Generation of Output Style Variation in the SAMMIE Dialogue System

Ivana Kruijff-Korbayová, Ciprian Gerstenberger
Olga Kukina
Saarland University, Germany
{korbay|gerstenb|olgak}@coli.uni-sb.de

Jan Schehl
DFKI, Germany
jan.schehl@dfki.de

Abstract

A dialogue system can present itself and/or address the user as an active agent by means of linguistic constructions in personal style, or suppress agentivity by using impersonal style. We describe how we generate and control personal and impersonal style variation in the output of SAMMIE, a multimodal in-car dialogue system for an MP3 player. We carried out an experiment to compare subjective evaluation judgments and input style alignment behavior of users interacting with versions of the system generating output in personal vs. impersonal style. Although our results are consistent with earlier findings obtained with simulated systems, the effects are weaker.

1 Introduction

One of the goals in developing dialogue systems that users find appealing and natural is to endow the systems with contextually appropriate output. This encompasses a broad range of research issues. Our present contribution concerns the generation of personal and impersonal style.

We define the personal/impersonal style dichotomy as reflecting primarily a distinction with respect to *agentivity*: personal style involves the explicit realization of an agent, whereas impersonal style avoids it. In the simplest way this is manifested by the presence of explicit reference to the dialogue participants (typically by means of personal pronouns) vs. its absence, respectively. More generally, active voice and finite verb forms are typical for personal style, whereas impersonal style often, though not exclusively, employs passive constructions or infinite verb forms:

(1) Typical personal style constructions:
   a. I found 20 albums.
   b. You have 20 albums.
   c. Please search for albums by The Beatles.

(2) Typical impersonal style constructions:
   a. 20 albums have been found.
   b. There are 20 albums.
   c. The database contains 20 albums.
   d. 20 albums found.

The dialogue system SAMMIE developed in the TALK project uses either personal or impersonal output style, employing constructions such as (1a–1c) and (2a–2d), respectively, to manifest its own and the user's agentivity linguistically. We ran an experiment to assess the effects of the system output style on users' judgments of the system's usability and performance and on their input formulation.

In Section 2 we review related work on system output adaptation and previous experiments concerning the effect of system output style on users' judgments and style. We describe the SAMMIE system and the generation of style variation in Section 3. In Section 4 we describe our experiment and in Section 5 present the results. In Section 6 we provide a discussion and conclusions.

2 Previous Work

Although recently developed dialogue systems adapt their output to the users in various ways, this

usually concerns content selection rather than surface realization. There is to our knowledge no system that varies the style of its output in the interpersonal dimension as we have done in SAMMIE. Work on animated conversational agents has addressed various issues concerning agents displaying their *personality*, but this usually concerns emotional states and personality traits, rather than the personal/impersonal alteration. (Isard et al., 2006) model personality and alignment in generated dialogues between pairs of agents using OpenCCG and an over-generation and ranking approach, guided by a set of language models. Their approach probably could produce the personal/impersonal style variation as an effect of personality or a side-effect of syntactic alignment.

The question whether a system should generate output in personal or impersonal style has been addressed by (Nass and Brave, 2005): They observe that agents that use "I" are generally perceived more like a person than those that do not. However, systems tend to be more positively rated when consistent with respect to such parameters as personality, gender, ontology (human vs. machine), etc. On the basis of an investigation of a range of user attitudes to their simulated system with a synthetic vs. a recorded voice, they conclude that a recorded voice system is perceived as more human-like and thus entitled to use "I", whereas a synthetic-voice system is not perceived as human enough to use "I" to refer to itself (Nass et al., 2006).

