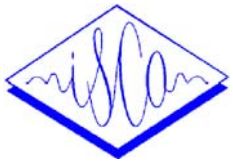

SPECOM'2004: 9th Conference
Speech and Computer
St. Petersburg, Russia
September 20-22, 2004

Turn-taking in social talk dialogues: temporal, formal and functional aspects

Louis ten Bosch[1], *Nelleke Oostdijk*[1], *Jan Peter de Ruiter*[2]

(1) Radboud University Nijmegen, The Netherlands
{l.tenbosch, n.oostdijk}@let.kun.nl
(2) Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
janpeter.deruiter@mpi.nl

Abstract

This paper presents a quantitative analysis of the turn-taking mechanism evidenced in 93 telephone dialogues that were taken from the 9-million-word Spoken Dutch Corpus. While the first part of the paper focuses on the temporal phenomena of turn taking, such as durations of pauses and overlaps of turns in the dialogues, the second part explores the discourse-functional aspects of utterances in a subset of 8 dialogues that were annotated especially for this purpose. The results show that speakers adapt their turn-taking behaviour to the interlocutor's behaviour. Furthermore, the results indicate that male-male dialogs show a higher proportion of overlapping turns than female-female dialogues.

1. Introduction

Turn-taking is one of the basic mechanisms in all types of dialogues and multilogues (conversations involving more than two people). It is also a crucial mechanism in human-system interaction. Some theories of turn-taking assume that dialogues essentially adhere to a half-duplex communication protocol, in which the interlocutors yield and take turns, guided by turn-taking cues (Duncan & Fiske, 1977). In their influential work, Sacks et al. (1974) have built a framework to describe the process of turn-taking, in which it is viewed as guided by a set of rules that speakers in a conversation adhere to. In this framework, turn transfers are assumed to occur at certain points (so-called Transition Relevance Places) and not at others. The smooth alternation of speaker and listener roles in a natural dialogue would then be the result of the aim of the interlocutors to minimize both the duration of overlapping speech overlaps and the time lapses between the turns.

From another perspective, dialogues are commonly described as the result of a joint activity between two speakers (e.g. Clark, 1996). The turn taking behaviour of speakers is the result of an interaction in which both speakers have a common goal. In this context, the term 'alignment' (Garrod and Pickering, 2004) is used to refer to the (essentially unconscious and interactive) process that smoothes the communication between speaker and listener, while making efficient turn-taking possible and contributing to facilitate mutual understanding.

Recent studies have attempted to investigate the turn taking mechanism by looking at certain features in the speech signal that correlate with the moment of turn changes. As a result, more is known about the relation between turn-taking and syntax, and about turn-taking and paralinguistic features of the utterance (e.g. Ford and Thompson, 1996; Koiso et al., 1998; Caspers, 2001). Some of these studies are based on dialogues in special situations, e.g. the Map Task (Anderson et al., 1991; Carletta et al., 1996), or under controlled conditions (Caspers, 2001). In Map Task dialogues, syntactic properties are shown to play an important role in the turn-taking mechanism (Koiso et al., 1998; cf. Selting, 1996). Moreover, it has become clear that turn-taking behaviour depends on whether speakers have a specific task and role and whether speakers may also communicate via other channels than speech. In the Map Task dialogues, speakers have a non-symmetric role in the conversation; one speaker is supposed to provide information about a certain task, while the other speaker is to follow the instructions. Recent studies (e.g., de Ruiter, submitted) aimed at investigating dialogues in which the roles of the speakers are more symmetric than in Map Task-like dialogues, and dialogues in which speakers could see each other under controlled conditions. De Ruiter showed that the distribution of duration of pauses and overlaps depends on the task given to the dialogue participants. A task involving a cognitive load leads to different turn-taking behaviour than the turn-taking observed in free conversation. In another study, ten Bosch et al. (2004) have shown that telephone conversations have much shorter inter-turn silences than face-to-face conversations.

While human-human dialogues show complex turn-taking phenomena, many speech-driven human-system interfaces impose a strict half-duplex protocol, due to the technical problems of echo cancellation and barge-in handling. But even when the technical problems are overcome, it appears that there are many fundamental problems, such as whether the "interrupted" speech output should stop immediately upon interruption, or continue, at least for a while, and how to handle the information that that the system had planned to convey, but was not rendered because output was aborted.

Should that information be kept on a stack, and if so, with high priority? Or should it rather be discarded as irrelevant (because the user did not let the system complete it)? Answers to these questions are very difficult to give as long as we do not improve our understanding of the formal and functional features of human-human dialogues.

