The DCIEM Map Task Corpus: Spontaneous Dialogue under Sleep Deprivation and Drug Treatment

E. G. Bard[1], C. Sotillo[1], A. H. Anderson[1], M. M. Taylor[2]

[1]Human Communication Research Centre, Universities of Edinburgh and Glasgow
[2]Defence and Civil Institute of Environmental Medicine, Canada

ABSTRACT

This paper describes a resource for the study of spontaneous speech under stress, a corpus of 216 unscripted task-oriented dialogues conducted by normal Canadian adults in the course of a sleep deprivation experiment under 3 drug conditions. Speakers carried out the route-communication task (see [1]) in alternation with a battery of other tasks over a 6-day study which included a 60-hour sleepless period. Each speaker participated in 12 dialogues. The design permits comparisons within speakers for sleep deprivation (baseline, deprived, post-recovery), and between speakers for drug condition (placebo, d-amphetamine, Modafinil) and number of conversational partners encountered (1, 2). Preliminary examination of dialogue length, task performance, and aspects of dialogue strategy indicate effects of all these variables. Effects of sleep-deprivation and drug condition are less severe than those found in simpler tasks [7].

1. INTRODUCTION

Stress potentially affects both the quality of speech and the performance of the speaker on speech-based tasks. This paper describes a corpus of spontaneous task-oriented dialogues collected in the course of a sleep-deprivation study which will make it possible to study both. The corpus is a large-scale elicitation experiment with factorial combinations of degree of sleep-deprivation, number of partners encountered, and drug treatment. The sleep-deprivation study [7, 4] was designed both to reveal the properties of Modafinil, a stimulant without many of the undesirable effects known in amphetamines, and to examine the effects of sleep-deprivation combined with continuous work. The study also attempted to extend the finding that dialogue tasks profit from speakers' contact with multiple partners [5].

2. METHOD

2.1. The Dialogue Task

The corpus uses the map task as developed in the HCRC Map Task Corpus [3, 1]. Speakers participate in pairs. In each pair, one speaker, the 'Instruction Giver' has a schematic map with a route printed on it. The other, designated the 'Instruction Follower', has a similar map with no route. Despite these designations, both speakers are allowed to say whatever is necessary to reproduce the Giver's route on the Follower's map. Neither can see the other's map and no hand gestures are allowed.

The maps themselves portray purely imaginary locations by means of labelled cartoon landmarks which do not always match in name, number, or site across Giver's and Follower's maps. Participants know that but not where maps differ. These design features ensure, first, that all participants begin by being equally ignorant of what the maps portray; second, that the experimenters can assess the accuracy of participants' descriptions independently of the way the dialogues themselves develop; third, that the local difficulty of the route communication task can be varied while the global difficulty is controlled.

The landmarks labels are designed to allow opportunities for well-known connected-speech phenomena, like nasal assimilation (*green bay*), /d/-deletion (*gold mine*), /t/-deletion (*east lake*), glottalization (*white mountain*), and weak syllable reduction (*buffalo*). At the end of the study, each speaker reads a list of landmark names covering the maps just used. The DCIEM CORPUS used 12 different map pairs, and 156 landmark names.

The resulting dialogues are thus task-oriented but unscripted. Because the maps differ, the task is not straightforward and subjects pursue it attentively. In other corpora using this task, the HCRC Map Task Corpus [1] and the Chiba University Map Task Corpus, normal young adult subjects not exposed to additional stresses show a range of abilities to pursue the goal effectively [2]. sufficiently high instrumental analysis

2.2. Talkers

The speakers in the present corpus were 35 Canadian adults, 33 male and 2 female, all army reservists volunteering for a sleep-deprivation experiment. A thirty-sixth subject declined to participate at the last minute and was replaced by

a series of experimenters. One participant became ill during the study and was temporarily replaced by experimenters. Subjects' military ranks ranged from private to captain, their ages from 20 to 48 years ($\overline{x} = 26.1$, $s.d. = 8.3$).

2.3. Recording

Dialogues took place in an internal room under quiet conditions. Subjects faced each other, each sitting by the long side of different padded 6ft x 3ft table, with the tables separated by a portable room-divider and bordered at left and right ends by further room dividers. Stereo DAT recordings (Panasonic SV-3500) used one close-talking head-worn microphone (Shure model SM10A), with levels boosted (Mackie Designs model CR1604 mixer), and one recording channel per speaker.

