LAST MINUTE: a Multimodal Corpus of Speech-based User-Companion Interactions

Dietmar Rösner (1), Jörg Frommer (2), Rafael Friesen (1),
Matthias Haase (2), Julia Lange (2), Mirko Otto (1)

1: Otto-von-Guericke Universität, Institut für Wissens- und Sprachverarbeitung,
Postfach 4120, D-39016 Magdeburg
{roesner, friesen, miotto}@ovgu.de
2: Otto-von-Guericke-Universität, Universitätsklinik für Psychosomatische Medizin und Psychotherapie
Leipziger Straße 44 D-39120 Magdeburg
{joerg.frommer, matthias.haase, julia.lange}@med.ovgu.de

Abstract

We report about design and characteristics of the LAST MINUTE corpus. The recordings in this data collection are taken from a WOZ experiment that allows to investigate how users interact with a companion system in a mundane situation with the need for planning, re-planning and strategy change. The resulting corpus is distinguished with respect to aspects of size (e.g. number of subjects, length of sessions, number of channels, total length of records) as well as quality (e.g. balancedness of cohort, well designed scenario, standard based transcripts, psychological questionnaires, accompanying in-depth interviews).

Keywords: User-Companion-Interaction, Multimodal, Wizard-Of-Oz

1. Introduction

"Really natural language processing" (Cowie and Schröder, 2005), i.e. the possibility that human users speak to machines just as they would speak to another person, is a prerequisite for many future applications and devices. It is especially essential for so called companion systems.

Wilks et al. describe companion systems as follows: "By Companions we mean conversationalists or confidants – not robots – but rather computer software agents whose function will be to get to know their owners, who may well be elderly or lonely, and focusing not only on assistance via the internet (contacts, travel, doctors etc.) that many still find hard to use, but also on providing company and companionship, by offering aspects of personalization" (Wilks, 2010).

Companion systems are investigated in a number of consortia, among them are the EU funded Companions project[1] and the special research area (Sonderforschungsbereich/Transregio 62; SFB/TRR 62) 'A companion technology for cognitive technical systems' funded by the German National Science foundation (DFG)[2].

There is broad agreement that recording humans interacting in an environment of interest (e.g. SAL scenario (Douglas-Cowie et al., 2008) or companion scenario (Legát et al., 2008; Webb et al., 2010)) is a fundamental step towards assessing machine-human interactions within such scenarios (McKeown et al., 2010) .

In the following we report about the LAST MINUTE corpus. The recordings in this data collection are taken from a WOZ experiment that allows to investigate how users interact with a companion system in a mundane situation with the need for planning, re-planning and strategy change.

---

[1] http://www.companions-project.org/
[2] http://www.sfb-trr-62.de/

2. A WOZ experiment in UCI

2.1. The WOZ dialog

In the following we give a detailed look at the course of the interaction between user and a WOZ simulated companion system in the experiment. We first discuss the normal unproblematic course of dialog turns. Then we give an analysis of various error or problem situations during the user companion interaction.

2.1.1. Global structure

The overall structure of an experiment is divided into

- a personalization module, followed by

- the 'last minute' module.

These modules serve quite different purposes and are further substructured in a different manner (cf. below).

The personalization module is decisively organised completely independent of the 'last minute' module and could easily be combined with different tasks or problem solving modules for other WOZ szenarios.

An abstract view of the last minute module:

1. A cover story is provided to the subject and the companion system is initialised by a personalization dialogue. Some limitations such as the available time of the experiment are told, some not. The cover story is provided very detailed to stimulate ego involvement and imagination of the subject. The system's speech is focussed primarily on the ideational metafunction.

2. The subject starts with the actual task while being supported by the system. The system alerts the subject when limitations are infringed and provides information about the task's status.

3. Further information is provided that let the subject realize that he had aimed a wrong goal and has now to change his strategy. This might lead to frustration and anger.

4. The system addresses the subject using the interpersonal and textual metafunctions with increased intensity. Selected dialogue strategies create the chance for reflection and expression of anger. The dialogue strategies provide empathic help based on the principles of Rogers' paradigm of client centered psychotherapy (Rogers, 1959).

5. The subject gets the chance to revise a limited number of former decisions in this stage. The system's speech is focussed primarily on the ideational metafunction again.

6. The system informs the subject about the end of the session. At this point the subject has to rate his own performance. The goal for the evaluation in this stage is to evaluate the style of attribution as well (self or external).

2.1.2. Personalization module

The subjects are instructed that they will interact with a speech driven system and that they can begin the interaction by saying that they want to start.

