Layered structures in dialogues

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

FLE/uh 89/235
20.12.1989

Rapport no. 726

Layered structures in dialogues: from what to how and vv.

F.L. Engel
R. Haakma
L.J. Kraak
M.G.P. Mulders

Layered structures in dialogues: from *what* to *how* and vv.

Engel F.L., Haakma R., Kraak L.J. and Mulders M.G.P.

Abstract

To obtain information or to let others (men or machine) know or do things, one has to communicate one's aims by the exchange of messages. Messages describe these goals in coded form. Normal inter-human dialogues are characterized by simultaneous communication of multi-layered intentions over more than one communication channel. Also multi-modal man-machine interfaces tend to obtain these complex characteristics.

For improving the design of such multi-channel man-machine interfaces, we have tried to unravel characteristics of normal inter-human communication as a first step. In this report, messages exchanged during an information dialogue over telephone have been transcribed into a layered communication structure, the guiding principle at the sender's side being the recursive conversion from intention ('what') to implementation ('how') and vice versa at the receiver's side. This layered representation of human dialogue appears to offer a framework for the structuring of message exchange in inter-human as well as in man-machine communication.

Contents

Introduction

Task oriented dialogues

To obtain information or to let others (men or machine) know or do things, one has to communicate one's aims by the exchange of messages. Messages describe these goals in coded form, their coding being more or less adapted to the communication characteristics, including possible transmission errors, of the specific channels used.
The receiver of a message is assumed to decode it in the context of his knowledge and beliefs about the partner, the current dialogue and the domain of discourse, and to interpret the sender's aims in the context of his own goals.

Error prevention

Note that communication errors might arise from physical disturbances in the transmission channel, but also from errors in the coding and decoding of the message contents. For effective communication, coding and decoding conventions of sender and receiver have to match. A coder/decoder mismatch, as in the case of e.g. partners using a different mother language, constitutes a source of communication errors.
Mainly by feedback, viz. by observing the receiver's reactions (communicative as well as other e.g. manual actions), the sender can detect errors in the decoding and/or interpretation of his message.

Protection against communication errors is also realized by the use of redundancy in message coding. The communication conventions ('protocol'), such as the for coding applied message syntax, represent in general a certain amount of redundancy. More redundancy can be introduced, e.g. in human communication by using more words and/or by the simultaneous transmission of a the same message via more than one channel (e.g. by saying "yes" and nodding simultaneously). The receiver is assumed to use this redundancy for fault tolerant message recognition (decoding).

However, also a certain efficiency in coding (effort, speed) might play a role in the use of multiple channels for a single message. For instance in the spoken sentence "Put that there" combined with pointing, the sender prevents extensive verbal coding by pointing. However, the sender's gain in efficiency has to be payed by the receiver, who has to combine the partial messages for perception of the speaker's goal. Incorrect synchronization of these partial messages might again lead to certain errors in communication.

Intentional structure

As touched upon, with the 'put that there' example, more complex communication goals are in general split into a hierachy of sub-goals.
The activity of coding the goal into a message has been implemented by the splitting of the main goal into sub-goals and coding these into sub-messages, to be send sequentially and/or in parallel. The receiver then has to reconstruct the more global goals from the earlier decoded sub goals.

In fact, the conversion of the main goal into subgoals together with a timing recipe of how they are to be achieved in time, already constitutes an encoding activity. The abstraction of the main goal from the subsequently or simultaneously perceived sub-goals then represents a message decoding activity.
Note that this activity of splitting and transforming a message into sub-messages can

start at a rather abstract layer, e.g. the overall goal of the dialogue, and can be repeated recursively down to what is called the 'physical layer', where messages are to be coded in terms of temporal fluctuations of the physical parameters of the communication channels concerned (e.g. in fluctuations of air pressure in the case of sound).

Layers of communication

In the ISO reference model for computer-computer communication, the "Open System Interconnection (OSI) model", the decoding by the receiver is structured in a mirrored way compared to that of the sender, thus layering the communication into separate layers with matching coder-decoder pairs. The purpose of each layer being to offer certain communication 'services' to the higher layers, shielding those layers from the details of how the offered services are actually implemented, see e.g. Tanenbaum (1981). Layer 'n' on one machine thus carries a virtual conversation with layer n on another machine, this virtual communication being implemented by passing data and control information to the layer immediate below, until the lowest layer is reached. At the lowest layer there is real physical communication with the other machine.

It is a claim of Taylor's (1988) layered protocol model that also in inter-human communication, the decoding by the receiving partner is structured in a mirrored way compared to that of the sender, so that the at a given layer encoded message becomes decoded again at the corresponding layer of the receiver. Accordingly, also feedback and other error handling messages are assumed to become layer specific.

The idea of layer specific dialogue control has been described already by Bunt, van Katwijk, Muller and van Nes (1980), who considered verbal information dialogues and found that what they called 'dialogue control acts' could be distinguished into layers of functional processing, viz. in those of perception, processing and production of dialogue information. Dialogue control acts have been defined by them as bearing on the communication as such, rather than constituting direct information transfer. Their concept of layered dialogue control acts comes close to the layered message verification/correction protocol as proposed by Taylor (1988).

Certain designers of human-computer interfaces claim that optimal man-system interfaces should also be layered, see e.g. Buxton (1983), Norman (1984) and Nielsen (1986). The underlying idea apparently is that the behavioural interface of the machine (computer) reflects some 'frozen' communication partner, being fixed by the system's programme. No agreement exists, however, about the number of layers, nor about the precise message coding/decoding procedures these layers should represent.

Certain experimental evidence for a human message processing model based on hierachical perception and related response generation has been found in terms of differences in reaction times for tasks of different cognitive complexity, see e.g. Woodworth and Schlosberg (1965). More recently, Norman (1981) used a hierarchical action-sequence ('Activation-Trigger-Schema') model for the explanation of errors that occur when a person undertakes an action that is not intended, the so called 'action slips'.

Whether and how inter-human communication is best described in terms of layered communication remains to be analyzed experimentally. The experimental approach is especially needed as the related communication 'protocols' (the coding and decoding as well as the error detection and correction procedures) are overlearned habits, that function almost unconsciously.

