Crook, N, Field, D, Smith, C, Harding, S, Pulman, S, Cavazza, M, Moore, R and Boye, J
Generating context-sensitive ECA responses to user barge-in interruptions.

Crook, N, Field, D, Smith, C, Harding, S, Pulman, S, Cavazza, M, Moore, R and Boye, J (2012) Generating context-sensitive ECA responses to user barge-in interruptions. Journal on Multimodal User Interfaces, 6 (1-2). pp. 13-25.
Doi: 10.1007/s12193-012-0090-z

Generating context-sensitive ECA responses to user barge-in interruptions

Nigel Crook, Debora Field, Cameron Smith, Sue Harding, Stephen Pulman,
Marc Cavazza, Daniel Charlton, Roger Moore, Johan Boye

Abstract

We present an Embodied Conversational Agent (ECA) that incorporates a context-sensitive mechanism for handling user barge-in. The affective ECA engages the user in social conversation, and is fully implemented. We will use actual examples of system behaviour to illustrate. The ECA is designed to recognise and be empathetic to the emotional state of the user. It is able to detect, react quickly to, and then follow up with considered responses to different kinds of user interruptions. The design of the rules which enable the ECA to respond intelligently to different types of interruptions was informed by manually analysed real data from human–human dialogue. The rules represent *recoveries* from interruptions as two-part structures: an *address* followed by a *resumption*. The system is robust enough to manage long, multi-utterance turns by both user and system, which creates good opportunities for the user to interrupt while the ECA is speaking.

1 Introduction

In human–human dialogues, when a person is interrupted by another participant, the interruptee is able to halt what he was saying, to consider carefully the content of the interruption, and to respond to it, without forgetting what he was planning to communicate before he was interrupted. Having the goal of building an affective, sympathetic ECA, we decided to devise and implement a way of modelling this more human-like way of handling interruptions in our ECA system.

The 'How Was Your Day?' (HWYD?) prototype ECA discussed here was developed as part of the COMPANIONS project [1]. It is an affective dialogue system aimed at social conversation with the user about his working day at the office. The overarching goal of the HWYD? prototype in its social conversation is to develop and maintain a supportive relationship with the user. We would argue that the conversational rules and goals of social dialogue are more evasive and obscured from participants' rational awareness than those of task-based dialogues. Our system design therefore needed to account for phenomena that task-based spoken language dialogue systems (SLDs) (*e.g.,* [2, 3, 4]) are not typically concerned with.

The HWYD? system works towards its supportive goal by analysing and responding to acoustic, emotional, and linguistic information from the user's spoken turns. It uses backchannelling and gesture to show sympathy as it listens to the user's difficulties, and it also gives relevant encouragement and advice. It also attempts to mirror the user's perceived emotional state through facial expressions and tone of voice.

The supportive nature of the ECA motivated us to try to devise and implement appropriate system responses to user interruptions that went beyond simply stopping speaking, since human–human interruptions can be an indicator that a supportive relationship is failing [5, 6]. The HWYD? system is able to detect user barge-in, to react very quickly to it through facial expression and other gestures, and to respond verbally to the content of the user's interruption. Its response to barge-in relies on an analysis of the pragmatic and semantic content of the user's interruption.

The system has several different ways of responding to different types of interruption. The design of the rules determining this system behaviour was informed by manual analysis of real-world data containing human–human interruptions.

A characteristic feature of the HWYD? prototype is that it is robust enough to process long, multi-utterance user turns, and to enable the system to also produce long turns. The user has plenty of opportunity, therefore, to interrupt while the ECA is speaking.

2 Previous work

In the linguistics literature an interruption is a type of turn in a conversation [7]. A turn constitutes an interruption when "The next speaker begins to speak while the current speaker is still speaking, at a point in the current speaker's turn which could not be defined as the last word. Interruptions break the symmetry of the conversational model; the interrupter prevents the speaker from finishing their turn, at the same time gaining a turn for themselves" [8, p. 109], which we follow in this work. We also follow the now largely accepted view that interruptions are distinct from backchannelling

[9, 10, 11] and we also follow [7, 12] in considering interruptions a special case of simultaneous speech.

The noun phrase *barge-in* appears to have a very similar meaning to *interruption* from linguistics, but the term *barge-in* is only used when conversation with a machine is being discussed (in contrast to human–human conversation). This paper will discuss human–machine conversation as well as human–human conversation, and so both terms will be used accordingly. The noun *interrupt* will also be used. We have observed that the noun *interrupt* is generally used where acoustics technology is concerned, and refers to overlapping sounds that may or may not be speech, or made by a person, and if they are made by a person, they may or may not constitute an interruption.

Much of the previous work on handling user barge-in has been done in the context of task-oriented dialogues such as telephony-based spoken language systems [13, 14]. It falls into two broad approaches [15]: those that analyse the acoustic signals from the user's speech to detect features such as prolonged intensity or voicing, which indicate a sustained vocal output from the user [16]; and those that use ASR-based methods such as language models for barge-in detection [17, 18, 19]. The method of barge-in detection presented in this paper is based on an analysis of the acoustic signals of the user's speech, and so falls into the first of these groups.

Many SLD systems are described as supporting barge-in interruptions [17, 18, 19]. In those cases, *barge-in* usually refers to the ability of the system to stop issuing a prompt if the user starts to reply verbally to the system before it has finished issuing the prompt. This is a much simpler approach to handling user interruptions than the more human-like one we wanted our system to emulate. The approach to interruption handling implemented in the HWYD? prototype and presented here was inspired by cognitive architectures such as Rodney Brooks' subsumption architecture [20, 21] in line with the forward-looking proposals for future speech-based human-machine interaction outlined by Moore [22].

Recent work on interruptions has focussed on *listener responses* and *active speaking*. In [23] the design of an interactive Virtual Human (VH) is presented that seeks to combine conversational speech with gestures. The emphasis here is that while the VH is speaking it should be attentive to anything that the user says. If the user does say something whilst the VH is speaking, then the VH attempts to identify the competitiveness and cooperativeness of the user's response. If the user's response is classified as competitive, then the VH identifies this response an an interruption. The VH and the HWYD? prototype have a number of features in common: both

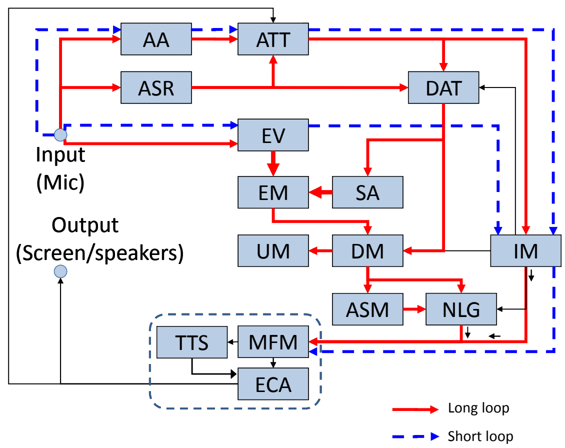

Figure 1: A schematic diagram of the system architecture showing long and short loops (the abbreviations used here are explained in Section 3).

combine conversational speech with gestures; both use the pitch and duration of user utterances to detect interruptions and to differentiate these from *backchannel* (see Section 4.1); both attempt to respond appropriately to user interruptions which includes replaning system utterances in some situations (see Section 5.2.2). In contract to the HWYD? prototype, however, VH does not take into account the different types of *barge-in* interruptions that occur in conversational speech between humans, neither does it have appropriate strategies for responding to each interruption type.