Another question is whether system output style influences users' input formulation, as would be expected due to the phenomenon of *alignment*, which is generally considered a basic principle in natural language dialogue (Garrod and Pickering, 2004).[1]

Experiments targeting human-human conversation show that speakers in spontaneous dialogues tend to express themselves in similar ways at lexical and syntactic levels (e.g., (Hadelich et al., 2004; Garrod and Pickering, 2004). Lexical and syntactic alignment is present in human-computer interaction, too. (Brennan, 1996) suggested that users adopt system's terms to avoid errors, expecting the system to be inflexible. However, recent experiments show that alignment in human-computer interaction is also automatic and its strength is comparable to that in human-human communication (Branigan et al., 2003; Pearson et al., 2006).

Early results concerning users' alignment to system output style in the interpersonal dimension are reported in (Brennan and Ohaeri, 1994): They distinguish three styles: anthropomorphic (the system refers to itself using first person pronouns, like in (1a) above, fluent (complete sentences, but no self-reference) and telegraphic, like (2d). They found no difference in users' perception of the system's intelligence across the different conditions. However, they observed that the anthropomorphic group was more than twice as likely to refer to the computer using the second person pronoun "you" and it used more indirect requests and conventional politeness than the other groups. They conclude that the anthropomorphic style is undesirable for dialogue systems because it encourages more complex user input which is harder to recognize and interpret.

The described experiments used either the Wizard-of-Oz paradigm (Brennan and Ohaeri, 1994) or preprogrammed system output (Branigan et al., 2003; Nass and Brave, 2005) and involved written communication. Such methods allow one to test assumptions about idealized human-computer interaction. Experimenting with the SAMMIE system allows us to test whether similar effects arise in an interaction with an actual dialogue system, which is plagued, among other factors, by speech recognition problems.

3 The SAMMIE System

SAMMIE is a multimodal dialogue system developed in the TALK project with particular emphasis on multimodal turn-planning and natural language generation to support intuitive mixed-initiative interaction.

The SAMMIE system provides a multimodal interface to an in-car MP3 player through speech and haptic input with a BMW iDrive input device, a button which can be turned, pushed down and sideways in four directions. System output is by speech and a graphical display integrated into the car's dashboard. SAMMIE has a German and an English version with the same functionality.

[1] This dialogue phenomenon goes under a variety of terms in the literature, besides alignment, e.g., accommodation, adaptation, convergence, entrainment or shaping (used, e.g., by (Brennan and Ohaeri, 1994)).

The MP3 player application offers a wide range of tasks: The user can control the currently playing song, search and browse by looking for fields in the MP3 database (song, artist, album, etc.), search and select playlists and construct and edit them. A sample interaction is shown below (Becker et al., 2006).

(3) U: Show me the Beatles albums.

S: I have these four Beatles albums. [shows a list of album names]

U: Which songs are on this one? [selects the Red Album]

S: The Red Album contains these songs [shows a list of the songs]

U: Play the third one.

S: [song "From Me To You" plays]

The system puts the user in control of the interaction. Input can be given through any modality and is not restricted to answers to system queries. On the contrary, the user can provide new tasks as well as any information relevant to the current task at any time. This is achieved through modeling the interaction as a collaborative problem solving (CPS) process, modeling the tasks and their progression as *recipes* and a multimodal interpretation that fits any user input into the context of the current task (Blaylock and Allen, 2005). To support dialogue flexibility, we model discourse context, the CPS state and the driver's attention state by an enriched information state (Kruijff-Korbayová et al., 2006a).

3.1 System Architecture

The SAMMIE system architecture follows the classical approach of a pipelined architecture with multimodal fusion and fission modules encapsulating the dialogue manager (Bunt et al., 2005). Figure 1 shows the modules and their interaction: Modality-specific recognizers and analysers provide semantically interpreted input to the multimodal fusion module (interpretation manager in Figure 1), that interprets them in the context of the other modalities and the current dialog context. The dialogue manager decides on the next system move, based on its CPS encoded task model, on the current context and also on the results from calls to the MP3 database. The multimodal fission component then generates the system reaction on a modality-dependent level