Most large scale studies of human-human dialogues that have been performed so far were based on tasks that induce considerable structure in the dialogues. In this study, we focus on conversational dialogues that were not constrained by specific tasks. Since we expect a substantial amount of ambiguity, we take an approach that is based on an phenomenological analysis of human-human dialogues, without the constraints of any specific theory. The study focuses on turn-taking phenomena in Dutch social talk telephone conversations. It consists of two parts. The first part provides a surface description of turn-taking and related phenomena as observed in 93 telephone dialogues. In this part of the study we are mainly interested in temporal phenomena, for example the distribution of pauses, turns, and overlaps. In the second part, we focus on functional aspects of utterances and turns in a subset of eight dialogues. To that end we have annotated all turns in those dialogues for their function, using an annotation scheme adapted from the most general results from the research conducted on the Map Task corpus. We then proceed to discuss in more detail the verbal expressions that are associated with the various functions that turns may have. In all cases, we have focussed on a factual description of the durational and functional aspects as observed in these dialogues. It is important to emphasize that we do not intend to support a specific theory about turn-taking or dialogues, nor do we intend to speculate about the cognitive implications of the observed phenomena.

In the next section, we will discuss the speech material, our working definition of turn, and the use of the labelling system applied to the speech material. Section 3 deals with the functional aspects of turn-taking, while in section 4 an analysis is presented of the verbal expressions that characterize different utterance types. The final section presents our conclusions and plans for future research.

2. Data, annotation scheme, turn

2.1. Data

Our dialogue corpus consists of data taken from the 9-million-word Spoken Dutch Corpus (Corpus Gesproken Nederlands, CGN; Oostdijk et al., 2002). For this study, we selected telephone dialogues that had been recorded via a switchboard, originated from the Netherlands, and for which an orthographic transcription was available as well as a manually verified segmentation on the word level. In all, our dialogue corpus comprises 93 dialogues, 10 of which are between men, 36 between women, while 47 dialogues involve a male and a female speaker. This corpus was used for the first part of this study. For the second part of this study, a subset of eight dialogues has been annotated in greater detail on an utterance-by-utterance level, where the annotation refers to the function of the utterances in the context of the discourse.

All dialogues are informal and spontaneous; speakers knew each other (they were relatives or friends) and they were free to talk about any subject. On average a dialogue lasts nearly 9 minutes (77 of the dialogues, i.e. 83%, are between 7 and 11 minutes; the shortest dialogue is only 2.8 minutes, the longest is 12.2 minutes). Due to the fact that these conversations are informal, the role of both speakers was identical.

2.2. Annotation scheme

The orthographic annotation that was already available in the CGN served as a starting point to identify the individual utterances. An utterance was pragmatically defined as a word sequence that was terminated by either a period (.), an ellipsis (…) or a question mark (?). Table I depicts a small part of a dialogue as it appears in our corpus. The entire fragment shown in Table I spans about 11 seconds. Each line in Table I represents a single word, uttered by one of the speakers. The lines are ordered according to the moment the associated word starts – this information is based on the manually verified segmentation on the word level. The first column in Table I contains the utterance index, an integer indicating the sequential position of the utterance in the dialogue. The same index sequence is used for the utterances of both speakers. For each word the start and end time are presented in the second and third column ('Begin', 'End', in seconds). The fourth column ('Tag') presents a broad classification of the utterance in terms of the temporal position of the utterance in relation to the other speaker's utterances. This classification uses three labels, viz. 'continuation', 'interruption', and 'turn change', abbreviated 'cont', 'interr', and 'turn ch', respectively. These labels correspond to the temporal organization of the utterances in relation to turns that is depicted in Figure 1. The final two columns in Table I show the utterances (word by word, listed vertically) of both speakers A and B. The symbols 'xxx' and 'ggg' represent unintelligible speech and non-verbal speaker sounds respectively.

The table shows a fragment of a Dutch dialogue between two female friends showing a particularly complex turn-taking behaviour. Both speakers change turns very often, by rapidly reacting to the interlocutor.

2.3. Data analysis of the 93-dialogue corpus

In the corpus of 93 dialogues, the annotations do not distinguish between types of utterances (such as back-channels or utterances with a propositional content). We therefore present a global, quantitative analysis of the temporal structure related to sequences of utterances and

speaker changes, with the focus on pauses. In sections 2.3.1 and 2.3.2, we present our findings on the durations of pauses, and the distribution of pauses and overlaps that was observed in the 93-dialogue corpus.