3 Overall system design

The HWYD? prototype comprises 15 modules that communicate using XML encoded messages through a loosely-coupled multi-hub architecture [24] (Figure 1). The user is fitted with a close-talking noise-cancelling microphone which enables a good separation of his voice and the system's text to speech (TTS) output. Signals from the user's speech are simultaneously processed by the Automatic Speech Recogniser (ASR—Nuance's Dragon NaturallySpeaking); by the Acoustic Analysis module (AA), which extracts low level features from the acoustic signal (pitch, average intensity, and probability that a section of auditory input is speech); and by EmoVoice (EV) [25], which analyses the arousal (active, passive) and valence (positive, negative, neutral) of the user's emotional state from the acoustic properties of his speech.

The Acoustic Turn-Taking module (ATT) also analyses the acoustic signals and decides when the user starts and stops speaking. The Dialogue Act Tagger (DAT) segments the ASR output and labels each segment with a dialogue act. These labeled segments are then processed by the Sentiment Analysis (SA) module [26], which tags each for positive, negative or neutral valence. The Emotional Model (EM) fuses the output of Emovoice with the output of the SA. The output of the EM is then passed on to the Dialogue Manager (DM), which extracts semantic information from the user's utterances. The DM identifies the principal topics of the user's utterances and tags them with the fused emotion label that it received from the EM. This information is then passed on to the Affective Strategy Module.

3.1 Affective Strategy Module design

The HWYD? prototype seeks to provide a conversational tone to the dialogue with the user by both attempting to understand lengthy, narrative user utterances and by replying in kind. These replies require both an understanding of the topics discussed and an appreciation of the emotional state of the user. The role of the Affective Strategy Module (ASM) is to take the information gathered for both of these aspects and to generate an appropriate narrative response, commonly in the form of a long, multi-utterance turn to be spoken by the ECA. The generation of long system turns by the ASM creates good opportunities for user barge-in, and these long ECA utterances may even provoke a user interruption since they often advise the user about how to deal with difficult or stressful situations that they encounter.

The ASM works to create a plan for a narrative response that will positively influence the user's emotional situation. It does so by first appraising the user's situation, including both the topic of discussion and the user's emotional response to this, and then using this to select an affective strategy which will guide the generation of the plan. For example this might consist of the selection of a reassuring strategy if the user shows an overly negative outlook on events. Our approach is inspired by Bremond's narrative theory of influence [27], in which a character's expectation of a given outcome can be used as a basis for influencing that character. For example, a character anticipating a loss is more readily influenced by discussion of how that loss can be averted or reduced.

The ASM employs a Hierarchical Task Network planner with a heuristic-based selection process to generate the plan for the narrative response, allowing the appraisal information to determine the plan while still affording variability in the plan's composition [28, 29]. The ASM's plan forms the input to the Natural Language Generator (NLG), which outputs surface form to the ECA (Figure 1) for performance to the user.[1]

3.2 The appraisal process

There are three principal factors which are determined by the appraisal process and consequently determine the selection of an affective strategy: the event type, the anticipated outcomes and the users emotional reaction. Each of these are decided in relation to the central topic of conversation, which becomes the main event, with any further details acting in support of this. The event type indicates the effect, positive or negative, that the main event will have on the user in terms of being either an improvement on their situation, such as a pay rise, or a deterioration of their situation, such as a pay cut. In situations where the event has not reached a conclusion (namely when a future event is discussed, such as potential redundancies), the anticipated outcomes contrast the users anticipation of what will happen with that of the agent and depends on the agents own predefined outlook (as either optimistic or pessimistic). Finally, the information from the EM (as a temporal fusion of information provided by the SA and EmoVoice modules) is compared to the event type to determine whether the users reaction is appropriate or inappropriate in terms of being emotionally consistent with it. These factors are then used to determine which of the six possible affective strategies are selected. Where anticipations match, a sympathetic strategy is selected for deteriorations and a congratulatory strategy for improvements. Where anticipations differ, a reassuring strategy is selected when the user is too negative and a cautionary strategy is selected when the user is too positive. Where the emotional reaction is judged inappropriately positive, a warning strategy is selected (such as with Example 1) and an encouraging strategy when inappropriately negative. The chosen affective strategy then influences the operator selection during the generation of the responses plan.

3.3 Some illustrative system–user dialogue samples

The ASM has three main considerations when constructing a plan. To illustrate, Example 1 shows actual system output from a user session.[2] The first consideration (1) deals with the user's emotional state and provides an

[1] Telefonica I+D in Madrid have designed the skin, lip-synching, and gestures for the HWYD? avatar (that we call *Matilda*)[30, 31], which is driven by a Haptek engine. Loquendo in Turin have innovated emotional speech for the HWYD? TTS module [32].

[2] All inter-component messages sent and received during run-time are recorded in a log file.

Example 1 A long turn ('tirade') planned by the ASM to influence the user's attitude

Try not to get too excited.	(1)
The takeover sounds terrible.	(2i)
Are you certain you have the right idea regarding the takeover being a good situation?	(2ii)
It's likely that your workload could be affected due to the merger.	(2iii)
It's obvious that mergers although unavoidable are likely to be bad.	(3)

appropriate response to acknowledge this; the second (2) provides comments on the main topic based on the selected affective strategy; and the third (3) provides a summary to reinforce the chosen affective strategy. The second area can be divided into further areas commenting on different aspects of the user's situation: comments on the appropriateness of the user's emotional reaction (2i); comments on the appropriateness of the user's anticipated outcome (2ii); comments on how the user should respond to various influencing factors (2iii); as well as comments on how the user should respond to the anticipated outcome.

Example 2 shows angry barge-in from an actual system-user session.[3] The beginning of overlapping speech is marked by a right- pointing arrow $\rightarrow$ (user barge-in turn (c)), and by a down-pointing arrow $\downarrow$ (the system's interrupted turn (b)).[4] The system log records the turn and segment number during which the ECA stopped speaking (here (b3)). Although the log does not yet record the last word that the ECA pronounced, we manually observed during this user session that the ECA's last pronounced word was 'them' in utterance (b3) of the ECA's tirade. The subsequent unspoken system words are shown in the example in italic font. Those words, though not uttered by the system, were nevertheless planned by the ASM, and so they appear in the log.

