Finding a Meta-dialogue Strategy for Multi-intent Spoken Dialogue Systems

Jakob Landesberger and Ute Ehrlich
Speech Technology, Daimler AG, Ulm, Germany

Keywords: Spoken Dialogue System, Multi-intent, Multitasking, Conversational Interface, HCI.

Abstract: Speech is an easily accessible and highly intuitive modality of communication for humans. Maybe that is the reason why people have wanted to talk to computers almost from the moment the first computer was invented. Today several consumer-level products developed in the last few years have brought inexpensive voice assistants into everyday use. The problem is that this speech interfaces are mostly designed for certain commands. But during demanding tasks like driving a car, it can be useful to talk about several things at once, to get back to the main task as fast as possible. While talking about different things in a single utterance it is important to give the user adequate feedback, like a meta-dialogue informing about which topic is discussed at the moment. In this paper we compare several meta-dialogue approaches for a speech dialogue system capable of handling multi-intents. The aim of our study is to reveal which strategies users prefer regarding metrics such as flexibility, joy, and consistency. Our results show that explaining topic transitions and topic introductions via speech receive a high user rating and is cognitively less demanding then visual cues.

1 INTRODUCTION

At least since Apple released its assistant Siri, Conversational Interfaces became a part of our everyday life. Siri offers its users the possibility to execute a huge range of functions and get information with input in natural language. Generally, such Conversational Interfaces have a big advantage as speech is an intuitive modality of communication for human (Lemon et al., 2002).

Conversational Interfaces do not only ease the use of smartphones, they can also increase safety. If integrated in cars they allow the driver to interact with the car without having to look away from the road. In such situations with a demanding task like operating a car, people tend to communicate in an efficient and economical way. To get back as fast as possible to the more demanding task people often talk about several things at once in one utterance. While utterances can contain multiple intents simultaneously, such as answering a question and providing feedback about the understanding of the question, intents can also be aligned sequentially (Bunt, 2011). E.g.: “Take the fastest way to work and please call my brother”. These utterances are called multi-intents (MIs). MIs do not only occur while people do several things at once, they are quite frequently used in a conversation. The ability of humans to easily process such MI statements and to react accordingly, allows for effective dialogue. MIs are often used to add topics to an ongoing dialogue e.g. in an over-answering scenario. Over-answering means to generate extended responses that provide more specific or additional information. This is a quite useful mechanism to make a dialogue swifter and more efficient. (Wahlster et al., 1983).

In order for a system to process, understand and react appropriately to MI statements, the first crucial step is to detect MIs and to distinguish the individual intentions. Some methods for MI detection exists, such as Xu and Sarikaya (2013) which approach the detection problem with MIs as a classification problem and use multi-label learning. Kim et al. (2017) provide a two stage method to detect MIs with only single-intent labelled training data. The MI data was fabricated by concatenating combinations of single intent sentences. Due to the general lack of MI data Beaver et al. (2017) propose a commercial customer service speech corpus containing MIs.

The above mentioned researched MI interactions are normally single-turn inputs which are dealt with independently. Intents where follow-up questions or further clarification is needed, are hardly considered. Interactions which span multiple turns of a dialogue

Landesberger, J. and Ehrlich, U.
Finding a Meta-dialogue Strategy for Multi-intent Spoken Dialogue Systems.
DOI: 10.5220/0008348701710176
In *Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications (CHIRA 2019)*, pages 171-176
ISBN: 978-989-758-376-6

are of course a part of natural conversation. Multiple turns are a requirement for essential communication aspects like receiving more details to generate a more precise answer, for the sake of grounding or resolving miscommunication (Clark and Brennan, 1991).

Miscommunication is often divided into misunderstanding and non-understanding. Misunderstanding means that one participant obtains an interpretation that he or she considers correct, but is not in line with the other speaker's intentions (Skantze, 2003). Resolving misunderstandings can be complicated and confusing (Georgiladakis et al., 2016). Therefore making sure both dialogue partners know they talk about the same topic is important.

Meta-dialogue which exceeds topic related content, e.g. to inform about a topic-switch or which topic is discussed first, is a way to assure this mutual understanding.

