Interaction Strategies for an Affective Conversational Agent

Cameron Smith[1], Nigel Crook[2], Johan Boye[2], Daniel Charlton[1], Simon Dobnik[2], David Pizzi[1], Marc Cavazza[1], Stephen Pulman[2], Raul Santos de la Camara[3] and Markku Turunen[4]

[1] School of Computing, Teesside University, Middlesbrough, United Kingdom
[2] Oxford University Computing Laboratory, Wolfson Building, Oxford, United Kingdom
[3] Telefonica I+D, C/ Emilio Vargas 6, 28043 Madrid, Spain
[4] Department of Computer Sciences, 33014 University of Tampere, Finland
{c.g.smith, d.charlton, d.pizzi, m.o.cavazza}@tees.ac.uk
{nigel.crook, simon.dobnik, stephen.pulman}@comlab.ox.ac.uk,
johan.boye@speechact.se, e.rsai@tid.es, mturunen@cs.uta.fi

Abstract. The development of Embodied Conversational Agents (ECA) as Companions brings several challenges for both affective and conversational dialogue. These include challenges in generating appropriate affective responses, selecting the overall shape of the dialogue, providing prompt system response times and handling interruptions. We present an implementation of such a Companion showing the development of individual modules that attempt to address these challenges. Further, to resolve resulting conflicts, we present encompassing interaction strategies that attempt to balance the competing requirements. Finally, we present dialogues from our working prototype to illustrate these interaction strategies in operation.

Keywords: Embodied Conversational Agents, Companion, Affective Dialogue, Conversational Dialogue, Interruptions, Interaction Strategies.

1 Introduction

An emerging concept in recent years has been that of a social agent which focuses more on the relationship it can establish with a human user than on the assistance or information it can provide for a practical task. This concept of a "Companion" is particularly significant for Embodied Conversational Agent (ECA) research where the notion of companionship emerges from the overall communicative abilities of the ECA (that is, embodied and conversational aspects feeding into affective dialogue). Yet, there are also significant technical challenges encountered here in the integration of linguistic communication and non-verbal behaviour for affective dialogue [1].

In this paper, we present the implementation of a companion ECA integrating all the above aspects into a single prototype, in a way which supports conversational phenomena one would expect from affective dialogue, namely lengthy utterances on both sides and interruptions. This presentation mainly focuses on the interaction

strategies supported by the agent, which support the principled integration of the large number of software components required to analyse user input, reason upon the situation, control the flow of dialogue and generate appropriate ECA responses and multimodal behaviours. Our main objective is to give an insight into these interaction strategies and to illustrate the Companion's performance with detailed examples from a fully-implemented prototype.

Fig. 1. The Companion during a typical dialogue.

2 System Overview and Application

The Companion (as shown in Figure 1) presents itself as an ECA with which the user can engage in a free conversation, albeit on a select set of topics. As an application scenario, we wanted an everyday life domain that would support conversation with some affective content. We opted for a scenario in which the user, a typical office worker, returns home and talks about the day's events. We refer to this as the "How was your day?" (HWYD) scenario. The system currently supports over 40 work-based conversational topics, with further discussion of a range of influencing factors and event outcomes, across a range of emotional situations. By definition, the conversation is not task-oriented (unless one considers a very high level task of supporting the user through positively influencing their attitudes) and follows a mixed-initiative paradigm. User initiative, as expected, takes a central role, but without reducing the Companion to a passive, although sympathetic, listener. As evidenced by the example dialogues of Figures 5, 6 and 7, the Companion will attempt to offer appropriate advice as soon as it has assessed the user situation and considers such advice as appropriate.

Our system integrates no less than 15 different software components covering aspects of multimodal affective input, affective dialogue processing, interruption management and multimodal affective output. The software architecture integrating these components follows a blackboard philosophy [2], which provides the control flexibility required to implement various interaction strategies (see below). The system (Figure 2) comprises speech, language, reasoning and animation modules. Automatic Speech Recognition (ASR) is provided by Nuance's Dragon NaturallySpeaking, whilst Text-To-Speech (TTS) is an extension of Loquendo's

commercial system developed as part of this project. The ECA appearance and animation are based on the Haptek™ toolkit. As expected, all dialogue and Natural Language Understanding (NLU) modules are proprietary. Emotional aspects are pervasive in these modules but their inclusion depends on the module itself: the animation module for the ECA naturally supports non-verbal behaviour and the expression of emotions, whilst our Text-To-Speech system has been specifically extended to support emotional markers. Finally, some modules are entirely dedicated to affective processing: the recognition of emotional categories from speech is based on the EmoVoice [3] system, the affective content of utterances' transcripts is uncovered using a Sentiment Analysis module [4]. Depending on the interaction strategy considered, these modules will be used separately or their output will be merged using an Emotional Model performing multimodal fusion of affective categories. In this system, multimodality is primarily dedicated to affective aspects, both in terms of input (emotional contents of speech/voice and transcribed utterances) and output (ECA speech, facial expressions and gestures).