[3]The user's turns are reproduced exactly as was output by the ASR, and therefore include some ungrammatical structures.

[4]Recall that we equate an interruption with overlapping/simultaneous speech (section 2).

Figure 2: The ECA

Example 2 An irritated user interrupts the ECA

U(a)	Of I'm afraid I may make the made redundant	(ASR out)
S(b)	It's good to not get ahead of yourself.	(b1)
	Looking at the redundancies from a different perspective might help.	(b2)
	Take care as your you could lose your motivation because of ↓ *[them.*	(b3)
	It seems clear that redundancies although unavoidable are likely to be bad.]	(b4)
U(c)→	I have lost my motivation already don't tell me that	
S(d)	How likely is it that this may happen?	

Table 1: Stages of managing a barge-in interruption

Stage 1:	ECA outputs an utterance
Stage 2:	User interrupts ECA
Stage 3:	Interruption Manager (IM) pauses ECA output
Stage 4:	MFM informs DM of where the ECA's utterance was stopped
Stage 5:	ATT signals the end of utterance
Stage 6:	IM and DM respond to the interruption
Stage 7:	ASM continues, replans or aborts the interrupted utterance

4 Architecture for handling user barge-in

The handling of user barge-in interruptions involves the cooperative effort of several modules. This is achieved through a staged message passing process. Barge-in is thus handled in the seven stages shown in Table 1. These roughly chronologically-ordered stages will be discussed in more detail in the following subsections.

4.1 Detecting user barge-in

The first problem to tackle in handling barge-in is to detect when a genuine user interruption has taken place. This requires the system to be able to differentiate an interruption from other acoustic phenomena such as sudden extraneous noises (like a door banging), as well as from backchannelling. In the HWYD? prototype, the Acoustic Turn Taking (ATT) module is informed whenever the ECA starts or stops speaking (Stage 1 above). If the user begins to speak while the ECA is still speaking, the ATT makes the decision as to whether this was a genuine interruption.

An interrupt is classed as barge-in when the intensity of the acoustic signal is such that the user is believed to be talking at the same time as the ECA, and has talked for more than a certain duration or above a set intensity threshold (the Interrupt Threshold). These conditions aim to avoid treating user backchannels as interruptions.

Figure 3 shows the conditions under which interruption detection occurs. A *Talking Threshold* is used to determine whether the user is speaking at all, while the *Interrupt Threshold* is used to determine whether the user is speaking at a high enough intensity to qualify for an interruption. The

Table 2: Dialogue Interruption Cases

Case 1:	ECA talks, then there is a pause, then user talks (*i.e.*, normal scenario)—not an interruption.
Case 2:	ECA talks, user starts talking over the ECA. But user's speech signal is temporally short with low intensity—user is assumed to be back-channelling, so not an interruption.
Case 3:	ECA talks, user starts talking loudly or shouting (whether temporally short or not) over the ECA—identified as an interruption.
Case 4:	ECA talks, user starts talking over the ECA and this is over a relatively long duration (longer than *Interrupt Duration C*)—identified as an interruption.

Interrupt Duration C is used to determine whether the user has talked over the ECA for long enough to constitute an interruption. Note that either of the conditions `high enough` or `long enough` will trigger the recognition of an interruption.

The four cases shown in Figure 3 correspond to the dialogue situations described in Table 2. If the ATT concludes that a genuine user interruption has taken place, then it informs the Interruption Manager (IM) module of this (Stage 2 above). When the ATT detects that the user has stopped speaking, the ATT informs the IM that the user's interrupting utterance has ended (Stage 5 above).

An analysis of the system logs of dialogue sessions with three different users that took place in a noisy conference hall reveals that 54 out of 64 user interruptions (83.7%) were correctly identified as interruptions by the HWYD? system. The remaining 10 false positive interruption detections (16.2%) were caused either by background noise, non-verbal sounds made by the user, or short low-intensity utterances made by the user.

4.2 A short loop for timely system reactions

Once a user barge-in has been detected, in order for the system's reaction to appear natural, it is necessary for it to very quickly acknowledge the interruption in some way. For example, when the user interrupts a system

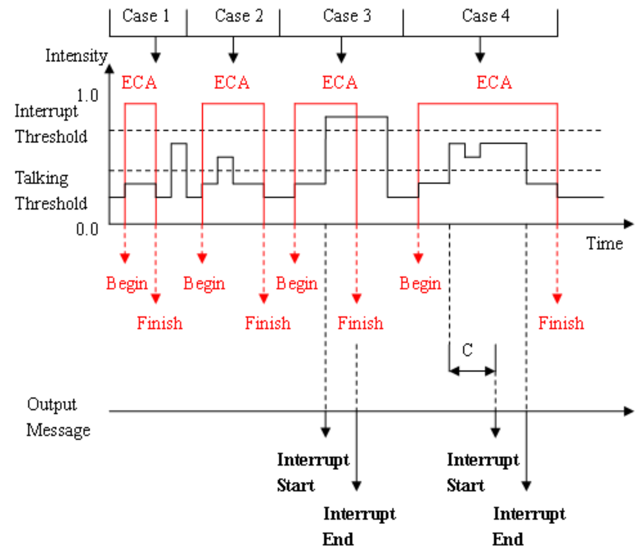

Figure 3: Interrupt detection.

turn, the system should stop speaking that turn with human-like timing and skill, and certainly should not continue speaking as though there has been no barge-in.

The need for fast reactions to user barge-in was what initially brought the Interruption Manager into the HWYD? system's design. The system architecture uses two distinct processing loops. The 'long loop' constitutes the normal full processing of a user turn, which invokes every system module in the manner described in section 3 (solid arrows in Figure 1). The average duration of the long loop measured from when the user stops speaking to when the ECA starts to speak is 1.6 seconds (this rises to an average of 2.8 seconds when the ECA is generating long multi-utterance turns). The 'short loop' (dashed arrows in Figure 1) bypasses the linguistic modules that attempt to understand the meaning of the user's utterances and to generate system responses from that understanding. Consequently the 'short loop' takes less time (about 20 milliseconds in total) to process a system response.

When user barge-in occurs, the Interruption Manager invokes the short loop to enable the system to exhibit an initial reaction to the barge-in within a naturally short timeframe. (The short loop responses are not only employed for reacting to barge-in, they are used continually throughout the dialogue to implement ECA gestures and verbal backchannels within a very short time of the end of a user turn.) The choice of particular gesture and verbal backchannel used in the short loop depends on EmoVoice's classification of the valence of the user's interrupting utterance (positive, negative, neutral). Each short loop gesture may or may not be accompanied by verbal backchannel (randomly decided). The backchannel utterances that are generated are randomly selected from one of three sets of brief utterances corresponding to positive (*e.g.*, 'yes', 'sure'), neutral (*e.g.*, 'I see', 'Ok') and negative (*e.g.*, 'I understand', 'Oh') sentiment, chosen to align with EmoVoice's classification of the user's interrupting utterance. Following the system's initial fast reactions generated by the short loop (stopping speaking, showing surprise in gesture and facial expression, verbal backchannel, ...) processing of the long loop completes, and the system gives a verbal response which *has* been processed by the linguistic modules and takes into account the linguistic content of the user's interruption.