In the following chapters is described how we have converted an earlier recorded telephone information dialogue, as transcribed by Cramer (1985), into an explicitly layered structure of intention communication, where the communication layers correspond with the related layers of intention. The experiences gained with this analysis will be indicated.

Experimental data

Dialogue task

The dialogue concerned aimed at the acquisition of flight and related train departure data from a flight information office via telephone. For the precise conditions maintained during the experiments, see Pruest, Minnen and Beun (1984) and Beun (1985).

The subject's task was as follows:

"You like to visit your daughter in Los Angeles. The plane (with flight number KL 401, or was it the KL 601?) departs Tuesday next week. You are leaving by train from Tilburg. What is the latest time to leave if you are to catch the plain?"

The informant had time tables and other data to her disposal, from which the inquired travel information could be retrieved.

Transcription

The transcribed source text, on which the following structure analysis is based, consisted of the literal quotation of what was said by both partners, together with certain temporal indications such as the duration of pauses, the beginning of speech utterances, etc. The dialogue was in the Dutch language, see Appendix A for an English translation.

Layered dialogue description

Information flow charts

Tainsh (1985) has used 'job process charts' to describe the internal activities and mutual communication of two interacting systems, being two machines or a machine and its user.

A job process chart contains three columns: viz. one for description of the user's task, one for the machine's task and the column in between for indicating the message flow between them. The task columns of the two dialogue partners contain program 'flow-charts' (see e.g. McCracken, 1965) which indicate the operations to be successively performed by the two systems, as a function of received and/or derived data. The operations of the two systems are synchronized by the passing of messages between them at certain moments of time.

In order to represent a dialogue between two partners linearly in time, the program loops that normally occur in flow-chart descriptions are linearized. Moreover, for being able to represent layered dialogue structures, we indicate the related task layer with help of indentation. To distinguish these flow charts from others, they will be called 'information flow charts' in the sequel.

From 'what' to 'how'

Fig. 1 gives a survey of the different classes of (virtual) messages distinguished in the communication of the multilayered goals of both partners. Characteristic for the introduced layering is that the goals of a given layer are expanded into more detailed subgoals at the next lower layer. In this way, 'What' has to be communicated becomes converted into 'How' it will be communicated at the layer below. With the applied terminology we tried to express the with goal expansion related increase of specificity.

For instance, by the how-decision of doing it by telephone instead of e.g. mail, the layer 0 activity of exchanging travel information becomes expanded into 1- initiating a telephone call, 2- exchanging flight and rail data (indicated in more detail compared to 'travel information') and 3- closing the telephone call. On their turn these activities become further expanded, e.g. the exchange of flight and rail data changes into the exchange of 1- discourse context, 2- message (inquiry) structure and 3- (again more detailed) plane and train data.

Note that the subject's layer 0 communication goal is in fact quite specific, viz. coming to know the latest train departure time from Tilburg, in contrast to the less detailed communication goal of Schiphol Information Service, viz. of being cooperative by providing travel information. However, for coming to know his train departure time the subject has to communicate a much broader context of transcontinental flights, etc.

Also note that the lowest layer, here layer 3, covers all relevant operations down to the 'physical layer', where in our case different physical communication channels are in-

Layer 0:	1 Exchange of travel information
Layer 1:	1.1 Initiation and closing the telephone call
	1.2 Flight and rail data exchange via telephone
Layer 2:	1.1.1 Making and closing the telephone connection
	1.1.2 Discourse opening and closing
	1.2.1 Exchange of discourse context
	1.2.2 Exchange of message structure
	1.2.3 Exchange of plane & train data
Layer 3:	1.1.1.1 Finding and sending telephone numbers
	1.1.1.2 Handling the telephone and its alarm / interrupt signals
	1.1.1.3 Telephone call stop phrases
	1.1.2.1 Caller & callee name exchange
	1.1.2.2 Discourse start & stop phrase exchange
	1.2.1.1 Exchange of discourse topic
	1.2.2.1 Exchange of question/answer structure
	1.2.2.2 Exchange of question/answer confirmation, repeat request, etc.
	1.2.3.1 Exchange of plane & train data context
	1.2.3.2 Exchange of plane & train departure times and rules, flightnumbers, etc.

Fig. 1: Survey of the applied message classes. The tree numbering indicates the mutual relations among the operations at the different layers. For instance, 1: the 'Exchange of travel information' at layer 0 is expanded at layer 1 into 1.1: the 'Initiation and closing of the telephone call' and 1.2: the 'Flight and rail data exchange via telephone'. On its turn, layer 1 operation 1.1: 'Initiation and closing of the telephone call' becomes expanded into the layer 2 operations 1.1.1: 'Making and closing a telephone connection' and 1.1.2: 'Discourse opening and closing', etc.

volved; for instance besides the voice channel also a manual channel for number selection, horn lifting, etc.

Layer 1

Fig. 2 gives the layer 1 information flow chart for the example dialogue 2 ('Situation A'), as have been transcribed by Cramer (1985). At this layer only 'virtual messages' are exchanged between both partners, which means that at this layer there is only indirect communication between them. Virtual communication is realized by the internal exchange of 'real messages' with the layers below. It is only at the lowest layer, here layer 3, that real communication occurs between both partners. Accordingly, the utterances exchanged in our example can be found at layer 3, see Figs. 4,5 and 6..

In the interaction column of Fig. 2, arrows indicate the flow of communication. At layer 0 the subject starts communication, this is indicated by 'b=>' at the upper left corner of the interaction column, and also ends it as is indicated by 'e=>' in the lower left corner of the interaction column. These directions in communication are derived bottom-up from the lowest communication layer considered, so layer 3 here. In fact, they indicate at the given layer who started and who ended the communication segment concerned.

The idea behind this direction analysis is that the 'leading partner' is assumed to initiate communication, see e.g. Weijdema, Dik, Oehlen, Dubber and de Blauw (1982). The lead can be taken over by the other partner through e.g. a 'sub-sequence' of messages (for instance by asking explanation), thus supporting the achievement of certain communication sub-goals. In our case, the subject initiated all communication chunks at the layers 0 and 1, what makes him there in terms of Grosz and Sidner (1986), the initiating conversational participant.