Affective dialogue processing is lead by the Dialogue Manager (DM), which supports traditional functions such as managing clarification dialogue and repair. It further makes use of the more specific Affective Strategy Module (ASM) for generating complex affective utterances and a Natural Language Generation (NLG) module for realising replies into utterances for the multimodal affective output stage. The multimodal affective output is coordinated by the Multimodal Fission Manager (MFM) which controls both the ECA and Text-To-Speech modules. This is all overseen by an interruption management layer coordinated by the Interruption Manager (IM). The necessity to control turn-taking and interruptions has led to the incorporation of specific speech modules: the Acoustic Analysis (AA) and Acoustic Turn Taking (ATT) modules, which input into a Dialogue Act Tagger (DAT).

Natural language processing was also adapted to the objectives of affective dialogue and free conversation. The techniques used, including tagging, shallow parsing, named entity identification and contextual reference resolution, resemble Information Extraction and provide a robust coverage of the longer utterances, compared to previous dialogue systems, found in non-task orientated conversations.

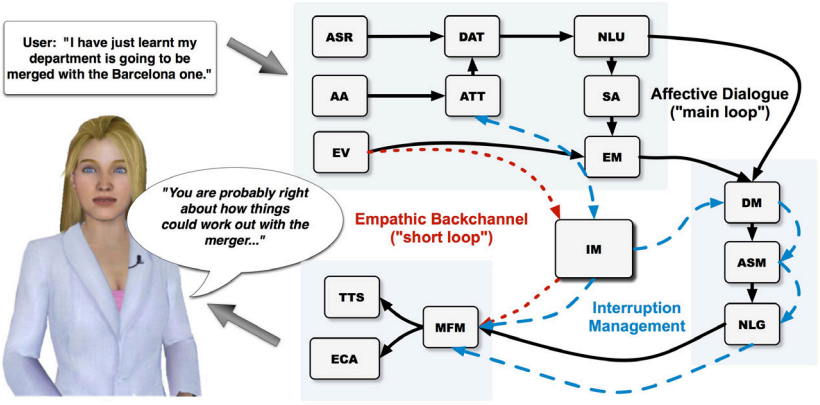

Fig. 2. System components with principal interaction loops (see text for details).

3 Interaction Strategies

The majority of language-enabled ECA have been developed in the context of task-based dialogue; this was dictated by both application constraints and linguistic coverage. However, the very idea of a companion agent assumes a level of conversation which is disconnected from any immediate task, and in particular is freed from strict constraints on the nature of dialogue.

Therefore several traditional assumptions which have presided over the formalisation of human-computer dialogue may need to be relaxed when exploring affective conversation. In everyday life, many inter-human conversations see one of the participants relating events through lengthy descriptions, without this corresponding to any specific request or encompassing speech act. Our objective was to support such free conversation, whilst still obtaining meaningful answers from the Companion in the form of advice appropriate both to the affective and informational content of the conversation.

In order to balance the constraints of free conversation with those of tractability, we have deliberately opted for a single-topic conversation, in contrast both to small talk [5] and 'chatterbot' approaches. It should be noted that even 'chatterbots' fail to depart from the conventions of human-computer dialogue, and most often feature dialogues in which user and agent utterances alternate rather strictly [6].

Our individual components seek to address some of the challenges of conversational dialogue: affective input, longer utterances, balancing clarification dialogue with long-form responses and the generation of these long-form responses. Yet individual optimisations only tackle part of the problem and can often introduce further problems of their own. As such, we additionally sought a more holistic approach; several interaction strategies allowing the different components to work together effectively, each strategy catering to different requirements of a Companion.

In the following sections we look in detail at the interaction strategies available before going on to provide examples from our implemented system showing the various interaction strategies in operation.

3.1 "Short Loop" Interaction: an Empathic Backchannel

Previous work has amply demonstrated the importance of backchannels in human-agent conversation [7] [8] [9]. In addition, the processing time required by the complete affective dialogue system, which includes reasoning upon the user's situation and the appropriateness of her emotional reaction, still exceeds recommended response time for dialogue systems, being on average over 3 seconds. This makes it essential to provide a real-time (< 700ms) yet relevant backchannel to the user, which is able to acknowledge user interaction and provide an initial response appropriate to the affective context even without a full analysis of the utterance.

The "short loop" implements a fast alignment between the perceived emotional state of the user and the ECA's expression, as well as acknowledging user utterances (see Figure 2). This is achieved by matching the ECA's non-verbal response to the emotional speech parameters detected by the emotional speech recogniser EmoVoice and including an appropriate verbal acknowledgement (on a random basis to avoid

acknowledging all user utterances). The short loop thus essentially aligns the ECA response on the user's attitude.

