2017 26th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
Lisbon, Portugal, Aug 28 - Sept 1, 2017.

Reprompts as Error Handling Strategy in Human-Agent-Dialog? User Responses to a System's Display of Non-understanding*

C. Opfermann and K. Pitsch

Abstract— **In speech based technical systems, a 'reprompt' can be deployed as a verbally non-explicit and semantically unspecific practice of making a failure-to-understand transparent. Users' repeats or rephrasings of their previous answers might lead to further non-understandings, resulting in further reprompts by the system. On the basis of a Wizard-of-Oz video corpus in a schedule management setting with an embodied conversational agent and the special user groups of elderly and mildly cognitively impaired persons, we investigate in a conversation analytical approach the interactional impact of three-fold reprompts on subsequent user actions to an appointment suggestion. We focus especially on the *type of user actions during the course of multiple reprompts in a confirmation/disconfirmation context*. Analysis reveals more fine-grained user response types, testifying that all users ratify the first reprompt. After the second and third one, users tend to either *add* problem manifestations or initiations of the relevant next move. Or they *substitute* their previous answer by these types of actions. While additional or substituting problem manifestations call for more specific and linguistically restricting error handling practices, the user-initiated next moves are technically exploitable as implicit cues for confirmation in the presented special yes/no-context.**

I. Introduction

Collaboration in a task requires at its basic interactional level achieving and maintaining mutual understanding [1]. In human-machine collaboration, the arriving at mutual understanding calls for any technical system to display its internal state of processing results in a comprehensible way to the user. Technical systems which operate with spoken language interfaces – be it a robot, a virtual agent or a telephone-based service system – have to deploy apt dialog strategies to this end, especially in cases of non-understanding in order to initiate a further user response which might help to resolve the understanding problem.

For error recovery by an external display of a system's non-understanding, various different dialog strategies are deployed in speech based technical systems [e.g. 2, 3, 4]. A repeat of the previous system prompt or so-called 'reprompt' is a strategy that displays non-understanding on a *pragmatic* level compared to *verbally explicit* notifications like "Sorry. I didn't catch that": It sets the relevance of the prompt anew, thereby implicating that the user reaction was not understood. As it does not identify any specific trouble source, it is the user who has to decide if and how to produce his/her second answer differently. The longstanding research tradition of investigating error handling strategies and corresponding user reactions has yielded diverse findings (see e.g. [5] for disadvantageous linguistic adaptations like hyperarticulation and [6] for cyclic errors due to user rephrasings versus [7] with a successful 83% error recovery rate for reprompts). Repeats and rephrasings are the most documented user responses to a system's non-understanding, rephrasings always being more frequent. A previous research study with the virtual agent system "Billie" made also evident that multiple reprompts can lead to user rephrasings and their expansions[1]. In the face of such diverse results and in the need of functional error handling strategies in the agent system "Billie" with its specific schedule domain, we investigate anew into the functionality of system reprompts as two questions have been neglected in previous studies so far: *the types of rephrasings* which are initiated by a reprompt, and the impact of multiple reprompts on *users' conduct over the course of time during an error spiral*. Such qualitatively more fine-grained analyses provide technically useful insights for a context-sensitive deployment of dialog strategies in cases of a system's non-understanding in a yes/no-context.

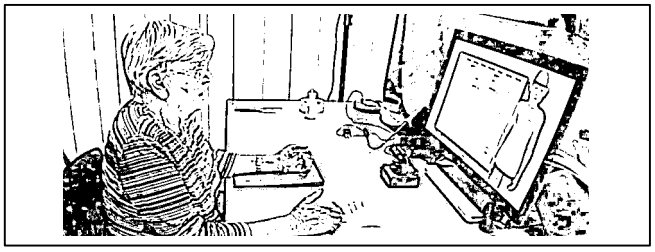

Figure 1. User negotiating an appointment suggestion with the virtual agent "Billie" in order to enter it into a virtual calendar

The presented work analyzes in a conversation analytical (CA) approach the forming of user response types to a system's reprompt and its multiple execution. The investigation was carried out in a semi-experimental Wizard-of-Oz study with the special user groups of senior (SEN) and mildly cognitively impaired (CIM) persons, and a group of students as controls (CTL). Within the interactional task of establishing an *appointment suggestion* (see section III), we specifically address the following questions:

(1) What are users' subsequent actions after each system reprompt?

(2) How do user actions change over time, especially in the face of lack of uptake of their actions by the system? How do users display if this halt to progressing with the task is troublesome?

(3) Can we find any specific differences between the three user groups (SEN, CIM, CTL)?

*The authors acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF) in the project 'KOMPASS' (FKZ 16SV7272). Christiane Opfermann and Karola Pitsch are with the University of Duisburg-Essen, Institute of Communication Studies, Universitätsstraße 12, 45141 Essen (phone: 0049-201-183-7000, christiane.opfermann@uni-due.de; karola.pitsch@uni-due.de).