Subject	Layer 0,1 Interaction	Schiphol Information Service
0 Obtain travel information	b=> Begin travel info.exch.	0 Provide travel information
0 => 1 Begin		0 => 1 Begin
1 Initiate tel. call Schiphol Info	b,e-> - - - - - - - - - - -	1 Receive tel. call
1 Flight & rail data exchange	b-> - - - - - - - - - - - <-e	1 Flight & rail data exchange
1 Close tel. call Schiphol Info	b,e-> - - - - - - - - - - -	1 Finish tel. call
0 <= 1 End (Info acquisition)	e=> End travel info.exch.	0 <= 1 End (Info. provision)

Fig. 2: Information flow chart of communication layers 0 and 1. The layer 0 interaction begins at the subject's side (indicated by 'b=>' at the lefthand top of the interaction column) and is also ended by him (indicated by 'e=>' at the bottom lefthand side of the interaction column).
In line, layer 1 interaction begins at the subject's side (the 'b,e->' arrow at the lefthand top of the interaction column) and also ends there (indicated by the 'b,e->' arrow at the bottom left side of the interaction column).

The following rules have been applied for deriving these communication directions.
- The direction of initiation of a communication chunk at a given layer corresponds with the direction of the first message one layer lower.
- The direction of the end of a communication chunk at a given layer corresponds with the direction of the last message one layer lower.
- The message direction at the lowest layer is determined by the partner who sends and receives the related physical messages.

Note that the direction of the end of a communication chunk depends on what is considered to be the beginning of the 'last' message a layer below.
The beginning and endings of communication are indicated in the information flow charts by arrows provided with a 'b' and/or 'e' respectively.

Subject	Layer 1,2 Interaction	Schiphol Information Service
1 Initiate tel.call Schiphol Info	**b=> Begin initiation tel. call**	**1 Answering tel. call initiation**
1 => 2 Begin		1 => 2 Begin
2 Make tel. connection	b-> - - - - - - - - - - - <-e	2 Handle tel. connection request
2 Handle discourse opening	e-> - - - - - - - - - - - <-b	2 Send discourse opening
1 <= 2 End	**e=> End init. telephone call**	1 <= 2 End
1 Flight & rail data exchange	**b=> Begin fligth & rail data**	**1 Flight & rail data exchange**
1 => 2 Begin		1 => 2 Begin
2 Supply of discourse context	b-> - - - - - - - - - - - <-e	2 Receive discourse context
2 Supply of message structure	b,e-> - - - - - - - - - - -	2 Receive message structure
2 Exchange plane & train data	b,e-> - - - - - - - - - - -	2 Exchange plane & train data
2 Supply of message structure	b,e-> - - - - - - - - - - -	2 Receive message structure
2 Exchange plane & train data	b-> - - - - - - - - - - - <-e	2 Exchange plane & train data
1 <= 2 End	**End flight & rail data <=e**	1 <= 2 End
1 Close tel. call Schiphol Info	**b=> Begin closing tel. call**	**1 Close tel.call Schiphol Info**
1 => 2 Begin		1 => 2 Begin
2 Discourse closing	b-> - - - - - - - - - - - <-e	2 Discourse closing
2 Closing tel. connection	b,e-> - - - - - - - - - - -	2 Close tel. connection
1 <= 2 End	**e=>End closing of tel. call**	1 <= 2 End

Fig. 3: Layer 2 expansion of the layer 1 tasks given in Fig. 2. Note that the layer 2 'Make tel. connection' operation in the (layer 1) 'Initiate tel.call Schiphol Info' procedure is bilateral (indicated by arrows in both directions).

Layer 2

The three layer 1 operations have been expanded further at layer 2, see Fig. 3.
Note that during the flight & rail data exchange the loop 'Supply message structure' - 'Exchange plane & train data' has been run two times in the example dialogue. After each loop Schiphol Information Service has to do a low layer check whether a new loop follows or that the beginning of the telephone call closing operation is received.

With regard to the turns taken by the two partners, it can be observed that all three layer 2 communication chunks are started by the subject. Only the flight & rail data exchange is ended by Schiphol Information Service.

Layer 3

Related to 'Initiate tel. call Schiphol Info' at layer 1, Fig. 4 gives the layer 3 messages exchanged between the subject and Schiphol Information Service.

Note that layer 3 communication between both partners occurs via speech as well as via manual operations. In fact, lower layer interaction with the telephone set itself and with the telephone switching center via the telephone set is involved. These interactions should become separated once communication lower than layer 3 is considered. For simplicity reasons, these layers are omitted for the moment, so that layer 3 covers here all communicative operations down to the physical layer(s).
Also note the subject's (in general required) interaction with a telephone book. This interaction is not elaborated further, as we focus on the communication between both partners here.

With regard to 'Flight & rail data exchange', Fig. 5 gives an information flow chart of the layer 3 message exchange. As already mentioned, the exchange plane & train data loop has been run two times, so correspondences between them can be observed.

A first observation might be that most enquiries start with the sending of context data. This can be observed at the beginning of the Flight & rail data exchange in Fig. 3, where the overall discourse context is supplied. In the layer 2 "Supply of discourse context" message first of all a sign is given that what follows should be considered as a larger piece of discourse (at layer 3: "Ik had een eh aantal vraagjes ... "), a 'Closed Discourse Unit' in the terminology of Houtkoop and Mazeland (1985), to be accomplished by the subject. After that the main question context is given (at layer 3: "over een vliegreis naar Los Angeles, die ik volgende week Dinsdag wil maken.").
The supply of question context also clearly appears in the second plane & train data exchange, where the train question is introduced and related to the earlier supplied context ("om 't vliegtuig te halen"). The context for the first plane & train data exchange already appeared in the introduction of the discourse context. Still a reference is made to that earlier mentioned context ("het vluchtnummer van die vlucht").