3.2 "Main Loop" Interaction: Affective Dialogue and Reasoning

The main interaction strategy consists in a complete end-to-end implementation of affective conversation (with a response time of under 3000 ms). It enacts the overall behaviour of the Companion as an affective dialogue system and involves its full response to the user utterance in terms of both verbal and non-verbal behaviour (both gestures and facial expressions).

The "main loop" (see Figure 2) thus corresponds to an end-to-end implementation of affective conversation between the user and the agent. It is based on the identification of office life events, together with the affective context in which they are introduced. Following an appraisal step that determines the adequacy of the user's response to the situation she is facing (e.g. difficulties with colleagues, restructuring, redundancies), the Companion will provide an affective response in the form of reassurance, advice, comfort (or, in some cases, warning) to positively affect the user's attitude. The content is however specific to the details of the situation reported and makes reference to the different causes and consequences of the reported events. Conversational dialogue further requires a degree of flexibility in juggling user utterances of varying lengths with shifting topics while accounting for affective aspects. The expectation is that the Companion will be able to provide a response of appropriate length and tone in reply to the topic provided by the user. However, in order to do this effectively the Companion may be required to clarify information and elicit further information to support a meaningful response. The dialogue management thus needs to find a balance between employing clarification dialogue and generating appropriate responses to the information provided by the user.

The overall conversational loop is under the supervision of a Dialogue Manager which controls the various phases of dialogue and their timing, as well as the level of system initiative, in an integrated fashion. One of the main decisions it has to make is when to trigger lengthier utterances (which we have termed 'tirades' – see e.g. Figures 5, 6 and 7), which correspond to an affective dialogue strategy aiming at influencing the user's attitude by means of a short narrative. The challenge for the DM is to shift between the various aspects of conversation: allowing long rants from the user, providing sympathetic feedback without shifting dialogue initiative towards itself, triggering clarification sub-dialogues, or regaining initiative through long utterances that provide advice and support in a more structured fashion. Some of these aspects may be covered by the identification of Dialogue Acts, but Dialogue Acts alone may not be able to deal with the contents of longer user utterances (> 30 words). This is why one of the integrating principles adopted in our system is to also base dialogue control on event instantiation, thus relating it to Information Extraction.

3.3 Information Extraction

Conversations may involve utterances of various lengths including utterances much longer (> 50 words) than those typically found in task-oriented dialogues. Sentences may be ill-formed or highly elliptical. Furthermore, speech recognition under realistic conditions frequently results in a high word error rate making the task of syntactic analysis even harder. The task of the Natural Language Understanding module is to recognise a specific set of events reported by the user. These events are formalised as objects consisting of feature-value pairs. The NLU (in collaboration with the DM) employs shallow processing methods that instantiate event templates. These methods resemble Information Extraction (IE) techniques [10] [11].

The NLU takes the 1-best output from the speech recogniser, which has already been segmented into dialogue-act sized utterances. The utterances are then part-of-speech tagged and separated into Noun Phrase (NP) and Verb Group (VG) chunks which denote concepts in our domain. VGs consist of a main verb and any auxiliary verbs or semantically important adverbs. Both of these stages are carried out by a Hidden Markov Model trained on the Penn Treebank, although some customisation has been carried out for this application (relevant vocabulary added and some probabilities re-estimated to reflect properties of the application). NP and VG chunks are then classified into Named Entity (NE) classes, some of which are the usual 'person', 'organisation', 'time' etc. but others of which are specific to the scenario, as is traditional in IE: e.g. salient events, expressions of emotion, organisational structures etc. NE classification, in the absence of domain specific training data, is carried out via hand-written pattern matching rules and gazetteers. The NPs and VGs are represented as unification grammar categories containing information about the internal structure of the constituents: for example, an utterance like "John will move to the Madrid office next month" would yield results like that on the left of Figure 3.

```
<sentence>
<np head="John" stem="John" netype="per" gndr="male">
   <w>John/NNP</w>
</np>
<vg head="move" stem="move" polarity="pos"
netype="event_nature" tense="future">
   <w>will/MD</w>
   <w>move/VB</w>
</vg>
<w>to/TO</w>
<np head="office" stem="office" netype="org_generic">
   <w>the/DT</w>
   <w>Madrid/NNP</w>
   <w>office/NN</w>
</np>
<np head="month" stem="month" netype="time_period">
   <w>next/JJ</w>
   <w>month/NN</w>
</np>
</sentence>
```

```
[ A: event
    participants = [B],
    nature = move_office,
    agent = C,
    temporal_reference = future,
    modal = will,
    adv = D]

[ B: office
    full_name = [Madrid,office],
    name = office,
    number = sing]

[ D: time_period
    time = month,
    number = sing]

[ C: person
    gender = male,
    name = John,
    number = sing]
```