When people use MIs to talk about multiple things at once, meta-dialogue seems to be a necessity to establish and maintain the required amount of understanding.

We want to investigate which meta-dialogue strategy shows the best result in a real-life speech dialogue system capable of handling MIs. To create a scenario where meta-dialogue is especially important we made sure the dialogue-topics require further clarification and misunderstandings have to be solved every now and then. To investigate the question which meta-dialogue strategy is perceived best, we implemented a MI, multi-domain Wizard of Oz study in a driving simulator with periodically forced misunderstandings.

2 EXPERIMENT SETUP

In existing dialogue systems which consider meta-dialogue strategies to introduce a particular topic, mostly topic interruption strategies were evaluated. Heinroth et al. (2012) looked at four different topic switching strategies. The explanation strategy explains what topic is about to be started, showed high scores regarding efficiency and user-friendless and supports the user to memorise the topics. Non-verbal behaviour, like visual cues plays an important role in topic switching, too (Sidner et al., 2005).

In order to find out what strategies a dialogue system should use to handle MIs we follow Glas and Pelachaud (2015) by testing a set of potential meta-dialogue strategies with respect to their effects on the user perception of the dialogue and the dialogue system. The tested system does not only allow MIs from the user, it uses MI-answers if convenient, too, e.g. if multiple topics need further clarification.

We distinguish three meta-dialogue strategies:

- NMD (no meta-dialogue): Only topic related answers and no meta-dialogue at all.
  "... I'll remind you directly before you have to take over. Should I raise the temperature ...?"
- SMD (spoken meta-dialogue) Explains topic transitions and topic introductions via speech.
  "... I'll remind you directly before you have to take over. And regarding the cold: Should I raise the temperature ...?"
- VMD (visual meta-dialogue) Topic transitions and topic introductions are explained with a graphical user interface.
  "... I'll remind you directly before you have to take over. Should I raise the temperature ...?" (Construction site symbol moves away; Temperature topic symbol moves in (see Figure 1))

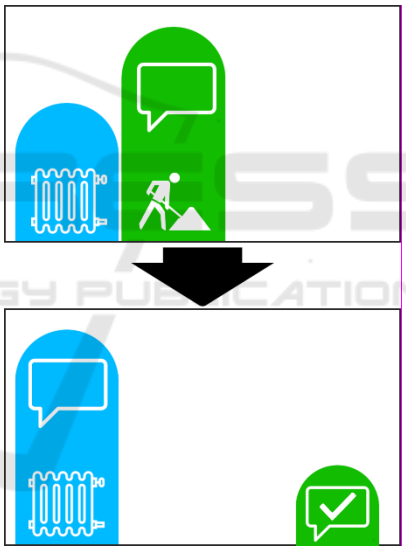

Figure 1: Visual meta-dialogue representation during the system prompt: *"... I'll remind you directly before you have to take over. Should I raise the temperature ...?"*.

The experiment described in this work is a Wizard of Oz study. The investigator simulates the behaviour of a speech dialogue system (SDS), whereas the subjects believe to interact with a real system. This procedure allows to conduct user studies before developing a real system. Analysing the user utterances provides a detailed view of how a user interacts with a SDS. This data will be necessary for the development of a real user centred system (Dahlbäck et al., 1993, Fraser and Gilbert, 1991).

3 USER EXPERIMENT

Each participant of the user study conducted six dialogues with the simulated SDS of an autonomous car. To keep the cognitive load and the distraction from the dialogue low there was no driving task. To keep the study controllable the system tries to clarify the user's need by asking closed questions. While the system was uttering a question, a picture regularly appeared on the screen in front of the participant. This picture represented one out of four user conditions likely to occur during a car ride such as the driver feels cold. Participants were instructed to answer the question and to respond to the shown picture in one turn.

User (U): "I can take over but I feel really cold!"

If both topics required further clarification or proper feedback the system used a MI statement, too.

System (S): "Ok. I'll remind you directly before you have to take over. And regarding the Cold: Should I activate the seat heating or raise the temperature?"