The asynchronous nature of the two loops can sometimes lead to incongruities between the backchannel generated by the short loop and the next immediate system utterance generated by the long loop (*e.g.*, 'I see', followed by 'What do you mean?'). The design of the HWYD? system, in common with the design of many SDSs, assumes that the utterance is the minimal unit of processing spoken input [36]. We acknowledge that incremental dialogue processing could have helped avoid some incongruities, had we chosen that design path. We are aware that there is growing interest in incremental dialogue processing, which reflects more accurately human cognitive processing procedures, in which analysis of the input begins before the utterance is complete. Systems that implement incremental dialogue processing have been successfully developed [33, 34, 35, 36, 37]. Such systems obviate the need for short loops that bypass the semantic processing, thereby reducing but not necessarily eliminating the occurrence of incongruous backchannel from the ECA.

The short loop is implemented as follows. When the IM receives notification of barge-in, it first sends a request to the Multi-Modal Fission Module (MFM) to pause the output of ECA speech (Stage 3 above) and to give the ECA a look of surprise or irritation at being interrupted. The MFM informs the DM of the identity of the sentence in the ECA's utterance during which the barge-in began (Stage 4). When the user finishes pronouncing the interrupting utterance (Stage 5), the IM broadcasts a notification of the barge-in to all modules (Stage 6). This is done for two reasons: (i) to alert modules to the fact that the pronouncing of the previous system turn was not completed (and therefore that certain things the ECA planned to say to the user have not in fact been said, which the DM must now take into account);

(ii) and to make modules aware that the next user utterance they process will be a special interrupting utterance. Once the DM receives the message containing the interrupting utterance, it must decide how to respond to the barge-in.

5 Context-sensitive system recoveries from user barge-in

It is a hard enough problem for the system to detect an interruption, and to react quickly to it by halting speech in a natural way (only one of many additional considerations has been to stop the speech of the ECA in a way that is human-like, by not stopping mid-syllable, for example, and fading out rather than abruptly cutting). It is yet another problem for the system to work out how to continue the conversation from that point. One of the more simple strategies to implement would be for the system to be passive, and to be silent and wait for the user to speak again. Another would be for the ECA to say something that takes no account of the linguistic content of the user's interrupting utterance ('Let's continue', for example). Having designed and implemented a first version of the system as described in the preceding sections, we decided that we wanted a more sophisticated, human-like strategy which would generate different types of appropriate verbal system responses to different types of interruptions. This first of all required us to classify interruptions and responses into different types.

Although we found a large body of work focused on the phenomenon of *interrupting* someone (what interruptions are [11], how people take turns [38, 39], overlapping/simultaneous speech [40, 12], functions of interruptions [41], the psychology of interrupting [42, 43], control of topic in conversation [44], effects of barge-in on ASR [45], prosodic cues of interruptions [46]), work addressing the problem of how a participant in a conversation approaches *responding* to an interruption was not much in evidence.

Rather than making up some rules from intuition alone, we decided to try and inform the system responses to interruptions from data. We have used an informal, publicly available corpus which comprises transcripts of the weekly BBC Radio 4 programme *Any Questions*.[5] The program consists of a panel of four public figures (including usually at least two politicians) who answer questions that are put to them by individuals in the audience. Unscripted discussion of issues between the panel members is

[5] `http://www.bbc.co.uk/radio4/news/anyquestions_archive_dated.shtml`

expected and encouraged. The discussion is overseen by a neutral chairman. Each programme lasts for 50 minutes.

We chose this corpus because it is untypically full of interruptions—the panel includes politicians, who by the nature of the job are eager to air their views and get their arguments across, and who appear very willing to seize the floor without waiting for their turn. The topics of discussion in the corpus are wide-ranging and revolve around items in the nationally broadcast news of that particular week.[6] There are 77 transcripts in the corpus we downloaded, one transcript for each broadcast programme (the collected programmes were broadcast between 12 January, 2007 and 20 March, 2009). The corpus contains approximately 564,500 words. The transcripts contain very sparse annotation. They do not mark exactly which words of the two speakers during an interruption are overlapping words. We also expect that some of the words that were spoken simultaneously are missing entirely from the transcript. However, the BBC transcribers have marked with an ellipsis each point at which an interruption began. Thus far we have fully analysed only three transcripts.[7] The three transcripts together constitute 420 turns. Of these 420 turns, we have identified 44 turns as interruptions, correspondingly 42 turns as recoveries from interruptions, and 20 additional turns as simultaneously an interruption and a recovery.[8]

5.1 Addressing and resuming from an interruption

On examination of the transcripts it was observed that the turn immediately following an interruption was often clearly *doing* two different things, one after the other. In many cases there is a clearly observable distinction between an interruptee's *addressing* of an interruption, and a second immediately following action taken by the interruptee once the interruption has been addressed, which we are calling a *resumption*. In *addressing* an interruption, the interruptee appears to focus on the content of the interruption, and often (but not always) gives the interrupter the response he is looking for. As for the *resumption*, the general direction of this part of the turn is a return to the argument that the interruptee was making immediately before the interruption. It is, however, not usually a strict resumption

[6] The domain of the corpus we used was different from that of the HWYD? system, whose domain was social dialogue focusing on one's day at the office. There was no opportunity at this late stage for us to collect a corpus in that domain, and no corpus available in this domain that was also full of interruptions.

[7] 20070112, 20070119, 20070126.

[8] Owing to the preliminary nature of the work, no inter-annotator agreement is yet in force.

Example 3 'Any Questions' example interruption

T34	If I could go through it—my office received the letter, it's in the criminal record bureau area rather than the areas I'm responsible for, so was passed on to my colleague Joan Ryan...
T35 →	And she read it?
T36	*And she read it and responded and I'll come back to that if I may.* **On the broad point I certainly do think the ministerial team are fit for purpose, I would say that wouldn't I, but I actually mean that.** [more]

of his account from the point at which it was interrupted, because the semantic content and emotional character of the interruption may well affect the interruptee's subsequent contributions, not to mention the break in flow of concentration. Additionally, once the interruptee has addressed the interruption, he often will add some comment on the address he has made. Example 3 is an extract from the corpus showing an interruption (beginning marked by a →), an address (in italics), and a resumption (in bold font).

