The following full text is an author’s version which may differ from the publisher’s version.

For additional information about this publication click this link. http://hdl.handle.net/2066/76306

Please be advised that this information was generated on 2017-11-18 and may be subject to change.
Comparing the Usability of a User Driven and a Mixed Initiative Multimodal Dialogue System for Train Timetable Information

Janienke Sturm\textsuperscript{1}, Ilse Bakx\textsuperscript{2}, Bert Cranen\textsuperscript{1}, Jacques Terken\textsuperscript{2}

\textsuperscript{1}Department of Language & Speech
University of Nijmegen, The Netherlands
\textsuperscript{2}Department of User Centered Engineering
University of Eindhoven, The Netherlands

{Janienke.Sturm,B.Cranen}@let.kun.nl, {I.H.M.Bakx,J.M.B.Terken}@tue.nl

Abstract
The aim of the study presented in this paper was to compare the usability of a user driven and a mixed initiative user interface of a multimodal system for train timetable information. The evaluation shows that the effectiveness of the two interfaces does not differ significantly. However, as a result of the absence of spoken prompts and the obligatory use of buttons to provide values, the efficiency of the user driven interface is much higher than the efficiency of the mixed initiative interface. Although the user satisfaction was not significantly higher for the user driven interface, by far most people preferred the user driven interface to the mixed initiative interface.

1. Introduction
The aim of the study described in this paper is to compare the usability of two multimodal train timetable information systems. Both systems present themselves to the user by means of the same graphical interface, but differ in the way speech is used to provide input to the system. One system can be considered user driven whereas the other is in fact a mixed initiative spoken dialogue system enhanced with a GUI.

The research was carried out within the framework of the MATIS project (Multimodal Access to Transaction and Information Services), which aimed at evaluating the usability of multimodal interaction for form-filling applications on small, mobile terminals [1]. To this end, a prototype multimodal form-filling interface was built that provides train timetable information. This interface adopts a mixed initiative dialogue strategy in which both the user and the system could take the initiative. The mixed initiative dialogue has been designed in such a way that it serves both novice and experienced users: the system initiates a spoken dialogue by asking questions, in order to help novice users in completing the form. The dialogue also helps solving errors and provides clarification in order to prevent users from getting stuck. More experienced users, on the other hand, may take over the initiative by interrupting the spoken questions by pressing buttons on a graphical display of the fill-in form. In this way, experienced users can use the system in a more efficient way. An experiment was carried out in order to establish the effect of extended use on the usability of the system. The results of this experiment showed that as their hands-on experience with the system grew users indeed started to use the multimodal system in a more efficient way by interrupting system questions and by using buttons rather than speech to provide values [2].

It may well be, however, that the form-filling application is so easy to use that even novice users do not really need the spoken system guidance [3]. In that case, an interface that is completely user driven and that does not apply a spoken dialogue may be just as effective and efficient for novice users as a mixed initiative system. To investigate this issue, in the current study the mixed initiative interface (MIMI) was compared to a completely user driven version of the same interface (Tap&Talk). This user driven interface was implemented as a tap-and-talk interface: the system does not ask the user any questions, instead the user must indicate which field he or she wants to fill in by pressing buttons on a graphical representation of the form on the screen. The present paper describes and compares the usability of the two interfaces.

2. Methods
2.1. Interfaces
As mentioned earlier, the two interfaces differ in the way the interaction is controlled. Whereas in the MIMI interface a spoken dialogue (which may be interrupted by the user) guides the user through the interaction, in the Tap&Talk interface the interaction is completely controlled by the user who uses the buttons on the screen to indicate which field he or she wants to fill in. Both interfaces show the same graphical representation of the fill-in form on the screen (Figure 1).

