Timelines, talk and transcription:
a chronometric approach to simultaneous speech

Richard Forsyth[3], David Clarke[1], Phoenix Lam[2]

[1] School of Psychology, University of Nottingham, U.K.
[2] formerly: Department of English, Hong Kong Polytechnic University, Hong Kong.
[3] formerly: School of Psychology, University of Nottingham, U.K.

Correspondence to: forsyth_rich [AT] yahoo.co.uk

[This is a preprint. Please cite as:
Forsyth, R.S., Clarke, D.D. & Lam, P. (2008).
Timelines, Talk and Transcription: a chronometric approach to simultaneous speech.
International J. of Corpus Linguistics, 13(2), 225-250.

Final published version at:
http://www.ingentaconnect.com/content/jbp/ijcl/2008/00000013/00000002/art00004
]

Abstract

Linguists and other social scientists have employed many transcription conventions to exhibit the temporal interleaving of multi-speaker talk. The existence of many different systems, which are mutually incompatible, is evidence that representing spoken discourse remains problematic. This study proposes a novel orthographic transcription layout based on word timings. To test this method, the Maptask corpus (Anderson et al., 1991) is used because it contains unusually precise information on the timings of vocal events. This makes it possible to evaluate a non-standard talk-division format (TST1) in which the alternation of speakers is not imposed by a transcriber's intuition but emerges from the empirical data. It highlights the prevalence of "echoing" in the joint production of dialogue. Moreover, lengths of speech segments and inter-speaker intervals as defined by this procedure show significant associations with a number of contextual and interactional variables, indicating that this approach has analytic as well as representational benefits.

Keywords: Corpus, Maptask, Simultaneous speech, Speech segmentation, Transcription.

1. Introduction

Representations of spoken discourse must allow for the phenomenon of simultaneous speech. Linguists and discourse analysts have employed numerous transcription conventions for exhibiting the temporal interleaving of multi-speaker talk (e.g. Atkinson & Heritage, 1984; Boden & Zimmermann, 1991; Schiffrin, 1994; Leech et al., 1995; MacWhinney, 1995; Carter, 2004; MICASE, 2007). Most of these conventions are mutually incompatible. Researchers have used a variety of symbols to mark the start and end of simultaneous speech, including braces, brackets, hash signs, and XML-style tags; in addition, they may or may not attempt to align overlapping speech segments vertically on the page. In other words, different researchers make different compromises with the ideal of a sequence of complete non-overlapping utterances.

The great majority of transcription schemes derive ultimately from "the familiar conventions of the playscript" (Payne, 1995: 206). The fact that different researchers make different compromises with this ideal is evidence that representation of turn-taking in natural dialogue remains a problematic issue.

For the English language, different transcription conventions have been developed to serve the needs of different disciplines. In the field of Conversation Analysis (Sacks et al., 1974) the so-called Jeffersonian notation (Atkinson & Heritage, 1984: ix-xvi) devised by Gail Jefferson is prevalent. However, even within this tradition there are variations, as noted by O'Connell and Kowal (1994). For example, Schiffrin (1994) gives three different styles of transcription in an appendix. Likewise, in Corpus Linguistics a variety of schemes have been developed (Svartvik & Quirk, 1980; Svartvik, 1990). The collection edited by Leech et al. (1995), written from a Corpus Linguistics perspective, contains at least four distinct systems of speech transcription. Many other notations have been proposed (e.g. Ehlich, 1993; Gumperz & Berenz, 1993; MacWhinney, 1995; Dahlmann & Adolphs, 2007). For further discussion of issues involved in transcribing speech see Edwards & Lampert (1993) or Leech et al. (1995).

As the title of an article by Ochs (1979), "Transcription as theory", makes clear, there can be no single objective method of transcription. Inevitably, "transcription is an act of interpretation" (Bucholtz, 2000: 1463). Thus a variety of systems for transcribing speech is to be expected. Nevertheless, while we agree "that there can be no privileged, objective position from which to transcribe speech" (Bucholtz, 2000: 1461), we believe that the bewildering profusion of schemes for rendering discourse on the page could benefit from a reduction in variety, and an increase in objectivity, particularly with regard to how the speech stream is divided into segments and assigned to the participating interlocutors.

The present investigation was performed to find out what might be gained by jettisoning the artificially constraining idea of the playscript, and representing in a straightforward chronological manner the order in which participants contribute to the stream of spoken events. Thus this study is mainly methodological: its prime objective is to introduce a simple yet automatic method of dividing transcribed speech into segments on a strictly chronological basis, and to make an initial test of the utility of that method by applying it to a publicly available spoken dataset.

For this purpose, the Maptask corpus (Anderson et al., 1991) was used because it contains unusually precise information on the timings of vocal events. Using the term *vocable* to denote words and certain short phrases spoken without internal silence which the corpus compilers have treated as units for timing purposes (such as "going to" and "you know"), there are over 150,000 vocables in the Maptask corpus. Each vocable has its onset and ending time recorded, accurate to the nearest 1/100th of a second. This painstaking attention to timing permits the examination of the interleaving and overlapping of spoken contributions in greater detail than has previously been customary.

Section 2 of this paper provides further information on the Maptask corpus. Section 3 outlines the talk-division procedure itself, labelled TST1 (Talk Slicing Technique, version 1). Section 4 illustrates this technique by contrasting some examples of its output with more conventionally transcribed extracts; while section 5 presents a number of findings showing that there are many significant associations between quantitative indicators derived from applying the TST1 procedure and other variables that characterize the participants in the dialogues and/or their sociolinguistic context. In section 6 the implications of this research are discussed.

2. The Maptask corpus

The Edinburgh Maptask corpus (Anderson et al., 1991) consists of a total of 128 two-person dialogues. These were produced by 64 undergraduates from the University of Glasgow, 32 females and 32 males. Each dyad was given a task, the "Map Task" (Brown et al., 1984). In this task both participants have a sketch-map with several landmarks on it. They cannot see each other's maps. One participant, the instruction giver, has a path printed on his or her map, and is required to guide the other participant, the instruction follower, in drawing that path on the follower's map, which has no path marked on it. These two maps are similar but not identical, and the participants are told this fact. The number of (paired) maps used was 16 and in each pair the number of mismatching features between giver's and follower's map was equal. Each participant took part in four conversations, twice as an instruction giver (with both a familiar and an unfamiliar partner) and twice as an instruction follower (with both a familiar and an unfamiliar partner). In half of the conversations the speakers could make eye contact while in the other half a screen prevented them from seeing each other.

A noteworthy feature of this task is that it provides a quantifiable measure of communicative success. This is realized as a deviation score, which is computed as the area, in square centimetres, between the route as shown on the instruction giver's map and the path drawn on the instruction follower's map. Thus higher values indicate lesser success in the task. Having an outcome measure, such as this one, made it possible for this study to examine associations between patterns of spoken interchanges and communicative success.

The compilers of the Maptask corpus monitored a number of variables describing attributes of the speakers and/or interaction. Of these a subset of variables appeared suited to the purpose of the present study, which is to test whether attributes derived from the TST1 segmentation procedure show significant associations with variables characterizing the speakers or the interaction, including its outcome. The variables selected as suitable for examination are listed in Table 1, below.

Table 1. Maptask Variables Selected for Examination.

Name	**Description**	**Range**
Eye-contact	Whether participants could see each other's faces	0, 1
Familiarity	Whether participants were already acquainted	0, 1
Genders	Sexes of the participants	F, M

Roles	Role in the task, giver or follower	F, G
Outcome	Path deviation score (higher values indicating worse results)	4 .. 227

The ages of the participants were also recorded, but these fell in a narrow range, with more than 89% being aged 18-22, so this variable was not selected for further analysis. Details of the relationships between the above variables and measures derived from the segmentation technique can be found in Section 5.

