The Tao of CHI: Towards Effective Human-Computer Interaction

Robert Porzel **Manja Baudis**

European Media Laboratory, GmbH
Schloss-Wolfsbrunnenweg 33
D-69118 Heidelberg, Germany
{robert.porzel,manja.baudis@eml-d.villa-bosch.de}

Abstract

End-to-end evaluations of conversational dialogue systems with naive users are currently uncovering severe usability problems that result in low task completion rates. Preliminary analyses suggest that these problems are related to the system's dialogue management and turn-taking behavior. We present the results of experiments designed to take a detailed look at the effects of that behavior. Based on the resulting findings, we spell out a set of criteria which lie orthogonal to dialogue quality, but nevertheless constitute an integral part of a more comprehensive view on dialogue *felicity* as a function of dialogue quality and efficiency.

1 Introduction

Research on dialogue systems in the past has focused on engineering the various processing stages involved in dialogical human-computer interaction (HCI) - e. g., robust automatic speech recognition, intention recognition, natural language generation or speech synthesis (cf. Allen et al. (1996), Cox et al. (2000) or Bailly et al. (2003)). Alongside these efforts the characteristics of computer-directed language have also been examined as a general phenomenon (cf. Zoeppritz (1985), Wooffitt et al. (1997) or Darves and Oviatt (2002)). The flip side, i. e., computer-human interaction (CHI), has received very little attention as a research question by itself. That is not to say that natural language generation and synthesis have not made vast improvements, but rather that the nature and design of the computer as an interlocutor itself, i. e., the effects of *human-directed language*, have not been scrutinized as such.

Looking at broad levels of distinctions for dialogue systems, e. g., that of Allen et al. (2001) between controlled and conversational dialogue systems, we note the singular employment of human-based differentiae, i. e., the degree of the restriction of the *human* interactions. Differentiae stemming from the other communication partner, i. e., the computer, are not taken into account - neither on a practical nor on a theoretical level.

In the past controlled and restricted interactions between the user and the system increased recognition and understanding accuracies to a level that systems became reliable enough for deployment in various real world applications, e. g., transportation or cinema information systems (Aust et al., 1995; Gorin et al., 1997; Gallwitz et al., 1998). Today's more conversational dialogue systems, e. g., SMARTKOM (Wahlster et al., 2001) or MATCH (Johnston et al., 2002), are able to cope with much less predictable user utterances. Despite the fact that in these systems recognition and processing have become extremely difficult, the reliability thereof has been pushed towards acceptable degrees by employing an array of highly sophisticated technological advances - such as dynamic lexica for multi-domain speech recognition and flexible pronunciation models (Rapp et al., 2000), robust understanding and discourse modeling techniques (Johnston, 1998; Engel, 2002; Alexandersson and Becker, 2001) combined with ontological reasoning capabilities (Gurevych et al., 2003; Porzel et al., 2003).

However, the usability of such conversational dialogue systems is still unsatisfactory, as shown in usability experiments with real users (Beringer, 2003) that employed the PROMISE evaluation framework described in Beringer et al. (2002), which offers some multimodal extentions over the PARADISE framework described in Walker et al. (2000) . The work described herein constitutes a starting point for a scientific examination of the whys and wherefores of the challenging results stemming from such end-to-end evaluations of conversational dialogue systems. Following a brief description of the state of the art in examinations of computer-directed language, we describe a new experimental paradigm, the first two studies using the paradigm and their corresponding results. Concluding, we discuss the ensuing implications for the design of successful and felicitous conversational dialogue systems.

2 Studies on Human-Computer Dialogues

The first studies and descriptions of the particularities of dialogical human-computer interaction, then labeled as *computer talk* in analogy to *baby talk* by Zoeppritz (1985), focused - much like subsequent ones - on:

- proving that a regular register for humans conversing with dialogue system exists, e. g., those of Krause (1992) and Fraser (1993),
- describing the regularities and characteristics of that register, as in Kritzenberger (1992) or Darves and Oviatt (2002).

