - a - \mathrm{y})^2}{1-a} \tag{11}$$

There are other sets of possibilities, besides those illustrated in Figure 5.5. These include where the magnitude of *b* is such that *a* must either be coincident with it at the beginning or end of the range of movement, depending on the y-position, and where *a* and *b* sum to more than 1 (i.e. there must always be some degree of coincidence). However, these give effectively the same formulae as described above[33] and so are not elaborated.

---

[33]For $a+b>1$, there are differences in sign for a term that is squared and so do not affect the result.

The assumption of a fixed first position for the proportion b has only limited value: Activity B is distributed just as is Activity A. A similar procedure may be followed to reason about any point on b moving varying through Y with respect to X (that is, the range of a) up to the limit of 1-b, through states B~A, AB and B~A again. However, now there are expressions to describe the movement of a, rather than having to assume a fixed x. These may be integrated between the relevant limits for b at state transitions. These integrals resolve into two final expressions for expected coincidence of Activities A and B, depending on whether their proportions sum to more than 1:

If $a + b <= 1$:

$$p_{ce} = \frac{a\,b}{1-a} - \frac{{}^{1}/_{3}\,a^{3}}{(1-a)\,(1-b)} \qquad \textbf{(12)}$$

If $a + b > 1$:

$$p_{ce} = \frac{a\,b}{1-a} - \frac{{}^{1}/_{3}\,a^{3}}{(1-a)\,(1-b)} + \frac{{}^{1}/_{3}\,(a+b-1)^{3}}{(1-a)\,(1-b)} \qquad \textbf{(13)}$$

Using the formulae for p_{ce} given in **(12)** and **(13)**, the original Synchronization formula **(2)** may be invoked and now standardized with respect to the square root of the variance of p_{ce} , the standard deviation of mean expected coincidence:

Calculation of the variance of expected coincidence requires two terms, the sum of square of coincidences and the square of the sum of coincidences. As described above, the term p_{ce} represents the sum of all coincidences, defined by possible coincident states given fixed and known Simple proportions of various magnitudes. The sum of squares, represented below as $(p)^2{}_{ce}$, thus requires raising each of the expressions leading to **(8)**, **(9)** and **(10)** to the power of two, prior to integration, and then following the same steps to give these expressions, again contingent on the magnitude of a and b:

If $a + b <= 1$:

$$(p)^2{}_{ce} = \frac{a^2 b - {}^{1}/_{3}a^{3}}{1-a} - \frac{{}^{1}/_{6}\,a^{4}}{(1-a)\,(1-b)} \qquad \textbf{(14)}$$

If $a + b > 1$:

$$(p)^2{}_{ce} = \frac{a^2 b - {}^{1}/_{3}a^{3}}{1-a} - \frac{{}^{1}/_{6}\,a^{4}}{(1-a)\,(1-b)} + \frac{{}^{1}/_{6}\,(a+b-1)^{4}}{(1-a)\,(1-b)} \qquad \textbf{(15)}$$

Variance of p_{ce} is thus given by the difference between (14) or (15), depending on the conditional, and the square of (12) or (13). Finally, using the formulae for p_{ce} given in (12) and (13), the original Synchronization formula (2) may be invoked and now standardized with respect to the square root of the variance of p_{ce}, that is, the standard deviation of mean expected coincidence:

$$S = \frac{p_c - p_{ce}}{\sqrt{\{(p)^2_{ce} - (p_{ce})^2\}}} \qquad (16)$$

5.5 Summary

The most promising dependent measure generated from an Activity Set Analytic perspective estimates the proportion of time in a state given by the constituents of two or more Activity Sets. Proportional data are problematic with conventional ANOVA procedures. ANOVA models total variance as the sum of a number of contributory influences. Proportions combine in a multiplicative rather than additive manner, so that differences in the levels of one factor condition the extent of association possible with the levels of other factors. A consequence is the spurious indication of statistical interactions between factors. Information Theory was considered in this application, modelling associations between Activity Sets as 'signals' amidst the noise of chance co-occurrence. It is conceivable that some research questions may be adequately addressed by applying Information Theory in this way. However, for the a number of other questions this is unsatisfactory since the *T* statistic loses the sense of the association (positive or negative). A requirement identified earlier in this thesis was to draw out contemporaneous relationships between temporal streams of behaviour as evidence of their synchronization. An alternative method was developed as Synchronization Analysis. Synchronization Analysis allows one to compare an observed proportion of time in a Co-activity state with a proportion expected by chance alone, given observed Simple Activity proportions uniformly distributed within the limits of the period. A new statistic, *S*, is proposed as a standardized measure of Synchronization, given a random distribution of Co-activity states determined by the relative distribution of Simple Activity states.

6 Synchronization in Gaze and Speech Co-activity in a video-mediated interaction

Chapter summary

The *S* statistic is applied to Co-activity data deriving from the VMI case study reported in Chapter Four. SpeechSpeech Co-activity was found to be near alternate, as would be expected, but indicative of a fluid and "informal" style of exchange. SpeechGaze Co-activity between-participants showed strong evidence of temporal organization whereas there was little to suggest within-participants SpeechGaze Co-activity Synchronization. Whilst one participant was speaking, their partner as a listener tended to Gaze at computer VDU and personal paper notes than chance. Listeners, so defined, were more likely than chance to look at speakers. Gaze at Computers was less likely than chance to occur when other participants were looking at their paper notes. These findings suggest a high degree of inter-Activity temporal organization, including both visual and spoken behaviours, taken as evidence for the video link having an important role in coordinating and exploiting the available resources.

6.1 Analysing Co-activity data with *S*

Chapter Four described a case study of VMI to introduce the application of Activity Set Analysis to interactive behaviour. Analysis of Co-activities was put in abeyance, pending the invention of an appropriate way to deal with the most promising measure, standardized time in Co-activity state. Chapter Five presented the *S* statistic as a conceptually sound and empirically viable metric of Co-activity temporal organization. This Chapter applies *S* to the possible Co-activities given by Speech and Gaze Activity Sets defined in Table 4.1, Chapter Four.

Section 4.2.2 outlined grounds for expecting to find evidence of certain forms of Co-activity organization, based on the discussion in Chapter 1. These are formulated below as hypotheses, although still with the exploratory rather than comparative purpose of this Chapter.

Kendon (1967) suggested that speakers tend not to look at listeners and listeners tend to look at speakers. Sellen et al. (1992, 1995) and Vertegaal (1997, 1998), citing Kendon's conclusions, have argued that one of the strengths of VMI systems should be to smooth turn transitions. One might therefore expect within- and between-participant differences Speech-GazeVideo Co-activities. *Within-participant* GazeVideoSpeechUtterance corresponds to a speaker looking at listener. *Between-participant* GazeVideoSpeechUtterance corresponds to a listener looking towards a speaker.

1a/ Other-directed visual behaviour, operationalized as GazeVideo Activity, should depend on whether or not the gazer is also speaking. That is, *within-participants* GazeVideoSpeechUtterance should be negatively Synchronized and GazeVideoSpeechSilence should be positively Synchronized.

1b/ A person's GazeVideo Activity should depend on the Speech Activity of their interlocutors. That is, *between-participants* GazeVideoSpeechUtterance should be positively Synchronized and GazeVideoSpeechSilence should be negatively Synchronized.

GazeComputer and GazeNotes are broadly comparable, as text-based resources, in a way that GazeVideo is not: Chapanis (1977) and others have observed that written media inevitably absorb more time than spoken media. Clark (1996) contends that the communication is reliant on participants' mutual and reciprocal knowledge, as common ground. Arguably, GazeComputer represents a consultation of a resource with guaranteed common ground and so it should be clearly differentiated from GazeNotes in relation to Speech Activity. Furthermore, the deictic capacity afforded by the mouse pointer offers considerable paralinguistic potential to participants. Consequently, GazeComputer is expected to have a more vital relationship with SpeechUtterance, as the primary engine of communication, than GazeNotes Activity.

2a/ *Within-participant's* Synchronization scores for the two GazeVideo-Speech Co-activities should differ from both GazeComputer-Speech Co-activities and from GazeNotes-Speech Co-activities.

2b/ *Between-participants* Synchronization scores for GazeVideo-Speech Co-activities should differ from both GazeComputer-Speech Co-activities and from GazeNotes-Speech Co-activities.

3a/ *Within-participant's* GazeComputerSpeechUtterance was expected to be more positively Synchronized than GazeNotesSpeechUtterance.

3b/ *Between-participants'* GazeComputerSpeechUtterance was expected to be more positively Synchronized than GazeNotesSpeechUtterance.

Much of the social psychological literature on interaction gaze mechanisms considers mutual gaze (Argyle, 1988; Rutter, Stephenson & Dewey, 1981; Wagner, Clark & Ellgring, 1983), or the GazeVideoGazeVideo Co-activity as it is operationalized here. If mutual gaze has such a privileged place in the gamut of visual behaviours of relevance to social interaction, one might expect its occurrence to exceed purely chance levels.

4/ GazeVideoGazeVideo Co-activity was expected to exceed chance levels (i.e. be positively Synchronized).

If the video-as-data hypothesis (Nardi et al., 1993; Whittaker, 1995) is to hold much weight here, GazeVideo should negatively Synchronize with both GazeComputer and GazeNotes. This is on grounds that GazeVideo serves to inform participants of the relevant communication activity of their remote interlocutors. Putting a Clarkian analysis together with a Distributed Cognition emphasis on the artefacts of joint significance in collaboration (Hutchins, 1995), one might expect both notes- and computer-screen-directed-gaze to be profitably coordinated. Furthermore, given the stronger communicative utility advocated for GazeComputer, via its explicit commonness and gestural capacity, GazeComputerGazeComputer Co-activity should show a stronger Synchronization than GazeNotesGazeNotes.

5a/ GazeVideoGazeComputer and GazeVideoGazeNotes should both be negatively Synchronized.

5b/ GazeComputerGazeComputer and GazeNotesGazeNotes should both be positively Synchronized.

5c/ GazeComputerGazeComputer should be more positively Synchronized than GazeNotesGazeNotes.

Four sections report the **within-subjects** SpeechGaze Co-activities; **between-subjects** SpeechGaze Co-activities; **between-subjects** GazeGaze Co-activities; and **between-subjects** SpeechSpeech Co-activities. In each case, mean *S* scores are reported as described in Chapter Five, together with:

(i) standardized time data for each Simple Activity contributing to the Co-activity state;
(ii) observed standardized time in Co-activity state; and
(iii) expected standardized time in Co-activity state.

All data are quoted both in seconds per minute and as proportions, to aid interpretation and to allow the reader to cross-reference against data tables in Chapter Four, where required.

Each person was seen as a participant in an interactive system of behaviour and so the unit of analysis for all Co-activity data reported here is the dyad. No operational distinction was made between participants as their experimental role and available resources were identical. Consequently, the Activity of one member is considered in much the same way as the other. Following on from this assumption, there two issues for Co-activity Synchronization analysis that stand to confuse. Using 'A' and 'B' as arbitrary labels for dyad members, Co-activities may appear to be analytically unique and yet not be independent. For example, the within-participants Co-activities 'A-GazeComputer, A-SpeechUtterance' and 'B-GazeComputer, B-SpeechUtterance' are conceptually indistinguishable[34]. Similarly, the between-participants Co-activities 'A-GazeComputer, B-SpeechUtterance' and 'B-GazeComputer, A-SpeechUtterance' are not materially distinct. In these cases, a composite *S* was generated for each dyad by calculating a mean of each *S* Co-activity variant. So, in these examples, analysis would be of a "*within-participants GazeComputerSpeechUtterance Co-activity S*" and a "*between-subjects GazeComputerSpeechUtterance Co-activity S*". Were there to have been some consistent ground for distinguishing participants, a within-dyads analysis of these types of Co-activities would have been appropriate.

S data were initially compared against zero with a t-test, as a baseline estimate of any consistent temporal relationship, since an *S* score of zero means that the observed and expected proportions of time in Co-activity state are identical. Where appropriate, comparisons are then made on the basis of the five hypotheses above. The Co-activities themselves are not independent and so *S*

[34]See Table 4.1 for defintion of terms.

comparisons between Co-activities in each section (as the conjunctions for pairs of Activity Sets) were carried out as paired-samples t-tests. In tables, statisitical significance is indicated by an asterisk (*) for p<.05 and two asterisks (**) for p<.01. Results are summarized and briefly discussed at the end of each section.

For brevity, the various Co-activities shall be referred to in the main by using the first letters of Activity Sets and Activities involved, i.e. GazeNotesSpeechSilence is GNSS, GazeVideoGazeNotes is GVGN, and so on.

6.2 Describing within-participants Gaze and Speech Co-activity Synchronization

Data reported here are examined for evidence of temporal relationships among each participant's own Activities, without consideration of the other dyad member. For the Activity Sets defined in Chapter Four , there is only one family of Co-activity to fall under this heading: Speech Gaze (see Table 4.15).

Speech Utterance Activity is first examined in relation to each Gaze Activity, followed by a consideration of Co-activities with SpeechSilence.

	GazeVideo	SpeechUtterance	Observed Co-activity GVSU (within)	Expected[35] Co-activity GVSU (within)
Seconds per minute	10.51	22.53	4.16	4.75
(SD)	(5.83)	(5.78)	(3.01)	(3.10)
Proportion of discussion	0.175	0.375	0.069	0.079
(SD)	(0.10)	(0.10)	(0.050)	(0.052)
	"S"=	-0.099	(0.196)	

Table 6.1: Within-participants GazeVideoSpeechUtterance (GVSU) Co-activity Synchronization

[35]It is important to understand that the Simple standardized times are included here to help with interpreting observed and expected Co-activity figures. As discussed extensively in Chapter Five, Simple Activity data have a non-linear relationship with the expected figure. For this reason, the mean expected Co-activity figures given here differ from the expected GVSU Co-activity figures reported in the between-participants analysis, despite being associated with identical mean Simple Activity statistics.

	GazeComputer	SpeechUtterance	Observed Co-activity GCSU (within)	Expected Co-activity GCSU (within)
Seconds per minute	28.53	22.53	10.20	13.79
(SD)	(11.55)	(5.78)	(6.55)	(7.29)
Proportion of discussion	0.476	0.375	0.170	0.230
(SD)	(0.19)	(0.10)	(0.109)	(0.121)
	"S"=	-0.690**	(0.327)	

Table 6.2: Within-participants GazeComputerSpeechUtterance Co-activity Synchronization

As Table 6.1 shows, GazeVideo and SpeechUtterance Activity of each participant appears to have co-occurred at no more nor less than chance levels (t (6) = 1.34; p=.230). However, as is clear from Table 6.2, GazeComputer and SpeechUtterance Activity of each participant was coordinated at well above chance levels (t(6)= 5.58; p<0.001). The mean *S* score is negative, showing that this coordination involved participants tending not to look at the computer VDU whilst in Utterance.

	GazeNotes	SpeechUtterance	Observed Co-activity GNSU (within)	Expected Co-activity GNSU (within)
Seconds per minute	19.82	22.53	7.63	8.95
	(10.96)	(5.78)	(4.44)	(5.37)
Proportion of discussion	0.330	0.375	0.127	0.149
	(0.18)	(0.10)	(0.074)	(0.089)
	"S"=	-0.227*	(0.196)	

Table 6.3: Within-participants GazeNotesSpeechUtterance Co-activity Synchronization

Table 6.3 shows that GNSU Co-activity was also coordinated by avoidance, suggesting that a participant was less likely to look at their paper notes when in Utterance than chance alone would predict (t (6) = 3.06; p<.05).

As Speech Activity is defined by a binary Activity Set, the pattern of Synchronization results evident for the Gaze Activity with SpeechSilence might be expected to mirror those for SpeechUtterance. Mathematically this does not necessarily follow: it would be to ignore the difference in opportunity for one Activity to coincide with another, were the binary set to take up markedly different proportions of the interaction. This follows from the possibility of SpeechSilence occurring both when the other dyad member is in Utterance and when they are also in Silence, suggesting quite different contexts for interpretation (both reading, for example).

	GazeVideo	SpeechSilence	Observed Co-activity GVSS	Expected Co-activity GVSS
Seconds per minute	10.51	35.47	6.36	7.24
(SD)	(5.83)	(10.25)	(3.29)	(5.16)
Proportion of discussion	0.175	0.591	0.106	0.121
(SD)	(0.10)	(0.17)	(0.055)	(0.086)
	"S"=	-0.081	(0.543)	

Table 6.4: Within-participants GazeVideoSpeechSilence Co-activity Synchronization

Co-activity data for GVSS show no evidence of Synchronization (t(6)=0.39; p=.708).

	GazeComputer	SpeechSilence	Observed GCSS(within) Co-activity	Expected GCSS(within) Co-activity
Seconds per minute	28.53	35.47	18.34	20.18
	(11.55)	(10.25)	(6.80)	(9.86)
Proportion of discussion	0.476	0.591	0.306	0.336
	(0.19)	(0.17)	(0.113)	(0.164)
	"S"=	-0.219	(1.382)	

Table 6.5: Within-participants GazeComputerSpeechSilence Co-activity Synchronization

GCSS Co-activity data differs markedly from the GCSU data above, showing no evidence of synchronization (t(6)=0.42; p=.689). GNSS also fails to reach significance, contrasting with the companion GNSU Co-activity data (t(6)=1.41; p=.207).

	GazeNotes	SpeechSilence	Observed Co-activity GNSS (within)	Expected Co-activity GNSS (within)
Seconds per minute	19.82 (10.96)	35.47 (10.25)	12.19 (7.76)	15.30 (10.43)
Proportion of discussion	0.330 (0.18)	0.591 (0.17)	0.203 (0.129)	0.255 (0.174)
	"S"=	-0.536	(1.003)	

Table 6.6: Within-participants GazeNotesSpeechSilence Co-activity Synchronization

> *H1a* *Within-participants* GazeVideoSpeechUtterance should be negatively Synchronized and GazeVideoSpeechSilence should be positively Synchronized.

Hypothesis 1a suggests that Gaze Activity varies with Speech Activity within participants. This has been supported, but the picture is complex. On the evidence of these data, GazeComputer and GazeNotes Activities occur less than chance alone would predict when a person is in Utterance, but GazeVideo occurs at about chance levels. No Synchronization, positive or negative, is in evidence between any of the levels of Gaze Activity and SpeechSilence. The apparent contradiction of these findings is easily resolved by considering the four Speech states for each Gaze Activity state: SUSU, SUSS, SSSU and SSSS.

Data presented below show that there is very little time spent in SUSU (see Table 6.21), thus the *S* data for Gaze-SpeechUtterance reported here overwhelmingly represent where a speaker was looking whist their partner was listening. In contrast, there was a very similar period spent in the state SSSS to SUSS and so *S* data for Gaze-SpeechSilence include both where one dyad member was looking whilst their partner was speaking and periods where both with in the state SpeechSilence, consulting materials, navigating information and so on.

> *2a/* *Within-participant's* Synchronization scores for the two GazeVideo-Speech Co-activities should differ from both GazeComputer-Speech Co-activities and from GazeNotes-Speech Co-activities.

Hypothesis 2a suggested that GazeVideo Activity would differ from GazeNotes and GazeComputer in its relationship with Speech Activity. This was indeed found to be the case, since GNSU and GCSU were both negatively Synchronized and GVSU was not.

> *3a/* *Within-participant's* GazeComputerSpeechUtterance was expected to be more positively Synchronized than GazeNotesSpeechUtterance.

It was expected that the Synchronization of GC and GN with Speech Activity would be different, by virtue of the computer's status as a shared and novel interaction device. *S* scores for GCSU and GNSU were thus compared directly and indeed found to differ ($t(6)=2.70; p<.05$). GCSU was thus rather more negatively Synchronized with SpeechUtterance than GNSU (see tables above). They were considered to be similar kinds of resources, since both should involve referring to written information resources. Indeed, it is notable that both GCSU and GNSU are negatively Synchronized, lending some face validity to the comparison.

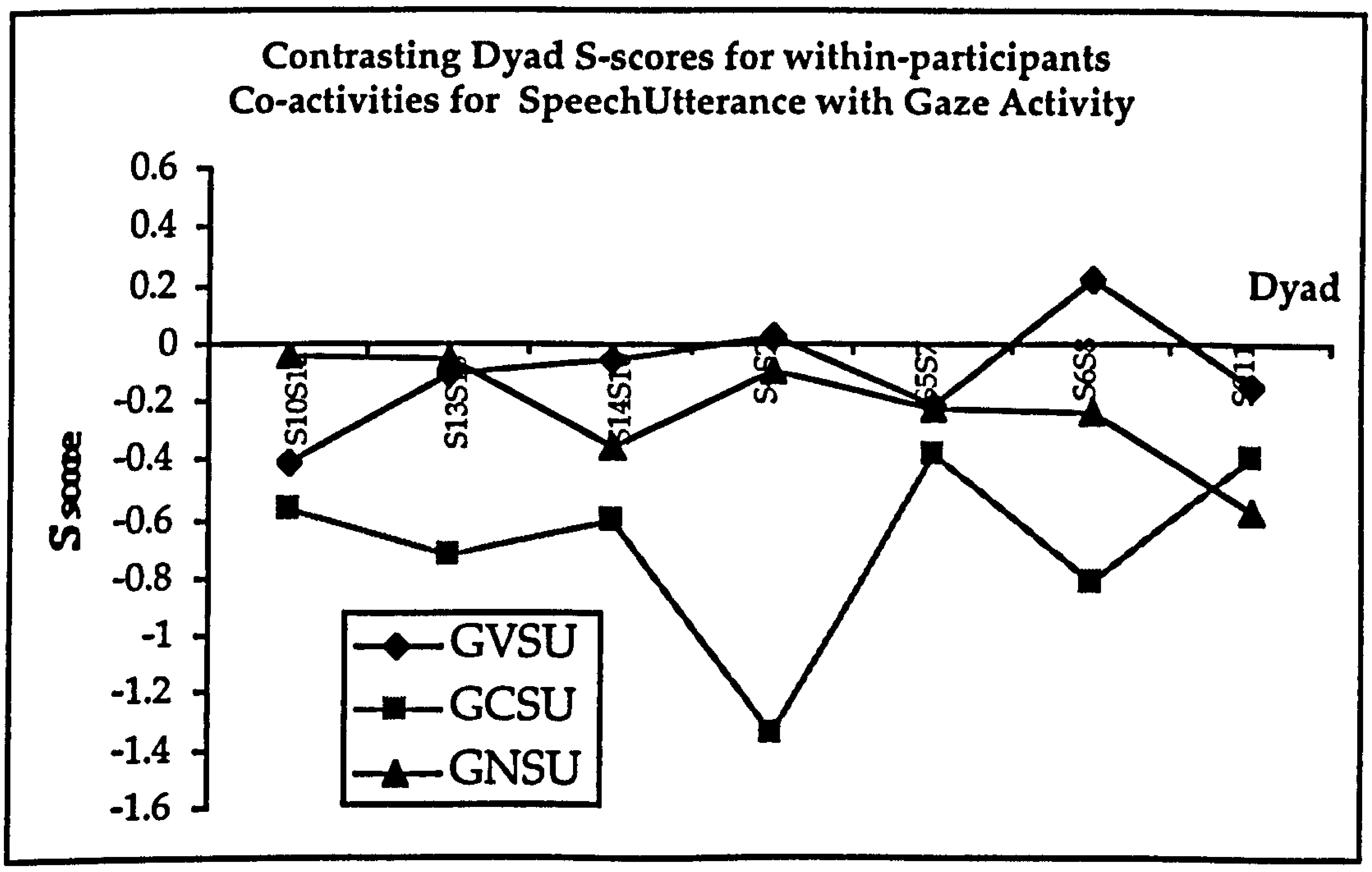


Figure 6.1: Comparing *S* scores for within-participants GVSU, GCSU and GNSU Co - activities

However, a consideration of the GazeVideo - Speech Co-activity with GazeNotes and GazeComputer indicates that their dynamic consultation was quite different. Direct comparison of *S* scores for GVSU with GCSU in a paired-samples t-test reaches significance ($t(6)=3.39; p<.05$) whereas a GVSU and GNSU contrast does not ($t(6)=1.13; p=.303$). Figures 6.1 plots the *S* values obtained for these three Co-activities, for each of the seven dyads, to underline the consistency of this relationship. This suggests that with regard to the interaction dynamic, the role of the computer was markedly distinct.

These findings should be interpreted with caution. On the one hand, it appears that when participants were speaking, the amount they looked towards video monitors was proportionally greater than their amount of looking towards computer VDU or paper notes. On the other, GazeVideo occurred simply at random levels in association with SpeechUtterance whereas GCSU and GNSU were part of an organized system of behaviour. Conservatively, it would seem that something mitigated against GazeComputer Activity whilst participants were in Utterance, rather more than GazeNotes, but there was no evidence of any temporal organization between Speech and GazeVideo Co-activities.

6.3 Between-participants Co-activity data

Data reported here represent combinations of one participant's Activities with Activities of their partner.

Following on from the preceding discussion, the data are examined here for consistent Co-activity organization in the Gaze Activities between individuals or, more properly, *within dyads* . For example, is it true that, on the whole, if a given individual is in the Gaze Activity 'Computer', their conversational partner is more or less likely also to be in the Gaze Activity 'Computer'? Or is there no evidence of a relationship between the Gaze Activities of dyads?

6.4 Between Subjects, Speech and Gaze Activity Sets

The data presented in Table 6.7 show the relationships between the Gaze Activity of dyad members against the Speech Activity of their partners.

Between-participants GazeSpeech Co-activity States	Proportion of Time in State, in Seconds per Minute (SD)		Frequency, in Number of Occurrences per Minute (SD)	
GVSU	5.434	(2.954)	3.98	(1.30)
GCSU	10.198	(4.766)	5.93	(2.33)
GNSU	7.070	(4.744)	5.14	(1.78)
GESU	0.391	(0.343)	0.639	(0.523)
GVSS	5.589	(3.281)	4.80	(1.63)
GCSS	18.471	(8.313)	7.10	(2.20)
GNSS	12.099	(6.728)	6.19	(1.99)
GESS	0.797	(0.648)	0.990	(0.653)

Table 6.7 Between-participants, Gaze Speech Co-activity, standardized time and frequency data (seconds per minute and occurrences per minute)

As with the within-participants Gaze Speech analysis, Co-activities for including the SpeechUtterance are first considered, followed by Gaze with SpeechSilence Co-activities.

	GazeVideo	SpeechUtterance	Observed GVSU (between)	Expected GVSU (between)
Seconds per minute	10.51	22.53	5.32	4.69
(SD)	(5.83)	(5.78)	(2.93)	(3.51)
Proportion of discussion	0.175	0.375	0.089	0.078
(SD)	(0.10)	(0.10)	(0.049)	(0.059)
	"S"=	0.229*	(0.199)	

Table 6.8 Between-participants GazeVideoSpeechUtterance Co-activity Synchronization

Analysis of the GVSU Co-activity shows that participants' GazeVideo Activity was positively Synchronized with the SpeechUtterance Activity of their dyad counterparts (t(6)= 3.05; p<.05).

	GazeComputer	SpeechUtterance	Observed Co-activity GCSU (between)	Expected Co-activity GCSU (between)
Seconds per minute	28.53	22.53	9.71	13.25
(SD)	(11.55)	(5.78)	(4.92)	(6.18)
Proportion of discussion	0.476	0.375	0.162	0.221
(SD)	(0.19)	(0.10)	(0.082)	(0.103)
	"S"=	-0.728**	(0.381)	

Table 6.9: Between-participants Gaze ComputerSpeechUtterance Co-activity Synchronization

Participants tended to look towards their computer VDUs less than chance predicts whilst their conversational partners were in Utterance (t(6)=5.05; p<.01).

	GazeNotes	SpeechUtterance	Observed Co-activity GNSU (between)	Expected Co-activity GNSU (between)
Seconds per minute	19.82	22.53	7.15	9.55
(SD)	(10.96)	(5.78)	(5.04)	(6.86)
Proportion of discussion	0.330	0.375	0.119	0.159
(SD)	(0.18)	(0.10)	(0.084)	(0.114)
	"S"=	-0.408**	(0.249)	

Table 6.10: Between-participants GazeNotesSpeechUtterance Co-activity Synchronization

Again, participants tended not to look at their notes when their partner was speaking as much as chance alone predicts (t(6)=4.34; p<0.01). However, the extent of departure from chance levels of Synchronization, given by *S* score, is very similar to the GazeComputerSpeechUtterance figure (t(6)= 1.40; p=.212). This reveals a slightly different relationship between GazeComputer and GazeNotes with respect to Speech Activity than was found for the within-participants analysis (see above).

	GazeVideo	SpeechSilence	Observed Co-activity GVSS (between)	Expected Co-activity GVSS (between)
Seconds per minute (SD)	10.51 (5.83)	35.47 (10.25)	5.20 (3.25)	7.36 (4.96)
Proportion of discussion (SD)	0.175 (0.10)	0.591 (0.17)	0.087 (0.054)	0.123 (0.083)
	"S"=	-0.513*	(0.513)	

Table 6.11: Between-participants GazeVideoSpeech Silence Co-activity Synchronization

	GazeComputer	SpeechSilence	Observed Co-activity GCSS (between)	Expected Co-activity GCSS (between)
Seconds per minute (SD)	28.53 (11.55)	35.47 (10.25)	18.82 (8.87)	20.58 (10.74)
Proportion of discussion (SD)	0.476 (0.19)	0.591 (0.17)	0.314 (0.148)	0.343 (0.179)
	"S"=	-0.142	(1.663)	

Table 6.12: Between-participants GazeComputerSpeechSilence Co-activity Synchronization

	GazeNotes	SpeechSilence	Observed Co-activity GNSS (between)	Expected Co-activity GNSS (between)
Seconds per minute (SD)	19.82 (10.96)	35.47 (10.25)	12.67 (6.96)	14.83 (9.38)
Proportion of discussion (SD)	0.330 (0.18)	0.591 (0.17)	0.211 (0.116)	0.247 (0.156)
	"S"=	-0.321	(0.750)	

Table 6.13: Between-participants GazeNotesSpeechSilence Co-activity Synchronization

Moving on to a consideration of between-participants SpeechSilence with the three Gaze Activities, a slightly different story emerges as was found for the corresponding within-participants analyses. As Table 6.11 indicates, *S* scores for the co-incidence of GazeVideo for one dyad member and SpeechSilence for the other member were negatively Synchronized ($t(6)=2.65; p<.05$). However, as was found in the within-participants analysis, there was no evidence of any Synchronization in GCSS ($t(6)=0.23; p=.829$) or GNSS ($t(6)=1.13; p=.301$) *S* scores (see Tables 6.12 and 6.13 respectively).

> *1b/* *Between-participants* GazeVideoSpeechUtterance should be positively Synchronized and GazeVideoSpeechSilence should be negatively Synchronized.
>
> *2b/* *Between-participants* Synchronization scores for GazeVideo-Speech Co-activities should differ from both GazeComputer-Speech Co-activities and from GazeNotes-Speech Co-activities.

It was expected that one person's Gaze Activity would vary as a function of another's Speech Activity. This was amply demonstrated. The stronger prediction was that GazeVideo Activity would differ as a function another's Speech Activity in a different way to GazeComputer and GazeNotes. This was indeed found to be the case. Participants' GazeComputer and GazeNotes Activities were negatively Synchronized with their partners' Speech Utterance, whereas between-participants GazeVideoSpeechUtterance Co-activity was positively Synchronized. Furthermore, GVSS was found to be negatively Synchronized whilst there was no Synchronization in evidence for either GCSS or GNSS.

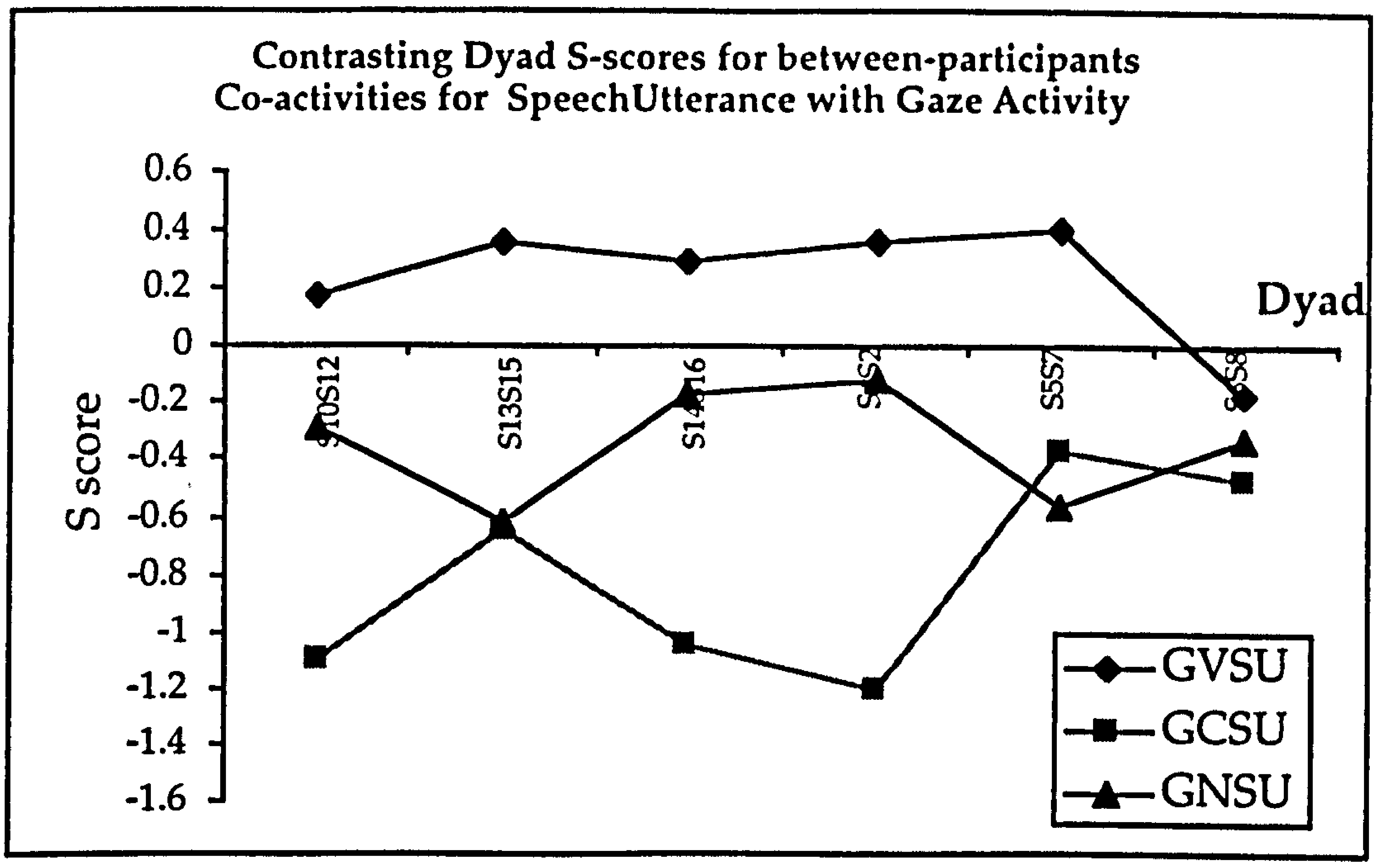


Figure 6.2: Comparing *S* scores for between-participant GVSU, GCSU and GNSU Co-activities

> *3b/ Between-participants'* GazeComputerSpeechUtterance was expected to be more positively Synchronized than GazeNotesSpeechUtterance.

Hypothesis 3b further refined expectations by asserting that between-participants GNSU and GCSU would each differ from GVSU in similar ways and yet still be distinguished from one another. Whilst differences from GVSU are clear and in the same sense (GCSU vs. GVSU, $t(6)=5.38$; $p<.01$, GNSU vs. GVSU, $t(6)=5.08$; $p<.01$), as noted above, the extent to which GCSU and GNSU differ from expected levels is statistically indistinguishable. Figure 6.2 plots the *S* values obtained for these three Co-activities, for each of the seven dyads. The difference in the sense of association between SU and the three Gaze Activities is clear. Remember: the between-participants analysis corresponds to speaker and listener states, rather than where a speaker is also looking. Thus, from a listeners perspective, notes and computer screen were treated dynamically in much the same way. From a speaker's perspective (given by the within-participants analysis, see Figure 6.1), the computer screen were dynamically recruited less than their notes, compared to baseline levels of use.

6.5 Between-participants Gaze Gaze Co-activities

Gaze Gaze Co-activities	Standardized time in secs. per min. (SD)		Standardized frequency, occurrences per min. (SD)	
Both Looking at the video monitor	3.06	(2.46)	2.681	(1.574)
A looking at video monitor, B looking at computer monitor	3.19	(2.34)	2.880	(1.957)
A looking at video monitor, B at written notes	5.21	(4.82)	2.543	(1.376)
A looking at video monitor, B at default	0.449	(0.385)	0.576	(0.479)
A looking at computer monitor, B looking at video monitor	3.28	(2.76)	3.551	(2.312)
Both looking at computer monitor	2.44	(1.67)	5.804	(3.018)
A looking at computer monitor, B looking at written notes	17.83	(8.99)	3.543	(1.001)
A looking at computer monitor, B at default	0.369	(0.190)	0.613	(0.356)
A looking at written notes, B looking at video monitor	6.04	(2.42)	2.180	(0.798)
A looking at written notes, B looking at computer monitor	6.12	(5.25)	3.249	(1.898)
Both looking at written notes	10.60	(8.79)	3.689	(1.755)
A looking at written notes, B at default	0.420	(0.710)	0.457	(0.457)
A looking at default, B looking at video monitor	0.426	(0.415)	0.457	(0.317)
A looking at default, B looking at computer monitor	0.491	(0.513)	0.613	(0.749)
A looking at default, B looking at written notes	0.336	(0.349)	0.582	(0.398)
Both looking at default	0.060	(0.032)	0.122	(0.049)

Table 6.14: Between-participicants, Gaze Gaze Co-activity. standardized time and frequency data (secs. per min. and occurrences per min.)

In principle, there being three a priori meaningful Gaze Activities per Set, there are nine potential Gaze Co-activity relationships to examine (see Table 6.14). Of these nine, three Co-activity state pairs are logically equivalent i.e. Gaze Computer A with Gaze Video B is conceptually the same Co-activity as Gaze Computer B with Gaze Video A. Thus, six potential relationships present themselves: Video Video (VV), Computer Computer(CC), Notes Notes(NN), Video Computer(VC), Video Notes(VN), and Computer Notes(CN).

	Gaze Video	Gaze Video	Observed GVGV Co-activity	Expected GVGV Co-activity
Seconds per minute	10.51	10.51	3.06	2.19
(SD)	(5.83)	(5.83)	(2.36)	(1.88)
Proportion of discussion	0.175	0.175	0.051	0.036
(SD)	(0.10)	(0.10)	(0.039)	(0.031)
	"S"=	0.272*	(0.239)	

Table 6.15: Between-participants GazeVideo GazeVideo Co-activity Synchronization

The Co-activity GazeVideoGazeVideo (i.e. when both participants looked at the video monitor at the same time) occurred for significantly more of the discussion period than chance predicts (t(6)=3.01; p<.05).

	Gaze Video	Gaze Computer	Observed GVGC Co-activity	Expected GVGC Co-activity
Seconds per minute	10.51	28.53	4.20	6.04
(SD)	(5.83)	(11.55)	(3.79)	(5.43)
Proportion of discussion	0.175	0.476	0.070	0.101
(SD)	(0.10)	(0.19)	(0.063)	(0.091)
	"S"=	-0.469**	(0.302	

Table 6.16: Between-participants GazeVideo Gaze Computer Co-activity Synchronization

Participants spent significantly less time in the state GVGC than would be expected on the basis of Simple looking levels (t(6)=4.11;p<.01). That is, whilst one participant looked towards the video link, the other was less likely to be looking at their computer VDU than chance would predict.

	Gaze Video	Gaze Notes	Observed GVGN Co-activity	Expected GVGN Co-activity
Seconds per minute	10.51	19.82	2.86	3.79
(SD)	(5.83)	(10.96)	(2.23)	(2.81)
Proportion of discussion	0.175	0.330	0.048	0.063
(SD)	(0.10)	(0.18)	(0.037)	(0.047)
	"S"=	-0.213	(0.283)	

Table 6.17: Between-participants GazeVideo Gaze Notes Co-activity Synchronization

The Co-activity GVGN showed no evidence of mutual sensitivity, *S* scores not differing significantly from zero (t(6)=1.99; p=.094).

	Gaze Computer	Gaze Computer	Observed GCGC Co-activity	Expected GCGC Co-activity
Seconds per minute	28.53	28.53	17.83	18.31
(SD)	(11.55)	(11.55)	(8.64)	(11.65)
Proportion of discussion	0.476	0.476	0.297	0.305
(SD)	(0.19)	(0.19)	(0.144)	(0.194)
	"S"=	-0.168	(0.802)	

Table 6.18: Between-participants Gaze Computer GazeComputer Co-activity Synchronization

Mutual computer-directed gaze neither more nor less likely than chance predicts (t(6)=0.55; p=.559).

	Gaze Computer	Gaze Notes	Observed GCGN Co-activity	Expected GCGN Co-activity
Seconds per minute	28.53	19.82	6.08	10.13
(SD)	(11.55)	(10.96)	(3.93)	(5.60)
Proportion of discussion	0.476	0.330	0.101	0.169
(SD)	(0.19)	(0.18)	(0.065)	(0.093)
"S"=		-0.776**	(0.241)	

Table 6.19: Between-participants Gaze Computer Gaze Notes Co-activity Synchronization

Table 6.19 shows a very marked departure from the expected model of Co-activity for GazeComputerGazeNotes, in that it was highly unlikely that one of the dyad would be looking at their paper notes whilst the other was looking at the computer (t(6)=8.53; p<.001).

	Gaze Notes	Gaze Notes	Observed GNGN Co-activity	Expected GNGN Co-activity
Seconds per minute (SD)	19.82 (10.96)	19.82 (10.96)	10.60 (8.45)	10.03 (9.36)
Proportion of discussion (SD)	0.330 (0.18)	0.330 (0.18)	0.177 (0.141)	0.167 (0.156)
	"S"=	0.201	(0.373)	

Table 6.20: Between-participants Gaze Notes Gaze Notes Co-activity Synchronization

Both dyad members looked at their notes much as chance predicts (t(6)=1.42; p=.205).

4/ GazeVideoGazeVideo Co-activity was expected to exceed chance levels (i.e. be positively Synchronized).

Hypothesis 4 suggested the GVGV Co-activity would be above chance levels and this was found to be the case. A mean *S* of approaching 0.3, representing the Z score for the expected Co-activity distribution, represents a modest difference and yet from a theoretical perspective is important. As discussed in Chapter One, gaze coordination has been at the centre of non-verbal behaviour studies for decades. It suggests that *S* is measuring coordination of Activity in a sensitive and appropriate manner.

5a/ GazeVideoGazeComputer and GazeVideoGazeNotes should both be negatively Synchronized.

GVGC Co-activity and GVGN Co-activity were expected to be substantially different from GVGV, and rather similar to each other. GVGV contrasts strongly with GVGC (t(6)=5.18; p<.01) and also with GVGN (t(6)=2.47; p<.05) lending support to this prediction. Furthermore, whereas GVGV was positively Synchronized, GVGC Co-activity negative Synchronization meant that when one participant looked towards their video monitor, they were significantly less likely than chance to see their counterpart inspecting the computer information. Although there was no clear evidence of negative Synchronization for GVGN Co-activity, comparing GVGC and GVGN directly showed no difference in Synchronization score (t(6)=1.61; p=.158). Figure 6.3 plots *S* scores for GVGV with GVGC and GVGN to clarify this relationship.

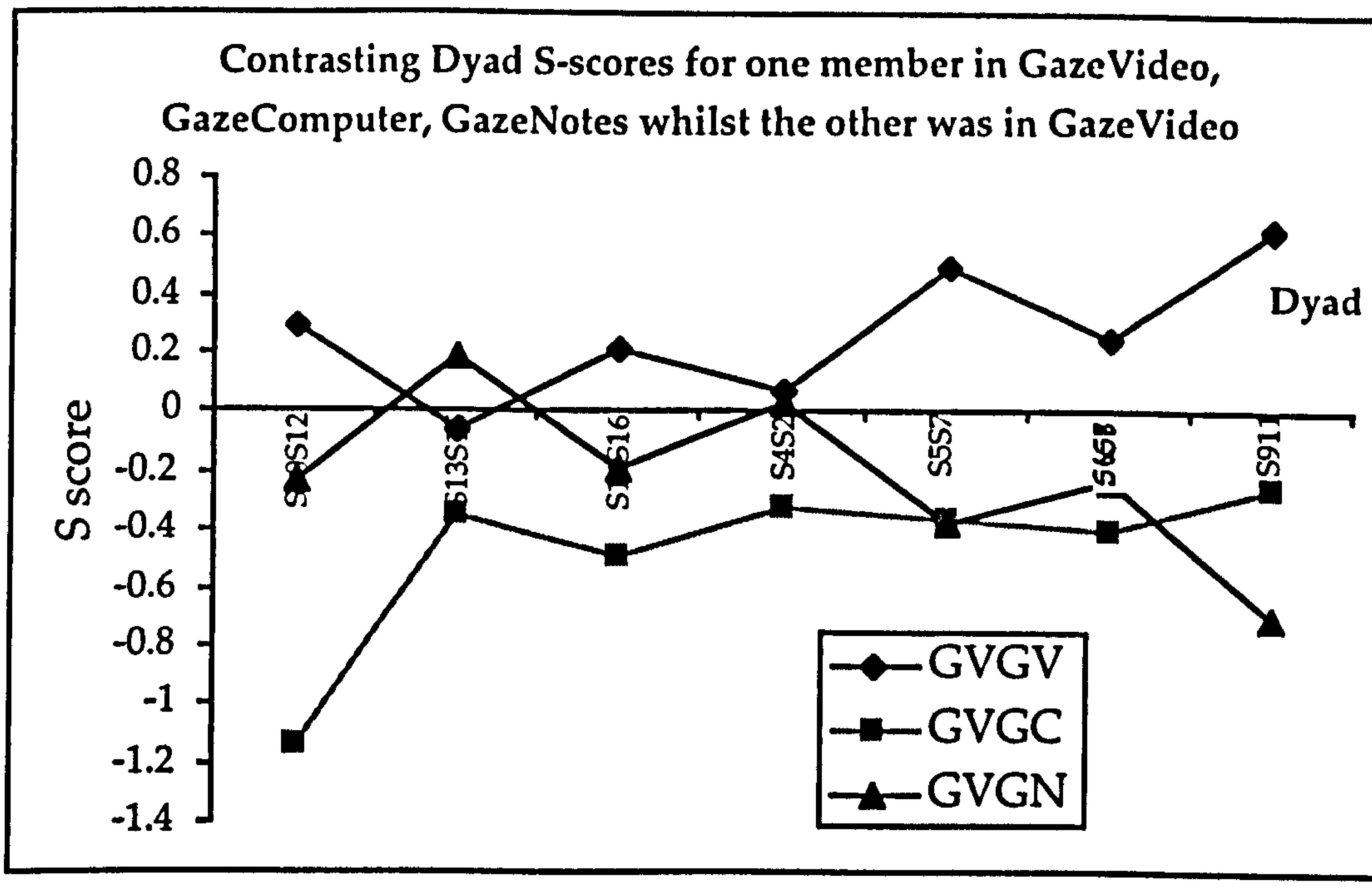


Figure 6.3: ***S*** **scores for Co-activities involving GazeVideo**

Figure 6.4 presents dyad *S* scores for GCGC, GCGN and GNGN Co-activity, as evidence of information-resource based coordination.

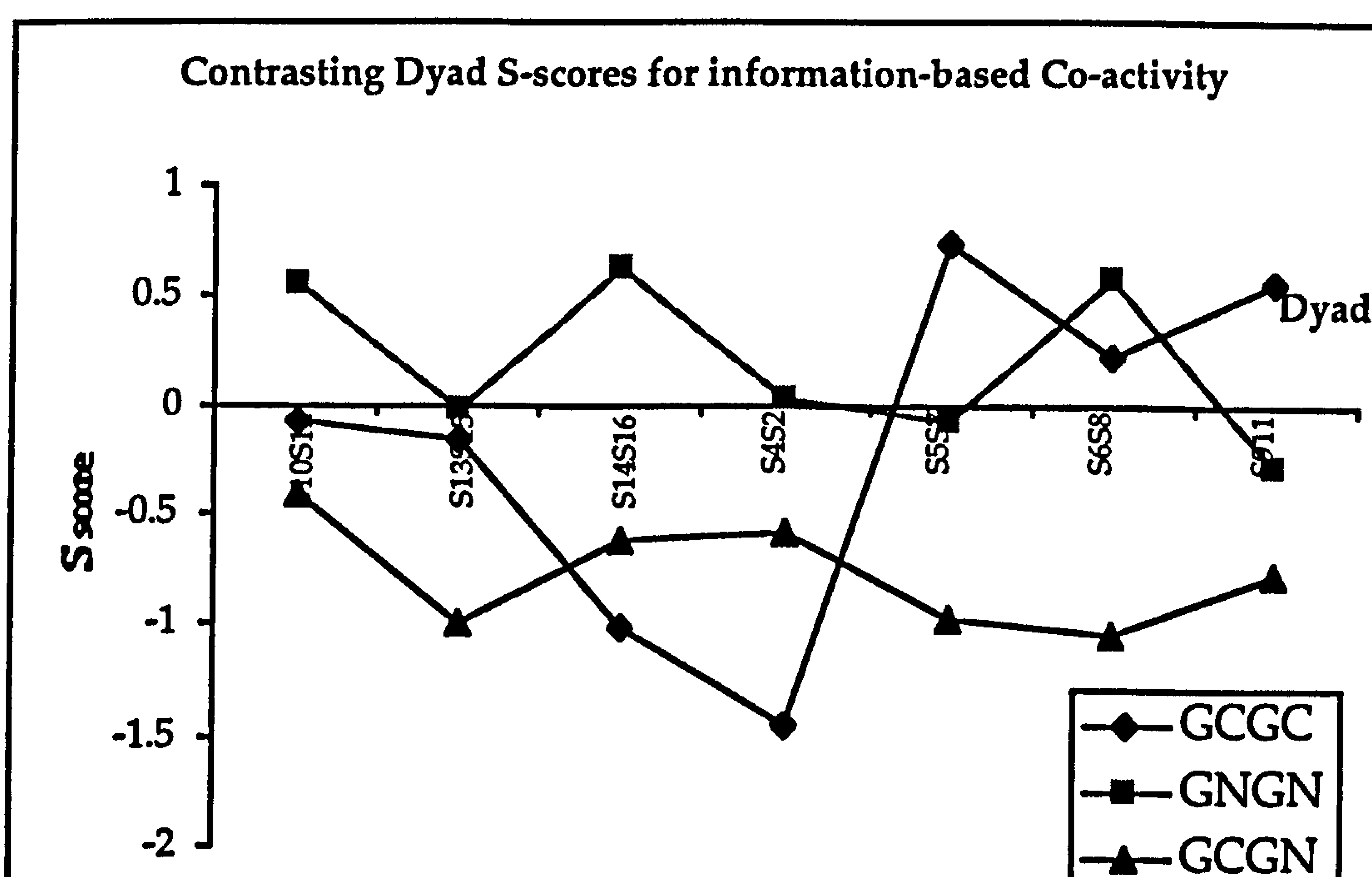


Figure 6.4: Comparing *S* scores across dyads for GCGC, GCGN and GNGN Co-activities.

5b/ GazeComputerGazeComputer and GazeNotesGazeNotes should both be positively Synchronized.

5b/ GazeComputerGazeComputer should be more positively Synchronized than GazeNotesGazeNotes.

It was expected that GCGC and GNGN should both be positively Synchronized. The former did not differ significantly from zero but the picture is somewhat confused, with a high degree of variability across dyads (see Figure 6.4). As corollary of Hypotheses 5a and b, one might expect GCGN Co-activity to be negatively Synchronized. Indeed, the data support this proposition: GCGN Co-activity Synchronization was notable and negative.

These data might be read as indicative of an uncertainty towards the status of the "live" computer link. The striking variability by dyad in GCGC Synchronization suggests that for some its synchronous scrolling and gesture-by-mouse made it rather more of communication coordination device. Conceiveably, for others it was in effect more of separately consulted information resource. The variability is all the more intriguing here given the information- similar GNGN Co-activity. Both require reading and both contain case detail for the arguments participants were developing. Furthermore, the stability of GCGN negative Synchronization suggests that the 10% of discussion in this state served a relatively fixed function(s) in all cases. GCGC, at 30%, made up by far the largest proportion of any Gaze Gaze Co-activity but is comparable to the GNGN 20% figure, underlining the difference in interaction dynamic.

Much has been made in this thesis of the vital relationship between people, tools and place as a system of Activity. Here, we see that Gaze coordination is not achieved simply through turn exchange but more importantly and more significantly, on these findings, through tools. That the computer link offered textual information casts it in a similar role to paper materials and yet the nature of the information as being overtly shared and role as a pointing facility suggested that it would be treated differently. These conceptual differences have been realized in objective measures of coordination.

6.6 Between Subjects, Speech Activity Sets

This section describes something of spoken dynamic of the interactions included in this case study. From Table 6.21 it may be seen that around three quarters of the time was spent with one partner in SpeechUtterance whilst the other was in SpeechSilence. Approximately a quarter of the time on average was spent with neither partner in Utterance. Inspection of the videotaped record shows that much of this time was spent appraising computer or papers notes, showing that the discussions required a combination of unearthing relevant information and constructing arguments. That there was as much as 5% of the time in overlapping speech suggests that discussions were fluid and easy to maintain despite the potential for interference. Again, inspection of the video record reinforces this view.

Interpretation of these figures requires careful examination of the criteria for coding sounds as an instance of utterance. For coding purposes, no distinction was made between backchannel utterance ('uh huh', 'yeah' etc.) and fuller forms of speech ('... but (0.5) candidate niiine'). The defining characteristic of vocal activity was that it should represent some kind of unitary vocal contribution to the talk exchanged between the individuals. So the data here included as simultaneous speech allows for occasions of turn transition, forced interruption, supporting speech and so on.

The four states possible with the Speech Activity Set definition amount to three meaningful Co-activities.

States, described by Activity conjunctions	Proportion of Time in State, in Seconds per Minute (SD)		Frequency, in Number of Occurrences per Minute (SD)	
aSUbSU	2.43	(1.06)	5.65	(2.07)
aSUbSS	18.32	(4.50)	9.32	(3.16)
aSSbSU	21.87	(6.63)	9.29	(1.37)
aSSbSS	17.38	(3.63)	12.91	(2.38)

Table 6.21: Between-participants Speech Activity, standardized time and frequency data (seconds per minute and number of occurrences per minute)

The first, SUSU, represents simultaneous speech, including backchannel vocalizations, floor-taking attempts and overlapping speech. The second is indicative of time with one participant as speaker and the other as listener (SUSS). The third, SSSS, shows how much time was spent with between turns at talk and reading or operating the computer.

	Speech Utterance	Speech Utterance	Observed SUSU Co-activity	Expected SUSU Co-activity
Seconds per minute (SD)	22.53 (5.78)	22.53 (5.78)	2.43 (1.06)	10.13 (2.39)
Proportion of discussion (SD)	0.375 (0.10)	0.375 (0.10)	0.041 (0.018)	0.169 (0.040)
	"S"=	-1.165	(0.245)	

Table 6.22: Between-participants Speech Utterance Speech Utterance Co-activity Synchronization

These figures are unsurprising from a theoretical viewpoint. From a methodological viewpoint, they demonstrate that the method of coding and recording employed has some face validity. As one would expect, simultaneous speech, as the Co-activity SUSU, occurred at far lower than chance levels ($t(6)=12.58$; $p<.001$). That there was enough simultaneous speech suggests that the interactions were not constrained formally in the way Chapter One described.

	Speech Utterance	Speech Silence	Observed SUSS Co-activity	Expected SUSS Co-activity
Seconds per minute (SD)	22.53 (5.78)	35.47 (10.25)	20.09 (5.75)	17.52 (7.71)
Proportion of discussion (SD)	0.375 (0.10)	0.591 (0.17)	0.335 (0.096)	0.292 (0.129)
	"S"=	0.671	(0.862)	

Table 6.23: Between-participants Speech Utterance Speech Silence Co-activity Synchronization

SUSS Co-activity coordination approaches but does not reach significance ($t(6)=2.06$; $p=.085$). This may seem surprising, suggesting at first sight that there is little coordination of SpeechUtterance and SpeechSilence Activity *during periods of spoken exchange*. However, that is not what these data represent. The simple proportion of time in SpeechSilence (see Table 6.23), upon which the *S* score is built, includes also the substantial part of interaction when both participants are engaged with information materials without talk. In another kind of interaction, perhaps with no reference to such resources, one might have stronger a priori grounds for expecting a substantial positive Synchronization score.

	Speech Silence	Speech Silence	Observed SSSS Co-activity	Expected SSSS Co-activity
Seconds per minute (SD)	35.47 (10.25)	35.47 (10.25)	17.38 (3.49)	24.93 (8.79)
Proportion of discussion (SD)	0.591 (0.17)	0.591 (0.17)	0.290 (0.058)	0.415 (0.147)
	"S"=	-1.033	(4.039)	

Table 6.24: Between-participants Speech Silence Speech Silence Co-activity Synchronization

Again, interpretation of the data presented in Table 6.24 is rather muddied by SpeechSilence including participants in listener status as well as out-of-talk reading or navigation. *S* scores show that the observed proportion of time in SSSS Co-activity was very similar to the chance expected figure (t(6)=0.65; p=.540).

6.7 Discussion

Chapters Four and Six have both reported on aspects of interactive behaviour evident during a set of video-mediated interactions, designed to put an emphasis on information-based opinion-forming and consensus. The findings are summarized and commented on here. Given that the video link was the sole source of both the reception of visual information about the distal partner and transmission to the partner, it was anticipated that looking behaviour targeted in this direction would demonstrate a relationship with the speech of the participants. The task was to use the information provided, on paper and electronically, to decide collaboratively the relative merits of ten applications for financial assistance. The three Gaze Activities might all serve as information resources. It was suggested that the GazeVideo Activity serves additionally and exclusively as a transmission medium. However, the electronic information source could be used as a form of gestural device. This was because, in addition to linked scrolling, the software supported joint mouse activity as pointing devices. The mouse pointers served no function, when outside of the scroll bar screen region, other than to indicate to the distal party points of interest or significance in the documents provided.

If it be the case that each source of visual information is of equivalent utility for the purposes of the participants, it might be expected that they would be attended to in much the same sort of way. However, quite clearly the two involving written information require something very different of the participants on this very basis. Reading is a linguistic activity and it takes a finite amount of time to do. Interpretation of video images may be instantaneous or require close scrutiny, with or independently of talk. Furthermore, the electronic source offers a *shared* repository of (in this scenario) undisputed fact; the paper version, a *personal* re-casting of the information into a form suitable for the building of arguments and making cases. The visual allows access to information concerning the interpersonal and interparty factors that might well include commitment, concern, frustration, disdain, humour, confusion, confidence in assertion, etc.

In the first place, the majority of time is spent in the GazeComputer Activity, at 30 seconds per minute, followed by GazeNotes, at 20 s/min, and then around 10 s/min in GazeVideo. These differences could be accounted for just by casting the former as a reading activity and the latter as turn-coordination and paralinguistic exchange. Certainly, standardized frequency data show that each Gaze Activity is engaged as often as the next, approximately 5 per minute, and duration data agree with standardized time.

Between-subjects SpeechSpeech Co-activities suggested that the discussions were fluid and consisted of dialogue for about two thirds of the interaction period. There was thus a considerable amount of time outside of spoken exchange, predominately dealing with text material. As predicted, there is a strong relationship between Gaze Activity and Speech Activity. In other words, Speech substantially contextualizes the use of Gaze, at least as operationalized in this study.

At around 4 s/min, within-participants GazeVideoSpeechUtterance extended for no more nor less than chance levels would predict. In contrast, within-participants GazeComputerSpeechUtterance, at 10 s/min, and GazeNotesSpeechUtterance, at 7s/min, take up significantly less time than would be expected by chance. Between-participants Gaze-Speech Co-activity analysis showed both that GVSU (5 s/min) was positively Synchronized and GVSS (also about 5 s/min) was negatively Synchronized. Considered alongside the foregoing discussion of SpeechSilence states, it indicates that participants with listener status were predisposed to look at the video link and, during periods of mutual Silence, participants were less likely to look towards the video link. This finding broadly agrees with Kendon's (1990) observations on gaze speech dynamics: speakers tend not to look at listeners, listeners tend to look at speakers. It differs in suggesting that the relationship is not quite reciprocal.

Whereas within-participants GCSU was found to be more negatively Synchronized that GNSU, between-participants analysis of GCSU and GNSU found a similar extent of negative Synchronization. The SpeechSilence Gaze Co-activities reveal no corresponding pattern of GC and GN Synchronization but must include a substantial period where both participants are in SpeechSilence. It seems that Computer and Notes are of less interest during periods of dialogue than one would predict on the basis of the Simple Activity data, and Computer especially so for speakers. Alternatively, one might conclude that from a listener's perspective, there is rather more value in GazeNotes and GazeComputer than from a speaker's perspective.

That the overall levels of GazeComputer Activity were so much higher than GazeNotes should not be forgotten. The power of the *S* measure is that it takes these differences into account in order to interpret where and how interactive behaviours relate to one another, not to make blanket judgements of their value. This result is a statement that GCSU diverged from chance levels rather more than GNSU for speakers, and to similar extents for listeners. It might then be taken to suggest that GCSU was rather more heavily invested during periods of mutual Silence.

This is not the same thing as asserting that GazeComputer and GazeNotes were of little interest whilst talk was in progress. It suggests that GazeVideo was of greater interest to participants during spoken exchanges than when mutual silence obtained. This supposition begs the question of why, then, within-participants GazeVideoSpeechUtterance shows no evidence of positive Synchronization. Wagner et al. (1983) argued that individual looking levels may be maintained with redundancy, such that a required degree of mutual gaze is achieved. They suggest that there are 'tonic' and 'phasic' aspects of gaze coordination and tonic gaze levels may simply be set to result in suitable phasic coordination. In other words, mutual gaze may not need to be finely coordinated, as it may be sufficiently valuable to people in conversation that they are prepared to bear a cost of redundant looking in order to achieve it. On the within-participants data, taken to refer to speaker Activity (i.e.

discounting within-participants SS states), the 'cost' appears to have been born by some erosion of both GazeComputer and GazeNotes Activity. Certainly, Wagner et al.'s proposed mechanism is feasible for any Co-activity. In terms of Gaze-Speech coordination, the evidence presented here supports a rather stronger statement about exploitation of the video link during interaction.

