

# Potential Benefits of Human-Like Dialogue Behaviour in the Call Routing Domain

Joakim Gustafson, Mattias Heldner, and Jens Edlund

KTH Speech Music and Hearing  
{jocke,mattias,edlund}@speech.kth.se

**Abstract.** This paper presents a Wizard-of-Oz (Woz) experiment in the call routing domain that took place during the development of a call routing system for the TeliaSonera residential customer care in Sweden. A corpus of 42,000 calls was used as a basis for identifying problematic dialogues and the strategies used by operators to overcome the problems. A new Woz recording was made, implementing some of these strategies. The collected data is described and discussed with a view to explore the possible benefits of more human-like dialogue behaviour in call routing applications.

## 1 Introduction

This paper discusses the possibility of making the interaction with *natural-language call routing systems* more human-like through different dialogue strategies, improved interaction control and utterance unit segmentation, and explores possible benefits of such dialogue behaviour in call-routing and similar applications. We use a large corpus of 42,000 Woz dialogues, where wizards were given the possibility to route problematic calls back to themselves and to act as operators. Based on an analysis of 82 such human-human intervention dialogues, we re-designed the system prompts in the Woz setup to capture the operators' behaviour in the intervention dialogues. We collected 188 new calls with the modified Woz setup in order to study the effect of adding more human-like dialogue behaviour to a call routing dialogue system. Specifically, we wanted to investigate if the new behaviour would elicit longer and semantically richer user utterances, and whether it would affect the quality of service.

## 2 Natural Language Call Routing Systems

The goal of customer care centres in large companies is to direct callers to the appropriate human operator or self-service application. In simpler cases, this can be automated using a voice controlled menu system, in which callers should ideally be presented with no more than four to six choices at each dialogue step. This is a limitation which makes voice controlled menu systems and system directed dialogues in general unsuitable for call centres where the number of possible reasons for calling is large, since it is difficult and time consuming to navigate large menu trees. Another problem is that the design of menu trees typically reflect the solution to the problem,

while the callers usually only know the effects of the problem – the symptoms – and call to find the solution [Acomb et al., 2007]. Making callers navigate a menu tree that reflects, from the point of view of the callers, an unknown problem solving structure is not likely to be optimal. Another possibility is to ask callers to describe their symptoms and let an automatic classifier decide to which tree node they should be directed. Systems automatically directing callers based on their verbal description of their concern are called *natural language call-routing systems*. Callers are either routed to the appropriate human operator or self-service application, or taken through additional dialogue steps allowing the system to obtain more information. This kind of call-routing is becoming increasingly more common in commercial settings. Well-known examples include the AT&T How May I Help You (HMIHY) [Gorin et al., 1997], Bell Canada's 310-BELL customer service line, and more recently the entrance to TeliaSonera residential customer care in Sweden [Wirén et al., 2007]. A related application is that of *automated technical support* [Acomb et al., 2007].

### 3 Human-Like Dialogues

In this paper we explore the possible benefits of adding more human-like conversational behaviour to commercial call routing systems, under the hypothesis that when confronted with more human-like system behaviour, callers are more likely to behave as if they were speaking to a human operator. This hypothesis is supported by other studies. For example, users of spoken dialogue systems generally produce shorter utterances than when they speak to human beings [e.g. Zoltan-Ford, 1991]. A finding from early data collections in the HMIHY project [Gorin et al., 1997] provides a relevant example. Gorin and colleagues noted a unimodal distribution and a very long tail in the histogram of utterance length. On closer inspection, the long utterances turned out to be interspersed with backchannel feedback by the operator. When similar interactions were recorded with pre-recorded prompts (and no backchannel feedback), the long tail disappeared, and a bimodal distribution appeared, with the extra mode on very short utterances. Gorin and colleagues call this particular computer-directed manner of talking *menu-speak*. Others have used other terms for this type of talk, for example *computerese* [Gustafson et al., 1997], *machine talk* [Martinovsky and Traum, 2003], and *computer talk* [Fischer, 2006]. The hypothesis that human-like system behaviour elicits user behaviour that is more like human-human dialogue and less menu-speak is also supported by a large number of studies on entrainment, showing generally that people adapt their speaking behaviour to the behaviour of their interlocutor, even if that happens to be a computer [Bell, 2000, Brennan, 1996, Garrod and Pickering, 2004]. Finally, the hypothesis is in effect revisited and put to the test in the experiments reported here.