[1] Documented in the final research report by Karola Pitsch et alii within the project VERSTANDEN (16SV6190), funded by the BMBF.

Analysis will show that users overwhelmingly stay cooperative to the multiple reprompts, mostly rephrasing their given answers. Users, however, tend to alternate *unpredictably* between reductions, expansions or variations of rephrasings. Following the second or third reprompt, users increasingly produce either additional actions like (i) *problem manifestations* or (ii) *initiations of the relevant next move* to their rephrased/repeated answer. Or they substitute their previous answer with these kinds of actions. This is implicative for (i) more specific and restricted error handling practices, and for (ii) implicit evidence of confirmation in this special context.

II. Error Handling Practices

A. Error Handling in Human-Machine Interaction (HMI)

Signaling non-understanding by a reprompt or other dialog strategy has already been thoroughly investigated with respect to impact on user responses during the last two decades [e.g. 5, 6, 7, 8, 9, 10], with much modern approaches promoting implicit error recovery by asking task related questions instead of notifying a non-understanding [4, partly 8], resulting in better user experiences [4] and higher error recovery rates [8]. However, we relate in more detail to the following studies with respect to their quantification of user responses after reprompts in order to compare later the results of our corpus analysis.

The studies of [7], [8], [9] and [10] all investigate the impact of a system's non-understanding on users' reactions on a lexico-syntactic level. Their corpora were obtained via autonomous system prototypes and in the domain of automatic booking lines. The corpus analyses of [7, 8, 10] coincide in the finding that a system's non-understanding leads to user rephrasings (ranging from 43% [8] to 48% [10]) and user repeats (18% [8] to 21% [10]) as most common response type. Among these studies, [7] and [8] picture clearly the interplay between specific linguistic prompt design and reactive user response types. Nevertheless, the question is left open of how users *exactly produce* and *form their responsive (repair) actions*, and among these their *rephrasings*. Our interest in this stems from the consideration if and when expansive rephrasings or out-of-domain user actions as reactions to (multiple) system reprompts might occur. A more differentiated qualitative analysis of users' subsequent actions might provide a slightly different outlook on reprompts in terms of their functionality to error handling or repair.

B. Repair in Human-Human Interaction (HHI)

Empirical research in the realm of human practices for achieving mutual understanding encompass various interactional practices specialized in displaying understanding or dealing with problems of speaking, hearing or understanding [11]. The latter are best known under the conversation analytical concept of 'repair' which entails the analytical distinction between the following constitutive components: (a) the trouble source, (b) the marking of non-understanding or understanding problems, i.e. the repair initiation, (c) the intent to resolve the understanding problem, i.e. the repair operation, and (d) a potential final ratification of the repair [11]. Concerning practices of initiating repair, an outstandingly informative study in this realm is the cross-linguistic investigation of formats of other-initiation of repair by [12]. It presents universal formats of repair initiation which are classified into two categories: (i) open class repair initiators (RI) and (ii) restricted RIs. These categories and related sub-categories differ in terms of their potential to locate the trouble source for the interlocutor (e.g. open class RIs like "huh" do not give any hint to the trouble source whereas open class RIs like the question word "who" identify a category of information that was not understood). They also differ in terms of marking the level of understanding problem (e.g. marking an acoustic problem versus a problem of identifying a certain reference).

Drawing on these findings, we decided to investigate system reprompts as repair initiation, as a reprompt establishes a clear display of non-understanding by the fact that it postpones the otherwise expected next step in interaction and renews the relevancy of the previously asked question. As it does not identify any specific trouble source (in contrast to a wh-question), it is apt for a technical spoken dialog system which usually cannot provide for this kind of specific information. On the other hand, this type of repair initiation leaves open any hypothesis to the interlocutor, or in our case the user, to identify what might be the problem. Thus, it also does not limit the types of responses. In conclusion, the question remains if this condition of a maximally "open" repair initiation can be functional in the setting of human-machine-interaction in the local context of an appointment suggestion and hence confirmation/disconfirmation context.

III. Study and Data

A. Virtual Agent System "Billie" in a WOz-Scenario

We investigated the interactional impact of multiple subsequent system repair initiations – in the format of a system reprompt – on user responses and their change in the course of time, with the virtual agent system "Billie" [13] in a Wizard-of-Oz (WOz) prototype version. The autonomous "Billie"-system is designed as an assistive system [14] for elderly persons and people with mild cognitive impairment who are experiencing a certain degree of reliance on assistance in everyday activities like schedule management, temporal orientation and maintenance of their social life.