Table I. A fragment of a telephone dialogue used in this study, showing the segmentation on word level, and the interaction between the two speakers. The first column ('Id') refers to the index of the utterance in the dialogue. On each line, the second and third column present the temporal information (in sec) about one word. Punctuation has been omitted in the verbatim representation (last two columns).

Id	Begin	End	Tag	Spk A	Spk B
261	307.700	307.800			gaan
261	307.819	307.910			we
261	307.910	308.132			daar
261	308.132	308.263			'ns
261	308.263	308.553			uh
261	308.553	308.698			in
261	308.698	309.263			Overijssel
261	309.263	309.394			of
261	309.394	309.606			zo
261	309.606	309.907			ergens
263	309.827	309.974	interr	nou	
261	309.907	309.960		in	
261	309.960	310.020			de
263	309.974	310.273		lekker	
261	310.020	310.300			buurt
263	310.273	310.664		toch	
264	311.079	311.171	turn ch		en
264	311.171	311.472			misschien
264	311.472	311.579			nog
265	311.497	311.575	interr	's	
265	311.575	311.720		niet	
264	311.579	311.661			'ns
264	311.661	311.712			een
264	311.712	311.987			keer
265	311.720	311.864		zo	
265	311.864	312.099		ver	
264	311.987	312.150			xxx
265	312.099	312.339		ook	
266	312.531	312.802	turn ch		nee
266	312.802	313.006			en
266	313.006	313.159			het
266	313.159	313.262			is
266	313.262	313.756			goedkoop
267	313.756	314.647	cont		ggg
268	314.096	314.917	interr	ggg	
269	314.647	315.546	interr		ggg
270	314.917	315.107	interr	ja	
270	315.107	315.244		ook	
270	315.244	315.532		dat	
271	315.532	315.713	cont	want	
272	315.546	315.746	interr		heel
271	315.713	315.801		m'n	
272	315.746	316.312			belangrijk
271	315.801	316.139		geld	
271	316.139	316.234		is	
271	316.234	316.442		op	
272	316.312	316.449			op
271	316.442	316.918		natuurlijk	
272	316.449	316.641			dit
272	316.641	317.146			moment
273	317.029	317.818	interr	ggg	
274	317.146	318.514	interr		ggg

2.3.1 Pause durations and distributions

We calculated the durations of all pauses (including all pauses between turns, and between utterances within turns), and the average pause duration per dialogue for all dialogues in our corpus. The results are shown in Figure 2. There appears to be a high correlation (0.894) between the average pause durations that two speakers in a dialogue maintain. This strongly suggests that speakers adapt their behaviour (in taking relatively longer or shorter pauses depending on their interlocutor). Speakers who participated in multiple dialogues in our material with different interlocutors also showed an adaptation in their average pause duration depending on the interlocutor. These findings support the notion that conversation has *rhythmic* properties that are maintained interactively.

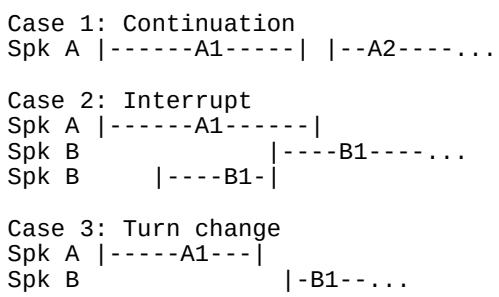

Figure 1. Temporal organizational patterns of utterances and turns as distinguished in the first part of this study. The symbols A1, A2, denote different utterances by speaker A, and B1 is an utterance by speaker B. The diagram visualizes the three temporally different situations that were distinguished while tagging the different turn transfers as presented in Table I. In the case of a 'continuation' (case 1), speaker A keeps the turn. The pause between A1 and A2 is optional. Case 2 is characterized by the fact that B1 starts before A1 terminates. There are two sub-cases, both labelled as 'interrupts' by speaker B, the difference being whether B1 ends before or after the end of A1. The situation found in case 3 (B1 starts after A1 ends) is labelled as a 'turn change'.

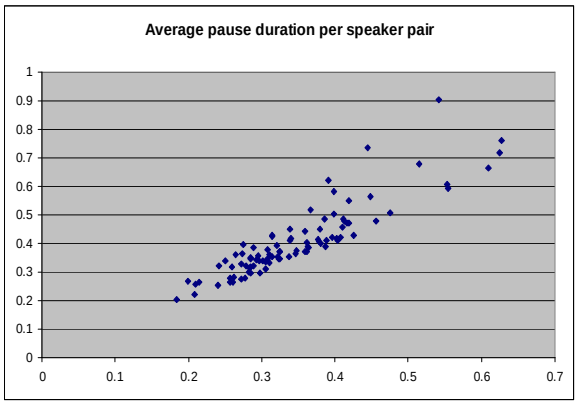

Figure 2. Scatter plot of the average pause durations. The axes represent durations in seconds. Each point represents one dialogue (and speaker pair), the coordinates of which are determined by the average pause duration (in sec) of each of the speakers. The

correlation of pause durations per speaker pair for all 93 dialogues equals 0.894.