As noted above, people generally produce longer and more semantically rich utterances when speaking to humans than when speaking to computers, and a principal reason for wanting callers to behave more like when speaking to humans in call routing applications is to elicit longer, richer utterances. As a call routing system analyses callers' verbal descriptions to automatically route the call, short menu-speak may force the system to prompt callers for many pieces of information, making it very similar to traditional menu based customer care systems. Longer and semantically richer descriptions stand a better chance of containing sufficient information for

appropriate routing. That this information can be utilised by human operators is evident, but whether current call-routing technology can make as much use of it is an open question. The technology used by Gorin and colleagues in the 90s was not helped by the longer human-human data (A. Gorin, personal communication, February 2nd 2006), and to the extent that the necessary information is present in the utterances, this represents a challenging task for the data miners.

## 4 Wizard of Oz Collections

Wizard-of-Oz (Woz) simulation is often used for collection of human-computer data when building a fully functional system is impractical or too expensive. In a Woz simulation, a human (the wizard) performs (part of) the system's functions unbeknownst to the subject, who is led to believe the system is fully automated and operational. As a typical example, Woz data collections are used in the early stages of a spoken dialogue system project to gather the first set of data on which the various models used by the system are based [Wooffitt et al., 1997]. Traditionally, wizards have been asked to produce output that resembles what can be expected from a talking computer, rather than from a person, so that the collected data will be representative for such interactions [Dahlbäck et al., 1993], and in many Woz collections, the wizard is the system designer and the subjects are friends or students that are given scenarios with tasks to solve with the system. Allwood & Haglund argue that the wizard and the subjects play roles at several possibly conflicting levels in Woz collections [Allwood and Haglund, 1992]. At the same time, the wizard has the role of the researcher and is playing the role of the system. Similarly, the user has the role of a subject in a scientific study whilst playing the role of a customer.

In a paper describing the development of the TeliaSonera call routing system [Wirén et al., 2007], Wirén et al. argue that the role playing aspects of traditional Woz collections make them unsuitable for use in the initial steps of developing a call routing system, since “we want to learn not just *how* callers express themselves, but also *what kind of tasks they have*, which obviously rules out prewritten scenarios.” To overcome the lack of realism in traditional Woz collections they conducted what they coined an *in-service Wizard-of-Oz data collection*, where real customer care operators acted as wizards handling calls from real customers with real problems. Using actual customer care operators as wizards provided valuable feedback on dialogue and prompt design. Furthermore, by allowing the wizards to route complicated calls to themselves, the in-service Woz setup yielded follow-up dialogues representative of how a human operator would sort out the problem at hand. These human-human dialogues are the basis for the present study. Porzel has proposed a similar variation of Woz simulations, which he called WOT (Wizard and Operator Test) [Porzel, 2006]. WOT involves a human operator acting as wizard. At a predefined moment, an obvious system breakdown is simulated, after which the operator stops acting as a wizard and takes the call in person, telling the caller that the system broke down and that (s)he will have to handle the remaining tasks. The result is that human-computer (wizard) and human-human (operator) data is collected within the same dialogue. In this paper, a combination of the in-service Woz and WOT methods is used to explore possible benefits of adding human-like conversational behaviour in a call routing system. The method is summarised in the following steps:

1. Perform an in-service Woz collection, in which the wizards are encouraged to intervene when the pre-defined prompt set is insufficient and causes communication problems.
2. Analyse the intervention dialogues to gain information about how the problem was solved and what linguistic and/or conversational resources were used.
3. Re-design the system prompts, adding dialogue features identified in the step 2.
4. Perform a second round of in-service Woz collections.
5. Compare relevant aspects of the interaction with the first collection.

Steps 2-5 can be iterated using all available intervention dialogues as a source of information when seeking more dialogue features to investigate. In this method, the wizards themselves decide when it is necessary to resort to intervention dialogues, making it more labour intensive than the WOT method, which always results in two types of dialogue for each recording. However, in an in-service Woz, the WOT method is not feasible, since deliberately causing system failure would annoy paying customers. More importantly, the wizards' decisions on when to intervene give valuable information about the limitations of the current set of system prompts. Finally, the method clearly identifies callers who have problems describing their errand in menu-speak, and allows us to investigate whether they also have problems describing it with human-directed talk.