Mutual gaze, represented by the GVGV Co-activity, showed good evidence of positive Synchronization, whereas GVGC and GVGN were negatively Synchronized. GCGN showed a clear negative Synchronization, suggesting that whilst one person was looking at their notes it was less likely than chance that the other would be looking at their computer VDU. A "mutual Computer" (GCGC) state did not Synchronize, although there was evidently some difference in the way computers were used by different dyads. "Mutual Notes" (GNGN) did not Synchronize either.

6.8 Conclusion

The overwhelming majority of the discussion time involved participants inspecting their computer screens and looking at one another rather less. That they did so with similar frequency suggests that whatever value was to be had from each of these Activities needed to be extracted on just as regular a basis. Examination of the contingencies between Gaze Activity and Speech Activity served to expose any organization between them. Given previous chapters' emphasis on integration of Activities, this was seen to be of particular interest. Where such organization was found, it implicates the constituents as a vital part of the interactive Activity system.

An Activity Set Analysis research agenda was set for an investigation of VMI. It has demonstrated some relationships between behaviours acknowledged to be significant for communication processes. Chapter One suggested that gaze distribution including a video link ought to show evidence of temporal organization with speech activity. This was found to be the case. That speaker and especially listener states seemed biased towards GazeVideo in competition with GazeComputer and GazeNotes lends support to arguments in favour of including video as a resource for electronically-mediated interactions. Equally, mutual gaze, in so far as the Co-activity state GazeVideoGazeVideo can be taken to represent it, was found to Synchronize positively rather than merely to reflect a random conjunction of automatic behaviour. In the context of a paucity of evidence to suggest that video has any role in mediated interaction, these are highly important findings.

These conclusions are strictly limited to interactions including reference materials. They are also subject to the important criticism that any of the Gaze-involved Co-activity coordination might in principle have been achieved through talk. The following chapter contrasts alternative configurations of video link with a variety of supporting materials, to expose how these Gaze-related Co-activities fare. If they are achieved purely through talk then changes to video arrangements should have no effect on their sense or level.

7 An experiment comparing effects of video image size and experimental task on Gaze distribution and Speech Activity

Chapter summary

The final piece of work reported in this thesis attempts to demonstrate the value of the approach developed to this point. Activity Set Analysis is applied to a comparative VMI experiment, involving 48 members of the general public. Communications are mediated via alternative video displays with separate, high-quality audio and also a shared artefact. Task dimensions are introduced to contextualize the Gaze and Speech Activity Sets described earlier. Three kinds of VMI support are compared: a full-motion relatively large video image, a full-motion small video image, and a still large video image, all with the same full-duplex audio. Tasks included "getting to know" the other person and a simple card game. "Getting to know you" was prompted by an on-screen questionnaire and the card game required one dyad member to look at each of ten playing cards in turn and either to name the card they saw or to attempt to deceive their remote partner by reading out a different card name. Results are discussed with reference to their consistency with the patterns of video-mediated interactive behaviour found in Chapter Six. They are further considered in the light of the discussion of VMI in Chapter One.

7.1 Introduction

This Chapter reports an experiment looking at two characteristics of video image provided as a picture-in-picture facility on a personal computer screen, as a resource to support remote pc-based communications. Chapter Six reported some significant evidence of visual behaviour coordination. It was found that participants in a dyadic discussion via VMI technology looked at one another together (a mutual gaze analogue) at greater than chance levels. As listeners, they also looked at their partner at above chance levels. Whether listener or speaker, accompanying computer and paper materials were the subject of gaze at less than chance levels.

Commercial providers of videoconferencing over local and wide area networks are now a firmly established part of the multimedia computing industry (Butters et al., 1994). Commercial desktop VMI systems differ from the configuration used in the Chapter Six case study in a number of ways. Typically a camera is mounted over or under a computer display screen and an image of the remote person(s) is displayed in an on-screen window. The quality of these video images varies greatly: even with modern high-capacity communication networks and advanced compression algorithms it is only possible to transmit a fraction of the information required for a fluid, high quality, video image. For this reason the images displayed are generally: (a) small in visual angle, (b) coarse grain in pixels and (c) updated infrequently. The case study configuration used a life-size image at high resolution and updated above flicker frequency. Furthermore, whereas the physical layout of the case study area allowed easy discrimination of a partner's point of gaze focus, the typical arrangement of desktop videoconferencing is of material in in one image plane and close in space. Camera position has a loose correspondence with viewer position, in relation to GUI window. The case study configuration minimized disparity between point of view and apparent position. It used very large displays and ensured that the camera was mounted as low as possible over the display, a fish-eye lens actually overlapping the edge of the picture area, and on the same vertical line as the image of the remote party. This Chapter asks whether the findings of Chapter Six replicate in a configuration comparable to desktop VMI and, is so, whether manipulations of desktop-analogue VMI device affect such Sychronization.

Image refresh rate and resolution are two of the major parameters for VMI "quality of service". As discussed in Chapter One, there is very little evidence to suggest that the "talking heads" model of VMI, almost universally adopted by these commercial systems, offers advantages over plain audio mediation. There is no definitive account of the minimum rate of change to an image necessary for a video link yet to have any value. Indeed it is difficult even to begin framing the minimum refresh-rate question, given that variety of grounds for supposing vision matters in communication, on the one hand, and the paucity of evidence in its support on the other. Speech reading would ideally require transfer rate of approximately phonemic intervals but a certain amount of information drop-out would be tolerable. Visual non-verbal cues vary in their rate and duration of expression so some would suffer from slow refresh rate more than others. A video image of a colleague is typically taken to be useful in a supportive role to speech in the process of communication, with paralinguistic cues. It may be that infrequently updated video images undermine any cue value in this sense. Similarly, at some level of image size or resolution, some or all of the paralinguistic content of non-verbal behaviour will be lost, whether by not being encoded in the video signal, displayed on the VDU, or by not being detectable to the observer as a function of viewing distance, task demands, or other display characteristics.

Nevertheless, a video image may be associated with less dynamic aspects of non-verbal communication, such as intimacy, as a function of interpersonal distance, and attributions from clothing or posture (Argyle, 1988). As Pagani and Mackay (1993) suggested in a discussion of a long-term "virtual office share" by video link, having a live image of ones conversational partner may serve to combat a sense of isolation. Resolution and image size are closely related, as a matter of determining image content. The size of a video image may influence the extent to which it can play a part in a communications process, by the extent to which it places demands on visual attention. A high resolution source displayed as a small image will occupy a small degree of an observer's visual angle and hence strain interpretation. A large image display of a low resolution source will not necessarily compensate of for the essential lack of fine-grain information, although possibly enhancing the sense of immediacy so generated. Heath and Luff (1992b) have observed that even quite gross movements conveyed by camera and video monitor, intended to attract attention, can fail to be registered by their intended recipient. Although Heath and Luff's example involves a failed attempt to instigate conversation, it is conceivable that visual behaviour during the course of an interaction will similarly fail, despite "having the ear" of the recipient.

This study poses two pragmatic questions relating to physical requirements for VMI display. The first relates to a simple design constraint: a simple matter of 'screen real-estate'. As discussed in Chapter One, working communication frequently involves co-reference to objects or documents. In a networked computer context, one might imagine that a display screen is the access point both for VMI and the relevant artefact to be shared during conversation. Although VDU technology is advancing apace, the high graphical content of most current computer interfaces, increasingly an intrinsic part of the office worker's environment, shall inevitably lead to pressures on the available space and hence some form of competition for on-screen display area.

The second issue informing the design of this study concerns the value of motion as a separate matter from the sense of the person given by an image. Poor movement may be worse than no movement, if the computational resources devoted to image refreshing are channelled instead into relatively few very high resolution images, or even only one 'scene setting' image. Thus, a 'visually augmented' full duplex audio communications configuration was employed as a basic control condition, by presenting a single captured still video image (the SVC condition). The conversational partners in this control condition saw a static image of one another for the duration of the task. In this way, the dynamics of a video image are partially dissociated from some generalized interpersonal effect of visual information of this kind. For the other conditions, interactions were mediated by two live-feed analogue video signals, presented either in a window occupying an appreciable segment of a typical computer VDU (the dynamic large video window - DVL - condition), the other in a somewhat smaller window (the dynamic small video window - DVS - condition. The maximum image size was determined by the screen area remaining after allowing for an accompanying graphical window used to display textual information as a shared artefact to the experimental task.

Two tasks were contrasted, a prompted introductory discussion, "get to know you", and a simple version of guessing game using playing-cards. The former was chosen as an informal computer-based "ice-breaking" activity, presenting an opportunity to talk for varying amounts around some simple lifestyle and personal history questions. The latter was chosen as potentially a more regulated activity, with no computer-based requirement besides the availability of an image of the remote participant, and to explicitly emphasize the content of the remote image. A set of playing cards were given to one participant, with instructions either to tell their remote partner the real suit and number of the card or to "lie" by reading out a different suit and number. As the card-holding participant's speech was effectively scripted by the contents of the card, information for the guesser was extremely limited. It was expected that the participants would both spend rather more time looking at the video as a function of this pseudo-deception aspect of the task.

7.1.1 Measures and Hypotheses

Measures of video-mediated interactive behaviour were of the form described in Chapter Four. Gaze and speech metrics, much as described in the case study, were used with Action Recorder (see Chapter Three) in order to derive mean state duration, proportion of time and frequency of engagement in state data. In this case, four binary Activity Sets were defined:

- North's speech: North Utterance
  North Silence
- South's speech: South Utterance
  South Silence

The case study data reported in Chapter Four indicated the "elsewhere" Gaze Activity to be of vanishing importance. For the purposes of the current experiment, it is more useful to conceive of this Activity as Gaze Materials or, in context, the shared software window in the first task phase and the cards or response sheet for North and South respectively in the second phase. Informally, inspection of these video recordings suggest that most of the Gaze Elsewhere Activity was spent, as suggested, focused on these materials. The quality of video recordings precluded observers from specifically and reliably distinguishing between gaze at materials and other aspects of the graphical user interface (GUI) or other immediate foreground foci. A gaze tracker would be required for such an analysis. Consequently, the hypotheses developed below primarily concern the use of a GUI video window, rather than of the other materials.

- North's gaze: North looking towards video image of partner
North looking elsewhere
- South's gaze: South looking towards video image of partner
South looking elsewhere

These Activity Sets were operationalized within *task phase*, differentiating behaviour for the first and second experimental tasks. One may conceive of these as constituting a super-ordinate Activity Set, framing the context of the Activity Sets operating within it, at higher pace and intrinsically connected with the state of dyadic Activity. However, at a data level, as there was only one instance of each task phase Activity, they are more easily seen simply as levels of an independent variable. This independent variable is named "Task Phase" to reflect that it combines order with different forms of collaborative effort:

- Task Phase: Introductory discussion
Card guessing game

7.1.2 Hypotheses

It was expected that task demands would have a fundamental effect on the "formality" of speech patterns, given by frequency and duration data, with the card task associated with more formality than the discussion task. As formality is usually described as longer utterances with fewer floor changes, duration of SpeechUtterance were expected to increase and frequency decrease leading to no prediction about proportion of time in state. However, as a social task, the "get to know you" phase might be expected to benefit from the social nuances of the video image. In so far as a larger image size could be said to correspond to closer interpersonal distance, one might expect some benefit in the overall amount of this social talk generated, over and above the formality of the talk given by duration and frequency data. So formality of Speech Activity in turn was expected to cross with experimental condition. The dynamic conditions, DVL and DVS, were expected to be associated with less formal speech than the control, SVC.. These differences stand to be considerably greater for the discussion than for the cards task.

H1: Speech Activity proportion, duration and frequency vary with task.

H2: Speech Activity proportion, duration and frequency vary by condition.

H3: Speech Activity interacts with task and experimental condition.

Gaze Activity was expected to be more "task oriented" in the SVC condition, evidenced by more time spent looking at the shared GUI window, overall and by frequency (and, consequently[36], by duration). Less Gaze Activity was also expected to be directed toward the video window in DVS condition compared to DVL, as a poorer resource. Frequency was not expected to differ, as until an image has been inspected its value or otherwise is indeterminate. Duration and hence proportion of time in state were expected to differ, as a function of the value of the image once inspected. No prediction was set on the direction of this change - it could increase as greater investment might be required for a given return, or decrease, as a perception of the limited value or greater effort potential pre-empts any such investment.

These differences were expected to be greater for the informal discussion than for the card task, as the latter was expected to "force up" the perceived value of the video image to extract whatever non-verbal information might be obtained to aid in the discrimination of honest from dishonest statements about card identity.

H4: Gaze Activity duration and proportion of time in state vary with task.

H5: Gaze Activity duration and proportion of time in state vary across experimental conditions.

H6: Gaze Activity frequency, proportion and duration of time in state interact with task and across experimental conditions.

Experimental role was expected to have a pronounced effect on Speech and Gaze Activities, regardless of task phase. Role is considered to be an

[36]As Chapter Four discussed, the frequency, proportion of time in state and duration in state are mathematically related. It nevertheless can make sense to consider all three from a theoretical viewpoint.

additional aspect of contextualization for the interactive behaviour under study in this experiment. In this respect, it fulfils the same investigative function as the two task phases. Dyad roles were identified with particular locations, North always occupying an anteroom and South always in a screened-off booth in an exhibition hall (see Figure 7.1). More interestingly, the "North" subject was responsible for managing the interaction through the materials provided. North had sole control of the shared software in the first phase, requiring a level of coordination with the system that might be expected to engender closer attention to the computer screen than to the video. In the second phase, North had a number of test cards to work through. For both phases, North was explicitly instructed to pace the task with regard to South's readiness and comprehension. NorthGazeVideo was thus expected to be engaged for shorter duration and for less time overall than SouthGazeVideo. However, as a part of maintaining the necessary coordination, in the DVL and DVS conditions, GazeVideo was anticipated to be more frequent for North than for South. Following on from North's requirement to secure agreement from South prior to advancing computer or cards, North was generally expected to engage in more Utterance than South. Speech Utterance Activity was anticipated to differ on the duration and proportion but not frequency measures, reflecting the differential explanatory load on North's speech content and, at the same time, interactive requirement for agreement.

H7: Duration, frequency and proportion of time in SpeechUtterance will be greater for North than for South.

H8: North's GazeVideo will be shorter and extend for less time overall than South but will occur more frequently.

Co-activities were analysed on the model presented in Chapters 5 and 6. Chapter Six reported finding some significant evidence of visual behaviour coordination. On this basis, Gaze Co-activities were expected to show more organization in the dynamic conditions than in the fixed-picture control, in as much as such coordination might be motivated by looking at the other person in relation to their interactive activities. Gaze Co-activity coordination is an interesting case here, as it could be argued of the findings reported in Chapter Six that all Gaze-related coordination was achieved through talk. As tasks, roles and audio facilities are identical in all three experimental conditions here, were this to have been true one would expect just as much Gaze coordination regardless of video configuration. Five hypotheses are set for Co-activities with Gaze components, as follows.

If the largest manageable window size is of markedly more benefit than a video window of nominal size, DVL would be associated with more Gaze Activity coordination than DVS. On grounds that DVL should have fewer problems interpreting the video image, one might expect it to have a rather more meaningful role to play in the interaction. The SVC image shows a static remote partner, unlikely to be of the same kind of value as the dynamic images. This thesis has contended that the meaningfulness of interactive resources should be reflected in their temporal organization with respect to other on-going behaviours. Chapter Six found GazeVideoGazeVideo to be positively Synchronized in this way and also that GazeVideo was negatively Synchronized with GazeComputer. It was additionally found that GazeComputerGazeComputer was negatively Synchronized. While in the current experiment GazeVideo maps on to the Chapter Six GazeVideo fairly well, GazeElsewhere is only a loose equivalent to GazeComputer. Methodological difficulties preclude a meaningful GazeElsewhereGazeElsewhere analysis in this case but GazeVideoGazeElsewhere, conceived as essentially a GazeVideo driven Activity, has rather more promise.

H9: GazeVideoGazeVideo and GazeVideoGazeElsewhere Co-activities for DVL and DVS conditions shall exceed chance levels whereas SVC shall not.

H10: GazeVideoGazeVideo and GazeVideoGazeElsewhere Co-activities will show greater organization in the DVL than in the DVS condition.

Gaze-Speech Co-activities between participants were also expected to show more evidence of coordination in the dynamic than control conditions. The case study Co-activity data (see Chapter Six) showed that SpeechUtterance was coordinated with remote party Gaze Activity. This coordination might have been achieved via the video link or substantially via the verbal content of the exchanges. The current experiment should serve to test this possibility by contrasting identical audio facilities with and without a video component. Greater coordination was predicted in DVL than in DVS on the basis of improved visual cue salience.

H11: GazeSpeech Co-activity *between* subjects was expected to exceed chance levels in the dynamic but not control conditions.

H12: GazeSpeech Co-activity *between* subjects was expected to be greater in DVL than DVS conditions.

As, somewhat surprisingly, no evidence of coordination of Gaze and Speech Activities within subjects was forthcoming in the case study Co-activity data (see Chapter Six), no expectation was set on such behaviour. However, as an unlikely result, it deserves to be tested again:

H13: GazeSpeech Co-activity *within* subjects shall exceed chance levels.

Anticipating task effects on Co-activity organization, SpeechSpeech Co-activity was expected to be greater for the more formal card task than for the discussion task, in so far as it is indicative of the amount of overlapping SpeechUtterance. Similarly, the formality dimension of communication media was expected to find expression in SpeechSpeech Co-activity organization, with dynamic conditions showing less organization than control and also differing from one another.

H14: SpeechSpeech Co-activity will show less "formality" in the discussion than the cards task.

H15: SpeechSpeech Co-activity will show less "formality" in the DVL and DVS conditions than in the SVC condition.

H16: SpeechSpeech Co-activity will show less "formality" in the DVL than the DVS conditions.

H17: SpeechSpeech Co-activity organization will interact with size and dynamism of video image and experimental task.

7.2 Method

7.2.1 Design

Three alternative configurations of VMI equipment were independently compared, supporting dyads in carrying out two successive tasks. Dyad role was differentiated by allocating responsibility for progress through the set tasks to only one of the dyad members. The task variable contrasts an initial discussion with a card game. So role and task phase were within-dyads and communications configuration between dyads factors, examined in a two by two by three factorial design respectively. Configuration conditions compare the provision of either a large dynamic, a small dynamic, or still a video window for the session. Motion and size were not fully crossed and task order was not controlled for pragmatic reasons, as later discussed.

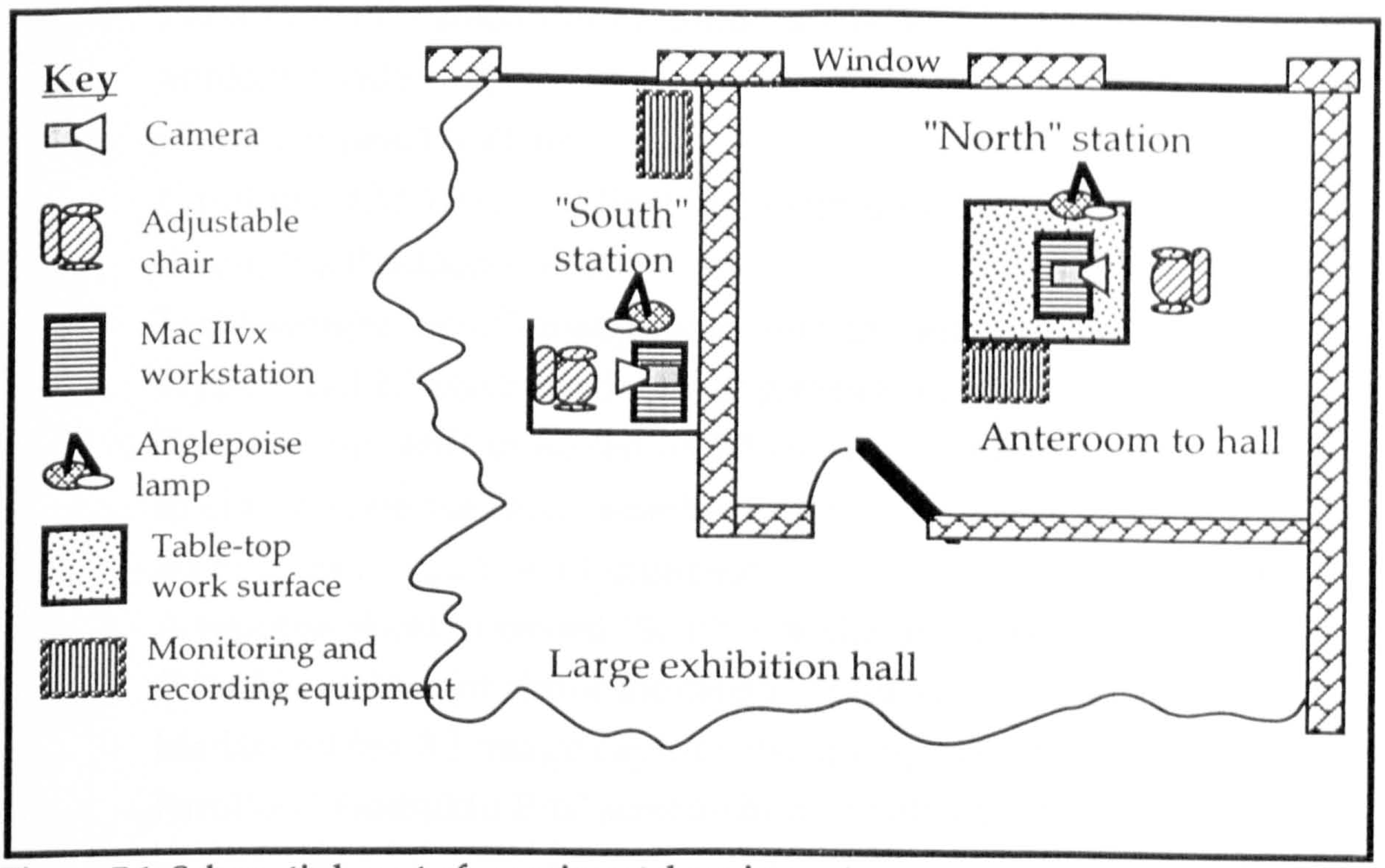


Figure 7.1: Schematic layout of experimental equipment

7.2.2 Subjects

Forty eight members of the general public were recruited at a public science exhibition. Ages ranged from approximately 10 to 65 years. Dyads were variously formed of people who had not previously met and prior acquaintances[37]. No payment was offered and consent was obtained, from parents where minors were concerned, for taking part in a short videophone experiment and for videotaping their conversations.

7.2.3 Materials

Two computer workstations were set up to allow dyads to communicate by audio and video, including VDU camera, microphone and audio headset. One workstation was labelled "North" and the other "South". The "North" workstation was additionally equipped with a computer mouse so that only north was able to interact with a HyperCard instruction stack. Changes to the display were relayed to south using screen sharing software running over a dedicated Ethernet connection.

The workstations comprised:

> Apple IIvx computer equipped with RasterOps 24MXTV video capture cards, driving Apple 16" computer VDU with analogue video feed from other workstation.

---

[37]Prior familiarity and gender of dyads, including mixed and same sex pairs, was closely balanced by condition, although group composition by condition was not identical.

Panasonic G1 camcorder mounted above VDU, supplying a PAL analogue video signal.
Height adjustable chair.
CamLink EM-100 omnidirectional clip-on microphones, amplified through a Panasonic VTR.
Light-weight "open" headphones with integral volume control.
HyperCard 2.1 stack for displaying experimental instructions.
Ten playing cards mounted on B5 size card with instructions to "North" to either name the card's number and suit, or to "lie, say the *different suit and number*"
A reponse sheet to record "South"s decision on the truth or falsity of "North"s statement about the card's identity.
Mediagrabber 3.1 image capture and manipulation software.
Farallon "Timbuktu Pro" screen-sharing software for Apple Macintosh.

For recording purposes: Panasonic AVE55 Video mixer and Panasonic Character Generator.

7.2.4 Procedure

Participants for the experiment were recruited as they visited a science exhibition. One of two VMI computer workstations was positioned in the main display hall behind 2m high screens and the other in an adjacent anteroom. Interested visitors to the exhibition were asked to participate in a short videophone experiment. The "South" workstation in the main hall was demonstrated with the aid of an experimenter at the "North" workstation (see Figure 7.1). Participants were informed that they were to be video recorded as a part of an investigation into videophone communication and their consent obtained. One of the pair was then ushered to the anteroom by the second experimenter whilst his or her experimental partner was taken behind the screens by the other experimenter.

Full-motion VMI involved no perceptible refresh problem as the image was supplied as a video signal directly to a special-purpose video-capture card installed in the computer. The video-capture card drove the display, including a video image from the remote workstation controlled by specialist software, in a re-sizeable graphical window, and an Apple HyperCard stack with instructions for the joint tasks they were to carry out. Audio was provided by tie-pin microphones clipped to each participant's clothing feeding light-weight headphones via audio amplifiers.

Participants were seated before the workstations chair height adjusted so that they were in front of the VDU at a distance of about 70 cm, and so that the display was at head height. In each case, a camera was positioned over each computer monitor to provide a head-and-shoulders view to the other workstation. Participants were informed that it was important to reach agreement on each of the items in the instruction stack before moving on, and that changes made by north took a few seconds to be received by South's screen. Participants were asked to "say hello" to one another in order to ensure the proper functioning of the sound channel. Ambient noise was highly variable in the exhibition hall so sound levels were checked and adjusted as necessary, enabling conversation at normal volumes. North received brief training on how to use the computer mouse to navigate the HyperCard instruction stack.

They were then informed that all the instructions they required were to be found on the computer display, and asked to begin. The tasks consisted of filling in a short screen-based questionnaire on their joint interests followed by a card game in which south attempted to deceive north.

In the full-motion, large-image (DVL) condition there was a 103mm tall x 140 mm wide image of the other participant aligned to the top left of the Apple 16" colour monitor. For the full-motion, small-image (DVS) condition, the video image was again aligned to the top left-hand corner of the display but measured 48 mm high by 65mm wide. Since both full-motion video conditions were sourced from the same PAL signal, the resolution was in-principle comparable. However, the VDU was configured to run at 640 x 820 pixels, so whereas the DVL condition could take advantage of the PAL 400-line TV image, the DVS condition was restricted to about half this vertical resolution. Horizontal resolution is governed by scanning rate on TV signals and thus did not differ markedly. The static, large image (SVC) condition was also set up with the same equipment. The video feed was frozen at the start of the joint task as a digital still, the subjects being asked to smile at the camera as if for a photograph. As with the DVL condition, the image was 103mm high and 140mm wide with comparable resolution.

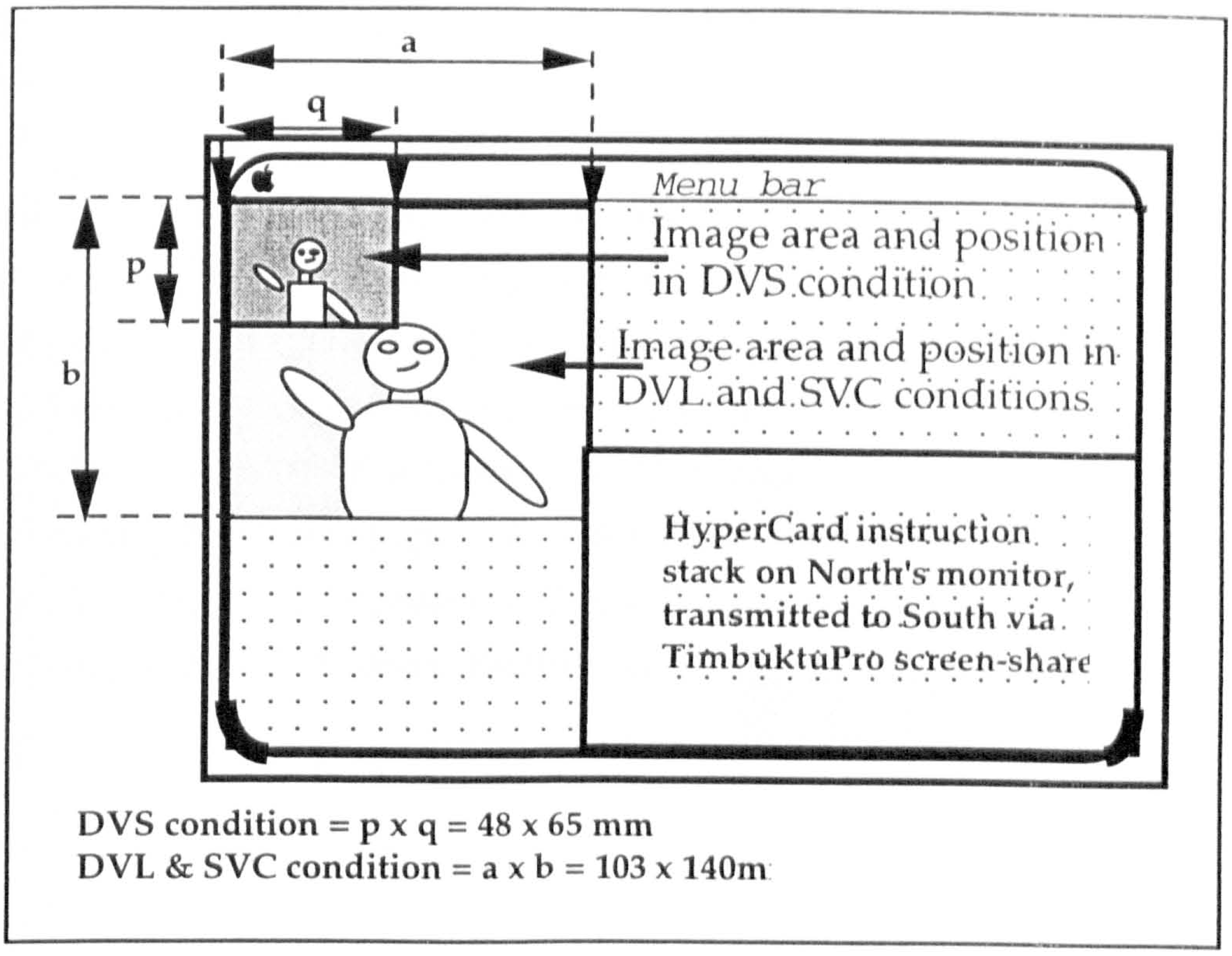


Figure 7.2: Video image size and position

7.3 Video analysis with Action Recorder

During the experiment videotapes were made by tapping the signals from the cameras used to provide the video link. There was thus one tape for each participant in a pair, containing exactly the images viewed by each on their respective monitors during their discussions. To make it possible to relate the data on the two tapes the same centisecond clock was mixed onto both at the time of recording.

The interactive behaviour of dyads was subject to an Activity Set Analysis of gaze deployment and speech activity. Action Recorder was configured for four Activity Sets, as described above. Videotaped recordings of interactions were then rated by a trained analyst. The analyst viewed each tape twice, recording Speech and Gaze Activity separately for each subject. The analyst was blind to the experimental condition in which the recording was made, each recording being cued by an experimenter. The time stamped key presses obtained in this way were transformed into state duration by Action Recorder, then aggregated using the SPSS statistical package to produce the data for analyses of Speech and Gaze Activity within dyads for the two experimental phases.

Inter-rater reliability was assessed by correlating SPSS Aggregated data, as previously described. In general, agreement for both Simple standardized time and frequency and Contingent data was found to be of the same very high order as reported in Chapter Four, Pearson's r invariably exceeding 0.8

and frequently better than 0.9. Mean duration data derived from Action Recorder ratings were again in good agreement, although at the lesser levels. Two of the eight dyads in the SVC condition and one from the DVS condition were excluded from all the analyses reported below, as metrics of their Gaze Activity were highly atypical of the rest of the sample. For the SVC dyads in the Cards task, mean duration of gaze at video link was 0.68 seconds (0.44 standard deviation) whereas the comparative figures for the excluded dyads were 19.7 seconds and 43.8 seconds. The DVS dyad was excluded as one of its member's mean duration of gaze of at the video link during the discussion task was 16.5 seconds, compared with the DVS group mean of 1.06 seconds (standard deviation 0.37). Some further comment on this is made in the discussion.

7.4 Overview of results

Activity Set data are reported in two sections: first Simple Activity analyses and then contingent, Co-activity analyses. Simple Activity analyses include duration and standardized time and frequency measures, as described in Chapter Four. Co-activities are subject to Synchronization Analysis on the *S* measure of departure from expected coincidence, as described in Chapter Five and exemplified in Chapter Six. *S* is treated in two distinct ways, one asking about the consistency and absolute polarity of Synchronization of a Co-activity, the other about the relative Synchronization of pairs of Co-activities. The former tests groups of *S* scores against zero in a paired-samples t-test. The latter compares S scores of nominated Co-activities from different sources, for example North versus South GazeVideoGazeVideo, with a split-plot ANOVA procedure.

Statistical tests are reported with two levels of significance, where differences estimated to have a chance probability, p<.05 are described as "significant" and p<.01 as "highly significant". These levels are indicated in tables by a single asterisk (*) and double asterisk (**) respectively. Where these levels are not reached, the calculated p value is given. As the nature of this thesis is exploratory, and yet a conservative attitude towards quantitative data is adopted by the author, it is left to the reader to decide the likelihood of Type II errors. This is especially relevant for the *S* analyses, as the *S* score is based on a considerably more conservative estimate of chance coincidence than earlier treatments in the literature.

Before embarking on these analyses, it is appropriate to make some brief mention of time taken to complete the experimental tasks. No predictions were made on this matter, given the lack of any evidence to suggest performance outcomes differ as a function of video configuration differences[38]. Indeed, at approximately 7.5 minutes, dyads took much the same time to go through the experiment, regardless of video condition (no main effect of configuration, F (2,18) = 0.43, p=.656). Task phases happened to take about the same time to complete (F (1,18)=2.80; p=.112) and also did not interact with condition (F(2, 18)=1.74; p=.204).

[38]The only other obvious measure of performance here is to do with the cards task. However, for the purposes of this experiment, guessing accuracy in the card task was entirely incidental and in any case performance was found to be at roughly chance levels for all conditions.

7.5 Simple Activity Set Analysis

This section reports data on how experimental condition, role and task phase are related to the two Activity Sets defined for this experiment. Speech Activity data is present for SpeechUtterance only as, for a binary Activity Set, data for the other Activity will agree with the first. Similarly, only GazeVideo Activity data is reported in this section.

For both SpeechUtterance and GazeVideo data, a prose summary is first given of findings for the three base Activity Set measures (mean duration, standardized frequency and standardized time). Data for each measure is then presented on successive pages, one per measure, including tabulated mean scores with standard deviations, ANOVA tables and bar graphs. In each case data is cast three-dimensionally, as condition by task by role.

7.5.1 SpeechUtterance Activity

The duration of SpeechUtterance Activity, corresponding crudely to turn length, was slightly longer in the Discussion task than in the cards task, thus H1 receives some support. A multivariate split-plot ANOVA showed that although the difference is of the order of only 0.1 seconds, it reaches significance as a main effect of experimental task (see Table 7.2). There is no evidence of any general difference in Utterance length between the three conditions, contrary to the H2 prediction. Similarly, there is no suggestion of a difference in SpeechUtterance duration by role, undermining in part H8.

Figure 7.3 suggests that there is a complex relationship between role, condition and task phase. North subjects in the DVL condition were in Utterance for longer their South partners, but only in the discussion task phase and not the cards task. This difference of role by task phase is not in evidence in the DVS and SVC conditions (three-way interaction between experimental condition, role and task phase - see Tables 7.1 and 7.2). The anticipated two-way interaction between task and experimental condition (H3) is not in evidence but may be masked by the role variable, given the higher-order interaction. Although the magnitude of this difference, again, is small at around 0.25 seconds, it should be noted that this difference is against mean SpeechUtterance duration of the order of one second. These very short Utterance duration suggest that spoken exchanges were informal, as suggested in H1.

The cards task phase was associated with a faster rate of SpeechUtterance than the discussion task phase, reaching significance with an ANOVA procedure (main effect of experimental task, see Table 7.4), adding weight to

H1. North roles engaged in SpeechUtterance more frequently than South, lending some support to H8. SpeechUtterance rates by role were of the order of 15 per minute, or on average either North or South beginning an utterance every two seconds. Given the duration data above, it is clear that talk was almost continuous over the course of the interaction. There is no suggestion of any influence of experimental condition on rate of Activity engagement, either in general (H2) or by interacting with task (H3).

Turning to the standardized time measure, H8 correctly anticipates North speaking for longer overall than South (Tables 7.5 and 7.6). No more time was spent in SpeechUtterance during the discussion than during the cards task phase, failing to support H2. However, as with the duration data, the relationship between task, role and condition is complex (three-way ANOVA interaction). From Figure 7.5, it seems that the discussion phase is generally associated with more SpeechUtterance for North than for the cards phase, but markedly more so in the DVL condition than either DVS or SVC. North's level of SpeechUtterance is much the same as in the cards phase however in the cards phase, all experimental conditions are associated with as much talk from North. Curiously, South seems to speak correspondingly less in these circumstances. That SpeechUtterance levels were way off ceiling, the overall amount of SpeechUtterance (summing North and South data) being 34 seconds per minute, suggests DVL puts South at a disadvantage to the point where they talk on average for half as much as North. A similar asymmetry for the cards task is less surprising as it seems to be independent of experimental condition, on the one hand, and the overt task demands are such that South simply has less to say than North here. It is quite striking to note how very similar these SpeechUtterance data for the DVS and SVC conditions.

In sum, task demands affected SpeechUtterance duration and frequency but overall amount of talk was relatively stable in the two task phases. Experimental role was associated with frequency and overall time differences in SpeechUtterance but Utterance length was relatively invariant. Whilst SpeechUtterance time did not vary directly with the VMI configuration, both SpeechUtterance mean duration and the overall amount of time spent in SpeechUtterance state are in a complex relationship with image type, role and task type.

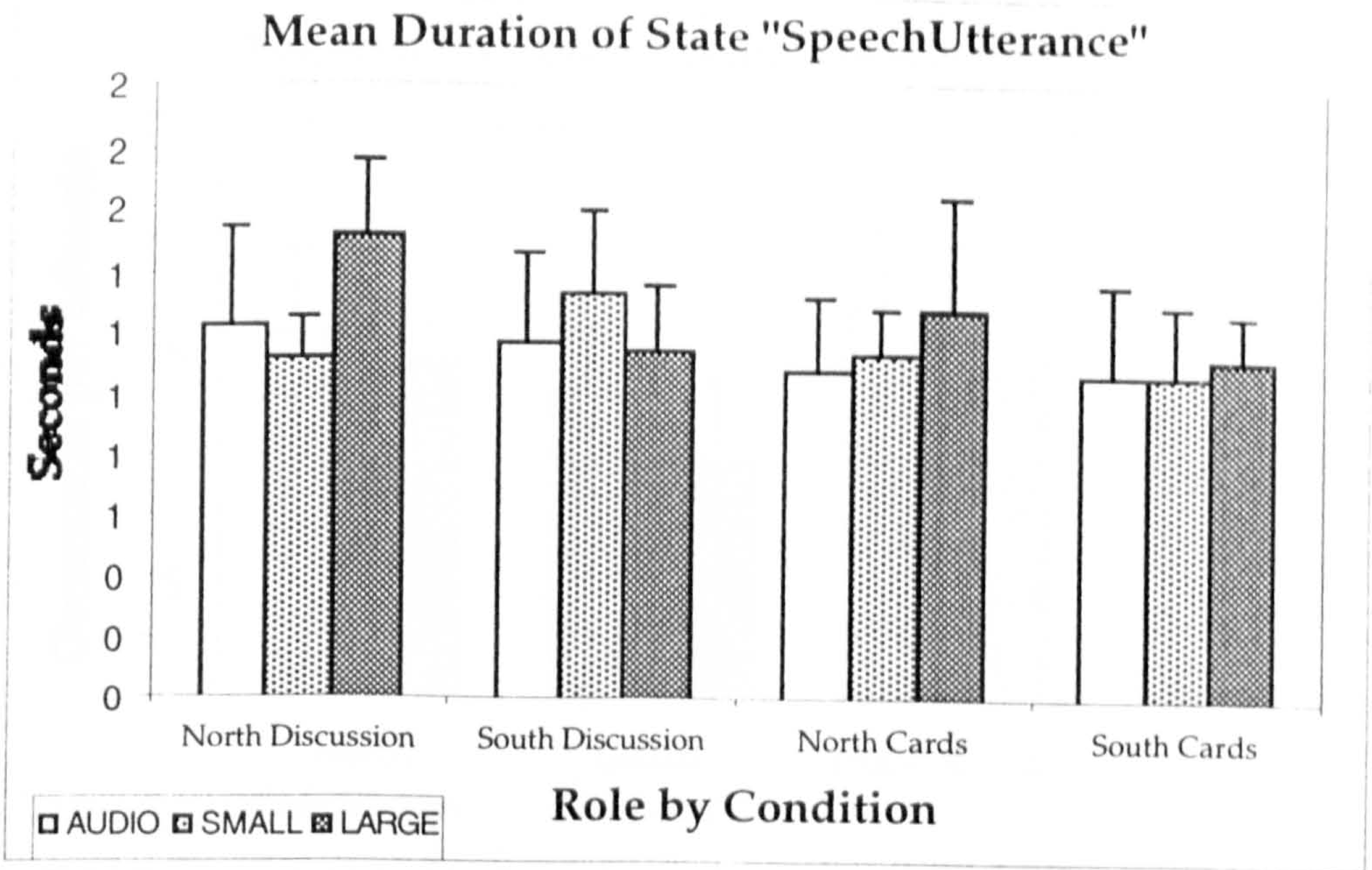


Figure 7.3: Mean duration of Gaze Activity contrasting experimental role and condition

Mean Utterance Duration (secs.)	North Intro.		South Intro.		North Cards		South Cards	
SVC	1.23	(0.33)	1.18	(0.30)	1.09	(0.25)	1.08	(0.30)
DVL	1.52	(0.36)	1.16	(0.21)	1.28	(0.24)	1.12	(0.15)
DVS	1.13	(0.13)	1.34	(0.26)	1.14	(0.14)	1.08	(0.23)
Grand Total	1.31	(0.33)	1.22	(0.26)	1.18	(0.22)	1.10	(0.22)

Table 7.1: SpeechUtterance Activity mean duration

MEAN DURATION OF "SPEECH UTTERANCE" ACTIVITY

Source of Variation	SS	DF	MS	F	Sig of F
Tests of Between-Subjects Effects.					
WITHIN+RESIDUAL	2.05	18	.11		
COND	.27	2	.13	1.17	.332
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN+RESIDUAL	1.81	18	.10		
ROLE	.11	1	.11	1.04	.320
COND BY ROLE	.45	2	.23	2.25	.135
Tests involving 'TASK' Within-Subject Effect.					
WITHIN+RESIDUAL	.39	18	.02		
TASK	.34	1	.34	15.48	.001**
COND BY TASK	.00	2	.00	.04	.964
Tests involving 'ROLE BY TASK' Within-Subject Effect.					
WITHIN+RESIDUAL	.21	18	.01		
ROLE BY TASK	.00	1	.00	.03	.855
COND BY ROLE BY TASK	.22	2	.11	9.44	.002**

Table 7.2: ANOVA table for SpeechUtterance mean duration data, including role and task as within-dyads and experimental condition as between dyad factors

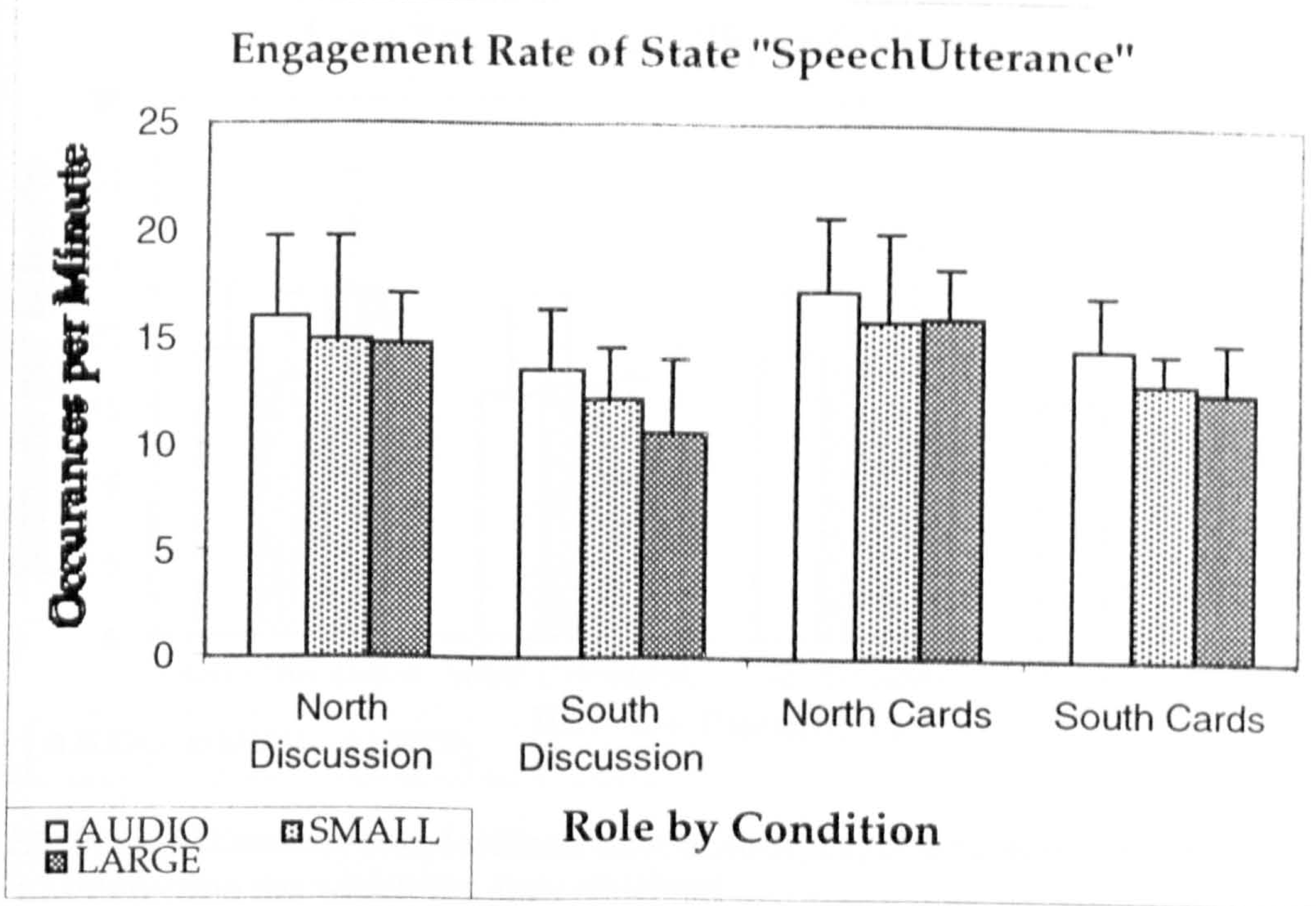


Figure 7.4: SpeechUtterance Activity state engagement frequency

Utterance Rate (n/min.)	North Intro.		South Intro.		North Cards		South Cards	
SVC	16.08	(3.84)	13.70	(2.79)	17.39	(3.28)	14.75	(2.39)
DVL	14.94	(2.28)	10.70	(3.43)	16.05	(2.31)	12.68	(2.31)
DVS	14.99	(4.89)	12.29	(2.39)	15.90	(4.21)	13.16	(1.38)
Grand Total	15.28	(3.60)	12.09	(3.06)	16.38	(3.21)	13.43	(2.15)

Table 7.3: Frequency data for SpeechUtterance Activity

```
RATE OF "SPEECH UTTERANCE" ACTIVITY ENGAGEMENT
Tests of Between-Subjects Effects
 Source of Variation          SS       DF        MS          F   Sig of F
 WITHIN+RESIDUAL            371.71     18      20.65
 COND                        50.67      2      25.34       1.23       .317
Tests involving 'ROLE' Within-Subject Effect.
 Source of Variation          SS       DF        MS          F   Sig of F
 WITHIN+RESIDUAL            213.41     18      11.86
 ROLE                       188.08      1     188.08      15.86       .001**
 COND BY ROLE                 7.12      2       3.56        .30       .744
Tests involving 'TASK' Within-Subject Effect.
 Source of Variation          SS       DF        MS          F   Sig of F
 WITHIN+RESIDUAL             70.38     18       3.91
 TASK                        29.95      1      29.95       7.66       .013**
 COND BY TASK                 1.62      2        .81        .21       .814
Tests involving 'ROLE BY TASK' Within-Subject Effect.
 Source of Variation          SS       DF        MS          F   Sig of F
 WITHIN+RESIDUAL             28.52     18       1.58
 ROLE BY TASK                  .18      1        .18        .12       .738
 COND BY ROLE BY TASK         1.31      2        .66        .41       .667
```

Table 7.4: ANOVA table for SpeechUtterance frequency data, including role and task as within-dyads and experimental condition as between dyad factors

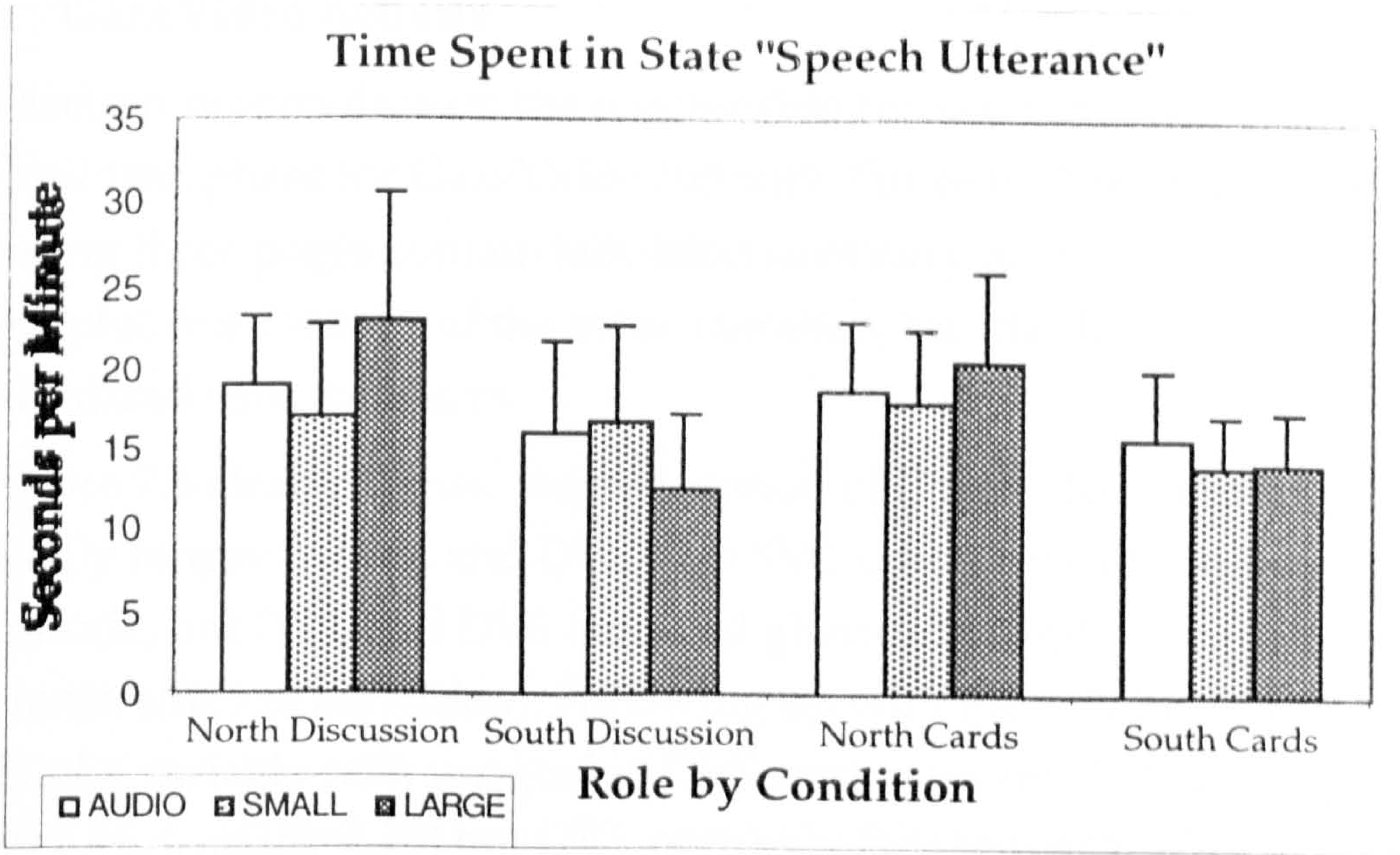


Figure 7.5: SpeechUtterance standardized time in state, expressed as seconds per minute of interaction for which the state obtained

Utterance standardized time	North Discussion		South Discussion		North Cards		South Cards	
SVC	18.94	(4.13)	16.09	(5.65)	18.66	(4.34)	15.68	(4.23)
DVL	22.99	(7.54)	12.55	(4.81)	20.55	(5.36)	14.27	(3.30)
DVS	16.96	(5.83)	16.74	(5.86)	17.96	(4.48)	14.13	(3.03)
Overall	19.82	(6.44)	14.96	(5.49)	19.15	(4.70)	14.63	(3.39)

Table 7.5: SpeechUtterance Activity standardized time in state (seconds per minute, SD)

STANDARDIZED TIME IN SPEECHUTTERANCE ACTIVITY					
Tests of Between-Subjects Effects.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	826.83	18	45.93		
COND	20.83	2	10.42	.23	.799
Tests involving 'TASKPHASE' Within-Subject Effect.					
WITHIN+RESIDUAL	198.26	18	11.01		
TASKPHASE	5.23	1	5.23	.47	.500
COND BY TASKPHASE	.96	2	.48	.04	.957
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN+RESIDUAL	770.09	18	42.78		
ROLE	406.88	1	406.88	9.51	.006**
COND BY ROLE	176.38	2	88.19	2.06	.156
Tests involving 'TASKPHASE BY ROLE' Within-Subject Effect.					
WITHIN+RESIDUAL	55.60	18	3.09		
TASKPHASE BY ROLE	.11	1	.11	.03	.854
COND BY TASKPHASE BY ROLE	56.72	2	28.36	9.18	.002**

Table 7.6: ANOVA table for SpeechUtterance standardized time data, including role and task as within-dyads and experimental condition as between-dyad factors

7.5.2 GazeVideo Activity

This section reports data on the relationship between experimental condition, role and task phase for GazeVideo Activity. For ease of reference, the following three pages contain tabulated summary data, ANOVA tables and bar graphs, one for each of the mean duration, standardized frequency and standardized time measures.

As Figure 7.6 clearly shows, mean duration of GazeVideo Activity was markedly longer in DVL and DVS than SVC conditions, at roughly double the magnitude, but DVL and DVS involved glances of similar length (see Table 7.8 - main effect of condition). Following up with Fisher's Protected *t*-test for levels of a variable with unequal *n*, DVS contrasts with SVC at the $p<.05$ level ($t(18)=2.14$, $t_{crit}(18)=2.10$) but DVL narrowly fails to reach this level of significance ($t(18)=2.08$). The card task phase involved longer instances of GazeVideo than the discussion (main effect of task). However, again as is clear from Figure 7.6, on average South roles' GazeVideo were rather longer then their North partners for the cards but much the same in the discussion task phase (role by task interaction). There is an indication that this role-task relationship is restricted to dyads in the DVS and DVL condition, however the condition-task-role interaction narrowly failed to reach significance (see Tables 7.7 and 7.8).

Plotting standardized frequency data (Figure 7.8), DVL and DVS participants engaged in GazeVideo considerably more frequently than SVC participants, at about twice the rate. A split-plot ANOVA confirmed a main effect of condition and, following up with Fisher's protected *t*-test, that DVL and DVS differ from SVC but not from one another (see Tables 7.8 and 7.9). GazeVideo frequency was also higher for the card task than for the discussion task (main effect of task). As Figure 7.8 suggests that this difference is limited to DVL and DVS conditions whereas, if anything, the opposite is true for SVC (task and video condition interaction[39]). During the cards task phase, North participants look at their South partners with about the same frequency as South looks at North. However, in the discussion task phase, North looks at South markedly less often whilst the frequency of South GazeVideo is much the same (task by role interaction).

Figure 7.9 plots the standardized time data for dyads. It suggests that, overall, participants in the dynamic video conditions spent longer in GazeVideo than

[39] Analysis of Simple Main Effects was not possible, due to unequal 'n' for levels of Gaze variable.

did those in the still control. (See Table 7.11 - condition main effect, follow up with Fisher's protected *t*-test). More time was spent gazing towards the video links during the cards task phase than for the discussion task phase (main effect of task phase). Figure 7.9 shows that the experimental condition difference was far more pronounced for the cards than for the discussion task phase (interaction between condition and task phase[40]). Although there is an indication that South roles spent more time in GazeVideo than North, contrasts by role narrowly failed to reach significance.

In sum, "investment" in GazeVideo seems to have been highly related to the dynamics of the video image. Participants in DVS and DVL looked at one another more often, for longer and did so for more time overall than those in SVC. There is little evidence that the size of the image influenced this behaviour (the difference between DVS and DVL configurations), the relevant matter seems to have been quite simply whether or it was "live". Task phase had a similarly profound effect on GazeVideo Activity, in glance rate and duration: North's rate increases and South's duration of GazeVideo increases in the cards relative to the discussion phase.. Higher levels were in evidence for the highly-formulaic cards than for the relatively free-flowing discussion phase.

[40]Again, unequal 'n' per condition precludes the use of a simple main effects follow-up.