3. A talk-slicing technique

A non-standard talk-division format was generated by an algorithm that simply sorted each vocable into temporal order, according to its onset time, and processed all the vocables sequentially, printing them out according to the following rule:

- if the current vocable was produced by the same speaker as the last one, append it to the current line;
- else print the current vocable on a new line preceded by a time stamp, in seconds, and speaker label.

Thus the arrangement of contributions by speaker is not imposed by preconceptions about well-behaved spoken interaction but emerges from the empirical data. This offers an *etic* perspective on the segmentation of spoken discourse, in contrast to the *emic* perspective of traditional approaches.

In the rest of this article we use the term *talkspan* to refer to a single segment of vocables produced by applying this algorithm, to avoid any misunderstanding that might result from the linguistic or interpretive connotations associated with pre-existing terms such as "turn" or "utterance".

Provided that each vocable is associated with an accurate onset time, this technique is extremely simple to implement. For the present investigation this procedure was implemented by a program in the Python language. This is referred to as TST1 (Talk Slicing Technique, version 1) in what follows.

The TST1 program also offers the option of displaying the time interval in seconds from the ending of the last vocable of the previous speaker to the start of the first vocable of the current speaker. If this number is negative, it indicates an overlap of voices (see, for example, Table 2).

4. Talk slicing on the page

The visual effect of this process can be illustrated by the extract below, from Maptask dialogue q8ec4. Table 2 gives the dialogue first as segmented in a conventional manner (2a), followed by the segmentation produced by the TST1 system (2b). The conventional utterance division is that produced on the demonstration page of the Maptask website when the option "plain" is selected, i.e. a using display mode not based on division into "moves", "games" or "transactions" (for details of these theoretical constructs, see Carletta et al., 1997).

In the conventional transcript **bold face** identifies overlapping speech. In the transcript produced by TST1, the interval coding is as follows:

> **Bold**: "Overlap" or negative inter-speaker interval (i.e. one speaker starts before the other has finished);
> Standard font: "Normal" interval between speakers of 0 to 0.999 seconds;
> *Italic*: "Long" inter-speaker interval of 1 second or more.

Table 2. Extract from Maptask dialogue q8nc4.

(2a) Extract from q8ec4, conventional utterance division. (Overlapping speech shown in bold face.)

GIVER: right , see the bottom line of the lake right ?
FOLLOWER: right .
GIVER: halfway along that .
FOLLOWER: halfway along ... okay .
GIVER: right ?
FOLLOWER: right .
GIVER: and then , then , straight down .
FOLLOWER: straight down .
GIVER: about er ... maybe five centimetres
FOLLOWER: right i've got some allotments here , so does it go ... to ...
your right again ... does it **go right** ?
GIVER: **it sort** of curves round .
FOLLOWER: curves round okay ... right .
GIVER: and then ... then .
FOLLOWER: along ... **to your right** ?
GIVER: **it's like an "s"** shape right but n-- not like a steep "s" shape
like a sort of ... er ... sort of ... stretched out "s" shape .
FOLLOWER: **right** .
GIVER: **a** wee bit ...

(2b) Extract from q8ec4, segmented by TST1, with timestamps in column 1 and inter-speaker intervals in column 2. (Overlaps shown in bold face with yellow background; gaps of under 1 second in normal font; gaps of 1 second or more in italics with cyan background.)

127.58 1.49 G: right see the bottom line of the lake right
130.43 0.51 F: right

132.50 1.58 G: halfway along that
133.67 0.22 F: halfway along okay
135.58 0.43 G: right
136.18 0.33 F: right
137.05 0.69 G: and then then straight down
141.04 0.37 F: straight down
142.27 0.26 G: about er maybe five centimetres
150.99 1.73 F: right i've got some allotments here so does it go to your right again does it go
156.67 -0.01 G: it
156.68 -0.10 F: right
156.77 -0.17 G: sort of curves round
158.35 0.22 F: curves round okay right
160.64 0.44 G: and then then
162.98 1.21 F: along to
163.96 -0.05 G: it's
164.01 -0.13 F: your right
164.14 -0.20 G: like an "s" shape right but n-- not like a steep "s" shape like a sort of er sort of stretched out "s" shape
172.08 0.40 F: right
172.64 -0.03 G: a wee bit

In this extract TST1 divides the conversation into more segments (21 lines) than the utterance-division of the human transcriber (17 lines). Hence the segments are shorter, on average, which is typical when manual transcript and TST1 output are compared.

Two noticeable features can be observed here. Firstly, the TST1 layout tends to include more stretches of "vertical" talk. This is reflected in the more frequent alternation of speakers and thus shorter segments in the second extract (2b) when compared to the first (2a). In the middle of the interaction, for instance, the first version represents "it sort of curves round" as a single line by the Giver whereas the in second version this is broken into two segments with the Follower's single-word contribution "right" in between.

Secondly, the layout produced by TST1 highlights the phenomenon of echoing, where words and short phrases are passed from one speaker to another. An example of this can be found early in extract (2b) where the Follower repeats almost verbatim at time 133.67 seconds what the Giver has just said at time 132.50 seconds. The next pair of lines (at 135.58 and 136.18 seconds) constitute an exact repetition. Another example can be found in the pair at 156.77 and 158.35 seconds, where the Giver's "sort of curves round" is partly replicated by the Follower's "curves round okay right".

Impressionistically, it can be said that the TST1 format emphasizes the joint production of talk rather more than is conventional and foregrounds the frequent short spans of one or two words that are often omitted in non-specialist transcription, e.g. journalistic reporting. Some of these might be treated as *backchannel* contributions by linguists (Yngve, 1970). If such vocal signals are noted in conventional transcriptions, they tend to be tacked onto the nearest content-bearing utterance, so their temporal placement is often imprecise.

Of course the two versions are not dramatically different, nor would we wish them to be. Thus there are plenty of examples where the two methods divide the stream of speech at the same points, as shown in the first few lines of both versions up to "about er maybe five centimetres" in Table 2. Nevertheless, it can be said that, relatively speaking, TST1 gives prominence to short contributions of one, two or three words. These are, so to speak, given equal status with the longer, more linguistically salient utterances.

Whether this "upgrading" of short contributions which may be paralinguistic affirmations, interjections or incomplete attempts at content-bearing utterances makes the dynamics of the interaction stand out more clearly on the page is for the human reader to judge. Whether it offers an analytical payoff is the subject of section 5.

A further example, illustrating a section of conversation with a greater incidence of simultaneous speech, is shown in Table 3.

Table 3. A section of dialogue q4nc2 illustrating overlaps.

(3a) Conversation conventionally segmented, with overlapping speech in bold face.
GIVER: do you have a , can i ... can you , do you do you have a field station marked **on the left-hand side** ?
FOLLOWER: **uh-huh , do i** go down that **far** ?
GIVER: **go** no , go down ... ... halfway between the ... the .
FOLLOWER: the **diamond mine ... and the field station .**
GIVER: **the diamond mine and the field station .**
FOLLOWER: okay .
GIVER: and then turn ... ... to ... the right .
FOLLOWER: the right ... okay , how far ?
(3b) Conversation divided by TST1 with timestamps and inter-speaker intervals. Overlaps are shown in bold face with yellow background; gaps longer than 1 second are in italics with cyan background; "normal" intervals are in standard font.
30.45 0.16 G: do you have a can i can you do you do you have a field station marked on

35.45	**-0.08**	**F:**	**uh-huh**
35.53	**-0.32**	**G:**	**the left-hand side**
36.34	**-0.17**	**F:**	**do i go down that far**
37.18	**-0.27**	**G:**	**go no go down halfway between the the**
41.94	*1.20*	*F:*	*the diamond*
42.27	**-0.30**	**G:**	**the diamond**
42.57	**-0.15**	**F:**	**mine**
42.73	**-0.32**	**G:**	**mine and**
43.08	**-0.12**	**F:**	**and**
43.21	**-0.03**	**G:**	**the**
43.24	**-0.03**	**F:**	**the**
43.27	**-0.03**	**G:**	**field**
43.30	**-0.24**	**F:**	**field**
43.53	**-0.01**	**G:**	**station**
43.54	**-0.52**	**F:**	**station okay**
44.87	0.30	G:	and then turn to the right
48.61	0.15	F:	the right okay how far

In Table 3, we see an even greater difference in the number of segments between the two versions than in Table 2. In extract (3b) the TST1-derived layout represents the interaction in eighteen lines whereas the traditional transcription format (3a) contains only nine lines. Further, the occurrence of echoing is even more prominent, as evidenced from time 41.94 to 43.54 seconds of the second version (3b) where literal repetition of words such as "diamond", "mine", "and", "the", "field" and "station" occurs as the participants speak almost in chorus.