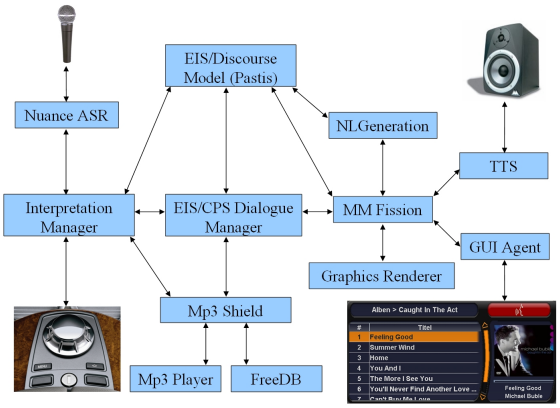

Figure 1: SAMMIE system architecture.

by selecting the content to present, distributing it appropriately over the available output modalities and finally co-ordinating and synchronizing the output. Modality-specific output modules generate spoken output and an update of the graphical display. All modules interact with the extended information state in which all context information is stored.

Many tasks in the SAMMIE system are modeled by a rule-based approach. Discourse modeling, interpretation management, dialogue management, turn planning and linguistic planning are all based on the production rule system PATE (Pfleger, 2004; Kempe, 2004). For speech recognition, we use Nuance. The spoken output is synthesized with the Mary TTS (Schröder and Trouvain, 2003).[2]

3.2 Generation of Natural Language Output with Variation

To generate natural language output in SAMMIE, we developed a template-based generator. It is implemented by a set of sentence planning rules in PATE to build the templates, and a set of XSLT transformations for sentence realization, which yield the output strings. German and English output is produced by accessing different dictionaries in a uniform way. The output is either plain text, if it is to be displayed in the graphical user interface (e.g., captions in tables, written messages to the user) or it is text with mark-up for speech synthesis using the MaryXML format (Schröder and Trouvain, 2003), if it is to be spoken by a speech synthesizer.

[2] http://mary.dfki.de/

The SAMMIE generator can produce alternative realizations for a given content that it receives as input from the turn planner. The implemented range of system output variation involves the following aspects, which have been determined by an analysis of a corpus of dialogues collected in a Wizard-of-Oz experiment using several wizards who were free to formulate their responses to the users (Kruijff-Korbayová et al., 2006b):

1. Personal vs. impersonal style: *Ich habe 3 Lieder gefunden* (*I've found three songs*) vs. *3 Lieder wurden gefunden* (*Three songs have been found*);
2. Telegraphic vs. non-telegraphic style: *23 Alben gefunden (23 albums found)* vs. *Ich habe 23 Alben gefunden (I found 23 albums)*
3. Reduced vs. non-reduced referring expressions: *der Song "Kinder An Die Macht"* (*the song "Kinder An Die Macht"*) vs. *der Song* (*the song*) vs. *"Kinder An Die Macht"* (*"Kinder An Die Macht"*);
4. Lexical choice for (quasi-)synonyms: *Song* vs. *Lied* vs. *Titel* (*song* vs. *track*)
5. Presence vs. absence of adverbs/adverbials: *Ich spiele jetzt den Song* (*I'll now play the song*) vs. *Ich spiele den Song* (*I'll play the song*).

The generation of alternatives is achieved by conditioning the sentence planning and realization decisions. The system can be set either to use one style consistently throughout a dialogue, or to align to the user, i.e., mimic the user's style on a turn-by-turn basis. For the purpose of experimenting with system output variation, the generator supports three sources of control for the available choices: (a) global (default) parameter settings (resulting in no variation); (b) random selection (resulting in random variation); (c) contextual information (resulting in variation based on the dialogue context).

The contextual information used by the generator to control realization includes (i) the grounding status of the content to be communicated (e.g., to decide for vs. against reducing a referring expression); and (ii) linguistic features extracted from the recognized user input (e.g., to make the corresponding syntactic and lexical choices in the output).