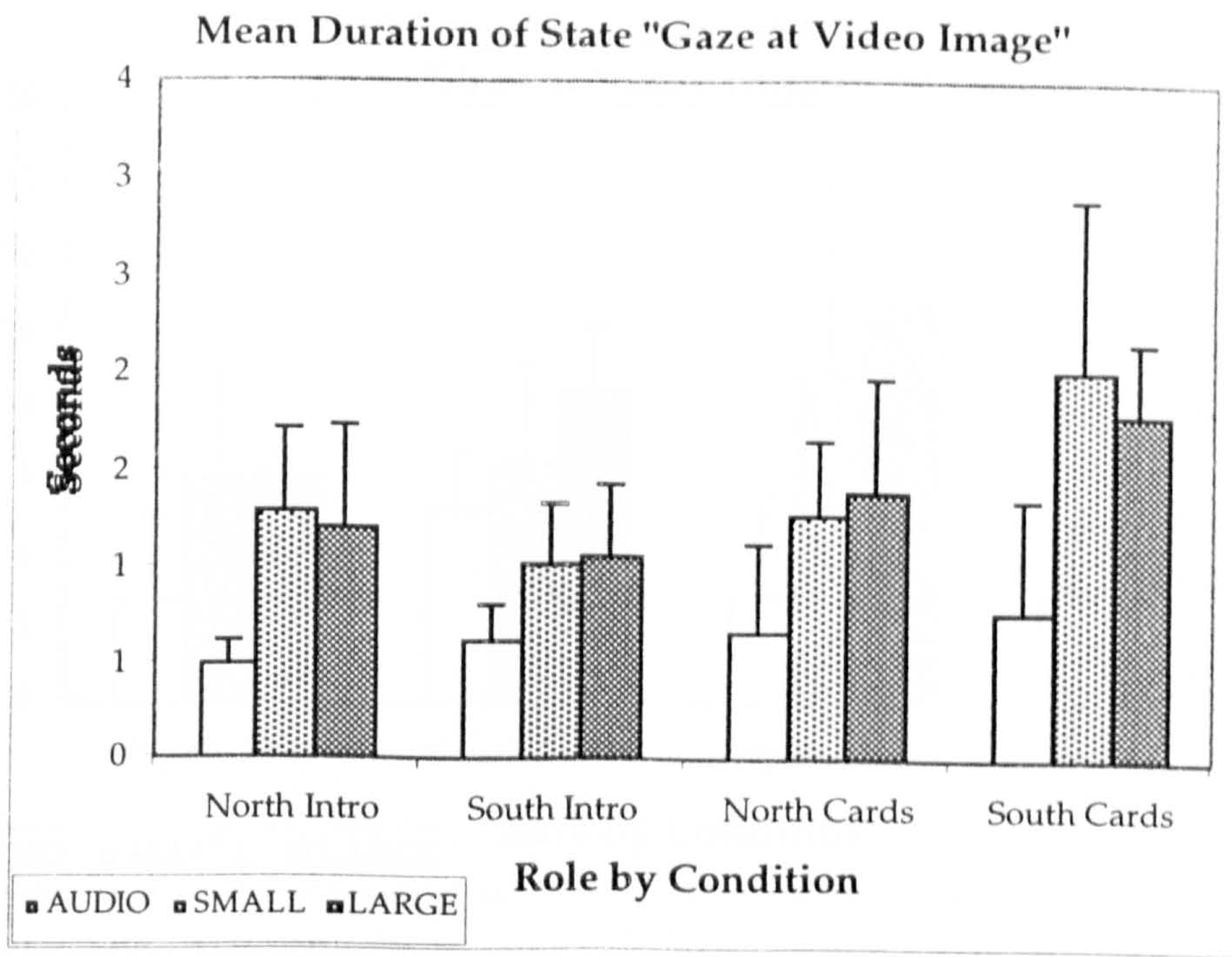


Figure 7.6: Mean duration of GazeVideo, contrasting role and task phase with condition

Duration of Gaze at Video (secs.)	North Discussion		South Discussion		North Cards		South Cards	
AUDIO	0.50	(0.11)	0.65	(0.17)	0.68	(0.45)	0.77	(0.57)
LARGE	1.22	(0.52)	1.06	(0.37)	1.39	(0.61)	1.82	(0.37)
SMALL	1.30	(0.43)	1.02	(0.31)	1.28	(0.38)	2.04	(0.85)
Overall	1.04	(0.53)	0.93	(0.34)	1.15	(0.56)	1.60	(0.80)

Table 7.7: Mean duration data for GazeVideo Activity

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	7.33	18	.41		
COND	9.41	2	4.70	11.55	.001**
Tests involving 'ROLE' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	3.55	18	.20		
ROLE	.59	1	.59	2.97	.102
COND BY ROLE	.06	2	.03	.15	.858
Tests involving 'TASK' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	2.31	18	.13		
TASK	2.89	1	2.89	22.56	.000**
COND BY TASK	.47	2	.24	1.84	.188
Tests involving 'ROLE BY TASK' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	2.92	18	.16		
ROLE BY TASK	1.42	1	1.42	8.74	.008**
COND BY ROLE BY TASK	.95	2	.48	2.93	.079

Table 7.8: GazeVideo mean duration split-plot ANOVA table, role and task by condition

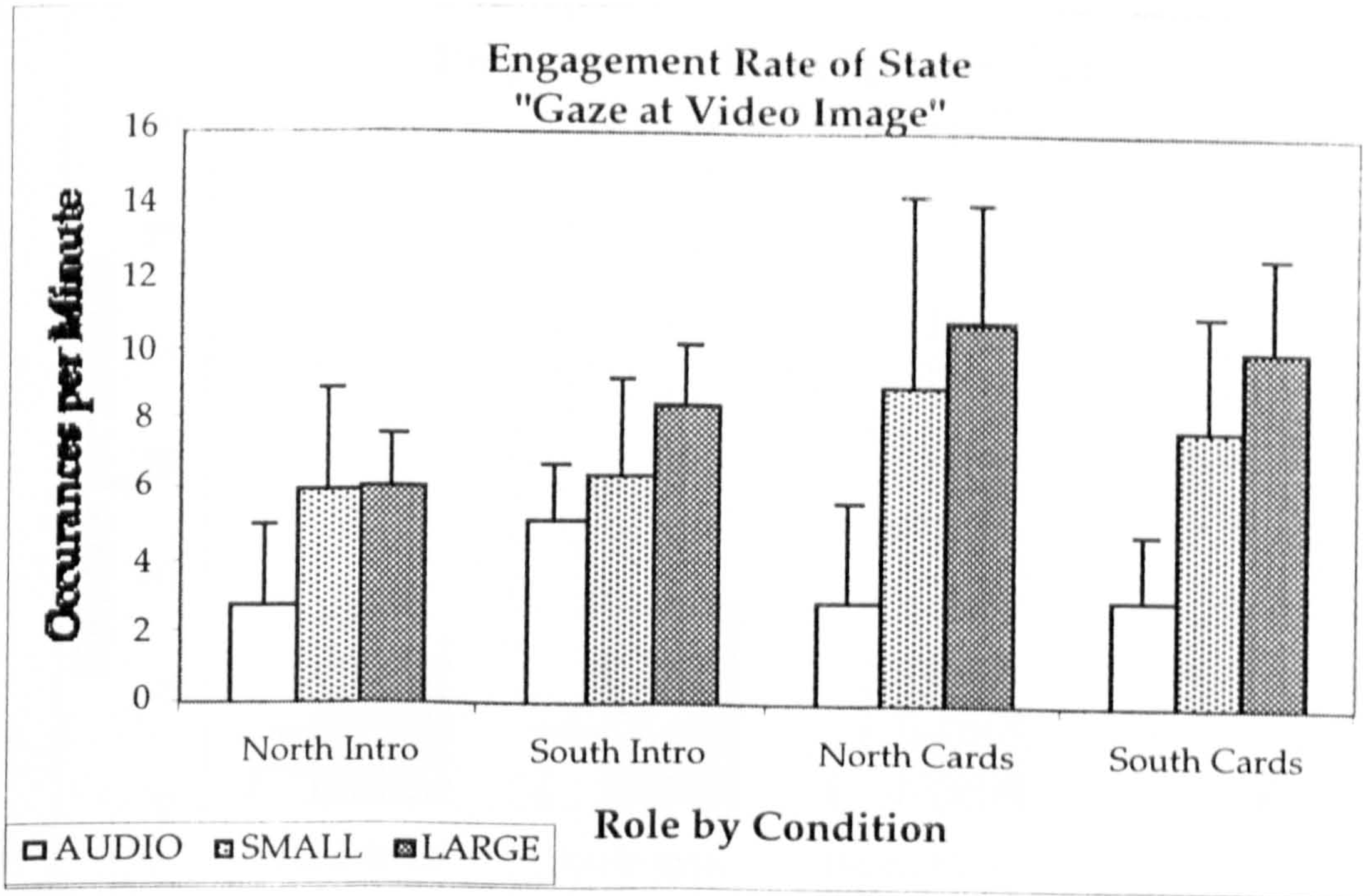


Figure 7.8: GazeVideo frequency data, contrasting role and task with experimental condition

Rate of Gaze (n/min.)	North Discussion		South Discussion		North Cards		South Cards	
AUDIO	2.72	(2.36)	5.11	(1.58)	2.90	(2.77)	2.97	(1.85)
SMALL	5.98	(2.87)	6.39	(2.74)	8.99	(5.25)	7.79	(3.28)
LARGE	6.13	(1.46)	8.45	(1.68)	10.82	(3.20)	10.11	(2.49)
Overall	5.10	(2.65)	6.81	(2.42)	7.95	(5.01)	7.30	(3.89)

Table 7.8: GazeVideo frequency, experimental condition by role and task phase

Tests of Between-Subjects Effects.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	356.27	18	19.79		
COND	418.30	2	209.15	10.57	.001**
Tests involving 'ROLE' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	68.56	18	3.81		
ROLE	6.20	1	6.20	1.63	.218
COND BY ROLE	9.48	2	4.74	1.24	.312
Tests involving 'TASK' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	106.64	18	5.92		
TASK	44.66	1	44.66	7.54	.013*
COND BY TASK	62.16	2	31.08	5.25	.016*
Tests involving 'ROLE BY TASK' Within-Subject Effect.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	38.60	18	2.14		
ROLE BY TASK	27.95	1	27.95	13.03	.002**
COND BY ROLE BY TASK	1.87	2	.93	.44	.654

Table 7.9: GazeVideo frequency split-plot ANOVA table, role and task by condition

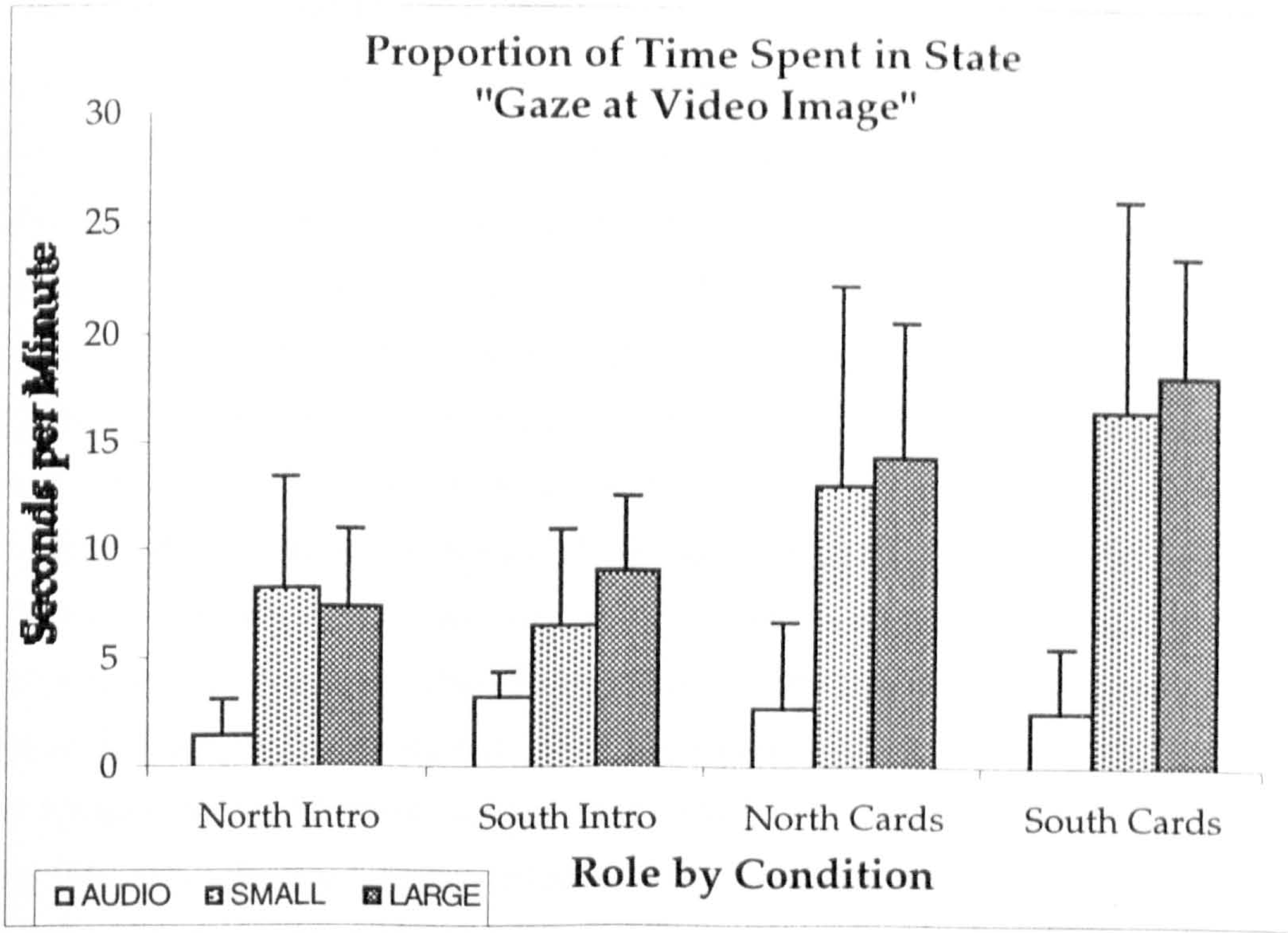


Figure 7.9: GazeVideo standardized time data, contrasting role and task with condition

Proportion of Gaze at Video	North Discussion		South Discussion		North Cards		South Cards	
AUDIO	1.54	(1.59)	3.30	(1.28)	2.88	(4.01)	2.70	(3.04)
SMALL	8.32	(4.98)	6.59	(4.34)	13.07	(9.23)	16.75	(9.39)
LARGE	7.45	(3.58)	9.09	(3.43)	14.49	(6.17)	18.30	(5.36)
Overall	6.05	(4.61)	6.60	(3.99)	10.70	(8.31)	13.33	(9.31)

Table 7.10: GazeVideo standardized time data, experimental condition by role and task phase

Tests of Between-Subjects Effects.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1338.64	18	74.37		
COND	1468.22	2	734.11	9.87	.001**
Tests involving 'TASKPHASE' Within-Subject Effect.					
WITHIN+RESIDUAL	371.88	18	20.66		
TASKPHASE	585.72	1	585.72	28.35	.000**
COND BY TASKPHASE	238.88	2	119.44	5.78	.012*
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN+RESIDUAL	220.32	18	12.24		
ROLE	46.37	1	46.37	3.79	.067
COND BY ROLE	16.83	2	8.42	.69	.515
Tests involving 'TASKPHASE BY ROLE' Within-Subject Effect.					
WITHIN+RESIDUAL	159.10	18	8.84		
TASKPHASE BY ROLE	18.21	1	18.21	2.06	.168
COND BY TASKPHASE BY ROLE	43.67	2	21.83	2.47	.113

Table 7.11 GazeVideo standardized time split-plot ANOVA, role and task by condition

7.5.3 Summary of Simple Analyses

Ten hypotheses were set for Simple Activity Set Analysis of SpeechUtterance and GazeVideo, based on Chapter One and the findings of Chapter Six. H1 suggested that frequency and duration of SpeechUtterance would vary with task, reflecting task formality differences. These differences were found but in the opposite direction to that predicted, almost certainly as a consequence of the very limited responses generated by the cards task. Overall amount of Utterance did not differ. H1 is thus disconfirmed. H2 made similar predictions for the effect of condition on SpeechUtterance Activity, for which there is also little evidence. H3 predicted an interaction between task and condition such that impoverished medium should combine with impoverished task to emphasize formality and amount of talk differences. Although such a pattern was not found simply between task and condition, a three-way interaction additionally including experimental role was found, for standardized time and duration data. A statistical interpretation of a three-way interaction is difficult. Inspection of Figures 7.3 and 7.5 suggest that North roles in the DVL condition in the discussion phase spoke longer turns and more overall than their South counterparts, in other conditions and in the other task. No firm conclusions may be drawn. However, H7 did anticipate role differences, so that North's SpeechUtterance would exceed South on all measures. Whilst there were no clear role differences in mean duration of Utterance, North did engage in more talk than South, measured both by overall time and frequency, and so H7 is broadly supported. Again, the complex interaction suggested that this prediction would have been more appropriate for DVL than the other conditions.

H4 anticipated profound task effects on duration and overall time measures of GazeVideo Activity and H5 similarly asserted that experimental condition would be associated with similar differences. Both are strongly supported, although both condition and task effects extended to unanticipated frequency differences. H6 suggested that task and experimental condition would interact on all measures. Good evidence for this was found for frequency and overall time measures but not for mean duration. H6 is thus partially supported. H8 foresaw shorter, more frequent GazeVideo Activity for North than for South. Although there was no overall effect of this kind, role by task interactions on duration and frequency but not overall time measures suggest that this was much the pattern during the cards task phase.

7.6 Co-activity Analyses

Data considered under this heading is taken to reflect the extent to which certain Activities relate to and to some extent depend upon conditions given by other Activities. Thus is it possible to say, for example, that the speech of one individual is contingent on the gaze of another (i.e. is there any evidence that they are more or less likely to speak when looked at than when not). Time in Co-activity states is given as the standardized "seconds per minute" measure and as proportion of interaction period. Eight Hypotheses were formed for Co-activity analyses, as described above (Section 7.1.2; H9 - H17). These are each tested in turn. Two forms of test are included, depending on the nature of the hypothesis concerned. One question is whether Sunchronization is greater than would be expected by chance. This is evaluated by comparing obtained *S* scores with zero using a paired samples t-test. Results then indicate whether there is evidence of consistent Synchronization, positive or negative, in the Co-activity under test. These are performed separately for each role by task phase and condition combination. The second question directly contrasts *S* scores for a Co-activity as a function of experimental condition, task phase and, where appropriate, experimental role by means of a three-way split-plot ANOVA. For the sake of brevity, the absolute Synchronization results are indicated in the tables by asterisks, using the convention described earlier. Where the result of the paired-samples t-test is material to discussion, it is given in the text or referred to in foot notes.

7.6.1 Was GazeVideoGazeVideo Co-activity Synchronized?

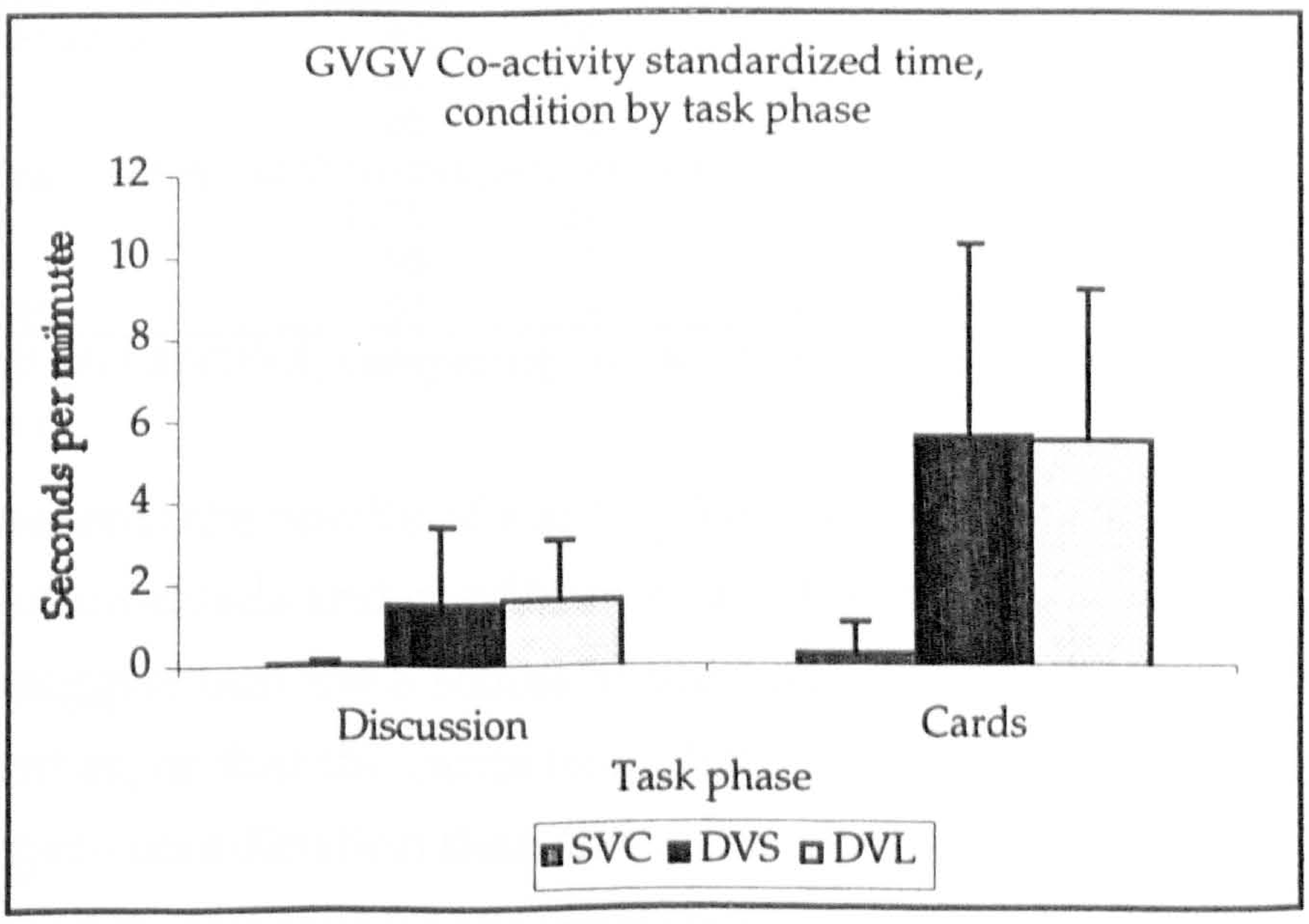


Figure 7.10: Observed standardized time in GazeVideoGazeVideo Co-activity, contrasting task with condition

As Figure 7.10 shows, GazeVideo behaviour followed expectations from the Simple proportions of time in this Activity state: SVC dyads almost never looked at one another's image at the same time whereas DVS and DVL dyads did so for a small but appreciable part of their interaction.

GVGV	Discussion			Cards		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	0.11 (0.10)	0.10 (0.11)	0.146 (0.367)	0.37 (0.78)	0.18 (0.28)	0.017 (0.405)
DVS	1.57 (1.86)	1.23 (1.38)	0.058 (0.288)	5.59 (4.68)	5.91 (5.94)	0.118 (0.406)
DVL	1.73 (1.37)	1.39 (1.04)	0.121 (0.254)	5.49 (3.67)	5.61 (4.05)	-0.026 (0.192)
Overall	1.22 (1.49)	0.97 (1.13)	0.107 (0.287)	4.06 (4.14)	4.16 (4.80)	0.034 (0.327)

Table 7.12 Standardized time in Co-activity state GazeVideoGazeVideo with mean S score (SD in parentheses), experimental condition by task phase.

Table 7.12 presents data for the GazeVideoGazeVideo Co-activity for each of the experimental conditions by task phase. There was no evidence of level of mutual looking at the video window exceeding chance levels for either role in either task phase in any condition. This indicated by an absence of asterisks in the table.

```
Analysis of Variance, GazeVideoGazeVideo Co-activity S scores
Tests of Between-Subjects Effects.
 Source of Variation          SS        DF        MS          F   Sig of F
 WITHIN CELLS               1.94        18       .11
 IMAGE                       .01         2       .01        .06       .938
Tests involving 'PHASE' Within-Subject Effect.
 WITHIN CELLS               1.75        18       .10
 PHASE                       .05         1       .05        .56       .466
 IMAGE BY PHASE              .09         2       .05        .48       .627
```

Table 7.13 Split-plot ANOVA, comparing effects of experimental condition and task phase on S score.

Table 7.13 presents the results of a split-plot analysis of variance, with task phase as a within-dyads and condition as a between-dyads factor. There is no evidence to suggest that the *S* scores of the three image conditions contrasted with one another, or that the cards task phase was associated with more or less mutual gaze coordination than the discussion task phase.

7.6.2 Did the GazeVideoGazeElsewhere Co-activity show any evidence of Synchronization?

A second aspect of Gaze coordination concerns the way in which participants looked at the materials in support of their collaborative activity. In the discussion task, that meant the shared screen and in the cards task, the cards themselves for North and a response sheet for South. As discussed above, it would be difficult to read too much into the data corresponding to this sort of activity, as GazeElsewhere, due to coding difficulties. However, the Co-activity GazeVideoElsewhere should be revealing of any Gaze organization beyond the rather limited GVGV state. As role discriminates the form of GazeActivity, there are two such Co-activity states: South in GazeVideo whilst North was in GazeElsewhere and vice versa. Tables 7.14 and 7.15 present observed and expected standardized time with corresponding mean *S* scores for the discussion and card task phases respectively. Figures 7.11 and 7.12 present the corresponding observed standardized time data.

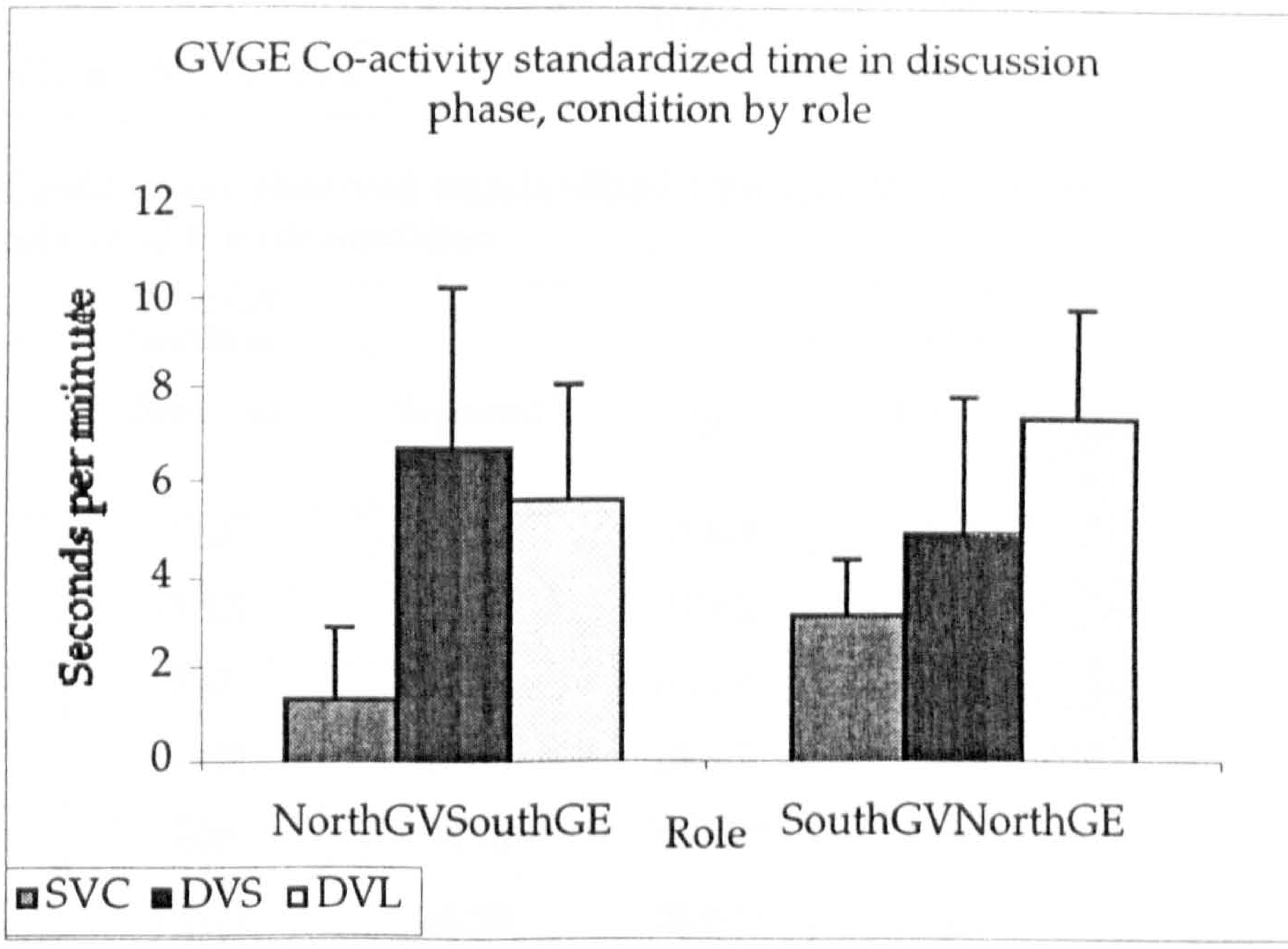


Figure 7.11: Discussion phase observed standardized time in GVGE Co-activity, contrasting role-in-GV with condition

GVGE Disc. phase	NorthGV SouthGE			SouthGV NorthGE		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	1.42	1.49	-0.317	3.17	3.27	-2.231
	(1.49)	(1.56)	(0.446)	(1.22)	(1.26)	(2.987)
DVS	6.76	7.92	-0.989*	5.02	6.05	-0.492
	(3.56)	(4.68)	(1.066)	(2.87)	(3.95)	(0.567)
DVL	5.72	6.87	-0.757**	7.37	8.64	-1.048**
	(2.47)	(3.27)	(0.519)	(2.45)	(3.18)	(0.744)
Overall	4.84	5.68	-0.709	5.39	6.24	-1.201
	(3.41)	(4.30)	(0.748)	(2.84)	(3.69)	(1.739)

Table 7.14 Standardized time in Co-activity state GazeVideoGazeElse with mean *S* score (SD in parentheses), experimental condition by role in GV

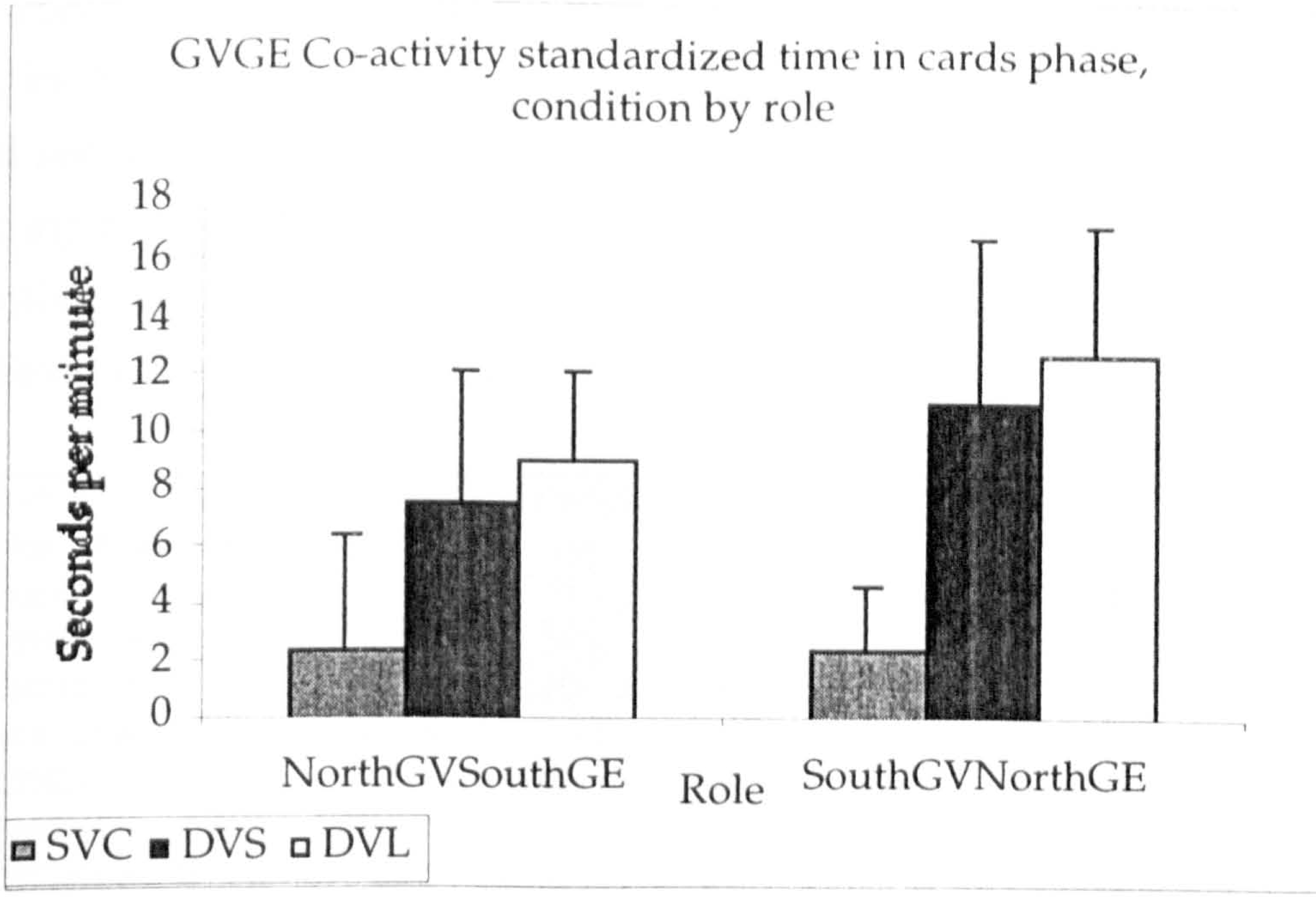


Figure 7.12: Cards phase observed standardized time in GVGE Co-activity, contrasting role-in-GV with condition

GVGE Cards phase	NorthGV SouthGE			SouthGV NorthGE		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	2.53	2.81	-0.649	2.32	2.63	-2.797
	(3.82)	(3.97)	(0.962)	(2.33)	(3.00)	(5.590)
DVS	7.47	10.24	-0.913**	11.17	14.47	-2.282
	(4.65)	(6.74)	(0.347)	(5.53)	(7.43)	(3.092)
DVL	9.00	11.95	-0.770**	12.81	16.24	-1.174**
	(3.19)	(4.88)	(0.445)	(4.26)	(4.55)	(0.380)
Overall	6.64	8.77	-0.783	9.27	11.76	-2.007
	(4.61)	(6.43)	(0.590)	(6.14)	(7.85)	(3.350)

Table 7.15 Standardized time in Co-activity state GazeVideoGazeElse with mean *S* score (SD in parentheses), experimental condition by role in GV

Both DVL participants in both discussion and cards phases showed highly significant Synchronization between GazeVideo and the others GazeElse Activity[41]. DVS North participants GazeVideo Synchronized with South's GazeElse, whereas although DVS Souths approach significance on this measure, their *S* scores showed no clear departure from chance co-occurrence[42]. Neither SVC participants showed any evidence of

[41]Paired samples t-test, p<.01 for all four comparisons, t values ranging from 4 to 8.73.

[42]South GV North GE, discussion, t(6)=2.30; p=.61. SouthGV NorthGE, cards, t(6)=1.95; p=.099.

Synchronizing their GazeVideo with the other's GazeElsewhere, regardless of condition[43]. These findings lend some strong support to Hypothesis 9. It would seem that participants in the DVL condition looked "elsewhere", including at their shared screens and paper-based information in the respective tasks, rather less than the Simple Activity Analysis would lead one to believe. DVS North participants also show negative Synchronization in this way.

Split-plot ANOVA for GazeVideoGazeElsewhere Co-activity					
Tests of Between-Subjects Effects					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	58.16	18	3.23		
CONDITION	4.32	2	2.16	.67	.525
Tests involving 'PHASE' Within-Subject Effect.					
WITHIN CELLS	72.42	18	4.02		
PHASE	4.35	1	4.35	1.08	.312
CONDITION BY PHASE	2.32	2	1.16	.29	.753
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN CELLS	76.16	18	4.23		
ROLE	18.23	1	18.23	4.31	.053
CONDITION BY ROLE	11.59	2	5.79	1.37	.280
Tests involving 'PHASE BY ROLE' Within-Subject Effect.					
WITHIN CELLS	74.69	18	4.15		
PHASE BY ROLE	2.82	1	2.82	.68	.421
CONDITION BY PHASE BY ROLE	3.39	2	1.69	.41	.671

Table 7.16: Split-plot ANOVA of *S* score for GVGE Co-activity, contrasting experimental condition by role and task phase

Hypothesis 10 was tested with a multivariate ANOVA for mixed design, comparing *S* scores by experimental condition, role and task phase (see Table 7.16). The prevalence of interactions between condition, role and task phase in the Simple analyses suggested that a t-test would not be an adequate procedure. There is no indication that the greater consistency in DVL Synchronization than in DVS translates into a difference between these conditions, or between them and SVC. Hypothesis 10 is thus rejected.

43Although the SVC condition included the fewest dyads, and hence was associated with the fewest degrees of freedom for this statistical procedure, the associated t values were comparatively small, ranging from 1.2 to 1.74.

7.6.3 Did the Synchronization of listeners' gaze at the image of the speaker vary by experimental conditions?

Hypotheses 11 and 12 were that the Gaze of one participant should Synchronize with the SpeechUtterance of their partner. Tables 7.17 and 7.18 and Figures 7.13 and 7.14 present data for discussion and cards task phases respectively.

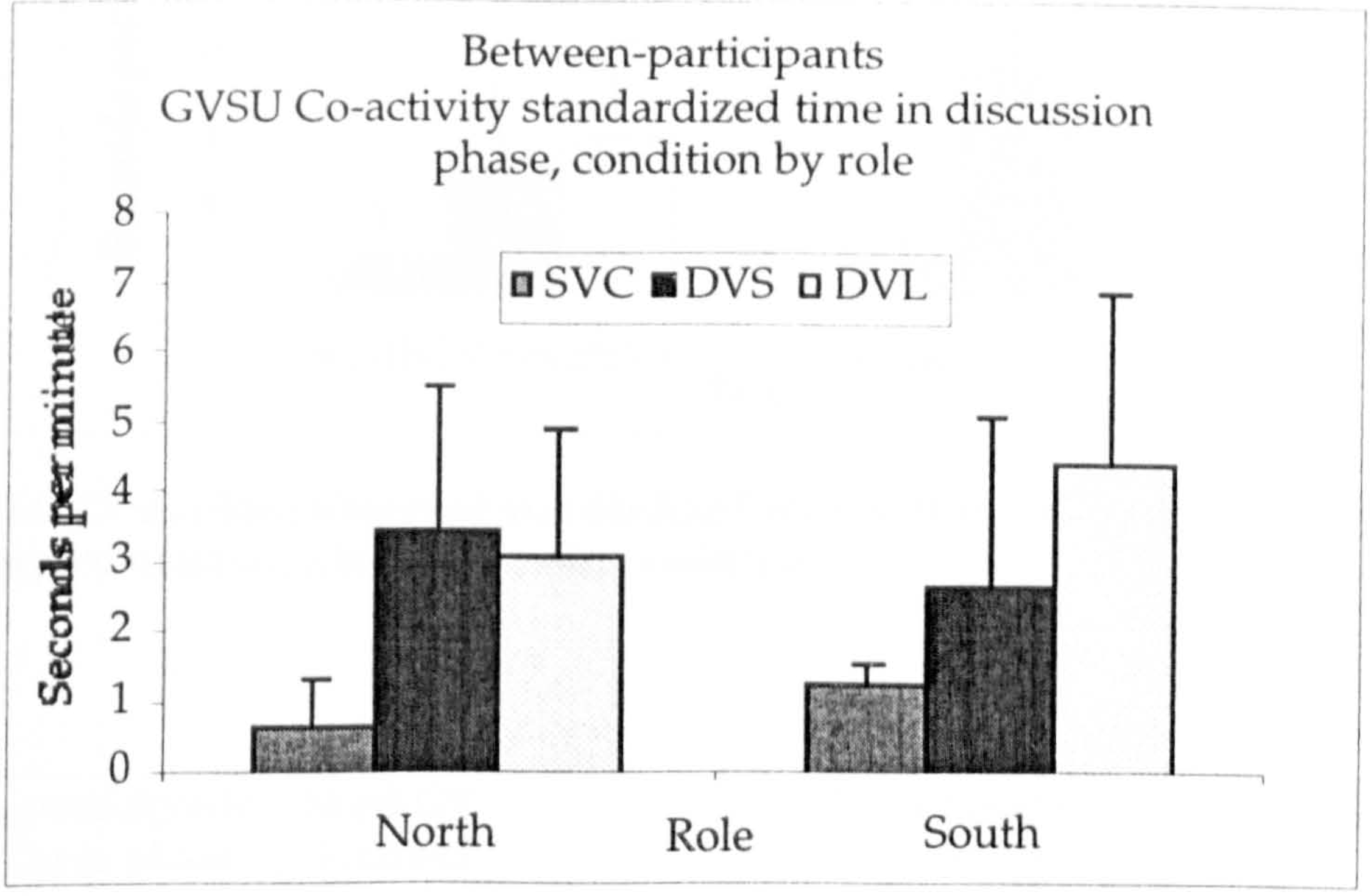


Figure 7.13: Discussion phase observed standardized time in between-participants GVSU Co-activity, contrasting role-in-GV with condition

Between-participants GVSU, Disc. phase	NorthGV SouthSU			SouthGV NorthSU		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	0.66	0.52	0.207	1.26	1.10	0.181
	(0.67)	(0.59)	(0.582)	(0.30)	(0.48)	(0.279)
DVS	3.46	2.73	0.364*	2.67	2.36	0.119
	(2.06)	(1.95)	(0.378)	(2.44)	(2.33)	(0.148)
DVL	3.06	1.97	0.478*	4.43	4.30	0.080
	(1.89)	(1.61)	(0.451)	(2.50)	(2.77)	(0.167)
Overall	2.51	1.81	0.363	2.93	2.74	0.122
	(2.02)	(1.71)	(0.460)	(2.40)	(2.49)	(0.194)

Table 7.17: Discussion task phase between-participants GVSU observed and expected standardized time and mean S score, experimental condition by role-in-GazeVideo

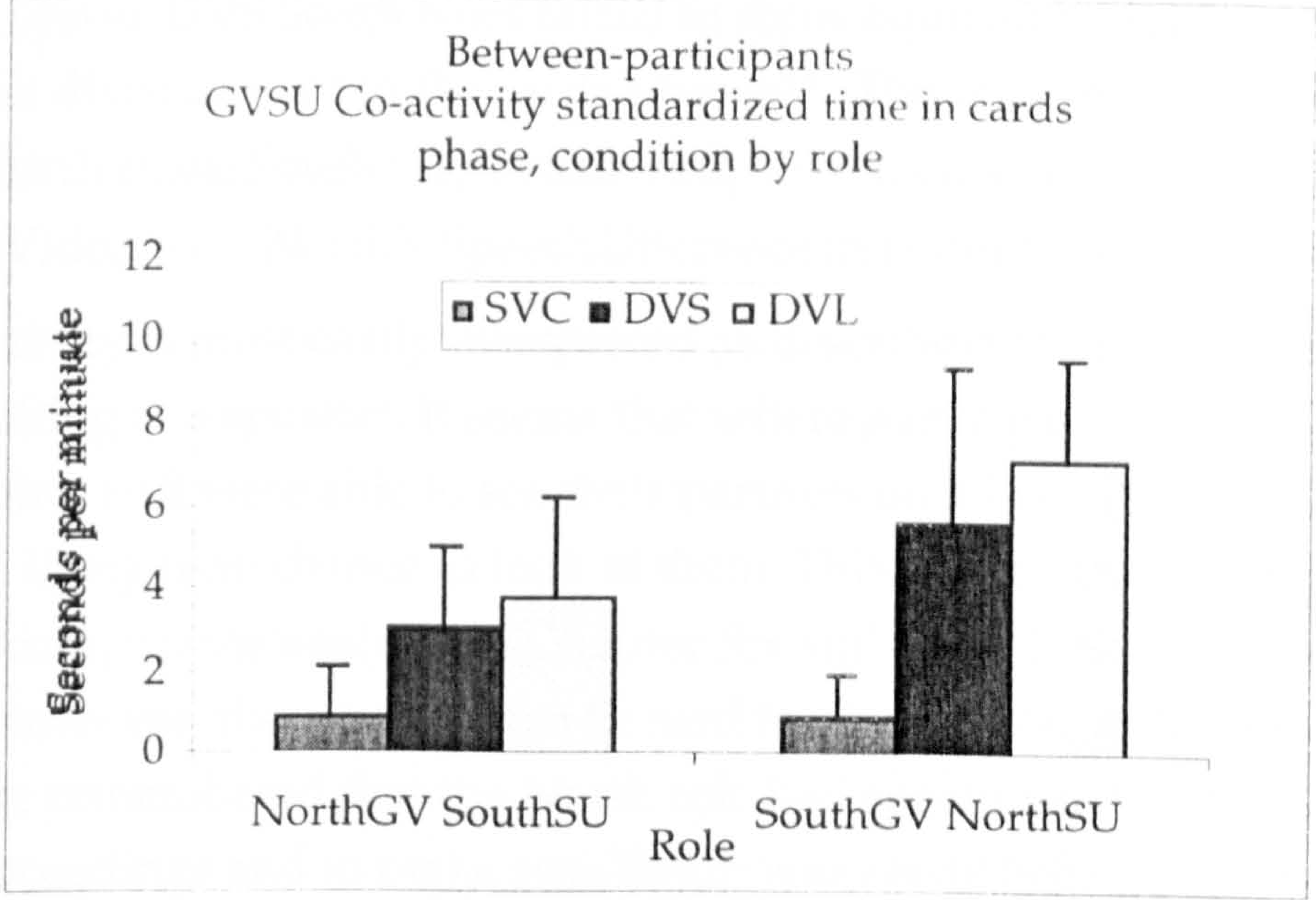


Figure 7.14: Cards phase observed standardized time in between-participants GVSU Co-activity, contrasting role-in-GV with condition

Between-participants GVSU, Cards phase	NorthGV SouthSU			SouthGV NorthSU		
Condition	Observed	Expected	"*S*"	Observed	Expected	"*S*"
SVC	1.01	0.94	0.310	0.99	0.93	0.280
	(1.10)	(1.36)	(0.582)	(0.99)	(1.19)	(0.457)
DVS	3.09	3.46	-0.134	5.67	6.12	-0.029
	(1.97)	(2.35)	(0.277)	(3.82)	(4.51)	(0.220)
DVL	3.89	4.27	-0.067	7.21	7.61	-0.056
	(2.36)	(2.51)	(0.146)	(2.44)	(2.62)	(0.211)
Overall	2.80	3.05	0.018	4.92	5.20	0.049
	(2.21)	(2.51)	(0.389)	(3.70)	(4.12)	(0.324)

Table 7.18: Card task phase between-participants GVSU observed and expected standardized time and mean *S* score, experimental condition by role-in-GazeVideo

SVC subjects showed no evidence of Synchronizing their Gaze at the video window with the SpeechUtterance of their partner, regardless of role or task[44]. DVS North roles showed some significant positive Synchronization of their GazeVideo with DVS South SpeechUtterance, in the discussion but not in the

[44]Paired-samples t-test, t(5) ranged from 0.87 to 1.59.

cards task phase. DVS South roles failed to show equivalent Synchronization in either the discussion or in the cards phases[45]. The same pattern was found for DVL North roles, Souths again showing no evidence of Synchronizing their GazeVideo with North's SpeechUtterance in either task context[46].

This Co-activity is most easily interpreted as describing the likelihood of a listener looking at a speaker. It seems that where participants in the North role were listening and were able to see their partners on a live video link, they were more likely than chance to look at them. This is finding replicates the result of a comparable analysis in Chapter Six and so H11 receives some support. However, the same cannot be said for South roles as listeners. Here, it should be remembered that the North role had specific instructions to control proceedings and to make sure South was ready before proceeding. In this sense, North may have expressed this degree of control over and obligation to South whereas South had no in-role obligation to reciprocate and was less sensitive to the artificial status difference here.

Comparing DVL and DVS failed to show any indication of a profound difference in Synchronization (see Table 7.19), contradicting the prediction of Hypothesis 12.

Split-plot ANOVA for GazeVideoSpeechUtterance Co-activity					
Tests of Between-Subjects Effects.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	2.39	18	.13		
CONDITION	.39	2	.20	1.48	.255
Tests involving 'PHASE' Within-Subject Effect.					
WITHIN CELLS	3.33	18	.18		
PHASE	.73	1	.73	3.93	.063
CONDITION BY PHASE	.81	2	.40	2.18	.142
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN CELLS	1.45	18	.08		
ROLE	.20	1	.20	2.42	.137
CONDITION BY ROLE	.11	2	.05	.67	.526
Tests involving 'PHASE BY ROLE' Within-Subject Effect.					
WITHIN CELLS	1.46	18	.08		
PHASE BY ROLE	.33	1	.33	4.04	.060
CONDITION BY PHASE BY ROLE	.16	2	.08	.99	.389

Table 7.19: Split-plot ANOVA of *S* score for GVSU Co-activity, contrasting experimental condition by role and task phase

---

[45]DVS SouthGV NorthSU Co-activity S scores: in discussion phase, t(6)=2.12; p=.078, in cards phase, t(6)=0.35; p=.740.

[46]DVL South GV NorthSU Co-activity S scores: in discussion phase, t(7)=1.35; p=.218, in cards phase, t(6)=0.75; p=.478.

7.6.4 Whilst speaking, did participants look at the image of their partner more than chance predicts?

Chapter Six produced the somewhat surprising result that a participant's SpeechUtterance does not Synchronize with their GazeVideo (i.e. looking towards the image of their remote partner). From Kendon's classic observations on gaze and turn exchange, one might have supposed that a speaker's gaze should negatively Synchronize with their speech. H13 thus proposed that GazeVideoSpeechUtterance should negatively Synchronize. Tables 7.20 and 7.21 and Figures 7.15 and 7.16 respectively present discussion and cards task phase data for the GVSU within participants Co-activity.

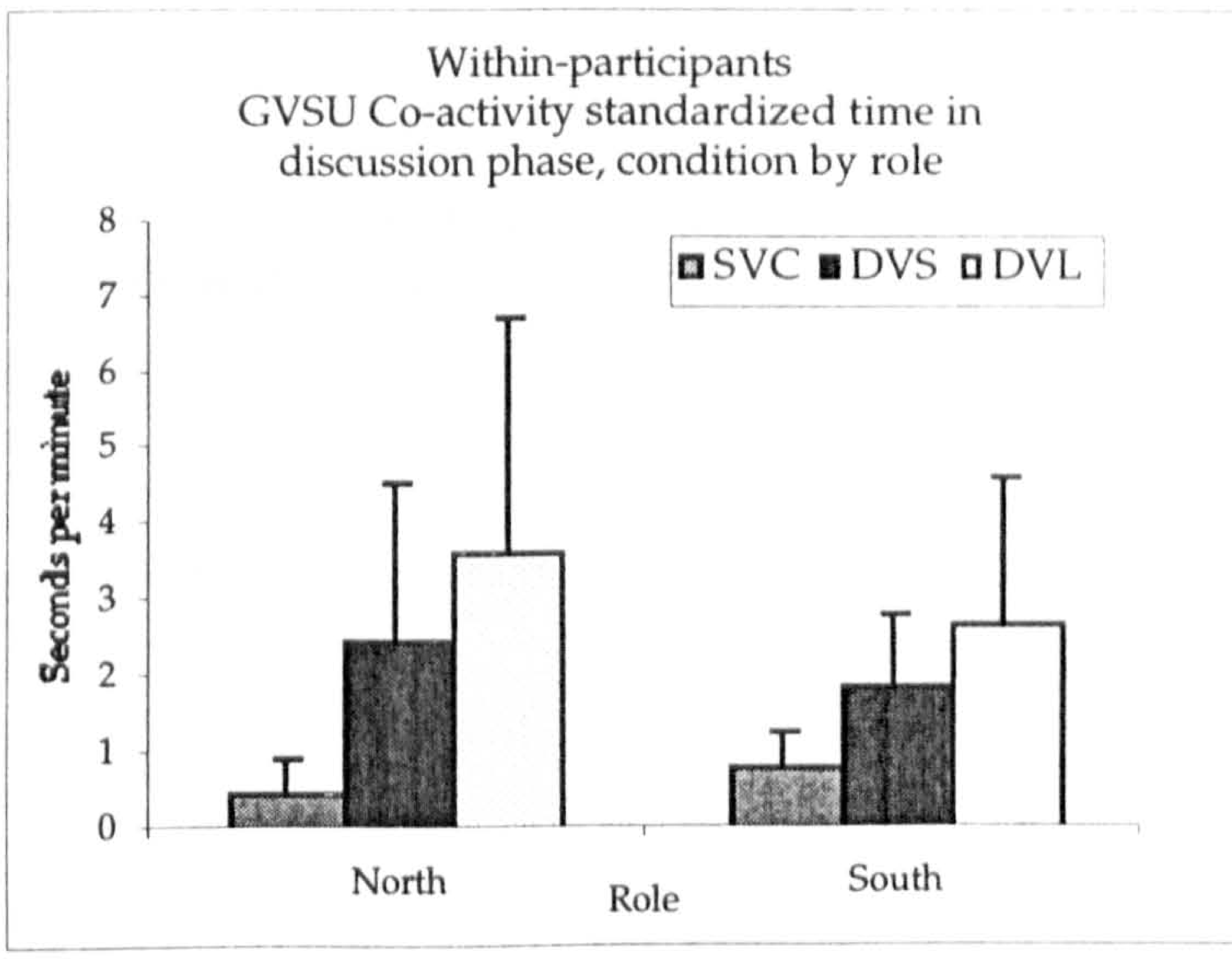


Figure 7.15: Discussion phase observed standardized time in within-participants GVSU Co-activity, contrasting role with condition

Within-participants GVSU, Disc. phase	North			South		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	0.48	0.44	0.019	0.83	0.94	-0.014
	(0.44)	(0.41)	(0.579)	(0.43)	(0.44)	(0.350)
DVS	2.46	2.98	-0.118	1.88	1.78	0.044
	(2.10)	(2.15)	(0.284)	(0.89)	(0.74)	(0.300)
DVL	3.59	3.63	0.016	2.70	2.43	0.104
	(3.16)	(3.16)	(0.290)	(1.88)	(1.59)	(0.299)
Overall	2.32	2.50	-0.028	1.89	1.79	0.050
	(2.56)	(2.60)	(0.376)	(1.45)	(1.21)	(0.302)

Table 7.20: Discussion task phase within-participants GVSU observed and expected standardized time and mean *S* score, experimental condition by role

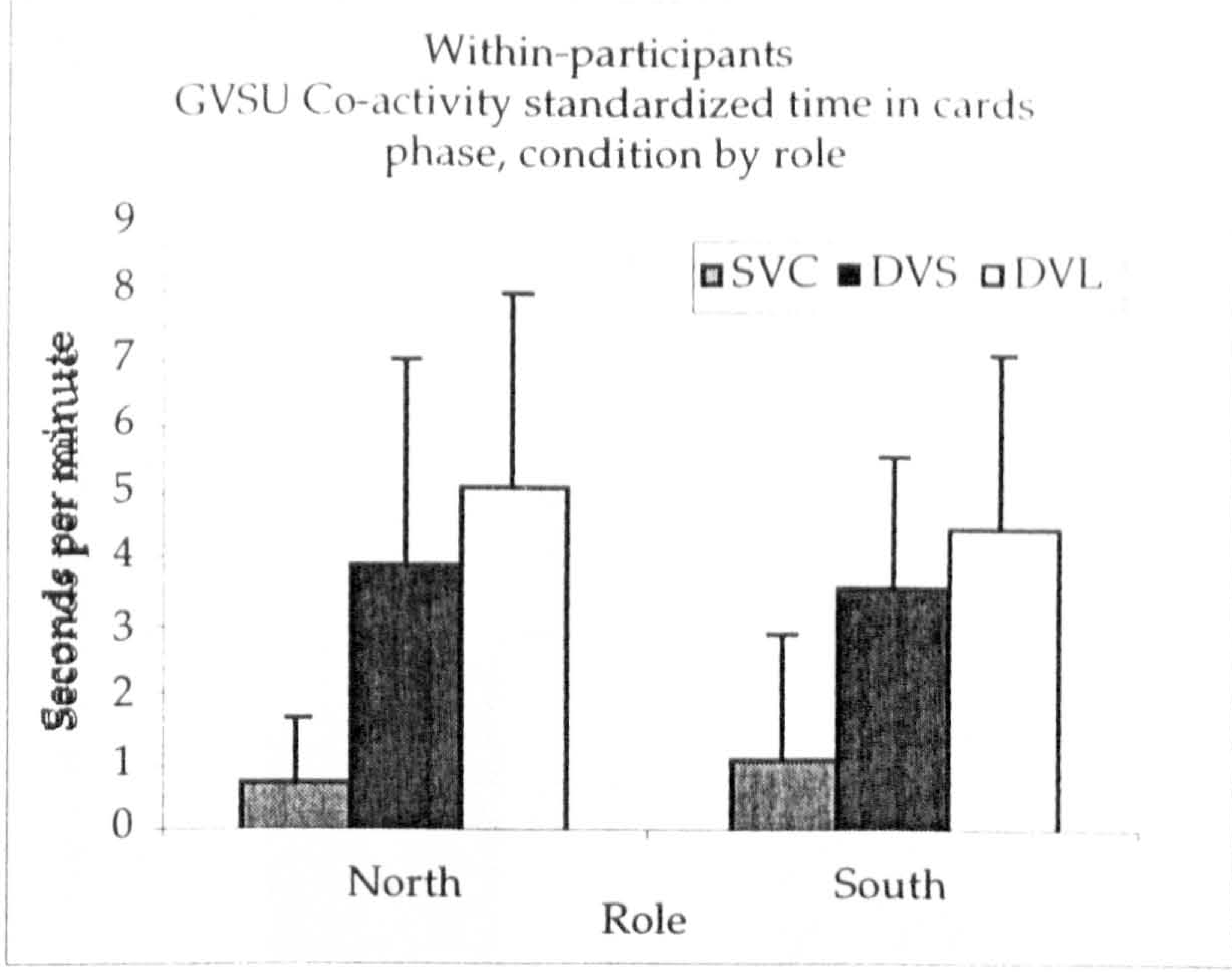


Figure 7.16: Cards phase observed standardized time in within-participants GVSU Co-activity, contrasting role with condition

Within-participants GVSU, Cards phase	North			South		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	0.77	0.79	0.085	1.09	0.86	-0.129
	(0.91)	(0.93)	(0.590)	(1.84)	(1.04)	(0.398)
DVS	3.93	4.92	-0.081	3.61	4.76	-0.183
	(3.12)	(4.19)	(0.450)	(1.97)	(3.09)	(0.362)
DVL	5.15	6.03	-0.169	4.51	5.20	-0.143
	(2.82)	(3.41)	(0.261)	(2.57)	(2.25)	(0.231)
Overall	3.49	4.16	-0.067	3.24	3.82	-0.152
	(3.05)	(3.82)	(0.428)	(2.53)	(2.94)	(0.313)

Table 7.21: Card task phase within-participants GVSU observed and expected standardized time and mean S score, experimental condition by role

Just as was found in the case study reported in Chapter Six, there was no evidence of any difference in the observed level of a person's GazeVideoSpeechUtterance than chance alone would lead one to expect. This is a valuable replication of an unlikely result, given that it was without regard to the quality of video image on offer, the task performed or the role of the participant. Since some evidence was found of between-participants Synchronization for these Activities, it is more striking still.

7.6.5 Does Speech Synchronization show "more formal" exchange as a function of video configuration?

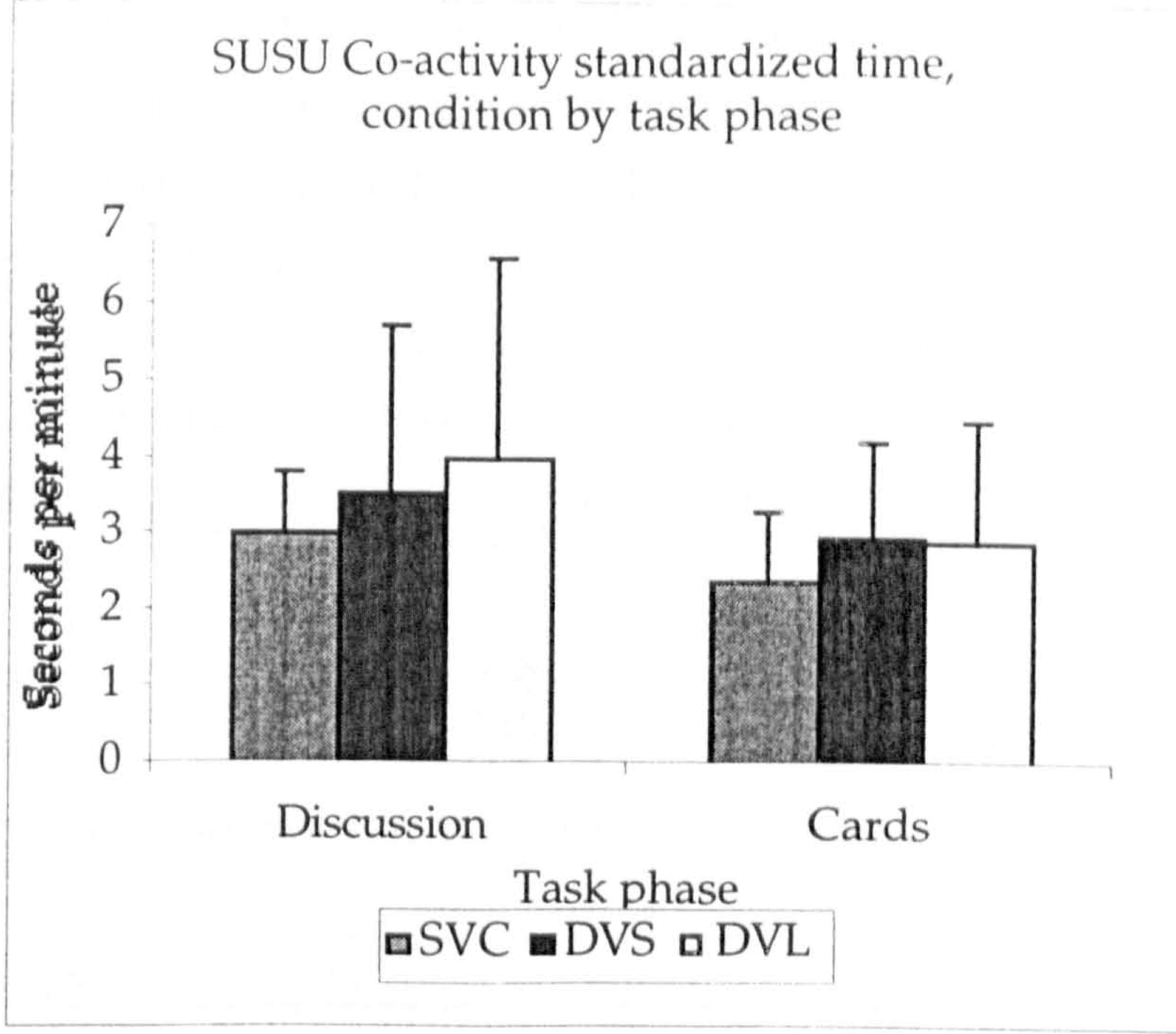


Figure 7.17: Observed standardized time in SpeechUtteranceSpeechUtterance Co - activity, contrasting task with condition

SUSU[47]	Discussion			Cards		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	3.00	5.99	-0.523**	2.41	5.76	-0.601**
	(0.85)	(2.25)	(0.228)	(0.93)	(1.88)	(0.207)
DVS	3.50	5.68	-0.418**	2.99	4.98	-0.375**
	(2.21)	(2.87)	(0.233)	(1.21)	(1.55)	(0.259)
DVL	3.98	6.35	-0.442**	2.94	5.89	-0.546**
	(2.62)	(4.40)	(0.306)	(1.54)	(2.06)	(0.117)
Overall	3.54	6.02	-0.457	2.80	5.55	-0.505
	(2.05)	(3.25)	(0.253)	(1.25)	(1.81)	(0.212)

Table 7.22: Observed and expected standardized time in SpeechUtteranceSpeechUtterance Co-activity state, experimental condition by task phase

The final four hypotheses make predications based on the common assertion in the VMI literature, that speech styles vary with the addition of a video

---

47 As one would expect, even in the most extreme circumstances, differences from chance levels of simultaneous speech, as represented by SpeechUtteranceSpeechUtterance Co-activity, are highly significant in all cases.

image to an audio communications facility. As discussed in Chapter One, statements of this kind are at best rather imprecise and in ignorance of the kind of collaborative activity concerned. The hypotheses for Speech Activity, as operationalized in the current experiment, suggest that the SVC condition (essentially, an audio-only facility) should show *more* negative Synchronization than the video conditions. More negative Synchronization shows less fluidity, and less shows greater overlap. It was additionally anticipated that the cards task would be associated with more "formal" styles of speech, in this sense, and interact with video condition.

Split-plot ANOVA of SUSU Co-activity *S* scores					
Tests of Between-Subjects Effects.					
Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1.48	18	.08		
CONDITION	.18	2	.09	1.10	.353
Tests involving 'PHASE' Within-Subject Effect.					
WITHIN CELLS	.48	18	.03		
PHASE	.02	1	.02	.86	.366
CONDITION BY PHASE	.04	2	.02	.83	.453

Table 7.23: Split-plot ANOVA of *S* score for SUSU Co-activity, contrasting experimental condition and task phase

There is no evidence that style of speech changed by condition, such that overlapping speech would be less (i.e. SUSU would be more negatively Synchronized) in any task or condition. It would seem, then, that the formality of spoken interaction was much the same despite the Gaze differences reported above, in association with task phase and experimental role. This is somewhat surprising but could be a consequence of the very high quality audio links employed, coupled with the "informality" of the testing situation. Even so, this is a null result and so caution should be exercised in drawing too many inferences from it.

Besides overlapping speech, one can ask about the coordination of SpeechUtterance with SpeechSilence, as the additional matter of time both partners were in SpeechSilence constituted between a half and two thirds of the interaction period. The Co-activity SUSS is informative about the distribution of Utterances between partners over the period.