Subject	**Layer 2,3 interaction**	**Schiphol Info. Service**
2 Make tel. connection	**b=> Begin make tel. connection**	**2 Handle tel. conn. request**
2 => 3 Begin		2 => 3 Begin
3 Find tel. number <-> tel.book		|
3 Lift horn		|
3 Send tel. number	-> *Telephone alarm signal* - - - -	3 Receive tel.alarm signal
3 Perceive End of Alarm signal	- - - - - *End of Alarm signal* <-	3 Lift horn
2 <= 3 End	**End make tel. connection <=e**	2 <= 3 End
2 Handle discourse opening	**Begin discourse opening <=b**	**2 Send discourse opening**
2 => 3 Begin		2 => 3 Begin
3 Receive name callee	- - - - *Schiphol Inlichtingen* <-	3 Send name callee
3 Send discourse start phrase	-> *Goede morgen* - - - - - - -	3 Rec. discourse start phrase
3 Send name caller	-> *met PvL* - - - - - - - - - -	3 Receive name caller
2 <= 3 End	**e=> End discourse introduction**	2 <= 3 End

Fig. 4: Layer 3 expansion of 'Initiate tel.call Schiphol Info', see Fig.3.

Subject	Layer 2,3 Interaction	Schiphol Info. Service
2 Supply of discourse context	**b=> Begin discourse context**	**2 Receive discourse context**
2 => 3 Begin		2 => 3 Begin
3 Send discourse topic	-> *Ik had een eh aantal vraagjes over een vliegreis naar Los Angeles, die ik volgende week Dinsdag wil maken.* - - - - - - - - - - -	| | |
| | |		3 Perc. discourse topic
3 Receive mess. confirmation	- - - - - - - - - - - - - - *Ja* <-	3 Send mess. confirmation
2 <= 3 End	**End discourse context <=e**	2 <= 3 End
2 Supply of message structure	**b=> Begin message structure**	**2 Receive message structure**
2 => 3 Begin		2 -> 3 Begin
3 Sending question structure	-> *Ten eerste* - - - - - - -	3 Rec. question structure
2 <= 3 End	**e=> End message structure**	2 <= 3 End
2 Exchange plane & train data	**b=> Begin plane & train data**	**2 Exchange plane & train data**
2 => 3 Ask flight number	-> *wat is eigenlijk het vluchtnummer van die vlucht, is dat KL 401 of KL 601 ...* - - - - - - - - - - - - -	2 => 3 Begin
| (Note discourse context ref.)		|
|		3 Rec. flight number request
3 Receive answer struct. info	- - - - - - - *Ogenblikje hoor* <-	3 Send answer structure
3 Send question confirmation	-> ... *kunt U mij dat vertellen* - - -	3 Receive quest. confirm.
|		3 Retrieve requested data
|		<-> database
3 Rec. question repeat request	...*en welke dag zei U dat het was?* <-	3 Request question repeat
3 Repeat question	-> *volgende week Dinsdag* - - - -	3 Receive question repeat
3 Receive question answer	- - - - - - - - - - - - *Kl 601* <-	3 Send requested flight data
3 Send confirmation	-> *Dank U* ... - - - - - - - - - -	3 Receive confirmation
2 <= 3 End	**e=> End plane & train data**	2 <= 3 End
2 Supply of message structure	**b=> Begin message structure**	**2 Receive message structure**
2 => 3 Begin		2 => 3 Begin
3 Sending question structure	-> *En dan een tweede vraag is eh* -	3 Rec. question structure
2 <= 3 End	**e=> End message structure**	2 <= 3 End
2 Exchange plane & train data	**b=> Begin plane & train data**	**2 Exchange plane & train data**
2 => 3 Begin		2 => 3 Begin
3 Send train data context	-> *ik ben van plan met de trein van uit Tilburg te komen* - - - - - - -	|
|		3 Rec. train data context
3 Send train departure request	-> *kunt U mij vertellen hoe laat ik dan uiterlijk eh* - - - - - - - - -	|
3 Retrieve question context		|
3 Send train dep. context	-> *vanuit Tilburg moet vertrekken om 't vliegtuig te halen* - - - - -	|
|		3 Rec. train dep. request
|		3 Retrieve flight departure
|		<-> database
3 Receive flight dep. time	*nou U vliegtuig vertrekt 14 uur 50* <-	3 Send flight dep. time
3 Receive flight trav. rule	*en u moet daar anderhalf uur van te voren zijn* - - - - - - - - - - <-	3 Send flight trav. rule
|		|
3 Receive answer context	*en de trein vanuit Tilburg die doet er* ... - - - - - - - - - - - <-	3 Send answer context
|		|
3 Receive answer struct. info	- - - - - *even kijken hoor* ... <-	3 Send answer structure
|	- *die doet er ongeveer 1 uur en 40 minuten over* ... - - - - - <-	3 Send train travel rule
3 Receive train travel rule		|
|	*en die vertrekt om 2 minuten over het hele en 2 minuten over het halve uur* - - - - - - - - - <-	3 Send train travel rule
|		|
3 Receive train travel rule		|
3 Rec. answer verif. request	- *kunt U het dan zelf uitrekenen* <-	3 Send answ. verif. request
3 Send answer verif.	-> *Ja dat zal wel lukken denk ik* - -	3 Receive answer verif.
3 Receive answer structure	- - - - - - - - - - - - - - *Oke* <-	3 Send answer structure
2 <= 3 End	**End plane & train data <=e**	2 <= 3 End

Fig. 5: Layer 3 expansion of the 'Flight & rail data exchange' operation, see Fig.3.

Interestingly, the subject made an error in planning and/or execution of his question sequence: following a backward reasoning procedure after the flightnumber request, he should have asked for plane departure times, check-in durations, etc. instead of asking for train departure times immediately.

In view of his "eh" in "kunt U mij vertellen hoe laat ik dan uiterlijk eh ..." , he probably became aware of his slip, but still continued his train inquiry (as it was his main communication goal, being specified in his task description?). As can be observed in Fig.5, Schiphol Information Service corrected his flaw by communicating the missing answers before giving an answer to the posed train departure question.

Also note that by simply giving the train travel rules available, instead of an elaborated version, Schiphol Information Service looses part of its cooperative behaviour. This is done quite sophisticatedly by shaping this 'determine it yourself' request ("kunt U het dan zelf uitrekenen ..") into a verification request of correct message-reception. Also note that Schiphol Information Service is leading the dialogue there by initiating and ending this communication part.