We argue that this representation gives more information about the temporal interleaving of the participants' speech than conventional layouts, without sacrificing the intelligibility of the linguistic content. Moreover, being a deterministic process it removes the need for consistency checking among transcribers as regards the beginnings and endings of utterances.

5. Relationships between variables derived from TST1 and other variables of interest

In this section we examine some of the more striking relationships between length of *talkspans* (i.e. segments of speech as defined by the TST1 procedure) and the other variables of interest (listed above in Table 1).

We first present some basic statistics arising from the segmentation of the corpus in section 5.1, then in section 5.2 we examine how talkspan length relates to selected contextual variables. Section 5.3 considers the relationships between inter-speaker intervals and the contextual variables, and pays particular attention to overlapping speech, i.e. cases where the inter-speaker interval is negative. In section 5.4 we consider the task-outcome measure, path deviation, and look at its association with talkspan length and with inter-speaker interval. Finally, section 5.5 presents an analysis of lexical items that are unusually common in the initial positions of talkspans, drawing attention to certain associations between the positioning of such vocabulary items and their functions.

5.1 Basic statistics of the corpus

The 128 dialogues between them consist of 26194 talkspans containing a total of 153,780 word tokens. For reference, Table 4 shows the numerical characteristics of the dialogues in the Maptask corpus in terms of talkspans (as defined by TST1), word tokens and deviation scores. Mean and median values are shown, along with standard deviations.

Table 4. Basic Statistics of Maptask Variables.

Variable	**Mean**	**Median**	**Standard deviation**
Talkspans per dialogue	204.64	169.50	122.68
Word tokens per dialogue	1201.41	1014.00	648.00
Talkspan length in word tokens	5.87	3.00	7.36
Talkspan length in seconds	1.99	1.01	2.78
Inter-speaker interval in seconds	0.28	0.09	0.78
Deviation score	71.83	56.00	49.17

None of these variables is symmetrically distributed, all being skewed to the right, so that the means are greater than the medians in each case. This asymmetry implies the existence of a small number of atypically high values (as is often found with linguistic phenomena) and indicates that only non-parametric tests are appropriate in comparisons involving these variables.

5.2 Correlatives of contribution length

This section presents findings concerning the associations between talkspan length as defined by TST1 and the four main contextual variables -- eye-contact, familiarity, gender and speaker role.

5.2.1 Instruction followers make shorter contributions than givers. Of the four main contextual features examined the variable with the strongest association with length of speakers' talkspans is speaker role. Figure 1 shows histograms of the lengths of talkspans of givers and followers separately. (These lengths, in seconds, have been subjected to a logarithmic transformation for display purposes in Figure 1, as the original distribution is very strongly skewed to the right.) It is obvious from the histograms in Figure 1 that the talkspan lengths of givers and followers differ dramatically. In particular, followers produce far fewer long contributions. Some difference is to be expected since they have different communicative goals; perhaps more worth noting is that both distributions are bimodal. This suggests that each distribution is a composite of two different types of contribution.

[Figure 1 about here.]

The density estimates plotted underneath the histograms emphasize the twin-peaked nature of these distributions. These plots also show that these timings are definitely not normally distributed: there is a clear preponderance of short talkspans, for both parties, though significantly more so for followers than givers.

As these variables are not normally distributed a non-parametric test, theWilcoxon rank-sum test (also known as the Wilcoxon-Mann-Whitney test), was performed to compare talkspan lengths between givers and followers. The results for both time in seconds and length in word tokens were very highly significant: for times, Wilcoxon equivalent z-score = -38.35, $p < 0.0005$; for tokens, Wilcoxon equivalent z-score = -49.32, $p < 0.0005$.

Another way of looking at this effect is by tallying the frequency of single-word talkspans between the two speaker roles, as presented in Table 5.

Table 5. Cross-tabulation of 1-word Talkspans by Speaker Role.

Contribution length:	1-word talkspan	multi-word talkspan
Instruction Giver	3767	9360
Instruction Follower	6554	6513

Over half (50.2%) the followers' talkspans consisted of just a single word, whereas only 28.7% of the givers' talkspans were single words. Thus the odds in favour of a follower's contribution being a 1-word utterance were 2.5 times as great as the odds for a giver's. Statistically, this difference is very highly significant indeed (Chi-squared = 1263.09, df = 1, likelihood ratio = 1275.28, $p < 10^{-278}$).

5.2.2 Familiar pairs produce more exchanges than unfamiliar pairs, but they are shorter. There is a clear tendency for pairs who are already acquainted to produce a larger number of talkspans per dialogue than those who are unfamiliar. The median number of talkspans for the familiar pairs was 191 while for the unfamiliar pairs it was 157.5. As these numbers were not normally distributed a non-parametric test of significance was performed, showing a highly significant difference: Wilcoxon test, equivalent z-score = -2.70, $p = 0.007$.

However the length of talkspans in seconds for familiar pairs was very significantly shorter: Wilcoxon equivalent z-score = -2.93, $p = 0.003$; although in terms of number of tokens per talkspan, this difference did not reach significance.

5.2.3 Pairs with eye-contact allowed produce longer talkspans than pairs without. Talkspan lengths both in terms of time and number of tokens are longer in the dyads with eye-contact permitted than those without. With eye-contact the median number of tokens per talkspan is 3, without eye-contact it is 2. The median length of talkspan is 1.12 seconds with eye-contact, and 0.92 without eye-contact. These differences are highly statistically significant. (For number of tokens, Wilcoxon equivalent z-score = -3.33, $p = 0.001$; for time in seconds, equivalent z-score = -8.30, $p < 0.0005$.)

Again, this effect can also be seen by tabulating the frequency of single-word talkspans: the rate is significantly lower in the group with eye-contact, as shown in Table 6.

Table 6. Cross-tabulation of 1-word Talkspans by Eye-Contact.

Contribution length:	1-word talkspan	multi-word talkspan
Eye-contact allowed	4535	7242
Eye-contact absent	5786	8631

The rate of single-word talkspans is 38.5% in the pairs with eye-contact and 40.1% in those without. Statistically, this difference is highly significant: Chi-squared = 7.18, df = 1, Likelihood ratio = 7.18, $p = 0.007$.

A possible reason for the higher ratio of single-word talkspans in pairs without eye-contact is that some interaction-management functions that can be managed by gaze if eye-contact is present have to be managed by (short) verbal/vocal contributions if it is absent, thus increasing the total of short contributions. As noted by Doherty-Sneddon et al. (1997: 113): "speakers attempt to confirm their listeners' understanding or agreement more often when they cannot see one another."

5.2.4 Males produce more tokens per talkspan than females (in about the same time). Overall, male speakers produce longer talkspans in terms of the number of tokens, though not in terms of the length of their talkspans in seconds. The median number of tokens per talkspan is 3 for males and 2 for females. The median length in seconds is 1.03 for males and 1.00 for females. The first of these differences is very highly significant (Wilcoxon test, equivalent z-score = -4.12, $p < 0.0005$) but the second fails to reach significance (equivalent z-score = -0.237, $p = 0.812$).