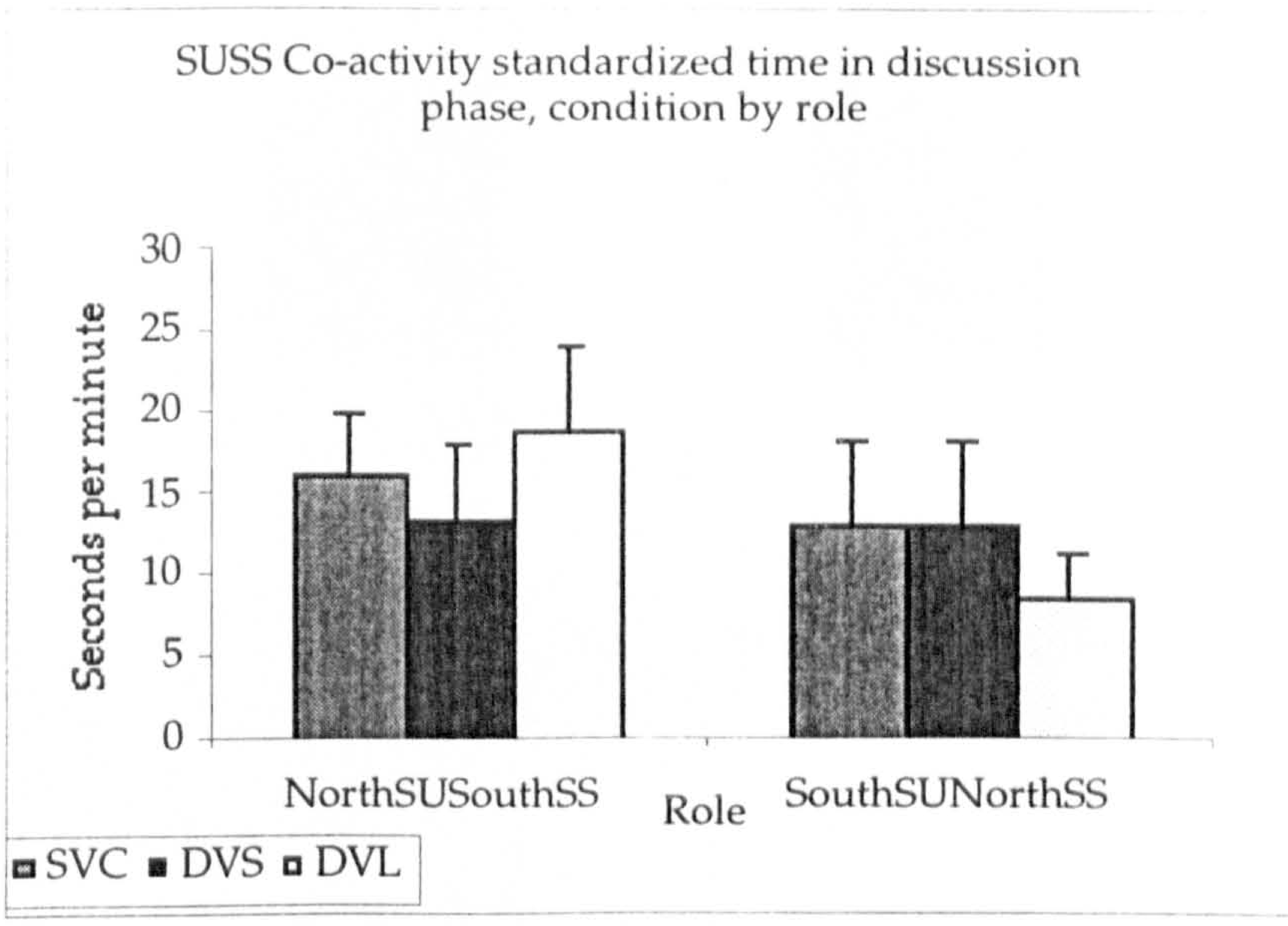


Figure 7.18: Discussion phase observed standardized time in between-participants SUSS Co-activity, contrasting role-in-SU with condition

Between-participants SUSS, Disc. phase	NorthSU SouthSS			SouthSU NorthSS		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	15.94	16.68	-0.348	13.10	13.44	-0.104
	(4.06)	(4.73)	(0.445)	(5.26)	(5.60)	(0.201)
DVS	13.44	14.61	-0.400	13.23	14.40	-0.317*
	(4.74)	(5.87)	(0.595)	(5.10)	(5.61)	(0.299)
DVL	19.02	21.11	-0.873**	8.57	8.72	-0.057
	(5.06)	(5.99)	(0.354)	(2.90)	(2.85)	(0.130)
Overall	16.28	17.68	-0.565	11.42	11.96	-0.157
	(5.07)	(6.06)	(0.511)	(4.79)	(5.20)	(0.238)

Table 7.24: Discussion task phase between-participants SUSS observed and expected standardized time and mean S score, experimental condition by role-in-SpeechUtterance

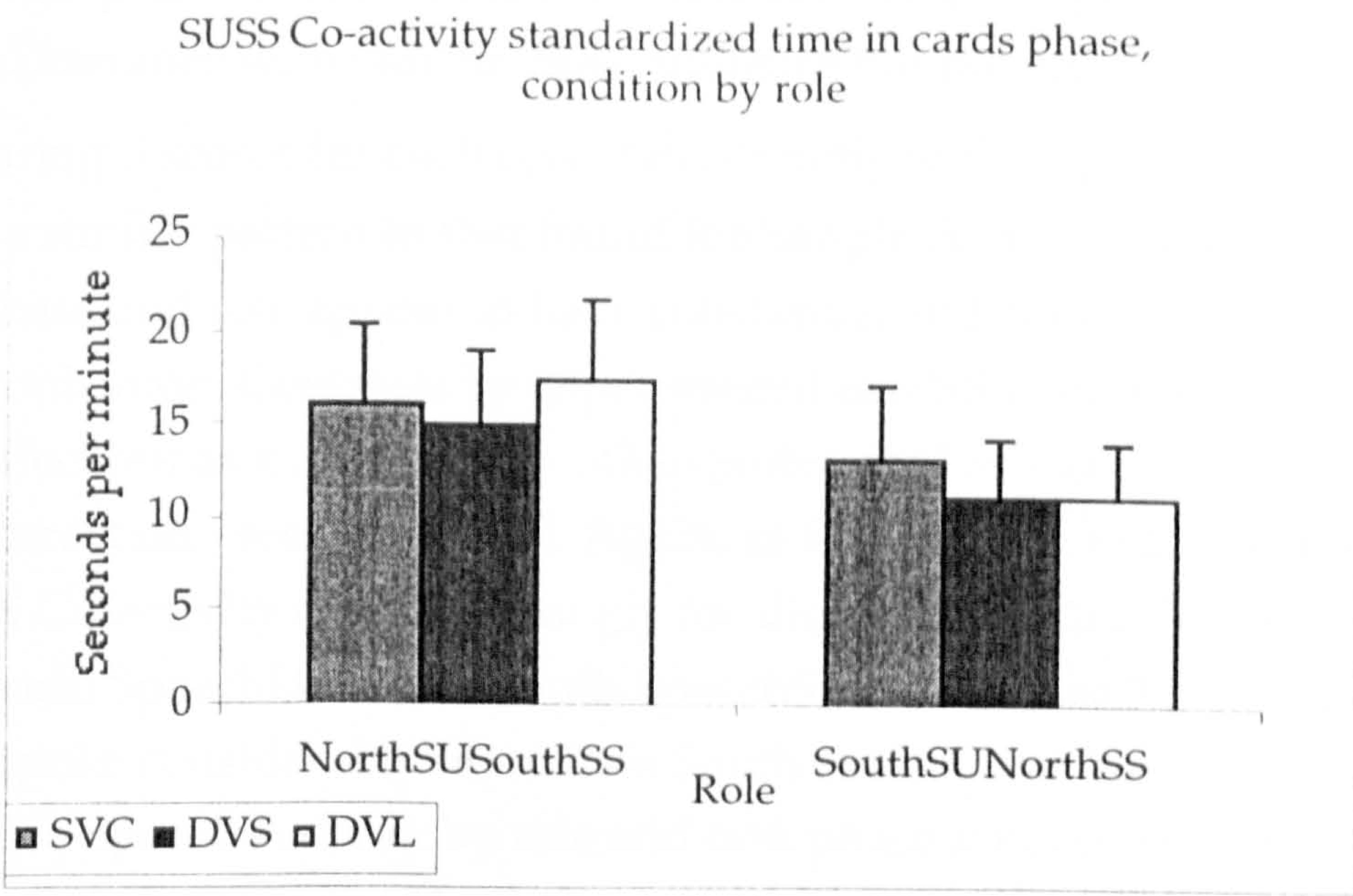


Figure 7.19: Cards phase observed standardized time in between-participants SUSS Co-activity, contrasting role-in-SU with condition

Between-participants SUSS , Cards phase	NorthSU SouthSS			SouthSU NorthSS		
Condition	Observed	Expected	"S"	Observed	Expected	"S"
SVC	16.29	16.59	-0.161	13.26	13.10	0.029
	(4.11)	(4.63)	(0.376)	(4.14)	(4.36)	(0.215)
DVS	15.01	16.31	-0.512*	11.13	11.81	-0.216
	(4.08)	(4.54)	(0.532)	(3.28)	(3.23)	(0.314)
DVL	17.61	18.72	-0.353*	11.33	11.25	0.016
	(4.23)	(5.14)	(0.349)	(2.83)	(3.35)	(0.220)
Overall	16.37	17.31	-0.351	11.82	11.97	-0.057
	(4.09)	(4.70)	(0.428)	(3.35)	(3.52)	(0.267)

Table 7.25: Card task phase between-participants SUSS observed and expected standardized time and mean S score, experimental condition by role-in-SpeechUtterance

As Figures 7.18 and 7.19 show, levels of SUSS Co-activity were markedly different for roles in the DVL condition but only in the discussion task phase. Tables 7.24 and 7.25 show that the SVC participants' levels of SUSS Synchronization do not differ from chance. DVS and DVL subjects on the other hand show good evidence of this form of Synchronization. In the cards phase, North participants negatively Synchronize their SpeechUtterance with South SpeechSilence. In other words, the turns at talk of participants are more closely coupled than chance alone predicts. This is also true for DVL in the discussion phase but not for DVS, where DVS Souths Synchronize their SpeechUtterance with Norths' SpeechSilence but not vice versa.

Comparing *S* scores for each condition directly with a split-plot ANOVA shows a similar pattern to that found for Simple Analysis of Speech Activity. Task phase and role appear to have substantial and general effects on SUSS Synchronization. Contrasts by experimental condition do not show up as a main effect but as a function of both experimental role and task phase (three-way interaction - see Table 7.26). Again, as Figures 7.18 and 7.19 show, levels of SUSS Co-activity contrast strongly for discussion-phase DVL Co-activity state North SpeechUtterance South SpeechSilence. Figure 7.5 shows that North spoke considerably more than South overall and that the general form of relative Speech Activity by role and task phase for each condition is similar to that in evidence in the SUSS Co-activity proportions. This Synchronization result shows that the Simple data do not tell the whole story. A similar explanation might be offered to that proposed for the GazeVideoSpeechUtterance Synchronization findings: North's role status and responsibilities may have been more influential on North's behaviour with a large dynamic image than with a small dynamic image. Since it is *negative* Synchronization between SpeechUtterance and SpeechSilence, it suggests that the utterance patterns of DVL dyads in the discussion phase were more fluid than their counterparts in SVC and DVS. This then does provide some support H17, that the formality of speech style interacts with video image and experimental task, but only as a function of role. Similarly, H16 suggested that DVL would be associated with more fluid exchange than DVS but this relationship was not found to be a straightforward.

Split-plot ANOVA for SpeechUtteranceSpeechSilence Co-activity

Source of Variation	SS	DF	MS	F	Sig of F
Tests of Between-Subjects Effects.					
WITHIN CELLS	4.14	18	.23		
CONDITION	.66	2	.33	1.43	.265
Tests involving 'PHASE' Within-Subject Effect.					
WITHIN CELLS	1.49	18	.08		
PHASE	.47	1	.47	5.66	.029*
CONDITION BY PHASE	.34	2	.17	2.05	.158
Tests involving 'ROLE' Within-Subject Effect.					
WITHIN CELLS	3.16	18	.18		
ROLE	2.30	1	2.30	13.10	.002**
CONDITION BY ROLE	.76	2	.38	2.16	.145
Tests involving 'PHASE BY ROLE' Within-Subject Effect.					
WITHIN CELLS	.47	18	.03		
PHASE BY ROLE	.05	1	.05	1.80	.197
CONDITION BY PHASE BY ROLE	.41	2	.21	7.93	.003**

Table 7.26: Split-plot ANOVA of S score for SUSS Co-activity, contrasting experimental condition by role and task phase

7.7 Discussion

This Chapter has attempted to use Activity Set Analysis to shed some light on the problem that set up the treatment of interactive behaviour promoted by this thesis. It adds to the "natural history" of video-mediated interactive behaviour reported in Chapters Four and Six by comparing operationalizations of gaze distribution and conversation in three alternative configurations of communication equipment. Forty eight members of the general public worked on two simple joint tasks, in pairs, via VMI links. All participants used the same high quality audio links and saw images of one another's head-and-shoulders in a graphical window on a computer's VDU. Sixteen subjects (eight dyads) saw a relatively large (103 x 140mm), live image; another eight dyads had a live, small (40 x 65 mm) image, and the remaining eight dyads saw a static and large (103 x 140mm) image. Dependent variables were Gaze Deployment (looking at the video image or elsewhere) and Speech (speaking or silent) Activity.

Two SVC dyads and one DVS dyad were excluded from the analyses due to behaviour that was considerably at odds with the other dyads in each condition. The South role in the DVS dyad engaged in state GazeVideo during the discussion very infrequently (mean=0.87 per min., group mean=6.37 per min., SD=2.74) but for extended durations (mean 16 seconds, group mean 1.02 seconds, SD=0.31). The two SVC dyads are rather more interesting. In both cases, North participants in the cards phase GazeVideo mean durations and standardized times were very extended (13.18 and 40.42 secs. per min., group mean=2.88 secs. per min., SD=4.00). This suggests that these North roles in the SVC condition during the cards task phase had a rather different attitude to the still image of their remote partner than their counterparts in other dyads. It is curious that similar differences were not also seen for South roles in SVC. Why this was is a matter for speculation but it does suggest that there may be some value in a static image. The way it was treated by these two participants was so very different from not only the other SVC but all other participants, it is clear that it represents a very different kind of resource.

In the first place, the dynamic configurations show massive differences in the number of times participants looked towards their images, the length of each glance and the overall amount of time "invested" in inspecting one another. This finding in itself provides some small indication that orienting towards a surrogate for another person is not an automatic process, driven by talk. Speech Activity proved to have a complex relationship with experimental condition, role and task. There is some indication that the DVL condition prompted more talk from North roles but in the discussion phase rather than the cards task phase.

There was no evidence of any Synchronization of mutual looking behaviour, regardless of experimental task or condition (no GVGV Synchronization). In contrast, for DVL participants there was strong evidence to suggest considerable negative Synchronization between looking at a partner whist they were not looking back, but elsewhere (GVGE). This effect was limited to the North partner in DVS and entirely absent for SVC. Whilst it is unsurprising that no GazeGaze Synchronization was found in SVC, that DVS and DVL contrast in this way is notable. However, direct comparison of *S* scores suggest that the difference between DVL and DVS Synchronization is subtle.

One of the strong findings reported in Chapter Six was of between-participants positive GVSU Synchronization: listeners look at speakers at greater than chance levels. SVC dyads showed no evidence of such Synchronization, regardless of task or experimental role. This is a very important finding in its own right, as it suggests that orientation towards one's partner in a "virtual meeting place", as represented by the computer workstation's image, is not an automatic, dialogue-driven process. One might argue no such conclusion may be reached, as differences could be fully accounted for by the significant audio configuration differences between those used in the previous study and those reported here. However, DVL and DVS North roles both showed just this positive Synchronization in the discussion task phase. Task demands for the cards phase were such that one might easily see how the interaction dynamic would be considerably different from a loosely structured discussion. It is less easy to account for South roles failing to show this sort of Synchronization. Inspection of Table 7.5 shows that North roles always spoke for longer than South roles and hence, as the Speech Activity Set is binary, South was in SpeechSilence for longer than North. Since the North role was "in the driving seat", dealing with the computer navigation and securing agreement from South, this is much as one might expect. However, GazeVideo levels were broadly comparable between roles (see Table 7.10). The social dominance literature reviewed in Chapter One suggested that dominant individuals in conversation tend to look at subordinate individuals rather more than vice versa. It may have been that the artificially raised status of the North role was reflected in this greater-than-chance gaze at South. South did not have instructions to submit to North's direction but was merely obliged to work at the pace of the intervening technology.

Following the failure to demonstrate within-participant GVSU in Chapter Six, the issue was re-examined. Again, there was no evidence to suggest that speakers (those in SpeechUtterance) looked at their co-participants (as GazeVideo) at greater or lesser levels than chance alone predicts.

The formality of speech style has much occupied the literature on VMI, with some contradictory findings (Boyle et al., 1994; O'Malley et al., 1996; Sellen, 1992). Synchronization measures did not show any broad-brush discrimination of task phase or experimental condition, as was anticipated. However, agreeing with the Simple Speech Activity data, a complex relationship between role, image and task was in evidence. In a certain context, dyads using a relatively large dynamic video image showed evidence of more fluid or "informal" interaction than dyads with either a small dynamic image or those with a still of their remote partner. The context was carried by participants with responsibility for controlling a shared device and securing agreement with a remote partner during an early period of sociable discussion but not during a late period of stereotyped exchange. It seems that under some circumstances only, the larger PC GUI dynamic image is capable of reflecting a conversational norm previously found for VMI with a life-size image.

Task type and order of experience were confounded as a single "task phase" variable. However, the nature of the discussion was inherently introductory. Conversations are usually associated with opening conventions, whether at work or socially, prior to activities of various kinds(Daly-Jones, 1998). In a mediated communication these can be problematic. The first phase was thus not simply identified by coming before the cards task, or by discussion, but as a first point of contact in a novel environment. That the results at least partially map onto findings from the twenty minute discussions reported in Chapter Six suggests some generality beyond the first three minutes of meeting. Nevertheless caution should be exercized in extending the findings from this study to other forms of interaction by VMI device. One might imagine that these short interactions, mixing introduction with a concrete if undemanding task, may bear similarities to service encounters such as visits to a bank. In these cases, control status and responsibility are usually clearly defined, perhaps in a similar way to the North/South distinction here.

8 *Conclusion*

To conclude, a summary is given of the contentions and findings of this thesis. Its limitations are then discussed and further work suggested.

8.1 Summary of thesis

Communication at a distance can be achieved in a number of ways: writing a letter, leaving a note, making a telephone call or with semaphore from ship to ship. In each of these examples, it is clear that effective communication behaviour is conditioned by the tools used to carry the information to those for whom it is intended. The constraints these tools impose upon communication, and to some extent the skills required to use them, are not difficult to describe. They include the effort required to set in motion and carry through the process of sending a message, determining the success of its reception, being able to tell whether or not the information has been understood by the receiver, and working out whether or not the recipient wishes to take it forward with a reply. These matters are usually invisible in ordinary, same-place same-time conversation. The skills are comprehensively overlearnt or even instinctive, so that their very fluency tends to mask their operation. This kind of fluid, powerful, subtle and very usable communication has been seen as a quality 'gold standard' for inventors of distance technologies. Consequently, there is a long history of attempts to achieve the gold standard by providing a moving picture of a remote person and two-way audio. Implicitly, they have intended that the well-adapted everyday apparatus of communication might be used without further ado. However, evaluations of such technologies have conspicuously failed to show any striking benefit for seeing the remote person over and above simply being able to converse with them.

This thesis began by reviewing studies of video-mediated communication and looking in greater depth at the kinds of communicative mechanism that might operate through a visual channel. Drawing from an extensive literature on non-verbal behaviour, gaze distribution was identified as a behaviour of acknowledged significance for regulating interactions and for creating and maintaining social structures. It was further noted that the role of gaze in human communication is intimately related with speech behaviour. Measures designed to expose the distribution of gaze with speech behaviour in interaction were implicated, on this account, and yet were conspicuously absent in the VMI literature.

The interactive work in seeing and being seen was argued to have been ignored due to the failure of the investigators' model of interactive behaviour, revealed through their impoverished operationalizations of performance within the time frame study. It was argued that the "fast cars" of gross performance measures were incapable of getting to the gold of visibility in interaction. They are simply the wrong vehicle for such a chase. A consideration of the characteristics of interactive behaviour followed, drawing heavily on HCI and CSCW research into human-information-artefact relationships. It was further argued that many previous investigators have failed to adequately differentiate the purposes of those engaged in communication, even within the experimental context. In other words, task context has rarely conditioned assessments of VMI interaction and yet is widely accepted as being quite fundamental to HCI and CSCW research. There are several schools of thought on interactive behaviour which take a strong position on matter of *integration* between person, tool and place. These include Distributed Cognition, activity theory and situated action theory. The question then arose of how one might go about understanding which elements of person, tool and environment are active elements in a given system of interactive behaviour; that is, how might one begin to map the parameters of an activity system. Commonly, an integrationist account of interactive behaviour places emphasis on intensive, unstructured participant observation and qualitative analyses. Indeed, these methods and mechanisms of report have an established place in contemporary design and requirements analysis, particularly in the discourse of CSCW.

Whilst these have much to offer, the empiricist foundations of Western psychology exact some specific requirements for investigative rigour that these methods struggle to meet. Methods should be inspectable, repeatable and subject to explicit safeguards against subjective interpretation. Commonly, these requirements are identified with laboratory control, strict operationalizations of concepts to frame hypotheses for testing, and analyses on statistical models. Armchair-ethnographers musing over their travel snaps might have a good story to tell but only about the house they happened to stay in. Yet an integrationist account of interactive behaviour suggests that some of these requirements are anathema: what is the point of "controlling environmental factors" when these take a full part in interactive processes? At best, a one-sided view is possible. At worst, a poorly informed account of a hamstrung activity system is presented as objective reality. What works for a model of a human organism as an independent mechanism cannot work when the mechanism extends to vitally include other reactive transformational instruments or places, or other people. More seriously still, to suggest that the outcome of an experiment run under strict control will have some value when translated to other circumstances is dangerously flawed. Unless some account is made of the manner in which the outcome came about, no predictions are realistic. Any endpoint reached by an interactive system is the consequence of a series of steps that might have taken a different direction at any juncture, as a function of any of the elements involved. Nevertheless, the charge of unchecked analytic subjectivity is very serious, and the typical riposte, that all claims to objectivity are flawed, deeply unsatisfactory.

An extension of a quantitative research paradigm was proposed as a complementary approach to interactive behaviour, to contribute to the insights only possible from situated intensive analysis. This is in itself a significant departure from the laboratory paradigm so central to Western psychology. It suggests that no findings are self-contained. A single study cannot produce definitive answers to any question; whether ethnographic or experimental, nuggets of truth are relativized. It can only contribute information to a discourse on the character of behaviour in certain contexts, revealing something about the methods by which the various aspects of an activity system adapted to one another. The discourse may then be appealed to, as the mass of literature, for guidance on design decisions. The days of the killer experiment are over.

To suggest that contemporary psychology rests on a notion of absolute truth is of course to make it a caricature, just as it is to suggest that sociology and allied participant-observation traditions are no more than idle opinionation. The substantive point here is that isolated study of individual production has no place in research on interaction: interactive behaviour is not an individual production. The challenge then becomes how to organize observations so that they might evenly account for process relationships between activity system elements. How might one begin to measure the dynamics of an activity system?

A method for observation was explained, based on a principle of grouping behaviours into separate temporal streams of activity. Grouping is determined on the basis of a relationship of state mutual exclusion, i.e. only one of the states described by the occurrence of a behaviour in the group can obtain at any given moment. Each of the behavioural groupings defined in this way is described as an Activity Set. Observations of behaviour are in principle at some distance from action from the actor's perspective. Activity Sets were so named as to emphasize their intended link with the material and integrated activity of an activity system. Activity Sets may then be examined in terms of their internal temporal characteristics, including frequency of occurrence, duration of member behaviour occurrence and overall amount of time spent in the state described by member behaviour. Taken together, the method for defining and examining data in this way, is presented as Activity Set Analysis. The process of data gathering and organization into composite transient states by Activity Set membership was formalized as a computer software tool, Action Recorder. Co-activity characteristics offer similar potential, with the additional potential to expose systematic co-occurrence of member behaviours. Activity Set Analysis was explored with a series of video recorded video-mediated interactions both for its analytic potential and to establish the practical application and reliability of Action Recorder-generated data. On the question used to frame this approach, the value of visibility in VMI, the occurrence of some gaze behaviour at any moment was examined for evidence of contingency on the states of those participating in an interaction. This contingency is expressed by the systematic co-occurrence of behaviours from different Activity Sets, given by the proportions of time in the composite state so defined against chance probability. The difference between chance and observed proportion of co-occurrence was presented as Synchronization Analysis, employing a new statistic, *S*. Synchronization is intended to expose where and how resources for interaction are recruited, subject to a variety of dependencies. An independent Activity might be engaged in parallel with other, collaborative Activities. Simply finding the presence of some Activity in association with an activity system is not therefore compelling evidence for its involvement. Synchronization Analysis is capable of revealing sensitive interrelation amongst on going Activities. Besides a demonstration of the approach, the case study additionally intended to provide a natural history of VMI in association with a live synchronous screen-share and accompanying documentation. Activity Set Analysis showed that, whilst most time was spent studying the shared screen, participants showed strong evidence of close coordination between their gaze and speech behaviour. When they looked at their video monitors, they were more likely than chance to be met by the gaze of their remote collaborator. As listeners and speakers, it was less likely than chance that they would consult computer or physical documents. So the role of the video link was systematically implicated as a resource in the interactive behaviour of these dyads.

Activity Set Analysis was then reported in application to an experimental study of audio-video communications equipment. Gaze and speech patterns were comparatively examined in communication mediated by high-quality audio and various images. The study contrasted qualities of video image defined by real-world desktop videoconferencing design constraints. A number of hypotheses were formulated as a basis for a functional discrimination of the value of the images provided by each experimental configuration. It was found that the dynamic video image configurations, DVS and DVL, were both associated with Synchronization amongst fundamental interactive behaviours: gaze distribution and utterance activity. In contrast, a plausible audio control failed to show many of the Synchronizations so in evidence. On this basis it would seem that seeing one's co-participant on a video link can be materially involved in a system of interactive behaviour.

Activity Set Analysis effectively discriminated between communications facilities hitherto having presented an almost intractable problem for investigators. Whether or not these findings constitute evidence of some gold in a VMI facility is an open question. I contend that they show how interactive behaviour is materially affected by the presence of at least some video-communications facilities, as at the very least the amount of time spent consulting written materials whilst in conversation is impacted. If it is reasonable, as I believe, to assume that this diversion of gaze time away from written materials during dialogue is indicative of those involved feeling a gain from seeing one another, then yes, Activity Set Analysis has managed to unearth some of Giacometti's gold.

The title of this work is "Understanding interactive behaviour: a quantitative approach". Thus far, the reader may feel satisfied that the subtitle has lived up to its promise but wonder about the contribution to "understanding" on offer. Activity Set Analysis is not presented as a panacea for research on interactive behaviour. It is one of many approaches that focus on the processes by which interactive behaviour may be carried out. Activity Set Analysis contributes to this battery of available analytic techniques in two ways. The first is to do with its *explicit and formal focus on contemporaneity* as a key to the interdependence of evident behaviour. There are no other formal techniques to expose such relationships, although it is not unusual to plot data from separate behavioural streams to "eyeball" for links or relationships between them. For example, Haccou (1987) suggests plotting and inspection as a means to identify inhomogeneity in an ethological record, to split the data into candidate segments for tests for sequence stationarity. Explicitly qualitative analyses, such Conversational Analysis, might transcribe a number of simultaneous behaviours, typically talk and intonation, or gesture (see for example, Heath & Luff, 1995). Informal inferences might then be made on the basis of co-incident behaviour and contextual knowledge of the situation in which the analysed conversation took place.

The second aspect is to do with its square commitment to a school of thought that emphasizes the very interdependence rather than independence of 'behaviours' in an interactive setting. This is an important matter as the distinction between statistical requirements for independence and theoretical requirements for the purpose of analysis are easily clouded. These two sets of requirements should be orthogonal to one another and yet are rarely so clearly separable (Silverman, 1985; Monk & Gilbert, 1995; Watts, Monk & Daly-Jones, 1996). The process of selecting criteria for dividing up the stream of activity is a vital part of ethological analysis and yet easily dismissed as merely a preliminary to the real work of measurement (Martin & Bateson, 1993). Ethograms purport to describe stereotypical behaviours of species and thus to provide ethologists with ready-made behavioural categories for those species. Unfortunately, even if ethograms were available off-the-shelf for all species, large within-species individual differences seriously threaten their validity. Martin and Bateson point out that any successful analysis hangs on effective preliminary observation, an exercise that is primarily qualitative. Activity Set Analysis is intended to be applied only in association with the kind of exercise of familiarization for which they argue.

A theoretical perspective on interactive behaviour is, by definition, fundamentally concerned with reciprocal connections: if some elemental behaviour has no connection with other on going behaviours, it is quite simply not interactive. That makes interactive behaviour especially sensitive to the matter of division into categories of 'atomic' independence. It is instructive to revisit the problem of contemporaneity for Markov models here. These models are premised on a capacity to identify behavioural categories that may not, under any circumstances, be allowed to overlap in time. A notion of overlap is impossible to represent in the state-space underlying the model. Thus the independence of the coded categories of behaviour is accompanied by an analytic imperative to maintain categorical independence, severely limiting the nature of inferences about relationships between categories. Activity Set Analysis is similarly based on the definition of a collection of exclusive and exhaustive categories of behaviour, as the collection of Activities comprising Set. The subsequent analysis is not then dependent on mutual exclusivity in the same way as the calculation of Markovian transition probabilities[48]. Combinations of Activity Sets, in principle, could fluidly identify any number of Co-activity states. Each of these might be taken to be observationally independent, in part satisfying criteria for building a continuous time Markov model. Unfortunately, there are strict mathematical dependencies given by the baseline proportions of time in Activity State, as discussed in Chapter Five. These are not tolerable for a Markov model, where state-categories must not in principle mutually constrain one anothers possibilities for expression (see Haccou, 1987, for extended discussion). Indeed, that was the problem that prompted formulation of the S statistic. Yet, regardless of these mathematical considerations, Activity Set Analysis was conceived as a technique to expose integration, where it occurs, rather than regularities in segmentation or isolated behaviour. It is not accidentally true that Activity Set Analysis has these properties.

So, if Activity Set Analysis is proposed as an addition to the "researcher's toolbox", as Haccou described it, how is its deployment anticipated? In the first place, Activity Set Analysis requires contextualization for the behavioural Co-activity states it addresses. Little has been said about this in the foregoing pages, since the primary aim was to demonstrate how to generate and analyse data in a novel way. There are many ways in which one might contextualize Activity Set data to contribute to building an understanding of the interactions concerned. Silverman (1985), for example, argues that the qualitative approaches common in sociology are lacking an important element

[48] The state-space must be finite and also it must be possible for any state to be visited from any other state, known as 'ergodicity'. Furthermore, the extent of any state within the period of observation must not be predictable from any other state. This predictability can be in terms of mutual constraint of expression, given by relative proportions of time in states. The Synchronization statistic, S, described in Chapter Five makes explicit provision to deal with this kind of predictability and so Activity Set Analysis in association with Synchronization Analysis is freed from this constraint.

of validity, unless supported by more transparently objective and replicable data. He thus takes a position that alternative analytic approaches should be commensurate with generating an integrated picture of the kind of behaviour under investigation. To put it another way, it should be possible to "triangulate" on a rich understanding of some interactive scenario by drawing on empirical information sources of different kinds. Claims about the role of some activity in an activity system, on the basis of intensive qualitative investigation of selective excerpts of interaction, might be partially validated by formulating expectations of certain kinds of behavioural patterns. These expectations might well be appropriately tested within an Activity Set Analytic framework.

Activity Set Analysis brings with it a degree of contextualization by virtue of its potential for nesting Activity Sets within one another. Chapter Seven attempted to do this by considering task phase as a "context" for interpreting the Speech and Gaze Activity Sets. Similarly, role was introduced as an element of analytic contextualization. The period and number of task elements could in principle be expanded indefinitely.

Finer-grain divisions of this kind are easily imaginable. For example, Conversational Games Analysis (CGA), developed at the Human Communication Research Centre, has been employed to examine video communication patterns in spoken dialogue (Boyle, Anderson & Newlands, 1994). The CGA strategy is to code segments of talk into well-defined functional units. The two levels of analysis they employ, "game" and "move", could be used as Activity Sets for talk without further ado. These might then be used to condition statements about gaze deployment as a refinement of the crude Speech Activity Set used in this thesis. It may be that some kinds of move in one kind of game are more gaze-intensive than others, in the context of a VMI investigation, implicating the exploitation of a video link as variable by the nature of communicative act involved.

In HCI, many descriptive strategies for interactive behaviour have evolved, as the basis of computational or analytic models of interaction process (see Monk, 1998, for a discussion). These decompose an interactive sequence into functional components according to a set of rules. As with CGA, recurrent sequences could be used to formulate Activity Sets. In both cases, an understanding of the interaction concerned would be a function both of the contemporaneity determined and the explanatory framework given by the modelling or CGA technique.

Activity Set Analysis represents a novel way of examining behavioural organization in time, with general applicability for any situation where relationships between co-occurring behaviours or environmental states are of interest.

8.2 Limitations

8.2.1 The performance and process distinction

An early motive for the work reported here was surprise at the negative outcomes of the majority of VMI research. They operationalized conceptions of better and worse communication as faster or more and slower or less. Failing to find evidence of this kind they decided that a video component has little benefit in addition to synchronous audio. Statistical tests do not allow for any statements to be made on the basis of a null result and thus the weight of evidence in this outcome-based set of measures is, in principle, very slim indeed.

However, outcome measures are motivated by a concern to find effects of material importance to the experimental tasks and thereafter generalizable to communication at work. One might just as easily turn the argument around to assert that the determination of process differences does not entail any benefit for the VMI participant. Differences are not always of practical significance. One must then fall back on a model of the behaviour concerned to form a judgement on the value of the findings.

The trouble with outcome measures is that they are invariably collected in artificial settings over short periods of time. The people taking part have no vested interest in either their success or failure, are unlikely to be concerned about putting themselves under considerable pressure for the limited duration of the study, and do not have to fit the requirements of the study into the many and varied demands running right through everyday life. In many forms of psychological investigation one might reasonably claim that the factors contributing to a result are beyond the control of those taking part. The behaviours appropriate for such investigation are very close to physiological matters. It is difficult to conceive of interactive behaviour in quite this mode. At work, the level of time or other pressure under which people operate is a basic priority. There are many kinds of work where an individual might have strong feelings about and influence over the outcome of their activity. Although, inevitably, certain resources recruited for the pursuance of interaction may be comprehensively overlearnt to the point where they are effectively automatic, the business of their recruitment itself cannot be automatic in quite the same way. Even if it were so for a particular individual, the factors determining recruitment and deployment of resources are extrinsic as well as intrinsic to them. For this reason, it is vital to have some way of understanding how and why certain kinds of behaviour have contributed to a result. One might then begin to ask questions that apply to real situations.

This thesis began by taking some trouble to inquire why seeing and being seen might be implicated in communication processes. These were used to help operationalize interactive behaviours and to formulate hypotheses about the forms and relationships between them. Of course, there is nothing radical about hypothesis formation on the basis of a literature review. On this occasion, the guiding principles in so doing were to take account of how one might relate process findings to different applied situations.

It would be quite inappropriate to contend that the findings of Chapter Seven entail measureable benefits from a business perspective, that having a fairly large GUI video window makes an important difference. For example, there was no difference in time taken to follow the introductory procedure. Chapter Seven indicated some effects sensitive to differential status in semi-formal time-limited encounters, to the effect that a larger video window seems to reflect status differences. One might then begin to argue about the importance of status and interaction control with regard to a number of issues and information yielded by other methods of enquiry in such a context. In any case, the approach promoted here is incapable of providing a definitive demonstration of the advantage of one set of interactive resources over another.

8.2.2 Observed Activity and systemic activity: Activity engagement and activity value

Three base measures of Activity were proposed with a simple model for their interpretation. Duration data were suggested to provide evidence of the length of time some Activity needs to be engaged such that some value is obtained from it, motivated within its interactive context. Frequency data were described as an expression of the number of occasions a motive for initiating such an Activity reached a point of execution. Standard time in Activity state was assumed to reflect the overall investment in each of the observed Activities, for the period of their recording. These may be expressed as:

$$\text{value}(A) \propto f(\text{time}).f(\text{number}) / f(\text{effort})$$

Were some reasonable formulation of the "effort" term possible, such a model would allow one to infer the value of observed Activity to an activity system. Yet this is a crude characterization of Activity engagement at best. Reading requires more time than surveying a scene, independent of the information value therein. One might question whether an Activity engaged once for four seconds is of the same, more or less value than another engaged twice for two seconds. Devoid of further context, the comparison does not make sense. Clearly, there are dangers in extrapolating information of this kind to compare different Activities. Activity Set data are only meaningful in so far as they contribute to a model of the underlying activity system, so value of Activities can only be inferred with reference to their place in the wider system.

8.2.3 Activity engagement and activity influence

Activity Set Analysis limits its treatment of activity to what may be seen. Activities are recorded as beginning and then as ending. In the underlying activity system, they may persist by function or intention for somewhat longer, or even be unbroken. That said, the coding rules for SpeechUtterance were explicitly designed to avoid a rigid categorization of vocalization. It was intended that the functional floor-holding period be recorded, including non-interrupting pauses and so on. In this way, the Activity definitions themselves might serve to distance the constraint of real time from functional time in Activity state.

8.2.4 Observed Activity and systemic activity: Interpreting Activity relationships

Thus far, it has been suggested that evidence of Synchronization provides a strong case for supposing that the Activities contributing to the Co-activity state are implicated in the underlying activity system. This is not at all the same thing as asserting that failure to find evidence of Synchronization entails lack of Activity involvement.

Wagner, Clark and Ellgring(1983) made a distinction between 'tonic' and 'phasic' aspects of gaze behaviour. They argued that participants in an interaction might well be prepared to maintain individual looking levels at a high enough level to optimize mutual gaze. In other words, there is a desirable level of mutual gaze understood by people in conversation and they simply continue to look at one another more or less at random until that level of eye contact has been achieved. There simply is no above chance eye contact, as chance occurrence has been "deliberately" employed to manage the process. The cost of investing effort in baseline interpersonal looking is born in preference to the cost of accurately judging moments when gaze of another person would be directed back. They suggest on this basis that finding a near-expected co-occurrence of gaze may mask the purposive nature of interpersonal gaze.

Another problem with Simple Activity and Co-activity interpretation concerns the independence of Simple Activity measures. Measured Simple time in state is taken as the basis for determining organization between Activities. The Activity Set Analytic approach to contemporaneity assumes that the Simple proportions of time in state have a value that is to some extent independent of their contribution to Co-activity states. This generalization is justifiable in so far as an assumption that the Activities concerned have some prior theoretical significance in themselves. So, in the case study example, it was assumed that there is some value to the Activity GazeNotes in itself, as apart from any value added by coordinating that Activity between individuals or with the Speech Activity of a particular individual. It was suggested that only observable behaviour likely to be related to the underlying activity system should be recorded. One might imagine some Activity that makes no sense in isolation but that is of great value in combination with a second. For example, were one to record foot Activity in driving a car as one Activity Set and upper body Activity as a second Activity Set, from an understanding of driving one would not expect ActivityEngageClutch to be independent of ActivityMoveGearstick. ActivityMoveGearstick cannot (normally) occur without ActivityEngageClutch and so it has no theoretical value in isolation.

8.2.5 Environmental Activity

The final limitation to the work reported in this thesis is more of an omission than a fundamental flaw. Chapter Two was at great pains to assert that tool and other environmental states were as important to a system of interactive behaviour as states of an individual. The empirical studies went on only to examine states of individuals. This was largely a consequence of the VMI paradox setting the context for the thesis. Much might have been gained by including computer system states, such as document display and mouse activity, with the screen-share in either case.

Interpretation of Activity measures, as discussed above, might be argued to have no translation where Activities cover entities such as "room", work system or computer system. I would argue that there is much to be gained by conceiving of any interactive entity in the same terms as a person. The complexities of interactive behaviour, the contingencies of action, modal constraints of operation and so on, mitigate against placing any activity system element in a privileged position. For understanding interactive behaviour, it makes better sense to step back from such matters, admitting a unit of analysis no smaller than the active elements therein.

8.3 Future work

8.3.1 Synchronization of 'n' level Co-activities

Chapter Six made some comment on the desirability of discriminating periods of dialogue from periods of silence, in relating Gaze Activity of the three defined types to interaction processes. As Synchronization Analysis is capable of answering questions only about pairs of Activities, this could only be achieved by inference across a number of Co-activity states. Not a very satisfactory state of affairs. It would seem that there is a strong case for extending 'S' to cover Co-activities made up of an indefinite number of Activity components. As the model for determining the extent of Synchronization against chance occurrence is strong, this would in principle be a useful step to take. However, the value of any data contributing to such an analysis might be rather more difficult to establish. Although inter-rater reliability was very good for the observed Co-activity proportions feeding into the *S* procedure, it is an open question as to whether or not three- four- or n-level combinations of derivative from an Action Recorder record would be capable of showing similar levels of agreement. That said, other interactive contexts may well not involve a pace of exchange as high as synchronous dialogue.

8.3.2 Simultaneous and sequential dependencies

No attempt was made to discuss intra-Activity Set relations. The 'S' measure is explicit in its emphasis on synchronous/antisynchronous consistency as positive and negative Synchronization, but is only possible for inter-Activity Set combinations. In part, this reflects the status of Activity Sets as merely observational conveniences and not as of any great theoretical significance. In other words, whether Activities come from the same or different Sets is an irrelevance for the underlying activity system. It is nevertheless quite possible that the sequential organization of behaviour could be revealing of interactive behavioural contingencies. Indeed, causal chaining has been promoted as the dominant explanatory model adopted in psychological explanation since its inception as an independent science. See Cummins (1983) for an alternative perspective.

As discussed in Chapter Two, Markov chaining approaches sequential dependencies by looking at the probability of next occurrence in a sequence. Thomas et al. (1983) have applied it to informal dyadic conversation, claiming that it is a tenable model for information transmission. They derived a model to describe a number of conversational sequences and strategies instrumental in gaining, retaining and relinquishing the speaking turn. Interestingly, they also claimed that the model was sufficiently flexible to produce accurate predictions of the order of conversational events, even when violating key mathematical assumptions behind the Markov model. As previously discussed, Markovity relies on sample homogeneity. On the one hand they showed, in a likelihood-ratio test for transition matrices, that nearly half their dyads fail for stationarity and additionally were not homogeneous, whilst on the other claiming that transitions are somehow consistent and meaningful. So although Thomas et al. seemed happy that their empirical data demonstrate the irrelevance of these assumptions, one might question how far it is possible to generalize Markov sequence modelling to other sequences similarly violating the basis for the procedure. Clearly, conversation in particular and interaction in general is characterized by change and phase. See, for example, the Conversational Games Analysis work of Boyle et al. (1994).

Besides stationarity these issues, Markovity seems to be bound up with application to series of events: even the duration information retained by continuous time approaches is not permitted to persist into the display of the next logged behaviour. As discussed in Chapter Five, amount of time in state is not to be lightly dismissed. A glance of 0.5 seconds is quite simply not easy to equate with a gaze of 20 seconds. How might one assess sequence when notional events constantly overlap? It might be possible to adapt Markovian process models to suit the chains of time-slice style state-descriptors (see Appendix 1d). Extreme caution would be needed as there are so many dimensions to the states that any predictions would be very difficult to generate. There is tremendous potential here for generating fundamentally uninteresting data. Treating each state descriptor as conceptually equivalent would be highly unsatisfactory. Similarity weightings would be required, not just in terms of the number of matching elements but also in psychological/theoretical terms. The problem of relating the number of Activities comprising a psychological activity is not straightforward. Similarly, the number of psychological activities contributing to the behaviour patterns of the individual, or group, are unknown. Even if there was a realistic solution to deciding on substates for inclusion in sequence models, the problem is likely to be extremely demanding in computational terms.

An alternative approach might be to condition the analysis of some Activities so that they conceptually persist beyond their coded end point. It could be argued that a given Activity have a temporal "zone of influence" beyond its actual state duration. Suppose you and I are having a chat at a table. There may be several objects on the table, including a number of documents. As I talk, I glance briefly at one of these documents. You react to my glance by glancing in turn to the document on the table and then back to me, having registered which one I have "indicated through glance". As it stands, Synchronization would only compute the duration for which we both were simultaneously gazing at the document. In terms of the conversation, I would like to suggest, we would have synchronized (small 's') on the document for somewhat longer. Formulating rules to determine such a 'zone of influence' would be far from straightforward, but there may be some reasonable grounds for usable approximations. For example, the duration of speech contemporaneous with the glance might serve to carry that influence. Adding a level of sophistication, it might also be reasonable to think of a limit to such a zone within the speaking time of the glancer, given by the occurrence of subsequent gaze activities not including the interpersonal gaze activity (e.g. to a second document, or out of the window/looking for inspiration).

Leon Watts

Kippax, West Yorkshire 26th September 1998

Appendix 1a: Full User Manual for Action Recorder

A1a.1. Overview of Action Recorder

Action Recorder is a computer tool for recording several sets of things that occur over a defined period of time. These "sets of things" are collections of various aspects of a situation under observation. In short, it is a tool for recording changes, as they occur.

Action Recorder collates time-series data (indicators of things that have happened, with the time at which they happened). By applying some limited knowledge about the status of each recorded event, Action Recorder can produce organised data, in a table format, suitable for use with SPSS™, a powerful and commonly used statistical analysis package.

Action Recorder has been designed primarily as a aid for the analysis of video recordings, although it may also be used in a limited way for real-time, "as it's happening" observation on a portable Apple machine. In this user manual, the focus on interactions between people is for convenience. The function of Action Recorder is to determine whether or not there *is* any evidence of an interaction[49], rather than some simple and unrelated co-occurrence of activity. For "individual", a reasonable and generic substitution might be "agent", which might be any identifiable source of influence. For one investigator, an agent might be a whole office, viewed as an integrated and coherent source of influence, in an organisation; for another investigator, a thermostat in a building might usefully be thought of as an agent, by influencing the temperature of a room by sensing that temperature.

A1a.1.1. A State-based Description Principle

The whole philosophy of Action Recording is organised around an idea that there is no such thing as 'nothing happening just now'. For this reason, it would be inappropriate to describe Action Recorder as an 'event recorder'. It is true, in the first instance, that this is just what Action Recorder does: allow

[49] The detection of interaction here referred to is with specific reference to the behaviours being observed, rather than in any more general sense.

the recording of events. But unlike most event recorders, Action Recorder really doesn't care about them as events in themselves. The most important thing about them is that *they represent the start of some state of activity* and hence are used only to bring up to date a record of the general state of the world.

Mutual Exclusivity and Event Recording

Observations are made by striking ordinary QWERTY keyboard keys. Each of the activities of interest is linked to a unique key, or key combination. Since Action Recorder is case sensitive (i.e. it can tell the difference between capitals and small letters), and the 'Option' key also makes a difference, Action Recorder can deal with up to around a hundred different activities for any given interaction.

The way that Action Recorder can interpret these keystrokes (the 'limited knowledge' referred to above) is through the use of a very simple principle, called mutual exclusivity. Mutual exclusivity means that, of a collection of things, only one of those things can be in place at a time. So, the very first step with Action Recorder is to decide what kind of behaviours are to be observed and then to group them together on the basis of mutual exclusivity. Take for example, the visual focus of a person and the things they say. A person can look in all sorts of place in a room, but only one place at a time. A person cannot both be speaking and silent at the same time. Groups of mutually exclusive activities are called "Activity Sets."

A1a.1.2. Understanding Activity Sets

Activity Sets are very convenient things for observational purposes since information about any one of the activities at any moment in a given Activity Set provides all the information there is about all of the activities in that Activity Set. Activity Sets are observational entities: they refer only to what is or is not mutually exclusive from an observer's point of view. So, whilst as an investigator I might be happy that John and Mandy's glances at one another are both instances of the same type of activity, and may very well have a strong relation to one another, as an observer I recognise that they may be observed independently of one another.

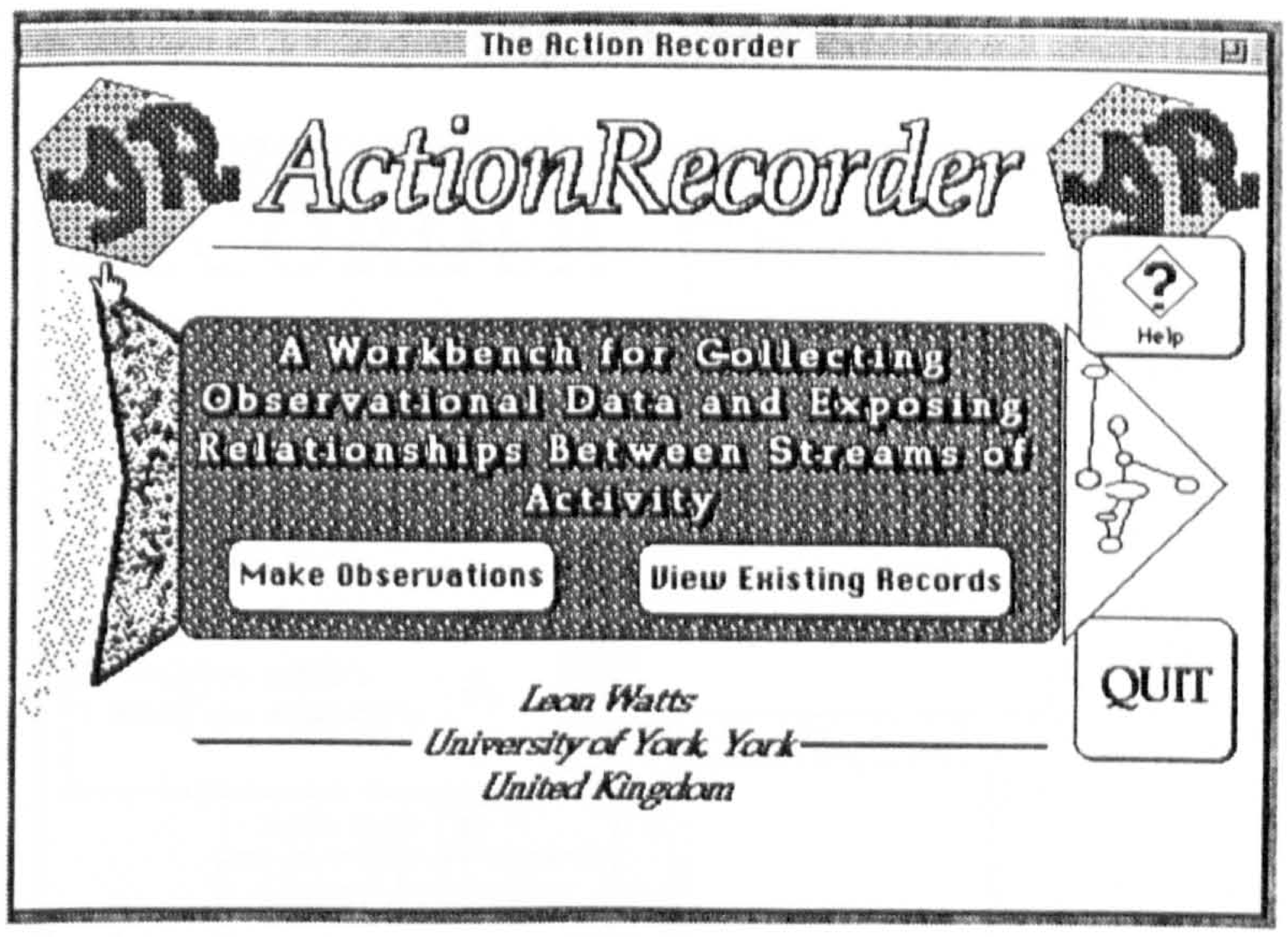
It may be helpful to consider the following example. Imagine a conversation between two people, John and Mandy, over a transatlantic videophone. John and Mandy have to prepare a report on a joint venture between the US and UK offices of their international corporation. They are able to talk to one another quite freely, but are additionally able to see one another and to show one another things over the video link. Their company is generally keen to find out whether or not the videophone offers anything over and above the far cheaper, telephone conference alternative. In this circumstance, it might be interesting to see how much time, and how often, John and Mandy make use of the video link versus the papers they have brought along to help them in their task. In this case, the room is set out so that looking at the papers precludes looking at the video link at the same time i.e. looking at the video link and looking at the papers are mutually exclusive. So these are two observable Activities which can belong in the same Activity Set.

Activity Sets "belong" to particular agents in an interaction: it is quite possible for John to be looking at Mandy whilst she is consulting her papers, or looking back at him, or somewhere else. So activities of various agents cannot be mutually exclusive with respect to one another.[50] For the purposes of this manual, John and Mandy's 'LOOKING' Activity Sets are identical to one another. This need not inevitably be true. Activity Sets refer to the potential activities of agents. If some agents have more varied activities, then their respective Activity Sets will be correspondingly more diverse. It may have been the case, for example, that Mandy had a computer terminal to access a company database that John did not. Were that true, Mandy's LOOKING Activity Set would have included looking at papers, looking at the video link and looking at the database.

[50]Again, it is important to understand that this independence is an *observational* matter from Action Recorder's perspective.

A1a.2. Starting Action Recorder

To start Action Recorder, position the mouse pointer over the  icon and press the mouse button twice in quick succession ("double click"). A welcome screen will appear which allows you to begin recording some new data, or to examine some previously recorded data.



There is also some on-line information about Action Recorder available from the "Help" button.

A1a.3. Three Phases of Data Handling

Action Recorder is a simple animal at heart, preparing data for a particular kind of analysis rather than doing any statistics in its own right. So in that sense, it is a data manager, and is capable of dealing with a great deal of data. It is a partial implementation of a process-oriented empirical methodology, known as the Activity Set Methodology (for further reading, see bibliography). Statistical summaries of data are then provided by SPSS™ for the Macintosh, version 4.0, via SPSS command and data files generated automatically by Action Recorder.

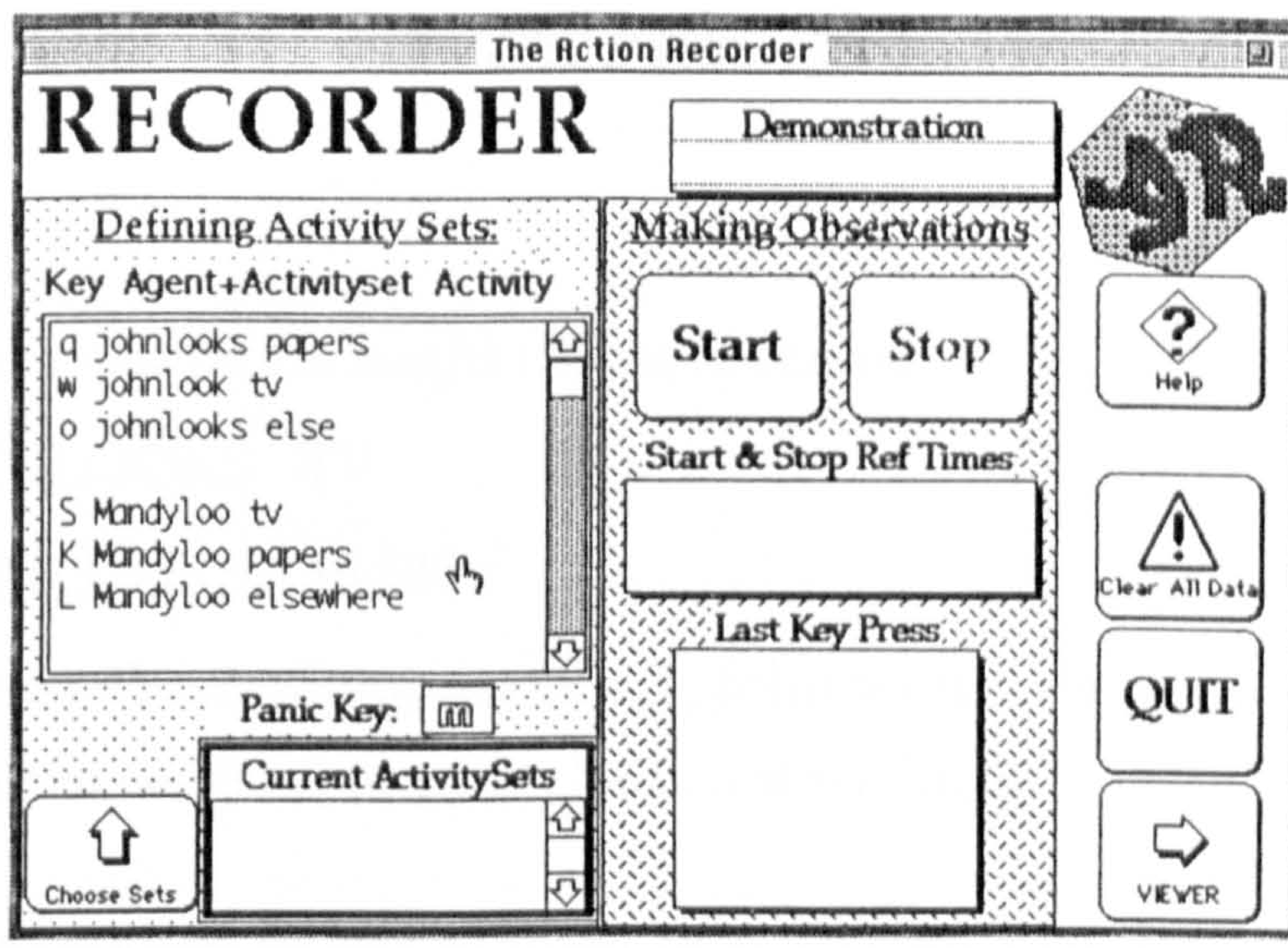
There are three phases of work with Action Recorder: actually doing the observation and relevant coding; organising the resultant (linear) codes into a state-based representation of the whole interaction; and finally, obtaining statistical summaries of the temporal and frequency characteristics of the activity streams analysed.

A1a.4. Phase One (a) - Recording Your Observations

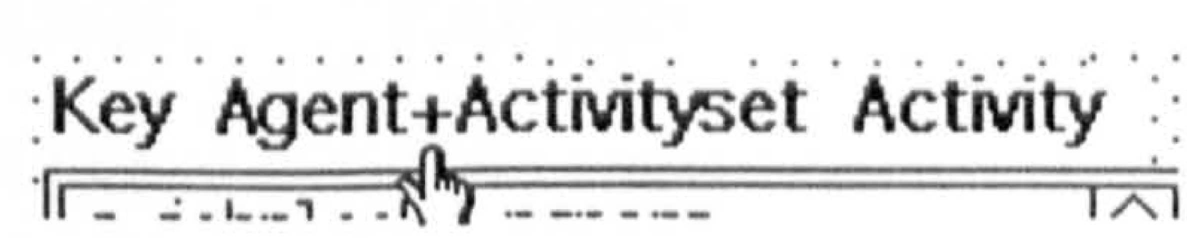
A1a.4.1. Preparing to Make New Observations

Defining Activity Sets

Activity Sets are defined on the RECORDER part of Action Recorder. The RECORDER may be reached from the welcome screen by clicking on the 'Make Observations' button, or from anywhere else by clicking on a button labelled 'RECORDER'.



On the left-hand side of the RECORDER are three boxes, under the collective heading of : **Defining Activity Sets:** One of these is beneath a key for the definition of Activity Sets:



This is the Activity Set definition area. This "key" tells you all you need to know about defining Activity Sets for use with Action Recorder. In this large box, first type a single character to stand for a type of Activity, for example the letter 'w'. This is for the key you will press when you see that Activity taking place. Then enter a space, followed by the name of the agent (for example, JOHN). Add the name of the Activity Set to the agent's name (for example, LOOKS), followed by a space and the name of the Activity itself (for example, TV, short for television monitor for the video link).

Our example would then appear as:

w JOHNLOOKS TV

There must be at least two Activities defined per Activity Set. Otherwise, when you come to make your observations, the single Activity must always be true (there would be no mutually exclusive alternative to it). If it be the case that you are interested in only one Activity within a potentially very large Activity Set, then you should define a 'not true' Activity within this Activity Set. So, in our example, if I were only interested in JOHN looking at the TV, and not particularly in anywhere else he might look, I would define a second Activity in the JOHNLOOKS Activity Set as NOT, or ELSEWHERE, or OTHER.

The Activity Set definitions might then appear as:

w JOHNLOOKS TV
o JOHNLOOKS ELSE

In this way, when observations are made, John would be recorded as looking at the video link in contrast to when he is not so doing.