2.4. Design

Sleep deprivation and task schedule. The map task was administered within a set of tasks repeated cyclically during a 6-day experiment. Subjects were run in groups of 6, with the subjects in a run isolated together for the full 6 days, sleeping as normal except for a 60-hour sleepless period which began on the morning of day 3. Subjects worked on various psychological tasks for almost all of their waking hours. Tasks were scheduled in 6-hour blocks of 3 2-hour sessions each, with approximately 15 minutes rest at the end of each session. The first day of the study was used for introduction to the tasks. Blocks began on the second day and continued throughout, except for sleep periods on first, second, fifth, and sixth nights. One pair of subjects participated in a map task dialogue as the last task in each 2 hour session. As a result, each subject participated in 1 practice dialogue (day 2), in 3 or 4 assessed dialogues before sleep-loss (days 2 - 3), in 7 or 8 dialogues during the sleep-deprived period (days 4 - 5), and in a single dialogue after post-deprivation sleep (day 6). The maps for all runs of the experiment, were taken from Quads 2 - 4 of the HCRC design.

Subjects regularly filled in mood and physiological feelings questionnaires. Other tasks included tracking some aspect of a screen display with a mouse over a 50-second period; time estimation; attention splitting over quadrants of the computer screen; mood and physiological feelings questionnaires; rapid logical reasoning; sequential subtraction; judgment of relative line length; efficient route planning; group decision making. For all tasks except this last and the map task, subjects were isolated in separate rooms.

Drug Treatments Two runs were assigned to each of 3 drug treatments (amphetamines: runs 1 and 6; Modafinil [6]: runs 3 and 5; placebo: runs 2 and 4) on a double-blind regime, with all subjects in a run receiving the same drug by mouth after 18 hours (Block 4, Session 2), 48 hours (Block 9, Session 2), and 58 hours (Block 11, Session 1) without sleep. Drug dosages were 20mg of dextro-amphetamine and 300 mg of Modafinil. The placebo was a standard inert substance.

Rotation patterns. Within each run, speakers were assigned either to a *foursome* or to a *pair*. The foursome were assigned to different maps in pairings which replicated those used in the HCRC Map Task Corpus. Over each set of 8 dialogues, each speaker in a foursome was Instruction Giver twice, using the same map with 2 different Instruction Followers, and Instruction Follower twice, using a different map each time. Each of these subjects worked with 2 of the other 3 subjects in his/her foursome. The members of the pair, on the other hand, worked only with one another, alternating in Giver and Follower roles. Over each set of 8 dialogues they encountered each map twice, once as Giver and once as Follower. The order of map pairs was the same for both subgroups. Foursomes collectively completed 24 dialogues per run (12 per speaker) and pairs 12 dialogues per run (12 per speaker).

3. PERFORMANCE MEASURES

The resulting dialogues include over 175,000 word tokens representing approximately 1,900 words. The first 10 or 11 dialogues in each run, 60-66 in total, precede the first lost night of sleep and can be used as a baseline. The next 22 dialogues from the middle of the first sleepless night until the end of the sleep-deprived period, should show progressive effects of sleep deprivation. Finally, the last three dialogues in each run follow a night of recovery sleep. In what follows we present preliminary explorations of gross effects of sleep deprivation, rotation condition and drug treatment. Though we report statistically significant results, the outcomes are markedly less robust than those reported [7] for physiological, mood, and simple task measures. Here we examine two issues, the effects of practice and the care with which speakers pursue the task.

We ask here whether subjects can achieve either savings in the amount of work they do to complete the task or improvements in the accuracy with which they perform it. Accuracy, measured as the deviation of the Follower's drawn route from the Giver's printed route, should improve as the subjects in a run work with one another repeatedly. Amount of work, measured very simply as amount of speech required to complete the task, should fall over repeated attemps, unless more work is actually needed to maintain accuracy under stress. Taylor [8] reports that the present subjects, particularly those in the placebo condition, tended to become less accurate in a tracking task as sleep deprivation continued. Modelling their performance, he shows that subjects' failures can be attributed to excessive attention to the expected results of their actions and insufficient attention to the actual results. Our initial examination of the date suggest an analogue in one aspect of commication.

3.1. Route accuracy

Followers' copied routes were overlaid with a 1cm grid with the squares covering the Giver's printed route blacked out. The *deviation score* was the number of grid squares covering
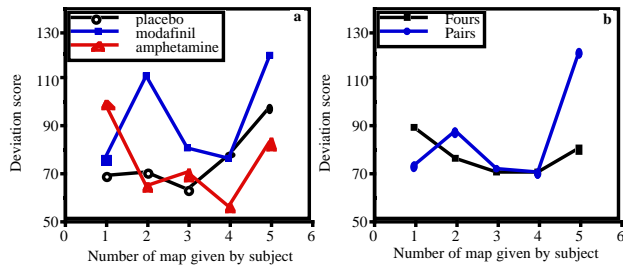
Figure 1: Average deviation of drawn route from model in cm squares over successive maps for each subject as Instruction Giver: a. for each drug treatment; b. for each rotation treatment.

both the parts of the Follower's still visible and the space between those visible sections and the model route.