Next, we categorized the pauses according to the points where they occurred in the discourse. Exploring the utterance labels that had been obtained in the labelling of the temporal organization patterns (cf. section 2.2, Figure 1), we distinguished three pause types: (1) pauses preceding turn changes, (2) pauses preceding continuations, and (3) pauses occurring within utterances. Table II presents an overview of the average pause durations (in secs) for these types of pauses, measured over all 93 dialogues.

Table II. Average pause durations and standard deviations for four types of pauses, measured over all 93 dialogues. For an explanation see the text.

Pause	mean	st.dev.
Turn changes	0.38	0.31
Continuations	0.52	0.38
Within utt.s	0.30	0.21

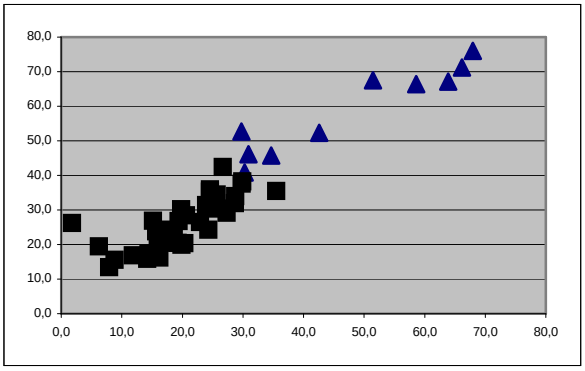

Figure 3. Scatter plot of proportion of turn transfers involving overlapping speech in male-male (triangles) and female-female (squares) dialogues. The axes represent percentages.

The data indicate that speakers in a free conversation take the opportunity to make longer pauses before continuations (i.e., between utterances in a single turn) than in any of the other situations (turn changes, and utterance internal pauses). Apparently, speakers are not afraid that their interlocutor grabs the floor if they make relatively long pauses, despite that fact that the number of interrupts is quite high. As a disclaimer, we must add that in these measurements no distinction is made between utterances that are back-channels (Yngve, 1970) or 'continuers' (Schegloff, 1982) on the one hand, and utterances with propositional content on the other hand.

2.3.2. Overlap

In the present study, both turn changes and interrupts relate to the transfer of the turn from one speaker to another. While turn changes occur without there being overlap, we here define 'interrupts' as utterances where the beginning of the utterance overlaps with the utterance of the other speaker (cf. section 2.2, Figure 1). In our data we observed a substantial gender effect in the *proportion* of turn transfers that were realized as interrupts. Figure 3 shows a scatter plot in which male-male speaker data (10 dialogues) are displayed together with female-female speaker data (36 dialogues). The male-male conversations consistently show a larger proportion of interrupts. The data also suggest an adaptation between the speakers with respect to the amount of interrupts compared to the total number of turn transfers.

3. Turn-taking and functional aspects

In the previous section, the analysis of the 93-dialogue corpus was based on the availability of the utterance as a unit in the orthographic transcription and the time stamps available from the manually verified segmentation at the word level. No attempt was made to distinguish between different types of utterances from a discourse-functional perspective. Evidently, such an annotation makes finer distinctions in turn-taking phenomena possible. In the present section (subsections 3.1 and 3.2), we report on the more elaborate tagging of a subset of the material that we carried out and the results we obtained from an analysis of the data using the refined annotation scheme.

3.1. Discourse-functional tagging

A subset of eight telephone dialogues (4 female-female, 2 female-male, 2 male-male) taken from the 93-dialogue corpus that was used in the previous section has been enriched with a finer-grained tagging. This additional tagging refers to the function of utterances in the discourse. For the annotation system that we designed for this purpose we took the Map Task annotation scheme (Carletta et al., 1996) as a starting point. Since the annotation used in the Map Task context was beyond what was required for the present purpose, and also because the scheme was dedicated specifically towards the annotation of instructional dialogues in which the roles of the speakers is not symmetric, we decided to adapt the original Map Task scheme. The resulting adapted scheme resembles that used in Walker and Whittaker (1990) in which prompts, commands, questions, and assertions were distinguished, and is close to the level of 'conversational moves' as described in Carletta et al. (1997). We maintained the top two levels of the tree structure in the Map Task annotation scheme, and added some more detail in the classification

of the back channels. Thus we arrived at the coding scheme presented in Figure 4.

