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

Dealing with 'Long Turns' Produced by Users of an Assistive System: How Missing Uptake and Recipiency Lead to Turn Increments

K. Cyra, and K. Pitsch*

Abstract— **Based on a user study, we start from the observation that 'long turns' uttered by users towards an assistive system constitute a challenge for the dialog management of a voice-operated system. Assuming an interactional perspective, we address the question as to how 'long turns' emerge in interaction. We suggest to conceive of these utterances as being co-constructed by both, the user and the multimodal conduct of the technical system. In this paper, we examine how such 'long turns' emerge step by step in terms of an initial utterance being expanded by so-called 'increments' as well as their specific structure. Analysis shows that such utterance expansions (causing 'long turns') react to the user facing problems with a lack of uptake resp. display of recipiency by the technical system. Combining qualitative micro-analysis with quantification, we discuss specific interactional contexts of turn increments, different actions performed by them and the role of uptake resources in the light of designing autonomous speech-based systems.**

I. Introduction

Designing the interactional conduct for robotic (and other technical) systems so that they could be used intuitively by humans, a central issue resides in the discrepancy between 'plans and situated actions' [1]: From the human perspective interaction does not follow a 'plan', but develops step-by-step in a more or less unpredictable, i.e. contingent, manner [1]. From the machine's perspective contingency poses a problem, as humans might produce utterances that cannot be processed easily because they are too long, unstructured or contain 'off-topic' information [2, 3]. Assuming an interactional approach, some HRI studies suggest to control these contingencies by equipping the robotic system with strategies to influence the user's conduct [4, 5] or to interrupt user speech verbally [6, 2] and multimodally [7]. In this paper (following ideas from [8, 9, 10]), we suggest to assume a complementary: we attempt to gain insights into the ways in which the users' utterances emerge step by step. Thus, 'long utterances' produced by the user and the contingent nature of human social interaction is not considered as being problematic per se, but as an interactional phenomenon which is co-constructed by all participants [9, 11, 12]. Hence, turns are not viewed as being 'long' from the outset, but as emerging successively in a way that step by step a next small unit (i.e. an 'increment') is added to the initial turn [13, 14]. The understanding of the dynamic nature of *turn increments* [13] as it is suggested in the field of interactional linguistics implies promising approaches to understand the structure and emergence of utterances and so, for technical systems to deal with interactional problems.

The research presented here starts from observations of users attempting to enter calendar information with an embodied conversational agent (ECA (HAI) [8]. The study shows how an autonomous assistive system contributes to the production of expanded utterances by the user and in the end overwrites information already ratified by the user. The previous study and the study presented here, are located in an assistive context focusing on users needing support in schedule management and orientation in time. The overall objective is to gain a better understanding of the interactional practices of both human user and technical system which contribute to the unintended production of long turns by the user, to inform system design. In order to understand why interaction ends up with problems caused by 'long turns' we reconstruct the interactional context of their production and take a close look on factors relevant for the emergence of turns. We address the following research questions by a qualitative analysis: (1) How do turns emerge step by step by successively adding 'increments' in a task-related interaction with a technical system and which practices contribute to the unintended production of long turns? (2) Which actions do participants perform by producing turn increments? Research questions addressed by a quantitative analysis are: (3) Are there specific contexts for the occurrence of turn increments? (4) Which of the system's multimodal resources are relevant for users while verbalizing appointments?

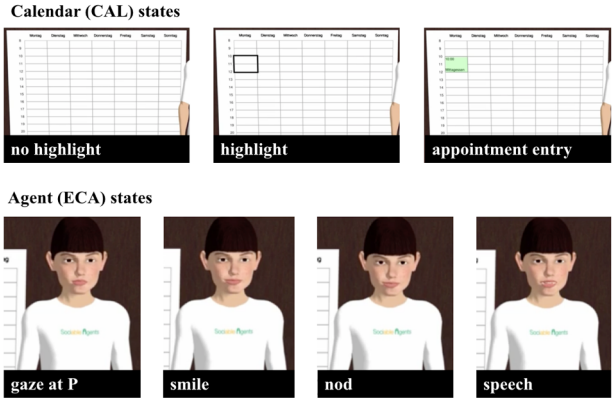

Figure 1. Overview of the system's uptake resources

* Research supported by the German Federal Ministry of Education and Research (BMBF) in the project 'KOMPASS'. K. Cyra and K. Pitsch are with the Communication Studies Department at the University of Duisburg-Essen, Germany (e-mail: katharina.cyra@uni-due.de, karola.pitsch@uni-due.de)