## 5 Three Call Routing Dialogue Collections

The call routing data collections and corpora described here are all in Swedish. The examples given are all translated from Swedish to English by the authors. The general statistics of the corpora are as follows:

Label	# dialogues	mean # turns/dialogue
ISWOZ-I	2228	4.6
INTERVENT	82	10.2
ISWOZ-II	188	6.4

All corpora have been labelled on the utterance level with a small label set: INITIAL HMIHY was used for the first open prompt, DESCRIPTION for the callers' description of the nature of their requests, HMIHY for a general request for more details, FOLLOW-UP QUESTION for requests for specific information, ANSWER for the callers' responses to FOLLOW-UP QUESTIONS, GREETING for greetings on both sides, CHANNEL for channel checks verifying that parties could hear each other, FEEDBACK for feedback, META for turns that discussed the dialogue itself (What did you say?) or the nature of the speaker (Are you a human or machine?). ROUTING was used when callers were informed that they were being directed to an operator. The corpora and their collection are described in detail in the following.

### 5.1 In-Service Woz I (ISWOZ-I)

In 2005, the Swedish telecom operator TeliaSonera developed a Swedish natural language call routing system for their main customer care line, a service which handles 14 million calls per year since its deployment in 2006. During the development, 42,000 calls were collected in an in-service Woz, in which the wizards were ten real

customer care operators that handled real incoming calls [Wirén et al., 2007]. The initial open prompt was designed to inform callers that they were talking to a machine, but that they could express themselves freely. The wording was: “Welcome to Telia. Here you describe the nature of your request in your own words instead of pressing buttons on your phone. If you say what you need help with I can direct you to the correct place in the customer care centre. What are you calling about?” After this initial open prompt, the system engaged in a system-driven, menu-based dialogue in those cases additional pieces of information was needed to route the call. In case the dialogue got stuck, the wizards had the option to let the system say “You are now being directed to an operator” (the same utterance that was used when routing was successfully achieved), whilst routing the call to themselves, effectively taking over the call in the role of the operator (done in about 5% of the calls). The design decision in this collection was to use system prompts that would signal that it was a machine talking, in order to limit the callers’ expectations on the system’s understanding capabilities [see also Boyce, 1999]. In particular, the aim was to achieve consistency between the initial open prompt and the subsequent system-driven disambiguation prompts. We will call the corpus collected in this design *in-service Woz I* (ISWoz-I), to contrast with the second in-service Woz described in this paper (ISWoz-II).

## 5.2 The Operator Intervention Dialogues (INTERVENT)

The second data set (INTERVENT) was obtained as an effect of the design of the in-service Woz. It consists of 82 of the ISWoz-I dialogues that lead to a communication breakdown. As a result of the breakdown, these dialogues have two parts: the first part is the Woz dialogue leading to the breakdown, and the second is the human-human dialogue taking place afterwards, as operators routed the calls to themselves. The dialogues are of particular interest since they give access to callers that had a hard time describing their reason for calling to a machine. It is worth noting that the problems leading to interventions were not insurmountable. The operators succeeded, without exception, in collecting the information needed to route these calls in the human-human dialogue. Furthermore, a small number of specific reasons causing the Woz dialogues to get stuck were discernible: (1) many dialogues never got started, as the caller was unsure whether the system was listening when they said “Hello?”, although the system responded “This is a voice controlled system where you describe the nature of your request in your own words. You can for example say...”; (2) callers found it hard to match their problem to the multiple-choice voice menus; (3) callers would occasionally answer “yes” to multiple choices given by the system; and (4) callers began by providing background information, either personal, as in “*This is Lars, calling from Stockholm*”, or task-related: “*I ordered broadband from you three weeks ago*”, for which the system had no response. After taking over the calls, operators solved all these issues with ease: for (1), they simply responded with “*Hello, what’s your problem?*”; (2) did not reappear, callers were perfectly able to describe their concerns to the operator; (3) was avoided by asking a yes/no question instead, such as “*Does it concern your land line?*”, using the most probable choice in that context, to which the caller typically responded “*Yes*” or “*No, my cell phone*” (there was not one instance of operators presenting multiple choices to get information in the 82 calls); and (4), finally, was typically handled with short utterances like “ok”, “hi”,

“uh-huh” or “yes” encouraging the caller to continue describing their concerns (these feedback utterances constitute 24% of all operator utterances in these dialogues). To conclude, the operators made use of basic conversational skills by responding promptly to channel checks and greetings and encouraging callers to keep talking by providing feedback.