![Figure 1: Screen shot of the fill-in form](image-url)
names, times and dates other than today or tomorrow. Other
nine female, between 20 and 71 years of age, with mixed

In both systems speech must be used to fill in the station
times and dates other than today or tomorrow. Other
names were recognized incorrectly.
values can be filled in by direct manipulation, using the
alternative stations in case the station
button is a form of implicit verification. (For detailed
recognized incorrectly.
switch between the start of

Finally, pressing the *Search* button forces the system to query
fields that need to be filled, the status of the dialogue
practice scenario and three test scenarios with each of the two

For the Tap&Talk interface the figures do not add up to 51

1 For the Tap&Talk interface the figures do not add up to 51
(17 times 3), because due to technical errors one subject could
not complete the first scenario and another subject could not
complete the third scenario.
Table 1 shows that overall the effectiveness of both interfaces is rather high and slightly higher for the MIMI system (96%) than for the Tap&Talk interface (88%). A Wilcoxon test showed that this difference is not significant (z = .63; N.S.). All unsuccessful dialogues were caused by persistent recognition errors, after which the subject hung up.

Table 1 shows that only for Scenario 3 the Tap&Talk interface is less effective than the MIMI interface. As this was the scenario with the most confusable station names, this seems to indicate that solving recognition errors is easier with the MIMI interface than with the Tap&Talk interface. This is probably caused by an extra feature in the MIMI interface that facilitates error correction: if a user has explicitly denied a certain station name, this station name will not be recognized in subsequent attempts to fill in a value for this field. In the Tap&Talk interface, as there is no explicit verification in this interface, the same station name may be recognized over and over again.

No significant effect was found of the order in which the two systems were tested.

### 3.2. Efficiency

The efficiency of the interfaces is measured as time to completion (of the successfully completed dialogues). For each scenario the mean duration of a dialogue was calculated in seconds measured from the start of the first user utterance to the query to the information database. Using time to completion as a performance measure in a user-driven interface is not trivial: people may take some extra time to figure out what the next step should be, when they are not being rushed by a spoken dialogue. We tried to diminish this effect by telling subjects that they were paying for this service. Nevertheless, some care should be taken in comparing the efficiency figures for the two interfaces.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>MIMI</th>
<th>Tap&amp;Talk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69.7</td>
<td>56.1</td>
</tr>
<tr>
<td>2</td>
<td>42.6</td>
<td>31.7</td>
</tr>
<tr>
<td>3</td>
<td>71.5</td>
<td>56.4</td>
</tr>
<tr>
<td>Average</td>
<td>60.9</td>
<td>47.1</td>
</tr>
</tbody>
</table>

Table 2 shows that on average dialogues are completed faster using the Tap&Talk interface than using the MIMI interface (the difference is on average 13.8 seconds). A Wilcoxon test showed that this difference is not significant (z = .63; N.S.).

### 3.3. User satisfaction

The results of the user satisfaction questionnaire are shown in Table 3. Most statements concern both systems, except for statements 15-19; those only concern the MIMI interface. For the negative statements 4, 9 and 13 both the statement and the scores have been inverted, so that high scores always denote the positive end of the scale.

<table>
<thead>
<tr>
<th>Statement</th>
<th>MIMI</th>
<th>Tap&amp;Talk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I consider the system easy to use</td>
<td>3.4</td>
<td>3.9</td>
</tr>
<tr>
<td>2. I always understood what was expected from me</td>
<td>3.8</td>
<td>4.7</td>
</tr>
<tr>
<td>3. I found it easy to correct errors</td>
<td>3.3</td>
<td>4.0</td>
</tr>
<tr>
<td>4. I thought the system was NOT slow</td>
<td>1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>5. I thought the travel advice was clear</td>
<td>4.5</td>
<td>4.7</td>
</tr>
<tr>
<td>6. The combination of speech and graphics was useful</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>7. The system reacted adequately to the combined input</td>
<td>3.4</td>
<td>4.1</td>
</tr>
<tr>
<td>8. Visualizing the filling form was useful</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td>9. I was NOT distracted by the display</td>
<td>3.1</td>
<td>3.4</td>
</tr>
</tbody>
</table>
In general, user satisfaction was rather high. Most statements were judged about equal for both interfaces, but in some cases there were minor differences mostly in favor of the Tap&Talk interface. Subjects understood significantly better what was expected from them using the Tap&Talk interface than using the MIMI interface (z = 3.13; p < .05) and correcting errors was considered easier in the Tap&Talk interface, although the effectiveness data suggest that correcting errors was easier using the MIMI interface. Also, subjects judged both systems as slow, but the MIMI system was judged significantly slower than the Tap&Talk system (z = 2.56; p < .05), which is in accordance with the efficiency results described in the previous section. According to the scores for statement 11, the learning effect for the MIMI interface was stronger than for the Tap&Talk interface. In general, subjects appreciated the visualization of the filling form (st. 8), and they thought the travel advice was clear (st. 5). Also, they judged the visualization of the travel advice useful (st. 10). Finally, most subjects would use this type of application if it were on their mobile device (st. 14).