II. Emergence and Management of Turns

A. HRI: Long turns as challenges for the system

Research shows that human strategies of handling interactional problems contradict the system's requirements regarding user input: Pitsch et al. [8] show how an autonomous technical system in the field of scheduling [15] adds to the production of a user's turn expansion. The system displayed calendar highlights as soon as the automatic speech recognition (ASR) had confident hypotheses on day or time, but the system's repeated rephrasing, omitted information and verbal pauses were handled as interactional problems by users: They expanded turns referring to the appointment topic as this was not displayed and the system repeatedly requested information about it. Eventually, the system overwrote ratified information and entered wrong topics based on the expanded turn.

Although there are approaches for incremental speech recognition and prosodic analysis [16], most systems analyze user utterances from a numeric view: problematic utterances are either too long (more than 4 sec. [2] consist of too many words [17]) or contain so-called off-task information [18]. HRI research only roughly describes human turns either as more structured when short, or less structured when long [3], not helping to understand turn emergence. Whereas most research focuses on speech as the principle dimension to analyze long turns, other approaches [3] consider multimodal resources (speech, users' motions) to describe the structure of human actions and show that a robot's non-understanding adds to the production of more and longer user utterances. Besides HRI, HAI research specifically backs to investigating long user turns [8, 7] by taking a multimodal perspective on managing human speech production and acknowledging the system's part in the emergence of turns [7].

B. HHI: Turn increments to handle interactional trouble

In interactional linguistics interaction is understood as a phenomenon that emerges in situated, local contexts [11], meaning that interaction is not detachable from its context, and that interlocutors naturally orient their actions towards co-participants and context. Building on that, we use the term *turn increments* [13] to describe the phenomenon of expanded or long turns. Research from HHI [13] indicates that turn increments emerge when speakers notice problems of recipiency (e.g. an inattentive interlocutor) and uptake [14]: by incrementally adding more to an initial turn participants provide added options for uptake to their interlocutors. These approaches emphasize the role of displaying recipiency and uptake as crucial for co-participants in interaction [11, 14]. Following HHI research [13] turn increments can be subdivided into extensions, that verbally continue an initial action, and so-called free Constituents that assess prior actions or missing uptake. Summarizing, humans produce initial turns and expand them incrementally, when interactional trouble arises [16].

As stated above, uptake is essential for interaction. In HHI uptake is realized by multimodal resources (gaze, speech, head movement, gestures etc. [11, 14]); if no uptake via at least one of possible resources is perceived, interactional trouble can be assumed. But even when immediate uptake is missing and discontinuities in talk occur [19], non-verbal actions are still performed and contribute to interaction [20]. Respecting Conversation Analysis' (CA) understanding of silence in talk (gap, lapse, pause [18]), we consider a multimodal understanding in which non-verbal behavior also shows participants' actions. We and treat silence in talk where uptake from an interlocutor might be expected, as *turn vacant pauses* [20].

III. Study Design And Interaction Structure

A voice operated assistive system was developed to enter appointments into a virtual calendar (CAL) with help of an ECA (Fig. 1) [21, 15]. The system was operated by a human wizard (WOZ) and included different tasks (Fig. 2, (i)) mainly consisting of appointment entries.

A. Assumed prototypical interaction

The interaction structure is based on the autonomous system (Fig. 2 (i)) with two appointment types: User input-based appointment entries (AE) and appointment proposals (AP). The focus will be on AE. Generally, appointments are defined by parameters for [DAY], [TIME {start}, {end}], [DURATION], and [TOPIC]. To enter appointments, at least information on [DAY], [TIME {start}] and [TOPIC] are required. A 2-hour-duration is set by default. AE states are based on a 3-phase interactional model (Fig. 2, (ii)) with Global Information Phase (GIP) to initiate appointment entry, Local Information Phase (LIP) to request specific parameters, and Entry Phase (EP) to enter the appointment into CAL. GIP and LIP can be initiated by system or user; EP is initiated by the system/WOZ only. To enter appointments, the system/WOZ follows an assumed prototypical interaction model. GIP and LIP are designed as stepwise procedures (Fig. 2, (iii), (iv)) with an optional system request, mandatory naming of appointment parameters by user, mandatory system verbal rephrase of parameters and highlight in calendar, and an optional user ratification. Vital phases to enter an appointment are GIP and EP, whereas LIP is needed, when necessary parameters are missing or need further negotiation.