Activity Sets can consist of any number of Activities, just so long as they fulfil the 'mutually exclusive' criterion. For our example, another type of looking behaviour of interest is the use John makes of the papers he has brought to the meeting. So the JOHNLOOKS Activity Set now consists of three Activities:

q	JOHNLOOKS	PAPERS
w	JOHNLOOKS	TV
o	JOHNLOOKS	ELSE

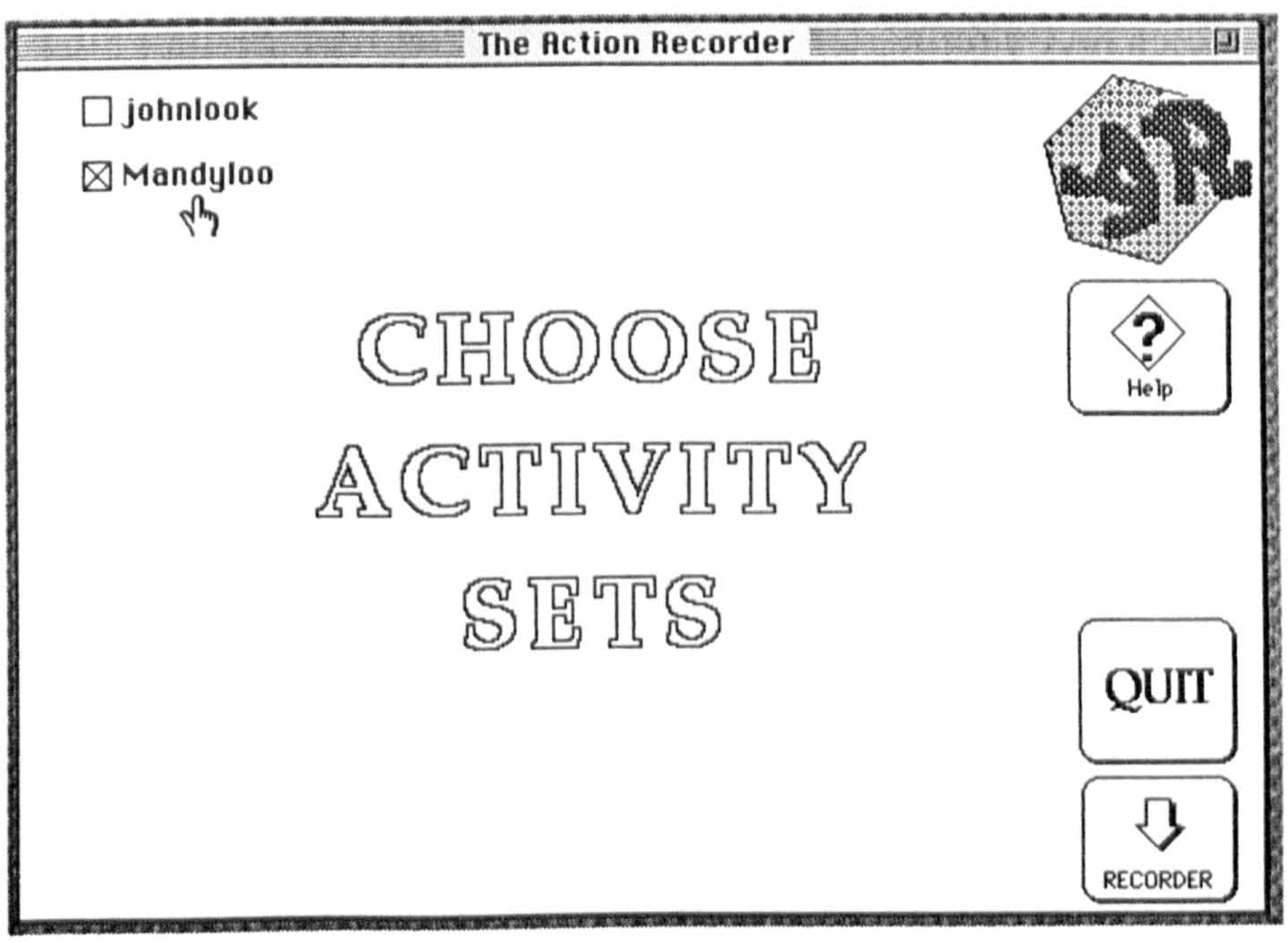
Each definition of an Activity should be on a separate line. Although there is no need to keep all definitions for a given Activity Set on adjacent lines, you may find it helpful to do so.

In the example RECORDER screen, a similar Activity Set is defined for Mandy. Capital letters were used for the Mandy's keystrokes, to help distinguish Mandy's activities from John's activities, although this is not necessary.

Selecting Activity Sets to Observe

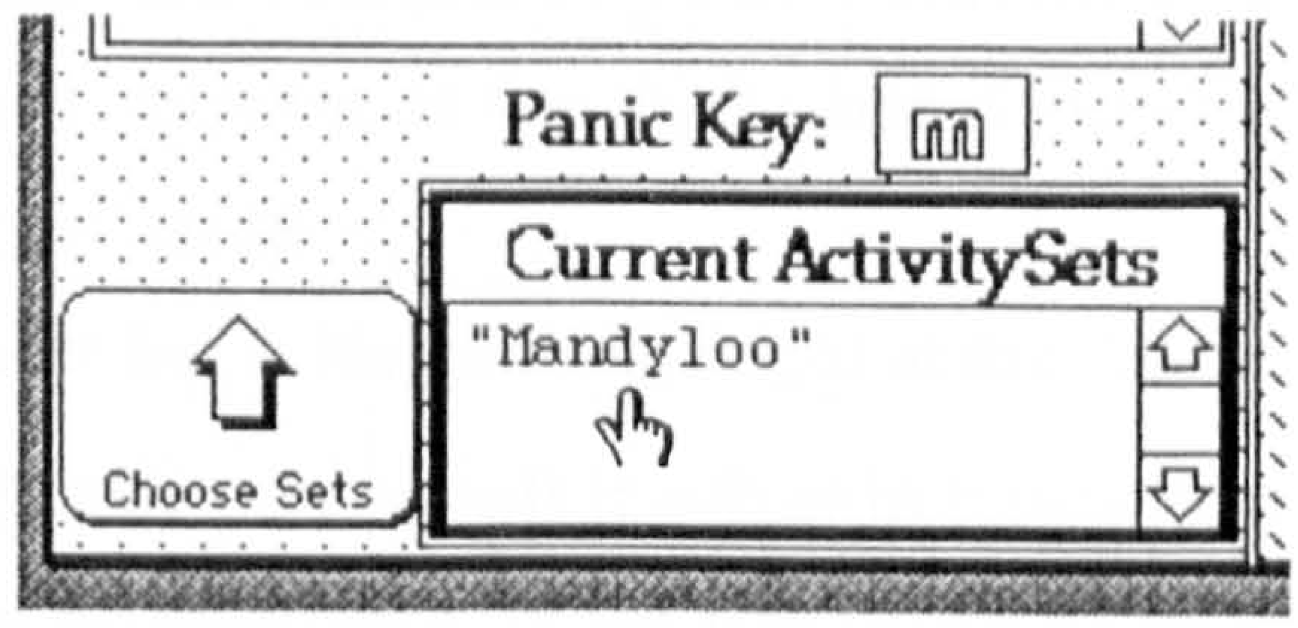
Potentially, there are any number of Activity Sets which could be defined. Almost certainly, even those that you decide are of importance for your analysis will overwhelm your capacity for concentration, if all were to be observed at once. For this reason, Action Recorder requires you to choose one or more of the Activity Sets you have defined at a time. This is done by pressing the button called "Choose Sets" at the bottom left-hand side of the RECORDER screen.

When you click this button, Action Recorder takes a look at your Activity Set definitions and puts together a list of the Activity Set names. These appear on a different screen. Next to each entry in this list is a box. Choosing Activity Sets to be observed on a particular occasion, or run through of a video recording, is achieved simply by clicking the relevant boxes. Just how many Activity Sets can be selected comes down to how well you, as an observer, feel you are able to accurately and reliably record the events that pertain to these Activity Sets. In the example below, one Activity Set has been selected, the "MANDYLOO" set.



Note: Action Recorder only takes notice of the first eight characters of an Activity Set name: although it is possible to enter names of any length in the Activity Set definition area, it is much better to stick to names only up to eight characters long, to avoid confusion. To emphasise the point, in the examples used, one of the Activities defined to the JOHNLOOKS Activity Set is written in full in the definition area, whilst the other three activities in this Set are written only as JOHNLOO.

Once you have checked boxes against each of the Activity Sets you wish to observe on this occasion, click on the RECORDER button to return to the RECORDER area. Action Recorder lets you know which Activity Sets it is ready to record by listing them in the box at the bottom left-hand side of the RECORDER area, next to the "Choose Sets" button. These are known as the 'current' Activity Sets.



The Panic Key

There is one more thing to set up before you go on to make your observations. This is called the "Panic Key". The Panic Key allows you some way to make corrections to your observations as you go, to avoid having to stop, correct and restart your observations. It is to be found on the left-hand side of the RECORDER, above the "Current Activity Sets" box. The Panic Key works as follows:

1/ Suppose you observe something happening and hit a key. You realise you hit the wrong key and know which key should have been pressed. Hit the Panic Key twice, and then the key you should have hit once. Action Recorder will substitute the correct key for the incorrect one, without affecting the time at which the original key was pressed.

2/ Suppose you think you have made some mistake, but you don't know exactly what your mistake was, or you think you may have missed something. Hit the Panic Key once. When you have finished your observations, and go to the VIEWER to see your data, Action Recorder will check the keystrokes you have made and see whether or not any of these were Panic Key keystrokes. If it finds any, it will tell you how many it found. It is then up to you to scroll through the VIEWER window to locate the errors, use the time stamp of the Panic Key to find the problem area of your videotape, and to correct the problem by hand.

Setting the Panic Key

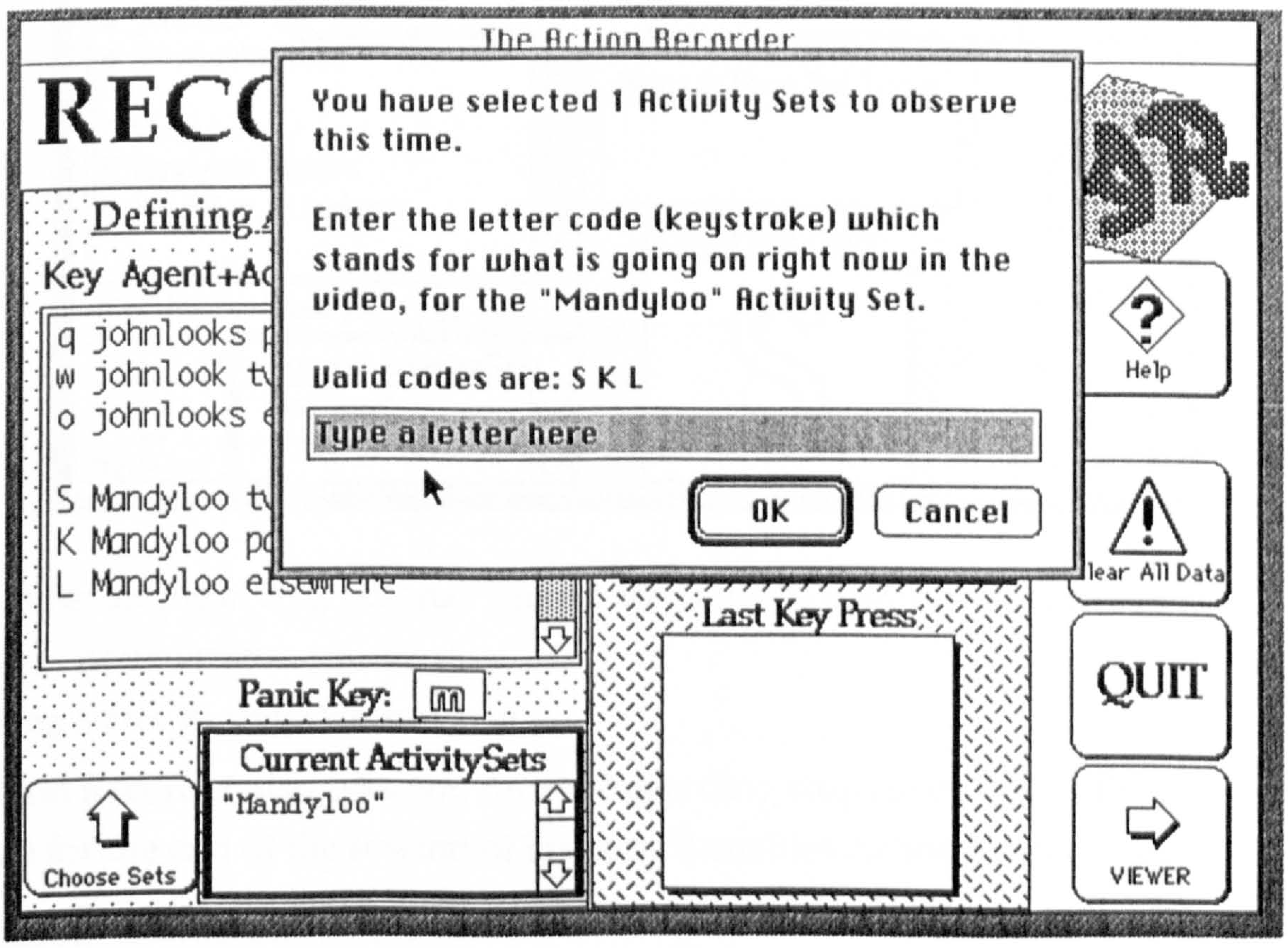
Type a single character into the box to the right of the Panic Key label. Note that this character must be unique to the Panic Key box, and not used in the Activity Set Definition, or Action Recorder will not know whether you have made a mistake or a correct keystroke. The Panic Key is pre-set to 'm' (for mistake), but it can be any character you like.

A1a.4.2. Making Observations

After setting up Action Recorder, making your observations is straight forward. Find the relevant section of your videotape and put your VCR on pause. Click the "Start" button at the right-hand side of the RECORDER, in the "Making Observations" area.

Telling Action Recorder What is Happening, Right at the Start

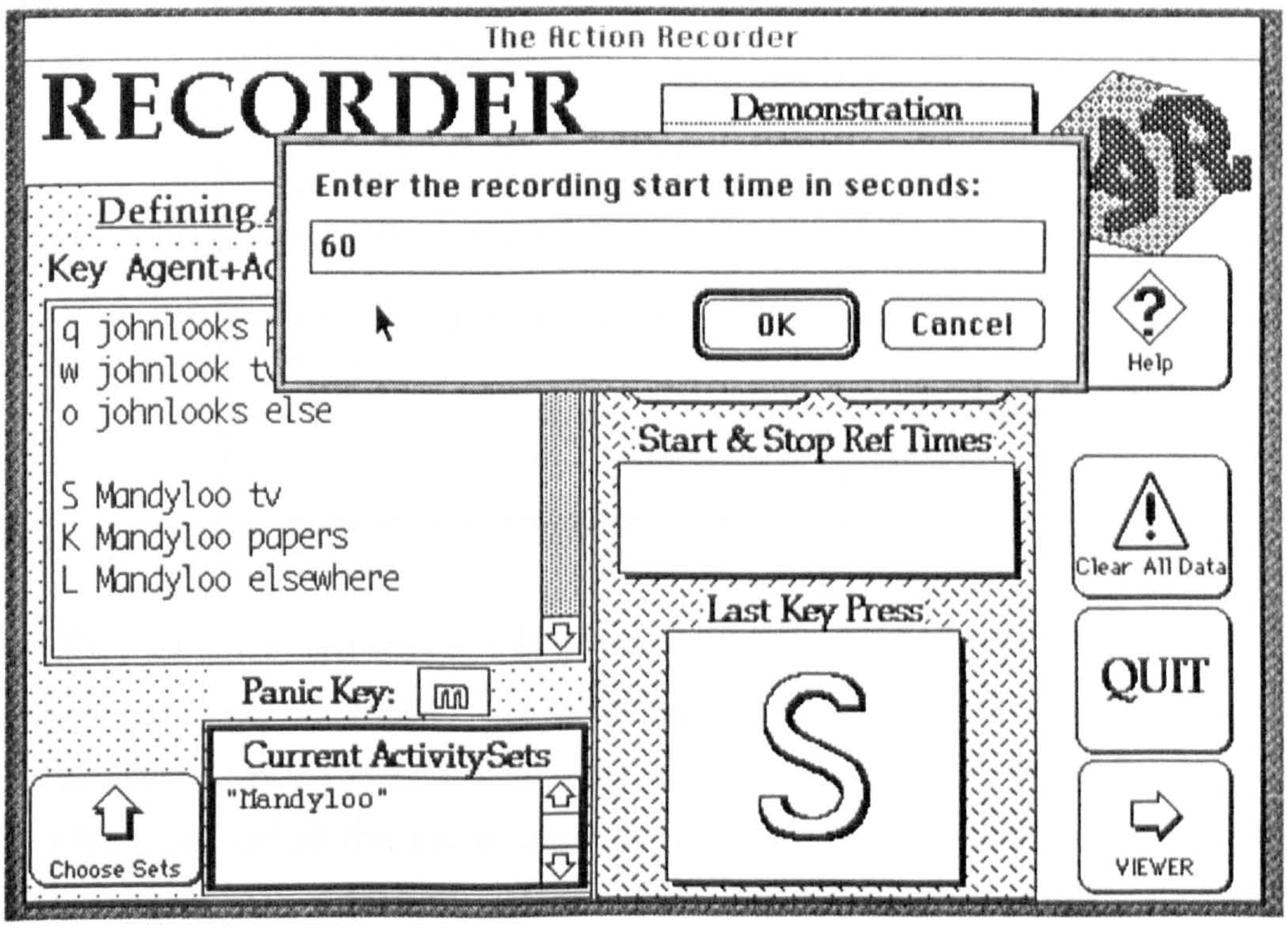
Action Recorder will ask you to tell it what is true about the Activity Sets you have chosen to observe this time, from the point at which you wish to record. This is because Action Recorder is all about things that are true in the world over time, rather than isolated events in the world. Action Recorder tells you how many Activity Sets you have chosen in a dialogue box, together with the name of one Activity Set, and a list of the keys you have defined as referring to Activities belonging to that Activity Set. Look at your video recording, where you have put it on pause. Decide which Activity in the named Activity Set is true at that time. Enter the relevant character, standing for the Activity which is going on where the video is paused. Click 'OK', or press ENTER.



In the example above, says that where I paused my video, Mandy was looking at her papers. So the MANDYLOO PAPERS Activity is true. I have defined the letter "S" to represent this Activity. Action Recorder tells me that this is indeed a valid code for this Activity Set, so the letter 'S' is entered in the box provided. "S" then appears in the 'Last Key Press' box, on the right-hand side of the RECORDER. Had the JOHNLOOK Activity Set also been chosen for observation the this time around, Action Recorder would also have requested for what is true about this Activity Set in turn, at the point where the videotape was paused.

Telling Action Recorder Where the Start Is

Everything is organised around the time code on your videotape. Once you have told Action Recorder those Activities going on at the start of the section of interest, it asks you to enter the time at which the section starts. This must be in seconds and tenths of seconds, rather than hours, minutes and seconds.[51]



In the example above, 60.0 has been entered in the dialogue box, which is of course one minute.

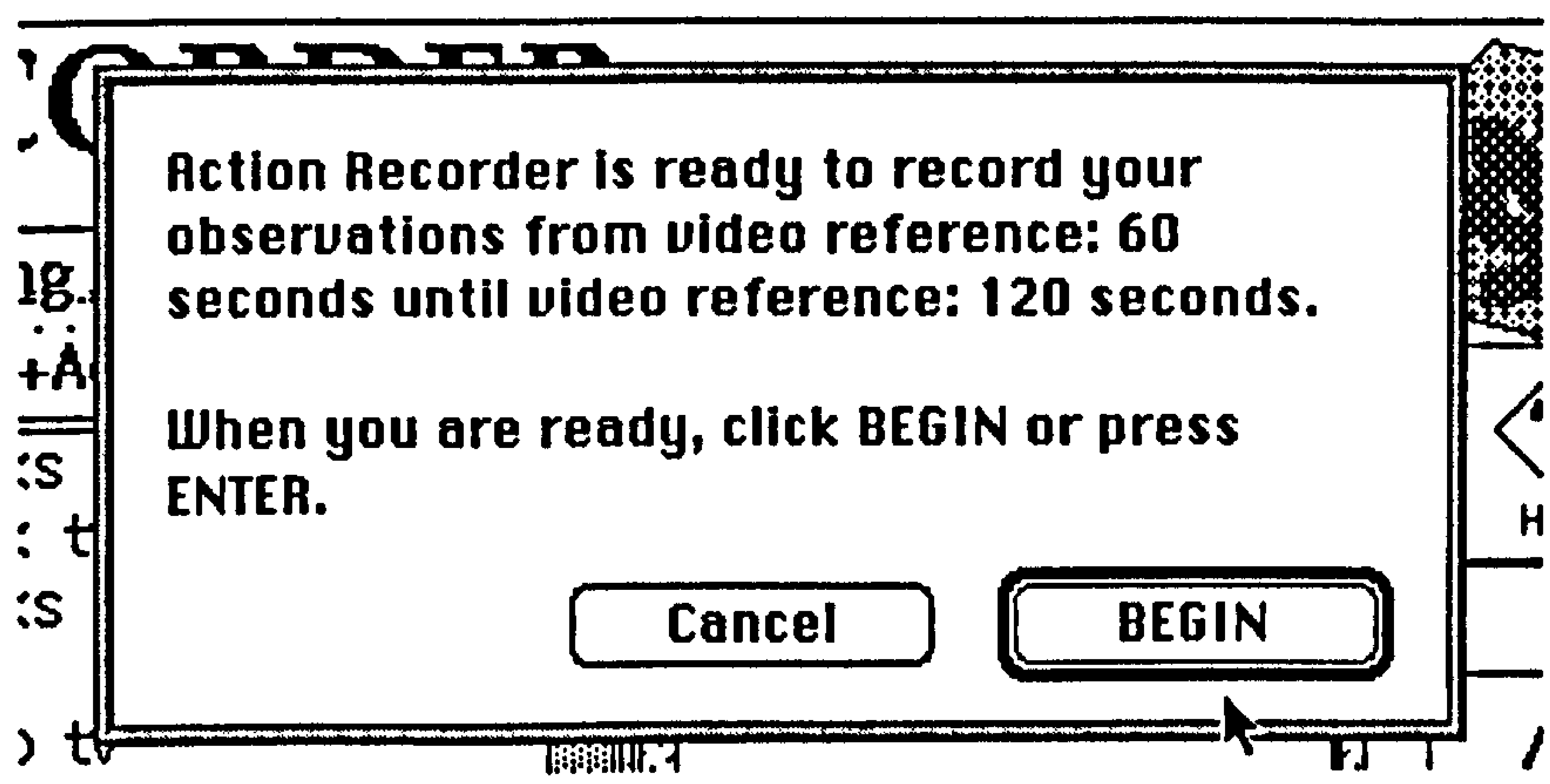
Telling Action Recorder When to Stop

Action Recorder also asks you for the 'recording stop time'. This is the time code for the end of the session of interest. It enables Action Recorder to

[51]The time code display on your VCR itself may not be very reliable. Some more expensive VCRs read the time codes directly from the video tape and display these on their front panels. A cheaper alternative is to use a 'Character Generator' at the time of recording, with a built-in stop watch. This box plugs in the camera and VCR, to mix a time code into the picture and so can later be read from your display/TV screen.

automatically stop recording keystrokes after a predetermined period, which means that you, as an observer, don't have to think about stopping Action Recorder yourself. When you have typed in the recording stop time and clicked "OK", you are ready to begin recording. In the example, the recording stop time was set for the two minute video reference, which is 120 seconds, making a one minute long observation period. Should you wish to stop recording your observations before the predetermined stop time for some reason, you may do so by clicking on the "STOP" button. After the "stop time" window is dismissed, a final dialogue box appears. This confirms the video reference times you have selected and informs you of Action Recorder's readiness to start recording your observations.

Starting observational RECORDings



When the "BEGIN" button is clicked, Action Recorder starts its internal clock to keep pace with the video. So, only click the "BEGIN" button when you are ready to start making key presses for observations. The "BEGIN" button should be pressed at the same time as the VCR is taken off of pause, and hence starts to play again. The ENTER key has the same effect as clicking on the "BEGIN" button and may be a more convenient method of starting Action Recorder's clock than the use of the mouse, especially if your VCR has a remote control unit. The remote control unit may then be positioned immediately adjacent to the Macintosh keyboard, where the ENTER key is at its extreme right-hand end.

So, when you are sure you are ready to start making your observations, simultaneously hit ENTER/click begin with the VCR pause/play button. It is better to go from pause to play than from stop to play, as the "activation inertia" of the VCR internals is far smaller in the former case.

Watching the RECORDER at work

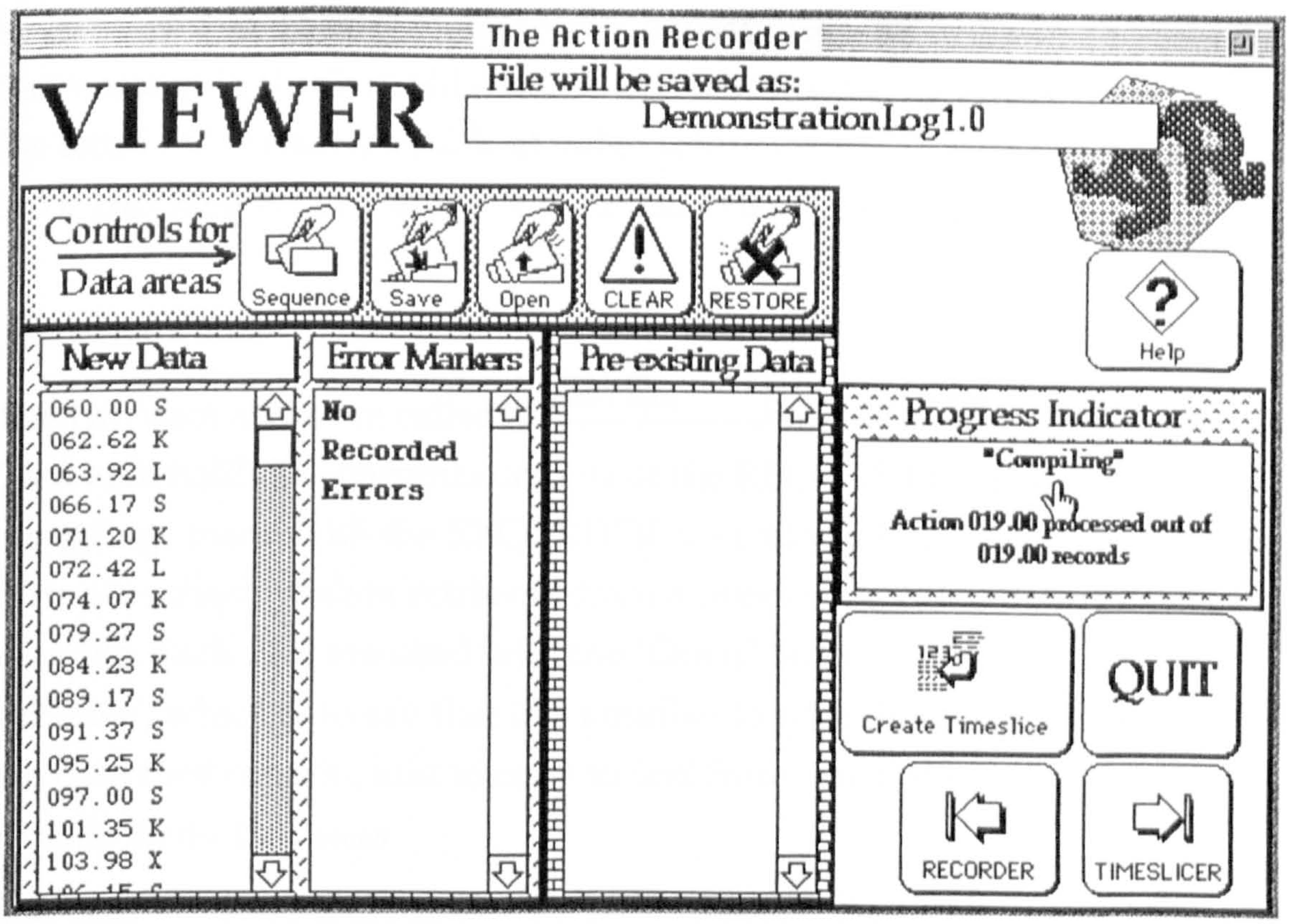
Action Recorder lets you know what it is doing whilst you make your observations in several ways. The video reference time from which your current observations commenced is shown in the Start & Stop Ref Times box. This box will also display the time at which the last set of observations stopped when you record several passes. Whilst a observation session is in progress, the "Start" button is greyed out, with the "Stop" button emboldened, and vice versa when not recording. Each key stroke you make during an observation session is shown in the "Last Key Press" box, to help when you are not sure whether or not you pressed the correct key.

Ending an observation session

Action Recorder continues to record your observations until the predetermined stop time, at the end of which the RECORDER is automatically returned to its resting state . Action Recorder then presents the VIEWER screen, allowing inspection and, if necessary, correction of the new data. Alternatively, you can stop recording your observations by clicking on the STOP button at any time. The effect of so doing is precisely the same as allowing Action Recorder to time out. The time at which the recording stopped is indicated in the "Start & Stop Ref Times" box, in either event.

A1a.4.3. Inspecting Observational Recordings

After an observational recording session has been completed, either by the RECORDER timing out, or ending a session with the "Stop" button, Action Recorder proceeds to the VIEWER screen.



In the above example, the VIEWER show some data collected for the MANDYLOO Activity Set. There are four boxes or areas on the VIEWER screen: two are for information and two for data.

Information Areas

THE PROGRESS INDICATOR

The box on the far right-hand side - under the Progress Indicator label - shows how far Action Recorder has got with processing the data. This processing often takes some time. Since Action Recorder will keep working, even if it's window is not 'in the foreground' on your Macintosh, you can do other work, say writing a report with a word processing package, whilst keeping an eye on how Action Recorder is getting on with the job.

Here, the VIEWER is shown after all the new data has been 'compiled', which is to say made human readable, with the correct video reference times next to each Activity letter code.

THE 'ERROR MARKERS' AREA

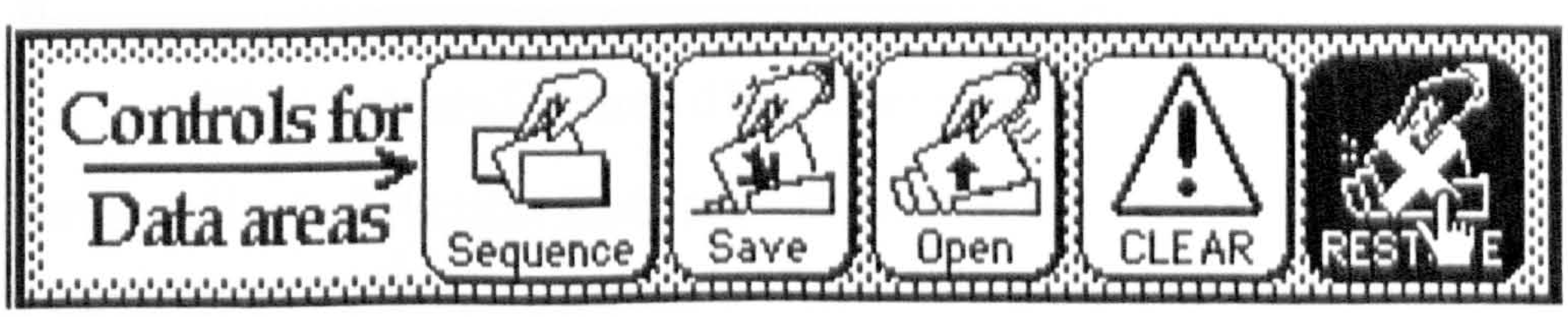
The other information box is headed Error Markers. The "Error Markers" box show the video reference times when the Panic Key was pressed once, during the latest RECORDER session. Note that it does NOT show instances of keystrokes that do not belong to any Activity Sets. So in the example above, the message in the Error Markers box is "No Recorded Errors", even though the letter "X" is clearly visible at video reference 103.98 seconds. This form of error checking is done at the "Create Timeslice" stage (see below).

Data Areas

The two data areas are called New Data and Pre-existing Data. The New Data area holds the current contents of the RECORDER, so all observational recordings made with the RECORDER are collected here. The Pre-existing Data area displays data retrieved from a previously created Action Recorder data file. Such files are used with the "Open" button. Both data areas are 'editable', which is to say that it is possible to type directly into them, amend the data they contain, and to copy to and from each area.

Controls for the Data areas

The data areas are managed with five buttons, collectively called "Controls for Data Areas". The "Sequence", "Save", "Clear" and "RESTORE" buttons may all apply to either data area. In each case, if both data areas contain data, use of any of these buttons will prompt Action Recorder to ask you which area to apply the button. The "Open" button only applies to the Pre-existing Data area.



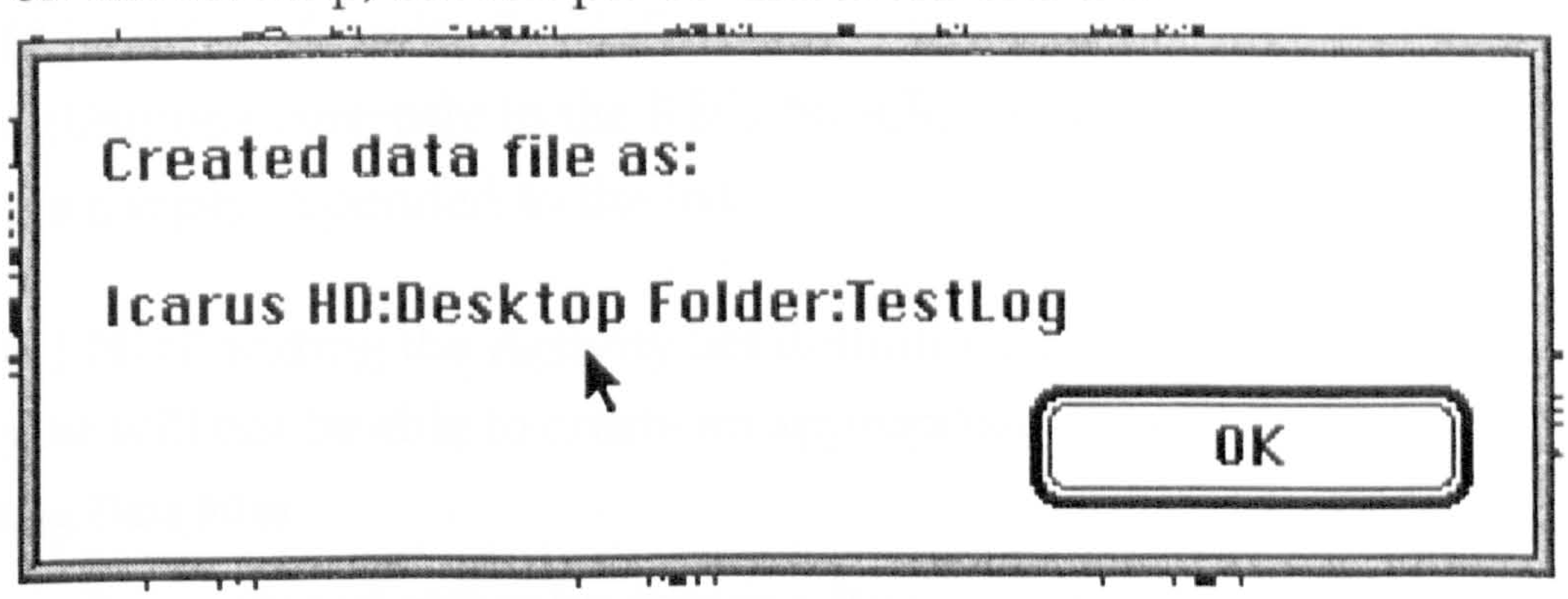
The RESTORE button can restores the contents of the New Data area after a session of RECORDER use, until a file is opened with the "Open" button. If the "Open" button is subsequently used, the 'focus' of the "RESTORE" button changes to the Pre-existing Data area. It then restores the contents of the Pre-existing Data area either to the original contents of the last file to be opened (via a dialogue box), or to the previous contents of the Pre-existing Data area. The "RESTORE" button has the potential to be very confusing. Just remember: you can always undo the "RESTORE" by pressing the button again, and

Action Recorder's "RESTORE" is provided in addition to the standard Apple "Undo" item in the "Edit" menu.

Saving Data

Once you are satisfied with the data you have recorded, you can save it out to a file, for future reference. Click the "Save" button in the "Controls for Data Areas", above the Data areas on the VIEWER:

A standard Apple Macintosh dialogue box appears, asking for a file name and allowing you to select a folder in which to save the file. After you have saved the file, Action Recorder confirms that the file has indeed been saved, and shows where it has been saved with a full path name. A "path name", is a list of all the folders enclosing the file, each separated by a colon (:). Note that the Apple Macintosh "Desktop" is really a folder in its own right, belonging to the drive containing the system folder. So if you save a file to the desktop, or in a folder on the desktop, the full path name looks like this:

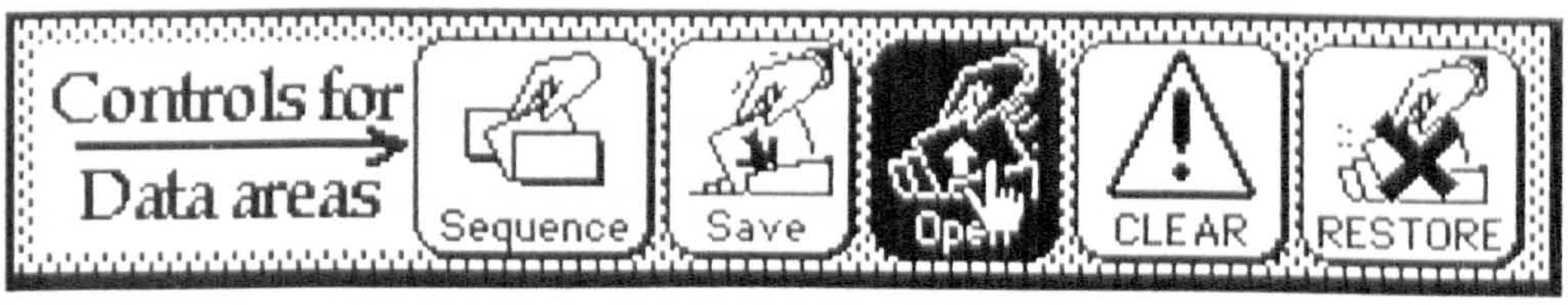


A1a.4.4. Viewing and Amending Previously Collected Data

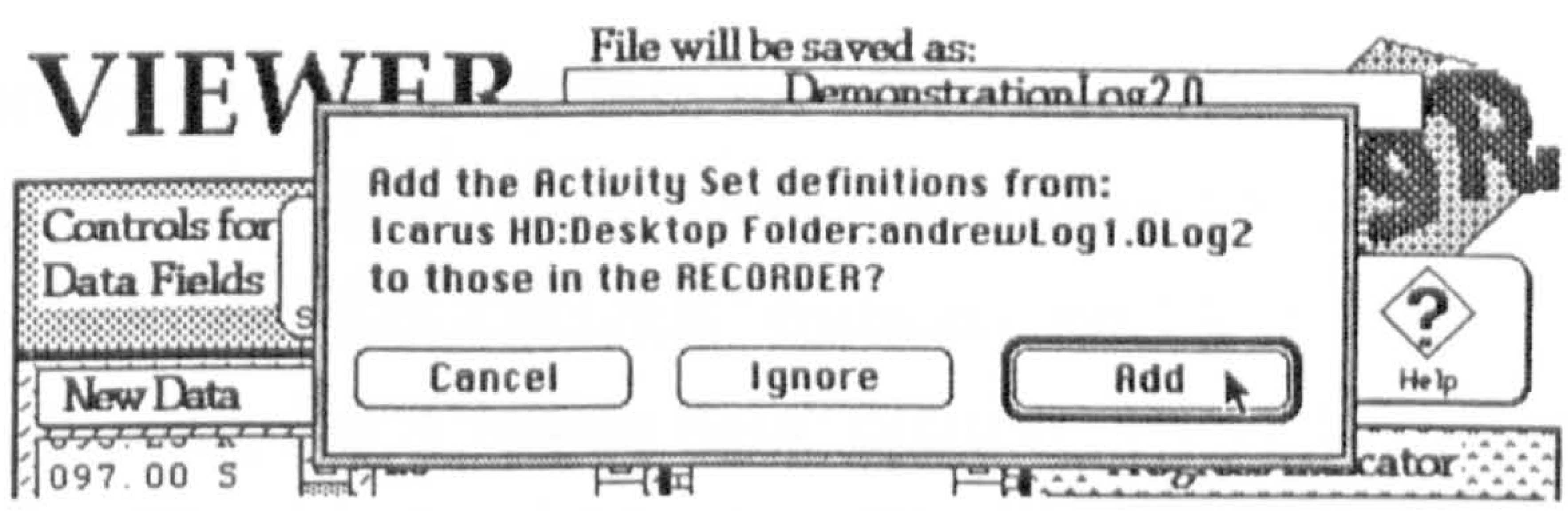
If you have collected Action Recorder data and saved it out to a file, you may bring it back in to Action Recorder's VIEWER. So doing allows simple, visual inspection; modification, for example error correction; or addition, with new data collected more recently.

Opening Data Files

To open a previously created data file, simply click on the "Open" button, from the "Controls for Data Areas", above the Data areas on the VIEWER:



A standard Macintosh dialogue box asks for the location of the file. Provided the file selected is an Action Recorder data file, you will be asked whether or not you wish to add the Activity Set definitions for this file to the Activity Set definitions area on the RECORDER.



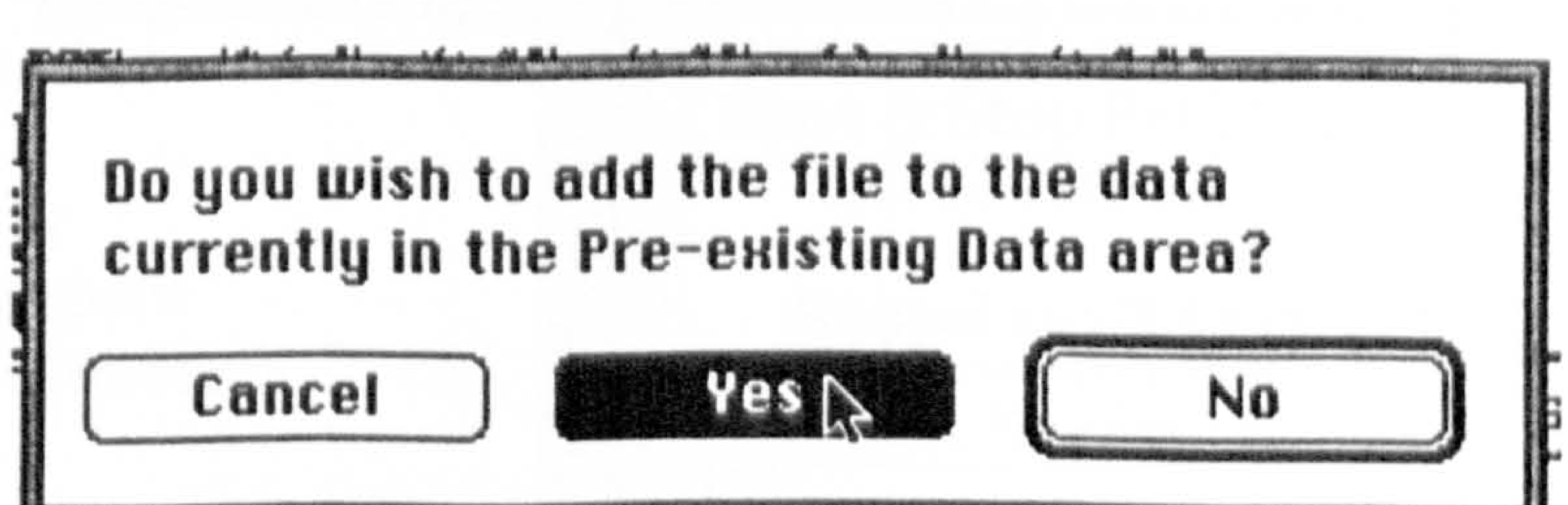
Note that:

(A) Adding Activity Set definitions will not remove any Activity Set definitions currently in the RECORDER; the definitions from the file are simply appended to the list.

(B) NOT adding the Activity Set definitions from the file will mean that you will not be able to create an appropriate Timeslice Table.

Combining Data Files

Data may be combined either by copying from one data area and pasting on the end of the other, or by successively opening data files. If the Pre-existing Data area contains data when the "Open" button is used, you will be asked whether or not you wish to combine the data from the file with the data already in the Pre-existing Data area:



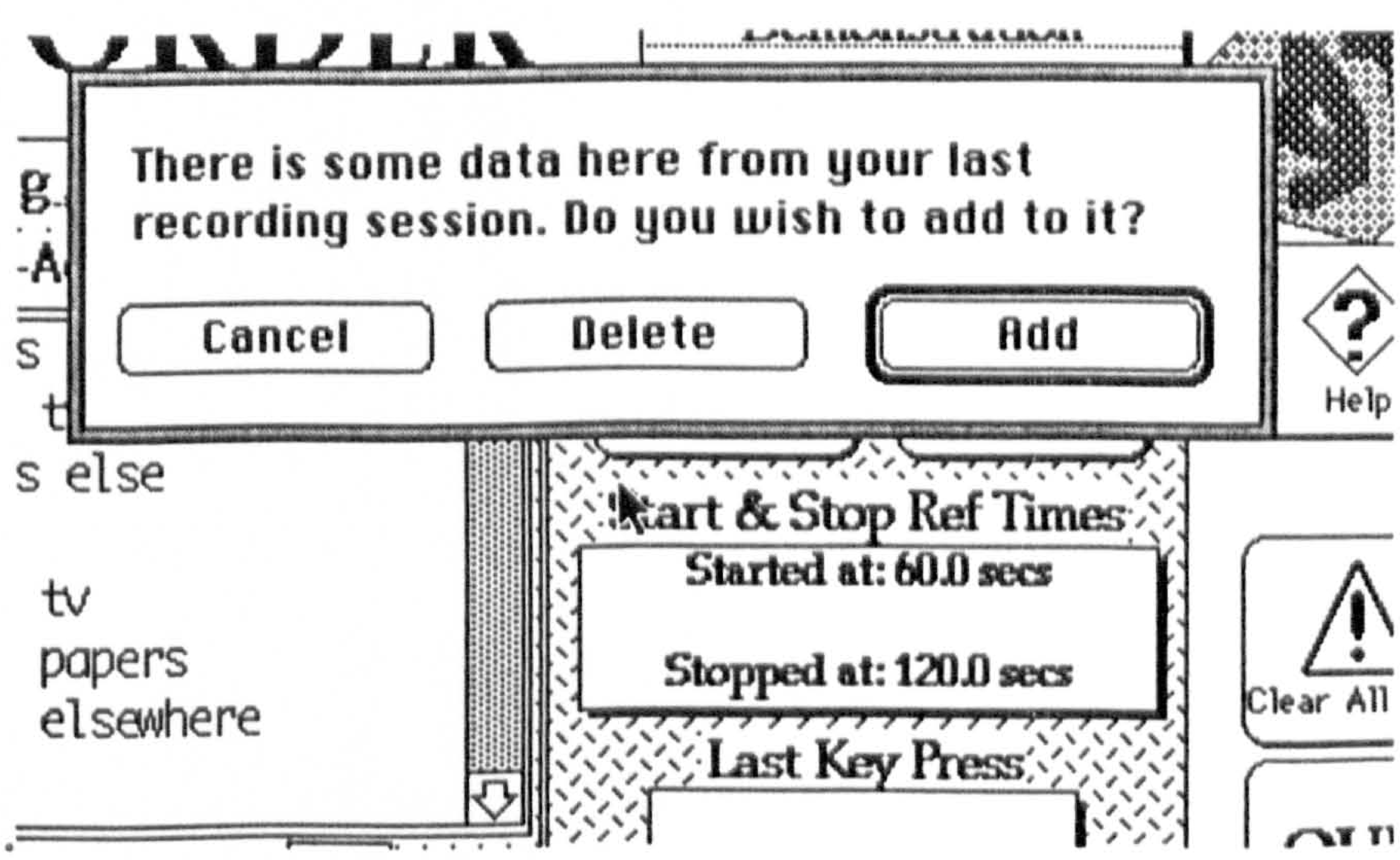
Clicking yes simply appends the data from the file to that already in the Pre-existing Data area.

Note that after a file has been opened, the "RESTORE" button returns the contents of the Pre-existing Data area either to the contents of this area prior to the opening of the last file, or to the contents of the most recently opened file. Remember: if some more new data are collected after the "Open" button has been used, "RESTORE" shifts focus to the contents of the New Data area. The "RESTORE" button has the potential to be very confusing. Once again, you can always undo the "RESTORE" by pressing the button a second time. Action Recorder's "RESTORE" is provided in addition to the standard Apple "Undo" item in the "Edit" menu, to help with those "Oh no, what have I done!!!" moments.

A1a.5. Phase One (b) - Making Further Observational Recordings

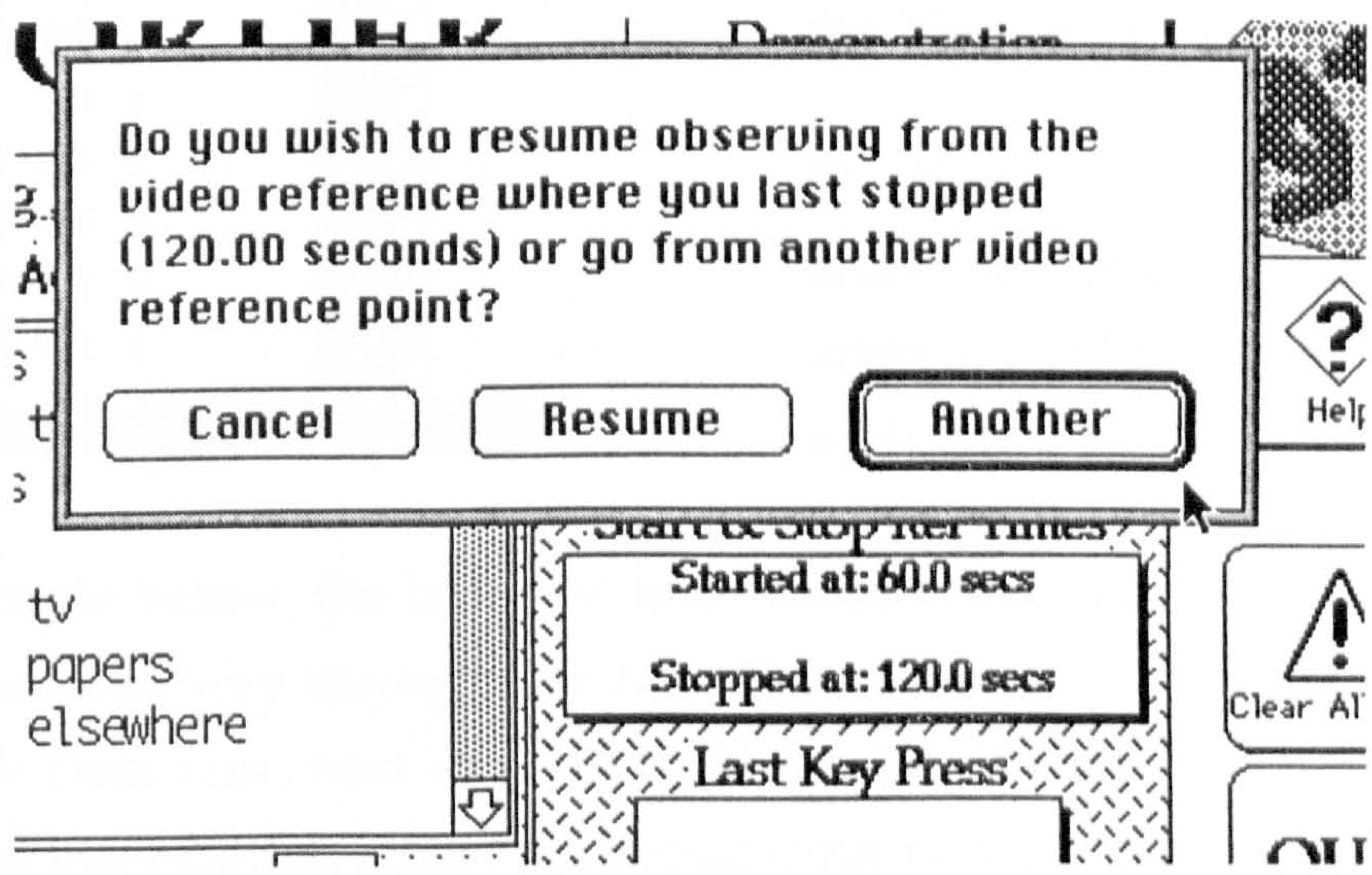
Should you not have recorded data for all of the Activity Sets you require, you will wish to add to the data in the New Data area. Do this by going back to the RECORDER (click the "RECORDER" button), and following the procedure as before , from clicking the "Choose Sets" button onwards (see above - "Recording your Observations; Making Observations; Telling Action Recorder What is Happening").

This time, after typing in the letter codes for the Activity Sets where the videotape is paused, you will see a dialogue box informing you that Action Recorder already has some recorded data. The dialogue box asks if you wish to either "Add" to these data, or "Delete" them. Click the "Add" button to add to the New Data. The "Delete" button clears the New Data and then begins a RECORD session precisely as before

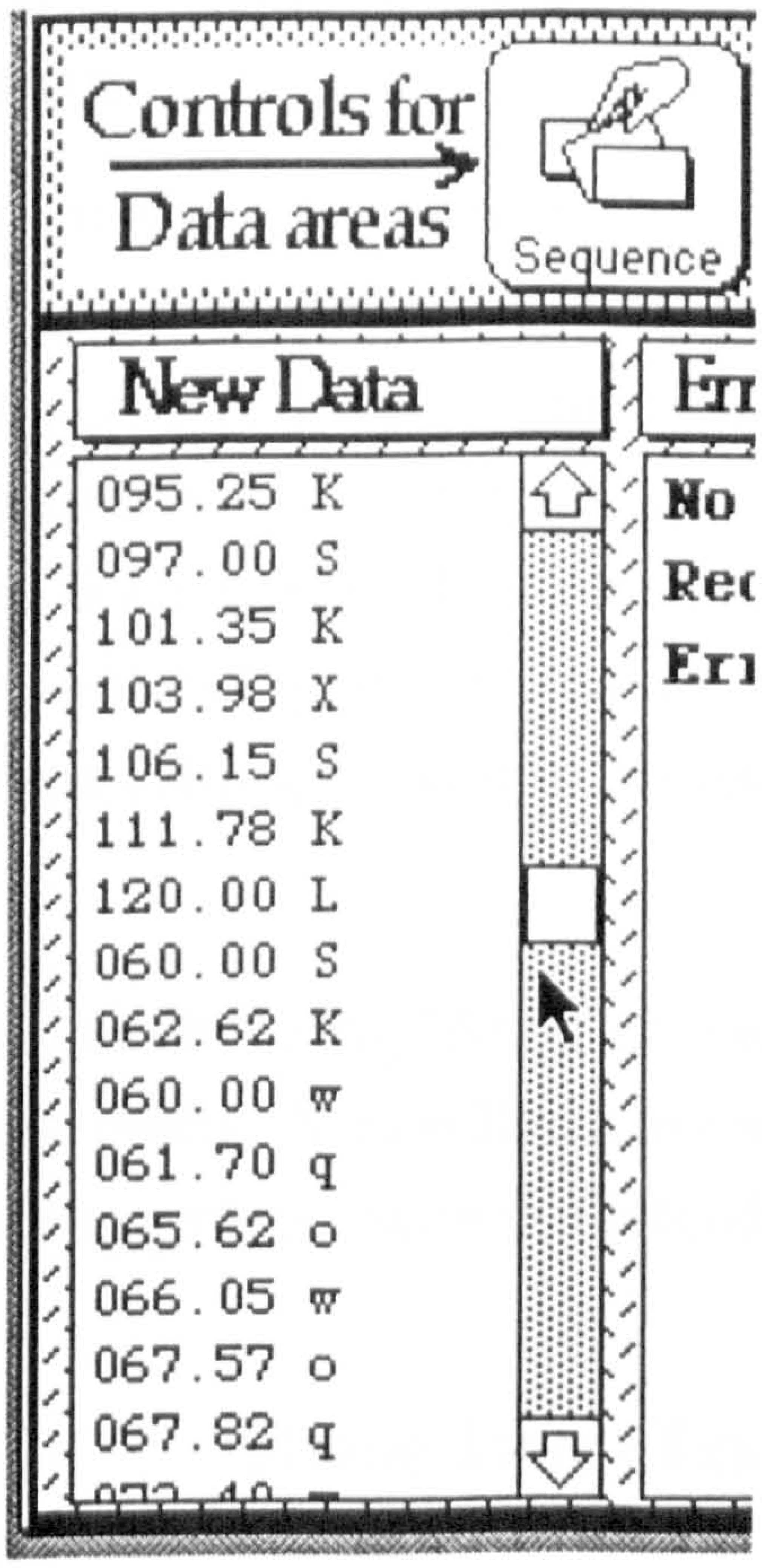


If you do not wish to Add or Delete, then click "Cancel", go to the VIEWER and save the data by clicking on the "Save" button.

Action Recorder next asks you if you wish to add to the New Data from the point at which you last stopped the RECORDER, or else from some other video reference point. It reminds you of the video reference where it was last stopped, to help make this clearer. Since, in this example, the MANDYLOO Activity Set has been observed from the 60 second video reference but the JOHNLOOK Activity Set has not, this RECORDER session should start from the same point at which MANDYLOO was begun i.e. 60 seconds. Since the "Resume" button would automatically restart the RECORDER's clock at 120 seconds, the "Another" button should be clicked.



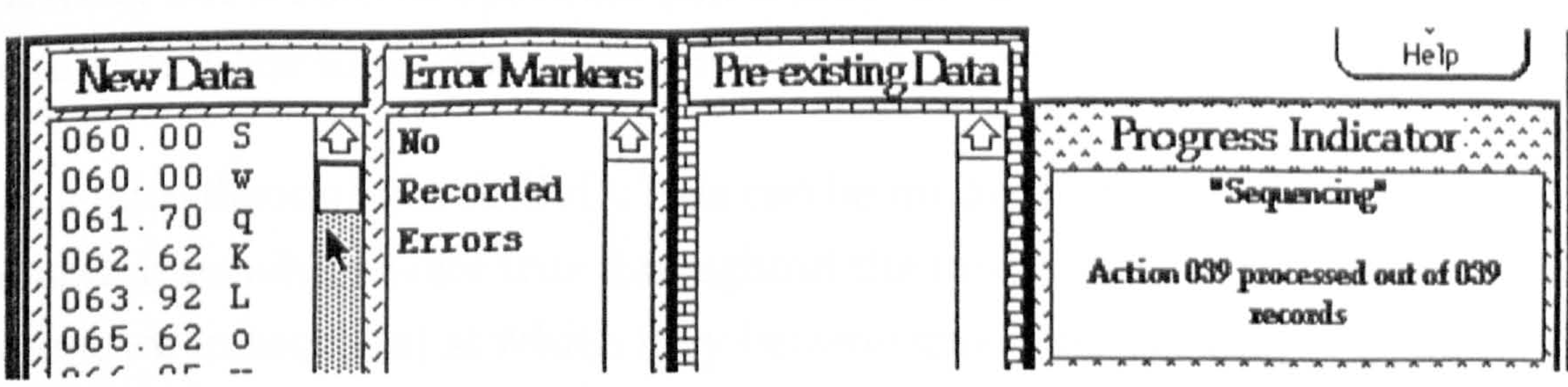
From this point on, the RECORDER behaves just as it does when there has been no data recorded. It either times out at a predetermined stop time, or is stopped by the use of the "Stop" button.



However, this time when the RECORDER stops, the VIEWER shows that the New Data area has been amended by the addition of the latest data from the RECORDER to the end of the data already contained in the New Data Area. As may be seen in this example, the two sets of data are not integrated, as they should be. Integration is achieved by pressing the "Sequence" key.

The "Sequence" key sorts two (or more) sets of data in one of the data areas according to the time with each letter code.

In the example below, the letter 'w' with video reference 60.00 has been moved from half way through the New Data area, after '120.00 L', to the start of the New Data area, next to '60.00 S'. Note also that the "Activity Indicator" now shows information about the sequencing process. Again, when there are many hundreds of keystrokes to sort, this may take some time so it may be useful to work on something else and come back to Action Recorder when it has finished.



A1a.6. Phase One (c) - Interrupting Your Observation Session

Has something happened during the observation of an Activity Set which interrupted the recording process, for example a colleague coming in or a fire alarm going off, simply click the "Stop" button. Then, when the cause of the interruption has been dealt with, rewind the videotape to the point indicated in the "Start & Stop Ref Times" box, set it to pause and click the RECORDER's "Start" button again. Follow the steps as described in Phase One (b), until the dialogue asking for you to choose to "Resume" or "Another" video reference time.

Instead of choosing "Another" video reference time, choose "Resume". When you are ready, Action Recorder will simply take up the observational recording process where it left off.

A1a.7. Phase Two - Transforming Your Observations into a Timeslice Table

The Timeslice Table

A Timeslice table is a *state-based representation of the changes occurring in Activity Sets over the duration of the interaction.* To put is another way, supposing the videotape was to be paused at any time in the interaction/session of interest. At that moment, it would be possible to say something about the behaviours of the various agents concerned, according to the Activity Set principle previously described. In this way, one activity per Activity Set would be specified (since only one activity per Activity Set can be true at any one time - the mutual exclusivity principle).

So, a slice through the Activity Sets can be made. The Timeslice table is a list of activities which were true throughout the interaction, with the time (video reference in seconds) at which they became true, and the length of time that this collection of activities stayed the same i.e. until the next change in one of the Activity Sets occurred.

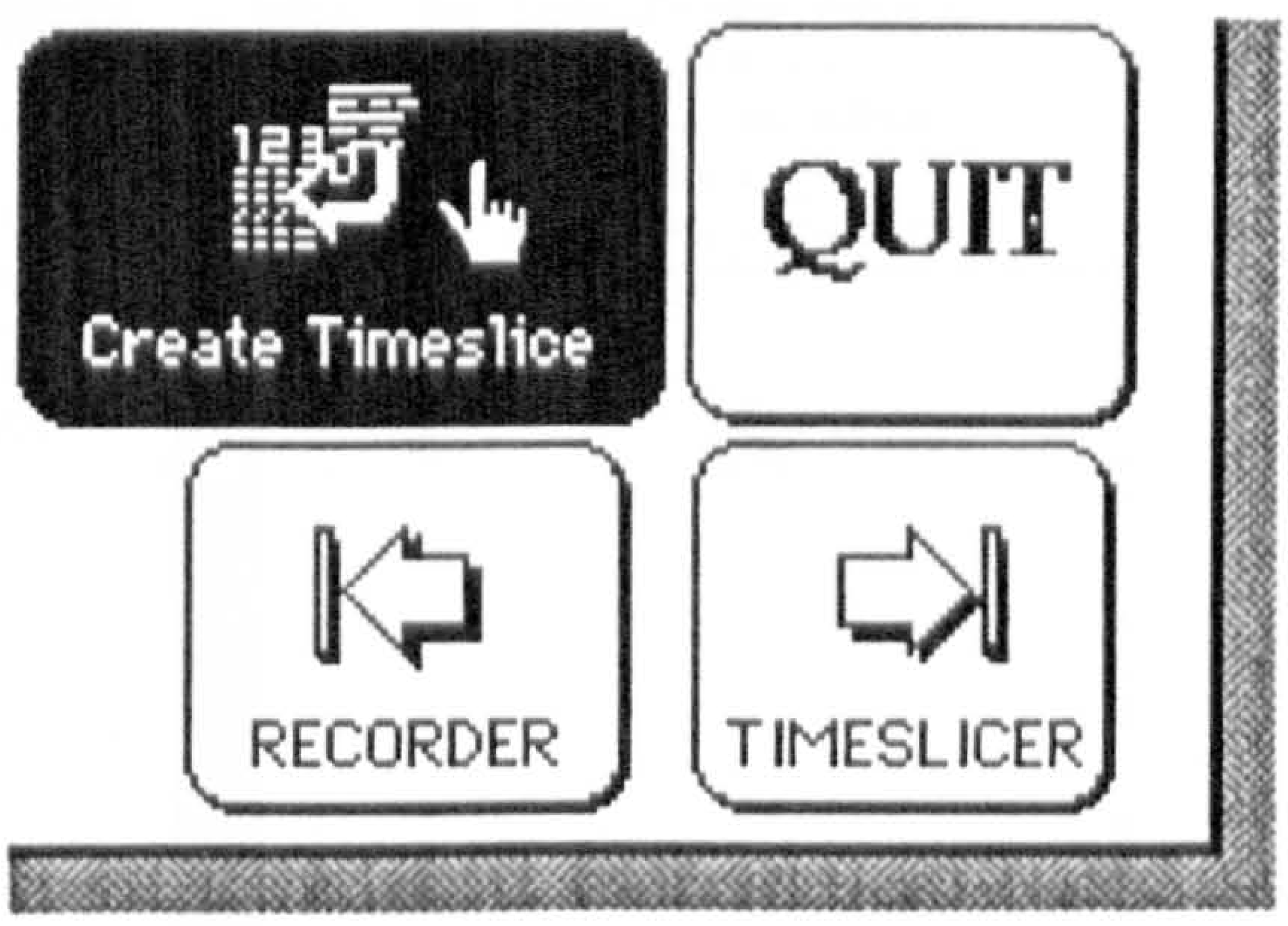
johnlook	Mandyloo	TIME	Duration
w	S	060.00	1.7
q	S	061.70	0.92
q	K	062.62	1.3
q	L	063.92	1.7
o	L	065.62	0.43

In the example above , you can see two columns of letter codes and two columns of numbers. There are two columns of letter codes because, in our example, there were two Activity Sets. The names of these Activity Sets may be seen at the top of the picture. The correct name for the column of letters is a read from left to right. Since "Mandyloo" is the second Activity Set name reading left to right, it means that the second column of letters reading left to right are those referring to the "Mandyloo" Activity Set.