To explore the functionality and aptness of certain basic turn designs and dialog structures under controlled conditions, the Wizard-of-Oz method was used [15]. The study comprised three specific schedule management tasks:

i. the collaborative *entering of appointments*

ii. *making a complete suggestion*, encompassing all relevant information items like "day-of-week", "start time", "end time" and "activity"

iii. *initiating a suggestion*, suggesting an "activity" and a special "day-of-week" which turns into entering the appointment collaboratively in case of acceptance (see section B, Fig. 2)

For each of the three tasks, a set of questions and formulaic beginnings of utterances were programed so that the wizard had just to type the user's answers into it for securing information transfer with the participant. Methodically, this kind of WOz-scenario permits to avoid unforeseen technical failure-to-understand incidents and to focus under controlled conditions on the questions if user responses match the assumed types of responses and courses of interaction. In addition, the deployment of the wizard's human understanding and interac-

tion competences provides the basis for analyzing incidents in which the human competences enable the achievement of mutual understanding and task resolution. This can reveal the (dis-)functionality of specific system moves and the need of further dialog and strategy requirements for a successful task resolution. Despite the human wizard's limited verbal resources, he/she had live access to the user's speech via headphones and to the user's nonverbal behavior via the webcam of the remote computer where the agent application was shown.

B. Dialog script and interactional resources for initiating a suggestion

The impact of multiple system reprompts on user response types was tested in the above-mentioned task (iii): *initiating a suggestion*. By suggestion, we draw on Couper-Kuhlen's definition of "an action type advocating a future action or activity to be carried out by the recipient that will benefit the recipient" [16, p. 633]. This applies to the agent's appointment suggestions as their enactment relies entirely on the user and is assumed to promote his/her leisure time activities.

The initiation of a suggestion consists of a fixed set of scripted utterances which were issued by the wizard in a predefined order, including the determined incident of prompting the same system suggestion four times in a row. Depending on the final decision of the user after the last reprompt to accept or resist the last suggestion, the wizard either requested the missing information items for finishing this entry, using the scripted questions and formulaic utterances of task (i) of entering appointments. Or he/she could confirm the user's resistance by a terminating phrase ("Then I won't enter it") or just move on with the next task.

Fig. 2 presents the scripted initiation of a suggestion with the structural and interactional functions possibly related with the system's prompts and with possible expectable responsive actions by users. All labels should be read as *potential* descriptions as their empirically data-based functional description relies on the response of a user in a real course of interaction.

In accordance to conversation analytic categories, the scripted dialog and possible user actions (Fig. 2) can be presented by three basic analytical dimensions:

(1) observable actions (e.g. verbal, nonverbal, medial activities),

(2) interactional function of the respective action, mostly on the basis of the deployed semiotic resources and shared conventions in a culture,

(3) structural function of the observed action in relation to preceding and subsequent actions.

Concerning the third dimension of *structural function,* CA-studies in HHI relate on the concept of action formation in so-called *pair sequences* which means e.g. that a suggestion sequence is usually enacted by two participants via two distinct but adjacent actions which relate type-adequately to each other. Thus, suggesting someone else to go for a walk constitutes the first pair part [FPP] of the pair sequence which makes relevant a responsive type-related action to that, or a so-called second pair part [SPP]. In this case, it can be either an acceptance as type-related second part, or a resistance. Accepting the suggestion would align with the FPP of a suggestion, a resistance would disalign with the initiated action. This type of

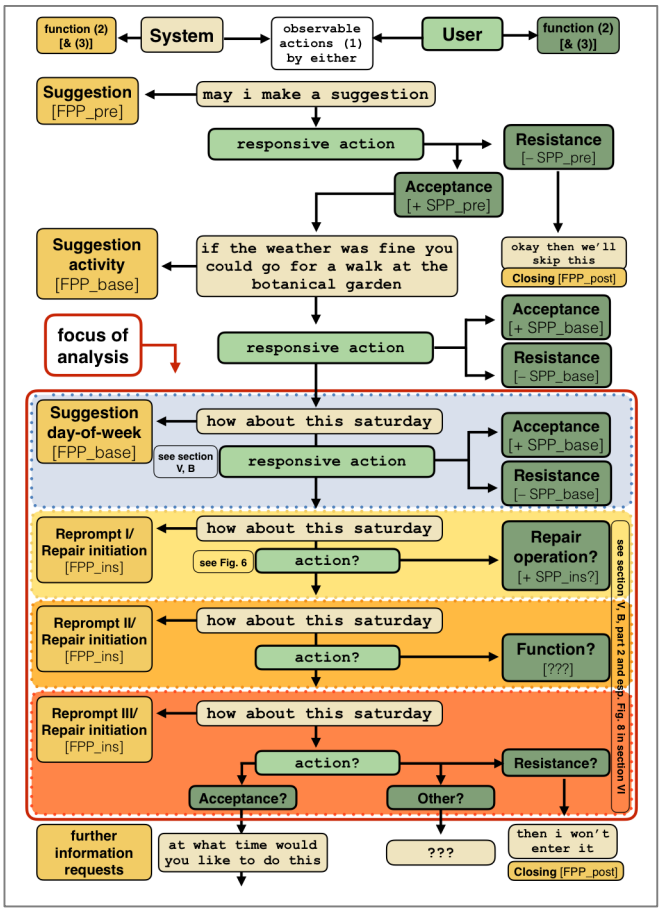

Figure 2. Scripted system utterances and possible user actions in a system initiated appointment suggestion; findings for coloured dotted outlines in section V and Fig. 6 to 11 in section VI.

relation is conventionally symbolized in CA as a (+)-type of SPP or (–)-type [17].