Fig. 3. NP and VG representation (*left*) and final semantic representation (*right*) used by the NLU

In the next stage of NLU processing, domain specific IE patterns are applied on NP and VG chunks which rely on their syntactic and semantic information to form constituents called objects. For example, "meeting with X about Y" where NE type of X is person, or "move to X" where NE type of X is org_generic. In the final stage reference resolution for pronouns and definite NPs is performed. This module is based partly on the system described by Kennedy and Boguraev [12], with the various weighting factors based on theirs. Each referring NP gives rise to a discourse referent, and these are grouped into coreference classes based on grammatical, semantic, and salience properties.

On its own the NLU module is a large-coverage system which can tag, shallow parse and resolve pronoun reference of any English sentence. Its coverage is most restricted by domain specific NE classes and IE patterns which must be introduced manually. The system covers more than 40 work-based topics of conversation, for example discussions of meetings, problems with office equipment, relationships with colleagues and even the weather. These are mostly represented as event objects. Complex objects such as these are created by a set of IE rules which attempt to cover a range of syntactic and semantic structures which denote identical content. In addition to event objects, the system covers objects of various NE types that relate to the events. For example, to refer to persons, the system may have to collect their names, gender and profession, organisation they work for, their colleagues and the location where they live. In contrast to events, these objects mostly rely on recognition of NE classes.

The final output from the NLU in the format expected by the DM for the utterance "John will move to the Madrid office next month" is shown on the right of Figure 3.

3.4 Dialogue Management

The DM is based on work described in Boye and Gustafson [13], Boye et al [14] and Boye [15] but has been substantially modified for the challenges of conversational dialogue. It receives user utterances from the NLU as semantic representations (right of Figure 3). The DM first checks which information addresses the previous question or comment posed by the system in the dialogue and which information opens up new topics. The information constituting answers to system questions is integrated into the information state of the DM (called the Object Store), while new topics give rise to new conversational goals.

The DM keeps track of all the topics under discussion by maintaining a set of conversational goals, e.g. (1) "Find out more about the possible office relocation to Madrid", or (2) "Make a comment about today's meeting". A number of goal-satisfaction rules (similar to the one on the left of Figure 4) specify how goals are broken down into sequences of sub-goals and system utterances. For instance, finding out more about the office relocation (1) might amount to asking specific questions about whether the relocation will indeed take place, what the consequences would be for the user, etc. The goal is considered satisfied when further information about the relocation has been collected.

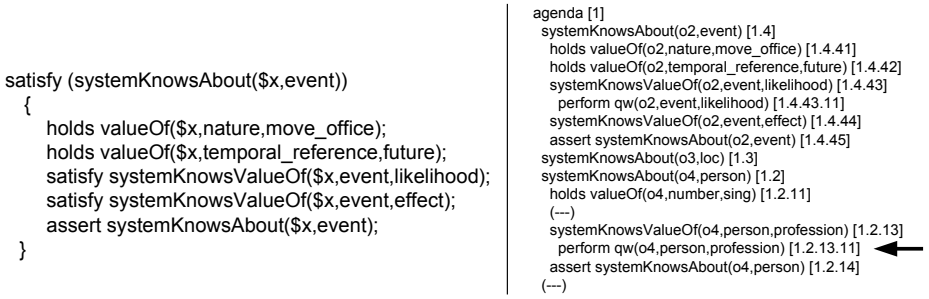

Fig. 4. Goal Satisfaction Rule (*left*) and Agenda (*right*) used by the DM

The various possible topics of conversation are organised as in an ontology, so that it is known what attributes can be expected to be present for a particular object. For example, the value of the "effect" attribute of the event object must be another object of type "event". Again this is reminiscent of Information Extraction, and the DM is in effect aiming to fill a template via clarification and supplementary questions (satisfy systemKnowsValueOf($x,event,effect)) to the point where it can be passed to the Affective Strategy Module.

The active goals are organised in a tree-structure, the so-called agenda, as shown on the right of Figure 4. At any given point in time, the agenda might contain many topics, some old, some new (systemKnowsAbout(o2,event)), some completed (---), some still open for discussion, and some not yet addressed by the system (systemKnowsValueOf(o2,event,likelihood). For each turn of the clarification dialogue, the DM chooses which topic to pursue next by considering all the currently un-satisfied goals on the agenda and heuristically rating them for importance. The heuristics employed use factors such as recency in the dialogue history, general importance, and emotional value associated with the goal. In the example in Figure 4, the system considered it more important to find out about the person (o4 or "John") than to find out about the event that the person is a participant of (o2 or "move_office"))[1].