5.2 Interruption and recovery types

In order to implement principled considered verbal responses to interruptions via the long loop (see section 4.2), we attempted to classify acts [47] of interruption, address, and resumption from observations made from the corpus, and to design some recovery-from-interruption rules based on these classifications (while making observations from the dialogue act classification literature (*e*.g., [48, 49, 50])). We classified the interruptions into types by inferring from the linguistic surface form of the turns the communicative intention that we think the speakers were trying to achieve by performing those turns, by which we mean the communicative action that the speaker wanted the hearer to recognise. During this process, the communicative actions appeared to fall neatly into two types: actions concerning the semantic content of what was said (such as the general argument a person is making, or a detail of his/her argument) versus actions concerning communicative actions (such as actions on the taking of turns or on the expressing of emotion). Appendix A(i) shows a table of recovery act types that we derived from the corpus. The two types 'Actions Concerning Content' and 'Actions concerning Actions' are made explicit.

In order to classify the communicative intention of each recovery, we also first had to classify the communicative intention of the immediately preceding interruption. Appendix A(ii) shows the set of interruption types that we identified. The commonest types of interruptions emerging so far are `elicit-detail` (28.13%), `ign-entire-t-content` (14.06%), `add-detail` (9.38%), `introduce-arg` (9.38%), `elicit-clarif` (7.81%), and `encourage-halt` (7.81%).

5.2.1 System recovery type: CARRY_ON

We identified a significant occurrence (20.97%) of recovery instances of an interruptee ignoring the entire content of the interruptor's turn.[9] Rather than seeing these as missing an address altogether, we choose to recognise them as *acts of ignoring* (with some similarities to Lakoff's "nonresponse" [6, pp.27–28])—we assume that the ignoring of the interruption is a conscious, willful act, and it is not the case that the interruptee is unaware he has been interrupted and has no attitude towards the interruption. We have called this address type `ign-t-content`.

After an `ign-t-content` address, we have observed two types (only) of resumption. Sometimes the interruptee continues speaking from the rough point at which he stopped speaking when he was interrupted, the impression being that he never lost the thread of his thought process that was in place just at the point of interruption. Following an interruption, sometimes the interruptee begins with what are apparently the words that were silenced or spoken over by the interruption. On other occasions he begins by repeating exact phrases that were uttered immediately before the interruption. And sometimes he begins by using different terms with similar meanings to those he was uttering just before the interruption. We have classed these together (for now) as `resume` types of resumption. Turn 20 of Example 4 from the corpus comprises a silent `ign-t-content` address followed by a `resume` resumption (in bold font), which is the kind where the speaker apparently continues exactly (word for word) as he would have if he had not been interrupted, and from the point that the interruption began.

There are also occasions when, after addressing the interruption, the interruptee continues with his account, but does not return to the rough

[9] Ignoring turn content (`ign-t-content`) was the second most frequent way of recovering from an interruption that we observed in the corpus, the most frequent being to supply information that was requested by the interruption (at 30.65%), which we also modelled (see section 5.2.3).

Example 4 'Any Questions' example of a silent `ign-t-content` address followed by a `resume` resumption

T18	I just wanted to pick up the rather weird comment which the Prime Minister then said about Iraq and how everybody would have to approve of what took place...
T19 →	No, no he didn't say that, I was proposing did he mean that, he certainly didn't say that everyone would have to approve.
T20	**It is an odd comment because I mean on the face of it the truth is our failure to carry out adequate peacekeeping in Iraq after the invasion, particularly by the failure of the United States, is one of the primary reasons why it has gone so badly for us and is a disaster, there's no other way of looking at it.** [more]

point at which he stopped speaking, and appears to have lost the thread. We consider these different enough from a `resume` type of resumption to warrant a separate category. We have therefore labelled these as `continue` types of resumption. Thus we devised from our observations two different resumption types which could justifiably follow the address `ign-t-content`.

5.2.2 Implementation of CARRY_ON

In considering the implementation of the recovery-from-interruption schema in the SLD system, we needed to bear in mind that the system would need to be able to distinguish between different types of interruption.

We also needed to identify under what conditions the system should perform an `ign-t-content` type of recovery. We observed that `ign-t-content` is sometimes preceded by cases where the interruptor asks the interruptee for some information which comprises a finer detail of an already-mentioned topic. We classified these as interruptions of type `elicit-detail`. (The interruption⇒address pair `elicit-detail`⇒`ign-t-content` has a frequency of 11.76% among the set of all interruption⇒address pairs observed in the corpus.)

Wishing to keep the NLU understanding of interruptions as easy as possible, we chose to recognise an `elicit-detail` interruption by identifying it as a WH question. We then defined a rule which said, if you spot an `elicit-detail` interruption, then do an `ign-t-content` address followed by a `continue` or by a `resume`. We call this the CARRY_ON recovery strategy, which is one of three main recovery strategies we have now devised from observations of the data and have implemented in the system.

To decide between `continue` and `resume`, we use the *duration* of the interruption in milliseconds (as measured by the IM) on the grounds that if an interruption is short, the interruptee is more likely to remember exactly what she was saying at the point of interruption, and to be able to return to it (`resume`). In a `resume`, the system returns to and repeats the last-spoken utterance of the long narrative (planned by the ASM) and carries on from there. In a `continue` the system returns to the narrative planned by the ASM, but does not repeat the last-spoken utterance, but instead jumps to the next planned utterance of the long narrative. These different strategies are facilitated by the fact that in Stage 4 of an interruption (see section 4) the MFM informs the DM of the point in a long narrative system turn (planned by the ASM) at which the user interrupted, also indicating how much of the long system turn the ECA managed to say before stopping in response to the user interruption. This is marked in the following extract from the system log of a session with a user. In the right-hand column opposite the arrow (the interrupting utterance) we see the utterance number (4) of the long turn (Turn (d)) during which the ECA stopped speaking. In Example 4 we also see a short loop back-channel (Turn (b)). And in Turn (f) we see the system repeat the last part of its narrative, making this a `resume` (but note that the surface form is slightly different, because the NLG does not give exactly the same output every time for the same input from the ASM).

5.2.3 System recovery type: REPLAN

The most common address+resumption pair that we have identified in the data is a `supply-info` address followed by an `elaborate-response` resumption (31.25%) of all observed address+resumption pairs (including those where the resumption is 'nil'). In a `supply-info` address, the interruptee typically supplies information that the interruptor has requested. In the following `elaborate-response` resumption, the interruptee elaborates on his own address. Bearing in mind the design of the system, and the planning capabilities of the ASM, we decided it was justifiable and sensible to frame this `supply-info+elaborate-response` recovery as a case where the system has a rethink, and generates an alternative plan which uses some but not all of the same variables as its most recent pre-interruption plan. On the basis of

Example 4 A CARRY_ON system recovery

U(a)	And it was a good journey there was no traffic	
S(b)	Uhuh	(short loop)
U(c)	So I got a good start working on my presentation next week	
S(d)	Try not to get too excited.	(d1)
	Yeah looking at the traffic from a different perspective might help.	(d2)
	In my ↓ *[opinion traffic is a bad thing!*	(d3)
	It seems clear that traffic although unavoidable is likely to be bad.]	(d4)
U(e)→	What you mean Matilda that I didn't say anything about the traffic being terrible	`elicit-detail`
S(f)	So as I was just saying, it's apparent that the consequences of traffic is likely to be bad.	`ign-t-content` then `resume`