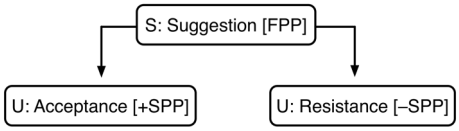

Figure 3. Schematic presentation of system's suggestion and user's possible response types in CA–categories

C. Setup and Procedure

The semi-experimental WOz-study with the agent system "Billie" entailed different schedule management tasks (see III, A) in which several induced incidents of interactional trouble had been scripted, one of them being the multiple reprompt described in section III, B. The interaction was recorded by three external cameras and a screen capturing for recording the monitor activities, too. A stationary eyetracker was installed beneath the monitor to record gaze conduct. Two wizards were employed and had been instructed to use the various predefined utterances according to their natural interaction competence. Participants were told that the virtual agent understands natural speech. They were advised to enter up to 10 appointments, either personal or fictitious ones. The latter had been prepared on cards lying on the table. Each participant dealt with a maximum of 13 schedule tasks, the tasks being presented in the same order to all participants.

D. Participants

53 subjects participated in the study as paid volunteers. They represented the two targeted special user groups and a control group. 18 elderly people (SEN) participated, aged between 67 and 92 years, and 19 people with mild cognitive impairments (CIM), which were all recruited via an institution specialized in providing various ranges of services individually adapted for citizens with various degrees of need of assistance. These two user groups are compared to a group of 16 students as controls (CTL). Due to the judicially and ethically vulnerable situation of the mildly cognitively impaired participants, the study was authorized by an independent ethics committee for research.

E. Corpus and Annotations

The video-corpus of interactions encompasses about 18 hours. The interactions were transcribed multimodally. Verbal actions are annotated according to the conversation analytic transcription system GAT2 [18], documenting also paraverbal features of speech. Nonverbal actions are transcribed, where relevant, according to conventions agreed-on by the research group at University of Duisburg-Essen.

Concerning the suggestion initiation task, 52 of 53 subjects joined the agent's suggestion throughout the whole task. One participant rejected the task by ignoring the pre-suggestion issued by the agent. In two interactions, the wizard did execute only two reprompts, so that these cases are not taken into account for corpus analysis. The resulting number is hence 50.

IV. Analytical Method

The conversation analytical method applied to this study consists of multimodal sequential micro-analysis of single cases [19], which is complemented by a basic quantification of the qualitative findings throughout the corpus. Starting from single case analysis for working out relevant categories of analysis, all cases are then coded and similar incidents are counted in order to depict the frequency and relevance of specific phenomena.

V. Findings I: User Response Types and Coding

Single case analyses throughout the corpus show that users' subsequent actions to the agent system's reprompts predominantly consist in the re-production of their first type-adequate action (acceptance or resistance) and hence second pair part [SPP] to the system's first pair part [FPP] of the suggestion "how about this saturday". Thus, users understand the reprompt as a repair initiation and carry out several successive repair operations, mostly by rephrasing their first acceptance or resistance (see VI, Fig. 6). This kind of persistency in action in face of a lack of progress in the system's state of understanding also demonstrates a continuous cooperative stance of users. However, several single case analyses show also user actions which mark that the persistent non-understanding constitutes itself a trouble source for the user. In some cases, this results in different additional or substituting actions to the previously produced acceptance or resistance. They are part of the data-based coding categories.

A. Case Analysis

The following case analysis of participant CIM-062 exemplifies the different types of user actions after multiple reprompts as repair initiators.

01		(0.7)	
02	Sys_ver	how about this SAturday;	
03		(2.4)	
04	**CIM-062**	**well THERE you give me a VEry good idea; (.)**	**pos.ass.I**
05		**that (.) is really VEry nice to do;**	**pos.ass.II**
06		(1.5)	
07	SYS_ver	how about this SAturday;	RP I
08		(0.4)	
09	**CIM-062**	**good iDEA;**	**reductive**
10		(1.7)	**rephrasing**
11	SYS_ver	how about this SAturday;	RP II
12		(1.1)	
13	**CIM-062**	**i already TOLD you;**	**P.M.**
14		**good iDEA- (.)**	**repeat**
15		**you may well ENter that;**	**R.N.M.**
1		(0.7)	
1	SYS_ver	how about this SAturday;	RP III
1		(0.3)	
		|*frowning*	**P.M.**
19	**CIM-062**	**|uh:- (-)**	**P.M.**
20		**hey: i have ALready told you-**	**P.M. (2x)**
21		**you CAN enter that;**	**R.N.M.**
22		(2.2)	
23	SYS_ver	oKAY,	