Similarly, the TIME, which is to say the video reference in seconds, is the third column and the Duration (again, in seconds) is the fourth, in this case.

Creating a Timeslice Table for your Data

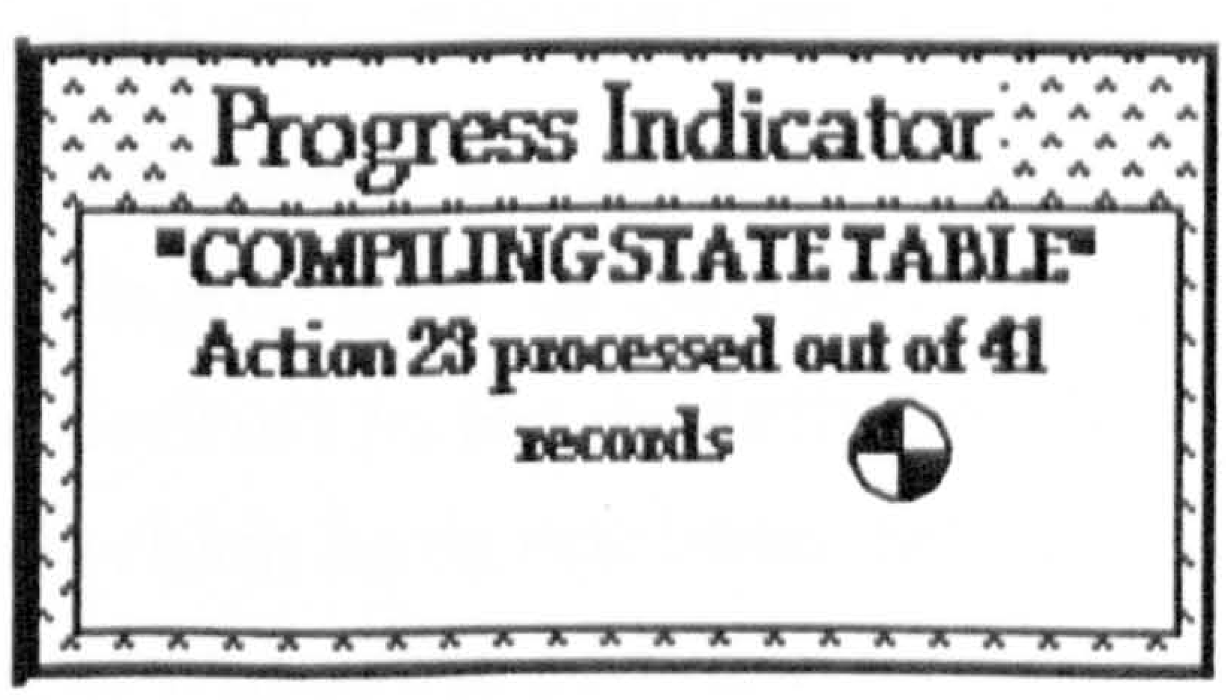
To make a Timeslice Table, click on the "Create Timeslice" button, in the bottom right-hand corner of the VIEWER screen:



If you have data in both the Pre-existing Data and New Data areas, Action Recorder asks you which you wish to use.

NOTE: It is VERY IMPORTANT that the correct Activity Set Definitions are in the RECORDER for the data you wish to turn into a Timeslice Table. This is because Action Recorder needs to look in the Activity Set Definitions area to work out which letter codes belong to which Activity Sets.

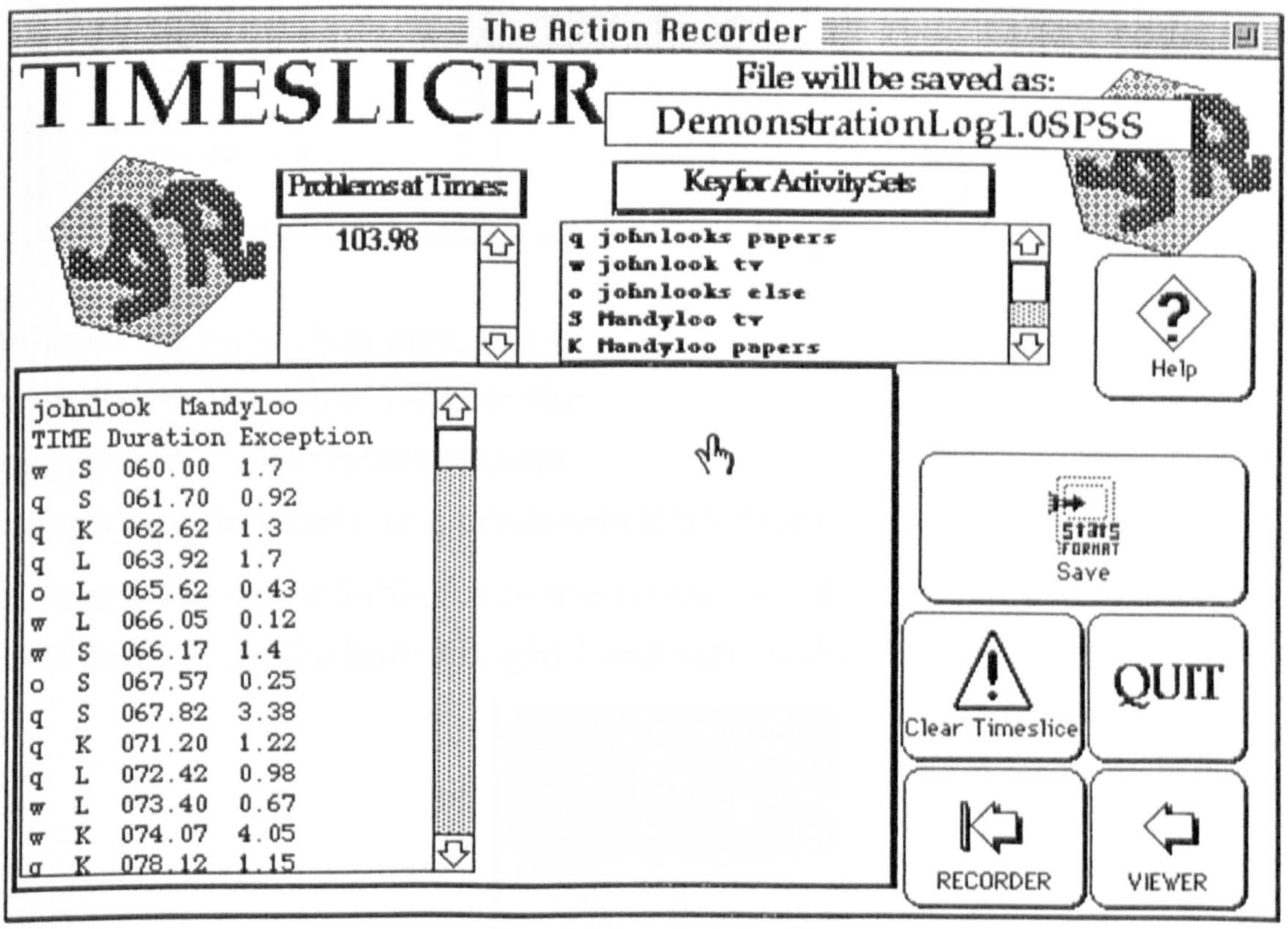
It can take a fair while to create a Timeslice Table. So, whilst this is being done, Action Recorder lets you know how it is getting on with the "Progress Indicator":



When the Timeslice Table is complete, Action Recorder chimes and flashes a dialogue box for you. It then takes you to the TIMESLICER, which is where the Timeslice Table is displayed.

Inspecting Your Timeslice Table

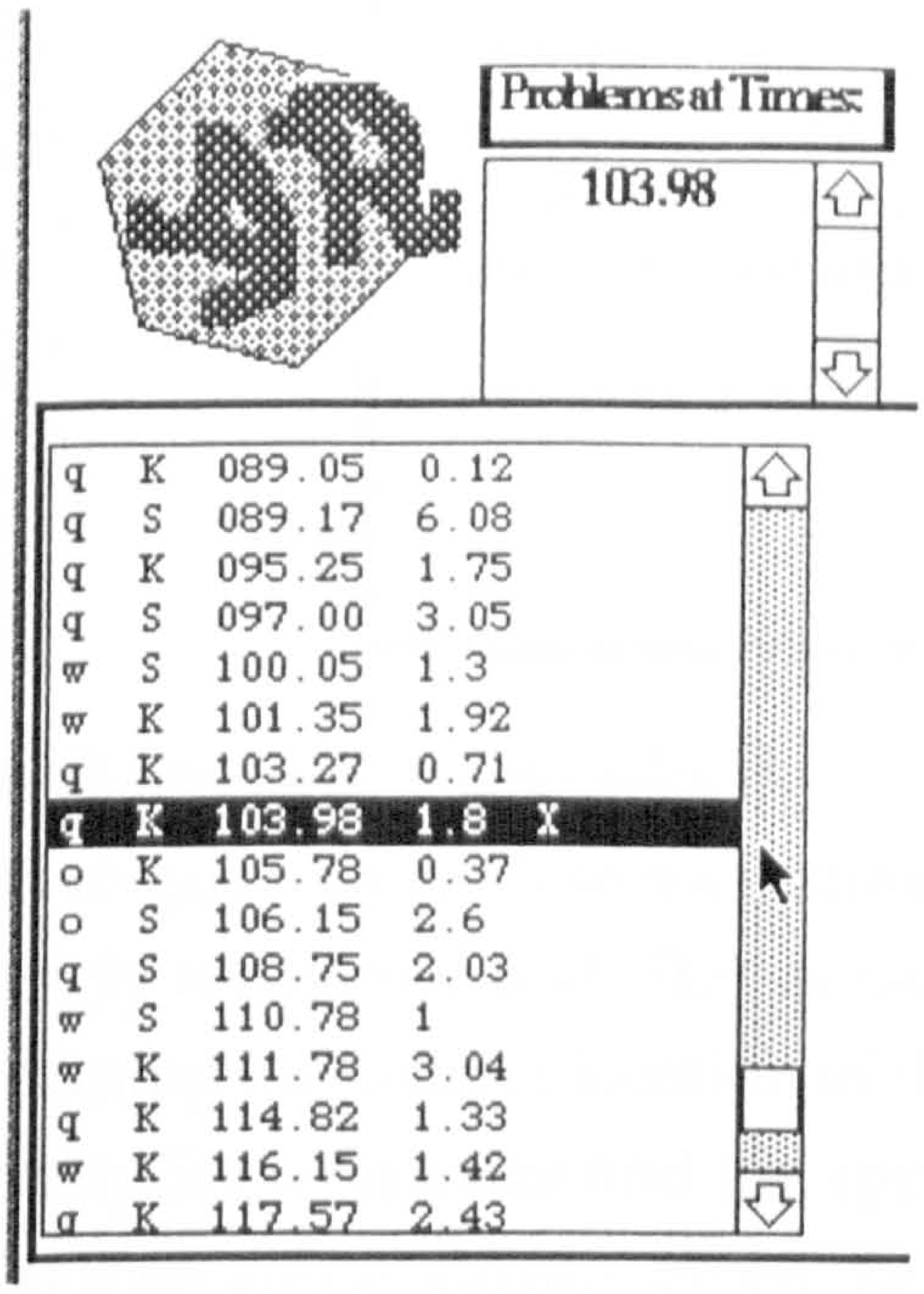
The Timeslice Table is displayed on the TIMESLICER screen. If you have a large amount of data, the Timeslice Table may actually be spread out over several viewing areas. If they overlap, just clicking on them brings them into the foreground. The Timeslice Table is to be found in the bottom left hand side of the TIMESLICER:



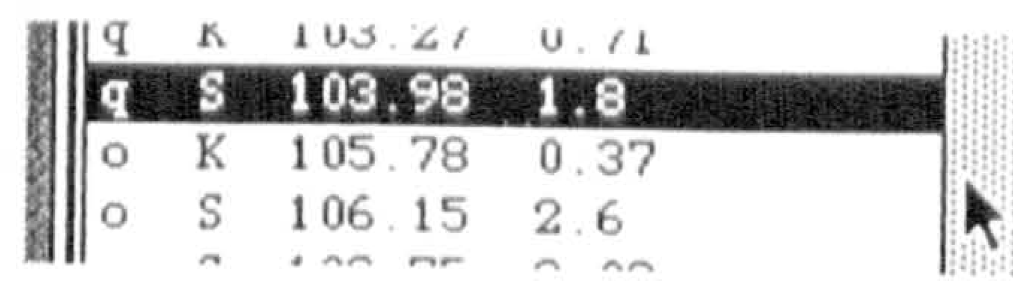
The Timeslice Table is not really meant for visual interpretation: it is a halfway house to obtaining statistical summaries about Activity Sets and Activity Set interdependencies with SPSS™. However, it is possible to interpret the table "by eye", by referring to the "Key for Activity Sets" area, in the top right-hand side of the TIMESLICER screen.

Correcting the Timeslice Table

As with the data produced by the RECORDER, there is an error checking process with the "Create Timeslice" process. This time, however, Action Recorder does check the "legitimacy" of each keystroke in the recorded data, against the Activity Sets defined in the Activity Set Definition area. Should any keystrokes be found which have not been defined to represent an activity, the video reference time at which the problem keystroke was found is displayed in the "Problems at Times:" area (top left of TIMESLICER).



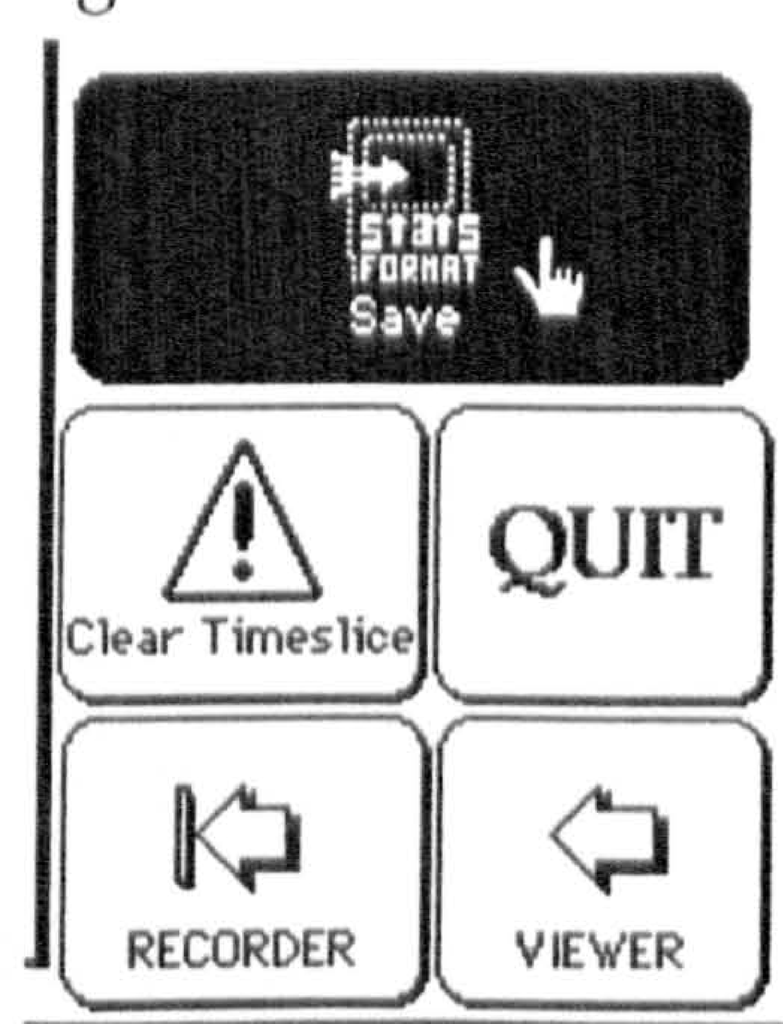
As with the New Data area, scroll through the Timeslice Table to the appropriate video reference point, find the relevant spot in the original videotape, and decide which key should have been pressed. Edit the Timeslice Table accordingly:



Here, there was no detectable change in the John or Mandy's activities, so it was decided that the "X" key was a slip on behalf of the observer.

Saving a Timeslice Table for Analysis with SPSS™ for the Macintosh® ,version 4.0

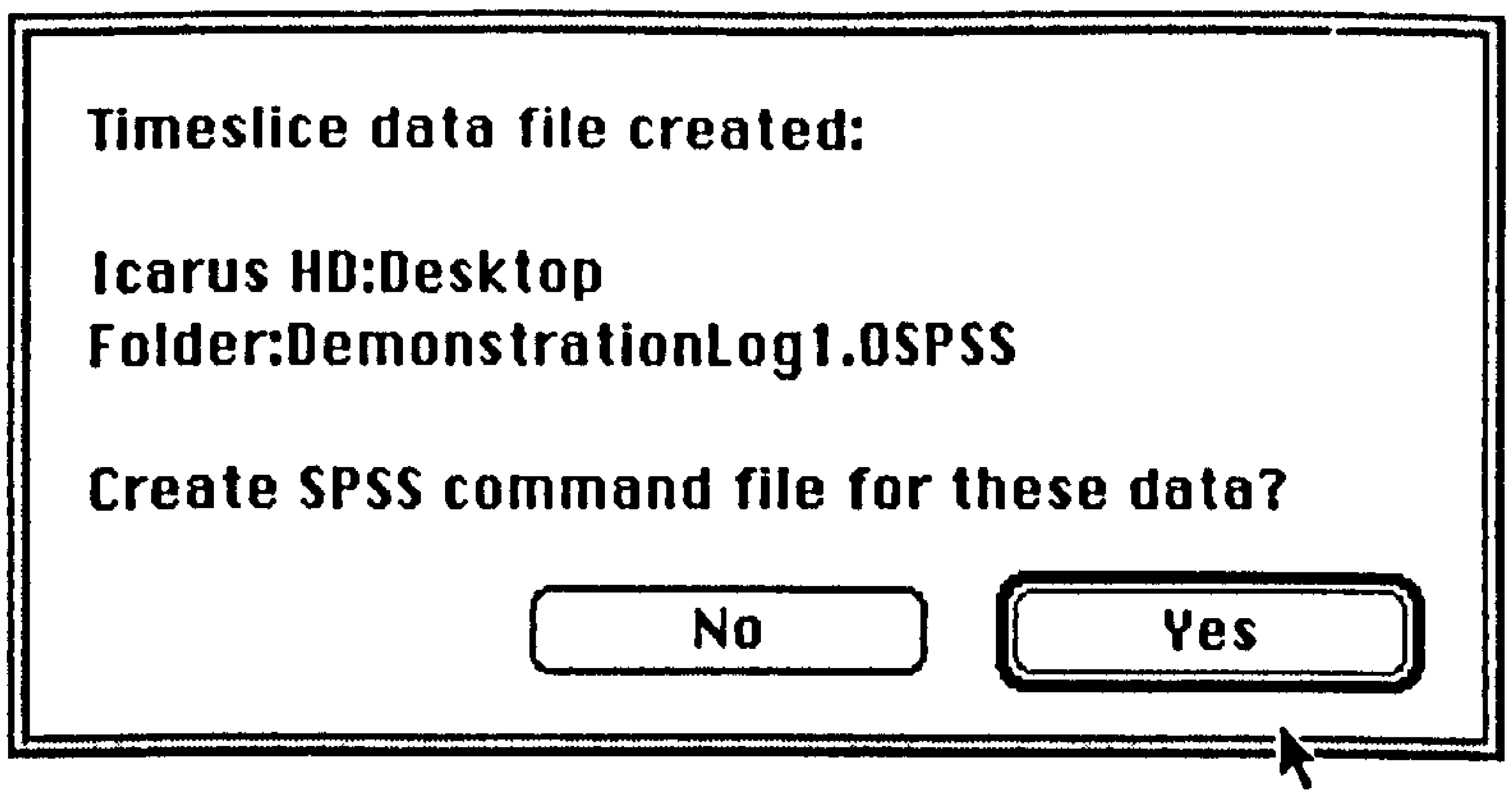
To save a Timeslice Table for future reference, click the large, "Stats Format Save" button, on the bottom right hand side of the TIMESLICER:



Action Recorder shows a standard Apple Macintosh file dialogue, asking for a name for this file. Action Recorder automatically offers the label originally used for the current data session, with the addition of "SPSS" to help identify the file as a Timeslice Table for SPSS. When the "Save" button on this dialogue is clicked, Action Recorder confirms the creation of this Timeslice data file, with the name and full path of the file[52].

---

[52]A "path name" is a list of all the folders enclosing the file, each separated by a colon (:) Note that the Apple Macintosh Desktop is really a folder in its own right, called "Desktop Folder",



This dialogue box also asks if you wish to create an SPSS command file. If you wish to go on to analyse the Timeslice Table with SPSS, clicking "Yes" is strongly recommended. This causes Action Recorder to put together a file containing the correct location of the Timeslice data file, with the correct Activity Set references and the appropriate SPSS commands. The command file is then automatically given the same name as the Timeslice data file you just saved, with the addition of ".bat".

Once a Timeslice data file and corresponding SPSS command file have been created, it is important not to move them around until you have actually used the command file to perform your analysis. Make sure you decide where you wish to keep your files before going on to create these files.

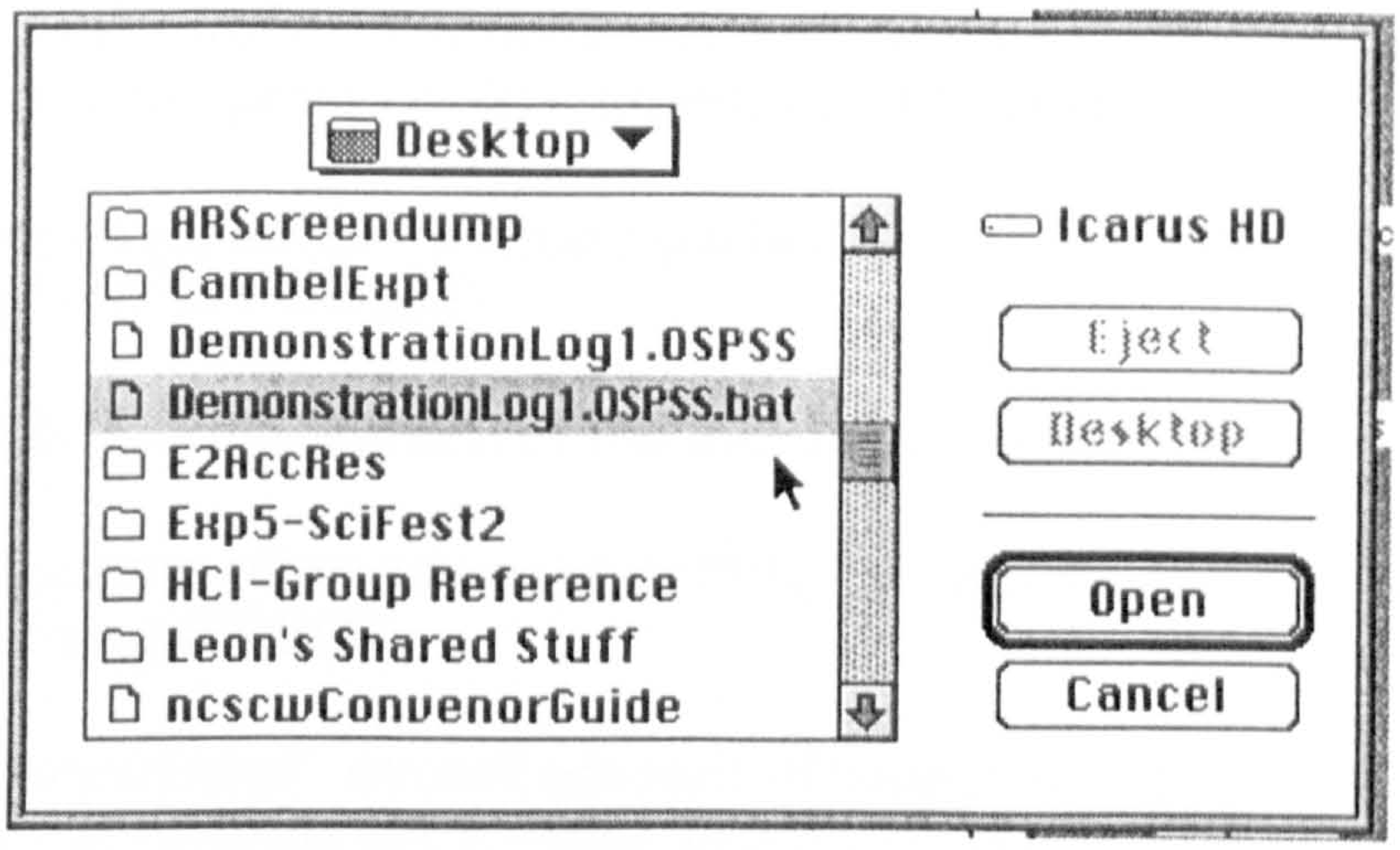
---

belonging to your computer's internal hard disk. In the above example, the computer's hard disk is called "Icarus HD", and the Timeslice data file is called "DemonstrationLog1.0SPSS".

A1a.8. Phase Three - Obtaining Statistical Summaries of Action Recorder Data

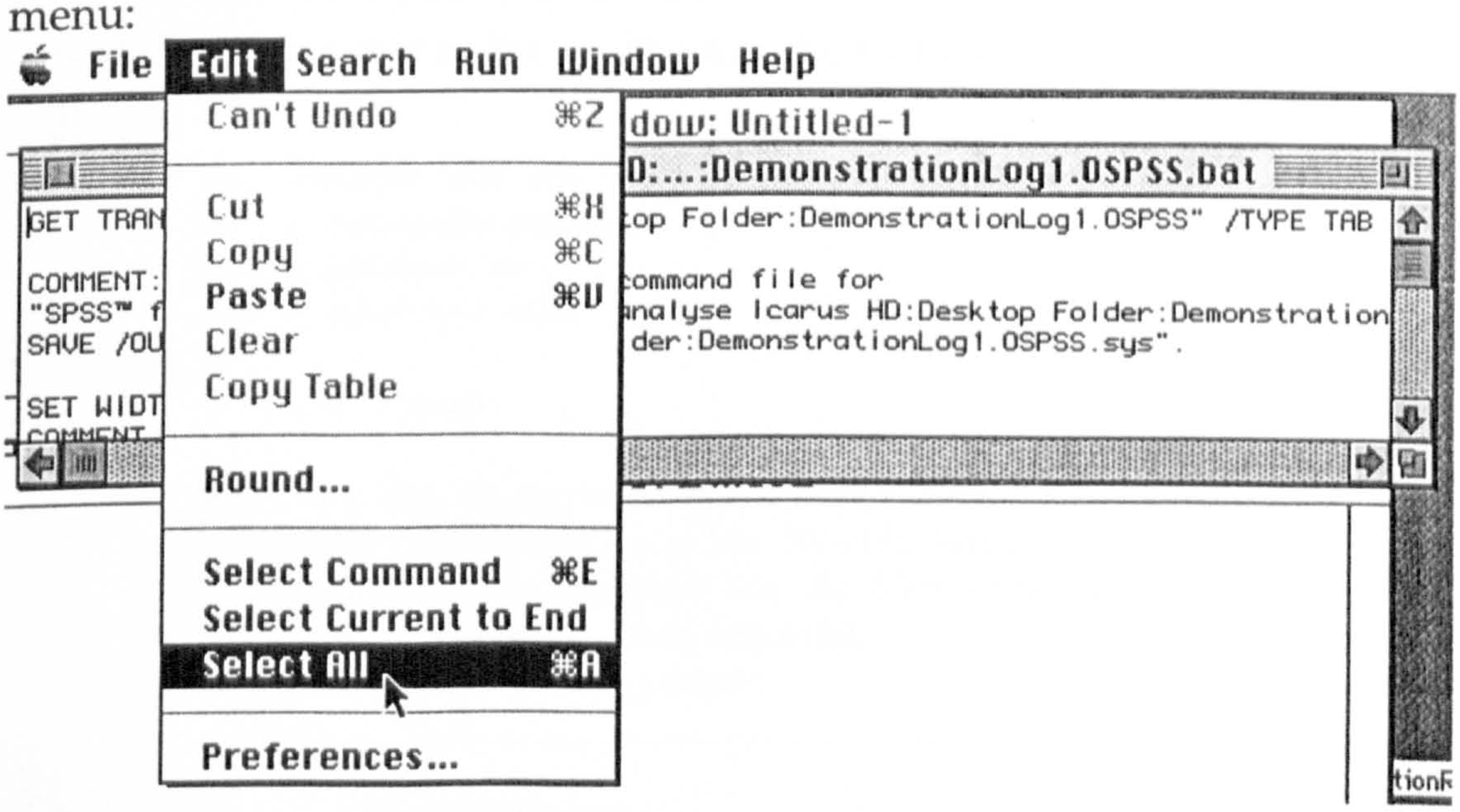
A1a.8.1. Automatic, Simple Summaries

The SPSS command file automatically generated by Action Recorder is very straightforward to use. Start SPSS in the usual way. From the "File" menu, select "Open" and then locate and select the command file:

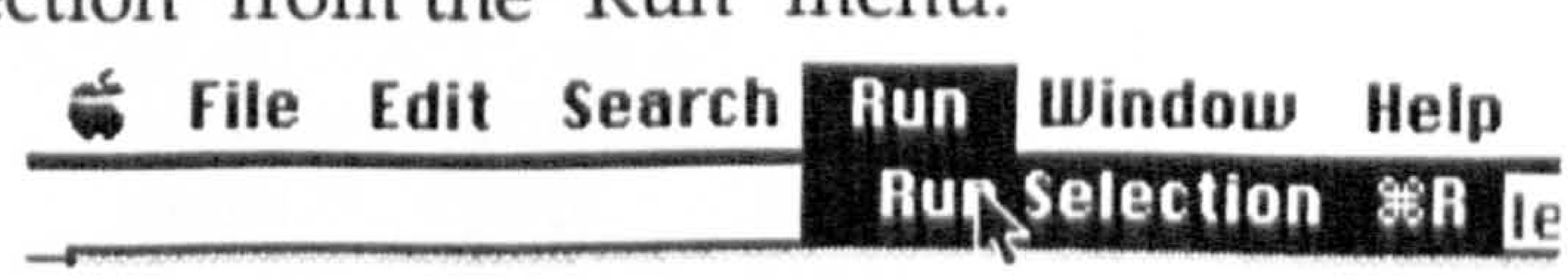


Next, select the SPSS "Input Window" showing Action Recorder's automatically generated command file. Click inside this window and choose "Select All" from SPSS's "Edit"

menu:



Finally, either type "R" with the command ('APPLE') key held down, or choose "Run Selection" from the "Run" menu:



SPSS will work for a few moments, and then show its "Output Window". The Output Window contains the statistical summaries of each of the Activity Sets you have defined.

A1a.8.2. Interpreting Action Recorder Data

The SPSS Output Window will contain a table of statistics for each of the Activity Sets you have defined. The statistics provided include:

the total amount of time (in seconds) a particular activity was engaged in, over the period of the recording(TOTTIME).

the average length of time a particular activity lasted, each time it occurred (AVGDUR).

the standard deviation of the average duration (STD).

the number of occasions an activity occurred, over the period of the recording (FREQ)

the *proportional* overall amount of time taken up by each activity, given in seconds per minute, in a particular activity (SPERMIN).

the *proportional* number of times each activity took places, also given per minute to aid comparison with time data (NPERMIN).

Above this table you will find a list of keystroke definitions, to help you see which of the letter codes applies to which of the activities you wish to know about:

```
->   COMMENT: KEY FOR THE ACTIVITIES
->   q johnlooks papers
->   w johnlook tv
->   o johnlooks else
->   .
->   list /cases.

There are 244,704 bytes of memory available.
The largest contiguous area has 244,472 bytes.
492 bytes of memory required for the LIST procedure.
240 bytes have already been acquired.
252 bytes remain to be acquired.
```

File: AGGREGATED FILE

JOHNLOOK	TOTTIME	AVGDUR	STD	FREQ	SPERMIN	NPERMIN
	.	.	.	0	.	.00
o	5.49	1.37	1.28	4	5.49	4.00
q	36.26	4.03	2.95	9	36.26	9.00
w	18.25	2.61	1.37	7	18.25	7.00

Number of cases read: 4 Number of cases listed: 4

In the example above[53], which describes the JOHNLOOK Activity Set, we can see that John spent a total of 36.26 seconds looking at his papers (letter code 'o'), compared to 18.25 seconds looking at Mandy, over the video link. However, we can also see that he looked at Mandy nearly as often as he looked at his papers.

```
->    COMMENT: KEY FOR THE ACTIVITIES
->    S Mandyloo tv
->    K Mandyloo papers
->    L Mandyloo elsewhere
->    .
->    list /cases.

There are 244,704 bytes of memory available.
The largest contiguous area has 244,472 bytes.
492 bytes of memory required for the LIST procedure.
240 bytes have already been acquired.
252 bytes remain to be acquired.
```

30-Jun-95   SPSS Release 4.0 for Macintosh
Page 5

File: AGGREGATED FILE

MANDYLOO	TOTTIME	AVGDUR	STD	FREQ	SPERMIN	NPERMIN
	.	.	.	0	.	.00
K	25.63	3.20	2.68	8	25.63	8.00
L	3.90	1.95	.42	2	3.90	2.00
S	30.47	4.35	1.58	7	30.47	7.00

Number of cases read: 4   Number of cases listed: 4

Mandy seems to have been just as interested in John as she was in her papers: each "look" at her papers lasted about the same time as each "look" at the tv/videophone; there were about as many of the looks, on the whole, and correspondingly, Mandy spent about as long overall looking at John on the tv/videophone as she did looking at her papers.

A1a.8.3. Producing Interdependency Summaries

To find out whether or not there is any evidence of interdependence between Activity Sets, some simple editing is required of Action Recorder's automatically produced 'simple summaries' command file.

---

[53]All the data in this user manual are fictitious.

There are five steps involved:

1) Decide which combinations of the Activity Sets you have defined are to be examined.

2) Go to the automatically generated command file in SPSS (*yourfilename* SPSS.bat) and scroll down until you find the word "AGGREGATE".

3) Just beneath the word "AGGREGATE" is a word "/break=", next to one of your Activity Set names. Replace this Activity Set name with a list of the Activity Set names forming the first of the combinations (if you intend doing more than one). Each Activity Set name should be separated from the next by a space.

4) Find the very next occurrence of the words "AGGREGATE" and "/break=". They should be only a couple of lines below the first occurrence. Do exactly the same here as you did in step (3).

5) Finally, select all the lines from one before the first occurrence of "AGGREGATE" (which should say "GET FILE" with a full file path name) to the line called "list /cases." Now choose "Run Selection" from SPSS' "Run" menu.

In our example, there are only two Activity Sets, hence only one combination is possible. So, to look for interdependency between John and Mandy's looking behaviour, change the command file as follows:

FROM:

```
GET FILE 'Icarus HD:Desktop
Folder:DemonstrationLog1.0SPSS.sys'.
AGGREGATE /OUTFILE *
 /presorted
 /break=johnlook
 /times=sum(duration).

AGGREGATE /OUTFILE *
 /break=johnlook
 /totTime=sum(times)
 /avgDur=mean(times)
 /std=sd(times)
 /Freq=NU(times).

compute SperMin=(totTime/60)*60.
compute NperMin=(Freq/60)*60.

COMMENT: KEY FOR THE ACTIVITIES
q johnlooks papers
w johnlook tv
o johnlooks else
.
list /cases.
```

TO:

```
GET FILE 'Icarus HD:Desktop
Folder:DemonstrationLog1.0SPSS.sys'.
AGGREGATE /OUTFILE *
 /presorted
 /break=johnlook mandyloo
 /times=sum(duration).

AGGREGATE /OUTFILE *
 /break=johnlook mandyloo
 /totTime=sum(times)
 /avgDur=mean(times)
 /std=sd(times)
 /Freq=NU(times).

compute SperMin=(totTime/60)*60.
compute NperMin=(Freq/60)*60.

COMMENT: KEY FOR THE ACTIVITIES
q johnlooks papers
w johnlook tv
o johnlooks else

S Mandyloo tv
K Mandyloo papers
L Mandyloo elsewhere
.
list /cases.
```

For clarity, the KEY FOR THE ACTIVITIES was also copied and pasted.

As with the simple summaries, SPSS will work for a few moments and then show you its "Output Window". This time you will see a more complicated table:

File: AGGREGATED FILE

JOHNLOOK	MANDYLOO	TOTTIME	AVGDUR	STD	FREQ	SPERMIN	NPERMIN
		.	.	.	0	.	.00
o	K	.37	.37	.	1	.37	1.00
o	L	.43	.43	.	1	.43	1.00
o	S	4.69	1.56	1.20	3	4.69	3.00
q	**K**	**13.20**	**1.47**	**.91**	**9**	**13.20**	**9.00**
q	L	2.68	1.34	.51	2	2.68	2.00
q	S	20.38	2.55	1.64	8	20.38	8.00
w	K	12.06	2.41	1.11	5	12.06	5.00
w	L	.79	.40	.39	2	.79	2.00
w	**S**	**5.40**	**1.35**	**.29**	**4**	**5.40**	**4.00**

Number of cases read: 10 Number of cases listed: 10

30-Jun-95 SPSS Release 4.0 for Macintosh

This table provide frequency and temporal information about all possible combinations of activity between two Activity Sets. Here, we can see that Mandy and John looked at each other simultaneously ('w' and 'S') for 5.4 seconds per minute, and on 4 occasions. The average duration of this state was 1.35 seconds (standard deviation = 0.29). Mandy and John looked at their papers simultaneously for more than twice this amount of time ('q' and 'K'), at 13.2 seconds per minute, and twice as often, at 9 times in the minute. However, the average duration of each simultaneous look at papers was about the same, at 1.47seconds (standard deviation = 0.91).

Just what the figures might mean to you, as an investigator, is a function of what you might expect to see. This is why step (1) of the five steps to interdependency analysis is probably the most important of all.

Further advice on how to get the most out of a statistical analysis of these data is available from the author.

A1a.9. A Note on Timer Accuracy

Action Record takes its time figures from the internal Macintosh timer. This timer works in units of one sixtieth of a second. For this reason, all Action

Recorder data is produce to two decimal places. The timings are not to centisecond accuracy - a decimal system requires two decimal places to capture sixtieths of a second.

There will inevitably be some delay between the making of a keystroke and the recording of the time at which it occurred. Every effort has been made to keep this lag to an absolute minimum. The lag will vary according to the type of Macintosh you have (and the CPU it uses) and the number of system extensions you use which make frequent call on the CPU. If accuracy is at an absolute premium for you, turn off all system extensions. In any case, you are especially advised to disable any clock programmes which show the time on your screen.

Lag in itself does not affect the type of statistical summaries produced by Action Recorder, since they are either frequency or of a duration-until-change nature. In other words, lag until the 'on' signal will be very nearly identical to lag until the 'off' signal, so the resultant duration calculation will be very nearly as exact as the responses of the observer allows.

Appendix 1b: Code Samples from Action Recorder

Action Recorder is written in HyperTalk 2.2, the scripting language of HyperCard 2.2. HyperCard has an object structure, comprising an entity called a "stack" containing objects in a hierarchy. The hierarchy begins at the stack and runs through "backgrounds" to "cards", with the additional complication that both backgrounds and cards can contain "button" and "field" objects. This is resolved in that everything at the card layer is higher than everything at the background layer. In this way, a stack is a set of backgrounds, are a set of buttons, fields and cards, are a set of buttons and fields. Buttons are intended to be used as direct-manipulation command interfaces and fields to collect, contain or display data.

All of these, including the stack root, can contain code scripts either as message handlers or as function definitions. Communication between objects and their code is achieved either through message passing or parameter passing via function calls. Message passing involves either user events, collected by high objects and processed and/or passed to lower objects, or the initiation of special message events by higher objects to be handled by lower objects. Action Recorder uses a combination of these methods.

Fields have the unusual property of a direct relationship to physical memory, so that any change to the contents of a field is immediately written to disk. This has advantages, in that data is relatively secure against system failure, and disadvantages, in that data manipulation in fields is considerably slower than via variables, as variables exist only in electronic memory. Since Action Recorder is required to perform operations that are both time-critical and where data integrity is paramount, both variables and fields are used for data capture and manipulation.

In the code below, some comments are included and marked by lines beginning " -- ". Conventionally all local variable names begin "lvar..." and all global variables begin "gvar...". Identification of fields is explicit in HyperTalk by the key words "field" or "fld". Fields belonging to backgrounds are prefixed by "background" or "bkgnd" and fields belonging to cards are prefixed by "card" or "cd". All objects in HyperTalk can be referred to by name (eg cd fld "Hello" of cd "Introduction"), object number (eg cd fld 5 of cd 5) or absolute object number (eg cd fld id 86). Failure to specify which card or background a particular object belongs to is not necessary but can be dangerous, where objects from several fields or backgrounds are invoked by a script. Action Recorder refers to objects in various ways, as a trade off between security, code readability and parsimony.

Five functions are included in this Appendix as the major components of Action Recorder (discounting all GUI code). These are for:

- logging keystrokes with time stamps
- sorting logged keystrokes by time code
- compiling a lookup table for Activity Set membership
- sorting logged keystrokes by Activity Set membership and time of occurrance to compile a "Time Slice Table': a list of Activity Set states over the course of the logged period.

A1b.1. Logging keystrokes with time stamps

Keystrokes are one class of event that may passed as messages. Whenever Action Recorder is the active window, any keystrokes pass through its objects, beginning at the card level. The following script is used to define a script for the RECORDER card so that any and all user keystrokes are logged with a time, from the computer's internal clock, until the RECORDER button "STOP" (cd btn id 8) receives a mouse click event, or the user pre-set time out value trips to send a mouse event to this button with the same effect. Time stamps are in Macintosh 'ticks' when they are recorded. To maintain temporal integrity, the absolute bare minimum of processing is carried out at this stage. No error checking or conversion to human-readable time is done until after the "STOP" button is activated.

The following script is defined on

```
-- Initiate the keystroke logging routines for the card
-- USING GLOBAL VARIABLES FOR DATA COMPILATION PERIOD
set the script of this card to ¬
"on keyDown theKey" && return && ¬
"global gvarRecord, gvarStopTime" && return && ¬
"put the ticks && theKey && return after gvarRecord" && return && ¬
"put theKey into card field Current" && return && ¬
"if the ticks >= gvarStopTime" && return && "then" && return && ¬
"send mouseUp to cd btn id 8" && return && ¬
"end if" && return && ¬
"end keyDown"
```

The script of the "STOP" button additionally allows for temporary interruptions by recording a reference time, in ticks, to allow an appropriate adjustment to the timestamps after resumption:

```
on mouseUp
  global gvarStartTime, gvarStopTime, gvarRecord, gvarCollectState, gvarUndoTarget
  set the script of this card to ""

  -- If a premature stoppage, record the time when recording was suspended in ticks
  if gvarStopTime > the ticks
  then
    put the ticks into gvarStopTime
  end if

  set enabled of me to false
  set enabled of cd btn "Start" to true
  put gvarRecord after card field Log

  -- display the reference time of the session when the recording was suspended
  set the numberFormat to 0.0
  put return & "Stopped at:" && (gvarStopTime-gvarStartTime)/60 && "secs" ¬
  into line 2 of cd fld "StartField"

  repeat with r = 1 to the number of lines in cd fld "CurrentSets"
    put gvarStartTime+(word 3 of cd fld "StartField")*60 into word 1 of line r of cd fld
"Log"
  end repeat

  -- delete line (the number of lines in cd fld "CurrentSets") +1 of cd fld "Log"
  put gvarStopTime into word 1 of the last line of cd fld "Log"

  -- copy the variable containing the keystrokes into a field
  set the cursor to busy
  put "new" into gvarUndoTarget
  go cd "DataLists"
  put cd fld "RawLog" of cd "DataLists" into cd fld "TempStore"
  send makeRaw to cd fld "RawLog" of cd "DataLists"
  set the cursor to hand
end mouseUp
```

A1b.2. Sorting logged keystrokes by timestamp

The keystrokes are sorted into temporal order with the "orderRaw" function, defined on the background to the RECORDER card. "orderRaw has to operate over a mixture of ASCII and integer data. It first converts the keystroke characters to ASCII codes, concatenates these to the "ticks" time code and then uses a built-in HyperTalk sort function to order them in ascending sequence. It then splits off the ASCII character code from the composite code and converts this back to an ASCII character. The parameter is simply to distinguish between the two possible fields, "New" or "Existing", the VIEWER card to which this function might apply.

```
-- *************v| THE ORDERRAW FUNCTION |v*************
function orderRaw lvarSortFac
  -- SEQUENCE BY CONVERTING KEYSTROKES TO ASCII AND CONCATENATING WITH TICKS ...
  set the numberFormat to 000
  put empty into lvarSortList

  put the number of lines in cd fld lvarSortFac into lvarLength

  -- Update the Activity Indicator 'feedback window'
  put quote & "Sequencing" & quote && lvarLength && "records" ¬
  into cd fld ActivityIndicator

  repeat with r = 1 to lvarLength
    if line r of cd fld lvarSortFac is empty
    then
      put lvarLength-1 into lvarLength
      next repeat
    end if
    put word 1 of line r of cd fld lvarSortFac & ¬
    the charToNum of (word 2 of line r of cd fld lvarSortFac)¬
    & return after lvarSortList
    set the cursor to busy
  end repeat
  -- sort lines of cd fld lvarSortFac ascending numeric
  sort lines of lvarSortList ascending numeric
  -- ... AND THEN CONVERTING THEM BACK AGAIN***
  repeat with r = 1 to lvarLength
    get the length of line r of lvarSortList
    put space before char (it - 2) of line r of lvarSortList
    put (the first word of line r of lvarSortList) && ¬
    the numToChar of (the last word of line r of lvarSortList)¬
    into line r of lvarSortList

    -- put r into word 2 of line 3 of cd fld ActivityIndicator
  end repeat
  put lvarSortList into cd fld lvarSortFac
  return "done"
end orderRaw
-- *******^| END FUNCTION ORDERRRAW |^************
```

A1b.3. Compiling a lookup table for Activity Set membership

The following function, "activitySetParse", is defined on the background to the RECORDER card. It uses the information from the "Define Sets" field on the RECORDER card (ActionsField1, cd fld id 11, cd id 4372) to compile a lookup table of Activity Set definitions (ActionSetList, cd fld id 4, cd id 7561). The lookup This table is used to sort logged keystrokes into a "Time Slice Table" of Activity Set state descriptions.

```
-- *******vl THE ACTIVITYSETPARSE FUNCTION  lv**********
function activitySetParse lvarAppender
 put empty into cd fld "ActionSetList" of cd id 7561
 put empty into lvarExistingActivitySets
 put cd fld "ActionsField1" of cd id 4372 into lvarFocusField
 put empty into cd fld "ActivityKey" of cd id 7561

 -- Establish the number of iterations required to deal with each
 -- of the activities defined in the Key on the CollectionControlPanel

 put the number of lines in lvarFocusField into lvarNumActivities

 -- ..........find out how many action sets have been defined......
 -- **  The cd fld "ActionSetList" of cd "TimeSlicedSort" is a list   ******
 -- **  of all the action set names stored as a space-separated array.  ******

 repeat with a = 1 to lvarNumActivities
  set the cursor to busy
  if line a of lvarFocusField is empty then next repeat
  put char 1 to 8  of (word 2 of line a of lvarFocusField) into lvarActionSetName

  if (space & lvarActionSetName & space) is not in cd fld "ActionSetList"  of cd id 7561
  then
   put (1 + lvarExistingActivitySets) into lvarExistingActivitySets
   put space & lvarActionSetName into line lvarExistingActivitySets¬
   of cd fld "ActionSetList"  of cd id 7561
   put lvarExistingActivitySets into lvarPosition
  else
   repeat with ax =  1 to lvarExistingActivitySets

    if lvarActionSetName is word 1 of line ax of cd fld "ActionSetList"  of cd
"TimeSlicedSort"
    then
     put ax into lvarPosition
     exit repeat
    end if
   end repeat
  end if

  put (space & (word 1 of line a of lvarFocusField))¬
  after line lvarPosition of cd fld "ActionSetList"  of cd "TimeSlicedSort"
```

[CONTINUED OVERLEAF]

```
[CONTINUED FROM OVER]

   if (space & lvarActionSetName & space) is not in cd fld "ActivityKey" of cd id 7561
   then
     put return & line a of cd fld "ActionsField1" of cd id 4372 after¬
     the last line of cd fld "ActivityKey" of cd id 7561
   else
     repeat with c = 1 to the number of lines in cd fld "ActivityKey" of cd id 7561
       if (word 2 of line c of cd fld "ActivityKey" of cd id 7561) contains lvarActionSetName
       then
         put return & line a of cd fld "ActionsField1" of cd id 4372 after¬
         line (c - 3 + (the number of words in line lvarPosition of ¬
         cd fld "ActionSetList" of cd id 7561)) of¬
         cd fld "ActivityKey" of cd id 7561
         exit repeat
       end if
     end repeat
   end if
  end repeat

  delete the first line of cd fld "ActivityKey" of cd id 7561
  repeat with r = 1 to the number of lines in cd fld "ActivityKey" of cd id 7561
   if line r of cd fld "ActivityKey" of cd id 7561 is empty ¬
   then delete line r of cd fld "ActivityKey" of cd id 7561
  end repeat

  return the number of lines in cd fld "ActionSetList" of cd "TimeSlicedSort"
end activitySetParse

-- **********^| END THE ACTIVITYSETPARSE FUNCTION |^************
```

A1b.4. Compiling Time Slice Table of Activity states, on the basis of Activity Set membership and time of state logging

Compilation of the Time Slice Table is initiated by activation of the "Create Timeslice" button on the VIEWER card.

First, a list of column headings is written in to the cd fld "SPSSList" on the TIMESLICER card (cd "TimeSlicedSort", cd id 7561), by taking the first item from each line in the Activity Set lookup table (cd fld "ActionSetList") on the TIMESLICER card. Additional column headings are added for the time at which a keystroke was made, the duration of each Activity state and any undefined keystrokes (as "exceptions"). Each of these column headings will serve as variable names for the SPSS statistics package, when the Time Slice Table is later exported to it.

The Activity Set state is then initialized by creating a first line of state values beneath the column headings, with initial time information. This is then a total state vector. There after, the ASCII character in each of the entries in the log field (either "New" or "Existing" on the VIEWER card) is compared against the

Activity Set lookup table. When the correct Activity Set is found, the previous state vector is copied down to the next line in the SPSSList field and the new character and its time are used to overwrite the appropriate vector components. It is then possible to calculate the duration of the previous set of states by subtracting the new time from the old time value.

Whenever a character is not found in the cd fld "ActionSetList", the old state vector is copied down with the new time and duration calculations as before. However the undefined character is written to the last vector position (i.e. into the "exceptions" column) and the time additionally written to the cd fld "ProblemList" (cd fld id 81) on the TIMESLICER card. On completion of the Time Slice Table, a dialogue is presented to the user directing their attention to this dialgue box if undefined keystrokes have been encountered.

There is a 30,000 character limit to HyperCard fields. Action Recorder dynamically creates new SPSSList cards whenever this character limit threatens to be exceeded.

```
-- ********************************************************
-- Create the list of Activities in each set comprising the time-slice
-- by looking at the ActionFields of the CollectionControlPanel
-- ********************************************************
push this card
go to card id 7561
set the numberformat to 0.##

-- ***** THE ACTIVITYSETPARSE FUNCTION IS DEFINED ON THE BACKGROUND
********
put activitySetParse("no") into lvarNumActivitySets

pop card

if ((12 * lvarNumActivitySets) + 140) < 300
then
  set the width of cd fld 5 of cd id 7561 to ¬
  ((12 * lvarNumActivitySets) + 160)
else
  set the width of cd fld 5 of cd id 7561 to 300
end if
set the topLeft of cd fld 5 of cd id 7561 to 5, 135
```

[CONTINUED OVERLEAF]

[CONTINUED FROM OVER]

```
-- *******************************************************
-- ******** WHEN THERE IS NO ALREADY SORTED DATA IN THE OUTPUT LIST ****
-- *******************************************************
-- ***** CREATE THE TEMPLATE STRUCTURE FOR THE TIME-SLICE TABLE ON CARD ***
-- ************          id 7561          ***********
-- *******************************************************
-- ACTIONSET1  ACTIONSET2 ... ACTIONSET'n'   TIME     DURATION   EXCEPTION
-- dummy val1  dummy val2 ... dummy val 'n'  start time   token      token
-- *******************************************************

put (the number of lines in cd fld gvarCurrentSPSS of cd id 7561) ¬
into lvarTableRow

if lvarTableRow is 0
then
  put 2 into lvarTableRow

  put char 1 to 8 of¬
  (word 1 of line 1 of cd fld "ActionSetList" of cd id 7561)&tab&space ¬
  into line 1 of cd fld gvarCurrentSPSS of cd id 7561
  put "0"&tab&space into line 2 of cd fld gvarCurrentSPSS of cd id 7561

  repeat with d = 2 to lvarNumActivitySets
    put char 1 to 8 of¬
    (word 1 of line d of cd fld "ActionSetList" of cd id 7561) &tab&space¬
    after line 1 of cd fld gvarCurrentSPSS of cd id 7561
    put "0" &tab&space after line 2 of cd fld gvarCurrentSPSS of cd id 7561
  end repeat

  put "TIME" &tab& "Duration" &tab& "Exception" after ¬
  line 1 of cd fld gvarCurrentSPSS of cd id 7561

  put "0" &tab&space& "0" after line 2 of cd fld gvarCurrentSPSS of cd id 7561

  -- ** ENTER THE VIDEO START TIME IN THE TIME-SLICE TABLE **
  put (word 1 of line 1 of cd fld lvarTimeSliceTarget)&tab into ¬
  word (lvarNumActivitySets + 1) of line 2 of cd fld gvarCurrentSPSS of cd id 7561
end if

-- *******************************************************
-- *******      See which Activities belong to each of the Activity Sets      *****
-- *******************************************************
-- Instantiating the repeat control variable, lvarCycles from
-- the number of logged events in the data field

put the number of lines in cd fld lvarTimeSliceTarget into lvarCycles
-- TO MAKE A NOTE OF ANY KEYSTROKES NOT BELONGING TO PREDEFINED SETS
put empty into lvarProblemList

-- --------UPDATE THE FEEDBACK WINDOW--------
put quote & "COMPILING TIMESLICE TABLE" & quote & return & ¬
"Action 1 processed out of" && lvarCycles && "records" & return ¬
into cd fld 5
-- ----------------------------------------
```

[CONTINUED OVERLEAF]

[CONTINUED FROM OVER]

```
  -- SET A LIMIT FIGURE FOR THE NUMBER OF CHARACTERS PERMITTED IN A FIELD
  put 30012-(lvarNumActivitySets*10) into lvarCharLim

  repeat with a = 1 to lvarCycles
   -- INSTANTIATE THE TARGET EVENT AND TIME
   put word 1 of line a of cd fld lvarTimeSliceTarget into lvarActionTime
   put word 2 of line a of cd fld lvarTimeSliceTarget into lvarActionTarget
   -- FIND TIME OF PREVIOUS EVENT, TO CALCULATE THE STATE DURATION
   put word (lvarNumActivitySets + 1) of ¬
   line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561 into lvarLastTime
   delete the last char of lvarLastTime
   put (lvarActionTime - lvarLastTime) into lvarDuration

   -- CHECK THAT THE OUTPUT TABLE FIELD IS NOT ABOUT TO OVERFLOW (30000 CHAR MAXIMUM)
   -- IF IT IS, INCREMENT THE FIELD REFERENCE NUMBER, TO SHIFT THE OUTPUT FIELD.
   if the number of chars in cd fld gvarCurrentSPSS of cd id 7561 > lvarCharLim
   then
    lock screen
    push this card
    go card id 7561
    doMenu New Field
    put gvarCurrentSPSS + 1 into gvarCurrentSPSS
    put lvarDuration &tab into word (2 + lvarNumActivitySets) of¬
    the last line of cd fld (gvarCurrentSPSS - 1) of cd id 7561
    put the last line of cd fld (gvarCurrentSPSS - 1) of cd id 7561 into lvarCarryOver
    set the style of the last cd fld to scrolling
    set the width of the last cd fld to the width of cd fld 5
    set the topleft of the last cd fld to (5*gvarCurrentSPSS), (105 +
(10*gvarCurrentSPSS))
    set the textFont of the last cd fld to Courier
    set the textSize of the last cd fld to 10
    set the lockText of the last cd fld to TRUE
    set the script of the last cd fld to ¬
    "on mouseUp" & return &¬
    "if the number of me = the number of cd flds" & return & "then" & return & ¬
    "set the locktext of me to false" & return & "else" & return &¬
    "select me" & return &¬
    "repeat for (the number of cd flds - 4)" & return &¬
    "set the cursor to busy" & return &¬
    "domenu Bring Closer" & return &¬
    "end repeat" & return & "end if" & return &¬
    "set the cursor to hand" & return &¬
    "type tab with commandKey" & return  & "end mouseUp"

    type tab with CommandKey
    put lvarCarryOver into cd fld gvarCurrentSPSS of cd id 7561
    pop card
    unlock screen
    put 1 into lvarTableRow
   end if
   -- --------UPDATE THE FEEDBACK WINDOW--------
   put a into word 3 of cd fld 5
   set the cursor to watch
     put FALSE into lvarFound
   put 1 into s
```

[CONTINUED OVERLEAF]

[CONTINUED FROM OVER]

```
  -- BEGIN CYCLING THROUGH THE SETS OF ACTIVITIES FROM THE PARSED TABLE,
"ActionSetList"
  repeat until lvarFound
   if s <= lvarNumActivitySets
   then
    -- PROPOSE AN ACTIVITY SET (using the loop variable 's')
    put the number of words in line s of cd fld "ActionSetList" of ¬
    cd id 7561 into lvarActivities

    -- PROPOSE AN ACTIVITY (using the loop variable 'n')
    -- ** Note: the first word of the line is the Activity Set label rather than an **
    -- ******        Activity itself and is therefore discounted         *********
    repeat with n = 2 to lvarActivities

     put word n of line s of cd fld "ActionSetList" of cd id 7561¬
     into lvarDefinedActivity
     -- SEE IF THE TARGET EVENT IS A MEMBER OF THIS SET OF ACTIVITIES
     if charToNum(lvarActionTarget) = charToNum(lvarDefinedActivity)
     then
      put TRUE into lvarFound
      -- ** If the new Activity does not change the state of the world
      -- ** it is a redundant keystroke, thus no action need be taken
      -- ** AND CHECKING FOR SIMULTANEOUS EVENTS, CORRECTING
      -- ** FOR THE CONSEQUENT ZERO-LENGTH STATES
      put chartoNum(word s of line lvarTableRow of cd fld gvarCurrentSPSS¬
       of cd id 7561) into lvarLastAction
      if chartoNum(lvarActionTarget) is not lvarLastAction¬
      and lvarDuration is not zero
      then
       put lvarDuration &tab into word (2 + lvarNumActivitySets) of ¬
       line ( lvarTableRow) of cd fld gvarCurrentSPSS of cd id 7561

       put return & word 1 to (2 + lvarNumActivitySets) of ¬
       (line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561) ¬
       after line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561

       put lvarActionTime &tab into word (lvarNumActivitySets + 1) of ¬
       line ( 1 + lvarTableRow) of cd fld gvarCurrentSPSS of cd id 7561
       put lvarTableRow + 1 into lvarTableRow

       put lvarActionTarget &tab into word s of the last line of ¬
       cd fld gvarCurrentSPSS of cd id 7561

     else if lvarDuration is zero
     then
      put lvarActionTarget &tab into word s of the last line of ¬
      cd fld gvarCurrentSPSS of cd id 7561
     end if
     exit repeat
    end if -- THIS ENDS THE CONTINGENCY FOR BEING A MEMBER OF A SET

   end repeat -- THIS ENDS THE 'PER ACTIVITY' LOOP
```

[CONTINUED OVERLEAF]

[CONTINUED FROM OVER]

```
        -- *THIS IS TO CATCH AND RECORD INSTANCES OF UNDEFINED CHARS IN THE
RAWLOG*
      else
        if lvarDuration is 0
        then
          put space & lvarActionTarget after ¬
          the last line of cd fld gvarCurrentSPSS of cd id 7561
        else
          put lvarDuration &tab into word (2 + lvarNumActivitySets) of ¬
          line ( lvarTableRow) of cd fld gvarCurrentSPSS of cd id 7561

          -- ONLY THE FIRST 'n' ITEMS ARE CARRIED FORWARD, TO MISS UNDEFINED CHARS
          put return & word 1 to (2 + lvarNumActivitySets) of ¬
          (line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561) ¬
          after line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561

          put lvarTableRow + 1 into lvarTableRow
          put lvarActionTime &tab into word (lvarNumActivitySets + 1) of ¬
          line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561

          put space & lvarActionTarget after ¬
          line lvarTableRow of cd fld gvarCurrentSPSS of cd id 7561
        end if
        put lvarActionTime & return after lvarProblemList
        exit repeat
      end if

      put s+1 into s
    end repeat -- THIS ENDS THE 'ACTIVITY SET' LOOP

  end repeat -- THIS ENDS THE SUCCESSIVE LOGGED EVENTS

  delete the last line of cd fld gvarCurrentSPSS of cd id 7561

  if lvarProblemList is empty
  then
    put "No undefined keys found. TimeSlice table OK" into cd fld "ProblemList" of cd id 7561
  else
    if lvarActionTime is word 1 of the last line of lvarProblemList
    then
      delete the last line of lvarProblemList
    end if
    put lvarProblemList into cd fld "ProblemList" of cd id 7561
  end if

  put (cd fld "LogName") &"SPSS"¬
  into cd fld "SessionTitle" of cd id 7561

  -- Audible signal for the end of the sort-by-action process
  play harpsichord 45 50 45 50 55
  answer "Timeslice Table Complete"
  go to card id 7561
-- ************ END SCRIPTING FOR TIME SLICE TABLE *********
```

Appendix 1c: Sample Action Recorder Activity logs, taken from the case study reported in Chapters 4 and 6

Key:

f = subject 15 looks at the video link.
d = subject 15 looks at the computer.
s = subject 15 looks at the paper notes.
a = subject 15 looks elsewhere.

p = subject 15 is speaking.
q = subject 15 is quiet.

h = subject 13 looks at the video link.
j = subject 13 looks at the computer.
k = subject 13 looks at the paper notes.
l = subject 13 looks elsewhere.

z = subject 13 is speaking.
x = subject 13 is quiet.