During three out of six dialogues a misunderstanding was simulated. The misunderstanding occurred always after the participant used a MI utterance.

S: "Do you want to postpone or cancel your appointment?"
U: "Please postpone the appointment and I have to visit the restrooms."
S: "I will cancel your appointment. And regarding the Restrooms: There is ..."

The participants received instructions beforehand to correct possible errors and no matter which way they chose, the wizard ensured that resolving the misunderstanding was successful.

After having successfully finished the dialogues the participants had to fill in a questionnaire. We adopted SASSI (Hone and Graham, 2000), a standardized method to assess whole dialogues. We tailored the comprehensive questionnaire to 12 questions related to the following quality factors:

- Flexibility (FLEX) – Is the system flexible enough to meet different requirements und is it able to react quickly to changes?
- Joy (JOY) – Is there any fun and joy while using the system?
- Irritation (IRRT) – Does the interaction with the system lead to confusion or are there comprehension problems?
- Consistency (CONS) – Does the system behave in a consistent way and does it allow the user to realise the discussed topic at any time?

In order to derive a valid estimation a semi-structured interview took place after the questionnaire. The participants were asked to answer several questions regarding understanding, perceived differences between the meta-dialogue strategies and if there were any difficulties using multi-intents.

4 RESULTS

We analysed data from 35 participants (15f/20m), with average age of 25.08 (SD: 4.2). Their experience with SDS is mediocre (6-Likert scale, avg: 3.17, SD: 1.23) as well as the usage of SDSs (5-Likert scale, avg: 2.24, SD: 1.22). In total, we built a corpus of interactions with 5h and 33min of spoken dialogues. It contains 1454 user utterances with 364 MI statements.

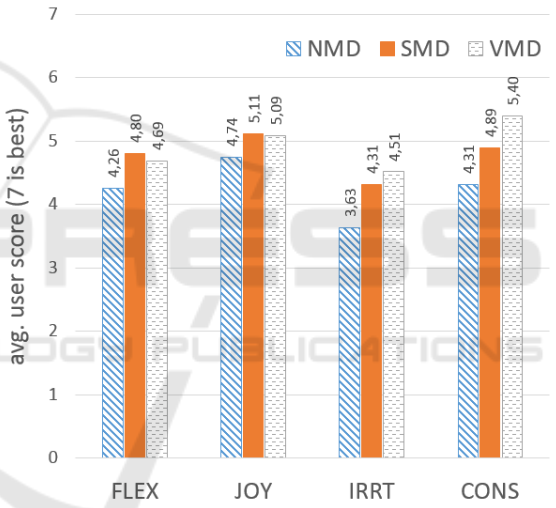

Figure 2: Average user score for each meta-dialogue strategy regarding the four tested categories of all participants.

Figure 2 shows the subjective results of the questionnaire. Overall the SMD and VMD strategies performed best regarding all categories (NMD vs. SMD p=.010; NMD vs. VMD p=.003). In contrast, NMD performed poorly in the IRRT category and worse than the other strategies in the remaining categories. The VMD strategy is assessed as being the least irritating and most consistent one but there is no overall significant difference to the SMD strategy. SMD received the best ratings for flexibility and joy.

Figure 3 shows the results of the questions how much cognitive effort was needed during the dialogue. Strategy NMD received the best rating and VMD the worst. While the differences between the

strategies are not significant, all strategies performed over all poorly.

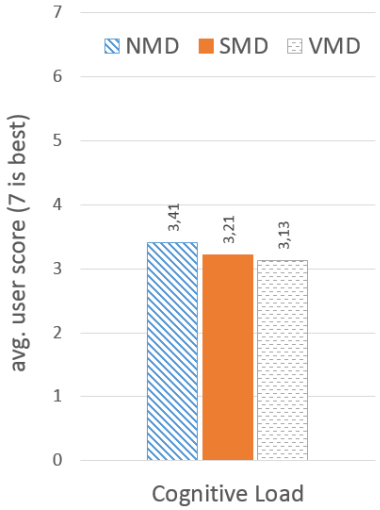

Figure 3: Average user score of all participants for each meta-dialogue strategy regarding the question how much cognitive effort was needed during the dialogue. (A higher score means less cognitive resources were needed.)