Figure 4. Persistent user acceptance with different accompanying or substituting actions (pos. ass.–positive assessment; RP–reprompt; P.M.–problem manifestation; R.N.M.–relevant next move)

After the system's first pair part of the suggestion of a day of the week (line 02), participant CIM-062 produces as type-adequate second pair part an **acceptance** which takes the form of a very extensive twofold **positive assessment** of the suggestion (lines 04 to 05). After the first reprompt in line 07, CIM-062 reproduces the acceptance again, but **rephrases** the previously verbose action to the two-word elliptical phrase "`good iDEA;`" (line 09) which constitutes a **reductive rephrasing**. After the 2nd reprompt, CIM-062 begins the turn with an **explicit and meta-communicative display** of the fact that the twofold lack of system-understanding is now **troublesome** [20]: "`i already TOLD you;`" (line 13). By that, CIM-062 claims the lack of uptake of the two given answers. After a **repeat** of the acceptance in form of the same elliptical phrase "`good iDEA;`" (line 14), CIM-062 initiates **the relevant next move** (R.N.M.) in settling and terminating the suggestion-acceptance sequence and produces a request to the agent to enter the appointment ("`you may well ENter that;`" line 15). The 3rd reprompt leads to even more trouble-oriented actions: CIM-062 produces an elongated hesitation marker at the beginning of the new turn, accompanied by a non-verbal display of confusion or maybe annoyance (see "`uh:-`" plus simultaneous frowning in line 19), which both establish another manifestation of finding the lack of the system's understanding troublesome. After these vocal and nonverbal **problem manifestations** (P.M.), CIM-062 rephrases the meta-communicative problem manifestation ("`i have ALready told you-`", line 20), intensified by a preceding summons of the agent ("`hey:`", line 20). CIM-062 skips the reproduction of an explicit acceptance and instead rephrases the request of appointment entry ("`you CAN enter that;`" line 21). In this vein, CIM-062 also demonstrates her ongoing willingness towards accepting the suggested appointment. Such analyses resulted in the following coding categories.

B. *Coding Categories for User Actions*

Types of user actions following system suggestion (see blue dotted outline in Fig. 2):

1. Types of adequate SPPs: Case analyses have shown, that type related SPPs of acceptance and resistance in the suggestion context had to be sub-classified into two more specific categories: an acceptance [+SPP] was regularly produced in the form of a **positive assessment** ("That's a good idea") and/or a **confirmation** ("yes/okay/I'll do that"), and a resistance [– SPP] was sometimes produced in form of an **account**, i.e. telling a reason for not being able to accept ("Saturday, I go to my cousins' "), and/or a **disconfirmation** ("No, Saturday is not so convenient for me"). Thus, Fig. 3 has to be adapted as follows:

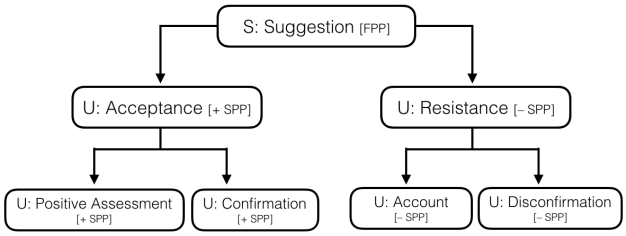

Figure 5. Sub-classes of user acceptance/resistance to a system's suggestion in CA–categories

2. User-initiated relevant next moves: (i) self-initiated information delivery ("From 10 o'clock to 11 o'clock"); (ii) self-initiated request of information ("At what time?"); (iii) entry allowance ("You may well enter that").

3. Other actions: Any other than the before mentioned actions.

Types of user actions following 1st to 3rd system reprompt (see yellow, amber and red dotted outlines in Fig. 2):

1. Rephrasings of type-related SPP: We take a rephrasing as a **rewording of the same pragmatic action** in a different lexico-syntactic manner. Rephrasings can be subdivided into:

(i) Reduction: rewording in a shorter utterance ("That's a good idea" —> "yes")

(ii) Expansion: rewording resulting in longer utterances ("no, nothing" —> "Well, there is nothing")

(iii) Variation: same extension of utterance in a different lexico-syntactical manner ("okay" —> "yes").

2. Repeats of type-related SPP.

3. Problem manifestations (P.M.): verbal ("I've told you three times"), nonverbal (rolling with eyes, drumming with fingers on the table), vocal (emotional interjections like "oh"/"ah"), laughter.

4. User-initiated relevant next moves (R.N.M.): (i) self-initiated information delivery ("From 10 o'clock to 11 o' clock"); (ii) self-initiated request of information ("At what time?"); (iii) entry allowance ("You may well enter that").

5. Other actions: Any other than the before mentioned actions.

The coding of each single case according to these categories led to a basic quantification of the various types of user actions following the system's successive reprompts. This allows for comparing our findings with existing literature and a discussion of the functionality of reprompts as error handling practice and its technical implications.