3.3 Personal/Impersonal Style Variation

The style variation in SAMMIE amounts to varying between active voice for personal style and passive voice or the "es-gibt" ("there is") construction for impersonal style whenever applicable, as illustrated for several typical dialogue moves below (where (i) always shows the impersonal, and (ii) the personal version).

(4) Search result:[3]
  i. Es gibt 20 Alben.
  There are 20 albums.
  ii. Ich habe 20 Alben gefunden.
  I found 20 albums.
  Sie haben 20 Alben. / Du hast 20 Alben.
  You have 20 albums
  Wir haben 20 Alben.
  We have 20 albums.

(5) Song addition:
  i. Der Titel Bittersweet Symphony wurde zu der Playliste 2 hinzugefügt.
  The track Bittersweet Symphony has been added to Playlist 2.
  ii. Ich habe den Titel Bittersweet Symphony zu der Playliste 2 hinzugefügt.
  I added the track Bittersweer Symphony to Playlist 2.

(6) Song playback:
  i. Der Titel Männer von Herbert Grönemeyer wird gespielt.
  The track Männer by Herbert Grönemeyer is playing.
  ii. Ich spiele den Titel Männer von Herbert Grönemeyer.
  I am playing the track Männer by Herbert Grönemeyer.

(7) Non-understanding:
  i. Das wurde leider nicht verstanden.
  That has unfortunately not been understood.
  ii. Das habe ich leider nicht verstanden.
  I have unfortunately not understood that.

(8) Clarification request:
  i. Welches von diesen acht Liedern?/Welches von diesen acht Liedern wird gewünscht?
  Which of these eight songs? / Which of these eight songs is desired?
  ii. Welches von diesen acht Liedern möchtest du / möchten Sie hören?
  Which of these eight songs would you like to hear?

[3] When referring to the user, personal style has several variants which differ in formality (formal and informal address) and first vs. second person reference.

Figure 2: Experiment setup

The personal/impersonal style variation is not applicable for some dialogue moves, e.g., (9), and for output in telegraphic style.

(9) Song interpreter:
Der Titel Bongo Girl ist von Nena.
The track Bongo Girl is by Nena.

4 Experiment

In order to assess the effects of style manipulation in the SAMMIE system, we ran an experiment in simulated driving conditions, comparing two versions of the system: one consistently using personal and the other impersonal style output.[4] The experiment employed the German version of SAMMIE. The setup (see Figure 2), participants, procedure and collected data are described in detail in (Kruijff-Korbayová and Kukina, 2008), and summarized below.

There were 28 participants, all native speakers of German. We balanced gender and background when assigning them to the style conditions. The experiment followed a fixed script for each participant: welcome, instruction, warm-up driving, 2 trial and 11 experimental tasks, evaluation questionnaire, payment and farewell. The participants were instructed to use mainly spoken input, although they could also use the iDrive button. It took them about 40 minutes to complete all the tasks. The tasks involved exploring the contents of a database of about 25 music albums and were of four types: (1) finding some specified title(s); (2) selecting some title(s) satisfying certain constraints; (3) manipulating the playlists by adding or removing songs and (4) free-use of the system.

The experimental tasks were presented to each participant in randomized order apart from the free use of the system, which was always the last task. The experimenter (E) repeated each task assignment twice to the participant, once in personal and once in impersonal style, as shown in the example below.

(10) E: *Bitte frage das System nach den Liedern von "Pur". Du willst also wissen welche Lieder von "Pur" es gibt.*
E: Please ask the the system about the songs by "Pur". You would like to know which songs by "Pur" there are.

The questionnaire was based on (Nass and Brave, 2005) and (Mutschler et al., 2007). It contained questions with a 6-point scale ranging from 1 (low grade) to 6 (high grade), such as *How do you assess the system in general:* technical (1) – human-like (6); *Communication with the system seemed to you:* boring (1) – exciting (6); *In terms of usability, the system is:* inefficient (1) —efficient(6).