Category	Abbrev.
Greeting	g
Back channel	b
Repeat	r
Continuer	c
Acknowledgement	ac
Phrases with propositional content	p
Questions	q
Answers	an
Statements	s

Figure 4. Utterance-based annotation scheme. For an explanation see the text.

The labels 'r', 'c' and 'ac' were given to utterances without propositional content that could be specified in a more precise way ('repeat', used for utterances that repeat a large part of the previous utterance by the other speaker, 'continuer', or 'acknowledgement'). Typical continuers are 'oh?', 'tsss', while 'hmm' often merely acknowledges the speaker about the attention paid by the listener. The parent label 'b' (back channel) was given to the other utterances without any propositional content. With respect to the utterances with propositional content, the label 's' (statement) is given to all phrases that were not clearly distinguishable as a question or an answer. On top of this, utterances that were labelled with 'q', 'a', and 's' may receive an additional sub-tag 'turn claim', 'turn keep, or 'turn yield' (abbreviated 'tc', tk' or 'ty').

A cross-validation on a subset of the material by two annotators has shown that this coding scheme can be applied without serious ambiguities (a Cohen's kappa value of 0.82 was found for the between-labeller agreement of a labelling 'question', 'answer', 'statement', and 'back-channel'). This reasonably high value is not surprising, as Carletta et al. (1997) report kappa values of 0.80 and higher among four annotators who applied the Map Task annotation scheme. However, despite the seemingly simple annotation scheme, some ambiguities were found difficult to resolve, and we believe that any annotation scheme is necessarily a compromise between optimisation of distinctiveness on the one hand and reduction of ambiguity in labelling on the other hand. Ambiguities that were encountered in the free conversation dialogues could be divided into three types. Firstly, very short utterances such as 'werkelijk?' [Eng. 'really?'] could be interpreted as back-channel or as question with real propositional content. Secondly, a high frequent word such as 'ja' (Eng. 'yes') is common as a genuine back-channel. However, in response to a question it functions as a clear affirmative answer. These are the two clear cases, but in practice we found that there are numerous instances where the interpretation of the function of 'ja' is less straightforward. A third type of ambiguity occurs in the labelling of repeats. Repeats often occur on the word level (e.g. 'mmm' – 'mmm', 'vandaar' – 'vandaar') and then usually function as back-channels. Occasionally (< 3% of the instances), however, they are realised as short phrases (e.g. 'ben je alweer thuis?' – 'ik ben alweer thuis', 'da's wel leuk dus' – 'da's wel leuk ja'; Eng. 'are you already at home?' – 'I'm already at home', 'so that's quite nice' – 'that's quite nice indeed'). Our strategy was to use high-level labels in ambiguous cases (i.e., the labels 'b' or 'p'), with preference of the 'propositional phrase' in case of ambiguity between the labels 'back-channel' or 'propositional phrase'.

3.2. Definition of 'turn' and 'turn transfer' refined

The turn-taking mechanism as described and used in the previous section is based on a mechanical interpretation of the starting and ending times of each utterance. The classification of the different turn transfers was based on the temporal organisation of the utterances, without taking into account their functional relevance in the dialogue. For example, turns falling under case 2 (cf. Figure 1) are classified as an interrupt, regardless the fact whether B1 is a back channel or an utterance with propositional content.

For the purpose of this section, we have refined the concept of 'turn' and 'turn transfer'. The discussion about what a turn change is will be clarified on the basis of the cases as shown in Figure 5.

```
Case 1: Continuation
Spk A |------A1-----| |--A2----...

Case 2: Interrupt
Spk A |------A1------|
Spk B             |----B1----...
Spk B     |----B1-|

Case 3: Turn transfer
Spk A |-----A1---|
Spk B              |-B1--...

Case 4:
Spk A |-----A1---|
Spk B  ...B0--| |-B1--...
Spk B  ...B0---|  |-B1--...
```

Figure 5. Elaborated version of Figure 1. This figure shows the same cases as depicted in Figure 1, but a special case 4 is added, in which speaker B starts a new utterance B1 after a short pause after B0.