The system then welcomes the subject, gives a short self description and prompts the subject to tell and spell his name.[3] All system output is pronounced via a text to speech system (TTS). The language quality of the TTS has decisively been chosen in such a way that the voice is clearly identified by human hearers as 'computer voice'.[4] [5] [6]

> guten tag und herzlich willkommen (.) sie sprechen hier mit dem prototypen eines computerprogramms (.) dieses soll nutzer in der bewältigung von alltagsaufgaben unterstützen (.) das besondere an diesem neuen computerprogramm ist (.) dass es sich individuell an seinen nutzer anpasst (.) zu diesem zweck werden im verlauf dieser sitzung einige aufgaben und testsituationen durchlaufen (1.02) bitte nennen und buchstabieren sie zunächst ihren vor und zunamen
> *[Welcome (.) you are talking to a prototype of a computer program (.) it can support users in the handling of mundane tasks (.) the remarkable thing about this new computer program is (.) that it adjusts individually to its user (.) for this purpose some tasks and test situations will be done during this session (1.02) please give and spell your name]*

This mode of interaction with system initiative only, i.e. the system asks a question or gives a prompt, is dominant throughout the whole personalization module. In other words this module is a series of dialog turns that are made up by a system question or prompt followed by the user's answer or reaction. In some sense this module thus resembles more an investigative questioning than a symmetric dialog.

Collecting personal data When the subject has given and spelled his name, the system prompts the subject to (further) introduce himself.

> zu beginn sind einige angaben zu ihrer person notwendig (.) können sie sich bitte vorstellen
> *[first some information about you is required (.) can you please introduce yourself]*

Depending on the amount of detail given by the subject in his self introduction the system then prompts for missing data.

> damit sich das computerprogramm individuell an sie anpassen kann (.) sind einige konkrete informationen zu ihrer person erforderlich (.) können sie bitte zu folgenden punkten angaben machen (–) ihr wohnort (-) ihre familie (-) ihre körpergröße (-) ihre konfektionsgröße (-) ihre schuhgröße
> *[in order to adjust the system individually to you (.) some information about you is necessary (.) can you please give information about (–) your home (-) your family (-) your body size (-) your dress size (-) your shoe size]*

If the subject does not give all requested information he is reprompted for missing details. Very often male subjects do not know their size of clothing ('konfektionsgröße').

> bitte ergänzen sie angaben zu ihrer konfektionsgröße
> *[please complete the information about your dress size]*

When all information is given the system utters a summary of the collected information and asks for a confirmation that all data is correct. [7] [8]

> sie heißen yyyyyyyy xxxxxxx (.) sie sind dreiundsiebzig jahre alt (.) sie wohnen in magdeburg (.) zurzeit sind sie in rente (.) ihre familie besteht aus einer frau (.) sie sind ein meter und zweiundsiebzig zentimeter groß(.) ihre konfektionsgröße ist fünfzig (.) ihre schuhgröße ist einundvierzig (–) sind diese angaben korrekt
> *[your name is yyyyyyyy xxxxxxx (.) you are thirty seven years old (.) you live in magdeburg (.) currently you are retired (.) your family consists of a wife (.) you are one meter and seventy two centimeters tall (.) your dress size is fifty (.) your shoe size is forty one (–) is this information correct]*

If the user confirms correctness, then this subphase ends, otherwise wrong data can be corrected.

Relevance of system feedback The summary of personal information that the system has acquired and is then presenting to the user for confirmation or correction is the first explicit feedback [9] uttered by the system about processing or 'understanding' results that the subjects experience. All preceeding (and most of the following) dialog turns in the personalization module are of the investigative type of system question followed by subject's answer with no explicit indication if the answer is processed by the system at all.[10] Please note, that the capabilities in automated speech recognition (ASR) and natural language understanding (NLU) needed for such a type of automated system response are easy to be realised with currently available technology and

---

[3]We give the original German texts used in the experiment and English glosses for convenience.

[4]This is a parameter to be systematically changed in future experiments.

[5]The gender of the current TTS voice is male, another parameter to be systematically changed in future experiments.

[6]All excerpts from transcript are given - unless otherwise noted - with the GAT 2 minimal coding.

[7]This anonymised verbatim example and - unless otherwise noted - all others are taken from the transcript of subject 20110221awb.

[8]The codes for subjects are comprised from the date of the experiment as yyyymmdd followed by a letter for the first (a), second (b) etc. experiment at this date and the intitials of the subject.

[9]Except for displaying the subject's name on the screen after the subject has introduced himself

[10]The only exceptions - variants of indirect feedback - are the reprompting for personal information still missing, cf. above, or the stimulus to tell more when a user's answer to one of the open questions is too short, cf. below.

remain within the scope of standard slot-filler based approaches to information collection in speech based dialog systems.

Data correction subdialog If the subject denies that the data mirrored by the system are correct the system prompts for a correction (like here for subject 20110627abs).

> bitte korrigieren sie
> *[please do correct]*

When the user has corrected incorrect data, the system again gives feedback (cf. above) and prompts for confirmation.

For N = 130 experiments in N_1 = 16 a data correction subdialog was necessary.