### 5.3 In-Service Woz II (ISWoz-II)

The analysis of the INTERVENT data led us to make a second in-service Woz collection (ISWoz-II) to investigate how well the wizards would perform their task if given a redesigned prompt piano with which they could generate the behaviour they displayed in the intervention dialogues. Furthermore, we wanted to see how callers would react when faced with a dialogue system displaying such dialogue strategies. The setup with real customer care operators acting as wizards was kept and used to collect 188 new calls. The new collection was explicitly designed to make the system more human-like. In order to avoid moral indignation in the callers by pretending to be a real human operator, the calls began with a different voice announcing that “You are now being directed to an automatic voice controlled operator” followed by a 7 second silence. Then the newly designed initial open prompt “Welcome to Telia how may I help you?”, read in a casual manner, was presented. The multiple-choice voice menus were replaced by the kind of yes/no-questions the operators had used. In addition, there was a prompt to be used to follow up no-answers: “So what is it about?” “Yes, hello?” and “Hello, what is you reason for calling?” were added as responses to channel checks and greetings; Finally, we added generic requests for more information such as “Could you tell me more” and a repertoire of feedback utterances (e.g. “uh-huh”, “ok”, and “yes”) to be used to encourage callers to continue speaking. The wizards were only given 15 minutes to get acquainted with the new layout of the prompt piano before starting to handle real calls from the customer care line. Table 1 contains a typical ISWoz-II dialogue, showing that the wizards frequently used the feedback options that had been added to the prompt piano.

**Table 1.** Labelled dialogue example from ISWoz-II

	<i>Utterance</i>	<i>Label</i>
<b>Wizard</b>	<i>Welcome to Telia how may I help you!</i>	INITIAL HMIHY
<b>Caller</b>	<i>Yes hello?</i>	CHANNEL?
<b>Wizard</b>	<i>Hello!</i>	CHANNEL!
<b>Caller</b>	<i>Oh sorry, the thing is that I got a bill from you...</i>	DESCRIPTION
<b>Wizard</b>	<i>Mm...</i>	FEEDBACK
<b>Caller</b>	<i>Amounting to 800 something...</i>	DESCRIPTION
<b>Wizard</b>	<i>Okay...</i>	FEEDBACK
<b>Caller</b>	<i>But the thing is that I have moved my subscription to X, and my subscription with you should have ended in Jan.</i>	DESCRIPTION
<b>Wizard</b>	<i>Yes...</i>	FEEDBACK
<b>Caller</b>	<i>So I don't understand why I got a bill from you and it is on 800 kr!</i>	DESCRIPTION
<b>Wizard</b>	<i>Is it for your phone at home?</i>	FOLLOW-UP QUESTION
<b>Caller</b>	<i>Yes.</i>	ANSWER
<b>Wizard</b>	<i>Okay, wait while I connect you...</i>	ROUTING

## 6 Effects of Human-Like Dialogue Behaviour

An analysis of ISWOZ-II showed a rate of 94% successful calls; 3% where callers hung up prematurely; and 3% in which the human operator intervened – in all cases either because the caller did not speak Swedish or because they had to inform the caller to call another phone number to get help. This can be compared with 89% successful calls in ISWOZ-I, where 9% hung up during or immediately after the initial open prompt, while 2% lead to intervention.

### 6.1 Effects on Caller Talkativity

Another effect of human-like dialogue strategy is reflected by differences in turn lengths in the three collections (see Fig. 1). For example, whereas the callers’ descriptions were considerably longer in ISWOZ-II than in ISWOZ-I, they were equally long in ISWOZ-II and in the human-human dialogues (INTERVENT). Also, it seems that one to two-word turns were more frequent in ISWOZ-I than in the other datasets. The cumulative distributions of turn length were furthermore very similar for ISWOZ-II and INTERVENT, but quite different from that of ISWOZ-I (see Fig. 2). The fact that one and two word utterances were common in the descriptions, suggest that many speakers use *menu-speak* in ISWOZ-I.