The results of these studies clearly show that such a register exists and that its regularities can be replicated and observed again and again. In general, this work focuses on the question: what changes happen to human verbal behavior when they talk to computers as opposed to fellow humans? The questions which are not explicitely asked or studied are:

- how does the computer's way of communicating affect the human interlocutor,
- do the particulars of computer-human interaction help to explain why today's conversational dialogue systems are by and large unusable.

In this paper we claim that this shift of perspective is of paramount importance, for example, to make sense of the phenomena observable during end-to-end evaluations of conversational systems. We designed our experiments and started our initial observations using one of the most advanced conversational dialogue research prototypes existing today, i. e., the SMARTKOM system (Wahlster et al., 2001). This system designed for intuitive multimodal interaction comprises a symmetric set of input and output modalities (Wahlster, 2003), together with an efficient fusion and fission pipeline (Wahlster, 2002). SMARTKOM features speech input with prosodic analysis, gesture input via infrared camera, recognition of facial expressions and their emotional states. On the output side SMARTKOM employs a gesturing and speaking life-like character together with displayed generated text and multimedia graphical output. It currently comprised nearly 50 modules running on a parallel virtual machine-based integration software called *MULTIPLATFORM* (Herzog et al., 2003). As such it is certainly among the most advanced multi-domain conversational dialogue systems.

To the best of our knowledge, there has not been a single publication reporting a successful end-to-end evaluation of a conversational dialogue system with naive users. We claim that, given the state of the art of the dialogue management of today's conversational dialogue systems, evaluation trials with naive users will continue to uncover severe usability problems resulting in low task completion rates.[1] Surprisingly, this occurs despite acceptable partial evaluation results. By partial results, we understand evaluations of individual components such as concerning the word-error rate of automatic speech recognition or understanding rates as conducted by Cox et al. (2000) or reported in Diaz-Verdejo et al. (2000). As one of the reasons for the problems thwarting task completion, Beringer (2003) points at the problem of *turn overtaking*, which occurs when users rephrase questions or make a second remark to the system, while it is still processing the first one. After such occurrences a dialogue becomes asynchronous, meaning that the system responds to the second last user utterance while in the user's mind that response concerns the last. Given the current state of the art regarding the dialogue handling capabilities of HCI systems, this inevitably causes dialogues to fail completely.

We can already conclude from these informal findings that current state of the art conversational dialogue systems suffer from

a) a lack of turn-taking strategies and dialogue handling capabilities as well as

b) a lack of strategies for repairing dialogues once they become *out of sync*.

In human-human interaction (HHI) turn-taking strategies and their effects have been studied for decades in unimodal settings from Duncan (1974) and Sack et al. (1974) to Weinhammer and Rabold (2003) as well as more recently in multimodal settings as in Sweetser (2003). Virtually no work exists concerning the turn-taking strategies that dialogue systems should pursue and how they effect human-computer interaction, except in special cases such as in Woodburn et al. (1991) for the case of conversational computer-mediated communication aids for the speech and hearing impaired or Shankar et al. (2000) for turn negotiation in text-based dialogue systems. The overview of classical HCI experiments and their results, given in Wooffitt et al. (1997), also shows that problems, such as turn-overtaking, -handling and -repairs , have not been addressed by the research community.

In the following section we describe a new experimental paradigm and the first corresponding experiments tailored towards examining the effects of the computer's communicative behavior on its human partner. More specifically, we will analyze the differences in HHI and

[1] These problems can be diminished, however, if people have multiple sessions with the system and adapt to the respective system's behavior.

HCI/CHI turn-taking and dialogue management strategies, which, in the light of the recent end-to-end evaluation results described above, constitutes a promising starting point for an examination of the effects of the computer's communicative behavior. The overall goal of analyzing these effects is, that future systems become usable by exhibiting a more felicitous communicative behavior. After reporting on the results of the experiments in Section 4, we highlight a set of hypotheses that can be drawn from them and point towards future experiments that need to be conducted to verify these hypotheses in Section 6.