Finally, Fig. 6 gives the layer 3 expansion of the telephone call closing procedure. Probably for reasons of politeness, the answers of Schiphol Information Service are longer than needed for straight communication ("Geen dank hoor" instead of "Geen dank" and "Daag" instead of "Dag")

Turn taking

As mentioned, the main goal of the subject is to know the latest time for leaving by train from Tilburg, while that of Schiphol Information Service is to be cooperative by providing travel information. From this situation it follows that the subject is expected to lead the dialogue, while Schiphol Information Service should follow. Characteristic for the leader is that he opens the dialogue segment concerned and frequently closes it too.
Related to turn taking, Sacks and Schegloff (1973) introduced the 'adjacent pair' as a basic concept for discourse analysis. Adjacent pairs are understood to be related speech acts that necessarily belong together, like question-answer, greeting-greeting, offer-acc-

Subject	Layer 2,3 Interaction	Schiphol Info. Service
2 Discourse closing	**b=> Begin discourse closing**	**2 End discourse closing**
2 => 3 Begin		2 => 3 Begin
3 Send discourse stop phrase	-> *Dank U wel* - - - - - - - - - -	3 Receive disc. stop phrase
3 Rec. disc. stop confirmation	- - - - - - *Geen dank hoor* <-	3 Send discourse stop conf.
|		**Note:** 'Inter pers.' mess.?
2 <= 3 End	**End discourse closing <=e**	2 <= 3 End
2 Closing tel. connection	**b=> Begin closing tel. conn.**	**2 Closing tel. connection**
2 => 3 Begin		2 => 3 Begin
3 Send tel.call stop phrase	-> *Dag* - - - - - - - - - - - -	3 Rec. tel.call stop phrase
3 Receive tel.call stop conf.	- - - - - - - - - - - - *Daag* <-	3 Send tel.call stop conf.
|		**Note:** 'Inter pers.' mess.?
3 Put down horn	-> *Tel. line interrupt signal* - - - - -	3 Rec. line interrupt signal
|		3 Put down horn
2 <= 3 End	**e=> End closing tel. connection**	2 <= 3 End

Fig. 6: Layer 3 expansion of 'Close tel.call Schiphol Info' operation, see Fig.3.

etance/refusal. However, the so called 'three-part exchanges', containing a confirmatory feedback message as the third step, occur more frequently (Weijdema, Dik, Oehlen, Dubber and De Blauw, 1982), thus improving reliable communication and also keeping the discourse initiator on lead.
The expectations about the characteristic initiatives of the leading subject are confirmed indeed at layer 0, where the subject initiates the dialogue (Fig. 2: b=> "Begin travel info.exch.") and ends it too (Fig. 2: e=> "End travel info.exch.").

Also at layer 1 (Fig. 2 and 3) we can observe these effects, both "Initiate tel. call Schiphol Info" and "Close tel. call Schiphol Info" are inititated as well as closed by the subject. Similarly the "Flight & rail data exchange" is inititated by the subject, however it is closed by Schiphol Information Service (Fig. 3: "End flight & rail data <=e").
The reason for the latter deviation can be understood by taking a closer look at the layer 3 expansion (Fig. 5) of the second "Exchange of plane & train data". It reveals that the last three-part exchange has to do with the earlier indicated somewhat non-cooperative discourse segment leaded by Schiphol Information Service ("kunt U het dan zelf uitrekenen - Ja dat zal wel lukken denk ik - Oke"). Instead of giving the required train departure time, Schiphol Information Service supplied the available rules (trains depart 2 minutes before the full and before the half hour, etc.), with the result that the subject had to calculate the needed departure time himself. Note that the subject additionally needs information here about the time needed for his transfer from the train station to Schiphol Airport. With the provided information he can only very roughly determine the train departure time himself.

In Fig. 5 it can be observed that all layer 3 "Flight & rail exchange" become initiated by the subject. The same holds for the discourse segments in the "Close tel.call Schiphol Info" interaction (Fig. 6). The only exception can be found in the "Initiate tel.call Info" segment (Fig. 4), where "Handle discourse opening" is initiated by Schiphol Information Service by mentioning callee's name. This is part of the European telephone call opening protocol that requires the callee's name to be mentioned first, in fact as a reply to the caller's telephone alarm signal. See Schegloff (1979) for an extensive elaboration of non-European caller - callee identification protocols.

Error prevention
Contrary to the American opening protocol that generally starts with some 'hello' message from the side of the caller, probably in an attempt to check out the acoustic intactness of the connection and to adapt to the speaker's voice and language used, or the preamble "you are speaking with ..." in the European protocol, here (Fig. 4) Schiphol Information Service directly identifies itself. It is not excluded that this behaviour is the result of the experimental set-up, where it is known by the information provider that the functioning of the acoustic layer as well as the language use is without problems.

During brakes in message transmission, a certain time-out is exercised by the hearer, before concluding that he is on turn. To prevent undesired turn taking, a number of behaviour protocols are in use. As indicated by Schegloff (1981), vocalizations such as 'uh huh', 'yeah' as well as head gestures are used for that purpose by the sender. In Fig. 5 a number of such turn keeping signals can be observed: "ik had een eh aantal vraagjes", "... Ogenblikje hoor ...", "En dan een tweede vraag eh ik ben van plan ", "kunt U mij vertellen hoe laat ik dan uiterlijk eh vanuit Tilburg ", "nou ... U vliegtuig vertrekt 14 uur 50", "... even kijken hoor ...".

It might be of interest to experimentally determine the values of these time-outs, to investigate the influence of turn keeping signals and to see whether the durations are possibly larger for the higher layers of communication.

Probably the "... kunt U mij dat vertellen" represents a perseveration of the subject being interrupted by Schiphol Information Service ("Ogenblikje hoor"), or just serves as a pause filler.

Finally, also the repair of a flaw in (short-term) memory processing can be observed in Fig. 5: "en welke dag zei U dat het was?". This question indicates that the informant of Schiphol Information Service remembered that the subject earlier mentioned the day of flying ("Ik had een aantal vraagjes over een vliegreis naar Los Angeles, die ik volgende week Dinsdag wil maken"), but already forgot the precise day.
Note that because of inaccuracy in the repair question (it is unclear whether the relevant week was also forgotten), the subject also repeats that information, to prevent further proliferation of the problem.