Activity Codes and Times In Seconds

```
k 04.35
s 04.37
p 04.37
z 04.37
f 05.65
s 07.13
h 08.10
x 09.20
q 09.47
p 09.93
q 10.67
k 10.88
h 12.05
f 12.55
z 12.82
x 13.22
s 13.23
k 14.12
h 17.48
k 17.87
h 18.63
p 18.92
z 19.47
x 19.75
k 19.98
q 20.10
h 20.37
f 21.90
k 22.62
z 24.32
x 24.60
p 25.02
z 25.62
x 27.20
q 27.30
p 27.55
s 28.15
q 28.27
h 29.87
f 30.42
s 30.85
f 31.87
k 32.72
z 33.00
p 33.07
q 33.80
p 34.13
s 34.63
x 36.38
h 36.90
z 36.98
x 38.18
k 38.27
q 38.70
p 43.13
q 43.77
f 45.70
p 46.37
z 46.85
s 47.13
q 47.17
x 47.22
d 47.52
s 47.83
f 49.65
p 50.38
q 50.82
p 53.82
z 55.07
x 55.47
z 55.73
x 56.50
z 56.80
s 57.42
d 57.88
q 58.95
p 59.22
x 59.35
f 59.47
z 59.62
d 60.30
x 62.03
z 62.52
f 63.58
d 64.02
x 64.12
z 64.38
s 64.92
d 65.63
h 67.10
x 68.17
q 68.18
p 68.52
k 68.70
s 69.98
z 70.20
d 71.18
s 74.72
d 75.37
x 78.00
z 79.47
s 83.13
h 84.15
k 84.72
d 84.87
s 85.82
d 86.42
x 86.95
h 87.27
z 87.27
k 88.08
x 89.43
z 91.92
h 91.95
k 92.70
h 98.12
k 98.65
f 99.27
s 99.92
```

Appendix 1d - Sample Action Recorder "Time Slice Table" for case study subjects 15 and 13

The data given below are an excerpt of the Time Slice Table generated by Action Recorder from the Activity log for subjects 15 and 13 (see Chapters 4 and 6, and Appendix 1c).

The first two columns represent the Gaze Activity and Speech Activity of subject 15, and the second two represent subject 13's Gaze and Speech. The fifth column gives the time at which an Activity state became true, represented by the Activity codes on that row. The sixth column gives the duration of the state..

Key:

f = subject 15 looks at the video link.
d = subject 15 looks at the computer.
s = subject 15 looks at the paper notes.
a = subject 15 looks elsewhere.

p = subject 15 is speaking.
q = subject 15 is quiet.

h = subject 13 looks at the video link.
j = subject 13 looks at the computer.
k = subject 13 looks at the paper notes.
l = subject 13 looks elsewhere.

z = subject 13 is speaking.
x = subject 13 is quiet.

Looks15	Speaks15	Looks13	Speaks13	TIME	Duration
s	q	k	z	4.37	1.28
f	q	k	z	5.65	1.48
s	q	k	z	7.13	0.97
s	q	h	z	8.1	1.1
s	q	h	x	9.2	0.27
s	p	h	x	9.47	0.46
s	q	h	x	9.93	0.74
s	p	h	x	10.67	0.21
s	p	k	x	10.88	1.17
s	p	h	x	12.05	0.5
f	p	h	x	12.55	0.27
f	p	h	z	12.82	0.4
f	p	h	x	13.22	0.01
s	p	h	x	13.23	0.89
s	p	k	x	14.12	3.36
s	p	h	x	17.48	0.39
s	p	k	x	17.87	0.76
s	p	h	x	18.63	0.29
s	q	h	x	18.92	0.55
s	q	h	z	19.47	0.28
s	q	h	x	19.75	0.23
s	q	k	x	19.98	0.12
s	p	k	x	20.1	0.27
s	p	h	x	20.37	1.53
f	p	h	x	21.9	0.72
f	p	k	x	22.62	1.7
f	p	k	z	24.32	0.28
f	p	k	x	24.6	0.42
f	q	k	x	25.02	0.6
f	q	k	z	25.62	1.58
f	q	k	x	27.2	0.1
f	p	k	x	27.3	0.25
f	q	k	x	27.55	0.6
s	q	k	x	28.15	0.12
s	p	k	x	28.27	1.6
s	p	h	x	29.87	0.55
f	p	h	x	30.42	0.43
s	p	h	x	30.85	1.02
f	p	h	x	31.87	0.85
f	p	k	x	32.72	0.28
f	p	k	z	33	0.07
f	q	k	z	33.07	0.73
f	p	k	z	33.8	0.33
f	q	k	z	34.13	0.5
s	q	k	z	34.63	1.75
s	q	k	x	36.38	0.52
s	q	h	x	36.9	0.08
s	q	h	z	36.98	1.2
s	q	h	x	38.18	0.09
s	q	k	x	38.27	0.43
s	p	k	x	38.7	4.43
s	q	k	x	43.13	0.64
s	p	k	x	43.77	1.93
f	p	k	x	45.7	0.67
f	q	k	x	46.37	0.48

f	q	k	z	46.85	0.28	d	q	k	z	65.63	1.47
s	q	k	z	47.13	0.04	d	q	h	z	67.1	1.07
s	p	k	z	47.17	0.05	d	q	h	x	68.17	0.01
s	p	k	x	47.22	0.3	d	p	h	x	68.18	0.34
d	p	k	x	47.52	0.31	d	q	h	x	68.52	0.18
s	p	k	x	47.83	1.82	d	q	k	x	68.7	1.28
f	p	k	x	49.65	0.73	s	q	k	x	69.98	0.22
f	q	k	x	50.38	0.44	s	q	k	z	70.2	0.98
f	p	k	x	50.82	3	d	q	k	z	71.18	3.54
f	q	k	x	53.82	1.25	s	q	k	z	74.72	0.65
f	q	k	z	55.07	0.4	d	q	k	z	75.37	2.63
f	q	k	x	55.47	0.26	d	q	k	x	78	1.47
f	q	k	z	55.73	0.77	d	q	k	z	79.47	3.66
f	q	k	x	56.5	0.3	s	q	k	z	83.13	1.02
f	q	k	z	56.8	0.62	s	q	h	z	84.15	0.57
s	q	k	z	57.42	0.46	s	q	k	z	84.72	0.15
d	q	k	z	57.88	1.07	d	q	k	z	84.87	0.95
d	p	k	z	58.95	0.27	s	q	k	z	85.82	0.6
d	q	k	z	59.22	0.13	d	q	k	z	86.42	0.53
d	q	k	x	59.35	0.12	d	q	k	x	86.95	0.32
f	q	k	x	59.47	0.15	d	q	h	z	87.27	0.81
f	q	k	z	59.62	0.68	d	q	k	z	88.08	1.35
d	q	k	z	60.3	1.73	d	q	k	x	89.43	2.49
d	q	k	x	62.03	0.49	d	q	k	z	91.92	0.03
d	q	k	z	62.52	1.06	d	q	h	z	91.95	0.75
f	q	k	z	63.58	0.44	d	q	k	z	92.7	5.42
d	q	k	z	64.02	0.1	d	q	h	z	98.12	0.53
d	q	k	x	64.12	0.26	d	q	k	z	98.65	0.62
d	q	k	z	64.38	0.54	f	q	k	z	99.27	0.65
s	q	k	z	64.92	0.71	s	q	k	z	99.92	0.58

Appendix 1e: Annotated version of an SPSS Syntax file, generated by Action Recorder

The following is a generic copy of a command or "syntax" file, as generated by Action Recorder to process Time Slice Tables with the SPSS statistics package. Time Slice Tables are created as a set of tab delimited values separated by line breaks. This format allows Activity Set names generated by Action Recorder from users' Activity definitions to be used as variable names in SPSS. For this sample command file, italicised entries would have values given by Action Recorder.

```
GET TRANSLATE /FILE "drive:folder:filename" /TYPE TAB /FIELDNAMES.
```

Depending on how many Activity Sets are included, the Time Slice Table may have to be read many times. After the original Time Slice Table file has been read in to SPSS, it is converted into a file of type ".sys", a special fast format for SPSS.

```
COMMENT: This is an ActionRecorder command file for
"SPSS™ for the Macintosh® 4.0", to analyse drive:folder:filename.
SAVE /OUTFILE "drive:folder:filename.sys".
SET WIDTH 100.
COMMENT (get full system file).

GET FILE "drive:folder:filename.sys".
```

The first use of the "AGGREGATE" command is with the "PRESORTED" parameter. Presorted forces Aggregate to collapse only neighbouring cases of Activity state with the same value of the specified variable (i.e. Activity Set). The Activity State durations are summed across successive cases whilst the current Activity remains unchanged, the combined values assigned to a variable called "times".

```
AGGREGATE /OUTFILE *
  /presorted
  /break= ACTIVITYSET1
  /times=sum(duration).
```

The "AGGREGATE" command without the "PRESORTED" parameter collapses all of the Activity state table together for each of the Activities in the specified Activity Set (given by the "BREAK VARIABLE"), regardless of their moment of occurrence. Summary statistics are calculated for the total amount of time spent in each Activity state, the mean and standard deviation of each Activity duration and the number of occasions on which they were engaged over the interaction period.

```
AGGREGATE /OUTFILE *
 /break= ACTIVITYSET1
 /totTime=sum(times)
 /avgDur=mean(times)
 /std=sd(times)
 /Freq=NU(times).
```

The length of interaction, *SESSIONLENGTH*, is calculated in advance by Action Recorder. It is used to calculate standardized versions of the summary time-in-state and count data, by expressing each as parts of a minute.

```
compute SperMin=(totTime/SESSIONLENGTH)*60.
compute NperMin=(Freq/SESSIONLENGTH)*60.
```

Finally, an interpretation key is included as a comment, so that the SPSS results for each Activity Set can be more easily understood. The **list** command gives all values for all variables in the "active file", or the final aggregated version of the Time Slice Table.

```
COMMENT: KEY FOR THE ACTIVITIES
A ACTIVITYSET1 ACTIVITYOne
B ACTIVITYSET1 ACTIVITYTwo
C ACTIVITYSET1 ACTIVITYThree
D ACTIVITYSET1 ACTIVITYFour
.
list /cases.
```

Appendix 2a: master copy of hardship application

Participants reviewed fictitious case information for a ten to fifteen minute period prior to their VMI discussion. All information was presented electronically but in a skeletal form. They were told that all student information was committed to a computer database. They were given a blank master copy of the form applicants would have originally have filled in, in order to relate the electronic numeric headings with the content of each statement (see Appendix 2b). The blank form is reproduced below[54].

1. Personal Details
Surname First Names Title
Nationality Age
2. Income Self £ Partner £

Salary / earnings
Savings
Grants / sponsorship
Other Income

3. Dependents
a) Are you responsible for the financial support of a relative? Yes/No
b) Do you have to make maintenance payments as a result of a court order? Yes/No
c) Indicate below the extent of your financial obligations towards any dependents.
d) Are you in receipt of any financial support for your dependents from either;
i) The State? Yes/No
ii) A partner? Yes/No

4. Disability / Special needs
Do you have a disability as a result of which you incur additional expenditure? Yes/No
What is the nature of your disability?

Please outline the resultant expenditure below.

5. Any other information (additional details pertinent to your application)

6. Declaration
I declare that to the best of my knowledge the information provided is true and complete.
Signed... Date..

---

[54]The version given to participants was on two pages.

Appendix 2b: Case Study information, presented by screen-sharing software

Two independent sets of ten cases were devised. All information was presented in list form under numbered headings in Group Technologies' "Aspects" screen-sharing software. Five cases are reproduced below with added headings, for illustrative purposes.

Case No:1

1. Personal Details.

Surname: Jackson First names: Karl Title: Mr

Nationality: English Age: 19

---

2. Income.	Self £	Partner £
Salary / Earnings	-	-
Savings	-	-
Grants / Sponsorship	£3000	-
Other Income	-	-

---

3. Dependants.

a) No b) No

c) /

d i) / d ii) /

---

4. Disability/special needs.

a) Yes

b) Missing fingers

c) 'I cannot write, so must buy a computer/typewriter to do coursework.'

---

5. Any other information.

'Since an agricultural accident I have been left with no fingers on my writing hand. If I were able to save the cost of fees, I would be able to channel some of these funds into buying the equipment to help me write college work.'

Case No:3

1. Personal Details.

Surname: Turner | First names: John | Title: Mr

Nationality: English | Age: 18

---

2. Income.	Self £	Partner £
Salary / Earnings	-	-
Savings	-	-
Grants / Sponsorship	£3000	-
Other Income	-	-

---

3. Dependants.

a) No **b)** No

c) /

d i) / **d ii)** /

---

4. Disability/special needs.

a) Yes

b) Unable to walk

c) 'difficulty in transport, extra effort, expense doing things others take for granted.'

---

5. Any other information.

'Due to being wheelchair bound and unable to drive, I am hoping to conduct my studies via computer as a tool for remote learning. The cost of computer equipment is considerable, also if I were saved the cost of fees, the money could buy the equipment and cover increased phone bill costs for my parents.'

Case No:4

1. Personal Details.

Surname: Adrianopolus First names: Diana Title: Miss

Nationality: Greek Age: 21

---

2. Income.

	Self £	Partner £
Salary / Earnings	-	-
Savings	-	-
Grants / Sponsorship	£6000	-
Other Income	-	-

---

3. Dependants.

a) No b) No

c) /

d i) / d ii) /

---

4. Disability/special needs.

a) No

b) /

c) /

---

5. Any other information.

'The fees for study in this country are much higher than at home, but I wish to study here, though it would be hard if I cannot afford to travel back to my family. Help with the fees would be a great relief.'

Case No:5

1. Personal Details.

Surname: Gray   First names: Marion   Title: Mrs

Nationality: English   Age: 23

---

2. Income.

	Self £	Partner £
Salary / Earnings	-	-
Savings	£1000	-
Grants / Sponsorship	£4000	-
Other Income	-	-

---

3. Dependants.

a) Yes   **b)** No

c) One child of two years, needs day care if I study.

d i) Yes   **d ii)** No

---

4. Disability/special needs.

a) No

b) /

c) /

---

5. Any other information.

'Since my boyfriend left I have been solely responsible for my child, yet I wish to study and further myself. This is difficult on my own, I live away from family so could only hire a child minder while I studied, saving even some of the cost of fees would be of great help.'

Case No:7

1. Personal Details.

Surname: Korby First names: Brenda Title: Mrs

Nationality: English Age: 27

---

2. Income.

	Self £	Partner £
Salary / Earnings	-	£12500
Savings	-	-
Grants / Sponsorship	£4000	-
Other Income	-	-

---

3. Dependants.

a) Yes **b)** No

c) 'Costs of daycare, special equipment (stair lift etc.) for disabled son.'

d i) Yes **d ii)** Yes

---

4. Disability/special needs.

a) Yes (not personally, but expense because of disabled son).

b) /

c) /

---

5. Any other information.

' I would dearly like to study but the extra burden of fees would not allow us to hire the care our son would need whilst I attended the Institute, a grant towards even part of the cost of fees would help in this regard.'

Case No:9

1. Personal Details.

Surname: Bellamy First names: Sarah Title: Mrs

Nationality: English Age: 28

---

2. Income.

	Self £	Partner £
Salary / Earnings	-	-
Savings	£500	-
Grants / Sponsorship	£3600	-
Other Income	-	-

---

3. Dependants.

a) Yes **b)** No

c) Disabled mother, costs of nursing, needs constant care / help with eating, washing etc.

d i) Yes **d ii)** No

---

4. Disability/special needs.

a) No

b) /

c) /

---

5. Any other information.

'A contribution towards the costs of tuition fees would enable me to hire some form of home help whilst I attended college.'

Appendix 2c: participants' general task instructions

There were two sets of instructions for each participant, one common to both dyad members, reproduced below, and one specific to each, reproduced in Appendix 2d.

Application for Special Support Fund.
Notes for Assessors

To try to cover the full range of students' financial problems would stretch the funds to the point where nobody would receive any significant help. For that reason it has been decided to try to ensure that the money is directed to those in greatest need as a result of pressing special difficulties.

The budget for the present year is set at **£5000** for the purpose of meeting the fees of potential candidates. In line with the above statement it is recommended that applications be considered with a view to meeting the cost of fees in full (currently £2000) or 50% thereof (£1000).

It is understood that the various colleges of the institute have stated priorities for broadening access to the Institute, where at all possible representative assessors of the colleges are urged to reach mutual agreement on the allocation of special support fund resources.

Task Summary:-

1. In considering the application forms be aware of the priorities of your particular college (ie. are you concerned with overseas, disabled or dependent issues.)

2. Applications to the Special Support Fund may be met with one of three responses, 100% funding (where £2000 is awarded), 50% funding (where £1000 is awarded) and no funding. The limits of the budget must not be exceeded, however assessors are free to divide up the budget as necessary. Thus 5 applications may receive 50% funding or 2 applications 100% funding with one at the 50% level, and so on.

3. Assessors have up to half an hour to consider the application forms and allocate the fund with the above constraints in mind. Where possible negotiate with other assessors to reach mutual agreement on fund allocation.

Appendix 2d: participants' specific task instructions

Each dyad member received additional instructions besides those presented in Appendix 2c. These described their specific role as a college representative on a sub-committee for aleviation of student hardship. Their instructions were designed to stimulate debate over the fictitous case information (see Appendix 2a), in that their priorities could not both be accommodated without some concession.

Version 1 - Disabled student priority

Rydale College

RE :- Application for Special Support Fund.

The college regards the broadening of access to disabled candidates as a priority. As the representative of Rydale College in the current round of Special Support Fund allocation, we would particularly like to emphasize that provision be made for candidates who, owing to the additional costs incurred as a result of disability or special need, would otherwise find the expense of academic fees prohibitive.

Version 2 - Family dependents priority

GoodFellow College

RE :- Application for Special Support Fund.

The college regards the broadening of access to candidates with families as a priority. As the representative of Goodfellow College in the current round of Special Support Fund allocation, we would particularly like to emphasize that provision be made for candidates who, owing to family dependents, the costs of child care and so on, would otherwise find the expense of academic fees prohibitive.

Version 3 - Overseas students priority

Darwin College

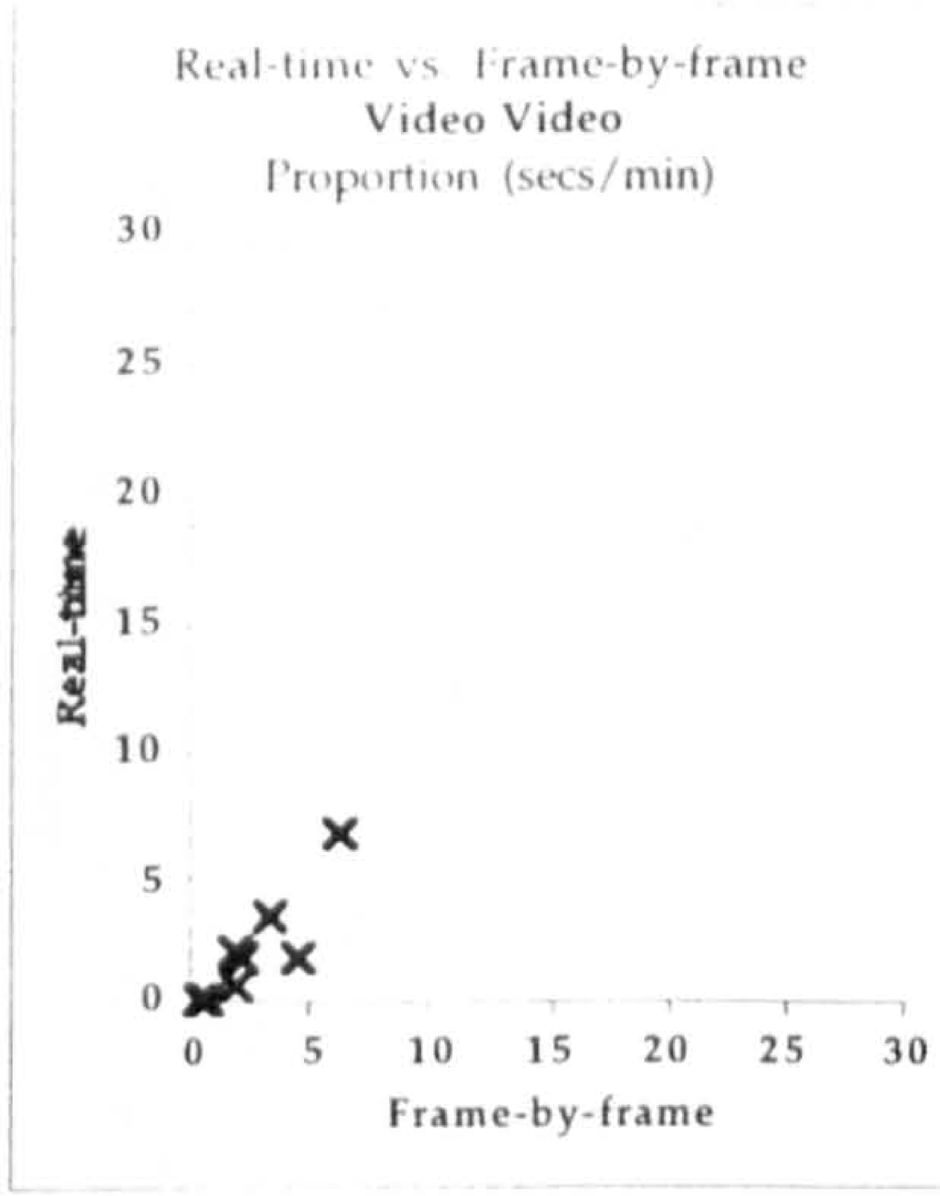
RE :- Application for Special Support Fund.

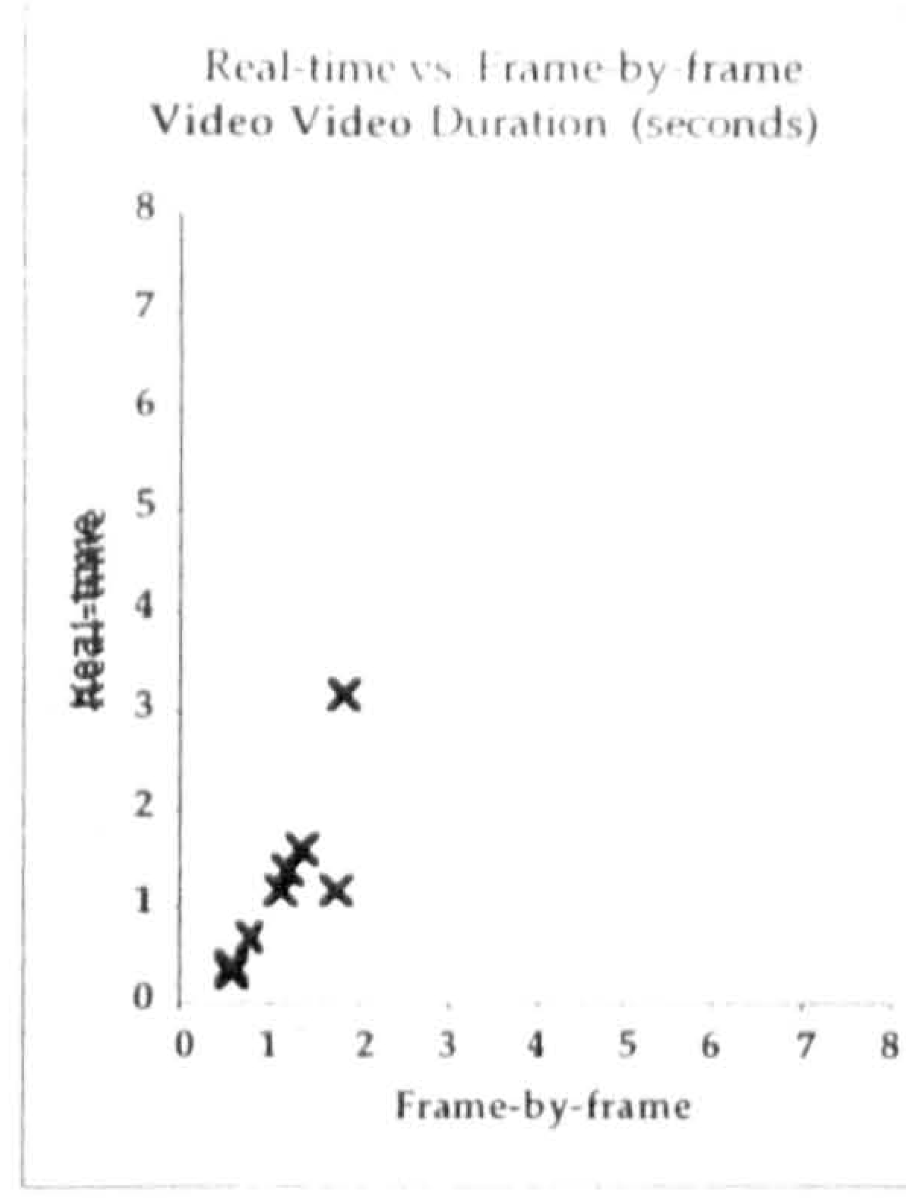
The college regards the broadening of access to overseas candidates as a priority. As the representative of Darwin College in the current round of Special Support Fund allocation, we would particularly like to emphasize that provision be made for candidates who, owing to the additional investment required for study in a foreign country, would otherwise find the expense of academic fees prohibitive.

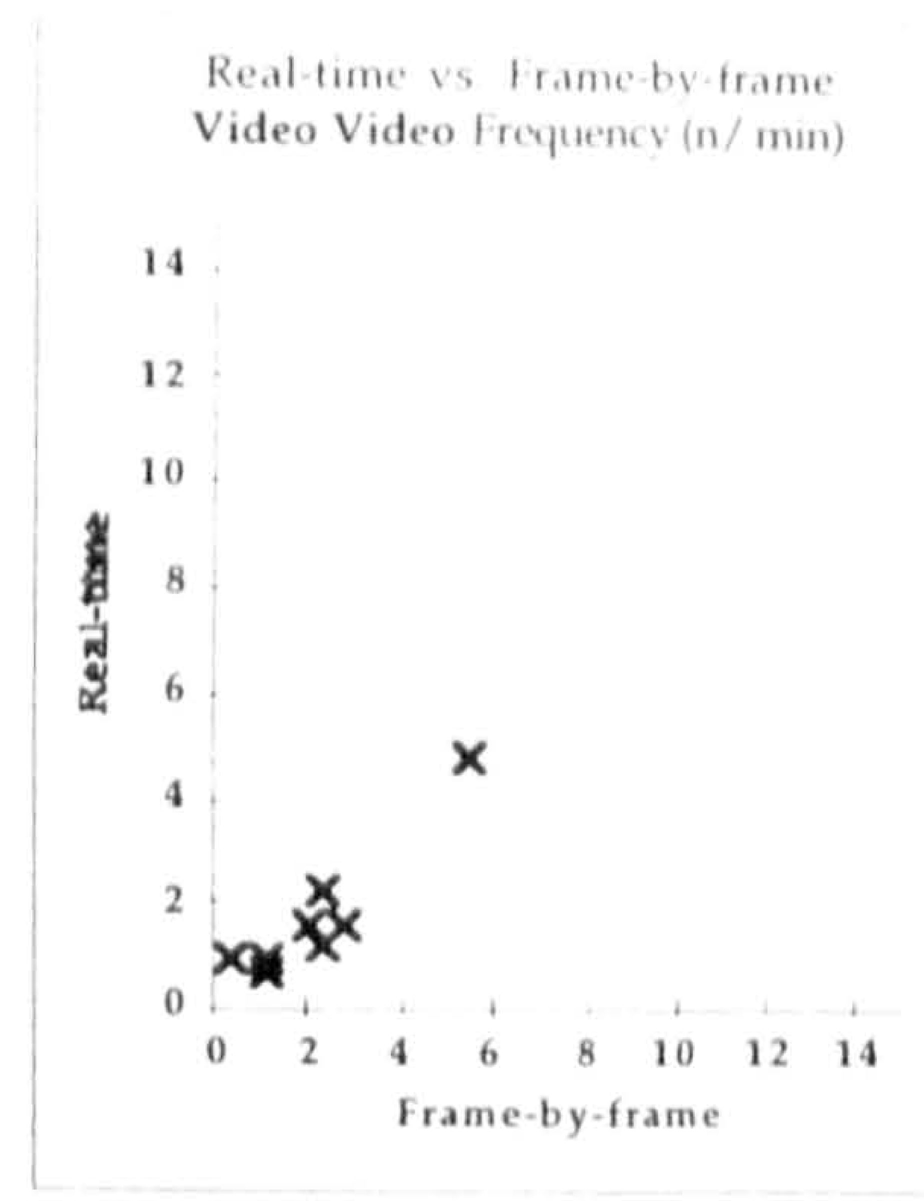
Appendix 3: inter-rater reliability for Co-activity data

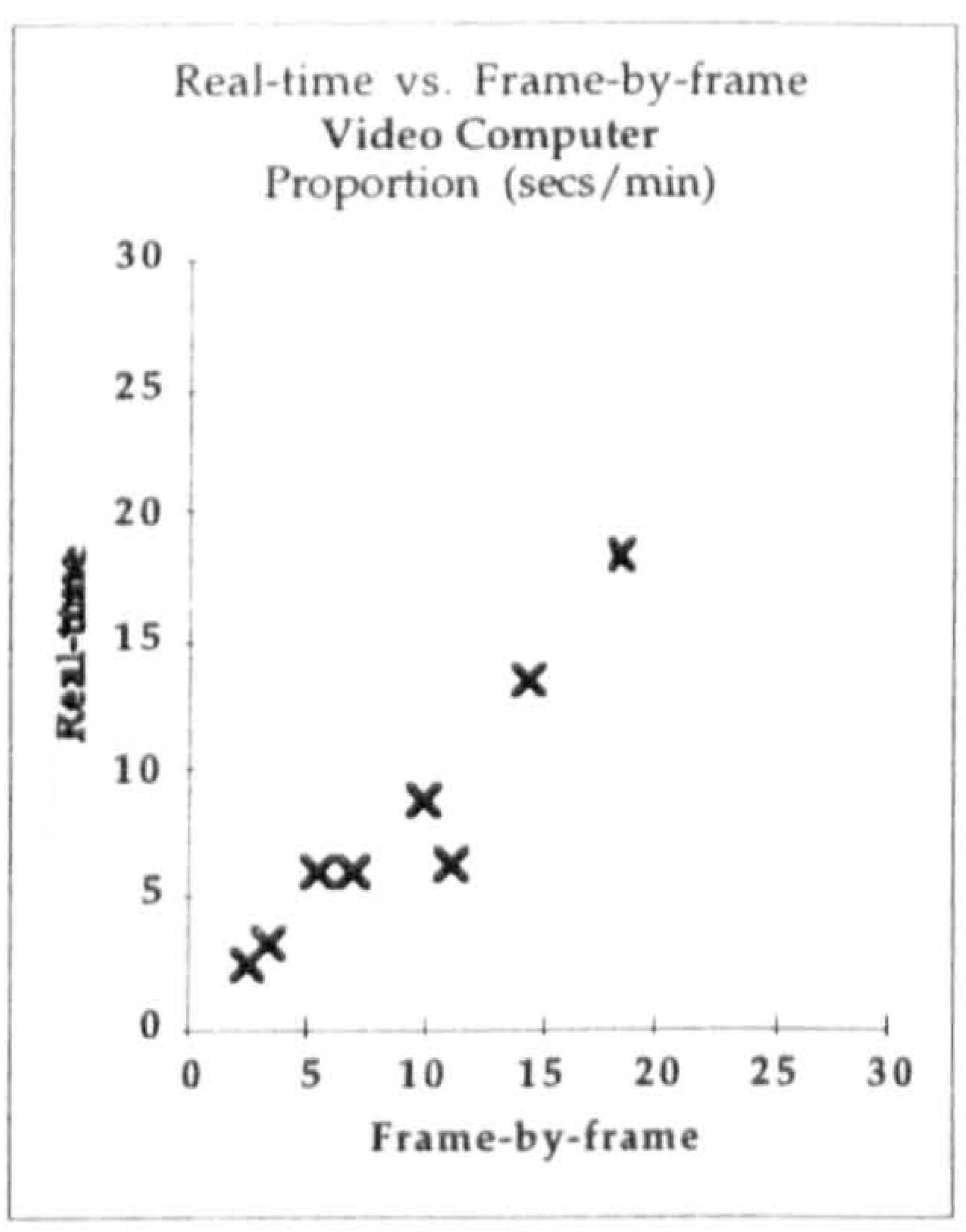
GazeVideo: GazeVideo Co-activity	Duration (Seconds)		Frequency (n / min.)		Time Proportion (s / min.)	
Dyad code	Real-time	Frame-by-frame	Real-time	Frame-by-frame	Real-time	Frame-by-frame
.2/4	0.67	0.74	1.33	2.33	0.90	1.73
.3/1	1.16	1.69	1.67	2.67	1.93	4.50
.5/7	1.59	1.33	2.33	2.33	3.70	3.10
.6/8	0.34	0.50	1.00	1.00	0.34	0.50
.9/11	3.20	1.73	0.67	1.00	2.13	1.73
.10/12	1.42	1.17	5.00	5.33	7.12	6.23
.15/13	0.42	0.50	1.00	0.33	0.42	0.17
.16/14	1.16	1.08	1.67	2.00	1.94	2.17
Mean	1.25	1.09	1.83	2.12	2.31	2.52
Standard Dev.	0.91	0.49	1.38	1.53	2.23	2.04
Diff	0.15		-0.29		-0.21	
Diff %	13.05		-14.66		-8.55	
m	1.52		0.84		0.96	
c	0.45		0.13		0.22	
r2	0.66		0.87		0.76	
r	0.81		0.93		0.87	

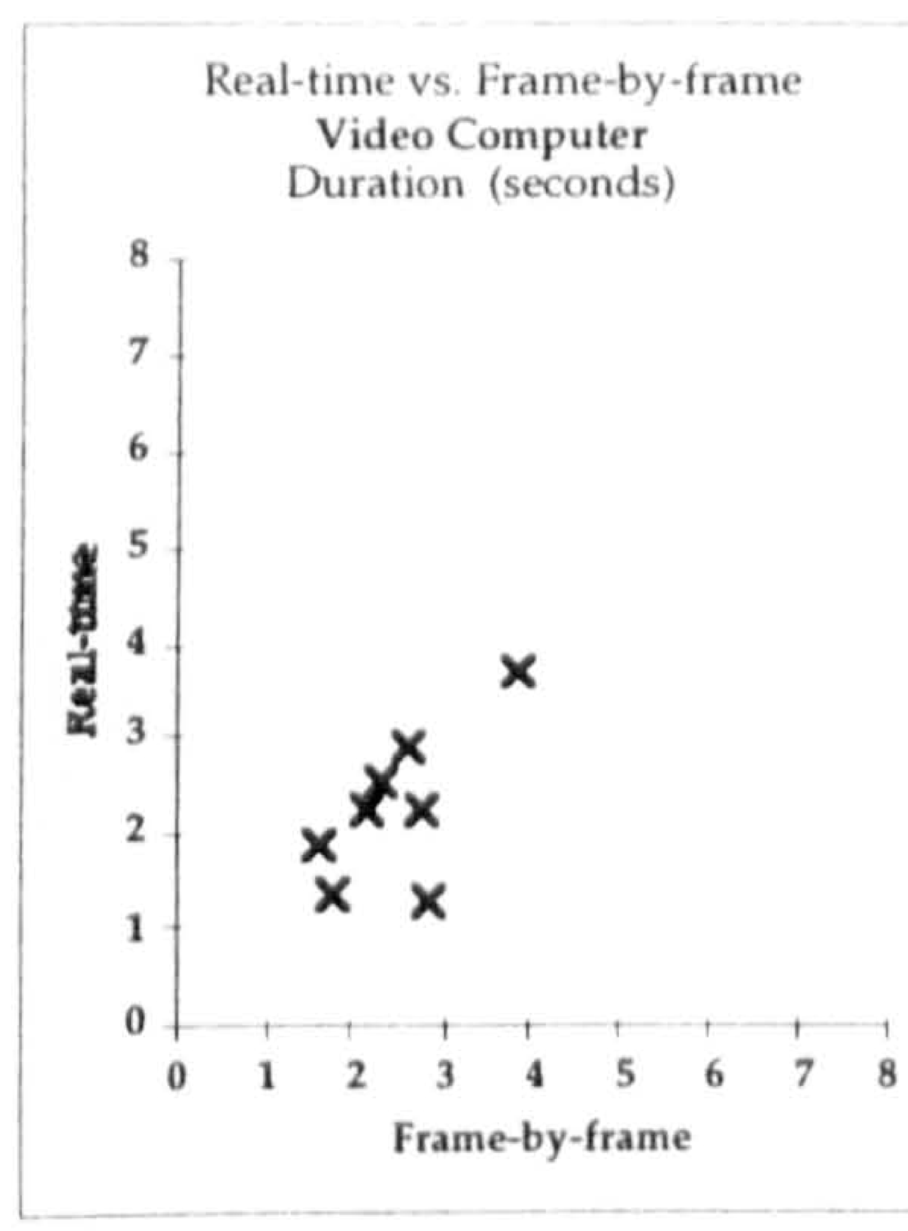
Simple Activity Set validity showed the lowest correlation for GazeVideo duration data. It might then be assumed that GazeVideo:GazeVideo Co-activity durations would prove the sternest test for statistical validity. The data presented above show that in fact the two recording methods agree very well. The following two pages present scattergrams that compare the full range of Co-activity data types, derived from each rater's Action Recorder observations.

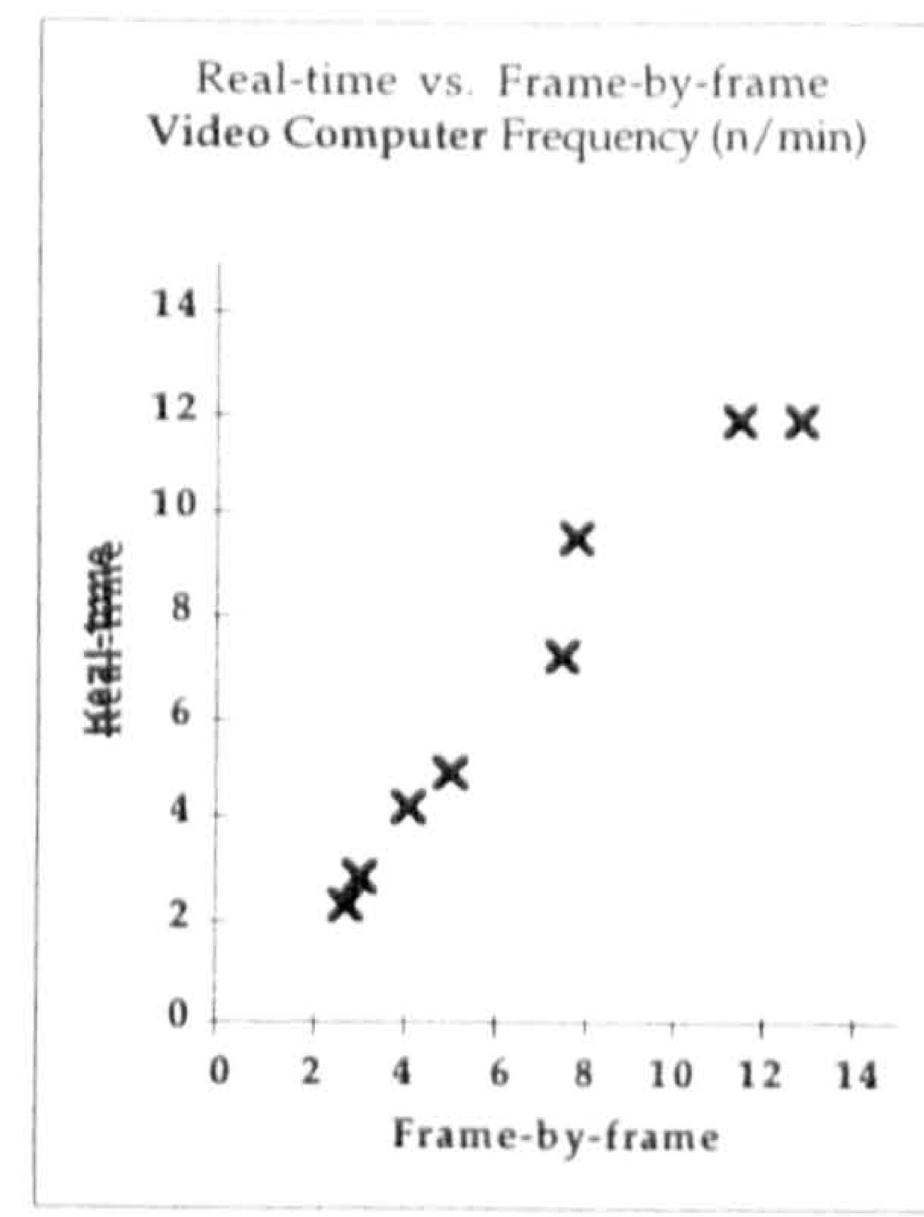

Real-time vs. Frame-by-frame
Video Video
Proportion (secs/min)
Real-time
Frame-by-frame

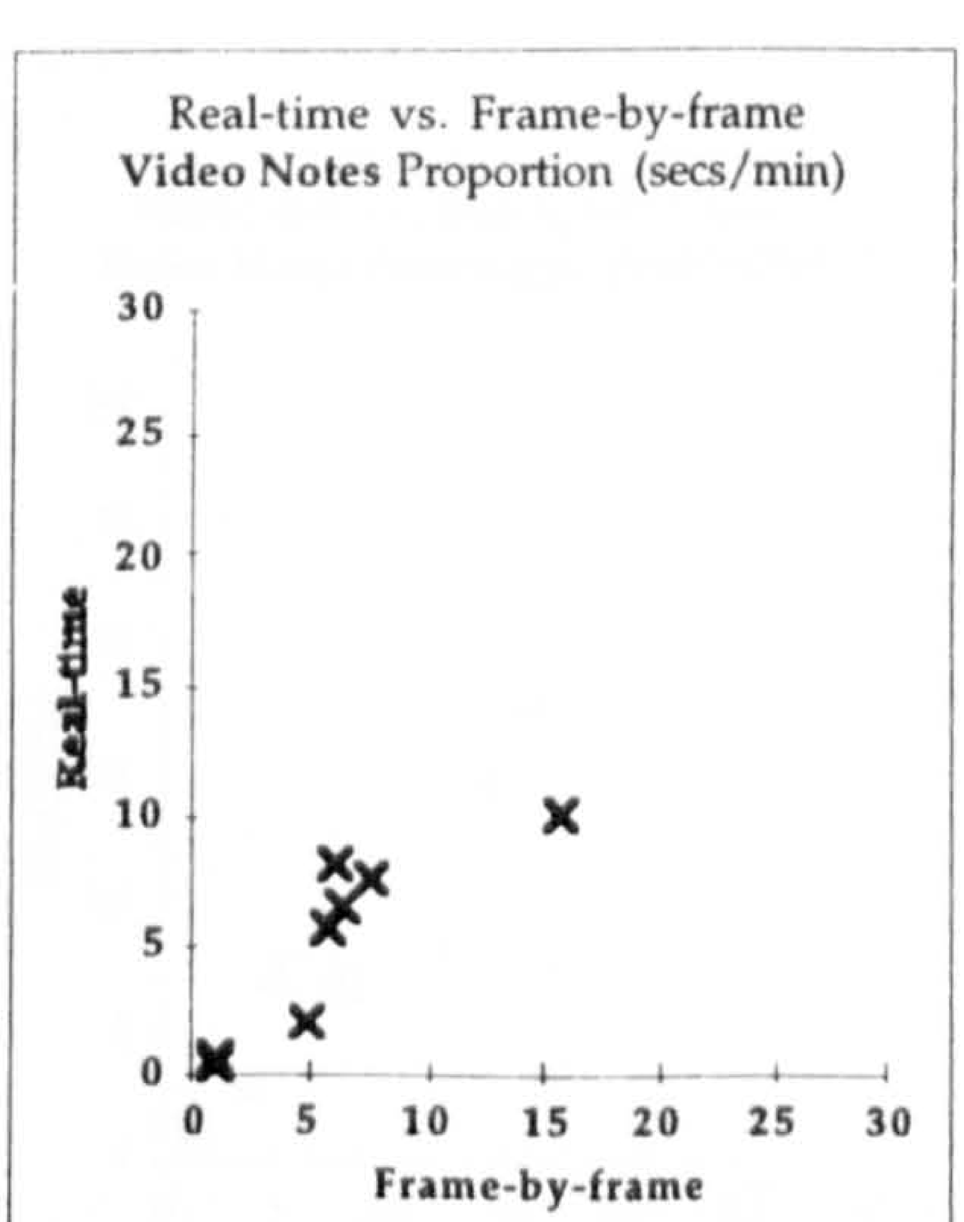

Real-time vs. Frame-by-frame
Video Video Duration (seconds)
Real-time
Frame-by-frame

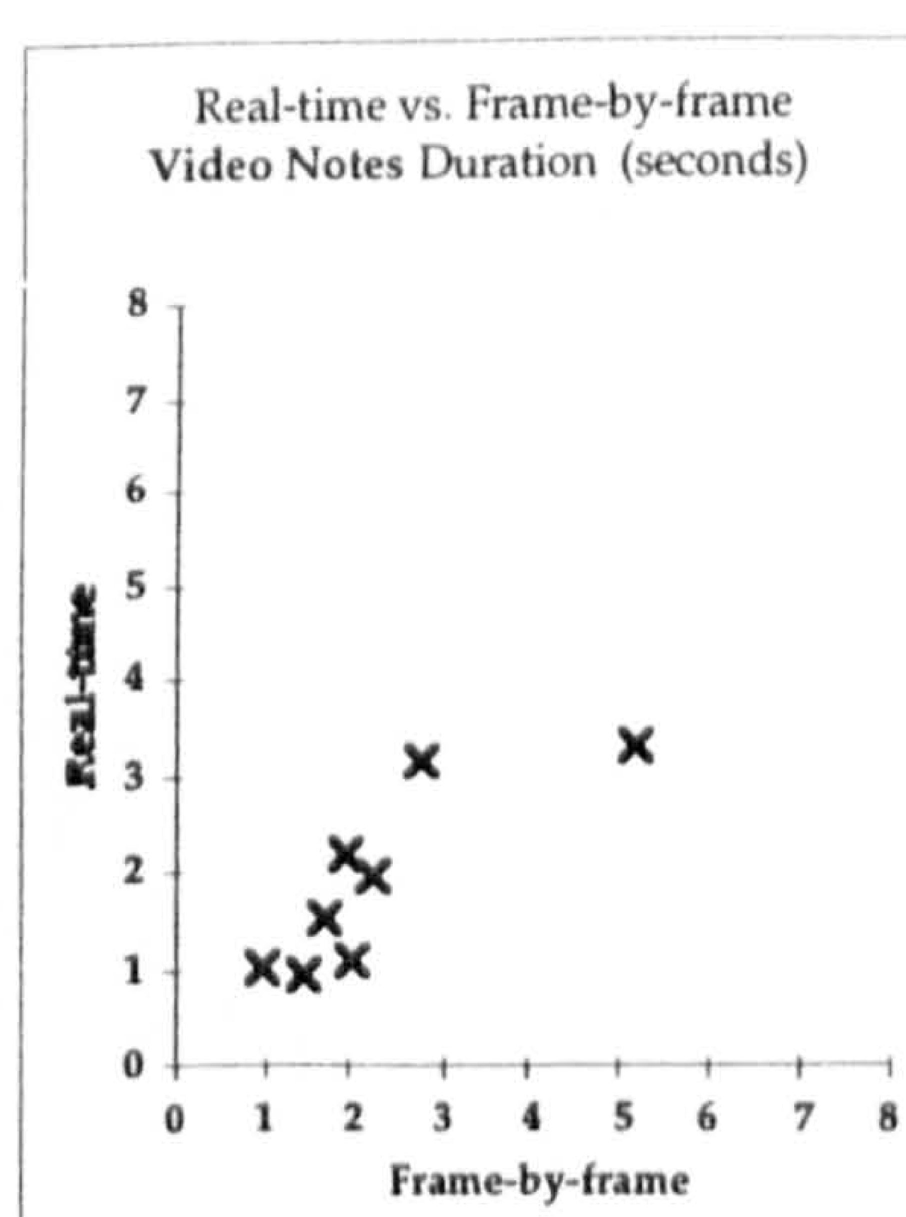

Real-time vs. Frame-by-frame
Video Video Frequency (n/min)
Real-time
Frame-by-frame

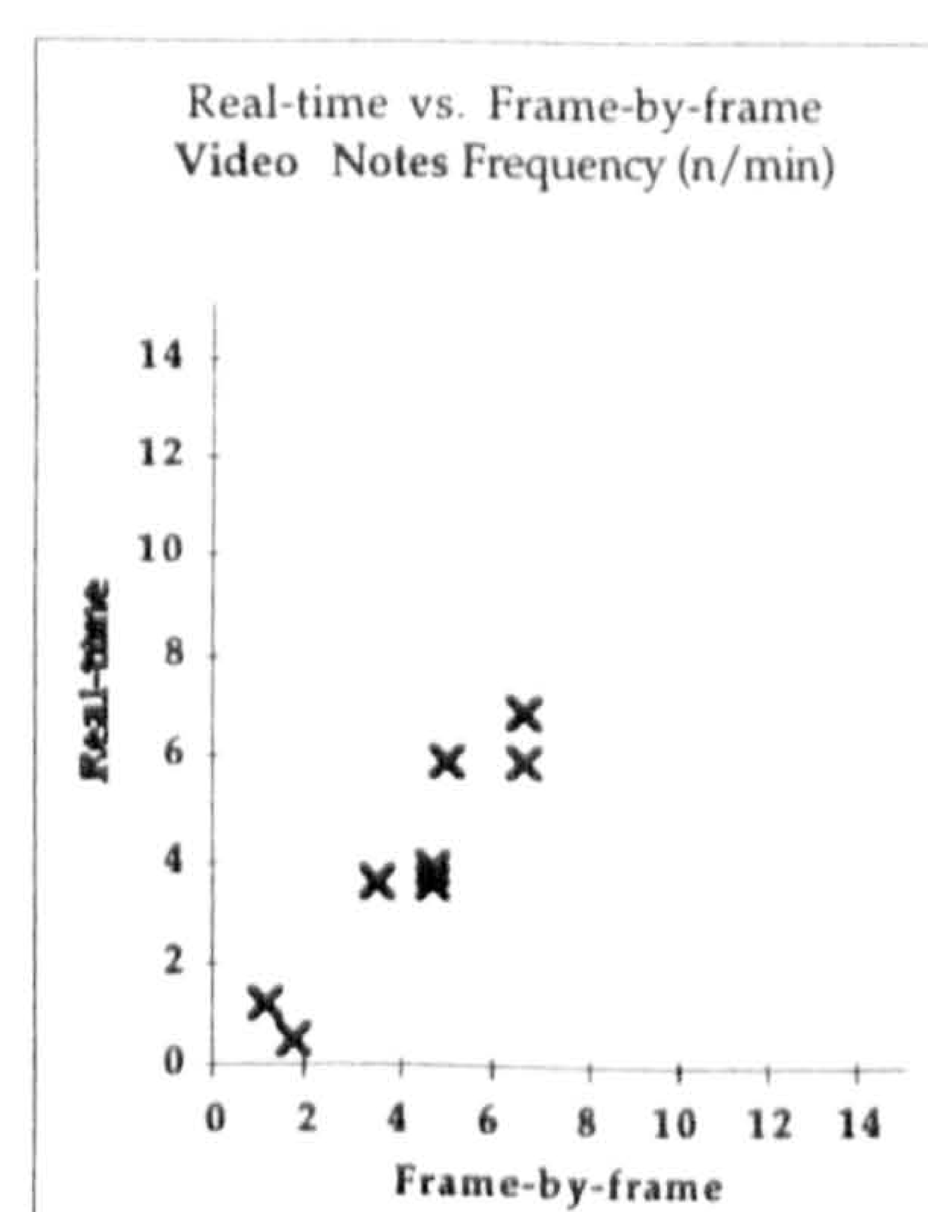

Real-time vs. Frame-by-frame
Video Computer
Proportion (secs/min)
Real-time
Frame-by-frame

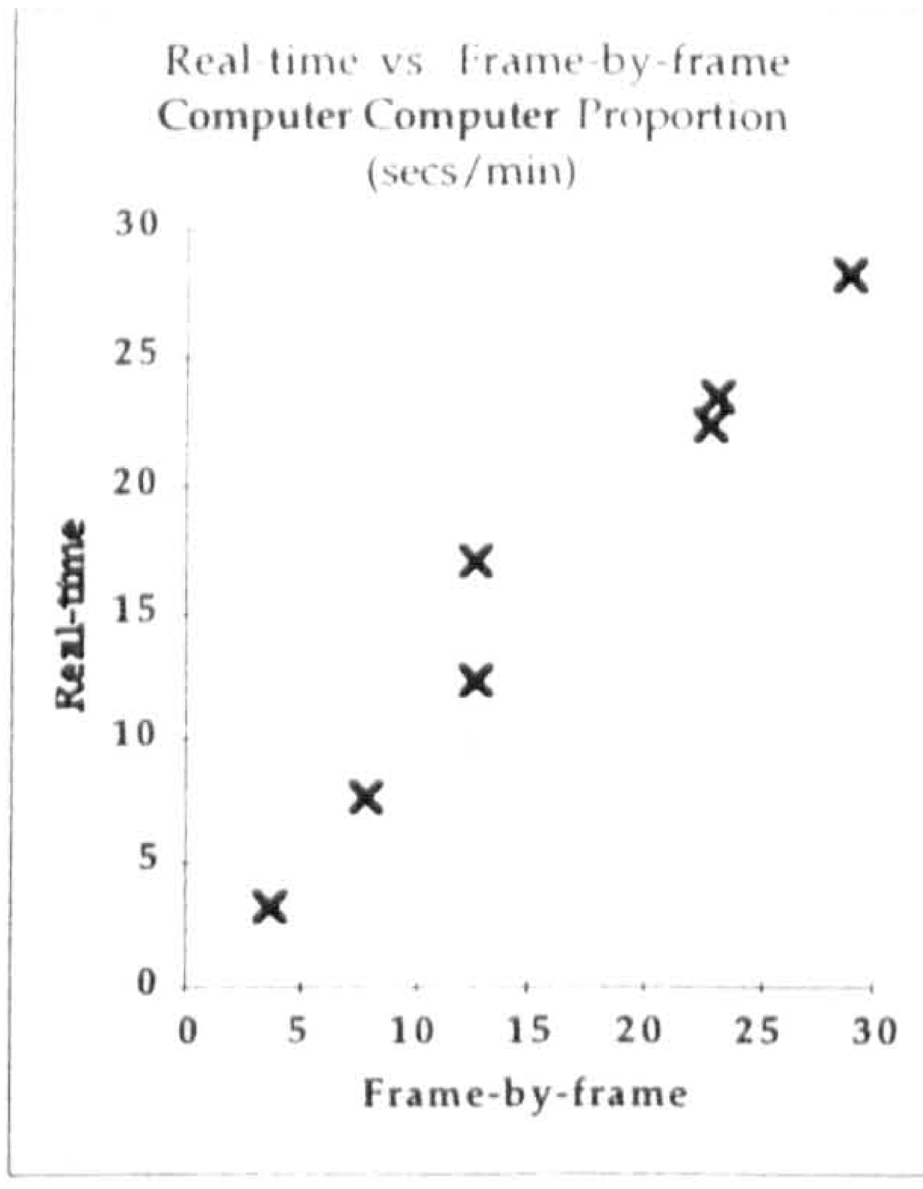

Real-time vs. Frame-by-frame
Video Computer
Duration (seconds)
Real-time
Frame-by-frame

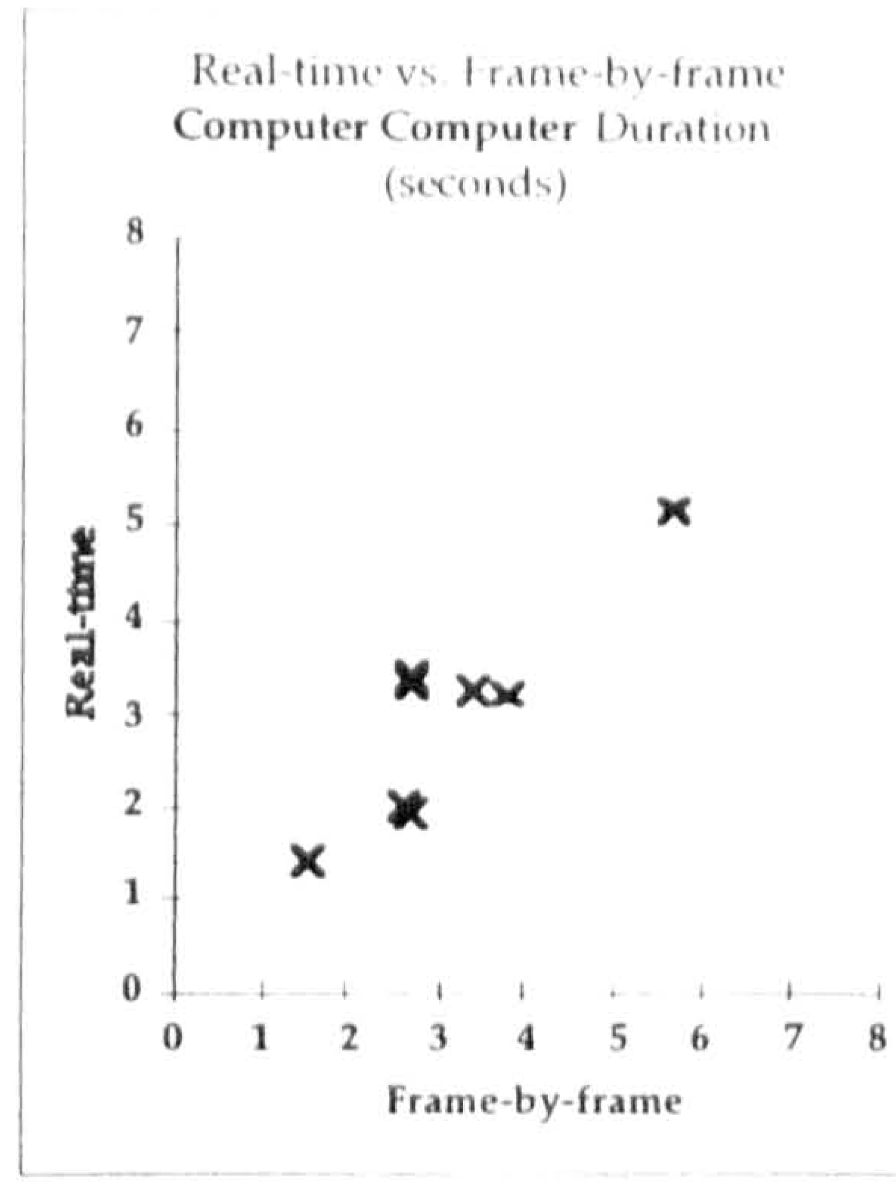

Real-time vs. Frame-by-frame
Video Computer Frequency (n/min)
Real-time
Frame-by-frame