Recalling experiences After the subphase of collecting personal data from the subject a subphase with a number of prompts to recall events or experiences is entered. In this subphase the system stipulates user narratives on the following topics:

- a recent event that the subject enjoyed very much

- a recent event that made the subject very angry

- the hobbies of the subject

- which technical devices are used by the subject for which purposes in daily life

- an event where one of these technical devices was especially helpful and where the subject made good experiences with the device

- an event where one of these technical devices was not helpful at all and where the subject made bad experiences with the device

For the last two questions the system may ask within a follow up question if additional technical devices exist for which the same or similar experiences hold.

Politeness In all modules the system uses the polite version, the German 'Sie' (polite, formal German version of 'you') when addressing the user.

How users approach the system differs significantly. Some subjects avoid any personal pronouns when adressing the system, others employ the German 'du' (informal German version of 'you') and only very seldom the German 'Sie' is used. This issue will be further investigated in detail because it is one of a number of indicators of the way how subjects experience the system.

2.1.3. The last minute module

The last minute module starts with a narrative of the system that shall stimulate user's imagination and ego involvement. This exposition ends with the system's question if the subject wishes a repetition of the information.

> es ist mitten im sommer (.) es regnet seit tagen (.) viele ihrer verwandten und freunde sind bereits in den sommerurlaub gefahren (.) überraschend erreicht sie die nachricht (.) dass sie eine reise gewonnen haben (.) es soll für vierzehn tage nach waiuku gehen (.) einem kleinen urlaubsort am meer (1.77) im gewinn ist enthalten (.) dass sie ihr reisegepäck aus einem katalog individuell zusammenstellen können (.) allerdings wird ihr flugzeug noch heute starten (.) ihr taxi zum flughafen ist bereits bestellt und wird sie in knapp fünfzehn minuten abholen (.) diese zeit bleibt zur auswahl ihres reisegepäcks

(.) der koffer mit den von ihnen ausgewählten artikeln wird am flughafen für sie bereit gestellt (.) mit gepacktem koffer steht ihrer reise nichts mehr im wege (1.0) wünschen sie eine wiederholung dieser information
> *[it is in the middle of summer (.) it is raining for days (.) many of your relatives and friends already left for summer holidays (.) surprisingly you are informed (.) that you have won a holiday trip (.) for fourteen days to waiuku (.) a small holiday location at the sea (1.77) the prize includes (.) that you can choose your luggage individually from a catalogue (.) your plane will take off today (.) your taxi to the airport is already appointed and will pick you up in fifteen minutes (.) this is the time that remains to choose your luggage (.) the suitcase with the chosen items will be prepared for you at the airport (.) with the packed suitcase nothing will get in the way of your holidays (1.0) do you want this information to be repeated]*

Depending on the subject's answer, either the information is repeated or further information about the task of packing the suitcase by choosing from a fixed series of presented categories is given.

> auf dem bildschirm sehen sie jetzt die im katalog enthaltenen rubriken (.) jede rubrik beinhaltet eine vielzahl von artikeln (.) die rubriken werden nacheinander aufgerufen (.) so dass sie aus jeder rubrik artikel für ihren vierzehntägigen urlaub auswählen können (.) bitte geben sie zu jedem artikel die bezeichnung und die gewünschte stückzahl an (–) in der mitte des bildschirmes sehen sie den zu packenden koffer
> *[you can now see the categories contained in the catalog on the screen (.) each category contains a number of items (.) the categories will be selected one after another (.) so you can choose items from each category for your fourteen day holidays (.) please give the name and the desired quantity for each item (–) you can see the suitcase to be packed in the center of the screen]*

{09:38} 079 W während der artikelauswahl werden weitere informationen zum urlaubsort waiuku eingeholt (1.0) ein wichtiger hinweis (.) es stehen nun knapp fünfzehn minuten zur auswahl von artikeln zur verfügug (.) bitte beachten sie bei der auswahl ihres reisegepäcks (.) dass sie für vierzehn tage verreisen (1.0) sie können jetzt aus der rubrik oberteile auswählen (.) wenn sie die auswahl aus dieser rubrik beendet haben (.) sagen sie bitte (.) dass sie zur nächsten rubrik übergehen möchten
> *[during the selection of the items additional information about the holiday destination waiuku are collected (1.0) an important hint (.) the selection of the items has to be done in fifteen minutes (.) please keep in mind while selecting your luggage (.) that you travel for fourteen days (1.0) you can now choose items from the category tops (.) when you finished your selection (.) please say (.) that you want to switch to the next category]*

Selection Now follows the main part of 'last minute'. The subject is expected to choose items from twelve different categories that are presented in a fixed order (cf. 2.1.3.). In a simplified view we thus have an iterative structure made up from twelve repetitions of structurally similar subdialogs each for the selection from a single category. The options of each category are given as menu (with icons, cf. fig. 1) on the subject's screen.