Table 4 shows the user preferences for the two interfaces. On most aspects subjects clearly preferred the Tap&Talk interface, as could be expected given the satisfaction data in Table 3. Once again, subjects indicated that correcting errors was easier using the Tap&Talk system than using the MIMI system, which is in contradiction with the effectiveness data. The only aspect on which there was no clear preference for one of the two interfaces concerned the travel advice; apparently, users were not helped by the spoken version of the travel advice in the MIMI interface. 71% of the subjects indicated that they would choose to use the Tap&Talk interface. In general, subjects appreciated the learning effect for the MIMI interface was stronger than for the Tap&Talk interface. Subjects understood significantly better what was expected from them using the Tap&Talk interface than using the MIMI interface (z = 3.13; p < .05), which is in accordance with the efficiency results described in the previous section. According to the scores for statement 11, the learning effect for the MIMI interface was stronger than for the Tap&Talk interface. In general, subjects appreciated the visualization of the filling form (st. 8), and they thought the travel advice was clear (st. 5). Also, they judged the visualization of the travel advice useful (st. 10). Finally, most subjects would use this type of application if it were on their mobile device (st. 14).

Table 4: User preferences

<table>
<thead>
<tr>
<th>Question</th>
<th>Preferred system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which system did you consider the easiest to use?</td>
<td>MIMI Tap&amp;Talk No preference</td>
</tr>
<tr>
<td>With which system did you know best which information you had to provide?</td>
<td>4 (24%) 13 (76%) 1 (6%)</td>
</tr>
<tr>
<td>With which system was correcting errors easiest?</td>
<td>2 (12%) 14 (82%) 1 (6%)</td>
</tr>
<tr>
<td>Which system did you consider the most fun to use?</td>
<td>6 (35%) 11 (65%)</td>
</tr>
<tr>
<td>With which system was understanding the travel advice easiest?</td>
<td>4 (24%) 5 (29%) 8 (47%)</td>
</tr>
<tr>
<td>Which system would you prefer to use in the future?</td>
<td>5 (29%) 12 (71%)</td>
</tr>
</tbody>
</table>

The results show that the effectiveness of the two interfaces does not differ significantly. However, due to the absence of spoken prompts and the obligatory use of buttons to provide values, the efficiency of the successful dialogues with the user driven interface is much higher than the efficiency of the mixed initiative interface. Apparently, also for novice users the user driven interface is much more efficient than the mixed initiative interface without a real loss of effectiveness. However, based on the results of a previous experiment, we expect that the learning curve will be much steeper for the mixed initiative interface than for the user driven interface, so that the differences will reduce once users get more experienced [2].

Finally, although the overall user satisfaction was not significantly higher for the Tap&Talk interface, by far most people preferred the Tap&Talk interface to the MIMI interface.

5. Acknowledgement

The MATIS project is funded by the Dutch Ministry of Economic Affairs through the Innovation Oriented Programme Man-Machine Interaction (IOP-MMI).

6. References

Proceedings ISCA Workshop on Multimodal Interaction in Mobile Environments, Kloster Irsee, Germany.
Proc. of International Conference on Intelligent User Interfaces; Orlando, Florida.
Proceedings LREC2002, Gran Canaria de Las Palmas, Spain.