Figures 1a and b plot average deviation score against the 5 successive dialogues in which each speaker serves as Instruction Giver. Error scores do fall until the fifth map late in the sleepless period when they rise sharply ($F = 3.92$, $d.f. = 4,120$, $p = .005$). Modafinil subjects and subjects working in pairs are especially susceptible to the effects of sleep deprivation on performance. The rise is strongest in the Modafinil treatment groups (Fig. 1a) ($F = 2.14$, $d.f. = 8, 120$, $p = .04$): although this group drew the least accurate routes overall (placebo 93.31, Modafinil 112.35, amphetamine 92.15: $F = 3.41$, $d.f. = 2, 30$, $p < .05$), they had not been the least accurate on their baseline attempts in map 1. As for rotation groups, only subjects working exclusively in pairs (Fig. 1b) show a marked rise in route deviation at the fifth map ($F = 3.03$, $d.f. = 4, 120$, $p = .02$). There is no 3-way interaction.

3.2. Dialogue Length

Amount of speech was measured by average dialogue length in turns and in word tokens. Turns are defined here as speech in a transcribed turn, running from the point where one participant begins to speak to the point where the other begins, whether or not the first has finished, or where one continues to speak during and after an interruption. Word tokens are dictionary-recognizable orthographic units between spaces or between punctuation and spaces. Figures 2 and 3 plot these measures against dialogue number (for all but practice dialogues), averaging over all dialogues in a drug (placebo, Modafinil, amphetamine) x rotation (foursome, pair) condition.

ANOVAs for day of study x drug x rotation showed that over all subjects dialogue length in turns did fall each day ($F = 2.71$, $d.f. = 4, 186$, $p = .031$, Scheffé tests at $p < .01$: 159.89 on day 2, 150.54 on day 3, 134.32 on day 4) before levelling out on day 5 (125.88), the last in the sleepless period, and then rising ($p < .05$) on the post-recovery day (146.61). The downward trend was never significant for placebo subgroups

(Figs. 2a and 3a). Length in turns did fall significantly (Figs. 2b and 2c) for Modafinil pairs ($r = -.425$, $N = 24$, $p = .038$) and foursomes ($r = -.300$, $N = 48$, $p = .038$) and for amphetamine foursomes ($r = -.430$, $N = 48$, $p = .002$). Length in tokens decreased significantly over trials for Modafinil ($r = -.299$, $N = 48$, $p = .039$) and amphetamine ($r = -.382$, $N = 48$, $p = .039$) foursomes only (Figs. 3b and 3c).
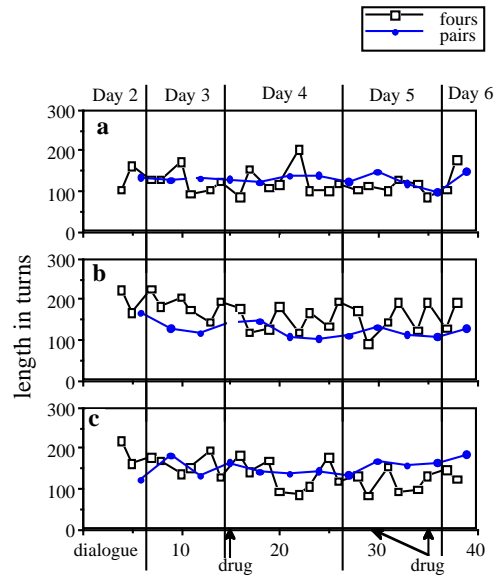

Figure 2: Average length of dialogues in turns: a. placebo subjects; b. Modafinil subjects; c. amphetamine subjects
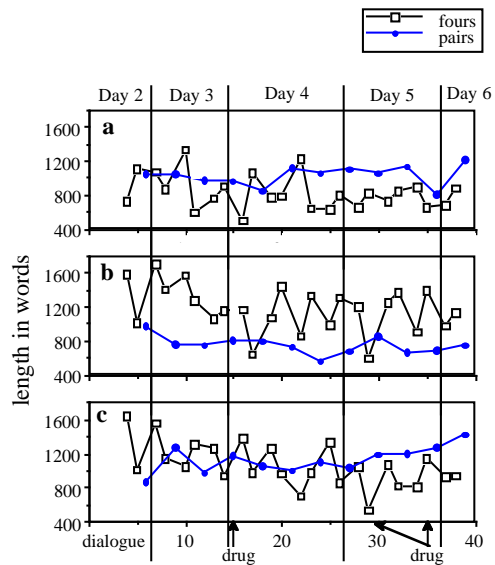