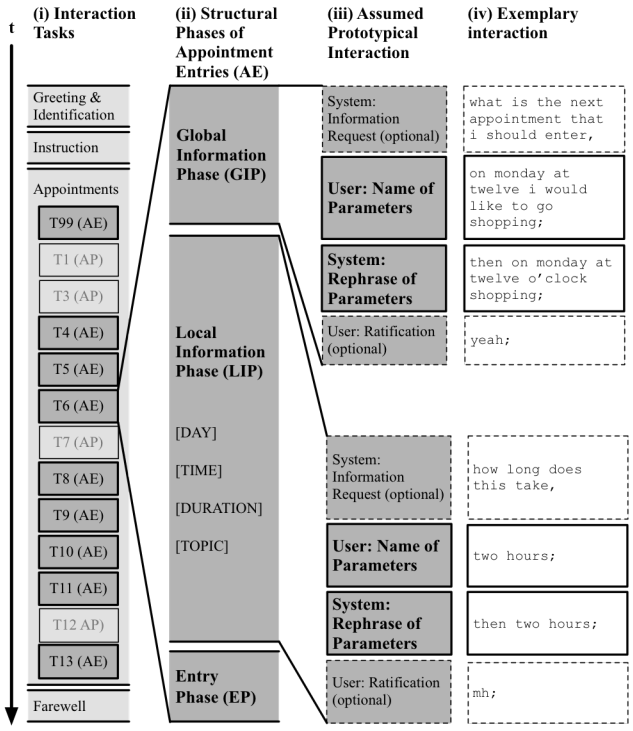

Figure 2. Overview of interaction structure

B. WOZ system resources for uptake

Basic uptake resources are CAL and ECA (Fig. 1) with CAL as the system's multimedia resource and three states of appointment display: no highlight, highlight of time slot and appointment entry. ECA has multimodal resources for uptake and display of recipiency: speech, gaze, head movement (nod) and facial expression (smile). To explore interactional resources for uptake the WOZ system was flexible regarding timing and verbal uptake options. Visual and verbal uptake were linked (e.g. highlight + rephrasing). The WOZ GUI contained verbal resources (Fig. 2 (iii)) to manage interaction via shortcuts. To avoid misinterpretations of the ASR, the WOZ entered parameters for [DAY], [TIME] and [DURATION] manually via shortcuts to avoid delays. Pre-defined [TOPIC] parameters from appointment cards, which were handed to participants, were offered as shortcuts. Users' individual [TOPIC] parameters were entered without shortcuts. To better react to user behavior the WOZ GUI contained additional verbal resources for appointment entry and uptake (e.g. paraphrase of initial request). Also, less specific resources were added to all interaction states (via shortcut e.g. "Yes", "One moment please" or click e.g. "Could you repeat this please?", "I did not quite understand you.") and a free-text field. The WOZ could not initiate CAL uptake independently from verbal resources. To explore the emergence of turn increments in contrast to the autonomous system [8], the wizard had to decide which parameters to pick up from the user's verbal input, when to rephrase them and so to visualize them in CAL. The WOZ system had no restrictions regarding overlap avoidance.

IV. Study and Data

We conducted a semi-experimental explorative Wizard-of-Oz (WOZ) user study with the ECA BILLIE [15, 21] to examine how users interact with the agent when entering appointments. The study was approved by an independent ethics committee. The WOZ system was set up on a smart TV device in a laboratory setting. The WOZ room was next door unnoticed by the participants. Participants could enter a maximum of 13 appointments. Six appointment states contained induced errors comparable to autonomous system behavior.

A. Participant and wizard instructions

Participants were asked to jointly enter appointments into the virtual calendar with the ECA using natural language (no keywords). The participants could choose to name own appointments or use appointment cards provided by investigators. After investigators left the ECA introduced the task of collaborative appointment entry (`"When you tell me a new appointment, the calendar will display day and time."`), referring to the calendar (gaze, speech). Participants were informed about the nature of the study (without information about the WOZ), that they will be recorded (audiovisual and eye tracker data) and that they could interrupt the study at any given moment. All participants gave written informed consent (where necessary, also by legal representatives).