3 Experiments

For conducting the experiments we developed a new paradigm for collecting telephone-based dialogue data, called *Wizard and Operator Test* (WOT), which contains elements of both Wizard-of-Oz (WoZ) experiments (Francony et al., 1992) as well as Hidden Operator Tests (Rapp and Strube, 2002). This procedure also represents a simplification of classical end-to-end experiments, as it is - much like WoZ experiments - conductible without the technically very complex use of a real conversational system. As post-experimental interviews showed, this did not limit the feeling of *authenticity* regarding the simulated conversational system by the human subjects (S). The WOT setup consists of two major phases that begin after subjects have been given a set of tasks to be solved with the telephone-based dialogue system:

- in `Phase 1` the human assistant (A) is acting as a wizard who is simulating the dialogue system, much like in WoZ experiments, by operating a speech synthesis interface,
- in `Phase 2`, which starts immediately after a system breakdown has been simulated by means of beeping noises transmitted via the telephone, the human assistant is acting as a **human** operator asking the subject to continue with the tasks.

This setup enables to control for various factors. Most importantly the technical performance (e. g., latency times), the pragmatic performance (e. g., understanding vs. non-understanding of the user utterances) and the communicative behavior of the simulated systems can be adjusted to resemble that of state of the art dialogue systems. These factors can, of course, also be adjusted to simulate potential future capabilites of dialogue systems and test their effects. The main point of the experimental setup, however, is to enable precise analyses of the differences in the communicative behaviors of the various interlocutors, i. e., human-human, human-computer and computer-human interaction.

3.1 Technical Setup

During the experiment S and A were in separate rooms. Communication between both was conducted via telephone, i. e., for the user only a telephone was visible next to a radio microphone for the recording of the subject's linguistic expressions. As shown in Figure 1 the assistant/operator room featured a telephone as well as two computers - one for the speech synthesis interface and one for collecting all audio streams; also present were loudspeakers for feeding the speech synthesis output into the telephone and a microphone for the recording of the synthesis and operator output. With the help of an audio mixer all linguistic data were recorded time synchronously and stored in one audio file. The assistant/operator acting as the computer system communicated by selecting fitting answers for the subject's request from a prefabricated list covering the scope of the SMARTKOM repertoire of answers, which - despite the more conversational nature of the system, still does not include any kind of dialogue structuring or feedback particles. These responses were returned via speech synthesis through the telephone. Beyond that it was possible for the wizard to communicate over telephone directly with the subjects when acting as the human operator.

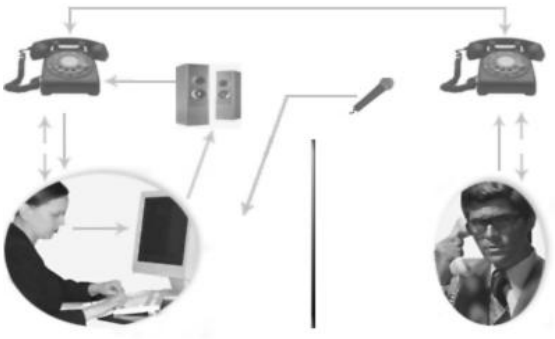

Figure 1: Communication in `Phase 1` goes from synthesized speech out of the loudspeakers into the operator room (left) phone to the subject room (right) and in `Phase 2` directly via the phone between the humans.

3.2 The Experiments

The experiments were conducted with an English setup, subjects and assistants in the United States of America and with a German setup, subjects and assistants in Germany. Both experiments were otherwise identical and in each 22 sessions were recorded. At the beginning of the WOT, the test manager told the subjects that they were testing a novel telephone-based dialogue system that supplies touristic information on the city of Heidelberg. In order to avoid the usual paraphrases of tasks worded too specifically, the manager gave the subjects an overall list

of 20 very general touristic activities, such as *visit museum* or *eat out*, from which each subject had to pick six tasks which had to be solved in the experiment. The manager then removed the original list, dialed the system's number on the phone and exited from the room after handing over the telephone receiver. The subject was always greeted by the system's standard opening ply: *Welcome to the Heidelberger tourist information system. How I can help you?* After three tasks were finished (some successful some not) the assistant simulated the system's break down and entered the line by saying *Excuse me, something seems to have happened with our system, may I assist you from here on* and finishing the remaining three tasks with the subjects.