It seems unlikely that males actually speak faster than females so this result might be explained by males speaking with fewer hesitation pauses, though this has not been verified.

Female speakers also produce a higher proportion of single-word contributions than males (40.6% versus 38.3%), as shown in Table 7. This association is very highly significant (Chi-squared = 14.55, df = 1, Likelihood ratio = 14.55, p < 0.0005).

Table 7. Cross-Tabulation of 1-word Talkspans by Speaker's Gender.

Contribution length:	1-word talkspan	multi-word talkspan
Female speaker	5186	7593
Male speaker	5135	8280

A plausible explanation for this difference is that women tend to give more active-listening signals such as "yeah" than men. This is supported by several American studies which report a higher rate of backchannel signals in the speech of women than men (Eckert & McConnell-Ginet, 2003: 110).

5.3 Gaps and overlaps

As well as the lengths of talkspans, we examined some relationships with the lengths of the intervals between talkspans (which, when negative, signify overlapping talk).

5.3.1 Overlaps are commonplace. In the corpus as a whole transitions where one speaker begins before the previous speaker has finished comprise 42 percent of all transitions. Table 8 shows the absolute and relative frequencies of overlaps (transitions where the inter-speaker interval is negative), "normal" intervals (non-negative but less than 1 second) and long gaps (where the interval equals or exceeds 1 second).

Table 8. Frequencies of three categories of inter-speaker interval.

Transition Type	**Frequency**	**Percentage**
Overlap	11011	42.04
"Normal"	12571	47.99
Long gap	2612	9.97

The proportion of overlapping speech in the present study is broadly comparable to the 45 percent found in the Japanese version of the Maptask corpus, as reported by Iwa et al. (1998).

5.3.2 Average gapsize is longer when eye-contact is allowed. Figure 2 depicts the distributions of mean gapsize (inter-speaker interval) for all 128 conversations, 64 with eye-contact allowed and 64 without eye-contact. It will be seen that when eye-contact is allowed the average size of inter-speaker intervals tends to be longer.

[Figure 2 about here.].

This difference is highly significant: Wilcoxon equivalent z-score = -2.93, n=128, p = 0.003. This finding is broadly compatible with the results of Bull and Aylett (1998), although they defined utterances differently.

This effect is also reflected in the relative frequencies of the three types of inter-speaker interval identified above (overlap, normal and long), as shown in Table 9, below.

Table 9. Frequencies of 3 types of inter-speaker interval with and without eye-contact.

Transition Type	**Eye-contact allowed**	**Eye-contact absent**
Overlap	4638	6373
"Normal"	5699	6872
Long gap	1440	1172

Overlaps are relatively more frequent without eye-contact, while long gaps are more frequent with eye-contact. This association is statistically very highly significant (Chi-squared = 145.74, df = 2, Likelihood ratio = 145.16, $p < 10^{-31}$).

5.3.3 Instruction Givers overlap more often than Followers. While there is no significant association between type of inter-speaker interval and familiarity of the participants or speaker's gender, there is an association with speaker role. Instruction givers are responsible for relatively more overlapping transitions and relatively fewer long gaps than instruction followers, as shown in Table 10. In other words, instruction givers are more likely to start speaking before the other party has finished and less likely to pause before replying than instruction followers.

Table 10. Relationship between inter-speaker interval and speaker role.

Transition Type	**Instruction Giver**	**Instruction Follower**
Overlap	5692	5319
"Normal"	6343	6228
Long gap	1092	1520

This association is statistically very highly significant (Chi-squared = 83.68, df = 2, Likelihood ratio = 84.00, p < 0.0005).

5.4 Associations with the outcome variable

This subsection examines the extent to which quantitative features derived from segmentation by TST1 are predictive of the outcome measure, path deviation score.

5.4.1 Longer talkspans by followers are associated with worse performance scores. A number of derived variables relating to talkspan lengths were computed and correlated with the outcome measure, path deviation score, for each of the 128 dialogues. These were: total number of talkspans, total time of talk, mean length of talkspan (in tokens and seconds), mean lengths of talkspans for giver and follower separately (in tokens and seconds), and the rate of 1-word talkspans (overall, for giver and for follower). Since the path deviation score was not normally distributed (Kolmogorov-Smirnov statistic = 1.695, p = 0.006) nor were several of the other variables, a non-parametric correlation, Spearman's rho, was used. Only one of these eleven variables was found to correlate significantly with outcome score, namely the mean length of the follower's talkspans in seconds (rho = 0.264, p = 0.003). This correlation is positive, meaning that longer talkspans by the follower (not the giver) are associated with worse results in the task. This may be because long contributions by the follower are a kind of "losing-the-plot" signal. In other words, when the interaction is going smoothly, the follower won't have anything very complicated to express, resulting in shorter talkspans.

5.4.2 In the worst-performing quartile, givers produce fewer 1-word talkspans. If the 128 dialogues are partitioned into those with a deviation score of over 100 (the upper, i.e. worse-performing, quartile) and the rest, the givers in the worse-performing quartile have a significantly lower mean rate of 1-word talkspans, 11.12% versus 13.23%, than the rest (t-test: t = 2.38, df = 126, p = 0.02). This may indicate a lower rate of monitoring the effect of their instructions by the givers in the less-successful group.

5.4.3 Overlaps can be beneficial. For the 128 dialogues taken as a whole, there are significant negative correlations between the total number of overlaps and the deviation score (Spearman's rho = -0.215, n=128, p = 0.015) as well as between the proportion of overlaps and the deviation score (Spearman's rho = -0.280, n=128, p = 0.001). The negative sign implies that more frequent overlapping speech is associated with lower deviation scores, i.e. better task performance. This is another blow for the view that overlapping talk is in some sense pathological. In the words of Tannen (1994: 60): "many instances of overlap are supportive rather than obstructive."

Figure 3 shows that this relationship is non-linear, despite generating a significant rank correlation coefficient. This L-shaped distribution might be better characterized as a logical NAND: it is possible to have high deviation scores or a large number of overlaps (or neither) but not both. The reason why the upper-right quadrant of the graph is, in effect, forbidden, is not clear.

[Figure 3 about here.]

5.5 Vocabulary and positioning

If TST1 were merely chopping up the speech stream at random, one would expect the words next to boundary-points to be a random selection of the words in the corpus, with the same probability of occurring near a boundary as anywhere else. On the other hand, if the technique is slicing the speech-stream at socially or linguistically meaningful junctures, one would expect position within a talkspan to have an effect on vocabulary. One way of investigating this issue is to look at the initial words of the 26194 talkspans.

5.5.1 Lexical items found in initial positions do not constitute a random subset of the overall vocabulary. Table 11 lists the thirteen words which occur most frequently at the beginning of a talkspan. Between them they account for over 53 percent of initial words in this corpus. The table gives their overall frequencies in the corpus, their frequencies as initial words, and their frequencies as initial (and only) words in single-word talkspans. The entries are ranked according to the middle column, frequency in initial position of a talkspan.

Table 11. Most common words in initial position.

	Whole Corpus		Initial Position		Solo Word	
Word	**Frequency**	**Percent**	**Frequency**	**Percent**	**Frequency**	**Percent**
right	6932	4.51	3838	14.65	1735	16.81
okay	2458	1.60	1467	5.60	889	8.61
yeah	1710	1.11	1342	5.12	783	7.59
uh-huh	1460	0.95	1286	4.91	957	9.27
and	3464	2.25	1085	4.14	103	1.00
no	1315	0.86	923	3.52	390	3.78
mmhmm	903	0.59	870	3.32	763	7.39
the	12870	8.37	751	2.87	145	1.40
so	1647	1.07	648	2.47	140	1.36
well	1090	0.71	505	1.93	85	0.82
to	4447	2.89	464	1.77	112	1.09
you	4657	3.03	459	1.75	90	0.87
oh	566	0.37	403	1.54	95	0.92
Total =	153780		26194		10321	

It can easily be seen that the relative frequencies of the words in initial positions do not distribute themselves according to the relative frequencies in the corpus as a whole. For example, the token "uh-huh" accounts for less than one percent of the tokens in the corpus, yet comprises almost five percent of the tokens found in initial positions. In the other direction, the commonest word in the corpus, "the", accounts for over eight percent of the word tokens, but only 2.87 percent of those in initial position.