When sufficient information has been gathered from the user through the clarification dialogue, the DM will invoke the Affective Strategy Module so it can generate a suitable tirade. The DM makes the decision to invoke the ASM using heuristics that take into account, amongst other things, the emotional value of the user's utterances and the recency of the latest ASM invocation[2].

3.5 Affective Dialogue Strategies

Previous dialogue systems [16] [17] have resorted to different models as a basis for influencing user behaviour, such as the Transtheoretical Model [18]. However, in our current scenario we are more interested in changes in attitudes rather than behaviour

[1] We are currently exploring the use of reinforcement learning with a reward function based on the emotional value of the user utterance to choose goals in a more natural way.

[2] This decision could also involve reinforcement learning.

[19]. In presenting a response to the user then, it is first necessary to understand, or appraise, the situation that the user presents to the Companion. This involves gaining an understanding of the events described and how these will affect the user. Further, the user's reaction to these events is also crucial in generating an appropriate tirade. The Affective Strategy Module (ASM) centres its response on a main event, generally the focal event selected by the DM, and its consequences for the user.

An appraisal process determines the nature of the main event in terms of both its impact on the user and the appropriateness of the user's reaction. The impact depends on whether the event constitutes an improvement (`promotion`, `payrise`) or a deterioration (`office-move`, `redundancy`, `increased-workload`) to the user's situation. This is determined by using the NLU information to instantiate an event template which indicates both the event type (e.g. improvement) and anticipated outcome based on what the event is and the information available. Every possible NLU event has its own event template within the ASM and default knowledge is used to instantiate these templates where information is not available from the NLU.

Next, the user's mood, provided by the Emotional Model, is used to determine whether the user is showing an appropriate or inappropriate emotional reaction to the event, given the anticipated outcome. This is essentially whether the user is reacting positively to improvements and negatively to deteriorations.

These details are then used to determine the strategy employed by the Companion. These strategies have been selected such that they cover the full range of possible situations a user can be in: a congratulatory strategy for when things are going well for the user, a sympathetic strategy for when they are not, encouraging or reassuring strategies for when the user's outlook is too negative and warning or cautionary strategies for when the user's outlook is too positive. The appraisal process also analyses additional influences, be they positive or negative, for the events at hand. These will be used to enrich the Companion's tirade, giving a more precise content to reassurance or warning statements.

In common with both narrative generation [20] and text generation [21], the ASM is based on planning technologies, more specifically a Hierarchical Task Network (HTN) planner [22], which works through recursive decomposition of a high level task into sub-tasks until a plan of sub-tasks that can be directly executed is produced. The HTN planning process uses the information from the event templates along with results from the appraisal as heuristics to guide its decomposition. Combined with the fact that this heuristic selection process occurs at multiple levels of the HTN, it allows for greater complexity and variance than is achievable with a scripted approach.

The resulting plan of operators provides a set of communicative functions, each targeting different aspects of the user's utterance but unified under the overall affective strategy. For instance, various operators can emphasise or play down the event consequences or comment on additional factors that may affect the course of events. The planner uses a set of 40 operators, each with multiple parameters. Overall this supports the seamless generation of hundreds of significantly different influencing strategies from the base set of influence operators.

This plan is passed to the NLG module where each operator is realised as a sentence forming part of the overall narrative utterance. The operators contain information supporting an FML-like language [23] which allows full multimodal output comprising affective TTS, gestures and facial expressions.

Figure 5 illustrates the operation of the ASM on an excerpt from an actual dialogue. The Companion first instantiates some basic information (a "bad day" event and discussion of "office politics") from the first user utterance. However, this is not enough to meet the threshold for generating an affective tirade so the DM triggers a clarification step ("tell me more …"), which actually prompts a longer and more detailed reply from the user. From this reply the system is able to instantiate further event templates, one about company restructuring, one about redundancies and one about relationships between colleagues, with the DM determining that the redundancies event template is the most prominent event. The ASM then appraises this main event, determining (from the instantiated event template) that the redundancies have not yet happened, and opting to perform a reassuring strategy. The ASM then generates a plan which shows different levels of empathy (one generic and one specific, mentioning the threat of redundancy), but also dissociates the two incidents by reminding the user that antagonistic colleagues will have no influence on redundancy decisions (this is achieved by looking for factors potentially influencing the key event, here company restructuring).