VI. Findings II: Response Types Quantification

A. *User Actions after First Reprompt*

Our corpus analysis of system reprompts in suggestion sequences confirms the findings of [7, 9, 10] who investigated user conduct after a system's non-understanding (see II, A): After the first reprompt of the suggestion "`how about this SAturday;`", the user groups react in 76% of all 50 cases with rephrasings and repeats of their previously produced acceptance or resistance to the suggestion.

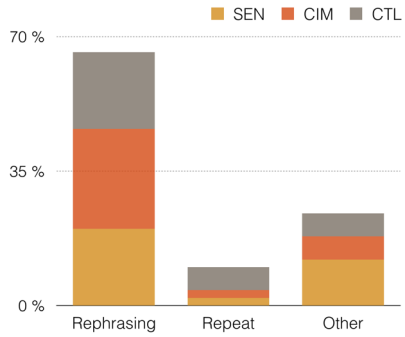

Figure 6. User response types following 1st system reprompt in functional relation to previously produced reponse (yellow outline in Fig. 2); categories P.M. and R.N.M not shown for simplification (but see Fig. 8)

However, contrary to the mentioned studies in the domain of booking lines [7, 8, 10], our user rephrasings do not double user repeats (see II, A; e.g. 43% rephrasings versus about 20% repeats in [8]), but equals almost to the septuple. This finding calls for a closer look at the interactional motivations for the *much higher amount of rephrasings in our corpus*. The qualitative findings show that this result is due to the special task of settling an appointment suggestion and its related user response types like accepting and resisting (see section V and Fig. 5). In contrast to system requests in booking activities that primarily consist of yes-no-questions, wh-questions or directives, which all orient to *facts and dates,* the suggestion-sequences *orient to the users' willingness, disposition and ability* [16, 21] towards the suggested activity and day for acting it out. This leads to user response types like positive assessments in the function of an acceptance, or accounts and trouble tellings in the function of a resistance to the suggestion (see section V for examples of user responses). In our corpus, 28% of user rephrasings occur due to a change from one action sub-category to the other within the same kind of action production (e.g. user first produced an account as a response, i.e. a type-adequate SPP and then an explicit disconfirmation as SPP after the 1st reprompt).

In contrast to the referred studies, our primarily qualitative micro-analytical approach also reveals another differing aspect: *the type of rephrasings* after the 1st system reprompt. In 58% of all rephrasings, users produce reductive rewordings, 30% produce expansive rewordings and 12 % vary their rephrasings (for a comparison between user groups see Fig. 7). This partial result suggests a positive outlook on a functional

deployment of reprompts, as the first reprompt mostly leads to shorter and more explicit user responses which might be easier detectable for spoken language interfaces in the context of a suggestion and its subsequent confirmation or disconfirmation. The results of the next section, however, draw a less favorable outcome for reprompts in yes/no-contexts.

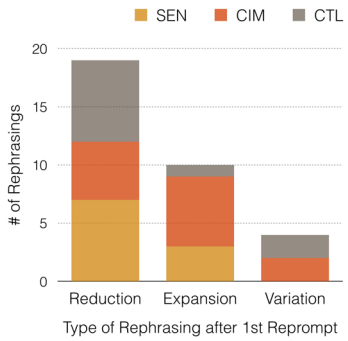

Figure 7. Type of user rephrasings following 1st reprompt

B. *User Actions after Multiple Reprompts*

As shown in the case analysis above, users may reproduce their first action to the system's suggestion and produce additionally different actions like **problem manifestations** (P.M.) or entry allowances after the 2nd and 3rd reprompt (see Fig. 4, lines 13–15 and 19–21), the latter establishing a user-initiated **relevant next move** (R.M.N) in settling the appointment. Other forms of relevant next moves are either an information request of a time for going for a walk ("at what time", CTL-002) or the delivery of the respective time for doing the suggested activity (e.g. "at three o'clock p.m.", SEN-043). In some cases, these types of user actions are also produced instead of a type-adequate SPP to the suggestion. So, there is an increase of additional or substituting actions of that kind to the main repair operations like rephrasings and repeats over the course of multiple system reprompts (see Fig. 8). Despite an increase in problem manifestations, there was only one user in the whole corpus (2%) who abandoned the task due to frustration.