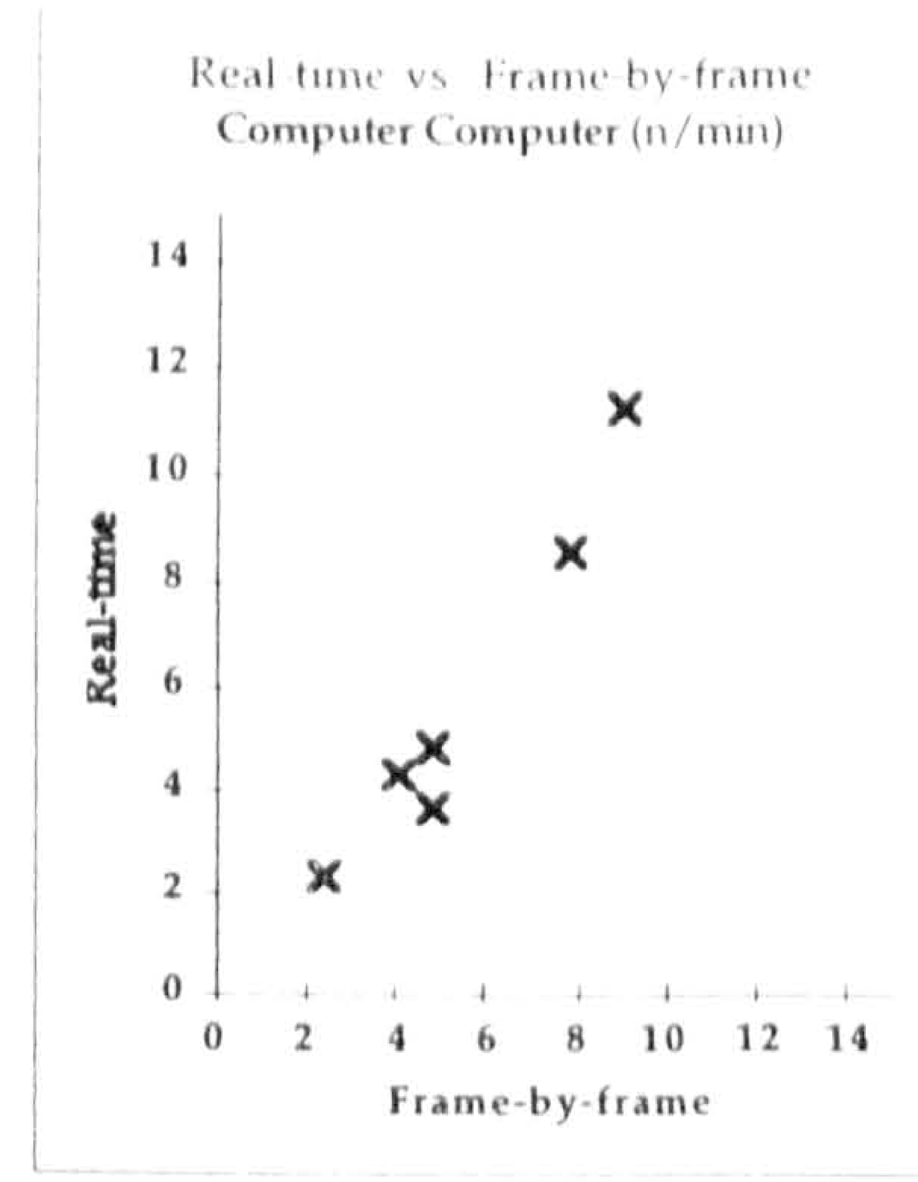

Real-time vs. Frame-by-frame
Video Notes Proportion (secs/min)
Real-time
Frame-by-frame

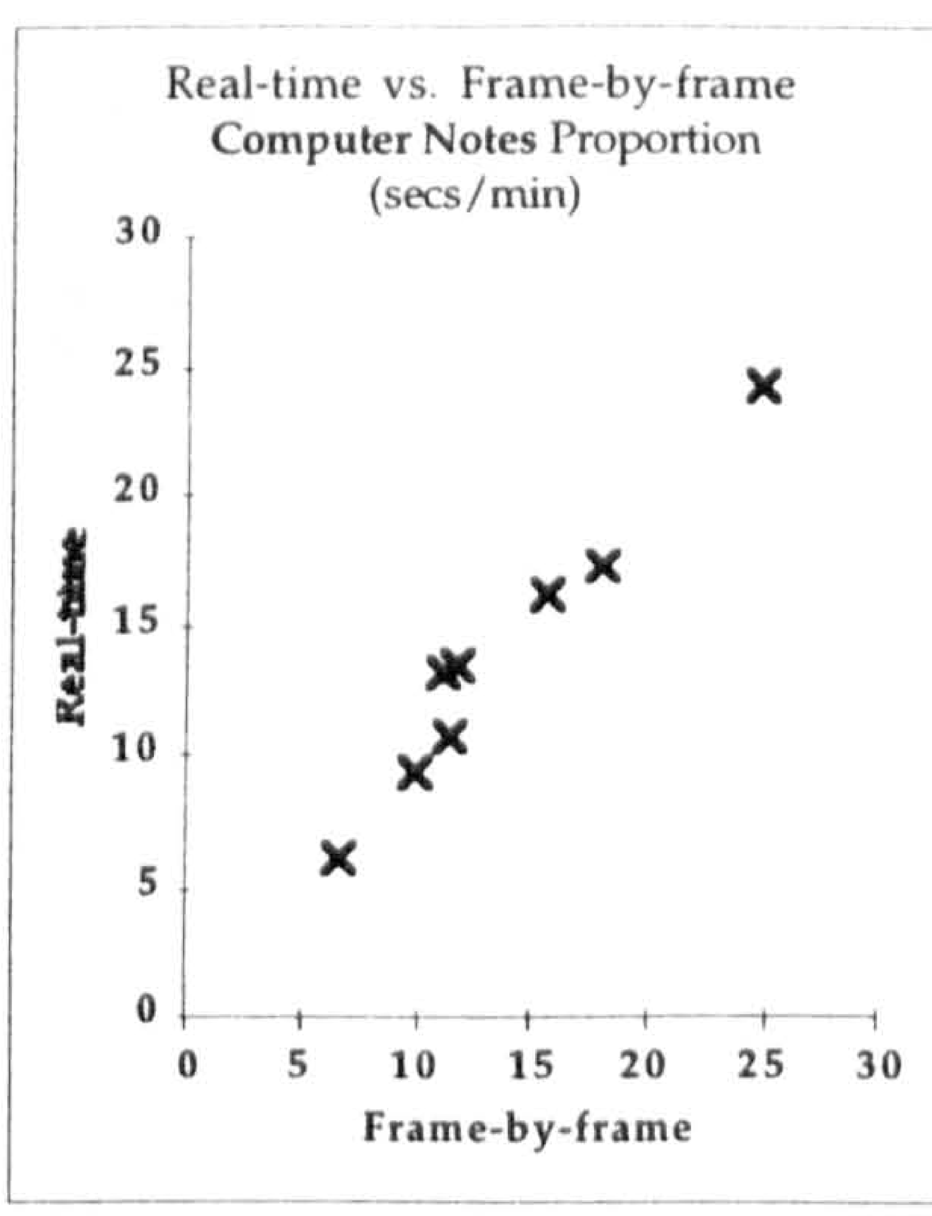

Real-time vs. Frame-by-frame
Video Notes Duration (seconds)
Real-time
Frame-by-frame

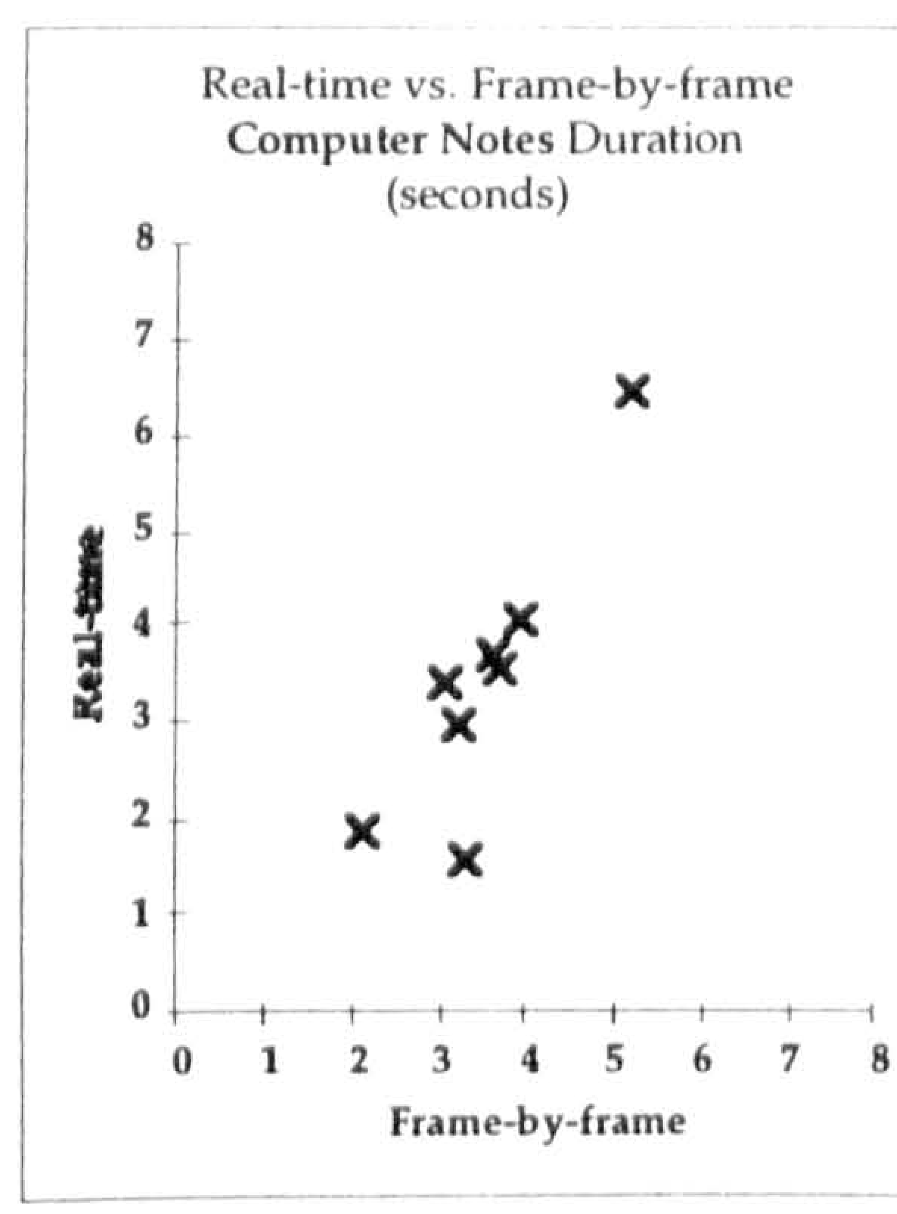

Real-time vs. Frame-by-frame
Video Notes Frequency (n/min)
Real-time
Frame-by-frame

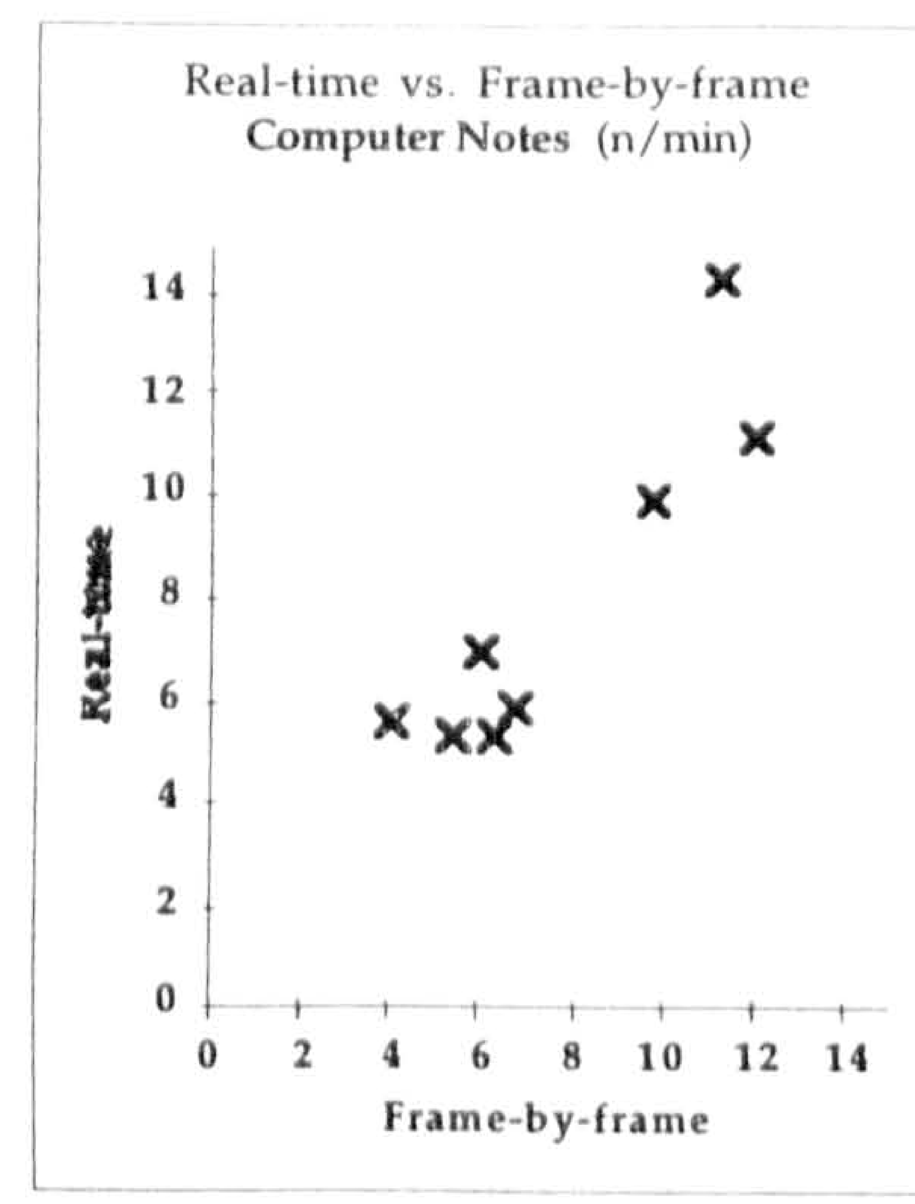

Real-time vs. Frame-by-frame
Computer Computer Proportion
(secs/min)
Real-time
Frame-by-frame

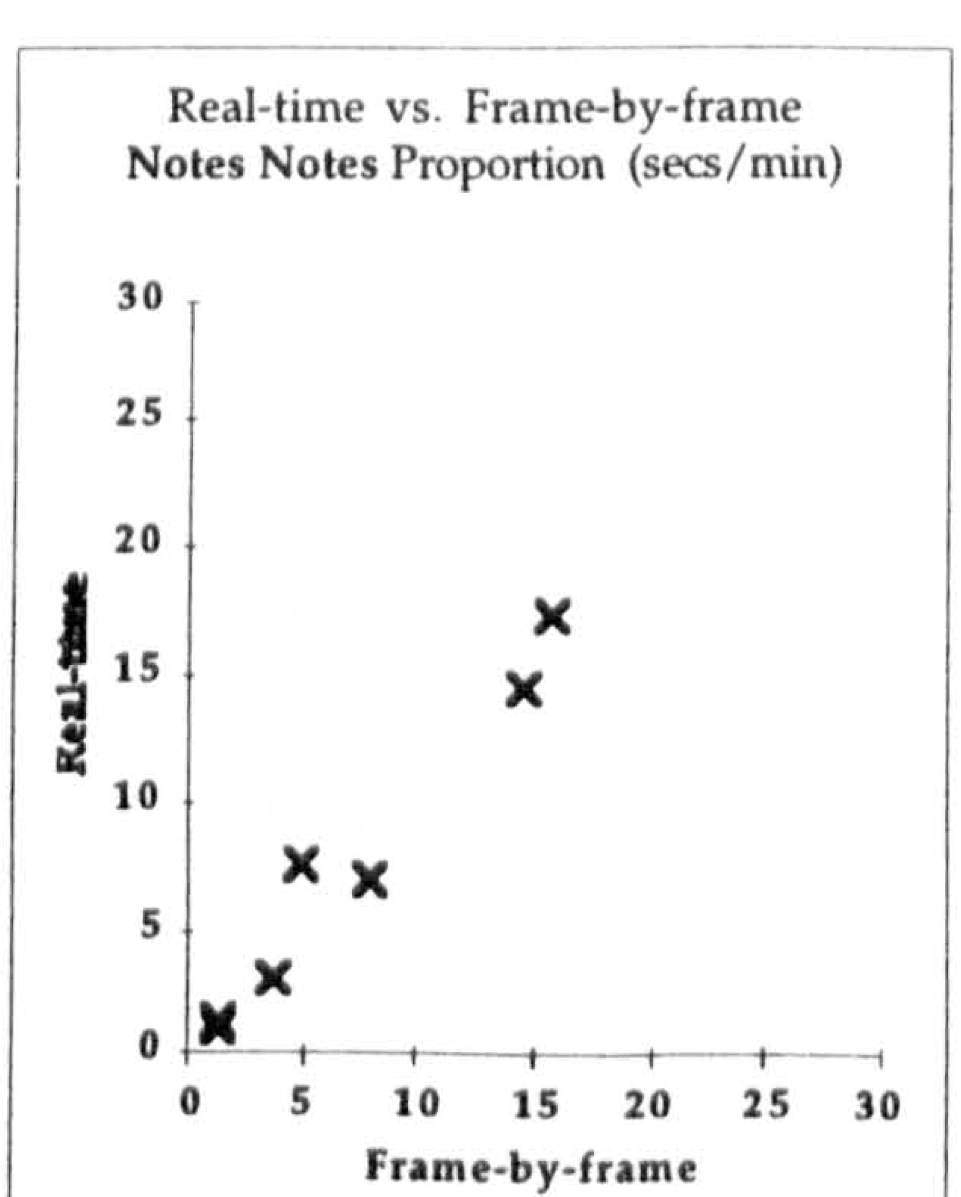

Real-time vs. Frame-by-frame
Computer Computer Duration
(seconds)
Real-time
Frame-by-frame

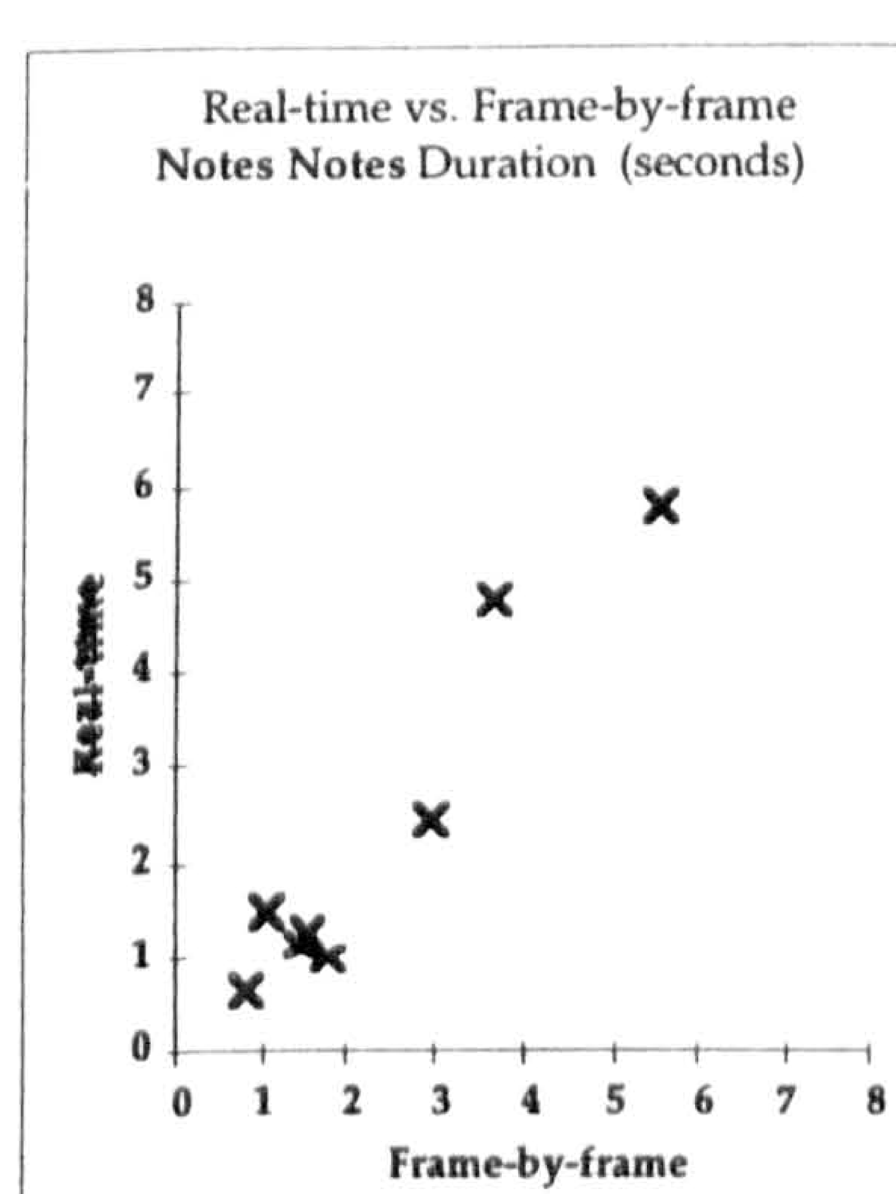

Real-time vs. Frame-by-frame
Computer Computer (n/min)
Real-time
Frame-by-frame

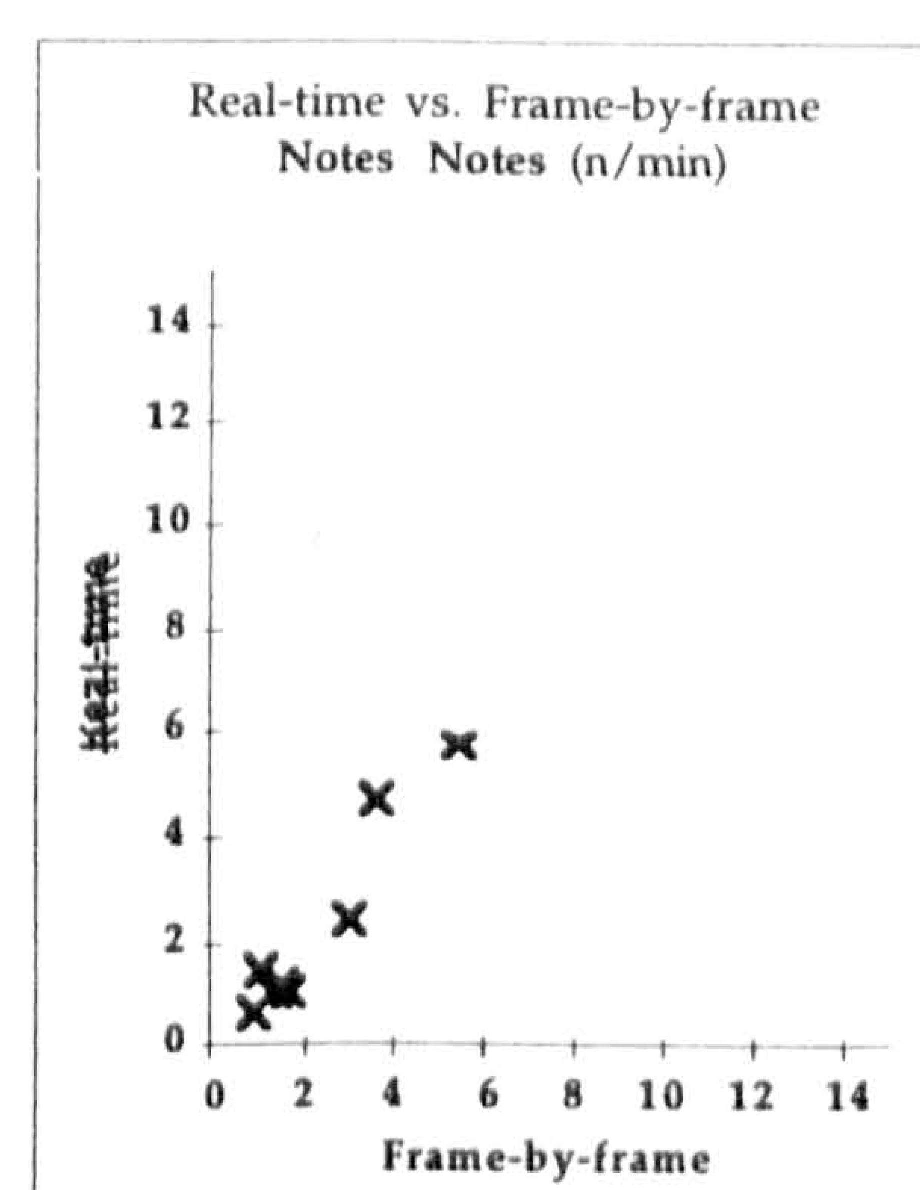

Real-time vs. Frame-by-frame
Computer Notes Proportion
(secs/min)
Real-time
Frame-by-frame


Real-time vs. Frame-by-frame
Computer Notes Duration
(seconds)
Real-time
Frame-by-frame


Real-time vs. Frame-by-frame
Computer Notes (n/min)
Real-time
Frame-by-frame


Real-time vs. Frame-by-frame
Notes Notes Proportion (secs/min)
Real-time
Frame-by-frame


Real-time vs. Frame-by-frame
Notes Notes Duration (seconds)
Real-time
Frame-by-frame


Real-time vs. Frame-by-frame
Notes Notes (n/min)
Real-time
Frame-by-frame

Appendix 4: Synchronization and the 'S' Statistic

Let X stand for the distribution of Activity A, given a known proportion of time spent in Activity A, a, and Y stand for the distribution of Activity B, given a known proportion of time spent in Activity B, b. In the interest of clarity, let Z stand for instances of co-incidence between Activities A and B, so that ΣZ corresponds to p_{ce} or the *proportion* of time expected to be spent in the Co-activity state, AB. Refer to Figure 5.5 for a representation of the four main cases included below (I to IV).

First take expectations over X for fixed Y = y

Case I $a < y < 1\text{-}b\text{-}a$

$$\sum(Z \mid Y=y) = \int_{y-a}^{y}(x+a-y)\tfrac{dx}{1-a} + \int_{y}^{y+b-a} a\tfrac{dx}{1-a} + \int_{y+b-a}^{y+b}(y+b-x)\tfrac{dx}{1-a}$$

$$= \left[\frac{\frac{1}{2}(x+a-y)^2}{1-a}\right]_{y-a}^{a} + a(b-a)\tfrac{1}{1-a} + \left[-\frac{1}{2}\frac{(y+b-x)^2}{1-a}\right]_{y+b-a}^{y+b}$$

$$= \frac{1}{1-a}\left\{\frac{1}{2}a^2 + a(b-a) + \frac{1}{2}a^2\right\}$$

$$= \frac{ab}{1-a}$$

Case II $0 < y < a < 1\text{-}b\text{-}a$

$$\sum(Z \mid Y=y) = \int_{0}^{y}(x+a-y)\tfrac{dx}{1-a} + \int_{y}^{y+b-a} a\tfrac{dx}{1-a} + \int_{y+b-a}^{y+b}(y+b-x)\tfrac{dx}{1-a}$$

$$= \frac{ab-\frac{1}{2}(a-y)^2}{1-a}$$

Case III $0 < a < 1\text{-}b\text{-}a < y < 1\text{-}b$

$$\sum(Z \mid Y=y) = \int_{y-a}^{y}(x+a-y)\tfrac{dx}{1-a} + \int_{y}^{y+b-a} a\tfrac{dx}{1-a} + \int_{y+b-a}^{1-a}(y+b-x)\tfrac{dx}{1-a}$$

$$= \frac{ab-\frac{1}{2}(1-b-a-y)^2}{1-a}$$

Case IV $0 < 1\text{-}b\text{-}a < y < a < 1\text{-}b$

$$\sum(Z \mid Y=y) = \int_{0}^{y}(x+a-y)\tfrac{dx}{1-a} + \int_{y}^{y+b-a} a\tfrac{dx}{1-a} + \int_{y+b-a}^{1-a}(y+b-x)\tfrac{dx}{1-a}$$

$$= \frac{ab-\frac{1}{2}(a-y)^2-\frac{1}{2}(1-b-a-y)^2}{1-a}$$

Case V $0 < y < 1\text{-}b\text{-}a < a < 1\text{-}b$

same formula as case II (but becomes case VI after 1-b-a)

Case VI $0 < 1\text{-}b\text{-}a < a < y < 1\text{-}b$

same formula as case III

Case VII $1\text{-}b\text{-}a < 0 < y < 1\text{-}b < a$

same formula as case IV although 1-b-a-y switches sign (which makes no difference as it occurs squared)

Now take expectations over Y

Case A $2a+b<1$

Use cases I, II and III above, getting:

$$\sum Z = \frac{ab}{1-a} - \int_0^a \frac{\frac{1}{2}(a-y)^2}{(1-a)} \frac{dy}{1-b} - \int_{1-b-a}^{1-b} \frac{\frac{1}{2}(1-b-a-y)^2}{(1-a)} \frac{dy}{1-b}$$

$$= \frac{ab}{1-a} - \frac{\frac{1}{3}a^3}{(1-a)(1-b)}$$

Case B $a+b<1$ but $2a+b>1$

Use cases IV, V and VI above, getting the same answer.

Case C $a+b>1$

Use case VII above, getting:

$$\sum Z = \frac{ab}{1-a} - \frac{\frac{1}{3}a^3}{(1-a)(1-b)} + \frac{\frac{1}{3}(a+b-1)^3}{(1-a)(1-b)}$$

Now find ΣZ^2, beginning with expectations over X for Y=y

Case I $\quad a < y < 1\text{-}b\text{-}a$

$$\sum(Z^2 \mid Y=y) = \int_{y-a}^{y}(x+a-y)^2 \tfrac{dx}{1-a} + \int_{y}^{y+b-a} a^2 \tfrac{dx}{1-a} - \int_{y+b-a}^{y+b}(y+b-x)^2 \tfrac{dx}{1-a}$$

$$= \left[\frac{\frac{1}{3}(x+a-y)^3}{1-a}\right]_{y-a}^{a} + a^2(b-a)\frac{1}{1-a} + \left[-\frac{1}{3}\frac{(y+b-x)^3}{1-a}\right]_{y+b-a}^{y+b}$$

$$= \frac{1}{1-a}\left\{\frac{1}{3}a^3 + a^2(b-a) + \frac{1}{3}a^3\right\}$$

$$= \frac{a^2b - \frac{1}{3}a^3}{1-a}$$

Case II $\quad 0 < y < a < 1\text{-}b\text{-}a$

$$\sum(Z^2 \mid Y=y) = \frac{a^2b - \frac{1}{3}a^3 - \frac{1}{3}(a-y)^3}{1-a}$$

Case III $\quad 0 < a < 1\text{-}b\text{-}a < y < 1\text{-}b$

$$\sum(Z^2 \mid Y=y) = \frac{a^2b - \frac{1}{3}a^3 - \frac{1}{3}(1-b-a-y)^3}{1-a}$$

Case IV $\quad 0 < 1\text{-}b\text{-}a < y < a < 1\text{-}b$

$$\sum(Z^2 \mid Y=y) = \frac{a^2b - \frac{1}{3}a^3 - \frac{1}{3}(a-y)^3 - \frac{1}{3}(1-b-a-y)^3}{1-b}$$

Cases V, VI and VII do not add to the derivation, as above.

Now find EZ taking expectations over Y

Case A $\quad 2a+b < 1$

Use cases I, II and III

$$\sum Z^2 = \frac{a^2b - \frac{1}{3}a^3}{1-a} - \int_{0}^{a} \frac{\frac{1}{3}(a-y)^3}{1-a}\frac{dy}{1-b} - \int_{1-b-a}^{1-b} \frac{\frac{1}{3}(1-b-a-y)^3}{1-a}\frac{dy}{1-b}$$

$$= \frac{a^2b - \frac{1}{3}a^3}{1-a} - \frac{\frac{1}{6}a^4}{(1-a)(1-b)}$$

Case B $\quad a+b < 1$ **but** $2a+b > 1$

Use cases IV, V VI above, getting same answer

Case C $\qquad a+b>1$

Use case VII above, getting:

$$\sum Z^2 = \frac{a^2b-\frac{1}{3}a^3}{1-a} - \frac{\frac{1}{6}a^4}{(1-a)(1-b)} + \frac{\frac{1}{6}(a+b-1)^4}{(1-a)(1-b)}$$

Now find the variance, given

$$VZ = \sum Z^2 - (\sum Z)^2$$

There are just two cases:

(1) If $a+b<1$ then:

$$VZ = \frac{a^2b-\frac{1}{3}a^3}{1-a} - \frac{\frac{1}{6}a^4}{(1-a)(1-b)} - \left\{\frac{ab}{1-a} - \frac{\frac{1}{3}a^3}{(1-a)(1-b)}\right\}^2$$

(2) If $a+b>1$ then:

$$VZ = \frac{a^2b-\frac{1}{3}a^3}{1-a} - \frac{\frac{1}{6}a^4}{(1-a)(1-b)} + \frac{\frac{1}{6}(a+b-1)^4}{(1-a)(1-b)} - \left\{\frac{ab}{1-a} - \frac{\frac{1}{3}a^3}{(1-a)(1-b)} + \frac{\frac{1}{3}(a+b-1)^3}{(1-a)(1-b)}\right\}^2$$

We find the square root of the variance, $\sqrt{VZ}$, and call it the standard deviation. Then a standardized value, S, may be expressed for an observed conjunction, p_{ab} is given by:

$$S = \frac{p_{ab} - \sum Z}{\sqrt{VZ}}$$

References

Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty-Sneddon, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R. (1991) The HCRC map task corpus. *Language and speech*, 34(4), 351-366.

Argyle, M. (1969) *Social interaction*. London: Methuen.

Argyle, M. (1988) *Bodily Communication* (2nd ed.). London: Methuen.

Argyle, M. & Dean, J. (1965) Eye contact, distance and affiliation. *Sociometry*, 28, 289-304.

Argyle, M. & Ingham, R. (1972) Gaze, mutual gaze, and proximity. *Semiotica*, 6, 32-49.

Attneave, F. (1959) *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Holt, Reinhart and Winston.

Austin, J. L. (1962) *How to do things with words*. Cambridge, Mass.: Harvard U P.

Bakeman, R. & Gottman, J.M. (1986)*Observing interaction:An introduction to sequential analysis* Cambridge: CUP.

Bales, R. F. (1983) SYMLOG: A practical approach to the study of group. In H. H. Blumberg, A. P. Hare, V. Kent, & M. Davies (Eds.), *Small groups and social interaction* (pp. 499 - 523). New York: Wiley.

Bannon, L. (1995) Editorial, Commentaries and a response in the Suchman-Winograd debate. *Computer Supported Cooperative Work*, 3(1), 29.

Bannon, L. J. (1991) From human factors to human actors: The role of psychology and human computer interaction studies in system design. In J. Greenbaum & M. Kyng (Eds.), *Design at work: cooperative design of computer systems* Hillsdale, NJ: Lawrence Erlbaum.

Barnard, P. & Harrison, M. (1989) Integrating Cognitive and System Models in Human Computer Interaction. In *Proceedings of the HCI'89 Conference on People and Computers V* (pp. 87-103).

Barnard, P., May, J. & Salber, D. (1996) Deixis and points-of-view in media spaces. *Behaviour and information technology*, 15(1), 37-50.

Barnard, P., Wilson, M. & MacLean, A. (1987) Approximate Modelling of Cognitive Activity: Towards an Expert System Design Aid. In *Proceedings of ACM CHI+GI'87 Conference on Human Factors in Computing Systems and Graphics Interface* (pp. 21-26).

Bateson, P. P. G. (1987) Ethology. In R. L. Gregory (Eds.), *The Oxford Companion to the Mind* Oxford: OUP.

Bellotti, V. & Sellen, A. (1993) Design for privacy in ubiquitous computing environments. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work* (pp. 77-92).

Birdwhistell, R. L. (1970) *Kinesics and context*. Philadelphia: Philadelphia University Press.

Blackler, F. (1995) Activity theory, CSCW and organization. In A. F. Monk & N. Gilbert (Eds.), *Perspectives on HCI: Diverse approaches* London: Academic Press.

Bødker, S. (1990) *Through the interface - a human activity approach to user interface design*. Hillsdale, NJ: Erlbaum.

Bødker, S. (1996) Applying Activity Theory to video analysis: How to make sense of video data in Human-Computer Interaction. In B. A. Nardi (Eds.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* (pp. 147-174). Cambridge, Mass: MIT Press.

Bowers, J. M. & Benford, S. D. (Ed.) (1991). *Studies in Computer Supported Cooperative Work: Theory, Practice and Design.* Amsterdam: North Holland.

Boyle, E., Anderson, A. & Newlands, A. (1994) The effects of visibility on dialogue and performance in a cooperative problem-solving task. *Language and Speech,* 37(1), 1-20.

Brittan, D. (1992) Being there: the promise of multimedia communications. In R. M. Baecker (Eds.), *Readings in Groupware and Computer-supported Cooperative Work: Assisting human-human collaboration* (pp. 57 - 67). San Mateo, CA: Morgan Kaufmann Pubs. Inc.

Bruce, V. (1996) The role of the face in communication: implications for video-phone design. *Interacting with Computers,* 8(2), 166-176.

Bull, P. E. (1987) *Posture and Gesture.* Oxford: Pergamon.

Bull, P. E. (1994) On identifying questions, replies, and non-replies in political interviews. *Journal of Language and Social Psychology,* 13, 115 - 131.

Bull, P. E. & Frederikson, L. (1994) Non-verbal communication. In A. M. Colman (Eds.), *Companion Encyclopaedia of psychology* London: Routledge.

Butters, L., Clarke, A., Hewson, T. & Pomfrett, S. (1994) The Dos and Don'ts of Video Conferencing in Higher Education: A report to the advisory group on computer graphics (Support Initiative for Multimedia Applications No. 4). *HUSAT Research Institute.*

Campbell, R. (1998) *Hearing by Eye II: Advances in the psychology of speechreading and auditory-visual speech.* The Psychology Press.

Campbell, R., Brooks, B., DeHaan, E. & Roberts, T. (1996) Dissociating face processing skills - decisions about lip-read speech, expression and identity. *Quarterly journal of experimental psychology Section A - Human Experimental Psychology,* 49(2), 295 - 314.

Camras, L. (1980) Animal threat displays and children's facial expressions: A comparison. In D. R. Omark, F. F. Strayer, & D. G. Freedman (Eds.), *Dominance relations: An Ethological View of Human Conflict and Social Interaction* (pp. 121 - 136). New York: Garland STMP.

Card, S. K., Moran, T. P. & Newell, A. (1983) *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Champness, B. G. (1972a) Attitudes towards person-person communications media (Unpublished paper No. E/72245/CH). *Communications Studies Group.*

Champness, B. G. (1972b) Feelings towards media in group situations (Unpublished paper No. E/72160/CH). *Communications Studies Group.*

Champness, B. G. & Davies, M. F. (1971) The Maier pilot experiment (E/71030/CH No. Cited in Short, Williams and Christie (1976)). *British Post Office Corporation, Long Range Research, Communications Studies Group.*

Chance, M. R. A. (1967) Attention structures as the basis of primate rank orders. *Man,* 2, 503 - 518.

Chapanis, A. (1975) Interactive Human Communication. *Scientific American,* 232(March), 36 - 42.

Cherry, E. C. & Taylor, W. K. (1954) Some further experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America,* 26, 554 - 559.

Chevalier-Skolnikoff, S. (1973) Facial expression of emotion in non-human primates. In P. Ekman (Eds.), *Darwin and facial expression: a century of research in review* New York: Academic.

Clark, H. H. (1996) *Using Language.* Cambridge: CUP.

Clark, H. H. & Brennan, S. E. (1991) Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127 - 149). Washington, DC: American Psychological Association.

Clark, H. H. & Schaefer, E. R. (1989) Contributing to discourse. *Cognitive Science,* **13**, 259 - 294.

Cohen, K. M. (1982) Speaker interaction: Video teleconferences versus face-to-face meetings. In L. A. Parker & C. H. Olgren (Eds.), *Teleconferencing and electronic communications: Applications, technologies and human factors* (pp. 189 - 199). Madison, WI: University of Wisconsin Extension, Center for Interactive Programs.

Cummins, R. (1983) *The nature of psychological explanation.* Cambridge, Mass.: Bradford Books/MIT Press.

Daly-Jones, O. (1998) *Visual and auditory information in electronically mediated communication.* M Phil, University of York, United Kingdom.

Darwin, C. R. (1892) *The Expression of Emotion in Man and Animals.* London: John Murray.

Diaper, D. & Johnson, P. (1989) Task analysis for knowledge descriptions: theory and application in training. In J. Long & A. Whitefield (Eds.), *Cognitive ergonomics and human computer interaction* Cambridge: Cambridge University Press.

Dix, A. (1992) Pace and interaction. In A. Monk, D. Diaper, & M. D. Harrison (Eds.), *People and Computers 7* (pp. 193-208). Cambridge: Cambridge University press.

Duncan, S. (1972) Some signals and rules for taking speaking turns in conversations. , **23**, 283 - 292.

Duncan, S., Brunner, L. J. & Fiske, D. W. (1979) Strategy signal in face-to-face interaction. *Journal of Personality and Social Psychology,* **37**, 301 - 13.

Duncan, S. & Fiske, D. W. (1977) *Face-to-face interaction: Research, methods and theory.* Hillsdale, New Jersey: Lawrence Erlbaum.

Duncan, S. & Niederehe, G. (1974) On signalling that it's your turn to speak. *Journal of Experimental Social Psychology,* **10**, 234 - 247.

Dunn, E. V., Conrath, D. W., Bloor, W. G. & Tranquada, B. (1977) An evaluation of four telemedicine systems for primary care. *Health Services Research,* **1**, 19 - 29.

Egido, C. (1988) Videoconferencing as a Technology to Support Group Work: A Review of its Failure. In *Proceedings of ACM CSCW'88 Conference on Computer-Supported Cooperative Work* (pp. 13-24).

Egido, C. (1990) Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In J. Gallegher, R. E. Kraut, & C. Egido (Eds.), *Intellectual teamwork: social and technological foundations of cooperative work* (pp. 351 - 371). Hillsdale, New Jersey.: LEA.

Ekman, P. & Friesen, W. V. (1969) The repertoire of nonverhal behaviour: Categories, origins, usage and coding. *Semiotica,* **1**, 49 - 98.

Ellsworth, P. C. &. L., E.J. (1976) Staring and approach: An interpretation of the stare as a nonspecific activator. .

Ellyson, S. L. & Dovidio, J. F. (1985) *Power, dominance and nonverbal behaviour*. New York: Springer Verlag.

Engestrom, Y. (1993) Developmental studies of work as a testbench of activity theory: the case of primary care medical practice. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: perspectives on activity and context* Cambridge: CUP.

Englebart, D. C. & English, W. K. (1968) A research centre for augmenting the human intellect. In *American Federation of Information Processing Societies, Fall Joint Computer Conference*, 33 (pp. 395 - 410). San Francisco, CA:

Exline, R. V. ;. G., D. & Schuette, D. (1965) Visual behaviour in a dyad as affected by interview content and sex of respondent. , 33, 117 - 122.

Finn, K. E. (1997) Introduction: An overview of video-mediated communication literature. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 3 - 21). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Fish, R. S., Kraut, R. E. & Chalfonte, B. L. (1990) The VideoWindow System in Informal Communications. In *Proceedings of ACM CSCW'90 Conference on Computer-Supported Cooperative Work* (pp. 1-11). New York: ACM Press.

Fish, R. S., Kraut, R. E., Root, R. W. & Rice, R. E. (1992) Evaluating video as a technology for informal communication. In *Proceedings of ACM CHI'92 Conference on human factors in computing systems* (pp. 37 - 48). New York: ACM Press.

Fish, R. S., Kraut, R. E., Root, R. W. & Rice, R. E. (1993) Video as a technology for informal communication. *Communications of the ACM*, 36(1), 48-61.

Flor, N. V. & Hutchins, E. L. (1991) Analyzing Distributed Cognition in Software Teams: A Case Study of Team Programming During Perfective Software Maintenance. In *Empirical Studies of Programmers: Fourth Workshop* (pp. 36-64).

Frohlich, D. (1995) Requirements for interpersonal information management. In P. J. Thomas (Eds.), *Personal information systems: business applications* (pp. 133-153). Cheltenham: Stanley Thornes and Unicom Seminars.

Gale, S. (1990) Human Aspects of Interactive Multimedia Communication. *Interacting with Computers*, 2(2), 175-189.

Gaver, W., Moran, T., Maclean, A., Lovstrand, L., Dourish, P., Carter, K. & Buxton, W. (1992) Realizing a video environment: Europarc's RAVE system. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems* (pp. 27-35). New York: ACM.

Gaver, W., Sellen, A., Heath, C. & Luff, P. (1993) One is Not Enough: Multiple Views in a Media Space. In *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems* (pp. 335-341).

Gaver, W., Smets, G. & Overbeeke, K. (1995) A virtual window on media space. In *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems* (pp. 257-264).

Goffman, E. (1964) *Behaviour in public places*. Glencoe, Illinois: The Free Press.

Golomb, S. W., Peile, R. E. & Scholtz, R. A. (1994) *Basic concepts in information theory and coding*. New York: Plenum Press.

Greatbatch, D., Heath, C., Luff, P. & Campion, P. (1995) Conversation analysis: human-computer interaction and the general practice consultation. In A. F. Monk & N. Gilbert (Eds.), *Perspectives on HCI: diverse approaches* London: Academic Press.

Haccou, P. (1987) *Statistical methods for ethological data* Doctoral thesis, Institute of Theoretical Biology, University of Leiden, The Netherlands. Published by Centrum voor Wiskunde en Informatica, Amsterdam.

Harrison, B. L. & Baecker, R. M. (1991) Video analysis in collaborative work (Technical Report No. *CSRI/Department of Computer Science, University of Toronto.*

Heath, C. & Luff, P. (1992a) Disembodied conduct: interactional asymmetries in video mediated communication (Technical report No. EPC-92-119). *Rank Xerox.*

Heath, C. & Luff, P. (1992b) Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human-Computer Interaction*, 7(3), 315-346.

Heath, C. & Luff, P. (1995) Ethnography: aligning field studies of work and system design In A. F. Monk & N. Gilbert (Eds.), *Perspectives on HCI: Diverse approaches* (pp. 175-198) London: Academic Press.

Hinde, R. A. (1974) *Biological basis of social behaviour*. New York: McGraw Hill.

Howes, A. (1994) A Model of the Acquisition of Menu Knowledge by Exploration. In C. Plaisant (Eds.), *Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems* (pp. 232). New York: ACM Press.

Howes, A. & Payne, S. J. (1990) Display-based competence: towards user models for menu-driven interfaces. *International Journal of Man Machine Studies*, 33, 637 - 655.

Hutchins, E. (1994) *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E. & Klausen, T. (1991) Distributed cognition in an airline cockpit. In Y. Engeström & D. Middleton (Eds.), *Cognition in communication at work* Cambridge: CUP.

Ishii, H. (1990) TeamWorkStation: Towards a Seamless Shared Workspace. In *Proceedings of ACM CSCW'90 Conference on Computer-Supported Cooperative Work* (pp. 13-26).

Ishii, H. & Arita, K. (1991) ClearFace: Translucent Multiuser Interface for TeamWorkStation. *ACM SIGCHI Bulletin*, 23(4), 67-68.

Ishii, H. & Kobayashi, M. (1992) Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems* (pp. 525-532).

Johnson, H. & Johnson, P. (1991) Task knowledge structures: Psychological basis and integration into system design. *Acta Psychologica*, 78, 3 - 26.

Kaptelinin, V. (1996) Computer-mediated activity: Functional organs in social and developmental contexts. In B. A. Nardi (Eds.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* (pp. 45-68). Cambridge, Mass: MIT Press.

Kendon, A. (1967) Some function of gaze direction in social interaction. *Acta psychologica*, 26, 22 - 63.

Kendon, A. (1990) *Conducting interaction: patterns of behaviour in focused encounters*. Cambridge: CUP.

Kieras, D. E. & Polson, P. G. (1985) An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22, 365-394.

Kirk, R. E. (1968) *Experimental design: Procedures for the behavioural sciences.*

Kleck, R. E. & Nuessle, W. (1968) Congruence between the indicative and communicative functions of eye contact in interpersonal relations. *British Journal of Social and Clinical Psychology*, 7, 214 - 216.

Kuutti, K. (1996) Activity theory as a potential framework for human-computer interaction research. In B. A. Nardi (Eds.), *Context and Consciousness: Activity Theory and Human-Computer Interaction* (pp. 17-44). Cambridge, Mass.: MIT Press.

Kuzuoka, H. (1992) Spatial workspace collaboration: A SharedView video support system for remote collaboration capability. In *CHI'92 Conference on Human Factors in Computing Systems,* Proceedings (pp. 533 - 540). Monterey, California: ACM Press.

Kuzuoka, H. & Shoji, H. (1994) Results of observational studies of spatial workspace collaboration. *Electronics and communications in Japan, Part 3,* 77(8), 58 - 68.

Larson, R. R. (1976) Charisma: A reinterpretation. In M. R. A. Chance & R. R. Larson (Eds.), *The social structure of attention* London: John Wiley and Sons.

Levinson, S. C. (1983) *Pragmatics.* Cambridge: Cambridge University Press.

Lindbergh, D. (1996) The H.324 multimedia communication standard. *IEEE Communications Magazine,* 34(12), 46 - 51.

Losada, M. & Markovitch, S. (1990) GroupAnalyzer: a system for dynamic analysis of group interaction. In *23rd Annual Hawaii International Conference on System Sciences,* Proceedings (pp. 101 - 110). Hawaii: IEEE Computer Society.

McCarthy, J. (1995) Applied experimental psychology. In A. F. Monk & N. Gilbert (Eds.), *Perspectives on HCI: diverse approaches* (pp. 75-96). London: Academic Press.

McGurk, H. & McDonald, J. (1976) Hearing lips and seeing voices. *Nature,* 264, 746 - 748.

Martin, P. & Bateson, P. (1993) *Measuring Behaviour* (2nd Ed.) Cambridge: CUP.

Mee, A. (1898) The pleasure telephone. *The Strand Magazine,* 16, 339-345.

Monk, A. F. (1998) Cyclic interaction: a unitary approach to intention: action and the environment. *Cognition.*68, 99-110.

Monk, A. F. (In press) Modelling cyclic interaction. *Behaviour and Information Technology.*

Monk, A. F. & Gilbert, N. (1995) *Perspectives on HCI: diverse approaches* London: Academic Press.

Moore, G. T., Willemain, T. R., Bonanno, A. B., Clark, W. D., Martin, A. R. & Mogielnicki, R. P. (1975) Comparison of television and telephone for remote medical consultation. *The New England Journal of Medicine,* 13, 729 - 32.

Morely, I. E. & Stephenson, G. M. (1969) Interpersonal and interparty exchange: a laboratory simulation of an industrial negotiation at the plant level. *British Journal of Psychology,* 60, 543 - 545.

Morley, I. E. & Stephenson, G. M. (1977) *The social psychology of bargaining.*

Nardi, B., Kuchinsky, A., Whittaker, S., Leichner, R. & Schwarz, H. (1996) "Video-as-data": Technical and social aspects of a collaborative multimedia application. *Computer Supported Cooperative Work,* 4, 73 - 100.

Nardi, B. A. (Ed.) (1996). *Context and consciousness: activity theory and human-computer interaction.* Cambridge, MA: MIT Press.

Nardi, B. A. & Miller, J. R. (1991) Twinkling Lights and Nested Loops: Distributed Problem Solving and Spreadsheet Development. *International Journal of Man-Machine Studies,* 34(2), 161-184.

Nardi, B. A., Schwartz, H., Kuchinsky, A., Leichner, R., Whittaker, S. & Sclabassi, R. (1993) Turning away from talking heads: the use of video-as-data in neurosurgery. In *InterCHI'93 Conference on Human Factors in Computing Systems,* Proceedings (pp. 327-334). Amsterdam, Netherlands: ACM Press.

Nayak, N. P., Mrazek, D. & Smith, D. R. (1995) Analyzing and communicating usability data: Now that you have the data what do you do? Report on the CHI'94 Workshop. *Communications of the Association of Computing Machinery.*

Neisser, U. (1976) *Cognition and Reality: Priniciples and Implications of Cognitive Psychology.* San Francisco: W.H.Freeman and Company.

Newlands, A. J. S. (1998) *The Effects of Computer-Mediated Communication on the Processes of Communication and Collaboration.* D.Phil, University of Glasgow.

Niemark, E. D. (1970) Model for a thinking machine: An information processing framework for the study of cognitive development. *Merrill-Palmer Quarterly,* **16**, 345 - 368.

Noldus, L. P. J. J. (1991) The Observer: a software system for dynamic analysis of group interaction. *Behaviour Research Methods, Instruments and Computers,* **23(3)**, 415 - 429.

Noll, M. A. (1992) Anatomy of a failure: picturephone revisited. *Telecommunications policy,* **16**, 307-316.

O'Conaill, B., Whittaker, S. & Wilbur, S. (1993) Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human-Computer Interaction,* 8(4), 389-428.

O'Malley, C., Langton, S., Anderson, A., Doherty- Sneddon, G. & Bruce, V. (1996) Comparison of face-to-face and video mediated interaction. *Interacting with Computers,* 8(2), 177-192.

Olson, J. & Storrosten, M. (1990) Finding the golden thread: Representations for the analysis of videotaped group work (Technical Report No. *CSMIL, University of Michigan.*

Omark, D. R., Strayer, F. F. & Freedman, D. G. (Ed.) (1980). *Dominance relations: An ethological view of human conflict and social integration.* New York: STMP Press.

Owen, R. N., Baecker, R. M. & Harrison, B. (1994) Timelines, a tool for gathering, coding and analysis of temporal HCI usability data. In C. Plaisant (Ed.), *CHI'94 Conference on Human Factors in Computing Systems,* Companion . Boston, Mass.: Addison-Wesley.

Pagani, D. S. & Mackay, W. E. (1993) Bringing media spaces into the real world. In G. D. Michelis, C. Simone, & K. Schmidt (Ed.), *Third European Conference on Computer Supported Cooperative Work,* (pp. 341 - 356). Milan, Italy: Kluwer Academic Publishers.

Puopolo, M.; Venerosi, A.; Chiarotti, F. & Alleva, E. (1998) Continuous time Markov chain models: a sueful tool in understanding behavioral patterns In: *Proceedings of Measuring Behavior'98,* The 2nd International Conference on Methods and Techniques in Behavioral Research, 18-21 August 1998, Groningen, The Netherlands.

Radley, A. (1991) *The body and social psychology.* New York: Springer Verlag.

Reid, A. A. L. (1971) The telecommunications impact model (unpublished technical report, cited in E.Williams (1974) "A summary of the present state of knowledge regarding the effectiveness of the substitution of face-to-face meetings by telecommunicated meetings" Communications studies group paper no.: P/74294/WL No. P/71161/RD). *Communications Studies Group.*

Robinson, M. (1993) Design for unanticipated use. In G. D. Michelis, C. Simone, & K. Schmidt (Ed.), *Third European Conference on Computer Supported Cooperative Work,* (pp. 189 - 202). Milan, Italy: Kluwer Academic Publishers.

Russell, J. A. (1991) Culture and the categorization of emotions. *Psychological Bulletin,* **110(3)**, 426 - 450.

Rutter, D. (1984) *Looking and Seeing: The role of visual communication in social interaction.* Chichester: John Wiley & Sons.

Rutter, D. R., Stephenson, G. M., Ayling, K. & White, P. A. (1978) The timing of looks in dyadic conversations. *British Journal of Social and Clinical Psychology*, **17**, 17 - 21.

Rutter, D. R., Stephenson, G. M. & Dewey, M. E. (1981) Visual communication and the content and style of conversation. *British Journal of Social Psychology*, **20**, 41 - 52.

Ryle, G. (1949) *Concept of Mind.*

Sacks, H., Schegloff, E. A. & Jefferson, G. (1974) A simplest systematics for the organisation of turn taking in conversation. *Language*, **50(4)**, 696 - 735.

Sanderson, P. M., McNeese, M. & Zaff, B. (1994) Knowledge elicitation and observation in engineering psychology: MacSHAPA and COGENT. *Behavior Research Methods, Instruments, and Computers*, **26**, 117 - 124.

Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Watanabe, L. M. & James, J. M. (1994) MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). *International Journal of Human-Computer Studies*, **41(633 - 68)**.

Schaphorst, R. A. (1996) Status Of H.3214 - The Videoconferencing Standard For The Public Switched Telephone Network And Mobile Radio. *Optical Engineering*, **35(1)**, 109 - 112.

Searle, J. R. (1969) *Speech Acts*. Cambridge: Cambridge University Press.

Seifert, C. M. & Hutchins, E. L. (1992) Error as Opportunity: Learning in a Cooperative Task. *Human-Computer Interaction*, **7(4)**, 409-435.

Sellen, A. J. (1992) Speech patterns in video-mediated conversations. In *CHI'92 Conference on Human Factors in Computing Systems*, . Monterey, California: ACM.

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

Shackle, B. (1996) Ergonomics *The Psychologist*, July 1996, **9(7)**, 307-312.

Shannon, C. E. & Weaver, N. (1949) *The mathematical theory of communication*. Urbana: University of Illinois Press.

Short, J. (1974) Effect of medium of communication on experimental negotiation. *Human relations*, **27**, 225 - 234.

Short, J., Williams, E. & Christie, B. (1976) *The social psychology of telecommunications*. London: John Wiley and Sons.

Silverman, D. (1985) *Qualitative methodology and sociology*. Vermont: Gower Publishing.

Smith, R., O'Shea, T., O'Malley, C., Scanlon, E. & Taylor, J. (1991) Preliminary experiments with a distributed, multimedia, problem solving environment. In J. Bowers & S. Benford (Eds.), *Studies in computer-supported cooperative work* (pp. 31-48).

Snowdon, C. T. (1983) Ethology, comparative psychology and animal behaviour. *Annual Review of Psychology*, **34**, 63 - 94.

Strongman, K. T. (1970) Communicating with the eyes. *Science*, **6**, 47 - 52.

Strongman, K. T. & Champness, B. G. (1968) Dominance Hierarchies and Conflict in Eye Contact. *Acta Psychologica*, **28**, 376 - 86.

Stroop, J. R. (1935) Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, **18**, 643 - 662.

Suchman, L. (1993) Do categories have politics? The Language/Action perspective reconsidered. In G. de Michelis, K. Schmidt, & C. Simone (Ed.), *Third European Conference on Computer-Supported Cooperative Work,* (pp. 1 - 14). Milan, Italy: Kluwer Academic Publishers.

Suchman, L. A. (1987) *Plans and Situated Actions: The Problem of Human-Computer Communication*. New York: Cambridge University Press.

Tang, J. C. (1991) Findings From Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies,* 34(2), 143-160.

Tang, J. C. & Minneman, S. L. (1990) VideoDraw: A Video Interface for Collaborative Drawing. In *Proceedings of ACM CHI'90 Conference on Human Factors in Computing Systems* (pp. 313-320).

Tang, J. C. & Minneman, S. L. (1991) VideoWhiteboard: Video Shadows to Support Remote Collaboration. In *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 315-322).

Thomas, A. P., Bull, P. J. & Roger, D. (1982) Conversation exchange analysis. *Journal of language and social psychology,* 1(2), 141 - 155.

Thomas, A. P., Roger, D. B. & Bull, P. E. (1983) A sequential analysis of informal dyadic conversation using Markov chains. *British Journal of Social Psychology,* 22, 177 - 188.

Tinbergen, N. (1963) On the aims and methods of ethology. *Zeitschrift fuer Tierpsychologie,* 20, 410 - 33.

Vertegaal, R. (1997) Conversational awareness in multiparty VMC. In S. Pemberton (Ed.), *CHI'97 Conference on Human Factors in Computing Systems,* Extended Abstracts (pp. 6 - 7). Atlanta, Georgia: ACM Press.

Vertegaal, R., Vons, H. & Slagter, R. (1998) Look who's talking: the GAZE groupware system. In *CHI'98 Conference on Human Factors in Computing Systems,* Summary (pp. 293 - 294). Los Angeles, California: ACM Press.

Wagner, H., Clarke, A. H. & Ellgring, J. H. (1983) Eye contact and individual looking: The role of chance. *British Journal of Social Psychology,* 22, 61 - 61.

Watts, L. A. & Monk, A. F. (1997) Telemedical consultation: task characteristics. In S. Pemberton (Ed.), *CHI'97 Conference on Human Factors in Computing Systems,* Proceedings (pp. 534 - 535). Atlanta, Georgia. USA.: ACM Press.

Watts, L. A. & Monk, A. F. (In Press) Telemedicine: what happens in teleconsultation. *International Journal of Technology Assessment in Health Care.*

Watts, L. A., Monk, A. F. & Daly-Jones, O. (1996) Inter-personal awareness and synchronization: assessing the value of communication technologies. *International Journal of Human Computer Studies,* 44(6), 849 - 875.

Watzlawick, P., Beavin, J. H. & Jackson, D. D. (1968) *Pragmatics of human communication.* Faber & Faber.

Weber, K. & Poon, A. (1994) Marquee: a tool for real-time video logging. In B. Adelson, S. Dumais, & J. Olson (Ed.), *CHI'94 Conference on Human Factors in Computing Systems,* Proceedings . Boston, Mass.: Addison-Wesley.

Whittaker, S. (1995) Rethinking video as a technology for interpersonal communications: theory and design implications. *International Journal of Human-Computer Studies,* 42, 501-529.

Wiener, M., Devoe, S., Robinson, S. & Geller, J. (1972) Nonverbal behaviour and nonverbal communication. *Psychological Review,* 79, 185 - 214.

Wiener, N. (1948) *Cybernetics*. New York: Wiley.

Williams, E. (1974) A summary of the present state of knowledge regarding the effectiveness of the substitiution of face-to-face meetings by telecommunicated meetings: type allocation revisisted (Long Range Research Paper 67 No. P/74924/WL). *British Post Office, Communication Studies Group*.

Williams, E. (1977) Experimental comparisons of face-to-face and mediated communication: a review. *Psychological Bulletin*, 84, 963 - 976.

Williams, E. (1981) The future of the media. In R. Williams (Eds.), *Contact: Human communication and its history* London: Thames & Hudson.

Yngve, V. H. (1970) On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society* Chicago: Chicago Linguistic Society.

Young, M. F. & McNeese, M. (1993) A Situated Cognition Approach to Problem Solving with Implications for Computer-Based Learning and Assessment. In *Proceedings of the Fifth International Conference on Human-Computer Interaction* (pp. 825-830).

Young, R. M. & Barnard, P. J. (1991) Signature Tasks and Paradigm Tasks: New Wrinkles on the Scenarios Methodology. In *Proceedings of the HCI'91 Conference on People and Computers VI* (pp. 91-101).

Young, R. M., Howes, A. & Whittington, J. (1990) A knowledge analysis of interactivity. In D. Diaper, D. Gilmore, G. Cockton, & B. Shakel (Ed.), *Interact'90*, (pp. 115-120). Cambridge: North-Holland.

Young, R. M. & Simon, T. (1987) Planning in the Context of Human-Computer Interaction. In *Proceedings of the HCI'87 Conference on People and Computers III* (pp. 363-370).

Yule, G. (1985) *The study of language: an introduction*. Cambridge: Cambridge University Press.

Zinchenko, V. P. (1995) Activity theory psychology and Cultural-historical psychology. In J. V. Wertsch, P. Del Rio, & A. Alvarez (Eds.), *Sociocultural Studies of Mind* Cambridge: CUP.