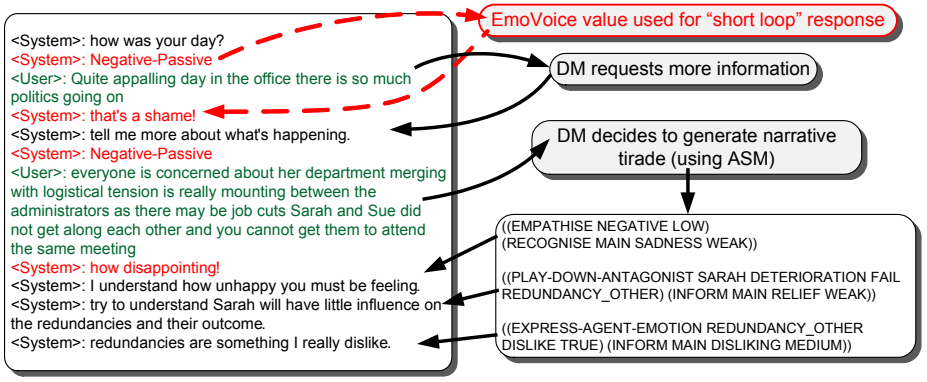

Fig. 5. An example dialogue where the user[3] discusses a negative situation and shows a correspondingly negative emotional state. Yet the Companion detects this is just a potentially bad situation and employs a reassuring affective strategy.

3.6 Handling Interruptions

Conversational flow in natural dialogues tends to be quite fluid, with partners frequently interrupting each other rather than observing the strict turn-by-turn structure of most current spoken language dialogue systems. Further, the generation of long, multi-sentence utterances by the ASM creates opportunities for the user to interrupt the Companion whilst it is speaking. Indeed, the long ASM utterances may even provoke a user interruption given that they often include advice on dealing with difficult or stressful situations the user has experienced. To resolve this, our Companion includes interaction strategies for dealing with both "barge-in"

[3] Note that user utterances show the result recognised by ASR, hence the inclusion of speech recognition errors.

interruptions and “non-barge-in” interruptions. When a user starts talking at the same time as the Companion, interrupting the Companion’s reply, this is classed as a “barge-in” interruption. We now describe the handling process (see also Figure 2).

(1) As the user may speak at any time, the Acoustic Turn Taking module must decide whether this constitutes a ‘genuine’ user interruption (as opposed to, say, backchannel). This decision is based on both the intensity and duration of the voice signal with the Interruption Manager being informed when an interruption is detected.

(2) The IM then requests that the ECA stop speaking and be given a look of surprise or irritation at being interrupted before broadcasting a notification of the interruption to all modules so they know the previous turn was not completed.

(3) The DM determines how much of the ASM response was completed.

(4) The ATT informs the IM when the interruption has ended. The IM then tracks the processing of the interrupting utterance through the system using a System State Model implemented as a two-level Finite State Machine [24]. Tracking the processing is necessary to ensure that the Companion responds within a realistic time frame.

(5) When triggered the DM must decide how to respond to that interruption.

(5a) The DM would choose to continue the interrupted utterance if the user's utterance does not provide any new information. For example, if the interrupting utterance was “I couldn't agree with you more”, then it would be reasonable for the DM to decide to continue the Companion's planned utterances from the point where the interruption took place. In Figure 6 the user interrupts the tirade in Figure 5 causing the system to stop the tirade and process the interruption. After the short loop response, the DM determines that it is not necessary to revise information and so will just ‘continue’, acknowledging the interruption and resuming the tirade from the point of interruption (that is, repeating the interrupted utterance).

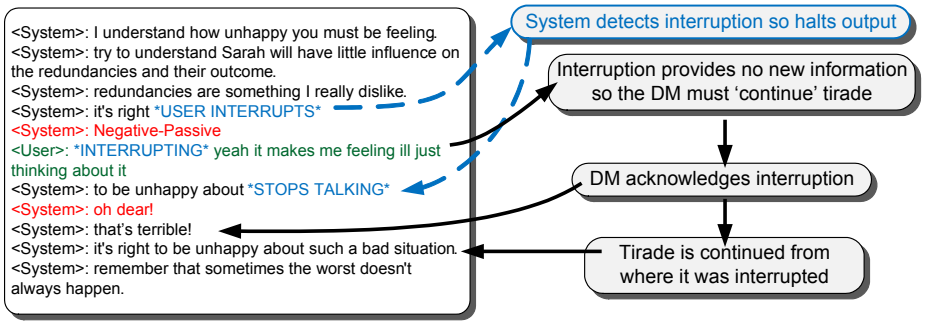

Fig. 6. An example dialogue where the user interrupts without providing new information. The Companion responds with ‘continue’ interrupt handling.