Compared to Figure 1, we have added a new case: case 4. Case 4 is added because it sheds light on the inherent complexity of any turn-taking taxonomy. If, in case 4, the pause made between the utterances B0 and B1 is long – long enough to be significant given the discourse – then B1 can be regarded to induce a turn

transfer. In this case, B takes over the turn from A. However, if the pause between B0 and B1 is short, it is evidently not straightforward to interpret case 4 as a turn change per se. In this case, it probably depends on the semantic content of B1, whether or not B1 can be related to a turn transfer.

Given this problem, we have defined a 'genuine' turn change, i.e. a turn change in which propositional phrases are the major determinants, to take place in the following specific cases. Obviously, case 1 is never related to a turn change. In case 2, only the first option leads to a 'genuine' turn change if B1 is a proposition. Also in case 3, a turn change takes place if B1 is a proposition. In case 4, there is only a turn change if the pause between B0 and B1 exceeds a certain threshold (see below) *and* B1 is a proposition.

Under the refined scheme almost all cases receive a modified interpretation depending on whether utterances have the status of a proposition or of a back-channel. We argue that a 'genuine' continuation (case 1) is defined by A1 and A2 both being propositions. As indicated, 'genuine' turn transfers only occur under the condition that the utterance B1 is a proposition.

Based on the labelling (in terms of the taxonomy in Figure 5) of the utterances in the eight dialogues, the 'genuine' turn changes were found using a computational approach, in which the tags in both speaker tiers were systematically investigated in order to categorize the local temporal patterning into one of the four cases. Based on an investigation of the dialogues, the threshold necessary in case 4 was set to one second. This one-second setting can be justified by two arguments. Firstly, in this material, the duration of one second must be considered as a long pause – most pauses between turns are substantially shorter (cf. Figure 2). This is in line with an observation described by the 'one-second rule' (Jefferson, 1989), according to which conversants usually avoid *intra-turn* pauses longer than one second. The second reason is that the short propositional utterances found in the corpus are often shorter than one second, which indicates that a second is functionally long enough to construct a genuine turn change.

In the subset of eight dialogues, a total of 3046 turn changes occur when no distinction is made between the functional types of the utterances. Of these turn changes, a number of 1398 (about 45 percent) appear 'genuine turn changes', i.e. meeting the additional constraints that were explained above.

Figure 6 presents more information about the difference between turn changes and 'genuine turn changes'. It presents the histograms related to the Floor Transfer Offset or FTO (see De Ruiter, submitted), which is defined as the duration of a pause between two turns by different speakers (FTO > 0), or the negative value of the duration of the overlap (FTO < 0) of two turns by different speakers. The upper histogram shows the FTO in the situation where the durations are measured regardless of the functional tag of the utterance. The lower histogram shows the distribution of the FTO in the case of the 'genuine turn transfers'. The mean FTO-value is 0.16 and 0.34 sec, respectively, while the median values are 0.11 and 0.24 sec, respectively. A comparison of the two histograms shows that the temporal distribution of the back-channels substantially decreases the mean and median of the FTO – in line with one of the roles of back-channels as 'pause-fillers'. Furthermore, the back-channels are the cause of the large contribution of FTO values between –0.5 sec and 0 sec in the upper histogram, i.e. for an interruption. This suggests that the group of back-channels, as a subset of the set of all utterances, obeys other rules than the rules that speakers adhere to when producing full-content propositions. The difference in statistical structure is exemplified by the right-hand tails of the histograms: The proportion of durations longer than one second is different in both situations (12.8 percent for all turn types, versus 17.3 percent for 'genuine' turn changes).

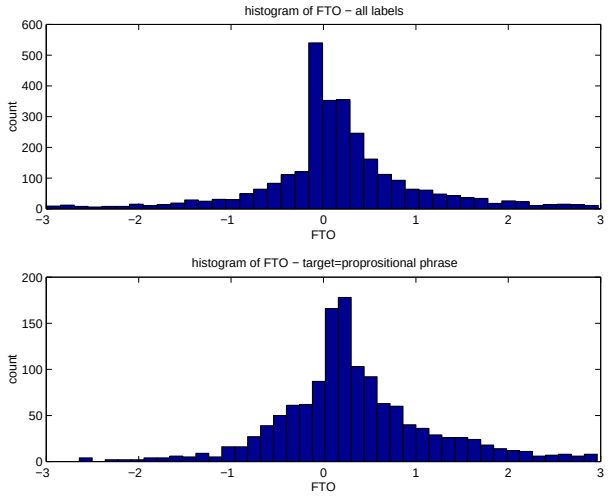

Figure 6. **Histograms of the Floor Transfer Offset (FTO, see text) in two different situations**. The upper histogram shows the FTO in the situation where the durations are measured regardless the functional tag of the utterance. The lower histogram shows the distribution of the FTO in the case of the 'genuine turn transfers' as defined in section 3.2. In the histograms, a total of 3046 and 1398 turns have been taken into account. The y-axes represent absolute counts.