To obtain an interaction as natural as possible, wizards were instructed to spontaneously decide on verbal resources provided by the WOZ GUI. The WOZ had to decide if, when and how to request information about appointments and to process user inputs as understood. He/she also had to decide when appointment parameters were complete to initiate the EP. To avoid delays WOZs were instructed to use the provided short cuts whenever possible. The WOZ could make use of all interactional resources at any time, had access to user's speech (conveyed via headphones), and could see a frontal shot of the user and user's gaze direction as detected by the eye tracker on a separate monitor next to the WOZ computer.

B. Participants

53 participants out of two user groups – seniors (SEN *N*=18) and people with cognitive impairments (CIM *N*=19) – and a student control group (CTL *N*=16) were recruited. The analysis will focus on data from the senior group, who were aged between 67 and 92, had diverse social and professional backgrounds, and lived in various residential environments (from own homes to outpatient assisted living) and had different experiences and knowledge about technology (some owned novel consumer electronics like smartphones/tablets). Two wizards with explicit knowledge about HRI/HAI were trained in advance to use the WOZ GUI.

C. Data

The study was conducted in a 3-week-period. The data corpus comprises 53 interactions recorded by 3 HD video cameras, audio data, eye tracker data, log files of the dialogue states, and a screen capture of the system interface. Changes of the WOZ system between interaction states, updates of agent and calendar resources, and segments for user speech (based on voice detection) were transferred to Elan Annotation Software [22]. The analyses are based on one exemplary case study and *N*=5 participants from the SEN participant group.

V. Analytical Method

The two-step analytical approach is based on Conversation Analysis (CA) [23] and quantification [4, 5]. First, a fine-grained micro-analysis of video-based interactional data of an exemplary case is conducted, to understand the interactional structures and procedures that become relevant for co-participants in interaction. By doing so, we reconstruct the participant's view and understanding of the situation and the system's actions. This is the groundwork for the development of empirically based analytic categories which are applied onto the sub corpus of turn increments. This basic quantification helps to validate categories and to show their relevance and frequency.

VI. Qualitative Analysis: Co-constructing a Turn

Micro-analysis reveals procedures of interlocutors, that lead to turn increments, and helps to understand how they emerge in the situated context. By inspecting the user's orientation towards (missing) system uptake, and inconsistent display of recipiency, we reconstruct interactional resources relevant for the production of turns and point out structural and functional elements of the turn in progress.

The extract WOZ1-SEN-042 / T99 (Fig. 3) shows the interaction in the GIP and EP of state T99 (Fig. 2 (i)). After ECAs task instruction interaction is initiated by participant (P) with a parameter for [DAY']: `"from: on MONday; (.)"` ending on a falling intonation and followed by a micro-pause while P's gaze is directed at CAL (see red area in Fig. 3). As no uptake by ECA or CAL is noticeable – ECA gazes at P and CAL shows no change – P continues with a first turn increment: `"from ten to twelve- (.)"` After a turn vacant pause

of 1.188 sec, in which P breathes in and monitors CAL, P continues with a second turn increment: "er: yes then i prepare myself for LUNcheon;". Having stated all relevant parameters to enter an appointment, P's turn comes to a possible turn completion: intonation is falling and the turn is syntactically complete. The inspection of P's gaze shows that P constantly orientates towards CAL in this phase. This first part illustrates how P's turn emerges: Each turn increment contains relevant information for the appointment, and each is followed by a pause: P orients towards pauses as transition relevance places on the auditory level. By adding further parameters to [DAY], P incrementally produces extensions of the initial turn, and provides renewed options for uptake yet they remain unanswered. To mask the issue of lack of uptake, P specifies the appointment, i.e. performs a *continued action*. In this section P's orientation towards CAL shows its relevance as a visual resource where uptake is expected.

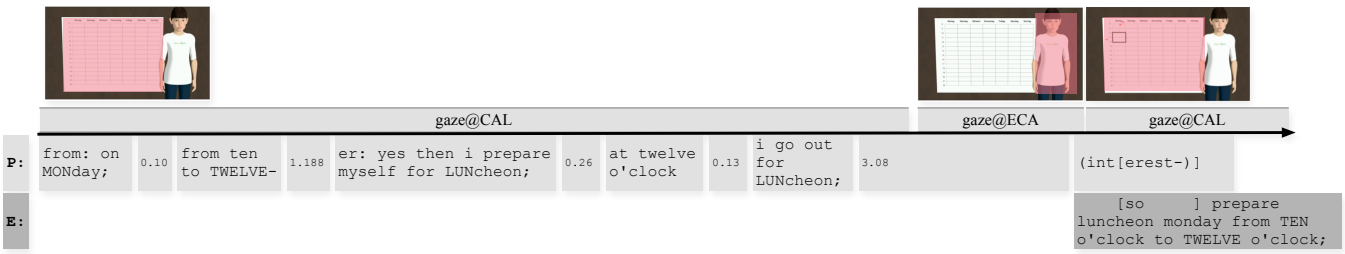

Figure 3. Exemplary turn increment in temporal development: Extract of WOZ1-SEN-042 / T99