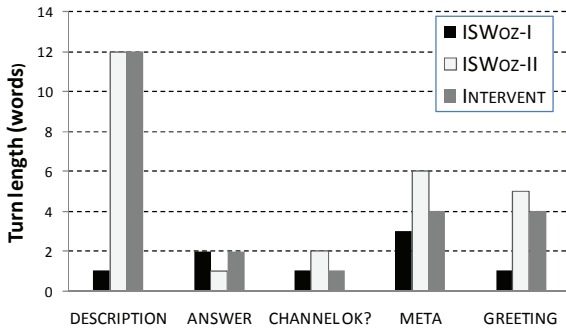


Fig. 1. Median caller turn length (in number of words) for different turn types

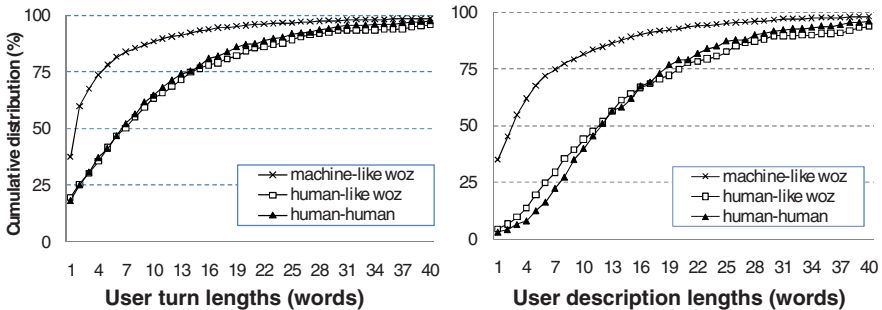
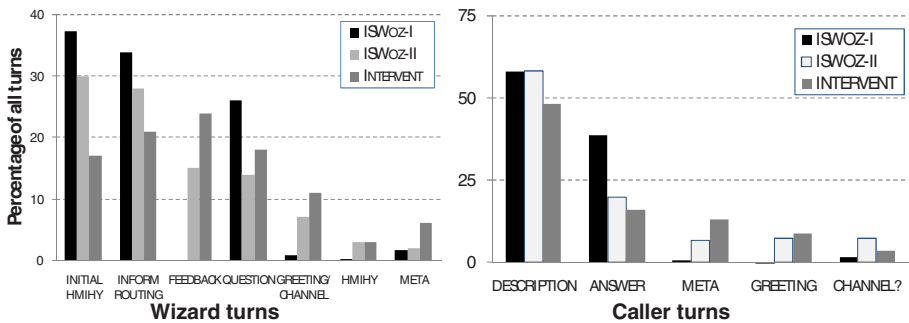


Fig. 2. Cumulative distribution of turn lengths in the two Woz collections and in the human-human intervention dialogues for all turn types and for the description type separately

## 6.2 Effects on Turn Type

The effects of human-like dialogue is also reflected by differences in the distribution of turn types in the corpora (see Figure 3). Figure 3 suggests that the dialogue strategies in ISWOZ-II resulted in dialogues that were more like those in INTERVENT than those in ISWOZ-I. It is worth noting that the amount of FOLLOW-UP QUESTION from the wizard/operator and the corresponding ANSWER from callers were considerably higher in ISWOZ-I than in the other corpora, although the dialogues in these corpora were as successful as the ones in ISWOZ-I. This suggests that the FEEDBACK utterances used in ISWOZ-II and INTERVENT is a successful strategy to elicit the required information. Finally, the dialogue strategies in ISWOZ-II elicited more greetings, channel checks and questions about the nature of the speaker, which the wizards were able to respond to using the new prompts added for this purpose.



**Fig. 3.** Distribution of turn types (%) in the three corpora for the wizard/operator side of the conversations and the caller side

In order to examine in more detail how the wizards handled the calls, the flow of turn types (including both participants) in ISWOZ-II was analysed. Figure 4 shows a dialogue flow chart that covers 90% of all caller/wizard turns in ISWOZ-II. Arrows representing less than 5% of the utterance have been excluded to avoid cluttering. Most callers (65%) described their reason for calling immediately after the initial open prompt. After such descriptions, the wizards did one of three things: if they had enough information they routed the call (34%); if they lacked a certain piece of information they posed a follow-up question (32%); and if the description contained too little information they used feedback like “ok” (28%). The distribution of these choices is approximately an even three-way split. Figure 4 also shows that feedback proved to be a very efficient way of getting callers to provide further descriptions (92%). Responding appropriately to greetings, channel checks and meta questions had a similar effect and also proved efficient for making callers describe their reason for calling.