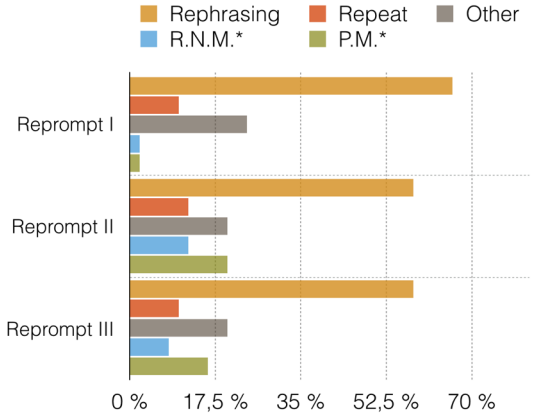

Figure 8. Occurrences of user action types after multiple successive system reprompts (yellow, amber and red outline in Fig. 2); P.M.–problem manifestation; R.N.M–relevant next move; the asterisk symbolizes that these actions are only reactions or additional ones to rephrasings or repeats

A further picturing of rephrasing types after multiple reprompts shows that whereas reductive rephrasings constitute the majority of rewordings after the 1st reprompt, users tend to expand more frequently their rewordings after the 2nd one and tend to vary rephrasings more after the 3rd – which is not necessarily favorable for user speech understanding if the preceding utterance was a longer one (see Fig. 9).

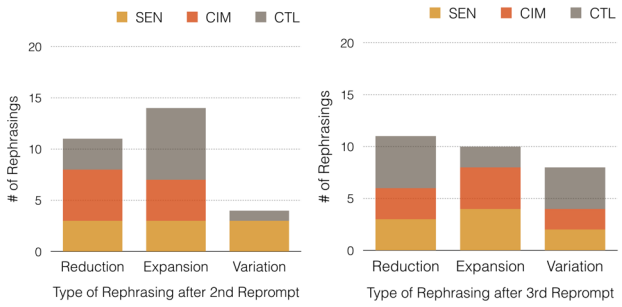

Figure 9. Type of user rephrasings following 2nd (left) and 3rd (right) reprompt

A comparison between the three user groups results in the finding that CTLs produce in general more reductive rephrasings, and especially more reductive rephrasings in form of a one-word utterance like a simple "yes" or "no" that stands alone in the turn. These *one-word reductive rephrasings* occur in 31% of CTL-rephrasings after 1st and 3rd reprompt (19% after 2nd one). SENs show this conduct only in 12% of rephrasings after 1st, 18% after 2nd and 6% after 3rd reprompt. Within the CIM-group, there is a constant decline: 18% after 1st and 2nd reprompt, and 6% after 3rd reprompt (see Fig. 10).

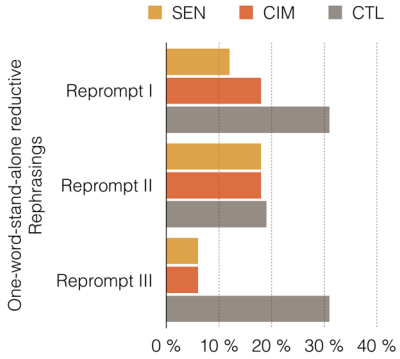

Figure 10. One-word-stand-alone **reductive rephrasings** after reprompts

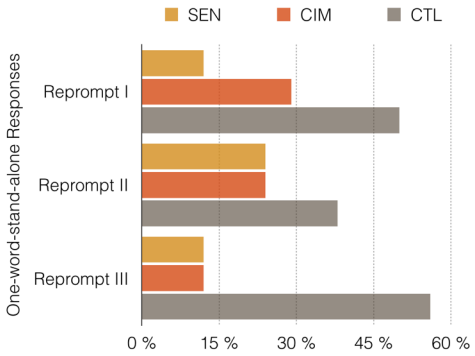

Figure 11. One-word-stand-alone **reponses** after reprompts

Fig. 11 shows the occurrence of one-word-stand-alone user responses in general, irrespectively of the action type it is related with. The user group specific difference becomes even clearer: CTLs produce by far more one-word turns than SENs and CIMs, with CIMs showing a decline with each issued reprompt. Single case analyses show that CIMs tend to produce more problem manifestations after the 3rd reprompt.

VII. Summary and Implications

In the presented study, we investigated the interactional impact on user actions when a technical spoken dialog system displays a persistent state of non-understanding by successive reprompts. The reported analyses show that users' subsequent action types to the reprompts may become too complex for a technical system with progressing number of reprompts. After the first reprompt, however, users' reactions consist mainly in reductive and linguistically more explicit rephrasings which is favorable in a context of confirmation and disconfirmation like that of the presented suggestion format as it improves user intention interpretation. After the second reprompt users tend to produce more verbally explicit or nonverbal manifestations of finding the persistent lack of system understanding troublesome. Moreover, some of the participants who confirmed the system's suggestion initiate a relevant next move in order to progress with the suggestion, thereby jumping to the next sequence. Together, these practices result (i) in unfavorably longer turns and/or (ii) in user actions that are out-of-domain utterances and thereby not easily to handle for language understanding modules.

Concerning the deployment of reprompts as error handling strategey in a technical system, we conclude therefore two practical applications of the research findings:

(1) It is obviously functional to have reprompts be issued one time, but not multiple times. In the context of suggestion sequences, the system utterance makes relevant a confirmation/disconfirmation-type user response which entails the usage of affirmative or negative communicative means. If the system stays in a state of non-understanding after a first reprompt and user reaction, a clarification request could be initiated. It could take the form of a yes/no-question plus a "you can say"- move [8] in order to prime the user for explicit yes/no-answer: "Do you mean yes? Please say yes or no."