Normal packing subdialog In a normal packing subdialog we essentially have a series of adjacency pairs made up of a user request for a number of items (more precisely: a user request for a number of instances from an item type) from the current selection menu (e.g. 'ten t-shirts') followed by a confirmation of the system (e.g. 'ten t-shirts have been added').[11]

---

[11]The confirmation is more detailed when uttered the first time by the system. It then contains as well an explicit request 'ten t-shirts have been added. please continue'. All subsequent confirmations use the shorter version.
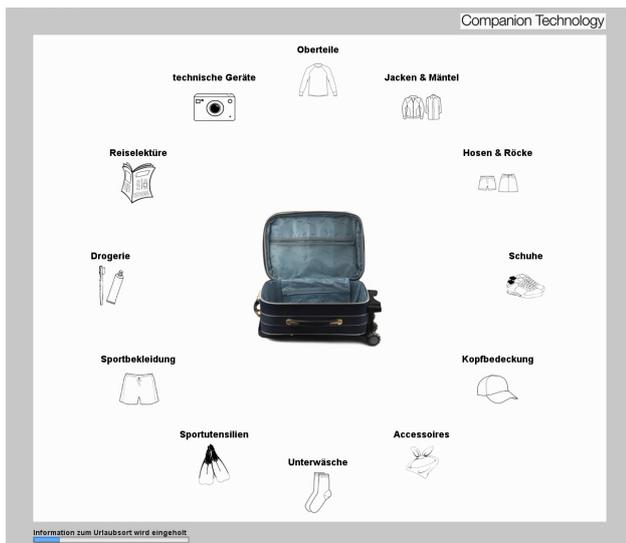
Figure 1: The subjects screen showing the category overwiev.

Temporal constraints For the whole packing a total of 15 minutes are allocated (global time constraint). Given twelve categories and three optional repeated categories during re-packing the average time for a single category must not exceed one minute (local time constraint). The global time constraint (and thus the local time constraint implicitly) is explicitly given twice to the subjects in the initial exposition of the task ('. . . the taxi to the airport will pick you up in approximately fifteen minutes . . . ', '. . . an important hint (.) you now have approximately fifteen minutes for the selection of items . . . '). The local time constraint is enforced and made explicit in case that the user spends too much time within a single category (cf. below).

The categories The categories offered to the subjects are in sequence of appearance (cf. fig. 1):
tops, coats, trousers and skirts, shoes, hats, accessory, underwear, sports equipment, sportswear, drugstore products, travel reading, technical devices.

Change of category There are two ways to finish the current category and to proceed to the next one:

- the user explicitly asks for a change of the category,

- the system changes the category because a time limit is reached.

In the latter case the system informs the user that the selection from the current category has to end and that the following category is now available.

> die auswahl von artikeln aus der rubrik sportbekleidung muss jetzt beendet werden (.) um die aufgabe in der zur verfügung stehenden zeit beenden zu können (–) sie können jetzt aus der rubrik drogerieartikel auswählen
>
> *[the item selection from the category sportswear has now to be finished (.) in order to complete the task in time (–) you can now choose from category drugstore products]*

Barriers The normal sequence of repetitive subdialogs with choices from a total of twelve categories is modified for all subjects at specific time points.
These modifications are:

- after the sixth category, the system informs the user that it will take more time to get information about the target location, in addition the current contents of the suitcase are listed verbally (listing barrier),

  > das beschaffen zusätzlicher informationen zum urlaubsort verzögert sich um einige minuten (.) bitte haben sie noch etwas geduld (-) die hälfte aller rubriken wurde von ihnen bearbeitet (.) es folgt ein zwischenstand ihrer artikelauswahl (.) folgende artikel wurden bereits ausgewählt (.) ein tshirt (-) vier achselshirts (-) zwei pullover (-) zwei langarmshirts (-) eine regen und windjacke (-) eine sommerjacke (-) eine strickjacke (-) ein anorak (-) eine jeans (-) eine kurze hose (-) ein paar badelatschen (-) ein paar turnschuhe (-) ein paar halbschuhe (-) ein paar wanderschuhe (-) eine schirmmütze (-) eine sonnenbrille (-) sie können jetzt mit der auswahl aus der rubrik unterwäsche fortfahren
  >
  > *[the collection of additional information about the holiday destination will be delayed a few minutes (.) please be patient (-) half of the categories are now completed (.) an overview of your current item selection follows (.) the chosen items are (.) a tshirt (-) four tops (-) two pullovers (-) two longsleeve shirts (-) a rain and wind coat (-) a summer jacket (-) a cardigan (-) an anorak (-) a jeans shorts (-) one pair bathing shoes (-) one pair sport shoes (-) one pair low shoes (-) one pair hiking boots (-) a cap (-) sunglasses (-) you can now choose from category underwear]*

- during the eighth category, the system refuses to pack selected items for the first time because the weight limit for the suitcase is reached. The user is informed that first other items have to be unpacked and that on demand a listing of the current contents can be given (weight limit barrier).