## 6.3 Effects of Turn-Taking Behaviour

Wizards often refrained from responding when a caller stopped talking, so there is a number of *pauses* within caller turns in the corpora. In the following we compare ISWOZ-II and INTERVENT with respect to such pauses, and any speaker internal silence of more than 200ms is considered a pause. We also discuss the duration of *gaps*, that



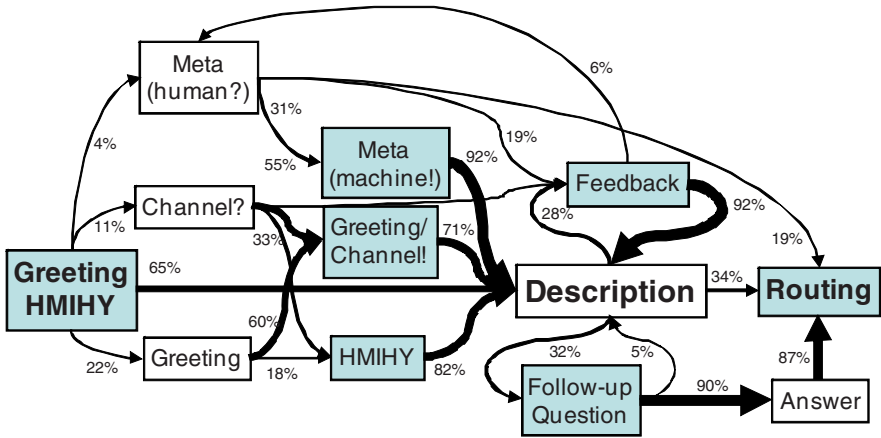


Fig 4. Flow chart of the dialogues in ISWOZ-II. System turns are marked with shaded boxes

is silences at speaker changes. 40% of all caller utterances in ISWOZ-II contain one or more pauses, which can be compared with 25% in INTERVENT. The median gap duration in INTERVENT is about 200ms in both directions, whereas the median pause duration is almost 500ms. In ISWOZ-II, the median gap duration in speaker changes from caller to wizard increases to 950ms, which seemingly influences the callers to some degree, as the median gap duration from wizard to caller is almost 500ms. The median pause duration in ISWOZ-II is almost 700ms. Figure 5 shows the cumulative distribution of pauses and gaps in ISWOZ-II and INTERVENT. Note that in INTERVENT, the distribution of gap duration (right panel) is quite similar for speaker changes in both directions.

Although the distribution of gap durations in caller-wizard speaker changes in ISWOZ-II is different to that in wizard-caller speaker changes and both are slower than the same changes in INTERVENT, the distributions seem to lend support to the notion that gap duration is a feature that interlocutors mimic from each other, as the callers' turn-taking is much slower when speaking to the slower system (Figure 5, right panel). Figure 5 corroborates earlier findings that silence duration thresholds are insufficient to create good turn-taking in dialogue systems, as pauses are as long or longer than gaps [e.g. Edlund et al., 2005].

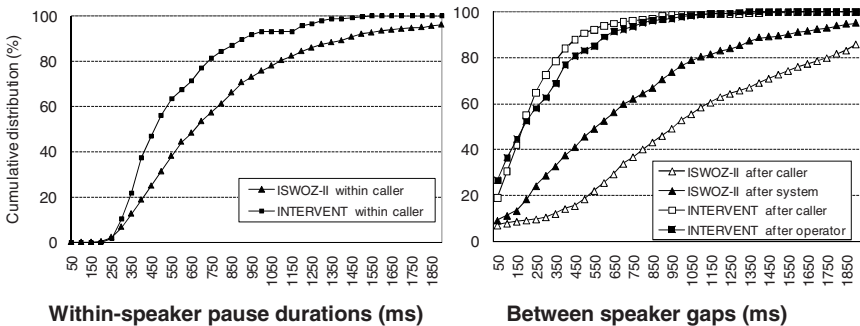


Fig 5. Cumulative distribution of pause durations and gaps in ISWOZ-II and INTERVENT