4 Results

The PARADISE framework (Walker et al., 1997; Walker et al., 2000) proposes distinct measurements for dialogue quality, dialogue efficiency and task success metrics. The remaining criterion, i. e., user satisfaction, is based on questionaries and interviews with the subjects and cannot be extracted (sub)automatically from logfiles. The analyses of the experiments described herein focus mainly on dialogue efficency metrics in the sense of Walker et al. (2000). As we will show below, our findings strongly suggest that a felicitous dialogue is not only a function of dialogue quality, but critically hinges on a minimal threshold of efficiency and overall dialogue management as well. While these criteria lie orthogonal to the Walker et al. (2000) criteria for measuring dialogue quality such as recognition rates and the like, we regard them to constitute an integral part of an aggregate view on dialogue quality and efficiency, herein referred to as *dialogue felicity*. For examining dialogue felicity we will provide detailed analyses of efficiency metrics *per se* as well as additional metrics for examining the number and effect of pauses, the employment of feedback and turn-taking signals and the amount of overlaps.

The Data: The length of the dialogues was on average 5 minutes for the German (G) and 6 minutes for the English (E) sessions.[2] The subjects featured approximately proportional mixtures of gender (25m,18f), age (12< >71) and computer expertise. Table 1 shows the duration and turns per phase of the experiment.

Measurements: First of all, we apply the classic Walker et al. (2000) metric for measuring dialogue efficiency, by calculating the number of turns over dialogue length. Figure 2 shows the discrepancy between the dialogue efficiency in `Phase 1` (HCI) versus `Phase 2` (HHI) of the German experiment and Figure 3 shows that the same patterns can be observed for English.

Phase	HHI G	HHI E	HCI G	HCI E
Average length	1:52 min.	2:30 min.	2:59 min.	3:23 min.
Average turns	11.35	21.25	9.2	7.4

Table 1: Average length and turns in `Phase 1` and `2`

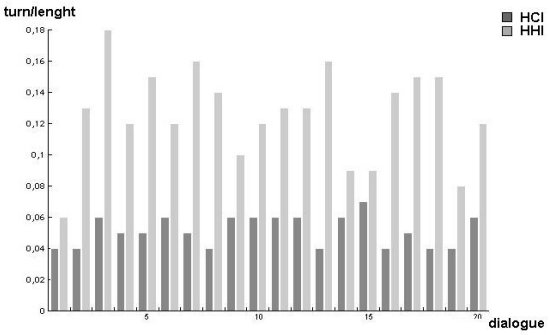

Figure 2: Dialogue efficiency (German data)

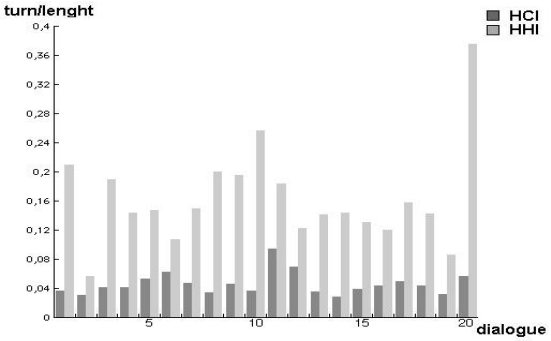

Figure 3: Dialogue efficiency (English data)

As this discrepancy might be accountable by latency times alone, we calculated the same metric with and without pauses. For these analyses pauses are very conservatively defined as silences during the conversation that exceeded one second. The German results are shown in Figure 4 and, as shown in Figure 5 we find the same patterns hold cross-linguistically in the English experiments. The overall comparison, given in Table 2, shows that - as one would expect - latency times severely decrease dialogue efficiency, but that they alone do not account for the difference in efficiency between human-human and human-computer interaction. This means that even if latency times were to vanish completely, yielding actual real-time performance, we would still observe less efficient dialogues in HCI.

[2]The shortest dialogues were 3:18 (E) and 3:30 (G) and the longest 12:05 (E) and 10:08 (G).

While it is obvious that the existing latency times increase the number and length of pauses of the computer interactions as compared to the human operator's interactions, there are no such obvious reasons why the number and length of pauses in the human subjects' interactions should differ in `Phase 1` and `Phase 2`. However, as shown in Table 3, they do differ substantially.