With such obvious discrepancies, a statistical test of significance is hardly necessary, but for completeness a Chi-squared test was conducted on the 13-by-2 matrix containing frequency counts for these words in initial and non-initial positions. As expected this yielded a very highly significant result, Chi-squared = 15547.69, df =

12, log-likelihood = 16769.39, $p < 10^{-300}$. It can be stated with extreme confidence that words in initial positions of talkspans are not a random selection of the vocabulary of the corpus as a whole.

5.5.2 Some words in initial positions are very much more likely to stand alone than others. In addition, a further test was performed on the 13-by-2 matrix of counts of these words (all initial) which did or did not have subsequent words following them in the talkspans which they initiated, as shown in Table 12.

Table 12. Initial words with and without following words.

Word	**Solo**	**Initial but not Solo (i.e. followed by more words in the same talkspan)**
right	1735	2103
okay	889	578
yeah	783	559
uh-huh	957	329
and	103	982
no	390	533
mmhmm	763	107
the	145	606
so	140	508
well	85	420
to	112	352
you	90	369
oh	95	308

Here again the result was very highly significant (Chi-squared = 2670.57, df = 12, log-likelihood = 2908.08, $p < 10^{-300}$). This shows that some initial words, such as "mmhmm", tend to stand alone while others, such as "well", tend to inaugurate a contribution of more than a single word.

5.5.3 Words with different functions behave differently with respect to positioning. The results in 5.5.1 and 5.5.2 alone would be sufficient to show that the division of speech produced by TST1 is very far from arbitrary, but that is merely a first step. It is perhaps more interesting to take a look at these words individually in respect of their positioning and see what implications emerge about vocabulary choice by the participants. Figure 4 displays each of these word tokens as a point on a two-dimensional graph in which the horizontal axis measures the degree to which that word is preferentially found in initial positions (as compared to its occurrence rate in the whole corpus) and the vertical axis is the degree to which the word is preferentially found as a single-word contribution (as compared to its occurrence rate among words in initial positions).[1] These words fall into three main groupings, as shown in Table 13.

Table 13. Three kinds of initializing words.

Continuers	High-frequency words that appear less often in initial positions than expected according to their overall frequency	the, to, you
Initializers	Words relatively frequent in initial positions that usually initiate a multi-word talkspan	and, oh, no, right, so, well
Solitaries	Words very frequent in initial positions that frequently constitute a single-word talkspan in themselves	mmhmm, okay, uh-huh, yeah

[Figure 4 about here.]

As shown in Figure 4, the word that is by far the commonest in initial position, "right", also has a high rate of appearance as a 1-word talkspan, but it is less likely to appear alone than some words that are less frequent as initials. In this respect it is far outdone by "mmhmm" and "uh-huh"[2], which are transcription conventions for the kind of paralinguistic vocal signals of assent or confirmation known as backchannel communications (Yngve, 1970; Schegloff, 1982; Tottie, 1989).

Thus differences of function and meaning manifest themselves as differences in positioning within talkspans. From a functional point of view, it could be said that the three-way grouping in Table 13 divides a continuum into three sections: solitaries such as "uh-huh" and "yeah" have primarily pragmatic meaning at one end; continuers such as "to" and "you" carry primarily propositional meaning at the other end; while in the middle initializers such as "so" and "well" are notoriously polysemous and can slide between propositional and pragmatic functions.

5.5.4 Are some overlaps more helpful than others? It is possible to use this vocabulary of high-frequency talkspan-initial words for a preliminary investigation of whether there may be different kinds of overlapping speech, some of which are more helpful to the goal of the interaction than others. For this purpose we term any talkspan which overlaps with the previous speaker's (i.e. begins with a negative inter-speaker interval) an *interpolation*. To investigate this matter, the conversations were divided into three subgroups on the basis of their path deviation scores: those with scores in the lower quartile were labelled “good”; those with scores in the upper quartile were labelled “poor”; while the rest, those in the middle range, were ignored in order to sharpen the contrast. Then a tabulation was made of the 13 most common talkspan-initializing words, listed above in Table 11, according to how often they initiated an interpolation in "good" dialogues and in "poor" dialogues. (Talkspans with positive inter-speaker intervals, i.e. gaps between speakers, were ignored.) The result is shown in Table 14, which is ordered by the final column, the ratio of occurrences in dialogues with good scores divided by occurrences in dialogues with poor scores.

Table 14. Initializers of overlapping talk in dialogues with good versus poor scores.

Initial word of talkspan	Frequency as an **overlapping** initializer in dialogues with poor scores	Frequency as an **overlapping** initializer in dialogues with good scores	Ratio (Good / Poor)
mmhmm	22	55	2.50
so	32	68	2.13
to	39	81	2.08
okay	62	109	1.76
and	48	82	1.71
uh-huh	71	121	1.70
the	89	149	1.67
right	176	266	1.51
you	46	59	1.28
yeah	78	97	1.24
oh	24	29	1.21
no	43	49	1.14
well	25	22	0.88
Sums:	755	1187	(norm) 1.57

The overall ratio is greater than 1, which confirms the point already noted that overlapping talkspans are more common in dyads which perform well on the task, but the table shows some interesting differences among the vocabulary items that initiate such interpolations. The two initializers that may be said to be most supportive are "mmhmm" and "so". The former is no surprise since it represents an affirmative feedback signal, though the fact that "mmhmm" appears more supportive than "uh-huh" may be of interest. The high position of "so" is not so predictable. This word is multi-functional, but one of its major functions is as a discourse particle "to establish an inferential connection" (Blakemore, 1988: 193). It is tempting to conjecture that interpolated speech is most helpful when it signals either assent or inferential connection.

On the other hand, the two least supportive initializers are "no" (which explicitly marks dissent) and "well", often used as a discourse particle marking a reply that is "not fully consonant" (Schiffrin, 1987: 100) with what has gone before or prefacing a dispreferred response (Lam, 2006). The third item from bottom is "oh" which typically used at the moment where the hearer suddenly realizes something.

Thus we find relatively high rates of interpolations that signify agreement or understanding in dyads which do well on the task of the interaction, and high rates of interpolations signifying dissent or lack of understanding in those that perform poorly. Here we have the beginnings of an approach to distinguishing helpful from harmful interpolations.

6. Discussion

This investigation is primarily designed to test a technique, so we do not claim that any of the findings reported in Section 5 is momentous in itself. Taken together, however, they tend to indicate that the technique has merits both as a representational and an analytic tool. Some implications of this are discussed in subsection 6.2. First, however, in 6.1, a particular issue which is highlighted by the segmentation technique, namely the frequency of occasions when one participant starts to speak before the other has finished, is worth remarking upon.

6.1 Talking together

Our results suggest that simultaneous speech is quite common and, arguably, beneficial. Yet humans are not very well able to speak and listen at the same time, so the alternation of speaker and listener role in conversation has been widely viewed as a linguistic universal. Pioneering research by Jaffé and Feldstein (1970) showed that the percentage of time that conversational participants were talking simultaneously was less than 5%, and normally much less. "Asynchrony is thus the rule and synchrony (simultaneous speech) the violation in conversational vocal patterns" (Jaffé & Feldstein, 1970: 11). Furthermore one of the "grossly apparent facts" about conversation listed by Sacks et al. (1974: 700) is: "Transitions (from one turn to a next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions".