As we are aiming for a more human-like system in ISWOZ-II, we want that the increase in gap length in speaker shifts should appear at places where one would expect longer gaps from human interlocutors. To test this, we analysed the effect of the discourse context on gaps in INTERVENT and ISWOZ-II. The flow-chart in Figure 4 illustrates the shifts described in the following. Regarding the behaviour of the wizard/operator, both corpora have more long gaps between DESCRIPTION and FOLLOW-UP QUESTION or ROUTING. Conversely, short gaps are more common between DESCRIPTION from the callers and FEEDBACK from the wizard, as well as after channel checks from the caller. The effect on the callers is also similar in the two corpora. In both corpora, the callers have more long gaps after requests for information from the wizard/operator, and before channel checks. In ISWOZ-II callers often made long gaps before inquiring whether it was a human or machine talking. These inquiries were often followed by a very swift reply from the wizard stating that it indeed was a machine, which again was often followed by a long gap. Finally, short gaps are more common between FEEDBACK from wizard/operator and further DESCRIPTION from the caller.

## 7 Discussion

In the data collections discussed here, the wizards did things current call-routing technology generally cannot handle. In both in-service Woz collections, wizards did not base their turn-taking behaviour strictly on silence duration thresholds, like spoken dialogue systems commonly do. In particular, they did not barge into hesitation pauses. Refraining from barging in when speakers hesitate or pause can be achieved in several ways. The problem has been approached using combinations of semantic and dialogue state information, as in [Bell et al., 2001, Nakano et al., 1999, Skantze and Edlund, 2004]. These approaches require that input be processed incrementally, in chunks much smaller than turns, as recognised by Allen et al., who argue that incremental interpretation of user input is necessary for the interaction with spoken dialogue systems to become more natural [Allen et al., 2001]. An alternative or complementary approach that distinguishes pauses (in speech) from gaps between speakers is using prosodic information [Edlund et al., 2005]. Such approaches can avoid violating in-speech pauses, while at the same time making it possible for the system to respond considerably faster when appropriate.

In ISWOZ-II, wizards frequently used the feedback options (e.g. “uh-huh”, “ok”) to encourage the callers to continue their description, and such feedback can indeed be viewed as a fast way of saying “tell me more”. In order to use feedback in the manner of the wizards, the system needs to know when to respond with ROUTE, FOLLOW-UP QUESTION, and FEEDBACK. The contents of the calls collected give a clear impression that the operators utilised prosody, semantics, and pragmatics in making this three-way decision. A preliminary analysis of the content in caller descriptions preceding each of these choices suggests that in excess of 50% of the FEEDBACK choices were preceded by descriptions that can be classified as background information, such as “*I called you before*”. Conversely, less than 5% of the FEEDBACK decisions were preceded by ellipses and other condensed utterances, although these were not uncommon in the material, giving rise to 30% of both ROUTE and FOLLOW-UP QUESTION decisions.

The use of feedback after background information makes sense, since it is not possible for the wizards to route and it would be hard to know which follow up question to choose without more information. This problem faces fully automatic systems as well, and brief feedback utterances could help matters considerably. Deciding when the system should provide brief feedback could be done by other means as well. If the categoriser used for the routing provides confidence measures, one could use feedback responses until a certain confidence is reached or until some time-out is exceeded. Feedback could also be used until the caller has spoken a certain preset number of words, given that we have statistics of how many words it usually takes to get good categorisation results.

We may also note that the whole idea of using brief feedback responses is associated with increasing callers' trust in the system. By creating dialogue that appears more human-like, callers' talkativeness is increased, and we occasionally need callers to talk more than they generally do when faced with a machine. Note, however, that we are not suggesting we build spoken dialogue systems that behave as human operators in every respect. It is sufficient that the system exhibits behaviour that *elicits the same kind of descriptions* as those found in *caller-human operator calls*, or to quote Cassell: "a machine that acts human enough that we respond to it as we respond to another human" [Cassell, 2007]. We have shown similarities between caller behaviour in INTERVENT, the human-human dialogue corpus, and ISWoz-II, the second in-service Woz data collection in which we aimed for more human-like dialogue. If these similarities are anything to judge by, we might say that we at least in part succeeded with this ambition.

## Acknowledgement

This had not been possible if it wasn't for the ASR 90 200 pilot team at TeliaSonera that provided us with the data collection tools, and for the skilled Wizards that showed us how to do human-like dialogue. Finally, Anders Lindström at TeliaSonera also participated in the design and collection of ISWoz-II.