  > eine badehose kann nicht hinzugefügt werden (.) anderenfalls würde die von der fluggesellschaft vorgeschriebene maximale gewichtsgrenze des koffers überschritten werden (.) bevor weitere artikel ausgewählt werden können (.) müssen sie für genügend platz im koffer sorgen (.) hierfür können bereits eingepackte artikel wieder ausgepackt werden (.) auf nachfrage erhalten sie eine aufzählung der bereits ausgewählten artikel
  >
  > *[a pair of swimming trunks can not be added (.) otherwise the maximal luggage weight limit of the airline would be exceeded (.) before further items can be chosen (.) you have to care for enough space in the suitcase (.) for this you may unpack already packed items (.) you can get a listing of all chosen items upon request]*

- at the end of the tenth category, the system informs the user that now more detailed information about the target location Waiuku is available. This information is again given verbally in detail together with some illustrations on the subject's screen (Waiuku barrier).

Intervention Nearly half of the subjects - randomly chosen - get an empathic intervention designed according to the principles of Rogerian psychotherapy (Rogers, 1959) after the Waiuku barrier. It comprises three system utterances. First the subjects are asked if they had chosen other items when the weather conditions would have been available earlier.

> (1.92) wegen einer unterbrechung der datenleitung konnten die informationen über den zielort nicht schneller beschafft werden (.) dadurch hat sich die situation für sie möglicherweise überraschend geändert (.) die ausgewählten artikel lassen darauf schließen (.) dass sie sich auf anderes wetter eingestellt haben (.) wenn ihnen die witterungsverhältnisse am zielort bekannt gewesen wären (.) hätten sie sich dann

womöglich für andere artikel entschieden (.) mich interessiert ihre meinung dazu
[because of an interrupted data line the information about the holiday destination could not be delivered earlier (.) because of this the situation may have changed for you (.) the chosen items show (.) that you prepared for different weather (.) if you knew the weather conditions at the travel destination (.) would you have packed different items (.) i am interested in your opinion]

Then they are asked if they have experienced unpleasant feelings due to the situation.

sind durch diese situation auch unangenehme gefühle aufgetaucht (.) wenn ja (.) können sie diese beschreiben
[did uncomfortable feelings occur because of this situation (.) if yes (.) can you describe them]

Finally the system expresses the hope that the subject will still be engaged in the subsequent experiment.

ich hoffe (.) dass ihre lust (.) an dieser aufgabe mitzuwirken (.) darunter nicht allzu sehr leidet
[i hope (.) your willingness (.) to participate in this task (.) did not suffer too much]

Finalization After the optional intervention or, when no intervention was given, immediately after the Waiuku barrier the subjects work through the two remaining categories. Then they get the opportunity to choose again from up to three categories of their own choice.

{22:12} 298 W die auswahl von artikeln aus den rubriken reiselektüre und technische geräte ist nun abgeschlossen (.) ihnen stehen insgesamt noch drei minuten zur auswahl ihres reisegepäcks zur verfügung (.) es können nun rubriken ihrer wahl ein zweites mal bearbeitet werden (.) wenn sie dies möchten (.) welche änderung ist ihnen am wichtigsten
[the item selection from the categories travel lecture and technical devices is now finished (.) you have three minutes left to choose items (.) you can now choose from categories a second time (.) if you want this (.) which change is most important to you]

In addition the remaining time is running short and the system informs the user about this.

{23:16} 318 W ihnen stehen insgesamt noch zwei minuten zur auswahl ihres reisegepäcks zur verfügung (-) aus welcher rubrik möchten sie nun auswählen
[you have two minutes left to choose items (-) from which category do you want to choose now]

We distinguish two types of end of the selection phase:

- the user ends the selection on his own or

- the system ends the selection due to the global time limit reached.

{24:44} 353 W die auswahl von artikeln wird hiermit beendet (.) das taxi zum flughafen wartet bereits vor der tür (.) der koffer mit den von ihnen ausgewählten artikeln wird am flughafen für sie bereit gestellt (—) abschließend noch einige fragen (.) bitte beschreiben sie (.) wie sie ihren urlaub unter den gegebenen bedingungen gestalten wollen
[the item selection is now finished (.) the taxi to the airport is already waiting at the door (.) the suitcase with the item you chose will be prepared at the airport (—) at last a few questions (.) pleases describe how you will arrange your holidays unter the given circumstances]

From the total of N = 130 subjects, $n_1 = 40$ subjects (with $n_2 = 20$ from the elderly and $n_3 = 20$ from the young group) ended the selection on their own, in the other $n_4 = 90$ cases the system closed the session and blocked additional user input.

Reflection of result After the end of the selection phase the system prompts the user with three final questions:

- How will you organise your holidays under the given constraints?

- How content are you with the contents of the packed suitcase?

- Would you travel to Waiuku with this suitcase?

Last words The system closes the session, thanks the user for his cooperation and says goodbye. Many users answer with (variants of) goodbye as well.