3.3. Transitions between utterance types

On the basis of the dialogues with functional annotation, it is possible to derive a statistical patterning of subsequent proposition types by considering bigrams of the corresponding tags. Statistics were collected for the FTO and then these were split out according to the functional tags associated with the utterances with propositional content before and after the turn change.

The regularity of this patterning is shown in the form of a matrix in Table III.

In Table III, each row corresponds to specific type of utterance of the first speaker, while the columns present the types of utterance in reaction by the second speaker. The table is based on a total of 1395 genuine turns – the same set as used for the lower histogram in Figure 6. (Three data points are missing here due to a slightly different processing). For example, the table shows that 86 percent of the questions are followed by an answer (possibly preceded by a back channel). The matrix has been reduced for reasons of brevity – therefore not all row sums equal exactly 100 percent.

Table III. **Percentages of the 'genuine' turn changes in terms of the pair of functional tags.** The last column contains the number of occurrences. For an explanation see the text.

From\To	s	q	an	tc	r	tot
s	69	17	2	3	7	649
q	8	5	86	0	1	149
an	52	39	2	0	6	54
s+tc	75	25	0	0	0	62
s+ty	70	19	4	4	4	77
b	81	10	2	3	2	303
r	56	11	22	0	11	18
ac	84	8	1	4	1	15
ag	79	18	0	0	4	28
tot						1395

4. Analysis of verbal expressions

Until now, we have presented a description of the turn-taking phenomena with emphasis on the temporal aspects. In this section, we give a brief overview of verbal expressions as observed in the data. In agreement with Ten Bosch et al. (2004), an analysis by hand led to the observation that utterances can be broadly classified into 4 types:

1) back-channels (very short, one to five tokens: *um, mmm, ja, goh zeg, dat zal wel ja; Eng. Um, mmm, yes,* …). We also considered short repeats to fall in this category. Repeats are utterances that literary share a fragment of the utterance spoken by the other speaker.
2) Failed attempts to take over the turn (usually rather short: e.g. *ik ben uh …, maar da's uh …, hé maar …; Eng.. I'm uh …, but that's uh* …). Utterances of this type often start with a sequence of back-channel like words.
3) Short propositional utterances that provide some feedback to the previous utterance or turn (mostly content-based, e.g. *grappig, da's wel substantieel; Eng. funny,* …).
4) Longer actual propositional phrases. Directly after a turn change, 32 percent of all propositional phrases in our material start with a word sequence that, without any following propositional content, would have been labelled as back-channel. That means that 32 percent of these longer propositional phrases starts in the same way as a back-channel does. Just before a turn change, that is, at the end of a speaker's turn, the situation is slightly more complex. Utterances with propositional content can either end without any back-channel-like words (e.g. most questions), or they may end with typical back-channel-like sequences (e.g. most hesitations), or they may contain turn yield cues which are often semantically based ('tja … ik weet het ook niet …', 'I don't know either …').

5. Discussion and conclusion

The analyses described in this paper lead to a number of interesting observations. An analysis of durations of varies types of pauses in 93 telephone dialogues has shown that speakers adapt their turn-taking behaviour with as a result a high correlation between average pause durations in the speech produced by the two speakers. Especially in the case of interrupts, there is a clear gender effect in the proportion of interrupts compared to turn changes in general. Both effects (adaptation of pause duration, and adaptation of interrupt behaviour) can be interpreted as a form of mutual 'alignment' between both speakers.

We believe that, although there are as yet few independent results to substantiate this claim, these findings are stable and reproducible in a larger collection of dialogues of similar type. The fact that the dialogues are spontaneous and without specific task for any of the speakers of course increases the homogeneity of the material.

The definition of pause was based on a 'mechanical' interpretation of beginning and end of 'utterance', without taking the difference between back channels and 'propositions' into account. Evidently this is a drawback, since the function in the discourse of back-channels is different from the phrases with propositional content. To overcome this drawback, we have explored the possibility to enrich a subset of the material with a tagging that refers to the high-level function of utterances in the context of the discourse. To that end, a labelling system was used that was inspired by the Map Task tagging tree, but simplified and slightly extended to meet the requirements of the task. Furthermore, the definition of 'turn' and 'turn change' has been refined using this functional tagging. By using this tagging, in combination with a more strict definition of turn change, we found that about 1395 of the 3046 (about 45 percent) of all 'turn transfers' are genuine.