## References

- [Acomb et al., 2007] Acomb, K., Bloom, J., Dayanidhi, K., Hunter, P., Krogh, P., Levin, E., Pieraccini, R.: Technical support dialog systems: Issues, problems, and solutions. In: Proc. Bridging the Gap: Acad. and Ind. Res. in Dialog Tech., Rochester, USA (2007)
- [Allen et al., 2001] Allen, J., Ferguson, G., Stent, A.: An architecture for more realistic conversational systems. In: Proc. IUI 2001, Santa Fe, USA (2001)
- [Allwood and Haglund, 1992] Allwood, J., Haglund, B.: Communicative activity analysis of a Wizard of Oz experiment. Technical Report, Göteborg Univ., Gothenburg (1992)
- [Bell, 2000] Bell, L.: Linguistic adaptations in spoken and multimodal dialogue systems. Licentiate thesis, KTH, Stockholm (2000)
- [Bell et al., 2001] Bell, L., Boye, J., Gustafson, J.: Real-time handling of fragmented utterances. In: Proc. Adaptation in Dialogue Systems, Pittsburgh, USA (2001)

- [Boyce, 1999] Boyce, S.J.: Spoken natural language dialogue systems: User interface issues for the future. In: Gardner-Bonneau, D. (ed.) *Human factors and voice interactive systems*. Kluwer, Boston (1999)
- [Brennan, 1996] Brennan, S.: Lexical entrainment in spontaneous dialog. In: *Proc. ISSD*, Philadelphia, USA (1996)
- [Cassell, 2007] Cassell, J.: *Body language: Lessons from the near-human*. In: *Genesis Redux*. University of Chicago Press, Chicago (2007)
- [Dahlbäck et al., 1993] Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of Oz studies – why and how. In: *Proc. Intelligent User Interfaces* (1993)
- [Edlund et al., 2005] Edlund, J., Heldner, M., Gustafson, J.: Utterance segmentation and turn-taking in spoken dialogue systems. In: *Computer Studies in Language and Speech*, Peter Lang, Frankfurt am Main (2005)
- [Fischer, 2006] Fischer, K.: The role of users' preconceptions in talking to computers and robots. In: *Proc. How People Talk To Computers Robots And Other Artificial Communication Partners*, Delmenhorst, Germany (2006)
- [Garrod and Pickering, 2004] Garrod, S., Pickering, M.J.: Why is conversation so easy? *Trends Cogn. Sci.* 8, 8–11 (2004)
- [Gorin et al., 1997] Gorin, A.L., Riccardi, G., Wright, J.H.: How I help you. *Speech Commun.* 23, 113–127 (1997)
- [Gustafson et al., 1997] Gustafson, J., Larsson, A., Carlson, R., Hellman, K.: How do system questions influence lexical choices in user answers? In: *Proc. Eurospeech*, Rhodes (1997)
- [Martinovsky and Traum, 2003] Martinovsky, B., Traum, D.: The error is the clue: Breakdown in human-machine interaction. In: *Proc. Error Handling in Spoken Dialogue Systems*, Château d'Oex-Vaud, Switzerland (2003)
- [Nakano et al., 1999] Nakano, M., Miyazaki, N., Hirasawa, J., Dohsaka, K., Kawabata, T.: Understanding unsegmented user utterances in real-time spoken dialogue systems. In: *Proc. ACL* (1999)
- [Porzel, 2006] Porzel, R.: How computers (should) talk to humans. In: *Proc. How People Talk To Computers Robots And Other Artificial Communication Partners*, Delmenhorst, Germany (2006)
- [Skantze and Edlund, 2004] Skantze, G., Edlund, J.: Robust interpretation in the Higgins spoken dialogue system. In: *Proc. Robust*, Norwich, UK (2004)
- [Wirén et al., 2007] Wirén, M., Eklund, R., Engberg, F., Westermark, J.: Experiences of an in-service Wizard-of-Oz data collection for the deployment of a call-routing application. In: *Proc. Bridging the Gap: Acad. and Ind. Res. in Dialog Tech.*, Rochester, USA (2007)
- [Wooffitt et al., 1997] Wooffitt, R., Fraser, N.M., Gilber, N., McGlashan, S.: *Humans, computers and wizards*. Routledge, London (1997)
- [Zoltan-Ford, 1991] Zoltan-Ford, E.: How to get people to say and type what computers can understand. *Int. J Man-Mach Stud.* 34, 527–547 (1991)