The next part shows the emergence of a parallel action: After a pause of 0.26 sec, with P gazing at CAL without noticeable uptake (ECA gazes at P, CAL shows no change), P continues the turn by producing increments: Another parameter for [TIME {start}"]: "at twelve o'clock" and, after a pause (0.13 sec) a parameter for [TOPIC"]: "i go out for LUNcheon;" This turn increment ends on a falling intonation indicating another possible turn completion. Although we find similarities with *continued action* (turn increments divided by a pause providing renewed options for uptake) a close look at the parameters shows that P initiates a consecutive competitive appointment, a *parallel action*. For WOZ, there is competitive information to enter at least for [TIME {start}] and [TOPIC]. By P's continuous orientation towards CAL its role as a visual resource for turn production is affirmed.

After that a *turn vacant pause* [20] of 3.98 sec in which from a verbal perspective, 'nothing' happens, the analysis of gaze shows P's multimodal activity: After turn completion, P gazes at CAL for 0.98 sec (no noticeable uptake of CAL). After a 0.16 sec phase of gaze change, P orients towards ECA for 2.94 sec. By gazing at ECA, P indicates that this is the resource that is expected for uptake. but ECA again shows no uptake. This constitutes a discrepancy between the system's recipiency display (ECA's gaze at P) and missing uptake of the user's previous action (visual or verbal uptake).

The next part shows how P initiates another increment to handle missing uptake: With gaze directed at ECA, P starts an increment ("(int[erest-)]") that is interrupted multimodally in CAL: a highlight of [DAY'] and [TIME {start}' + TIME {end}'] becomes visible that immediately leads to P's gaze change towards CAL. Shortly afterwards and in overlap ECA verbally rephrases parameters ("[so] prepare luncheon monday from TEN o'clock to TWELVE o'clock;"). So, both system resources for uptake (ECA, CAL) are relevant resources in this context as P stops speaking. Again a closer inspection of the expanded turn shows that P initiates a different action to handle missing uptake which is *separate action*, i.e. this action can be marked as being off-topic, representing a challenge for the task-oriented system.

Case Analysis Summary: The inspection of the 'long turn's' emergent structure shows that the view on P's turn as being long from the outset can be reinterpreted as being turn increments that are produced systematically in line to the system's 'behavior'. To reconstruct the incremental structure on a verbal level, we assume prosody (mainly intonation) and pauses as structuring elements that are options for uptake, i.e. transition relevance places. With this framework the turn can be segmented into so-called turn constructional units (TCUs) [24]. Also, it shows that each TCU contains one action i.e. one parameter resulting in a well-structured utterance. Analysis is based on audio-visual activity of the system interface, that illustrates a practice of missing uptake options at transition relevant places which lead to turn expansion by the user, i.e. turn increments. We see, that the WOZ does not apply any resources that indicate information processing, system status or understanding, but sticks to the interactional structure of AE. Following [13] we can differentiate the user's *turn increment typology* for the context of HAI and task-related interaction that adds the system perspective: Case analysis shows that turn increments can be subdivided into *continued action*, *parallel action* and *separate action*. Within *continued* and *parallel action*, the user is oriented towards the initial task but adds specifications to an initial turn, or verbalizes related or consecutive information. These actions could be grouped as *extended actions* (Fig. 3). In contrast, *separate actions* perform new actions, like comments or assessments [13]. Analysis shows that the visual orientation towards visual (semiotic) resources [25] differs depending on the interactional task being worked on: For planning and verbalizing parameters CAL is the relevant visual resource, whereas when interactional trouble occurs, ECA becomes the relevant resource that is expected for uptake. A specific gaze conduct could be the basis for *task-related uptake resources* in the system's interface.

VII. Quantitative Analysis

The categories found in qualitative analysis are applied on a sub-corpus of interaction data. First, the interactional context of long turns is analyzed, i.e. the expectancy of long turns for the specific task of appointment entry. Second, the turn increment typology will be investigated and the resulting challenges for technical systems will be discussed. Finally, the relevance of multimodal uptake resources will be investigated.