Example 5 A REPLAN system recovery

S(a)	I have experienced a similar situation.	(a1)
	I understand it's unlikely that you could keep motivated due to the redundancies.	(a2)
	The dispute could influence them which isn't exactly helpful.	(a3)
	I'm not sure you have the abilities necessary to handle the outcomes ↓ *[arising from the redundancies.*	(a4)
	It's apparent that redundancies although unavoidable are likely to be bad.]	(a5)
U(b)→	And what do you mean I don't have the ability is necessary	`add-detail`
S(c)	Alright I'm nervous that the dispute may have an affect on you.	`supply-info`
	It's obvious that doing your best can help you face up to your day.	`elaborate-response`

the frequency of the interruption type `add-detail` (9.38%), and the design and function of the DM, we also decided that an `elaborate-response` resumption should be made in response to user utterances in which the NLU identifies new objects or events that have not been mentioned before in the current conversation (are not yet part of the common ground ('conversational record' [51], 'scoreboard' [52], 'shared presuppositions' [53], *etc.*). We call this the REPLAN recovery strategy. Example 5 is an extract from a user–system session log which illustrates a REPLAN. Turn (b) is classified by the system as an `add-detail` type of interruption.[10] Turn (c) shows the system using the new information about the user's ability in its recovery from the interruption.

Whenever the ASM receives a REPLAN signal from the DM it uses the updated information from the DM to re-generate the plan for the current system turn. This does not consist of any changes to the previous plan but rather the previous plan is abandoned and a new plan is generated

[10]The system did not recognise Turn (b) as a WH question because it did not *begin* with a WH word. This is one of many undesirable shortcomings that are currently being addressed.

including a full reappraisal of the information from the DM to ensure no false assumptions are retained from the previous plan. However, since part of the previous plan has already been relayed to the user (thus provoking the interruption), it may not be appropriate to relay a complete plan in its entirety as this would risk repetition. Instead the ASM uses the timing of the user's interruption (with respect to how much of the previous plan has been relayed) to determine the length of the new plan and thus what items should be discussed. As such, an interruption at the start of the previous plan will result in another full plan being generated while an interruption at the end of the previous plan will limit the newly generated version to a much smaller, more focussed plan without repetition of extraneous details. This is illustrated in Example 5 which shows that the new plan consists only of the salient point of (a3) and a concluding summary because the interruption came at the end of the previous plan,.

5.2.4 System recovery type: ABORT

We have devised a third recovery strategy—ABORT—the idea being that the system is aborting its own immediate pre-interruption communicative intentions. In an ABORT the system first performs an `apology` (constituting the address). After the apology, the system has two choices of resumption, depending on the nature of the user interruption. For cases in which the interruption is found to be a `rejection` of what the system has said, the system tries a different subject or a different approach to see if that pleases the user better. We are calling this a `re-focus` resumption. If the interruption is not found to be a `rejection`, and no WH question is discerned (*i.e.*, not a CARRY_ON), and no new objects are retrieved (*i.e.*, not a REPLAN), the system asks the user to repeat what he just said. Apologies, rejections and repetitions are all in evidence in the data, and so we felt that this was a justifiable approach to further broadening the ways in which the system can respond to user interruptions.[11] Example 6 is an extract from a user/system session log which illustrates an ABORT (preceded by a REPLAN). From Turn (d) we can see that the interruption is a `rejection`, and from Turn (e) we can see the system's apology and a strong focus on the content of the user's interruption rather than the system returning to

[11] Additionally, we take note that many interruptions are considered by much of the interruptions literature to be hostile, in that the interruptor snatches the conversational floor before it is his turn. By implication it is reasonable to posit that if a user interrupts the system, the system may have said something that has not been well received by the user, and this adds weight to the appropriateness of an apology in an interruption recovery.

Example 6 An ABORT system recovery (after a REPLAN)

S(a)	It's apparent that sometimes ↓ *[the worst doesn't happen]*.	
U(b)→	But still missing the deadline will be very bad and I'm even afraid I may get fired	
S(c)	It's obvious that being prepared can help you ↓ *[face up to your day]*.	
U(d)→	ahhh not that prepared because my computer crashed	`rejection`
S(e)	I'm sorry.	`apology`
	Who was responsible for this?	`re-focus`

its pre-interruption intentions.

6 Conclusion

This paper has described the mechanism that has been implemented in the COMPANIONS HWYD? prototype for responding to user barge-in interruptions during conversations with an ECA. Rather than being task-based, the domain of the ECA is social conversation with the over-arching goal of developing and maintaining a supportive relationship with the user by means of natural conversation. This aim precluded a distinctive design feature of the system, which is that it is robust enough to handle and produce long, multi-utterance turns. The long system turns create good opportunities for user barge-in interruptions, and so invited special attention to how the system might respond to these. The system now incorporates a novel mechanism for recognising and responding to interruptions in a variety of ways, which relies on a categorisation of interruptions and *recoveries* from interruptions, and on rules based on those categorisations, whose design has been informed by manual analysis of real data. Recoveries are two-part structures that begin with an *addressing* of the interruption, and follow with a *resumption* of the interruptee's account from immediately before the interruption occurred. The HWYD? system requires improvement and development in a number of directions. The corpus of interruptions requires full investigation to better inform the theory of how to recover from an interruption.

7 Acknowledgments

This work was partially funded by the COMPANIONS project (http://www.companions-project.org) sponsored by the European Commission (EC) as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434. We thank the University of Augsburg (Prof. Elisabeth André) for supplying a version of the EmoVoice [25] system. Other contributors to the prototype described in this paper are [list of names deleted for blinded version].

References

[1] M. Cavazza, R. Santos de la Cámara, M. Turunen, and the COMPANIONS consortium. How was your day ? a companion ECA. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, accepted for publication (demonstration paper)*, Toronto, 2010.

[2] S. Young. Still talking to machines (cognitively speaking), 2010. In *Proc. Interspeech* 2010, Chiba, Japan, 26–30 September, 2010.

[3] O. Lemon, K. Georgila, J. Henderson, and M. Stuttle. An ISU dialogue system exhibiting reinforcement learning of dialogue policies: generic slot-filling in the TALK in-car system. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, EACL '06, pages 119–122, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[4] J. Allen, N. Chambers, G. Ferguson, L. Galescu, H. Jung, M. Swift, and W. Taysom. Plow: a collaborative task learning agent. In *Proceedings of the 22nd national conference on Artificial intelligence - Volume 2*, pages 1514–1519. AAAI Press, 2007.