On the basis of the two different histograms of the Floor Transfer Offset, one related to all utterances and one based on utterances with propositional content only, the structure of a free conversational dialogue might well be explained by assuming two processes in which one processes overlays the other. The basic, underlying process governs the orderly alternation of the genuine

propositions. On top of this regular patterning, the production of back channels interferes with the regular consecutive sequences of propositions which substantially affects the FTO distribution. The extent to which this change is caused by back-channelling will probably be dependent on the type of dialogue. For example, in a strict question-answer dialogue, the type of back-channelling will probably have a different statistical character, with a different eventual FTO pattern as result.

The results make clear that in free human-human dialogues speakers adapt their turn-taking behaviour to the behaviour of the interlocutor. This convergence may be interpreted as the result of an 'alignment' process in the sense of Garrod and Pickering (2004). For voice-driven human-machine interaction, this implies that such adaptation processes may be an important factor in the design of an interface supporting an interaction that is as natural as possible.

A number of questions remain unanswered. Is the definition of genuine 'turn transfer', as proposed in section 3, appropriate? What is the correct interpretation of the start of a turn if a speaker turn consists of a back-channel followed by a propositional phrase? How can the temporal phenomena of turn-taking be described in terms of a target function that is to be optimised as result of a joint activity by both speakers?

These questions are intriguing, since after close inspection of the speech material, we are convinced that the human turn-taking behaviour is much more complex than the mechanism of an 'exchange of turns' suggests. Turn-taking, i.e. the alternation of the role of speaker and listener, is to be considered a quite particular situation during a social-talk role-symmetric dialogue. How to address these complex phenomena with a theoretical framework that goes beyond the turn-taking protocol? We intend to return to these issues in future research.

Acknowledgements

The work of Louis ten Bosch and of Jan Peter de Ruiter is made possible by the European IST project COMIC (IST-2001-32311). Thanks are due to Peter Beinema for preparing the data that were used in the analysis of pause durations and distributions of pauses and overlap in the 93-dialogue corpus.

6. References

Anderson, A.H., M. Bader, E. Gurman Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowto, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4): 351-366.

Burger, S. (1997). Transliteration Spontansprachlicher Daten, Lexikon der Transliterationskonventionen in Verbmobil II. Munich, Verbmobil Technical Report 56-97.

Carletta, J., A. Isard, S. Isard, J. Kowto, G. Doherty-Sneddon, and A. Anderson (1996). *HCRC Dialogue Structure Coding Manual*. Technical Report HCRC/TR-82. Human Communication Research Centre, University of Edinburgh.

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

Caspers, J. (2001). Testing the perceptual relevance of syntactic completion and melodic configuration for turn-taking in Dutch. *Proceedings Eurospeech Conference*, pp. 1395-1398.

Clark, H.H. *Using Language*. Cambridge MA, Cambridge University Press.

de Ruiter, J.P. (2004). Macroscopic explorations in dyadic turn taking. *Submitted.*

Duncan, S.D., and Fiske, D.W. (1977). Face-to-face interaction: Research, Methods and Theory. Hillsdale, New Jersey: Lawrence Erlbaum.

Ford, C.E. and 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*, Cambridge: Cambridge University Press, pp. 134-184.

Garrod, S., and Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, vol. 8, nr 1: 8-11.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In: D. Roger & P. Bull (Eds.), *Conversation, and interdisciplinary perspective* (vol 3, pp. 166-196). Clevedon: Multilingual Matters Ltd.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and 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.

Oostdijk, N., et al. (2002). *Het Corpus Gesproken Nederlands*. Collection of papers about the Corpus Gesproken Nederlands. LOT Summer School, Netherlands Graduate School of Linguistics, 2002.

Sacks, H., Schegloff, E,.A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50:. 696-735.

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn. Constructional units and turns in conversation, *Pragmatics* 6: 357-388.

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

ten Bosch, L., Oostdijk, N., De Ruiter, J.P. (2004). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. *Proceedings of the 7th International Conference on Text Speech and Dialogue*, Brno, Sept. 2004.

Walker, M., and Whittaker, S. (1990). Mixed initiative in dialogue: An investigation into discourse segmentation. In: *Proceedings of the 28th meeting of the ACL*, pp. 70-78.

Weilhammer, K., and Rabold, S. (2003). Durational aspects in Turn Taking. *Proceedings of the International Conference of Phonetic Sciences*, Barcelona, Spain.