It is natural therefore to assume that simultaneous speech is abnormal and indeed probably dysfunctional. Nevertheless several researchers have reported that simultaneous speech is far from uncommon (e.g. Iwa et al., 1998) and in many cases supportive of the goals of the interaction (e.g. Tannen, 1994). The key to resolving this apparent contradiction undoubtedly lies in the length and nature of the overlapping utterances. The present investigation may shed some light on this issue by providing one means of distinguishing helpful from unhelpful interpolations (5.5.4, above).

6.2 Towards a transcription tool

Our primary objective in this study has been to introduce a simple yet automatic method of dividing transcribed speech into segments on a strictly chronological basis, and to make an initial test of the utility of that method (TST1) by applying it to a publicly available spoken dataset, the Maptask corpus.

Digital recording, and analysis, of speech is becoming more important in a large number of fields, and its importance is likely to grow. Anyone who has been involved in transcribing recorded speech will know that dealing with simultaneous utterances by more than one speaker is one of the most problematic and time-consuming aspects of the process. From a representational point of view, the advantage of TST1 is that it is automatic. It sidesteps the normative considerations inherent in the concept of turn-taking, and merely sorts all the vocables of a conversation into sequence according to their onset-time -- noting by a line-feed and a speaker-prefix when the current vocable has been produced by a different speaker from the last one.

Thus it frees the transcriber from the problem of deciding which speaker "has the floor" and the associated problem of how to deal with cases where a speaker other than the one who is deemed to "have the floor" says something.

Of course this means that the resultant layout does not directly correspond to some established notions of what constitutes turn-taking. Whether this is a serious loss or not depends on the analysts' purpose and cannot be decided generally. However, in practice, at least on one corpus in one language, the speaker-division patterns produced by TST1 do not look outlandish: they are intelligible as dialogue even if somewhat vertically "stretched" by the standards of a typical playscript.

The main question is whether this way of dividing speech into chunks offers any benefits to a researcher studying spoken interaction. We argue that the results in section 5 suffice to show that features derived from the division imposed by TST1 yield indicators that can serve as diagnostic or predictor variables for significant aspects of the interaction. To recapitulate our main findings:

- Speaker role is very strongly associated with length of talkspan and with the likelihood of initiating speech before the other speaker has finished;
- Familiar participants produce more talkspans than unfamiliar ones;
- Presence or absence of eye-contact is reflected in average talkspan length and average inter-speaker interval;
- Males produce more word-tokens per talkspan than females;
- Longer talkspans are associated with worse task performance;
- More overlapping speech is associated with better task performance.

In addition, very clear association between position within a talkspan and vocabulary selection shows that TST1-defined boundaries are linguistically meaningful. For instance, vocables considered as backchannel signals (such as "mmhmm" and "uh-huh") show strong positional preferences, being very much more common as initial tokens than at other positions and also, given initial position, much more common as sole-word contributions. An analysis of tokens that initiate overlapping talkspans also throws some light on the question of whether interpolations are supportive of the goals of the interaction or not.

The chief disadvantage of this method of talk-division is that it requires accurate timings of each of the vocables uttered by all participants in a conversation. However, as automatic word-segmentation software becomes cheaper and more reliable, this problem should become less severe.

Its two great advantages are (1) it is simple; and (2) it is objective. Most existing methods of talk-division are complex: a turn or utterance is typically defined in terms of a mixture of syntactic, semantic, prosodic and functional characteristics -- requiring human expertise with the costs in terms of time and validation that the exercise of human judgement requires (cf. Ford & Thompson, 1996; Carletta et al., 1997; Bull & Aylett, 1998; Koiso et al., 1998; ten Bosch et al., 2004). TST1 by contrast requires human judgement only in the identification and timing of word boundaries, an uncontentious, though onerous, task. Once that is done the segmentation of the speech stream is automatic, thus approaching closer to the ideal of "letting the data speak for themselves". As the present study demonstrates, they have plenty to tell us.

Acknowledgements

This work was partly supported by Grant R149230035 of the UK Economic and Social Research Council as part of the DReSS Project sponsored by the National Centre for e-Social Science and partly by the School of Psychology, University of Nottingham. The authors would like to thank Professor Jean Carletta and Dr Amy Isard, of the Human Communication Research Centre of the University of Edinburgh, for answering questions about the Maptask data and kindly supplying the path deviation scores which are not currently included on the Maptask website. We would also like to thank Professor Claire O'Malley, of the University of Nottingham, for encouraging us to investigate this corpus in the first place and Dr Shaaron Ainsworth, also of the University of Nottingham, for helpful comments on an earlier draft.

Notes

1 Position on the horizontal dimension is computed as $2 \times O_j \times \ln(O_j/E_j)$ where O_j is observed frequency of token j and Ej is expected frequency. This is the component formula of the log-likelihood statistic, here used as an index. Expected frequencies for the horizontal dimension were calculated from word-frequencies in the whole corpus; expected frequencies in the vertical dimension were calculated from word-frequencies in initial position. Thus vertical height indicates the degree to which a token, given that it is already in initial position, is the sole token in a contribution.

2 In North American English the token "mmhmm" might be thought to indicate a positive sign, of affirmation or permission to continue speaking, while "uh-huh"

could be considered a negative signal, e.g. of disagreement or surprise. However, in Scottish English, the sound rendered here as "uh-huh" is most often another positive or agreeing signal. See also: Tottie (1989).

References

Anderson, A.H., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C. & Thompson, H. (1991). The HCRC Map Task corpus. *Language & Speech*, 34, 351-360. [http://www.hcrc.ed.ac.uk/maptask]

Atkinson, J.M. & Heritage, J. (1984) eds. *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Blakemore, D. (1988). ‘So’ as a constraint on relevance. In R.M. Kempson (ed.), *Mental representations: the interface between language and reality* (pp. 183-195). Cambridge: Cambridge University Press.

Boden, D. & Zimmerman, D.H.. (1991). *Talk and Social Structure*. Cambridge: Polity Press.

Brown, G., Anderson, A.H., Yule, G. & Shillcock, R.. (1984). *Teaching Talk. Cambridge*: Cambridge University Press.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, 32, 1439-1465.

Bull, M.C. & Aylett, M.P. (1998). An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. *Proceedings of ICSLP-98 Sydney, Australia (4).* 1175-1178.

Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G. & Anderson, A.H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1), 13-31.

Carter, R. (2004). *Language and Creativity*. London: Routledge.

Dahlmann, I. & Adolphs, S. (2007). Designing multi-modal corpora to support the study of spoken language. *Third International Conference on e-Social Science,* Ann Arbor, Michigan, 7-10 October 2007.

Doherty-Sneddon, G., Anderson, A.H., O'Malley, C.E., Langton, S., Garrod, S. & Bruce, V. (1997). Face-to-face and video-mediated communication: a comparison of dialogue structure and task performance. *J. Experimental Psychology: Applied*, 3(2), 105-125.

Eckert, P. & McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge: Cambridge University Press.

Edwards, J. & Lampert, M.D. (1993) eds. *Talking data: transcription and coding in discourse research*. Hillsdale, NJ: Erlbaum.

Ehlich, K. (1993). HIAT: a transcription ssytem for discourse data. In J.A. Edwards & M.D. Lampert (eds). *Talking data: transcription and coding in discourse research* (pp. 123-148). Hillsdale, NJ: Erlbaum.

Ford, C.E. & Thompson, S.A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E.A. Schegloff & S.A. Thompson (eds). *Interaction and Grammar* (pp. 134-184). Cambridge: Cambridge University Press.

Gumperz, J.J. & Berenz, N. (1993). Transcribing conversational exchange. In J.A. Edwards & M.D. Lampert (eds). *Talking data: transcription and coding in discourse research* (pp. 91-121). Hillsdale, NJ: Erlbaum.

