# Investigating Strategies for Resolving Misunderstood Utterances with Multiple Intents

**Jakob Landesberger**
jakob.landesberger@daimler.com

**Ute Ehrlich**
ute.ehrlich@daimler.com

**Abstract**

To investigate how participants resolve misunderstood utterances, which contain more than one intent, we conducted a wizard-of-oz study, simulating a speech dialog system capable of handling multiple intents in one utterance with periodically simulated misunderstandings. Next to the strategy ignoring everything despite the misunderstanding, we found that two third of the participants resolved the error and answered a system question in one turn.

## 1 Introduction

Humans tend to structure their communication in an efficient, economic way (Lemon et al., 2002). Especially in situations when they have to fulfill also other tasks such as in a driving situation. This means that they often speak about different things in one utterance (called multi intents (MIs)) to get back as fast as possible to the more demanding driving task, e.g. *"Take the normal way to work and I wanna call my wife"*. 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 like in the provided example (Bunt, 2011). Communication problems will arise if the system summarization of the utterance contains a misunderstanding. Humans have different strategies to cope with such a problem. The aim of this paper is to find these error correcting strategies for partly misunderstood MI utterances. Therefore, we implemented a MI wizard-of-oz study with periodically simulated misunderstandings.

## 2 Topics and Experiment Design

Each participant of the user study conducted six dialogues with the speech dialogue system (SDS) of an autonomous car. 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. During three out of six dialogues a misunderstanding was simulated. The misunderstanding occurred always after the participant used a MI utterance. It only concerned the user answer, not the additional intent which was triggered by the picture. The participants received instructions to correct possible errors, and no matter which strategy they chose, the wizard ensured that resolving the misunderstanding was successful.

## 3 Correcting Misunderstandings

We distinguish between two main strategies which participants used to correct the simulated misunderstanding. In the first strategy (called MI correction (MIC)) the participant uses at least two sequential aligned intents in one utterance: one to fix the misunderstanding and one to respond to the system question. The sequence of these intents can also be switched: the participant choses to respond first to the system question and after that resolves the misunderstanding. Doing so he changes the sequence of topics used by the system. This pattern is labeled as topic sequence change (TSC) (see Table 1).

| **Classification** | **Example utterance** |
|---|---|
| MIC \| TSC \| $^{c}$Correction only | "Please activate the air condition. I want to refuel in Austria$^{c}$." |
| MIC \| TSC \| $^{r}$Rejection and $^{c}$Correction | "Please activate the air condition but I still$^{r}$ want to refuel in Austria$^{c}$." |
| SIC \| BI \| $^{r}$Rejection only | [System still speaking] "Correction$^{r}$!" |
| SIC \| $^{c}$Correction only | "I would like to refuel$^{c}$." |

Table 1: Classified examples of participants' responses to the system misunderstanding: "*Ok. We won't do any more refuelling stops and regarding the heat: Should I turn on the air condition or activate the seat ventilation?*" Corrections and Rejections are marked with $^{c}$ and $^{r}$.

The second strategy (called single intent correction (SIC)) occurs if the participant focuses only on fixing the misunderstanding and ignores system question, or interrupts the system while the sentence containing the misunderstanding is uttered. The interruption of the system is called barge-in (BI) (see Table 1). Furthermore, we analyzed if the participant only rejected or corrected the misunderstanding or did both. Rejection means the utterance does point out the wrong part of the utterance, but requires further clarification: *"No, I don't want to cancel my appointment."* If only the correction is realized, it can be difficult to detect miscommunication at all: *"I want to postpone my appointment."*

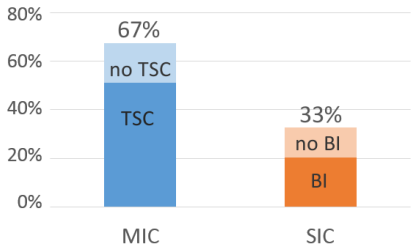


Figure 1: Overview of the distribution of the usage of correction strategies.

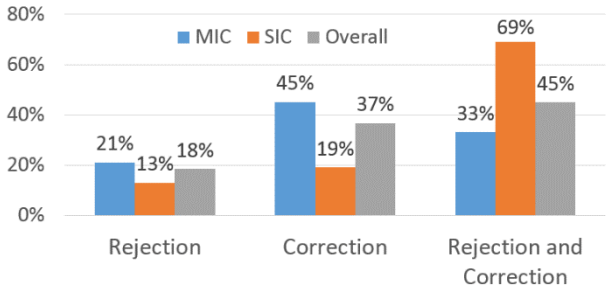


Figure 2: Distribution of the usage of rejections and / or corrections in MIC and SIC utterances.

## 4 Results

We analyzed data from 39 participants (15f/24m), with average age of 25.08 (SD: 4.2). Their experience with SDS range in the middle (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 33min of spoken German dialogues. It contains 1454 user utterances with 364 MI utterances.

Figure 1 shows the distribution of all classified utterances which were used to correct misunderstandings. In 67% of the correction utterances the misunderstanding is resolved and also an answer to the system question is provided. Most of them (76%) were labeled as TSC because the utterances contained first and foremost the answer to the system question and secondly the correction.

33% of the recognized misunderstandings were solved by handling only the error, according to the SIC strategy. Nearly two thirds (62%) interrupted the system at the moment the failure was realized. The other SIC utterances (38%) were uttered by participants who did not interrupt the system, listening to the whole prompt and decided afterwards to ignore the correct part.

Figure 2 shows the distribution of the usage of rejections and / or corrections. When considering MIC utterances most of them (45%) included only the correction whereas SIC utterances contained mainly both rejection and correction (69%). Overall a preference to give clear hints when miscommunication happens and correct the wrong utterance was observed (45%).

## 5 Conclusion

In a situation where users have to resolve a misunderstanding and answer a question, most of them do both in one turn. They mostly concentrate first on the question and focus the misunderstanding afterwards. If only the misunderstanding is addressed, they interrupt the system or ignore the additional question. Therefore, when developing a user-centred MI SDS it is necessary not only to consider the different strategies used but also variations like changing topic sequences or dropping topics. Additionally, users tend to express only the correction when using a MI utterance and give no obvious clues about the occurrence of a misunderstanding in the first place. Due to this reason it can be problematic to detect the miscommunication at all. It also seems, that if error recovery works properly, user do not hesitate to use multiple intents to get things done in one turn.

## Reference

O. Lemon, A. Gruenstein, A. Battle, and S. Peters. 2002. Multi-tasking and collaborative activities in dialogue systems. In Proceedings of the 3rd SIGdial workshop on Discourse and dialogue, pages 113–124. ACL.

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