General discussion

Layered model

While transcribing the dialogue into a layered structure, we noted that in general not all interaction will be with the other partner. For example, the layer-3 expansion of 'Initiate tel.call Schiphol Info' in Fig. 4 contains at the subject's side an interaction with the telephone book for finding the correct telephone number as well as one with the telephone apparatus for setting-up the telephone connection. In fact, it has to be assumed that both partners possess intention hierarchies of which only a minor part is related to communication with the other. A dialogue with e.g the telephone operator might give rise to a communication hierarchy comparable to the one with Schiphol Information Service, however, related in general to a lower layer in the subject's 'main intention' hierarchy. As a result communication between partners might occur among layers at different depth in the partners' main intention hierarchies. This asymmetry effect might also occur if the partners' communication goal structures differ in the amount of detail.
In our analysis we have tried to eliminate this asymmetry effect by introducing a certain layering in the abstractness of the messages exchanged, see Fig. 1, and to adapt the assumed layers of intention of both partners accordingly.
In dialogue practice, errors in inter-human as well as man-machine communication frequently reflect imperfections in the mutually matching of layers. Partners then cooperate by trying to fit their intention layers to each other during the conversation (e.g. by providing information about dialogue structure, see Fig. 5). It might be of interest to investigate the partners' tendency of matching coding / decoding levels experimentally by introducing mismatches in layering.

In this report little attention has been given to the 'below utterance' layers of communication. At these lower layers, intentions are less conscious and communication proceeds more automatically and often over different physical channels in parallel. Here, the higher levels also have to take into account the lower level communication facilities. Compare for instance the difference in coding redundancy needed for communication via written mail against that used for interaction via speech and gesture.
Probably these low-layer communication hierarchies are better defined bottom-up as the

available physical communication channels are assumed to dominate the message encoding there. Probably for similar reasons, Taylor (1988) distinguished layered 'internal processes' in parallel to the layered communication processes. However, to prevent the complexities arising from such opposite parallel hierachies, a simplification relevant for experimental verification might be to consider the above indicated model, in which the layers are grouped into two major categories, viz. a class of 'intention driven' higher layers and a class of communication or 'channel driven' lower layers.
In written conversation, characters, syllables, and words constitute the channel driven communication segments, while paragraphs and chapters might be the units of communication at the intention driven layers. Sentences are shared by both categories.

As illustrated in our telephone call transcription, also in spoken dialogues discrimination of discourse segments larger than single phrases and even larger than conversational turns are important structural elements. Polanyi and Scha (1984) acknowledge in their outline of a discourse grammar several levels of structure, constructed by means of sequencing and recursive nesting of discourse constituents. Amongst others they distinguish the 'topic' a discourse is dealing with, 'discourse units' (DU's) having a recognizable "point" or purpose (stories, descriptions, jokes, etc.), and 'discourse constituent units' (dcu's) containing consecutive clauses combined into one syntactic / semantic unit (lists, narrative structures, binary structures like "A because B", etc.). They consider the adjacency structures like Sacks and Schegloff's (1973) adjacent pairs to be a kind of discourse constituent units.
Recently, Scott and Sieckenius de Souza (1989) supplied certain text structure rules relevant for the ease of understanding rhetorical structure. These rules prevent the generation of e.g. a structure like "Since Owen has an American passport, he, who was born in Jamaica, is an American citizen" (instead of the more transparent structure "Since Owen, who was born in Jamaica, has an American passport, he's an American").

In their attempt to find explanations for interruptions and referring expressions such as anaphora, Grosz and Sidner (1986) assume the linguistic discourse structure to be composed of two additional components, an intentional structure, and an attentional state. Their attentional state, modelled as a 'focus space', contains information about objects, properties, relations and discourse intentions that are most salient at any given point in the discourse. It is assumed to relate to the current point of execution in the partners' columns of our layered information flow charts, together with the at that moment active higher level intentions covering the current low level action.
According to them, boundary markers such as the explicit use of certain phrases, intonation cues, or changes in tense and aspect indicate transitions between 'discourse segments' in their linguistic structure. Each of these discourse segment is assumed to be the linguistic realisation of a socalled 'discourse segment purpose' (dsp). In terms of Taylor's (1988) layered protocol model, the linguistic structure relates to the syntax of the at a certain layer incoming message, while the intentional structure has to do with the semantics of the decoded message. Both hierarchical structures are of course closely related. Only in the case of exceptions, such as when protocol messages have to be exchanged for handling errors at a certain level of communication, the two hierarchies might differ. In our analysis we made no distinction between both hierarchies. The question is in fact whether the intentional structure is assumed to be a static beforehand planned task hierarchy, or a dynamic structure adaptable at each moment to incoming events. We have chosen for the latter idea.

By showing that people are quite good in reconstructing the correct order of individual utterances as produced in certain types of dialogue, Beun (1989) suggested an experimental method for the study of intentional coherence between utterances in a dialogue. It might be of relevance to use this method for studying discrepancies between intentional and linguistic structures on coherence perception, for instance by opposing the intentional effects to the influences of artificially added linguistic boundary markers. Alternative coherence determination methods are indicated by Johnson-Laird (1983).
With regard to the attentional focus space, it may be worthwhile to investigate the possible relation between the allowed distance of the anaphoric/cataphoric reference around the current point of execution and the related distance in layering. Is it acceptable, or what type of precautions are needed, to refer to entities more than e.g. three layers remote?