[5] C. West and D. Zimmerman. Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. In B. Thorne, C. Kramarae, and N. Henley, editors, *Language, gender and society*, pages 102–117. Newbury House, Massachusetts, 1983.

[6] R. T. Lakoff. Cries and whispers: the shattering of the silence. In K. Hall and M. Bucholtz, editors, *Gender Articulated: Language and the Socially Constructed Self*, pages 25–50. Routledge, New York, 1995.

[7] H. Sacks, E. A. Schegloff, and G. Jefferson. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735, 1974.

[8] J. Coates. *Women, men, and language: A sociolinguistic account of gender differences in language*. London and New York: Longman., 2nd ed. edition, 1993.

[9] E. Bevacqua, S. Pammi, S. J. Hyniewska, M. Schröder, and C. Pelachaud. Multimodal backchannels for embodied conversational agents. In J. M. Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, and A. Safonova, editors, *IVA*, volume 6356 of *Lecture Notes in Computer Science*, pages 194–200. Springer, 2010.

[10] L.-P. Morency, I. de Kok, and J. Gratch. Predicting listener backchannels: A probabilistic multimodal approach. In H. Prendinger, J. C. Lester, and M. Ishizuka, editors, *IVA*, volume 5208 of *Lecture Notes in Computer Science*, pages 176–190. Springer, 2008.

[11] D. Zimmerman and C. West. Sex roles, interruptions and silences in conversation. In B. Thorne and N. Henly, editors, *Language and sex: Difference and dominance*, pages 10–129. Newbury House, Massachusetts, 1975.

[12] S. O. Murray. Toward a model of members' methods for recognizing interruptions. *Language in Society*, 14(1):31–40, 1985.

[13] A. Raux and M. Eskenazi. A multi-layer architecture for semi-synchronous event-driven dialogue management. In *ASRU*, pages 514–519, Kyoto, Japan, 2007.

[14] J. Barnett and M. Singh. Designing a portable spoken dialogue system. In E. Maier, M. Mast, and S. LuperFoy, editors, *ECAI Workshop on Dialogue Processing in Spoken Language Systems*, volume 1236 of *Lecture Notes in Computer Science*, pages 156–170. Springer, 1996.

[15] R. C. Rose and H. K. Kim. A hybrid barge-in procedure for more reliable turn-taking in human-machine dialog systems. In *Proceedings of the Automatic Speech Recognition and Understanding Workshop*, 2003.

[16] B. Balentine and D. P. Morgan. *How to build speech recogniton applications—A style guide for telephony dialogs*. Enterprise Integration Group, San Ramon, CA, 1999.

[17] A. R. Setlur and R. A. Sukkar. Recognition-based word counting for reliable barge-in and early endpoint detection in continuous speech recognition. In *Proceeding of the International Conference on Spoken Language Processing*, pages 2135–2138, 1998.

[18] K. Matsuyama, K. Komatani, T. Ogata, and H. G. Okuno. Enabling a user to specify an item at any time during system enumeration—Item identification for barge-in-able conversational dialogue systems. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009)*, pages 252–255, Brighton UK, 6-10 September 2009.

[19] K. Komatani and A. I. Rudnicky. Predicting barge-in utterance errors by using implicitly-supervised ASR accuracy and barge-in rate per user. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 89–92, Suntec, Singapore, August 2009. Association for Computational Linguistics.

[20] R. A. Brooks. A robust layered control system for a mobile robot. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA, 1985.

[21] R. A. Brooks. Intelligence without representation. In *Computation & intelligence: collected readings*, pages 343–362. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1995.

[22] R. K. Moore. Presence: A human-inspired architecture for speech-based human-machine interaction. *IEEE Trans. Computers*, 56(9):1176–1188, 2007.

[23] D. Reidsma, T. de Kok, D. Neiberg, S. Pammi, B. van Straalen, K. Truong and H. van Welbergen. Continuous Interaction with a Virtual Human. *Journal on Multimodal User Interfaces*, 4:97–118.

[24] R. Santos de la Cámara, M. Turunen, J. Hakulinen, and D. Field. How was your day? an architecture for multimodal ECA systems, 2010. *Proc. 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 24-?25 September, 2010, University of Tokyo, Tokyo, Japan, pp. 47-?50.

[25] T. Vogt, E. André, and N. Bee. Emovoice – a framework for online recognition of emotions from voice. In *Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies*

for Speech-Based Systems: Perception in Multimodal Dialogue Systems, PIT '08, pages 188–199, Berlin, Heidelberg, 2008. Springer-Verlag.

[26] K. Moilanen and S. Pulman. Sentiment composition. In *Proceedings of the Recent Advances in Natural Language Processing International Conference (RANLP-2007)*, pages 378–382, Borovets, Bulgaria, September 27-29 2007.

[27] C. Bremond. *Logique du Récit.* Editions du Seuil, Paris, 1973.

[28] M. Cavazza, C. Smith, D. Charlton, N. Crook, J. Boye, S. Pulman, K. Moilanen, D. Pizzi, R. Santos de la Cámara, and M. Turunen. Persuasive dialogue based on a narrative theory: an ECA implementation. In *Proceedings of the fifth International Conference on Persuasive Technology (Persuasive 2010)*, Copenhagen, Denmark, 7–10 June 2010.

[29] C. Smith, N. Crook, J. Boye, D. Charlton, S. Dobnik, D. Pizzi, M. Cavazza, S. Pulman, R. Santos de la Cámara, and M. Turunen. Interaction strategies for an affective conversational agent. In *Proc. of the 10th Int. Conf. on Intelligent Virtual Agents (IVA 2010)*, Philadelphia, PA, September 2010.

[30] A. Hernández, B. López, D. Pardo, R. Santos, L. Hernández, J. Relaño Gil, and M. Rodríguez. Modular definition of multimodal ECA communication acts to improve dialogue robustness and depth of intention, 2008. In *Proc. 1st Functional Markup Language Workshop, 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Estoril, Portugal, 12–16 May 2008.

[31] B. López, A. Hernández, D. Pardo, R. Santos, and M. Rodríguez. ECA gesture strategies for robust SLDS, 2008. In *Proc. Artificial Intelligence and Simulation Behaviour Convention (AISB 2008) Symposium on Multimodal Output Generation*, Aberdeen, UK, 1–4 April, 2008.

[32] M. Danieli and E. Zovato. The affective dimension of speech acts and voice expressiveness, 2010. In M. Pettorino, A. Giannini, I. Chiari and Francesca Dovetto (eds.) *Spoken Communication*, Cambridge Scholars Publishing, Newcastle upon Tyne, UK, pp. 191–204.

[33] S. Stoness, J. Tetreault and J. Allen. Incremental parsing with reference interaction. In *ACL Workshop on Incremental Parsing*, 2004, 18–25.

[34] G. Aist, J. Allen, E. Campana, C. Gallo, S. Stoness, M. Swift and M. Tanenhaus. Incremental understanding in human-computer dialogue and experimental evidence for advantages over nonincremental methods, In *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue*, Trento, Italy, 30 May - 1 June 2007, 149–154.