The recorded dialogues have been transcribed, the questionnaire responses tabulated. We manually annotated the participants' utterances (on average 95 per session) with the following features for further analysis:

- Construction type:
  - **Personal** (+/-) Is the utterance a complete sentence in active voice or imperative form
  - **Impersonal** (+/-) Is the utterance expressed by passive voice, infinite verb form (e.g., "Lied abspielen" (*lit.* "song play")), or expletive "es-gibt" ("there-is") construction
  - **Telegraphic** (+/-) Is the utterance expressed by a phrase, e.g., "weiter" ("next")
- Personal pronouns: (+/-) Does the utterance contain a first or second person pronoun
- Politeness marking: (+/-) Does the utterance contain a politeness marker, such as "bitte" ("please"), "danke" ("thanks") and verbs in subjunctive mood (eg. "ich hätte gerne")

[4] For the time being we have not evaluated the version of the system aligning to the user's style.

5 Results

The results concerning users' attitudes and alignment are presented in detail in (Kruijff-Korbayová and Kukina, 2008). Here we summarize the significant findings and provide an additional analysis of the influence of speech recognition problems.

5.1 Style and Users' Attitudes

The first issue addressed in the experiment was whether the users have different judgments of the personal vs. impersonal version of the system. Since the system used a synthetic voice, the judgments were expected to be more positive in the impersonal style condition (Nass and Brave, 2005). Based on factor analysis performed on attitudinal data from the user questionnaires we created the six indices listed below. All indices were meaningful and reliable. (A detailed description of the indices including the contributing factors from the questionnaires can be found in (Kruijff-Korbayová and Kukina, 2008).)

1. General satisfaction with the communication with the system (Cronbach's α=0.86)
2. Easiness of communication with the system (α=0.83)
3. Usability of the system (α=0.76)
4. Clarity of the system's speech (α=0.88)
5. Perceived "humanness" of the system (α=0.69)
6. System's perceived flexibility and creativity (α=0.78)

We did not find any significant influence of system output style on users' attitudes. Only for *perceived humanness of the system* we found a weak tendency in the predicted direction (independent samples test: t(25)=1.64, p=0.06 (one-tailed)), in line with the earlier observation that an interface that refers to itself by a personal pronoun is perceived to be more human-like than one that does not (Nass and Brave, 2005).

5.2 Style and Alignment

The next issue we investigated was whether the users formulated their input differently in the personal vs. impersonal system version. For each dialogue session, we calculated the percentage of utterances containing the feature of interest relative to the total number of user utterances in the session.

In accordance with the expectation based on style alignment in terms of agentivity, we observed a significant difference in the number of personal constructions across style conditions (t(19)=1.8, p=0.05 (one-tailed)). But we did not find a significant difference in the distribution of impersonal constructions. Not surprisingly, there was also no significant difference in the distribution of telegraphic constructions. An unexpected finding was the higher proportion of telegraphic constructions than verb-containing ones within the impersonal style condition (t(13)=3.5, p$<$0.001 (one-tailed)). However, no such difference was found in the personal style condition. Contrary to expectations, we also did not find any significant effect of style-manipulation on the number of personal pronouns, nor on the number of politeness markers.

Since alignment can also be seen as a process of gradual adjustment among dialogue participants over time we compared the proportion of personal, impersonal and telegraphic constructions in the first and second halves of the conversations for both style conditions. The only significant effect we found was a decrease in the number of personal constructions in the second halves of the impersonal style interactions (t(13)=2.5, p=0.02 (one-tailed)).

5.3 Influence of Speech Recognition Problems

Unlike an interaction in a Wizard-of-Oz simulation or similar, an interaction with a real system is bound to suffer from speech recognition problems. Therefore, we made a post-hoc analysis with respect to how much speech recognition difficulty the participants experienced, in terms of the proportion of participant utterances not recognized by the system relative to the total number of participant utterances in a session.