The high cognitive load during the VMD strategy is additionally illustrated by the answers to the interview question: "Does the graphic representation require too much attention?" 44% of the participants confirmed the statement, 48% did not want to commit themselves, 8% did not make a statement and not a single one rejected the question.

Another indicator for the overall high cognitive load are the answers to the question, if the participants can point out the difference between meta-dialogue strategy NMD and SMD. 67% did not notice any difference or said things like the system voice has changed – which is not true. 7% could not give an answer at all and only 26% noticed the spoken meta-dialogue in strategy SMD.

Figure 4 shows the subjective results of the questionnaire for the group of participants which recognized the difference between strategy NMD and SMD. Of particular note is the considerable worse result of strategy NMD regarding all tested categories (NMD vs. SMD p=.030; NMD vs. VMD p=.038).

The participants who did not recognize the difference, rated NDM overall higher, but still worse than strategy SMD or VMD. Figure 5 shows the results of this group.

We analysed the utterances used to resolve the misunderstanding if another MI was used or if one of the dialogue topics was dropped to focus on the misunderstood part.

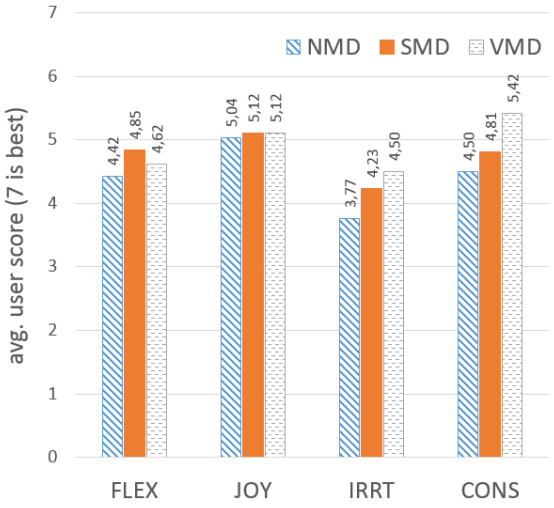

Figure 4: Average user score for each meta-dialogue strategy regarding the four tested categories of all participants who did recognize the difference between strategy NMD and SMD.

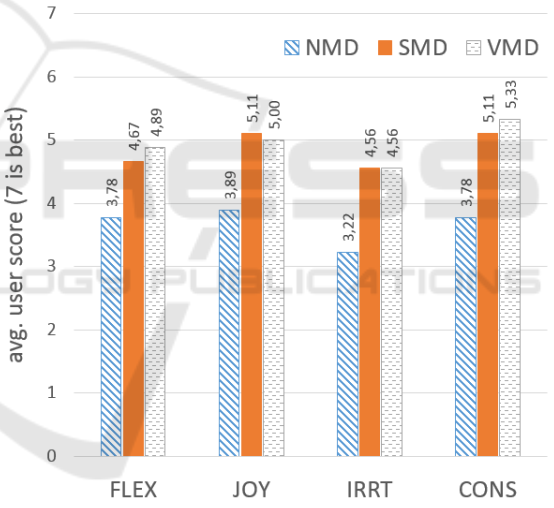

Figure 5: Average user score for each meta-dialogue strategy regarding the four tested categories of all participants who did not recognize the difference between strategy NMD and SMD.

33% of the recognized misunderstandings were solved by handling only the error. Nearly two-thirds (62%) interrupted the system at the moment the failure was realized.

S: Ok. We won't do any more refuelling stops and regarding the heat: Should ..."

U: "Stop! [Interrupting the system] I wanted to refuel."

The other participants (38%) did not interrupt the system, listened to the whole prompt but decided

afterwards to ignore the correct part and focus on the misunderstanding.