{26:50} 366 W mit abschluss dieser aufgabe wurde das ende der sitzung erreicht (.) vielen dank für ihre mitarbeit und auf wiedersehen
[with finishing this task the end of the session is reached (.) thank you very much for your cooperation and good bye]

2.2. Characteristics of the sample

The total cohort (N = 130) is balanced with respect to gender, age and educational level. The young group is 18-28 years, the elder group over 60 years old. The educated group has passed the Abitur (German university entrance qualification), the less educated group has a lower qualification.

A complete WOZ session takes approx. 30 minutes. The total lengths of sessions varies from 19 to 39 minutes.

3. The LAST MINUTE corpus

3.1. Multimodality

During the WOZ-experiment multimodal data is collected, i.e. video, audio and biopsychological data. Due to a few complications with recording many channels synchronously some experiments were recorded with different hardware, but most of the experiments were recorded with

- 4x HD camera (Pike F145C), 1388x1038px, 25fps

- 2x stereo camera (Bumblebee2 BB2-03S2C), 1280x480px, 25fps

- 2x directional microphone (Sennheiser ME66), mono, 44100Hz, wav

- wireless headset (t-bone earmic 500), mono, 44100Hz, wav

- skin reductance, heartbeat, respiration (NeXus32), 512samples/sec

- webcam for observation

- screencast of subject's screen

- TTS audio stream

- log files of the system's utterances

The subject room was furnished combining the devices necessary for the experiment with a livingroom-like setting. The arrangement was chosen to be emotional neutral and comfortable. The devices necessary for the experiment were placed on the desk as shown in figure 2.
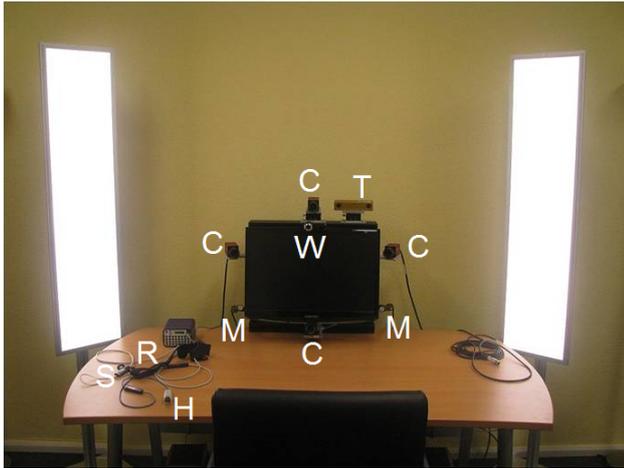
Figure 2: The hardware setting in the subject room. C=High resolution camera, H=Heart beat clip, M=Microphone, R=Respiration belt, S=Skin conductance clip, T=Stereo camera, W=Observation webcam. Not in the picture: Headwear microphone.

3.2. Questionnaires

An early established method for evaluation of personality traits and other psychological factors are psychometric questionnaires (Lienert, 1961). The subjects received the following psychological questionnaires after the experiment:

- German version (ASF-E) (Poppe et al., 2005) of the attributional style questionnaire (ASQ)

- NEO Five-Factor Inventory (NEO-FFI) (Borkenau and Ostendorf, 2008)

- Inventory of interpersonal problems (IIP-C) (Hoffmann et al., 2010)

- Stress Processing Questionnaire (SVF) (Erdmann and Jahnke, 2008)

- Emotion Regulation Questionnaire (ERQ (Gross and John, 2003))

- BIS/BAS (Carver and White, 1994)

- AttrakDiff (Hassenzahl et al., 2001)

- In addition to these psychometric instruments socio-demographic variables such as marital status, age, gender and computer literacy are collected.

Answering the questionnaire takes about 90 minutes.

3.3. Interviews

After the WOZ experiment about half of the participants undergo a semi-structured interview to determine the subjective experience of the experiment. The interviews took about 30 to 160 minutes and are audio recorded.

3.4. Transcripts

WOZ experiments and semi-structured interviews are transcribed by trained personnel using the rules for GAT 2-minimal transcript (Selting, 2009) and the transcription software FOLKER (Schmidt and Schütte, 2010). Minimal transcripts take dialect into account. Many of our subjects use elements of (German) dialects.Folker saves transcripts in an XML format which can be easily processed by analyzing software.

3.5. Comparison with related work

For a comprehensive discussion of other available corpora with naturalistic data cf. (McKeown et al., 2010).
Naturalistic data are either taken from sources like TV programs (e.g. (Grimm et al., 2008)) or are collected via designed and controlled experiments (e.g. (Douglas-Cowie et al., 2008)). Problems with the former approach are discussed in (McKeown et al., 2010).
The LAST MINUTE corpus has a number of distinguished features that go beyond other available corpora:

- large number of subjects: Whereas the cohort size of other data sets is in the range between 10 and 20 (cf. e.g. (McKeown et al., 2010)) our corpus comprises records from WOZ sessions with a total of N = 130 subjects.