On average, around 33% of participant utterances were not understood by the system.[5] We classified the participants into three groups according to the performance of speech recognition they experienced: the *good* group with less than 27% of input not understood (7 participants); the *poor* group

[5]This is admittedly rather bad performance, nevertheless it mostly does not prevent the participants from getting their tasks successfully completed within a reasonable time, as was shown in an rigorous usability evaluation of the system in normal driving conditions (Mutschler et al., 2007).

Speech Recognition Group		Satisfaction with Communication	Usability of the System	Perceived Flexibility of the System	Clarity of the System's Speech	Perceived Humanness of the System	Ease of Communication
good SR	Mean	3.8889	3.6444	3.2963	4.2222	3.7407	3.0889
	S. D.	.91287	.81104	1.12354	.97183	.54716	1.16237
poor SR	Mean	3.0000	3.0222	2.5556	3.2778	4.0000	2.8667
	S. D.	1.06719	.73106	.79931	1.03414	.66667	.93808

Figure 3: Judgments of the system by the "good" and "poor" speech recognition group

with more than 37% of input not uderstood (7 participants); the *average* group (the remaining 14 participants).

Speech Recognition and Attitudinal Data We suspected that speech recognition problems might be neutralizing a potential influence of style. Therefore we contrasted the judgments on all six factors between the good and the poor speech recognition group (see Figure 3). The "good" speech recognition group showed higher satisfaction with the communication ($t(16)=1.9$, $p=0.04$ (one-tailed)) and evaluated the clarity of the system's speech better ($t(16)= 2.0$, $p=0.03$ (one-tailed)). The good speech recognition group also showed a tendency to assess the usability and flexibility of the system higher than the poor speech recognition group ($t(16)=1.71$, $p=0.05$ and $t(16)=1.61$, $p=0.06$, respectively (marginally significant results)). The two groups did not differ with respect to their judgments of the ease of communication and perceived humanness of the system ($t(16)=0.45$, $p=0.66$ and $t(16)=0.90$, $p=0.38$). These results are not surprising. They confirm that speech recognition does have an effect on the user's perception of the system.

Speech Recognition and Style Alignment We also checked post-hoc whether differences in the experienced speech recognition performance had an influence on the style employed by the participants, again in terms of the proportion of utterances with personal, impersonal and telegraphic constructions, personal pronouns and politeness marking. However, we found no significant effect on the linguistic structure of the participant input across the groups (politeness marking: $F(2)=1.5$, $p=0.24$; all other $Fs<1$ (ANOVA)).

6 Discussion and Conclusions

We presented the generation of personal/impersonal style variation in the SAMMIE multimodal dialogue system, and the results of an experiment evaluating the influence of the system output style on the users' subjective judgments and their formulation of input. Although our results are not conclusive, they point at a range of issues for further research.

Regarding users' attitudes to the system, we found no significant difference among the styles. This is similar to (Brennan and Ohaeri, 1994) who found no difference in intelligence attributed to the system by the users, but it is at odds with the earlier finding that a synthetic voice interface was judged to be more useful when avoiding self-reference by personal pronouns (Nass and Brave, 2005).

Whereas (Brennan and Ohaeri, 1994) used a flight reservation dialogue system, (Nass and Brave, 2005) used a phone-based auction system which read out an introduction and five object descriptions. There are two points to note: First, the subjects heard system output that was a read out continuous text rather than turns in an interaction. This may have reinforced the activation of particular style features. Second, the auction task may have sensibilized the subjects to the distinction between subjective (the system's) vs. objective information presentation, and thus make them more sensitive to whether the system presents itself as an active agent or not.

Regarding the question whether users align their style to that of the system, where previous experiments showed strong effects of alignment (Brennan and Ohaeri, 1994), our experiment shows some effects, but some of the results are conflicting. On the one hand, subjects interacting with the personal style version of the system used more personal constructions than those interacting with the impersonal style version. However, subjects in either condi-

tion did not show any significant difference with respect to the use of impersonal constructions or telegraphic forms. We also found a higher proportion of telegraphic constructions than verb-containing ones within the impersonal style condition, but no such difference in the personal style. Finally, when we considered alignment over time, we found no change in construction use in the personal style, whereas we found a decrease in the use of personal constructions in the impersonal style. It is possible that dividing the interactions into three parts and comparing alignment in the first and the last part might lead to stronger results.