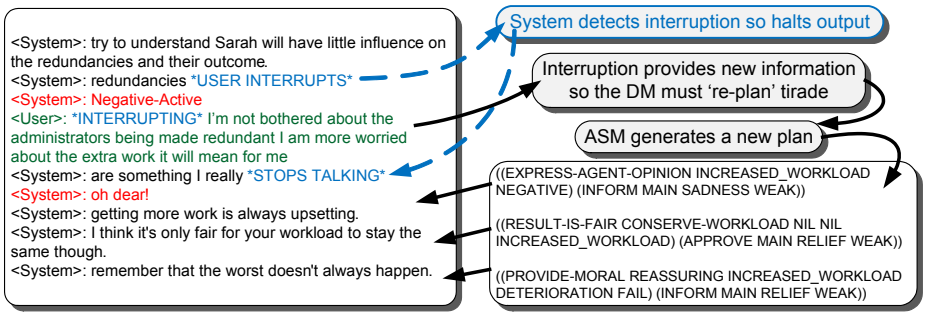

Fig. 7. An example dialogue where the user interrupts the Companion with new information. The Companion responds with 're-plan' interrupt handling.

(5b) The DM would choose to re-plan if the user's utterance provides new information. This would be the case, for example, if the user's interrupting utterance corrected what the system had just said. The re-plan is necessary because the current ASM plan was generated from a set of assumptions which have now been shown to be false or incomplete. In Figure 7 the user also interrupts the tirade in Figure 5. This time, after the short loop response, the DM determines that it is necessary to 're-plan'. The user interruption is understood as correcting the main topic to that of an increased workload for the user rather than discussion of redundancies. The tirade is then re-generated using this new main topic (with the strategy remaining reassuring). Note that it is not necessary to generate a full tirade for this new topic, as we have already relayed about half of the previous tirade, so we generate an equivalent to the remaining amount for the new tirade.

(5c) The DM chooses to abort if the user's utterance rejects the current dialogue strategy. An abort would be necessary if the interrupting utterance was something like "Don't talk to me about work, I'm not in the mood". An abort would discontinue the conversation until the user chose to continue by providing another utterance.

Handling "non-barge-in" interruptions is more straightforward as the user interrupts before the Companion has initiated its reply. The "non-barge-in" interruption can be summarised as follows:

1. The ATT detects an interrupt and informs the IM
2. The IM informs the affective dialogue processing modules
3. Affective dialogue processing modules disregard the current turn
4. The DM continues, incorporating the previous turn into the next

4 Conclusion and Results

We have presented a fully-implemented prototype of an ECA supporting affective dialogue under a truly conversational paradigm, which allows longer utterances both from the user and the agent, mixed-initiative as well as user interruptions. We conclude that our approach to the integration of conversational and affective aspects rests with the definition of interaction loops, all under the control of a top-level Dialogue Manager, orchestrating elementary dialogue steps (e.g. clarification),

narrative utterances for advice giving and user interruptions. The system has been extensively tested in the lab, in excess of a thousand sessions, and has demonstrated a regular ability to withstand meaningful dialogues of more than 10 minutes. It has reached maturity as a proof-of-concept system and is now the object of public demonstrations [25].

With respect to results, we have previously presented a validation of the affective output of our prototype [26] along with a more in-depth discussion of the generation of affective strategies, which has shown the affective content of ECA responses to be linguistically adequate in over 80% of cases. We continue to expand the linguistic coverage of our prototype and now seek to carry out extensive user evaluations involving prolonged use of the system. Such a systematic evaluation of our Companion will require the development of a specific methodology measuring the appropriateness of the ECA's responses locally as well as over the whole dialogue.

Acknowledgments. This work was funded by the Companions project (http://www.companions-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434. The EmoVoice system has been used courtesy of the Multimedia Concepts and Applications Group of the University of Augsburg.

Other contributors to the prototype described in this paper from the COMPANIONS consortium include: Wei Wei Cheng, Morena Danieli, Carlos Sanchez Fernandez, Debora Field, Mari Carmen Rodriguez Gancedo, Jose Relano Gil, Ramon Granell, Jaakko Hakulinen, Sue Harding, Topi Hurtig, Oli Mival, Roger Moore, Lei Ye and Enrico Zovato.

References

1. André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (Eds.): Affective Dialogue Systems, Tutorial and Research Workshop, In: Proceedings of ADS 2004, LNCS 3068, Springer, Kloster Irsee, Germany (2004)
2. Englemore, R. and Morgan, T.: Blackboard Systems. Addison-Wesley, New York (1988)
3. Vogt, T., André, E. and Bee, N.: EmoVoice – A framework for online recognition of emotions from voice. In: Proceedings of Workshop on Perception and Interactive Technologies for Speech-Based Systems, Springer, Kloster Irsee, Germany (2008)
4. Moilanen, K., Pulman, S.: Sentiment Composition. In: Proceedings of the Recent Advances in Natural Language Processing International Conference (RANLP 2007), Borovets, pp. 378--382 (2007)
5. Bickmore, T., and Cassell, J.: Small Talk and Conversational Storytelling in Embodied Interface Agents. In: Proceedings of the AAAI Fall Symposium on Narrative Intelligence, pp. 87-92. Cape Cod, MA. (1999)
6. De Angeli, A., Brahnam, S.: I hate you! Disinhibition with virtual partners. In: Interacting with Computers 20(3), pp. 302--310 (2008)
7. Morency, L.-P., de Kok, I., Gratch, J.: Predicting listener backchannels: A probabilistic multimodal approach. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 176--190. Springer, Heidelberg (2008)
8. Kopp, S., Stocksmeier, T., Gibbon, D.: Incremental multimodal feedback for conversational agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 139--146. Springer, Heidelberg (2007)