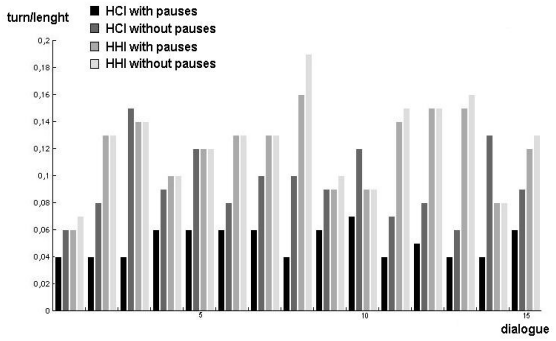

Figure 4: Efficiency w/out latency in German

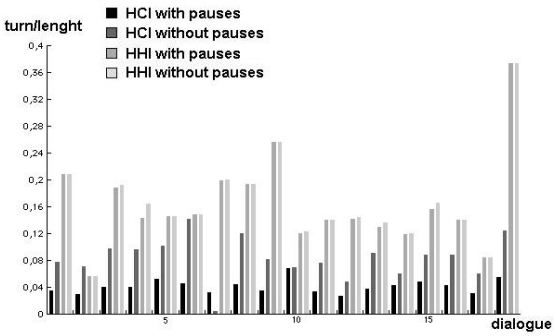

Figure 5: Efficiency w/out latency in English

Efficiency	HCI -p	HCI +p	HHI -p	HHI +p
Mean German	0.18	0.05	0.25	0.12
Standard-deviation	0.04	0.01	0.06	0.03
Mean English	0.16	0.05	0.17	0.17
Standard-deviation	0.25	0.02	0.07	0.07

Table 2: Overall dialogue efficiencies with pauses +p and without pauses -p.

Next to this *pause-effect*, which contributes greatly to dialogue efficiency metrics by increasing dialogue length, we have to take a closer look at the individual turns and their nature. While some turns carry propositional information and constitute utterances proper, a significant number solely consists of specific particles used to exchange signals between the communicative partners or combinations thereof. We differentiate between dialogue-structuring signals and feedback signals in the sense of Yngve (1970). Dialogue-structuring signals - such as hesitations like *hmm* or *ah* as well as expressions like *well*, *yes*, *so* - mark the intent to begin or end an utterances, make corrections or insertions. Feedback signals- while sometimes phonetically alike - such as *right*, *yes* or *hmm* - do not express the intent to take over or give up the speaking role, but rather serve as a means to stay in contact with the speaker, which is why they are sometimes referred to as *contact signals*.

Pauses	HCI-G	HHI-G	HCI-E	HHI-E
Number total	79	10	94	21
Number per dialog	3.95	0.5	4.7	1.05
Number per turn	0.46	0.05	0.64	0.05
total length	336sec	19sec	467sec	48sec
% of phase	9.37	0.84	13.74	1.75
% of dialogue	5.75	0.3	7.46	0.766

Table 3: Overall pauses of human subjects: `Phase 1` and `2` German (HCI-G/HHI-G) and English (HCI-G/HCI-E)

In order to be able to differentiate between the two, for example, between an agreeing feedback *yes* and a dialogue-structuring one, all dialogues were annotated manually. Half of the data were annotated by separate annotators, yielding an inter-annotator agreement of 90.61%. The resulting counts for the user utterances in phase one and two are shown in Table 4. Not shown in Table 4 are the number of particles employed by the computer, since it is zero, and of the human operator in the HHI dialogues, as they are like those of his human interlocutor.

Again, the findings for both German and English are comparable. We find that feedback particles almost vanish from the human-computer dialogues - a finding that corresponds to those described in Section 2. This linguistic behavior, in turn, constitutes an adaptation to the employment of such particles by that of the respective interlocutor. Striking, however, is that the human subjects still attempted to send dialogue structuring signals to the computer, which - unfortunately - would have been ig-

nored by today's "conversational" dialogue systems.[3]

Particles	structure HCI	particle HHI	feedback HCI	particle HHI
Number total	112 G 90 E	225 G 202 E	18 G 0 E	135 G 43 E
per dialogue	5.6 G 4.5 E	11.25 G 10.1 E	0.9 G 0 E	6.75 G 2.15 E
per turn	0.4 G 0.61 E	0.59 G 0.48 E	0.04 G 0 E	0.26 G 0.1 E