[35] T. Brick and M. Scheutz. Incremental natural language processing for HRI. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, Arlington, Virginia, USA, 2007, 263–270.

[36] G. Skantze and D. Schlangen. Incremental dialogue processing in a micro-domain. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 745–753, Athens, Greece, April 2009.

[37] D. Schlangen. and G. Skantze. A general, abstract model of incremental dialogue processing. In *Proc. of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 710–718, Athens, Greece, April 2009.

[38] D. Starkey. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23:283–292, 1972.

[39] J. M. Wiemann and M. L. Knapp. Turn-taking in conversations. *Journal of Communication*, 25:75–92, 1975.

[40] E. A. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63, 2000.

[41] C. W. Kennedy and C. T. Camden. A new look at interruptions. *Western Journal of Communication*, 47:45–58, 1983.

[42] D. Roger. 4: Experimental studies of dyadic turn-taking behaviour. In D. Roger and P. Bull, editors, *Conversation: An interdiciplinary perspective*. Multilingual Matters, Clevedon, England, 1989.

[43] I. Hutchby. Confrontation talk: Aspects of interruption in argument sequences on talk radio. *Interdisciplinary Journal for the Study of Discourse*, 12:343–372, 1992.

[44] M. Walker and S. Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proc. 28th Annual Meeting of the ACL*, pages 70–79, 1990.

[45] R. Heins, M. Franzke, M. Durian, and A. Bayya. Turn-taking as a design principle for barge-in in spoken language systems. *International Journal of Speech Technology*, 2:155–164, 1997. 10.1007/BF02208827.

[46] R. K. E. N. Oth, A. Kieling, T. Kuhn, M. Mast, H. Niemann, K. Ott, and A. Batliner. Prosody takes over: Towards a prosodically guided dialog system. *Speech Commun*, 15(15):155–167, 1994.

[47] J. L. Austin. *How to do Things with Words.* Oxford University Press, New York, 2 edition, 1962.

[48] H. C. Bunt. Dynamic interpretation and dialogue theory. In M. M. Taylor, F. Neel, and D. G. Bouwhuis, editors, *The Structure of Multimodal Dialogue*, volume 2, pages 139–166. North-Holland, Amsterdam, 2000.

[49] D. R. Traum. 20 questions on dialogue act taxonomies. *Journal OF Semantics*, 17:7–30, 2000.

[50] D. Jurafsky, E. Shriberg, and D. Biasca. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13, university of colorado, boulder, 1997. *Institute of Cognitive Science Technical Report 97-02.*

[51] R. H. Thomason. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics, 1990. In [?], pp. 325–63.

[52] D. Lewis. Scorekeeping in a language game, 1979. *Journal of Philosophical Logic* 8: 339–59. Reprinted in D. Lewis *Philosophical papers Volume I*, 1983, New York and Oxford: Oxford University Press, pp. 233–249.

[53] R. Stalnaker. Pragmatics, 1972. In D. Davidson and G. Harman. Editors. *Semantics of natural language (Synthese Library, Vol. 40)*, pp. 380–97. Dordrecht, Holland: D. Reidel.

Appendix A: Interruption and Recovery Types

(i) Types of Recovery from an Interruption and their Relative Frequencies

Actions concerning Content		
CONCERNING ARGUMENTS (POSED BY PARTICIPANTS)		
accept-arg	I accept that A is your argument (but I don't necessarily agree with A)	1.61%
concede-arg	I accept that A is your argument and I agree with A	1.61%
introduce-arg	I introduce a new argument (that is related to the current topic)	1.61%
**reject-arg*	I reject your argument[12]	1.61%
**retry-arg*	I retry my argument by trying a different surface form	1.61%
**ign-entire-t-content*	I ignore the arguments of your turn (but I do not speak over you)	20.97%
CONCERNING DETAILS OF ARGUMENTS		
**confirm-detail*	I confirm the detail that you have introduced	6.45%
**elicit-detail*	I ask you for some detail (that is not a clarification)	8.06%
**elicit-clarif*	I ask you to clarify an already introduced detail	1.61%
reject-detail	I reject the detail that you have introduced	6.45%
CONCERNING INFORMATION		
**supply-info*	I supply the information that you have requested	30.65%
withold-info	I decline to supply the information that you have requested	1.61%
Actions concerning Actions		
CONCERNING THE TAKING OF TURNS		
accept-t-completion	I accept your collaborative completion of my turn	1.61%
commit-hurry	I commit myself to hurrying my turn	1.61%
**encourage-halt*	I encourage you to quit your turn	1.61%
invite-turn	I invite a third party to take a turn	3.23%
CONCERNING THE EXPRESSING OF EMOTION		
ack-emo	I acknowledge your expression of emotion	1.61%
elicit-emo-happy	I inject humour to elicit happy emotion	1.61%
CONCERNING CHOOSING OF TOPIC		
permit-topicchoice	I permit you to choose the next topic	3.23%
**request-topicchoice*	I ask permission to choose the next topic	1.61%

[12]The examined transcripts contain relatively many turns that are simultaneously an interruption (I) and a recovery (R). A turn is both an interruption and a recovery when a speaker interrupts an interruption in order to make the recovery. Strictly speaking, then, three types of turn were being analysed in this work: I's, R's, and IR's. Four of the actions identified apply *only* to IR's, and not to any turn that is only an I or only an R. Those actions are marked by an asterisk.

(ii) Types of Interruption and their Relative Frequencies

Actions concerning Content		
accept-arg	I accept that A is your argument (but I don't necessarily agree with A)	1.56%
elaborate-arg	I elaborate on my earlier argument	1.56%
introduce-arg	I introduce a new argument (that is related to the current topic)	9.38%
**reject-arg*	I reject your argument	3.13%
**retry-arg*	I retry my argument by trying a different surface form	1.56%
reject-interp	I reject your interpretation of my argument	1.56%
**supply-info*	I supply the information that you have requested	3.13%
add-clarif	I add a clarifying detail to the argument you are presenting	1.56%
**elicit-clarif*	I ask you to clarify an already introduced detail	7.81%
**elicit-detail*	I ask you for some detail (that is not a clarification)	28.13%
add-detail	I add a detail (that is not a clarification) to the argument you are presenting	9.38%
**confirm-detail*	I confirm the detail that you have introduced	1.56%
correct-detail	I correct the detail that you have introduced	1.56%
**ign-entire-t-content*	I ignore the arguments of your turn (and I speak over you)	14.06%
Actions concerning Actions		
**encourage-halt*	I encourage you to quit your turn	7.81%
**request-perm*	I request permission from you to do something	1.56%
express-emo	I express emotion	1.56%
accept-apol	I accept your apology	1.56%
rush	I anticipate the words you are going to say and I say them for you	1.56%