9. Bevacqua, E., Mancini, M., and Pelachaud, C.: A listening agent exhibiting variable behaviour. In: Proceedings of IVA 2008, Lecture Notes in Computer Science, pp. 262--269, Springer Berlin/ Heidelberg (2008)
10. Grishman, R.: Information Extraction: Techniques and Challenges. In: Maria Teresa Pazienza (Ed.) Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, volume 1299 of Lecture Notes in Artificial Intelligence, pp. 10--27. Springer. International Summer School, SCIE-97. Frascati, Italy (1997)
11. Jönsson, A., Andén, F., Degerstedt, L., Flycht-Eriksson, A., Merkel, M., and Norberg, S.: Experiences from combining dialogue system development with information extraction techniques. In: Mark T. Maybury (Ed), New Directions in Question Answering, AAAI/MIT Press (2004)
12. Kennedy, C. and Boguraev, B.: Anaphora for everyone: Pronominal anaphora resolution without a parser. In: Proceedings of COLING 1996, ACL, pp. 113--118, Copenhagen (1996)
13. Boye, J. and Gustafson, J.: How to do dialogue in a fairy-tale world. In: Proceedings of the 6th SIGDial workshop on discourse and dialogue, Lisbon, Portugal (2005)
14. Boye, J., Gustafson, J. and Wirén, M.: Robust spoken language understanding in a computer game. In: Journal of Speech Communication, 48, pp. 335--353 (2006)
15. Boye, J. Dialogue management for automatic troubleshooting and other problem-solving applications. In: Proceedings of the 8th SIGDial workshop on discourse and dialogue, Antwerp, Belgium (2007)
16. Cavalluzzi, A., Carofiglio, V., de Rosis, F.: Affective Advice Giving Dialogs. In: Proceedings of ADS 2004, LNCS 3068, pp. 77--88, Springer, Kloster Irsee, Germany (2004)
17. Bickmore, T., Sidner, C.L.: Towards Plan-based Health Behavior Change Counseling Systems, In Proceedings of AAAI Spring Symposium on Argumentation for Consumers of Healthcare, Stanford, CA (2006)
18. Prochaska, J., Di Clemente, C., Norcross, H: In search of how people change: applications to addictive behavior. In: American Psychologist, 47 , pp. 1102--1114 (1992)
19. Tørning, K., Oinas-Kukkonen, H.: Persuasive system design: state of the art and future directions. In: Proceedings of PERSUASIVE 2009, vol. 350, New York, NY, USA (2009)
20. Cavazza, M., Charles, F., and Mead, S.J.: Character-Based Interactive Storytelling. In: IEEE Intelligent Systems 17(4), pp. 17-24 (2002)
21. Appelt, D.E.: Planning English sentences. Cambridge University Press, Cambridge (1985)
22. Nau, D., Ghallab, M., Traverso, P. Automated Planning: Theory & Practice, Morgan Kaufmann Publishers Inc., San Francisco, CA (2004)
23. Hernández, A., López, B., Pardo, D., Santos, R., Hernández, L., Relaño Gil, J., Rodríguez, M.C.: Modular definition of multimodal ECA communication acts to improve dialogue robustness and depth of intention. In: Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., and Vilhjálmsson, H. (Eds.), AAMAS 2008 Workshop on Functional Markup Language (2008)
24. Crook, N., Smith, C., Cavazza, M., Pulman, S., Moore, R. & Boye, J.: Handling User Interruptions in an Embodied Conversational Agent. In: Proceedings of the AAMAS International Workshop on Interacting with ECAs as Virtual Characters, pp. 27-33, Toronto (2010)
25. Cavazza, M., Santos de la Camara, R., Turunen, M., and the Companions consortium.: How was your day? A Companion ECA. In Proceedings of AAMAS 2010, accepted for publication (demonstration paper), Toronto (2010)
26. Cavazza, M., Smith, C., Charlton, D., Crook, N., Boye, J., Pulman, S., Moilanen, K., Pizzi, D., Santos de la Camara, R., Turunen, M.: Persuasive Dialogue based on a Narrative Theory: an ECA Implementation. In: Proceedings of PERSUASIVE 2010, Copenhagen (2010).