Table 4: Particles of human subjects: HCI vs. HHI

Before turning towards an analysis of this data we will examine the overlaps that occurred throughout the dialogues. Most overlaps in human-human conversation occur during turn changes with the remainder being feedback signals that are uttered during the other interlocutor's turn (Jefferson, 1983). The results on measuring the amount of overlap in our experiments are given in Table 5. Overall the HHI dialogues featured significantly more overlap than the HCI ones, which is partly due to the respective presence and absence of feedback signals as well as due to the fact that in HCI turn takes are accompanied by pauses rather than immediate - overlapping - hand overs.

Overlaps	HCI-G	HHI-G	HCI-E	HHI-E
Number total	7	49	4	88
per dialogue	0.35	3.06	0.2	4.4
per turn	0.03	0.18	0.01	0.1

Table 5: Overlaps in Phase 1 versus Phase 2

Lastly, our experiments yielded negative findings concerning differences in the type-token ratio (denoting the lexical variation of forms), speech production errors (false starts, repetitions etc.) and syntax. This means that there was no statistically significant difference in the linguistic behavior with respect to these factors. We regard this finding to strengthen our conclusions (see Section 6), that to emulate human syntactic and semantic behavior does not suffice to guarantee effective and therefore felicitous human computer interaction.

5 An Analysis of Ineffective Computer-Human Interaction

The results presented above enable a closer look at dialogue efficiency as one of the key factors influencing overall dialogue felicity. As our experiments show, the difference between the human-human efficiency and that of the human-computer dialogues is not solely due to the computer's response times. There is a significant amount of *white noise*, for example, as users wait after the computer has finished responding. We see these behaviors as a result of a mismanaged dialogue. In many cases users are simple unsure whether the system's turn has ended or not and consequently wait much longer than necessary.

The situation is equally bad at the other end of the turn taking spectrum, i. e., after a user has handed over the turn to the computer, there is no signal or acknowledgment that the computer has taken on the baton and is running along with it - regardless of whether the user's utterance is understood or not. Insecurities regarding the main question, i. e., *whose turn is it anyways*, become very notable when users try to establish contact, e. g., by saying *hello* -pause- *hello*. This kind of behavior certainly does not happen in HHI, even when we find long silences.

Examining why silences in human-human interaction are unproblematic, we find that, these silences have been announced, e. g., by the human operator employing linguistic signals, such as *just a moment please* or *well, I'll have to have a look in our database* in order to communicate that he is holding on to the turn and finishing his round.

To push the relay analogy even further, we can look at the differences in overlap as another indication of crucial dialogue inefficiency. Since most overlaps occur at the turn boundaries and, thusly, ensure a smooth (and fast) hand over, their absence constitutes another indication why we are far from having winning systems.

6 Conclusion and Future Work

As the primary effects of the human-directed language exhibited by today's conversational dialogue systems, our experiments show that the human interlocutor:

- ceases in the production of feedback signals, which has been observed before,
- still attempts to use his or her turn signals for marking turn boundaries - which, however, remain ignored by the system - and
- increases the amount of pauses, caused by waiting and uncertainty effects, which also is manifested by missing overlaps at turn boundaries.

Generally, we can conclude that a felicitous dialogue needs some amount of extra-propositional exchange between the interlocutors. The complete absence of such dialogue controlling mechanisms - by the non-human interlocutors alone - literally causes the dialogical situation to get out of control, as observable in the turn-taking and -overtaking phenomena described in Section 2. As witnessable in recent evaluations, this way of behaving does

[3] In the English data the subject's employment of dialogue structuring particles in HCI even slightly surpassed that of HHI.

not serve the intended end, i. e., efficient, intuitive and felicitous human-computer interaction.