That there is no difference in the use of telegraphic constructions across conditions is not surprising. Being just phrasal sentence fragments, these constructions are neutral with respect to style. But why does there seem to be an alignment effect for personal constructions and not for others? One way of explaining this is that (some of) the constructions that we counted as impersonal are common in both styles. Besides their deliberate use as means to avoid explicit reference to oneself, they also have their normal, neutral usage, and therefore, some of the utterances that we classified as impersonal style may just be neutral formulations, rather than cases of distancing or "de-agentivization". However, we could not test this hypothesis, because we have not found a way to reliably distinguish between neutral and marked, truly impersonal utterances. This is an issue for future work.

The difference between our results concerning alignment and those of (Brennan and Ohaeri, 1994) is not likely to be due to a difference in the degree of interactivity (as with (Nass and Brave, 2005)). We now comment on other differences between our systems, which might have contributed to the differences in results.

One aspect where we differ concerns our distinction between personal and impersonal style, both in the implementation of the SAMMIE system and in the experiment: We include the presence/absence of agentivity not only in the system's reference to itself (akin to (Nass and Brave, 2005) and (Brennan and Ohaeri, 1994)), but also in addressing the user. This concept of the personal/impersonal distinction was inspired by such differences observed in a study of instructional texts in several languages (Kruijff et al., 1999), where the latter dimension is predominant. The present experiment results make it pertinent that more research into the motives behind expressing or suppressing agentivity in both dimensions is needed.

Apart from the linguistic design of the system's output, other factors influence users' behavior and perception of the system, and thus might confound experiment results, e.g., functionality, design, ergonomics, speech synthesis and speech recognition.

A system with synthesized speech should be more positively rated when it does not refer to itself as an active agent by personal pronouns (Nass and Brave, 2005). (Brennan and Ohaeri, 1994) used a system with written interaction, the SAMMIE system employs the MARY text-to-speech synthesis system (Schröder and Trouvain, 2003) with an MBROLA diphone synthesiser, which produces an acceptable though not outstanding output quality. Our post-hoc analysis showed a tendency towards better judgements of the system by the participants experiencing less speech recognition problems. This is as expected. We did not find any statistically significant effect regarding the style-related features we analyzed. A future experiment should address the possibility of an interaction between system style and speech recognition performance as both factors might be influencing the user simultaneously.

One radical difference between our experiment and the earlier ones is that the users of the SAMMIE system are occupied by the driving task, and thus only have a limited cognitive capacity left for the interaction with the system. This may make them less susceptible to the subtleties of style manipulation than would be the case if they were free of other tasks. A possible future experiment could address this issue by including a non-driving condition.

Finally, the SAMMIE system has also the style-alignment mode, where it mimics the user's style on turn-to-turn basis. We plan to present experimental results comparing the alignment-mode with the fixed personal/impersonal style in a future publication.

Acknowledgments

This work was carried out in the TALK project (www.talk-project.org) funded by the EU as project No. IST-507802 within the 6^{th} Framework Program.

References

T. Becker, N. Blaylock, C. Gerstenberger, I. Kruijff-Korbayová, A. Korthauer, M. Pinkal, M. Pitz, P. Poller, and J. Schehl. 2006. Natural and intuitive multimodal dialogue for in-car applications: The SAMMIE system. In *Proceedings of ECAI, PAIS Special section*.

N. Blaylock and J. Allen. 2005. A collaborative problem-solving model of dialogue. In L. Dybkjær and W. Minker, editors, *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pages 200–211, Lisbon, September 2–3.

H. Branigan, M. Pickering, J. Pearson, J. F. McLean, and C. Nass. 2003. Syntactic alignment between computer and people: the role of belief about mental states. In *Proceedings of the Annual Conference of the Cognitive Science Society*.