Figure 3: Average length of dialogues in word tokens: a. placebo subjects; b.Modafinil subjects; c. amphetamine subjects

3.3. Feedback

Though preliminary and global, the accuracy and length analyses may suggest a profile for behaviour of individuals in different drug treatments as sleep deprivation progresses. The placebo subjects, who report considerable fatigue and show marked decrements on other tasks [7], maintain accuracy here, though they make no length savings over the study. The Modafinil group show the greatest performance damage late in the sleep deprivation period, but continue to abbreviate dialogues. The amphetamine subjects, reperted [8] to be working to finer criteria in the tracking task, suffer somewhat less in performance while reducing dialogue length.

In earlier work on the map task, we have found that one very good way to encourage accurate performance is to seek feedback actively at critical junctures in a dialogue, in particular at the point where landmarks are introduced for the first time [2]: introductions of new items within questions about the partner's awareness of the item correlate both with informative feedback from partners and with lower route deviation scores. To determine if Modafinil subjects were abbreviating their dialogues at the cost of omitting such useful exchanges, we examined Part 1 of the present Corpus, This provides a small-scale balanced factorial sample which includes one dialogue from each of the 3 successive days of the sleepless period for each speaker in each drug condition. The first mention of each landmark name used in each dialogue was excerpted ($N = 403$) and the utterance in which it occurred was classified either as a question about the listener's knowledge of the landmark or as some other kind of utterance. Results were summed across runs within drug condition and dialogue number.

Despite the tenency to abbreviate dialogues, raw numbers of question introductions per dialogue did not change reliably over time for any group. The only marked result of any kind in this small sample, was a significant upward trend in the number of non-question landmark introductions made by Modafinil subjects ($r = .80$, $N = 9$, $p = .009$). As a result, their ratio of question to non-question introductory mentions fell as the sleepless period progressed ($r = -.66$, $N = 9$, $p = .054$). This last result is anything but conclusive. It remains to be seen whether accuracy, length and feedback measures are related for individuals and whether those relationships differ with drug condition or sleep loss. We present these figures only to make the case that the Corpus invites further analyses.

4. THE PUBLISHED CORPUS

The DCIEM Map Task Corpus is distributed by the Linguistic Data Consortium and by the Human Communication Research Centre. It comprises 12 CD-ROMs. Part 1 of the DCIEM Corpus includes all materials for 54 dialogues which provide a small balanced study of drug treatment and sleep deprivation for use in small-scale pilot studies. Part 2 contains the rest of the materials.

For each dialogue, the Corpus includes a digitized speech file; an orthographic transcription including the HCRC Map Task Corpus SGML-type annotations with time-stamps at the onset of each turn; scanned model and copied maps; a NIST header file; and a TEI-entry point. All CDs also contain explanatory information and a detailed account of the experimental design and the speaker characteristics. Local and public domain software for accessing the speech on a several platforms is also included.

5. REFERENCES

1. A. H. Anderson, M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert. The HCRC Map Task Corpus. *Language and Speech*, 34:351–366, 1991.

2. A. H. Anderson and E. A. Boyle. Forms of introduction in dialogues: their discourse contexts and communicative consequences. *Language and Cognitive Processes*, 9:101–122, 1994.

3. G. Brown, A. H. Anderson, R. Shillcock, and G. Yule. *Teaching Talk*. Cambridge University Press, Cambridge, 1984.

4. A. Buguet, A. Montmayeur, R. Pigeau, and P. Naitoh. Modafinil, d-amphetamine and placebo during 64 hours of sustained mental work: .2. effects on 2 nights of recovery sleep. *J. Sleep Research*, 4:229–241, 1995.

5. S. Garrod and G. Doherty. Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53:181–215, 1994.

6. T. J. Lyons and J. French. Modafinil: The unique properties of a new stimulant. *Aviat. Space Environ. Med.*, 62:432–435, 1991.

7. R. Pigeau, P. Naitoh, A. Buget, C. McCann, J. Baranski, M. Taylor, M. Thompson, and I. Mack. Modafinil, d-amphetamine and placebo during 64 hours of sustained mental work: I: Effects on mood, fatigue, cognitive performance and body temperature. *J. Sleep Research*, 4:218–228, 1995.

8. M. M. Taylor. Effects of modafinil and amphetamine on tracking performance during sleep deprivation. Manuscript, 1996.