# Towards a Speaking Style-Adaptive Assistant for Task-Oriented Applications

**Maria Schmidt & Patricia Braunger**
**Daimler AG, University of Ulm**

While research on Spoken Dialog Systems (SDSs) has developed from using rule-based methods to employing learning algorithms (e.g., Reinforcement Learning/POMDPs, Neural Nets), the accessibility of voice assistants has grown simultaneously as these have become increasingly available on mobile devices. As these assistants are commonly known and used, both the perception as well as the expectations towards these changed among users. The developers shaped their assistants by giving them a personality that is conveyed by different characteristics. These include giving clever answers to questions or using backchannels in spoken responses (Müller et al. 2015). Such human-like behavior increases the attractiveness of voice assistants and the expectations on them.

Another means to become more human-like includes adapting to the interlocutor. Apart from dialog strategy and information content the system could adapt its speaking style to the users, cf. Schmitt & Minker 2012, Schmidt et al. 2018. Consequently, when designing an adaptive voice assistant, one has to decide on two things: First, which of the linguistic features in the system's output should be modeled user-specifically. Second, which of the features in the user input the system should react to. We call the latter **user input features** and the first **system output features**. More precisely, user input features include more or less static *user properties* such as age, gender, and level of experience with voice assistants, and the linguistics in the user's *speaking style*.

Complementarily, system output features are linguistic features the system is able to vary, e.g., politeness or length of voice output. That is, the system produces rather short or long utterances or more or less polite ones depending on the user's preferences.

In our work, we aim to identify potential triggers on the user's side for an adapted system speaking style. We divide this work into three parts:

1. Which linguistic cues do we see in the users' speaking style?
2. Which linguistic features in the system's speaking style are relevant to be implemented adaptively?
3. Which linguistic cues in the users' speaking style indicate how the system's output features should be realized adaptively?

For all three parts we have a closer look at these linguistic features:

a. Politeness
b. Utterance length
c. Vocabulary/Terminology
d. Casualness
e. Style of addressing

Our analysis is based on data collected during two previously conducted user studies.

First, we investigate which linguistic cues appear in the users' speaking style. As an example, we want to find out the way users verbalize politeness or how the vocabulary they use differs. The analysis is based on data from a study in which users had to freely speak to an in-car spoken dialog system within a Wizard of Oz experiment. The study is described in more detail in Braunger et al. 2017. The collected user utterances are examined in terms of the aforementioned linguistic features such as the style of addressing the system. In order to identify the linguistic cues that are associated with the respective

feature we rely on the findings from literature and on measures commonly used in literature. Additionally, we compare conversation-initial and non-initial user utterances.
Second, we analyze the results of an online study described in Schmidt et al. 2018. The study subjects were asked for their opinions on different system's speaking styles and on potential system output features. From the results we conclude whether the specific feature should be handled adaptively in the system's voice output or never/always occur in its utterances.
Third, we aim to identify on which triggers in the user's input the desired system's output depend and the challenges that arise by trying to extract those.

## References

Braunger, P., Maier, W., Wessling, J., & Werner, S. (2017). Natural Language Input for In-Car Spoken Dialog Systems: How Natural is Natural?. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*.

Müller, M., Leuschner, D., Briem, L., Schmidt, M., Kilgour, K., Stüker, S., & Waibel, A. (2015). Using Neural Networks for Data-Driven Backchannel Prediction: A Survey on Input Features and Training Techniques. In *International Conference on Human-Computer Interaction*. Springer, Cham.

Schmidt, M., Braunger, P., Hamidi, R., & Stier, D. (to appear 2018). Towards Desired Output Features of an Adaptive Spoken Dialog System - A Preliminary Study.

Schmitt, A., & Minker, W. (2012). *Towards adaptive spoken dialog systems*. Springer Science & Business Media.