S. Brennan and J.O. Ohaeri. 1994. Effects of message style on user's attribution toward agents. In *Proceedings of CHI'94 Conference Companion Human Factors in Computing Systems*, pages 281–282. ACM Press.

S. Brennan. 1996. Lexical entrainment in spontaneous dialogue. In *Proceedings of the International Symposium on Spoken Dialogue (ISSD-96)*, pages 41–44.

H. Bunt, M. Kipp, M. Maybury, and W. Wahlster. 2005. Fusion and coordination for multimodal interactive information presentation: Roadmap, architecture, tools, semantics. In O. Stock and M. Zancanaro, editors, *Multimodal Intelligent Information Presentation*, volume 27 of *Text, Speech and Language Technology*, pages 325–340. Kluwer Academic.

S. Garrod and M. Pickering. 2004. Why is conversation so easy? *TRENDS in Cognitive Sciences*, 8.

K. Hadelich, H. Branigan, M. Pickering, and M. Crocker. 2004. Alignment in dialogue: Effects of feedback on lexical overlap within and between participants. In *Proceedings of the AMLaP Conference*. Aix en Provence, France.

Amy Isard, Carsten Brockmann, and Jon Oberlander. 2006. Individuality and alignment in generated dialogues. In *Proceedings of the 4th International Natural Language Generation Conference (INLG-06)*, pages 22–29, Sydney, Australia.

Benjamin Kempe. 2004. PATE a production rule system based on activation and typed feature structure elements. Bachelor Thesis, Saarland University, August.

G.J.M. Kruijff, I. Kruijff-Korbayová, J. Bateman, D. Dochev, N. Gromova, T. Hartley, E. Teich, S. Sharoff, L. Sokolova, and K. Staykova. 1999. Deliverable TEXS2: Specification of elaborated text structures. Technical report, AGILE Project, EU INCO COPERNICUS PL961104.

I. Kruijff-Korbayová and O. Kukina. 2008. The effect of dialogue system output style variation on users' evaluation judgements and input style. In *Proceedings of SigDial'08*, Columbus, Ohio.

I. Kruijff-Korbayová, G. Amores, N. Blaylock, S. Ericsson, G. Pérez, K. Georgila, M. Kaisser, S. Larsson, O. Lemon, P. Manchón, and J. Schehl. 2006a. Deliverable D3.1: Extended information state modeling. Technical report, TALK Project, EU FP6, IST-507802.

Ivana Kruijff-Korbayová, Tilman Becker, Nate Blaylock, Ciprian Gerstenberger, Michael Kaisser, Peter Poller, Verena Rieser, and Jan Schehl. 2006b. The SAMMIE corpus of multimodal dialogues with an MP3 player. In *Proceedings of LREC*, Genova, Italy.

H. Mutschler, F. Steffens, and A. Korthauer. 2007. Deliverable D6.4: Final report on multimodal experiments Part I: Evaluation of the SAMMIE system. Technical report, TALK Project, EU FP6, IST-507802.

C. Nass and S. Brave, 2005. *Should voice interfaces say "I"? Recorded and synthetic voice interfaces' claims to humanity*, chapter 10, pages 113–124. The MIT Press, Cambridge.

C. Nass, S. Brave, and L. Takayama. 2006. Socializing consistency: from technical homogeneity to human epitome. In *P. Zhang & D. Galletta (Eds.), Human-computer interaction in management information systems: Foundations*, pages 373–390. Armonk, NY: M. E. Sharpe.

J. Pearson, J. Hu, H. Branigan, M. J. Pickering, and C. I. Nass. 2006. Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1177–1180, New York, NY, USA. ACM.

N. Pfleger. 2004. Context based multimodal fusion. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 265–272, New York, NY, USA. ACM Press.

M. Schröder and J. Trouvain. 2003. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, 6:365–377.