Recognition errors
As mentioned in the introduction, communication errors might arise from disturbances in the physical communication channel as well as from errors in coding and decoding of the messages. Message decoding or goal recognition is seen as a receiver process, by which aspects of the incoming submessages are used to categorize the entire message into one out of a set of possible interpretations. In this way, 'how' becomes converted back into 'what'. In general, the relevant interpretations will be preselected by expectations of the receiver, based on earlier perceived message context and through redundancy offered by the applied conversational protocol, thus deminishing the chance of generating irrelevant alternative interpretations.
More precisely, at each layer the receiver process is assumed to group the incoming message entities, along with their individual semantic attributes and their mutual structure (the syntax) into a higher level linguistic unit again with its specific semantic interpretation. For instance, the incoming letter units, with semantic attributes {a,b,k,n,~} and the following temporal structure: " ~ b a n k ~ ", are *detected* according to the linguistic syntax as a 'word' entry suitable as input to the word-lexicon. After word *classification* (recognition), the word-lexicon then provides the possible word meanings. In general, the lexicon gives several semantic interpretations for the recognized entry. For instance, "bank" might refer for instance to a building in the context of cities, to an organization in the financial context, to an obstacle for skippers, or to a piece of furniture in the parc. Message *interpretation* then consists of assigning the most probable interpretation to the entry concerned, by means of the available context (knowledge about the world concerned, discourse history, expectations following from message anticipation, etc.)

Generally speaking, the evidence for a specific classification will grow with time, in such a way that after some period a given class has received most as well as sufficient positive evidence for a classifying decision. In case of incoming errors, the accumulated evidence might generate a false classification or after some time out might be insufficiently strong for input classification. In the latter case, the receiver might initiate an error handling procedure, e.g. by asking for an explanation. Another error possibility is that a wrong interpretation set (context) becomes chosen because of incorrect expectations.
In that case the receiver might not discover the interpretation (semantic) error until a higher recognition process solves this problem by using available redundancy in the incoming messages to find as yet the correct interpretation, or by still starting a slower, higher layer error correction dialogue. It might be of interest to model these layered

speed/accuracy trade-offs and to investigate whether human dialogues are efficient in this respect.

Error classification
Related to the study of speech production mechanisms, errors in verbal communication have been widely studied, see e.g. the collection of articles in Fromkin (1973, 1980).
Also Levelt (1983), in his study of self-repair in speech monologues, categorized errors occurring in the planning and/or expression of utterances. Besides an unclarified rest category, he distinguished a class of A-repairs about the 'appropriateness' of the applied idea formulation against a class of E-repairs, having to do with lexical, syntactic and fonetic 'errors' the speaker discovers saying "often to his own surprise". The A-repairs seem to be related to the intention driven layers, while the E-repairs correspond to the more or less autonomous channel driven layers.
More general than just for speech production, Norman (1981) categorized what he called action slips. He distinguished three major categories of slips, viz. errors in the formation of intentions, faulty activation of action 'schemas' (action sequences controlled by sensori-motor knowledge structures) and errors in their temporal triggering. Here again the difference is made between intention and channel driven activities.
In view of the assumed layered processing, we expect that in inter-human communication, faulty activation of action schemas should be discovered earlier by the receiving partner then errors in the formation of intention. However, if the receiving partner is already aware of the intentions to be received, it might be that through the receiver's expectations intention planning errors are detected faster than possible underlying channel oriented errors.

Intelligent agents
The intention oriented layering in communication between men, suggests that such a layering might be of relevance also in man-machine communication. As indicated already in the introduction, certain designers claim indeed that optimal man-computer interfaces should be layered, implying the use of layered user commands as well as layered system feedback messages.
In man-computer interfaces layered commands are already applied for long, for instance by the use of intention oriented windows ('dialogue boxes'), simultaneously containing low layer form and menu like facilities with default parameter values for specifying the desired (higher layer) message to the computer.

With regard to communication from computer to user, layering is less far reaching. The bad example being "syntax error" as a kind of mystery message covering a large range of differently layered user input errors. Good examples of layered computer feedback are exceptional. This probably has to do with difficulties in goal recognition at the computer side. It is expected that this capability of goal recognition will improve with the growing of the computer from universal machine towards an application specific machine. For a first step in this direction, see the work of Jerrams-Smith (1989), whose "Smart User-System Interface" provides advice based on recognition of novice user error categories with Unix.
Intelligent help is assumed to become a major research topic in the field of man-machine communication. Given certain input errors, the machine should come up with layer specific error messages, adapted to the needs of the user. The latter aspect assumes that the computer maintains knowledge about the user's intentions and believes, the task at

hand, the interaction history, etc. For efficiency reasons, probably the this 'background knowledge' should be layered too. As a preliminary in this development, research could be done on 'agents' (communication mediators) with capabilities, up to a certain level of communication.

Conclusions

For improving the design of 'intelligent' multi-modal man-machine interfaces, we tried to unravel characteristics of normal inter-human communication.
In the previous sections we indicated how we structured an earlier recorded telephone information dialogue by repeatedly asking oneself what the message is to be exchanged and how this would be accomplished. From the how question then followed a more detailed chain of subgoals about what to achieve at the next lower level in the communication hierarchy. From the subject's goal to obtain certain information it then followed for instance that first a telephone connection should be made before the actual conversation with the informant can take place. So, at each moment during the performance of an interaction, multi-layered intentions are assumed to be active. It is the receiver's task to recognize these multiple intentions of the sender.

With regard to those tasks directly related to the vocal message exchange, the what-how approach supplied us with a hierarchical discourse description, from which certain stereotypical patterns of intention realisation evolved. For instance, for better understanding the speaker's aims, a certain amount of context is supplied before the actual question is put (see Fig. 5). The same idea holds at other levels of intention, for instance context is also supplied as an intro to a larger set of questions to be posed. At the receiver's side, context is assumed to help in recognizing the sender's intentions by disambiguating possible message interpretations or by emphasizing beforehand certain preferred categories of interpretation.
In current computer application programs context is supplied by the user only in a very limited form, e.g. by providing his user identification. Nevertheless, this already restricts significantly the for the interaction relevant file domain. Also by specifying the desired application program to run, a certain context becomes activated. However, this context is not user but programmer created, with the inherent problem of lack in familiarity to the user. The further application of user specified context might improve man-computer interaction.

As mentioned, it is the receiver's task in an interaction to recognize the sender's multiple intentions. It has been suggested most explicitly by Taylor (1988) that for efficient communication, this recognition process should be mirrored with respect to the sender's encoding processes. Accordingly, man-computer interfaces should be layered too. The communication 'protocol' of the given layer then defines how the message content is to be encoded at the sender's side and extracted again at the receiver's side. Besides knowledge about message form, protocol also includes layer specific feedback procedures for the correction of errors in transmission. For the recognition and further processing of perceived user intentions, the system's program has to be provided not only with layer specific knowledge for message recognition and related error handling, but also with (layer specific?) knowledge about the environment, the user, the dialogue history and the task at hand.