As future work we propose to take the Wizard and Operator Test paradigm introduced herein and to change and adjust the parameters of the computer-human interaction - while performing subsequent measurements of the ensuing effects - until an acceptable degree of dialogue efficiency is reached. That is, finding out just how much extra-propositional signaling is needed to guarantee a felicitous dialogue. Such communicative behavior, then has to be implemented in dialogue systems, to make their way of communicating more like that of their human partners. In our minds, achieving dialogue quality remains an important challenge for the scientific community, but - as we have shown herein and seen in recent evaluations - dialogue efficiency constitutes another necessary condition for achieving dialogue felicity.

Acknowledgments

This work has been partially funded by the German Federal Ministry of Research and Technology (BMBF) and by the Klaus Tschira Foundation as part of the SMARTKOM, SMARTWEB, and EDU projects. We would like to thank the International Computer Science Institute in Berkeley for their help in collecting the data especially, Lila Finhill, Thilo Pfau, Adam Janin and Fey Parrill.

References

Jan Alexandersson and Tilman Becker. 2001. Overlay as the basic operation for discourse processing. In *Proceedings of the IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. Springer, Berlin.

James F. Allen, Bradford Miller, Eric Ringger, and Teresa Sikorski. 1996. A robust system for natural spoken dialogue. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, USA.

James F. Allen, George Ferguson, and Amanda Stent. 2001. An architecture for more realistic conversational system. In *Proceedings of Intelligent User Interfaces*, Santa Fe, NM.

Harald Aust, Martin Oerder, Frank Seide, and Volker Steinbiss. 1995. The Philips automatic train timetable information system. *Speech Communication*, 17.

Gerard Bailly, Nick Campbell, and Bernd Möbius. 2003. Isca special session: Hot topics in speech synthesis. In *Proceedings of the European Conference on Speech Communication and Technology*, Geneva, Switzerland.

Nicole Beringer, Ute Kartal, Katerina Louka, Florian Schiel, and Uli Türk. 2002. PROMISE: A Procedure for Multimodal Interactive System Evaluation. In *Proceedings of the Workshop 'Multimodal Resources and Multimodal Systems Evaluation*, Las Palmas, Spain.

Nicole Beringer. 2003. The SmartKom Multimodal Corpus - Data Collection and End-to-End Evaluation. In *Colloquium of the Department of Linguistics*, University of Nijmwegen, June.

R.V. Cox, C.A. Kamm, L.R. Rabiner, J. Schroeter, and J.G. Wilpon. 2000. Speech and language processing for next-millenium communications services. *Proceedings of the IEEE*, 88(8).

Charles Darves and Shannon Oviatt. 2002. Adaptation of Users' Spoken Dialogue Patterns in a Conversational Interface. In *Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, U.S.A.

J. Diaz-Verdejo, R. Lopez-Cozar, A. Rubio, and A. De la Torre. 2000. Evaluation of a dialogue system based on a generic model that combines robust speech understanding and mixed-initiative control. In *2nd International Conference on Language Resources and Evaluation (LREC 2000)*, Athens, Greece.

Starkey Duncan. 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 3.

Ralf Engel. 2002. SPIN: Language understanding for spoken dialogue systems using a production system approach. In *Proceedings of the International Conference on Speech and Language Processing 2002*, Denver, USA.

J.-M. Francony, E. Kuijpers, and Y. Polity. 1992. Towards a methodology for wizard of oz experiments. In *Third Conference on Applied Natural Language Processing*, Trento, Italy, March.

Norman Fraser. 1993. Sublanguage, register and natural language interfaces. *Interacting with Computers*, 5.

Florian Gallwitz, Maria Aretoulaki, Manuela Boros, Jürgen Haas, Stefan Harbeck, R. Huber, Heinrich Niemann, and Elmar Nöth. 1998. The Erlangen spoken dialogue system EVAR: A state-of-the-art information retrieval system. In *Proceedings of 1998 International Symposium on Spoken Dialogue*, Sydney, Australia.

Allen L. Gorin, Guiseppe Riccardi, and Jerry H. Wright. 1997. How may I help you? *Speech Communication*, 23.

Iryna Gurevych, Robert Porzel, and Stefan Merten. 2003. Less is more: Using a single knowledge representation in dialogue systems. In *Proceedings of the HLT/NAACL Text Meaning Workshop*, Edmonton, Canada.