67% of the utterances after a misunderstanding included two intents. One regarding the misunderstanding and the other one an answer to the question. During the interview the participants confirmed that using a MI is nothing uncommon. The main issue, mentioned by 23% of the participants was, that they were not used talking to a system in such a natural way. A common statement included: *"I wouldn't have used Multi-Intents because I don't think a system can handle that."*

5 CONCLUSION

If there is a complicated task like resolving a misunderstanding users should be able to drop all the other topics and focus on the important task if needed. The need for such a mechanism was mentioned during the interview by 61% of the participants. If the user interrupts the system and drops a topic the system should allow the interruption and proactively raise the dropped topic again after the important task is finished. Another possibility which was mentioned during the interview is, that the system choses one topic to focus on and postpone the other one. But the decision has to be logical and comprehensible.

Users have no difficulties using MIs while talking to a simulated SDS. They even used MIs to solve misunderstandings and talk about other things in one turn. To maintain a consistent dialogue flow, an adequate meta-dialogue is a useful mechanism. The results presented in this work emphasise this statement and also indicate that a spoken meta-dialogue explaining what topic is about to be started, is preferred.

A visual realization of this additional information achieved a similar good rating, but is cognitively more demanding. In addition, a graphical representation for meta-dialogue is impracticable in scenarios with a visually demanding main task like driving a car.

6 FUTURE WORK

Despite the usefulness of MIs it seems that if the system uses MIs, too, to add topics or to try to clarify multiple topics at once, the whole dialogue becomes cognitive very demanding. In order to reduce the general high cognitive load, a system-side prioritisation of one intent could be useful, if the prioritisation is logical and comprehensible for the user. In future research, we will take a further look into prioritising one intent if multiple intents occur in a single utterance. We will investigate which parameters can be used to prioritize a topic and how the dialog and meta-dialog should be designed to enable intuitive and easily accessible communication for the user.

REFERENCES

Lemon, O. et al. (2002): Multi-tasking and collaborative activities in dialogue systems. In Proceedings of the Third SIGdial Workshop on Discourse and Dialogue.

Bunt, H. (2011). Multifunctionality in dialogue. Computer Speech & Language, 25(2), pp. 222-245.

Wahlster, W. et al. (1983): Over-answering yes-no questions: Extended responses in a NL interface to a vision system. In Proceedings of the Eighth international joint conference on Artificial intelligence, pp. 643-646.

Xu, P., & Sarikaya, R. (2013): Exploiting shared information for multi-intent natural language sentence classification. In Interspeech, pp. 3785-3789.

Kim, B., Ryu, S., & Lee, G. G. (2017): Two-stage multi-intent detection for spoken language understanding. Multimedia Tools and Applications, 76(9), pp. 11377-11390.

Beaver, I., Freeman, C., & Mueen, A. (2017): An Annotated Corpus of Relational Strategies in Customer Service. arXiv:1708.05449.

Clark, H. H., & Brennan, S. E. (1991): Grounding in communication. Perspectives on socially shared cognition, 13, pp. 127-149.

Skantze, G. (2003): Exploring human error handling strategies: Implications for spoken dialogue systems. In ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems.

Georgiladakis, S. et al. (2016): Root Cause Analysis of Miscommunication Hotspots in Spoken Dialogue Systems. In INTERSPEECH, pp. 1156-1160.

Heinroth, T., Koleva, S., & Minker, W. (2011). Topic switching strategies for spoken dialogue systems. In Twelfth Annual Conference of the International Speech Communication Association.

Sidner, C. et al. (2005): Explorations in engagement for humans and robots. Artificial Intelligence, 166(1-2), pp. 140-164.

Glas, N., & Pelachaud, C. (2015): Topic Transition Strategies for an Information-Giving Agent. In European Workshop on Natural Language Generation (ENLG), pp. 146-155.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993): Wizard of Oz studies—why and how. Knowledge-based systems, 6(4), pp. 258-266.

Fraser, N. M., & Gilbert, G. N. (1991): Simulating speech systems. Computer Speech & Language, 5(1), pp. 81-99.

Hone, K. S., & Graham, R. (2000): Towards a tool for the subjective assessment of speech system interfaces (SASSI). Natural Language Engineering, 6(3-4), pp. 287-303.