Our dialogue analysis also provided us with a number of research issues to be further pursued, for better understanding where to care for in multi-layered man-machine interfaces. Besides errors in transmission, imperfections in the mutually matching of layers between human partners frequently occur. Therefore, the interactive fitting of intention layers with those of a computer system or of another human partner in a dialogue, constitutes an important issue of further research. Closely related to this point is the further study of referring expressions (which might besides textual also be of e.g. manual origin) and intentional coherence. Finally, the study of layered speed/accuracy trade offs in communication, related to the use of multiple physical communication channels with different error rates, are worth to be considered more closely.

Acknowledgements

This report largely benefited from discussions with Joyce van de Vegte and Martin Taylor from DCIEM, North York and Jack Edwards from Artificial Intelligence Management Corporation, Toronto. Also the support of our IPO colleagues Robbert-Jan Beun and Kees van Deemter is greatly appreciated.

References

Beun, R.J. (1985) The function of repititions in information dialogues, IPO Annual Progress Report, 20, p 91-98.

Beun, R.J. (1989) Personal communication.

Bunt, H.C., van Katwijk, A.F.V., Muller, L.F. and van Nes, F.L. (1980) Dialogue control acts, IPO Annual Progress Report, 15, p 95-99.

Buxton, W. (1983) Lexical and pragmatic considerations of input structures, Computer Graphics, 17, p 31-37.

Cramer, Y.M. (1985) Transcriptie dialoogexperiment juni 1985, IPO Rapport no 513, p1.

Fromkin, V.A. (1973) Speech Errors as Linguistic Evidence. The Hague: Mouton, p 243-269.

Fromkin, V.A. (1980) Errors of liguistic performance: Slips of the tongue, ear, pen, and hands. New York: Academic Press.

Grosz, B.J. and Sidner C.L. (1986) Attention, Intentions, and the structure of Discourse, Computational Linguistics, 12, p 175-204.

Houtkoop, H. and Mazeland, H. (1985) Turns and discourse units in every day conversation, Journal of Pragmatics, 9, p 595-619.

Jerrams-Smith, J. (1989) An attempt to incorporate expertise about users into an intelligent interface for Unix, Int. J. Man-Machine Studies, 31, p 269-292.

Johnson-Laird, P.N. (1983) Mental Models. Cambridge: Cambridge University Press, p 356-395.

Levelt, J.M. (1983) Monitoring self-repair in speech, Cogniton, 14, p 41-104.

McCracken, D.D. (1965) A Guide to Fortran 4 Programming, New York: John Wiley.

Nielsen, J. (1986) A virtual model for computer-human interaction, Int. J. of Man-Machine Studies, 24, p 301-312.

Norman, D.A. (1981) Categorization of action slips, Psychological Review, 88, p 1-15.

Norman, D.A. (1984) Stages and levels in human-machine communication, Int. J. of Man-Machine Studies, 21, p 365-375.

Polanyi, L. and Scha, R. (1984) A syntactic approach to discourse semantics, Proceedings of the International Conference on Computational Linguistics, Stanford University, Stanford, Cal., July 2-6, p 413-419.

Pruest, H.H.A.M., Minnen, G.A.G. and Beun, R.J. (1984) Transcriptie dialoogexperiment juni/juli 1984, IPO rapport no 481.

Sacks, H. and Schegloff, E.A. (1973) Opening up and closings, Semiotica, 8, p 289-327.

Schegloff, E.A. (1979) Identification and recognition in telephone conversation openings; in Psathas, G. (Ed.) Everyday Language Studies in Ethnomethodology, New York: Irvington Publishers Inc., p 23-78.

Schegloff, E.A. (1981) Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences; in Tannen, D. (Ed.) Analyzing Discourse, Washington: Georgetown Universiry Press, p 71-93.

Scott, D.R. and Sieckenius de Souza, C. (1989) Getting the message across: a computational grammar of text, Philips Laboratoires Redhill, Technical Note no. 2885.

Tainsh, M.A. (1985) Job process charts and man-computer interaction within Naval command systems, Ergonomics, 28, p 555-565.

Tanenbaum, A.S. (1981) Computer Networks. London: Prentice Hall International, p 10-21.

Taylor, M.M. (1988) Layered protocols for computer-human dialogue. I: Priciples, Int. J. of Man-Machine Studies, 28, p 175-218.

Weijdema, W., Dik, S., Oehlen, M., Dubber, C. and de Blauw, A. (1982) Strukturen in Verbale Interaktie, Muiderberg: Coutinho.

Woodworth, R.S. and Schlosberg, H. (1965) Experimental Psychology, New York: Holt, Rinehart and Winston, p 8-42.

Appendix A

English translation

In Fig.7 a translation in English is given of the applied telephone dialogue. Because of the difference in syntax between English and Dutch language, the utterance timing patterns available in the original Dutch transcription are not precisely indicated here.

T = Subject
S = Schiphol Information Service
... = Pause

S: Schiphol Information
T: Good morning, P.v.L. speaking, I had a uh number of questions about a flight to Los Angeles that I want to make Tuesday next week.
S: Yes.
T: First, what is in fact the flight number of that flight, is it KL 401 or KL 601,
S: ... A moment please.
T: ... can you tell me that?
S: ... And what day did you say it was?
T: Tuesday next week.
S: KL601
T: Thank you, and then a second question, is uh I have the intention to take the train from Tilburg, can you tell me at what time uh I have to leave Tilburg at the latest to catch the plain?
S: Well, your plane is leaving at 14h50, and you have to be there one and a half houre before, and your train from Tilburg, he takes ... just have a look ... it will take about one hour and fourty minutes ... and he departs two minutes after the full hour and two minutes over the half hour, can you determine it yourself now?
T: Yes, that will succeed I think
S: OK
T: Thank you very much
S: No thanks
T: Bye
S: Good bye

Fig. 7: English translation of the Dutch telephone information dialogue used for the layered dialogue description