Iwa, J., Enomoto, M., Ohya, K., Shimano, K. & Tutiya, S. (1998). A study of speech overlap in Map-task dialogues. SIG-SLUD-9802-3, 15-21.

Jaffe, J. & Feldstein, S. (1970). *Rhythms of Dialogue*. New York: Academic Press.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech*, 41 (3-4), 295-321.

Lam, P. (2006). Well *but that's the effect of it*: The use of *well* as a discourse particle in talk shows. *Sprache und Datenverarbeitung (International Journal for Language Data Processing),* 30(1), 99-108.

Leech, G., Myers, G. & Thomas, J. (1995) eds. *Spoken English on Computer: Transcription, Markup and Applications*. Harlow: Longman.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

MICASE (2007). MICASE Manual.
http://legacyweb.lsa.umich.edu/eli/micase/MICASE_MANUAL.pdf
[accessed 27/12/2007]

Ochs, E. (1979). Transcription as theory. In E. Ochs & B.B. Schieffelin (eds). *Developmental pragmatics* (pp. 43-72). New York: Academic Press.

O'Connell, D.C. & Kowal, S. (1994). Some current transcription systems for spoken discourse: a critical analysis. *Pragmatics*, 4, 81-107.

Payne, J. (1995). The COBUILD spoken corpus: transcription conventions. In G. Leech, G. Myers & J. Thomas (eds). *Spoken English on Computer: Transcription, Markup and Applications* (pp. 203-207). Harlow: Longman.

Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 54(4), 696-735.

Schegloff, E.A. (1982). Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In D. Tannen (ed.). *Analyzing Discourse: Text and Talk* (pp. 71-93). Washington DC: Georgetown University Press.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Schiffrin, D. (1994). *Approaches to Discourse*. Oxford: Blackwell.

Svartvik, J. (1990) ed. *The London-Lund corpus of spoken English*. Lund, Sweden: Lund University Press.

Svartvik, J. & Quirk, R. (1980) eds. *A corpus of English conversation*. Lund, Sweden: CWK Gleerup.

Tannen, D. (1994). *Gender and Discourse*. Oxford: Oxford University Press.

ten Bosch, L., Oostdijk, N. & de Ruiter, J.P. (2004). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. *Proc. 7th Int. Conf. on Text, Speech & Dialogue*. Brno, Sept. 2004.

Tottie, G. (1989). What does uh-(h)uh mean? In B. Odenstedt & Persson, G. (eds). *Instead of flowers. Papers in honour of Mats Ryden on the occasion of his sixtieth birthday* (pp. 269-281). Stockholm: Almqvist & Wiksell.

Yngve, V.H. (1970). On getting a word in edgewise. *Papers from the 6th Regional Meeting of the Chicago Linguistic Society, Chicago*: University of Chicago, 567-578.

FIGURES:

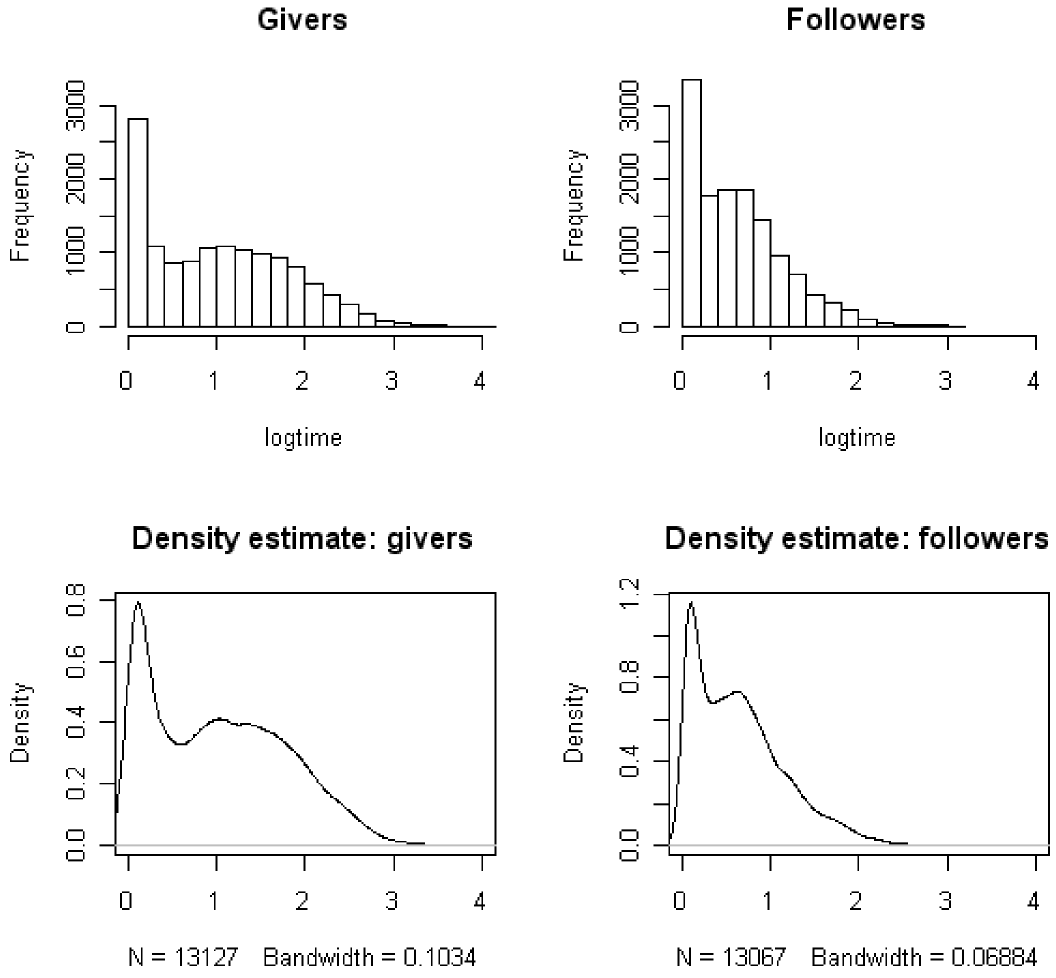

Figure 1. Histograms of logtime = ln(t+1) where t is length of talkspan in seconds.