- balanced participants with respect to different criteria: In many other samples there are only students involved in the experiments (cf. e.g. (McKeown et al., 2010)) whereas our cohort is balanced with respect to gender, age and educational level.

- length of sessions: A typical WOZ session (resulting in resp. recordings) takes approx. 30 minutes per user. This session size exceeds currently available material with record sizes of up to five minutes per user (cf. e.g. (McKeown et al., 2010)) and comprises different phases with a varying potential for arousal and for positive or negative experiences. In sum: the total length of records (synchronously taken in the diverse channels) for one channel (e.g. audio, video, ...) sums up for the experiments (N = 130) to more than 70 hours real time. For the interviews (N=73) an additional sum of 93 hours of audio only records are available.

- number and quality of recorded channels: cf. 3.1.

- data about subjects from psychological questionnaires: cf. 3.2. To our knowledge such additional sources are not available for any of the currently accessible corpora.

- for a subgroup: records from post hoc semi-structured interviews
  These interviews enable the subjects to reflect and explain his or her subjective experiences during the interaction in a free, nonrestricted way. The questions focussing issues relevant for the research aims concern

  – occurred user emotions,

- intentional ascriptions towards the CS to explain and predict the system's behaviour (like characteristics, aims, emotions etc.; (Dennett, 1987)),
- the speech based interaction,
- the intervention (if given),
- the role of technical systems in autobiography and
- the general evaluation of the system.

The presentation of these questions is handled flexible, i.e. the formulation and chronological order of the questions can be adapted to each individual interview.

To our knowledge such post hoc interviews focussing on an in depth reflexion of the subject's experiences during the experiments are not available for any of the other currently accessible corpora.

4. Usage scenarios for the corpus

4.1. Evaluation

4.1.1. Wizard logs

The wizards have been trained and their behaviour has been anticipated and prescribed as detailed as possible in a manual[12] (Frommer et al., 2010).

All dialog contributions from the system (i.e. wizard) were pronounced by a TTS. After a WOZ session all wizard contributions together with their timings are available as additional log file.

Evaluation of the wizard log files already allows to classify the overall interaction of different subjects with respect to a number of aspects. For other classifications NLP analysis of the contents of the subjects' utterances is necessary. In the following we will report about results from the former analyses.

4.1.2. Classifying outcomes

How to compare subjects with respect to the different outcomes of the experiments? What are appropriate measures of effectiveness and efficiency in their dialog behaviour and their problem solving?

The problem solving dialog in 'last minute' is organised as a series of (primarily system controlled) dialog turns made up from user requests and system reactions. Each turn can either be (locally) successful or it may fail. A user request (e.g. to pack or unpack a number of items or to switch to another category) that can be realised will always be explicitly acknowledged by the system. The system response in the failure case is dependent on the type or cause of failure. This cause may e.g. be a prompt for repetition or an information or a suggestion for an alternative action etc. In each case success or failure of an adjacency pair (dialog turn made up from a user contribution and a subsequent system reaction) can be easily decided based on the wording of the logged system response (i.e. there is no need for an NLP analysis of the user's contribution for this purpose). We distinguish two types of measures:

- domain related measures
- discourse related measures

4.1.3. Domain related measures

The evaluation of the contents of the packed suitcase allows to judge the overall success. Did the user manage to pack essential items (e.g. warm clothing) for the weather conditions?

4.1.4. Discourse related measures

Some dialog turns will fail by design for all subjects. For example, all subjects will reach a weight limit barrier in the course of the eigth (of twelve) categories from which they can choose items. If these 'unavoidable' failures are subtracted the other failed turns are indicators of real problems, e.g. they indicate errors or misunderstandings or other causes. The ratio of such failed turns to all turns (except those with failure by design, cf, above) serves as global measure for the relative 'faultiness' or success of the overall dialog.

The values for this ratio range within our corpus with N = 130 subjects from 9% till 73%.

4.2. Intentionality in UCI

Companion systems are designed for reacting individually to users, their actual emotional state and their situation. The post hoc interviews were conducted to examine if and when the user ascribes mental states to the simulated system and which ascriptions he made (Dennett, 1987). For companion-systems assumptions of positive intentions like helpfulness, trust-worthiness and empathy are desirable, assumptions of negative ones like malice, pursuit to dominance and poor willingness should be avoided. It can be assumed that the quality of the ascriptions and further issues of the subjective experience of the interaction account for the user's inner representation of the system and by that the relationship he or she develops towards the system.

4.3. User types

All subjects fill out as well a battery of well established psychometric questionnaires about various aspects especially of their personality. This will allow to correlate observed behavior and detected signs of affects and emotions with measured aspects of the personality of subjects and is expected to serve as a basis for defining a typology of users.