A. Occurrence of increments: Global information phase

To understand where turn increments occur in task-oriented interaction we searched for turn increments with a length of > 4.0 sec in the sub corpus of 5 senior participants (1 male, 4 female). This was done to merge HRI / HAI research perspective that defines 'turn increments' as 'long turns' by time or number of words [2] with qualitative CA position, which emphasizes the emergent structure of turns.

TABLE I. INTERACTIONAL CONTEXT OF TURN INCREMENTS

Appointment Entries (N=49)						**Appointment Proposals (N=11)**
Global Information Phase (N=32)		***Local Information Phase (N=17)***			***Entry Phase (N=0)***	
Primal Formulation (N=27)	Re-formulation (N=5)	[TIME] / [DURATION] (N=14)	[TOPIC] (N=2)	[TIME + TOPIC] (N=1)	[DAY] (N=0)	

N=60 turn increments were found, Data processing included manual transcripts of speech [26] and gaze, TCU-segmentation of turns, annotation of interaction phases and turn increment types. Turn increments were found in AE (*N*=49) and AP (*N*=11) states. As the case analysis data was focused on appointment entries (AE) only, we will inspect those closer. It shows that the most frequent interactional context of turn increments is the GIP (*N*=32) with 27 primal formulations and 5 reformulations of parameters after system errors. *N*=17 turn increments were located in the Local Information Phase for [TIME] or [DURATION] as the most prominent group (*N*=14). Turn increments do not appear in the LIP for [DAY] and EP. In contrast to [8] where [TOPIC] was the interactional context for problematic long turns, we found that turn increments are mostly located in the primal formulation of appointment entries in the GIP.

B. Increment types: Extended action as most frequent type

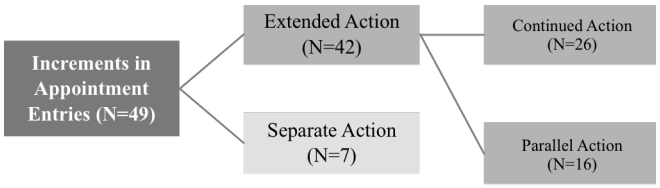

Figure 4. Distribution of turn increments within increment typology

The application of turn increment typology on turn increments in AE states (*N*=49) shows that the extended action increment type (*N*=42) can be found frequently, and continued action is the most common sub-type of increments (*N*=26) followed by parallel action (*N*=16). Separate action type increments were found in *N*=7 cases. So, by producing turn expansions users mostly stick to the initial task of appointment entry. The two extended action types can be rated as workable for the system as they are related to the task and the system's domain. In the cases of separate action, we found that these kind of actions might be challenging for the system as they contain (complex) information that is not directly related to the task of appointment entry. E.g. participants begin with an assessing separate action and then turn to appointment entry ("°h so BILlie; now i already told you THREE times that it SUITS me; (1.672) SATurday- (.) botanical GARden; (.)WALKover;" WOZ1-SEN-023/T7), or in other cases separate action is part of a self-assessing or self-talk activity ("the next appointment, (1.216) else do we have here; (2.328) that would be- (.) THURSday (---) from eight-een to half past eight cinema." WOZ1-SEN-021/T8).

C. Gaze: Calendar as the primal task-related resource

To examine relevant visual resources, we investigated *N*=8 turn increments. We could distinguish gaze at CAL and gaze at appointment cards. Following the qualitative analysis that differentiated CAL as planning resource, we summed up gaze at CAL and gaze at cards as CAL. Two substantial forms of gaze conduct were found (Fig. 4): 1. Orientation towards CAL only (*N*=4) and 2. orientation towards CAL and ECA (*N*=4), that can be subdivided into two groups of monitoring behavior: 2.a) a single gaze at ECA, and 2.b) in which gaze switches between ECA and CAL. Both main groups share the initial orientation towards CAL when users begin their turn, and the immediate orientation towards CAL, when system uptake (ECA speech or CAL highlight) is realized. The main difference is that participants of group 1. who only orient towards CAL seem to expect no uptake by ECA, in contrast to group 2.a) and b) who change gaze direction between CAL and ECA, when no uptake is noticeable. Although different timely onsets for gaze change can be seen, change of gaze direction appears at points of possible turn completion, or as turn-end projection [18]. This systematic procedure of orienting towards ECA indicates, that it is a relevant resource, when issues of uptake and recipiency seem to arise. Following the qualitative analysis, we can conclude that CAL is the primal task-related resource when appointment parameters are planned and verbalized, and ECA is the relevant resource when interactional trouble occur.