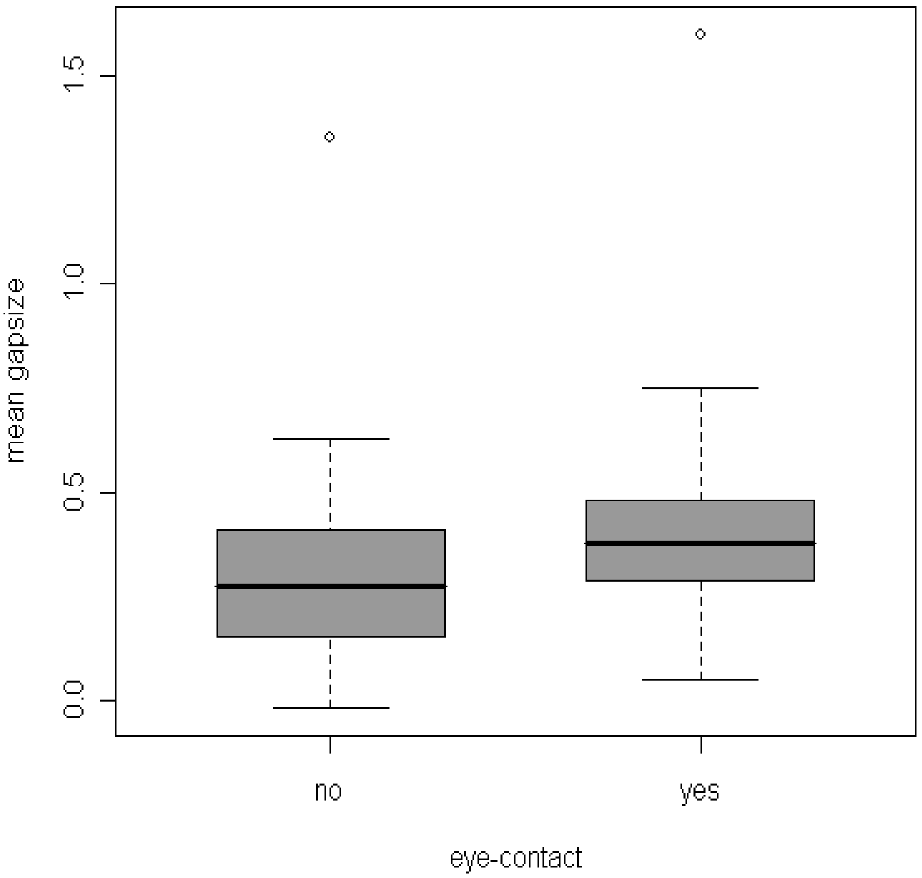

Figure 2. Boxplot of mean inter-speaker interval when eye-contact absent or present.

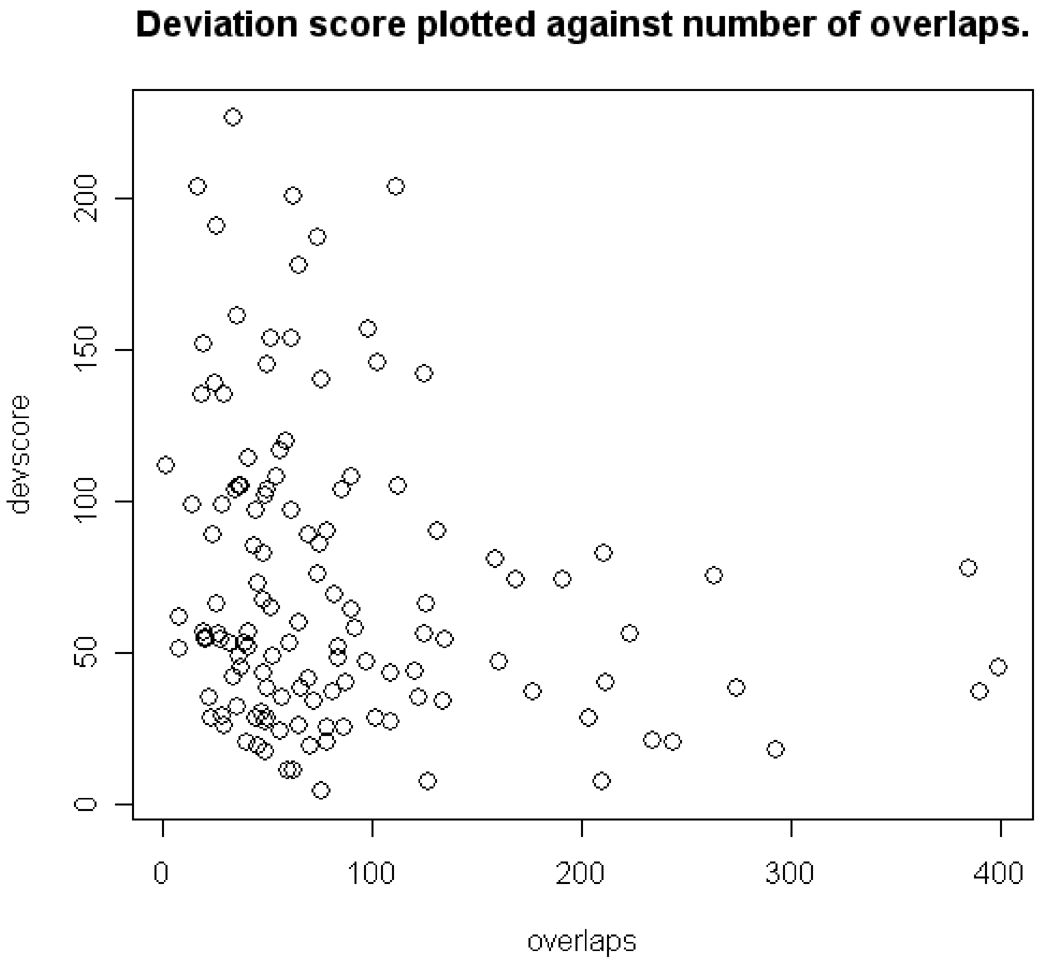

Figure 3. Relationship between number of overlaps and deviation score.

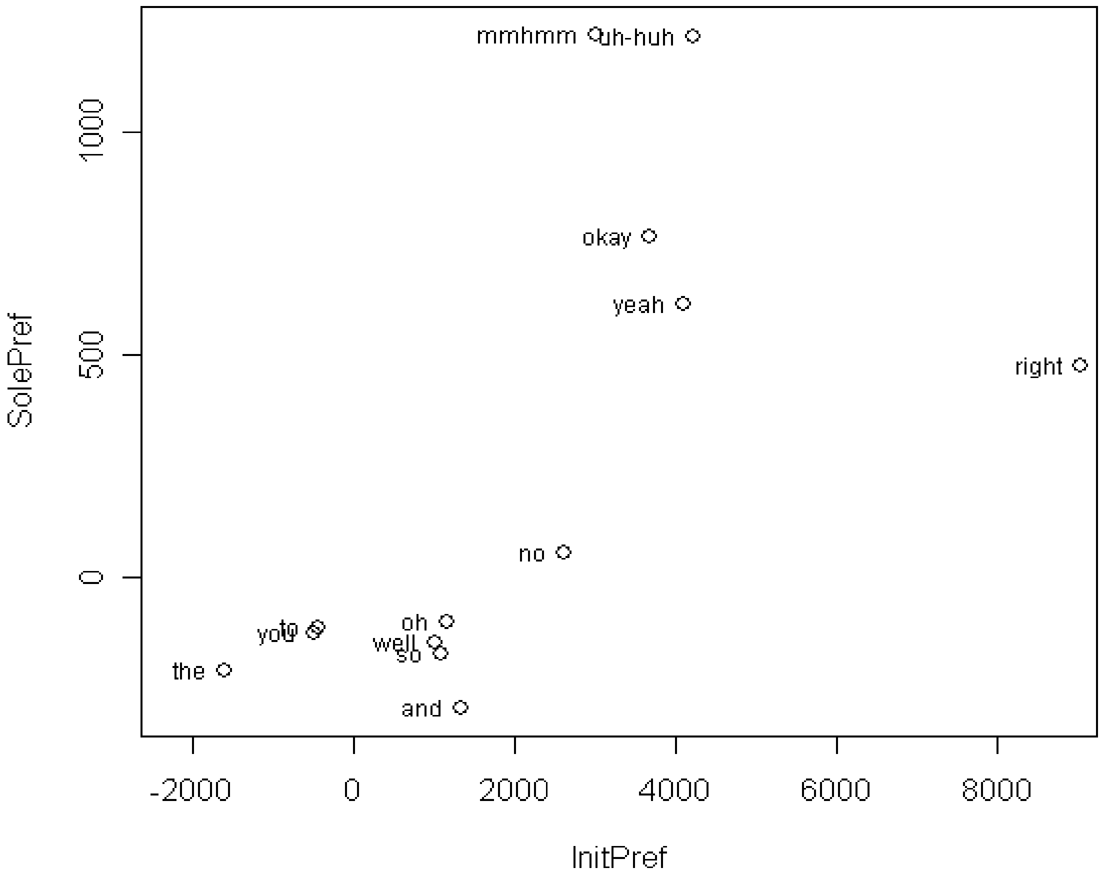

Figure 4. Plot of words in space of Solo versus Initial Preference.