5. Summary and Discussion

We have presented the current state of the LAST MINUTE corpus. This corpus of recordings from naturalistic interactions between humans and a WOZ simulated companion system excels available corpora with respect to cohort size, volume and quality of data and comes with accompanying data from psychometric questionnaires and from post hoc in depth interviews with participants. The material is a cornerstone for work in the SFB TRR 62 but is as well available for research in affective computing in general.

We are intensively collaborating with other groups from SFB TRR 62 that work on detecting and analyzing emotional and affective cues in the recorded data from the LAST MINUTE corpus.

The long term goal of our joint work is to develop robust classifiers that allow to reliably infer the users' emotional state during the interaction with a companion system thus

allowing the companion to appropriately react and to proactively intervene.

6. Acknowledgment

7. Availability

The LAST MINUTE corpus is available for research purposes upon written request from the authors. For the reviewers a sample from the corpus with anonymised data is available from the following URL `http://iws.cs.uni-magdeburg.de/a3/lrec2012/index.htm` with loginname `reviewer` and password `lrec2012`.

8. References

P. Borkenau and F. Ostendorf. 2008. *NEO-Fünf-Faktoren-Inventar nach Costa & McCrae: 2. neu normierte und vollständig überarbeitete Auflage*. Hogrefe, Göttingen.

C. S. Carver and T. L. White. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The bis/bas scales. *Journal of Personality and Social Psychology*, 67:319–333.

R. Cowie and M. Schröder. 2005. Piecing together the emotion jigsaw. *Machine Learning for Multimodal Interaction*, pages 305–317.

D. C. Dennett. 1987. *The Intentional Stance*. The MIT Press, Cambridge.

E. Douglas-Cowie, R. Cowie, C. Cox, N. Amier, and D. K. J. Heylen. 2008. The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation. In L. Devillers, J-C. Martin, R. Cowie, E. Douglas-Cowie, and A. Batliner, editors, *LREC Workshop on Corpora for Research on Emotion and Affect, Marrakech, Marokko*, pages 1–4, Paris, France. ELRA.

G. Erdmann and W. Jahnke. 2008. *Stressverarbeitungsfragebogen. 4. überarb. u. erw. Auflage*. Göttingen: Hogrefe.

J. Frommer, M. Haase, J. Lange, D. Rösner, R. Friesen, and M. Otto. 2010. Project A3 'prevention of negative dialogue courses' Wizard of Oz experiment operator manual. SFB-Trr-62 working paper, unpublished.

M. Grimm, K. Kroschel, and S. Narayanan. 2008. The Vera am Mittag German audio-visual emotional speech database. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 865–868, April.

J. J. Gross and O. P. John. 2003. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *J Pers Soc Psychol*, 85:348–362.

M. Hassenzahl, M. Burmester, and F. Koller. 2001. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler and G. Szwillus, editors, *Mensch & Computer 2003*, pages 187–196. Stuttgart: B.G. Teubner.

H. Hoffmann, H. C. Traue, F. Bachmayr, and H. Kessler. 2010. Perceived realism of dynamic facial expressions of emotion – optimal durations for the presentation of emotional onsets and offsets. *Cognition and Emotion*.

M. Legát, M. Grůber, and P. Ircing. 2008. Wizard of oz data collection for the czech senior companion dialogue system. In *Fourth International Workshop on Human-Computer Conversation*, pages 1 – 4, University of Sheffield.

G. A. Lienert. 1961. *Testaufbau und Testanalyse*. Beltz, Weinheim.

G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic. 2010. The SEMAINE corpus of emotionally coloured character interactions. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 1079–1084, July.

P. Poppe, J. Stiensmeier-Pelster, and A. Pelster. 2005. *Attributionsstilfragebogen für Erwachsene (ASF-E)*. Göttingen: Hogrefe.

C. Rogers. 1959. A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In S. Koch, editor, *Psychology: A Study of a Science*, volume 3: Formulations of the Person and the Social Context. New York: McGraw Hill.

D. Rösner, R. Friesen, M. Otto, J. Lange, M. Haase, and J. Frommer. 2011. Intentionality in interacting with companion systems – an empirical approach. In J. Jacko, editor, *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, volume 6763 of *Lecture Notes in Computer Science*, pages 593–602. Springer Berlin / Heidelberg. 10.1007/978-3-642-21616-9_67.

T. Schmidt and W. Schütte. 2010. Folker: An annotation tool for efficient transcription of natural, multi-party interaction. In N. Calzolari (Conference Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

M. Selting. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, pages 353–402.

Website of the Transregional Collaborative Research Centre SFB/TRR 62. http://www.sfb-trr-62.de/.

N. Webb, D. Benyon, J. Bradley, P. Hansen, and O. Mival. 2010. Wizard of oz experiments for a companion dialogue system: Eliciting companionable conversation. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA).

Y. Wilks. 2010. *Close Engagements with Artificial Companions: Key Social, Psychological and Design issues*. John Benjamins, Amsterdam.