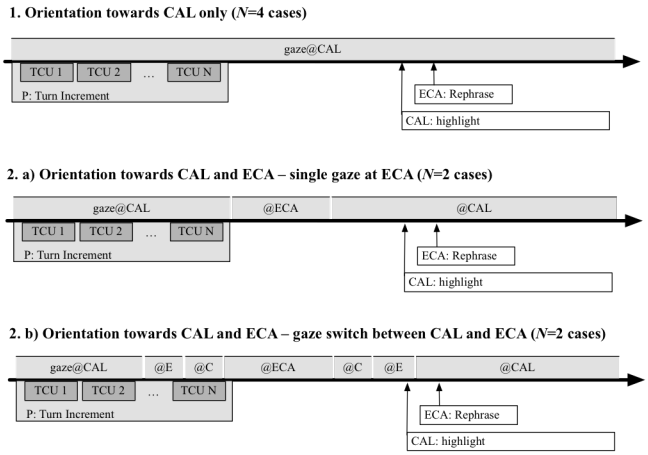

Figure 5. Forms of gaze conduct

VIII. DISCUSSION AND IMPLICATIONS FOR SYSTEM DESIGN

The study was set out to investigate long or expanded turns in task-related interaction of appointment entry and relevant interactional resources and practices. Results are discussed with respect to utilizing them to design autonomous systems.

1. Multimodal activity during verbal formulation: Users continually orient towards the system's multimodal behavior while verbalizing appointments. Their attention, indicated by gaze direction, exhibits a specific task-related orientation towards system resources with the calendar as a significant resource for planning and verbalizing appointments. Orientation towards the agent is found when uptake is expected but miss-

ing. The system's practice of initiating appointment entry without taking up information by relevant resources contributes to user's interpretation of the situation as being troublesome. The implication for system design is that system status, i.e. display of understanding, should be indicated by using relevant multimodal resources (e.g. representing ASR hypotheses in the calendar or a corresponding display of recipiency of the agent). Besides, typical gaze conduct could be assessed as areas where uptake is expected or as indicators for interactional trouble.

2. Incremental emergence of turns: Micro analysis shows that findings of HHI research can be applied to HRI as users incrementally add more to an initial turn when uptake is missing. By producing *turn increments* to an initial turn, participants provide options for uptake to their interlocutors and so, incremental turn expansion is a way of masking interactional trouble. Autonomous systems should therefore have strategies to handle turn increments, e.g. by multimodal interruption that displays understanding (or non-understanding) of user input.

3. Occurrence of turn increments in specific contexts: Turn increments were found predominantly at the beginning of an appointment entry (global information phase). This should be investigated and discussed further as it stands in contrast to previous findings [8] whereby turn increments were located in the context of verbalizing the appointment topic (local information phase). However, autonomous system should 'know' about the actual 'location' of interaction and that different user conduct can be expected in different phases of interaction. With this background the system could be alert of expectable or typical user conduct or orientation and provide uptake by relevant resources. Additional sensors (e.g. eye trackers) might help to validate the interaction state and user's orientation.

4. Different forms of actions performed by turn increments: We could show that turn increments can be subdivided into different actions: *Extended actions* continue the initial turn or action, by either *continued action*, e.g. by specifying an initial appointment, or by *parallel action*, that is, by performing a parallel task-oriented action like verbalizing a competitive appointment. *Separate actions* in contrast, do not extend the initial turn, but begin a different action (e.g. assessing missing uptake). Whereas the human WOZ could understand and handle these actions with provided resources, an autonomous system should have strategies to handle different types of actions especially when regarding parallel or separate actions, that initiate additional operations. Alternatively, the system could be equipped with strategies to avoid problematic turn increments as described above by relevant uptake or interruption.

REFERENCES

[1] L. A. Suchman, Plans and situated actions: The problem of human-machine communication. Cambridge University Press, 1987.

[2] F. Ghigi, M. Eskenazi, M. I. Torres, & S. Lee, "Incremental dialog processing in a task-oriented dialog." in INTERSPEECH 2014, pp. 308-312.

[3] M. Lohse, B. Wrede, & L. Schillingmann, "Enabling robots to make use of the structure of human actions-a user study employing Acoustic Packaging." in RO-MAN 2013 IEEE. IEEE. 2013, pp. 490-495.