(2) The user practice of delivering the time for acting out the suggested activity at the day they have already accepted in their own previous response could serve as cue for user intention interpretation. A flexible dialog system like [13] can use the information delivery of a start or end time in an appointment suggestion sequence as evidence for a produced acceptance and move on with settling the appointment.

Our findings of user repair practices and other actions to a system's persistent non-understanding are context-bound to the special task of appointment suggestions. This task is highly socially coined due to an inherent orientation to the willingness and ability of the recipient to comply with the suggestion [16, 21], therefore resulting in very verbose user actions which might be difficult to handle for spoken dialog systems [22]. With respect to the functionality of reprompts in HMI, the reported findings reaffirm the need of otherwise already applied approaches of more differentiated error handling [2, 8].

A next step in investigating into the impact of reprompts is a further research study with the autonomous agent system "Billie", contrasting reprompts with non-understanding notifications as display of non-understanding in the context of the facticity oriented task of entering appointments. This might reveal other interactional aspects of the special error handling practice of reprompts and subsequent user actions.

References

[1] H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[2] D. Griol, and J.M. Molina, "A framework for improving error detection and correction in spoken dialog systems", in *Soft Computing*, (20,11), 2016, pp. 4229–4241.

[3] D. Bohus, and A. Rucknicky, "The RavenClaw dialog management framework: architecture and systems", in *Computer Speech and Language* 23, 2009, pp. 332–361.

[4] G. Skantze. "Exploring human error recovery strategies: Implications for spoken dialogue systems", in *Speech Communication* (45,3), 2005, pp. 325–341.

[5] S. Oviatt, M. Mac Eachern, and G. Levow, "Predicting hyperarticulate speech during human-computer error resolution," in *Speech Communication* (24,2), 1998, pp. 87–110.

[6] D. Litman, J. Hirschberg, and M. Swerts, "Characterizing and Predicting Corrections in Spoken Dialogue Systems," in *Computational Linguistics* (32,3), 2006, pp. 417–438.

[7] J. Shin, S. Naryanan, L. Gerber, A. Kazemzadeh, and D. Byrd, "Analysis of User Behavior under Error Conditions in Spoken Dialogs," in *Proceedings of ICSLP* 2002.

[8] D. Bohus, and A. Rudnicky, "Sorry, I Didn't Catch That! An Investigation of Non-understanding Errors and Recovery Strategies," in *SIGdial* 6, 2005, pp. 128–143.

[9] A. Raux, B Langner, D. Bohus, A. Black, and M. Eskanzi, "Let's got public! Taking a Spoken Dialog System to the Real World," in *INTERSPEECH* 2005, pp. 885–888.

[10] S. Choularton, and R. Dale, "User Responses to Speech Recognition Errors: Consistency of Behaviour Across Domains," in *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 2004, pp. 457–462.

[11] J. Sidnell, *Conversation Analysis*. Chichester: Wiley-Blackwell, 2010.

[12] M. Dingemanse, J. Blythe, and T. Dirksmeyer, "Formats for other-initiation of repair across languages," in *Studies in Language* (38,1), 2014, pp. 5–43.

[13] R. Yaghouzadeh, H. Buschmeier, and S. Kopp, "Socially Cooperative Behavior for Artificial Companions for Elderly and Cognitively Impaired People," in *Proceedings ISCT 2015*, pp. 15–19.

[14] R. Yaghouzadeh, and S. Kopp, "flexdiam - Flexible Dialogue Management for Incremental Interaction with Virtual Agents," in *IVA 2016*, pp. 505–508.

[15] L. Riek, "Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines", in *Journal of Human-Robot Interaction* (1,1), 2012, pp. 119–136.

[16] E. Couper-Kuhlen, "What does grammar tell us about action?" in *Pragmatics* (24,3), 2014, pp. 623–647.

[17] E. Schegloff, *Sequence organization in interaction. A primer in conversation analysis I*. Cambridge, UK, New York: Cambridge University Press, 2007.

[18] M. Selting et alii, "Gesprächsanalytisches Transkriptionssystem," in *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 2009 (10), pp. 353–402.

[19] J. Sidnell, and T. Stivers (Eds.), *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, 2013.

[20] M. Selting, *Verständigungsprobleme*. Mannheim: Verlag für Gesprächsforschung, 2013.

[21] H. Houtkoop, *Establishing Agreement. An analysis of proposal-acceptance sequences*. Dordrecht-Holland: Foris Publications, 1987.

[22] C. Opfermann, K. Pitsch, R. Yaghoubzadeh, and S. Kopp, "The Communicative Activity of 'Making Suggestions' as an Interactional Process: Towards a Dialog Model for HAI", accepted for *HAI'17*.