Gerd Herzog, Heinz Kirchmann, Stefan Merten, Alassane Ndiaye, Peter Poller, and Tilman Becker. 2003. MULTIPLATFORM: An integration platfrom for multimodal dialogue systems. In *Proceedings of the HLT/NAACL SEALTS Workshop*, Edmonton, Canada.

G. Jefferson. 1983. Two explorations of the organistion if overlapping talk in coversation. *Tilburg Papers in Language and Literature*, 28.

Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker, and Preetam Maloor. 2002. Match: An architecture for multimodal dialogue systems. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Germany.

Michael Johnston. 1998. Unification-based multimodal parsing. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association of Computational Linguistics*, Montreal, Canada.

J. Krause. 1992. Natürlichsprachliche mensch-computer-interaktion als technisierte kommunikation: Die computer talk-hypothese. In J. Krause and L. Hitzenberger, editors, *Computer Talk*. Olms, Hildesheim.

H. Kritzenberger. 1992. Unterschiede zwischen mensch-computer-interaktion und zwischenmenschlicher kommunikation aus der interpretativen analyse der dicos-protokolle. In J. Krause and L. Hitzenberger, editors, *Computer Talk*, pages 122–156. Olms, Hildesheim.

Robert Porzel, Norbert Pfleger, Stefan Merten, Markus Löckelt, Ralf Engel, Iryna Gurevych, and Jan Alexandersson. 2003. More on less: Further applications of ontologies in multi-modal dialogue systems. In *Proceedings of the 3rd IJCAI 2003 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Acapulco, Mexico.

Stefan Rapp and Michael Strube. 2002. An iterative data collection approach for multimodal dialogue systems. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Spain.

Stefan Rapp, Sunna Torge, Silke Goronzy, and Ralf Kompe. 2000. Dynamic speech interfaces. In *Proceedings of the ECAI 2000 Workshop on Artificial Intelligence in Mobile Systems*, Berlin, Germany.

Sadock Sack, E Schegloff, and G Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50.

Tara Rosenberger Shankar, Max VanKleek, Antonio Vicente, and Brian K. Smith. 2000. A computer mediated conversational system that supports turn negotiation. In *Proceedings of the Hawai'i International Conference on System Sciences*, Maui, Hawaii, January.

Eve Sweetser. 2003. Levels of meaning in speech and gesture: Real space mapped onto epistemic and speech-interactional mental spaces. In *Proceedings of the 8th International Conference on Cognitive Linguistics*, Logrono, Spain, July.

Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. 2001. Smartkom: Multimodal communication with a life-like character. In *Proceedings of the 7th European Conference on Speech Communication and Technology*.

Wolfgang Wahlster. 2002. SmartKom: Fusion and fission of speech, gerstures and facial expressions. In *Proceedings of the Firsat International Workshop on Man-Machine Symbiotic Systems*, Kyoto, Japan.

Wolfgang Wahlster. 2003. SmartKom: Symmetric multimodality in an adaptive an reusable dialog shell. In *Proceedings of the Human Computer Interaction Status Conference*, Berlin, Germany.

Marilyn Walker, Diane Litman, Candace Kamm, and Alicia Abella. 1997. PARADISE: A framework for evaluating spoken dialogue agents. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain.

Marilyn A. Walker, Candace A. Kamm, and Diane J. Litman. 2000. Towards developing general model of usability with PARADISE. *Natural Language Engeneering*, 6.

Karl Weinhammer and Susan Rabold. 2003. Durational Aspects in Turn Taking. In *Proceedings of International Conference Phonetic Sciences*, Barcelona, Spain.

R. Woodburn, R. Procter, J. Arnott, and A. Newell. 1991. A study of conversational turn-taking in a communication aid for the disabled. In *People and Computers*, pages 359–371. Cambridge University Press, Cambridge.

Robin Wooffitt, Nigel Gilbert, Norman Fraser, and Scott McGlashan. 1997. *Humans, Computers and Wizards: Conversation Analysis and Human (Simulated) Computer Interaction.* Brunner-Routledge, London.

V Yngve. 1970. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, Chicago, Illinois, April.

Magdalena Zoeppritz. 1985. Computer talk? Technical report, IBM Scientific Center Heidelberg Technical Report 85.05.