[4] K. Pitsch., A.L. Vollmer, & M. Mühlig. Robot feedback shapes the tutor's presentation: How a robot's online gaze strategies lead to micro-adaptation of the human's conduct. Interaction Studies, 14(2), 2013 pp. 268-296.

[5] K. Pitsch, K.S. Lohan., K. Rohlfing, J. Saunders, C.L. Nehaniv, & B. Wrede. Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction. In RO-MAN, 2012 IEEE, pp. 974-981.

[6] M. ter Maat, K. P. Truong, and D. Heylen, How Turn-Taking Strategies Influence Users' Impressions of an Agent. Berlin, Heidelberg: Springer, 2010, pp. 441–453.

[7] R. Yaghoubzadeh &S. Kopp, "Towards graceful turn management in human-agent interaction for people with cognitive impairments." in SLPAT 2016. 2016, pp. 26-31.

[8] K. Pitsch, R. Yaghoubzadeh, & S. Kopp, S., "Entering Appointments: Flexibility and the Need for Structure?" in GSCL, 2015 pp. 140-141.

[9] K. Pitsch, "Ko-Konstruktion in der Mensch-Roboter-Interaktion. Kontingenz in Erwartungen und Routinen der Eröffnung", in U. Dausendschön-Gay, E. Gülich, U. Kraft (Eds.): Ko-Konstruktion als interaktive Verfahren, Bielefeld, Transkript, 2015, pp. 229-257.

[10] K. Pitsch, A.L. Vollmer, K. Rohlfing, J. Fritsch, & B. Wrede (2014). Tutoring in adult-child-interaction: On the loop of the tutor's action modification and the recipient's gaze. Interaction Studies, 15(1), 55-98.

[11] C. Goodwin, "The interactive construction of a sentence in natural conversation." in Everyday language: Studies in ethnomethodology, 1979, pp. 97-121.

[12] C. Goodwin, C. "Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning." in Sociological inquiry, 50(3-4), 1980, pp. 272-302.

[13] C. Ford, B. Fox, & S. A. Thompson, "Constituency and the grammar of turn increments." in The language of turn and sequence, C. Ford, B. Fox, S. A. Thompson, 2002, pp. 14-38.

[14] C.C. Heath, "The display of recipiency: An instance of a sequential relationship in speech and body movement." in Semiotica 42(2-4) 1982, pp. 147-168.

[15] R. Yaghoubzadeh, K. Pitsch, & S. Kopp, "Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users." in International Conference on Intelligent Virtual Agents 2015, pp. 28-38. Springer International Publishing.

[16] G. Skantze, & D. Schlangen, "Incremental dialogue processing in a micro-domain." In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics 2009, pp. 745-753.

[17] A. H. Oh, & A. I. Rudnicky, "Stochastic language generation for spoken dialogue systems. In Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems-Volume 3, 2000, pp. 27-32.

[18] D. DeVault, K. Sagae, & D. Traum, "Can I finish?: learning when to respond to incremental interpretation results in interactive dialogue." in SIGDIAL 2009, pp. 11-20.

[19] H. Sacks, E. A. Schegloff, & G. Jefferson. "A simplest systematics for the organization of turn-taking for conversation." in language, 1974, pp. 696-735.

[20] R. Schmitt, "Die Gesprächspause: Verbale" Auszeiten" aus multi-modaler Perspektive." in Deutsche Sprache 32(1), 2004, pp. 56-84.

[21] R. Yaghoubzadeh, M. Kramer, K. Pitsch, & S. Kopp, "Virtual agents as daily assistants for elderly or cognitively impaired people." in International Workshop on Intelligent Virtual Agents, 2013, pp. 79-91.

[22] Elan Annotation Software: https://tla.mpi.nl/tools/tla-tools/elan/

[23] J. Sidnell, T. Stivers, The handbook of conversation analysis, in Blackwell Handbooks of Linguistics, Wiley, 2012.

[24] M. Selting, "On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation." in Pragmatics, 6, 1996, pp. 371-388.

[25] C. Goodwin, "Action and embodiment within situated human interaction. " in Journal of pragmatics, 32(10), pp. 1489-1522.

[26] M. Selting, P. Auer, D. Barth-Weingarten, J.R. Bergmann, P. Bergmann, K. Birkner et al., "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)." In Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion. pp. 353-402.