

# CLASSiC

---

## D6.1.1: Domain-limited TTS corpus for expressive speech synthesis and Wizard-of-Oz Data for NLG Strategies

---

Cedric Boidin, Verena Rieser, Srinivasan Janarthanam, Oliver Lemon

Distribution: Consortium

---

CLASSiC

Computational Learning in Adaptive Systems for Spoken Conversation  
216594 Deliverable 6.1.1

October 2009



Project funded by the European Community  
under the Seventh Framework Programme for  
Research and Technological Development



CogSys  
Cognitive Systems



*The deliverable identification sheet is to be found on the reverse of this page.*

<b>Project ref. no.</b>	216594
<b>Project acronym</b>	CLASSiC
<b>Project full title</b>	Computational Learning in Adaptive Systems for Spoken Conversation
<b>Instrument</b>	STREP
<b>Thematic Priority</b>	Cognitive Systems, Interaction, and Robotics
<b>Start date / duration</b>	01 March 2008 / 36 Months

<b>Security</b>	Consortium
<b>Contractual date of delivery</b>	M18 = August 2009
<b>Actual date of delivery</b>	October 2009
<b>Deliverable number</b>	6.1.1
<b>Deliverable title</b>	D6.1.1: Domain-limited TTS corpus for expressive speech synthesis and Wizard-of-Oz Data for NLG Strategies
<b>Type</b>	Report
<b>Status &amp; version</b>	final 1.0
<b>Number of pages</b>	16 (excluding front matter)
<b>Contributing WP</b>	6
<b>WP/Task responsible</b>	WP6, leader FT
<b>Other contributors</b>	FT: Laurent Roussarie, Didier Cadic, Thierry Moudenc, UEDIN: Paul Crook
<b>Author(s)</b>	FT: Cedric Boidin, UEDIN: Verena Rieser, Srinivasan Janarthanam, Xingkun Liu, Oliver Lemon
<b>EC Project Officer</b>	Philippe Gelin
<b>Keywords</b>	data collections, TTS, expressive speech databases, NLG, Wizard-Of-Oz data collection for Self-Help and TownInfo system

The partners in CLASSiC are:

<b>Heriot-Watt University</b>	HWU
<b>University of Cambridge</b>	UCAM
<b>University of Geneva</b>	GENE
<b>Ecole Supérieure d'Electricité</b>	SUPELEC
<b>France Telecom/ Orange Labs</b>	FT
<b>University of Edinburgh HCRC</b>	EDIN

For copies of reports, updates on project activities and other CLASSiC-related information, contact:

The CLASSiC Project Co-ordinator:

Dr. Oliver Lemon  
School of Mathematical and Computer Sciences (MACS)  
Heriot-Watt University  
Edinburgh  
EH14 4AS  
United Kingdom  
O.Lemon@hw.ac.uk  
Phone +44 (131) 451 3782 - Fax +44 (0)131 451 3327

Copies of reports and other material can also be accessed via the project's administration homepage,  
<http://www.classic-project.org>

©2009, The Individual Authors.

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

# Contents

<b>Executive Summary</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Domain-limited TTS corpus for expressive speech synthesis</b>	<b>2</b>
2.1 Existing “Standard” database . . . . .	2
2.2 “Gentle” and “expressive” databases collection . . . . .	2
2.3 F0 features . . . . .	3
<b>3 A NLG Wizard-of-Oz data collection for learning Referring Expression Generation in the SelfHelp domain</b>	<b>4</b>
3.1 WoZ setup for data collection . . . . .	4
3.2 Corpus . . . . .	6
<b>4 A NLG Wizard-of-Oz data collection for learning Information Presentation Strategies in the TownInfo domain</b>	<b>7</b>
4.1 WoZ data collection . . . . .	7
4.2 WoZ environment . . . . .	8
4.2.1 NLG prompt generation . . . . .	10
4.2.2 Noise modelling . . . . .	10
4.3 Experimental setup and Data collected . . . . .	11
4.3.1 Experimental setup . . . . .	11
4.3.2 Data . . . . .	13
<b>5 Conclusion</b>	<b>14</b>

## **Executive summary**

This document is the deliverable 6.1.1, due at month 18 of the CLASSiC project. It describes two data sets collected during the first half of the project: a domain-limited TTS corpus for expressive speech synthesis and two Wizard-of-Oz corpora (one in the SelfHelp and one in the TownInfo domain) collected towards learning NLG Strategies.

Some aspects of this work were published at the European Conference on Natural Language Generation (ENLG) 2009, [Janarthanam and Lemon, 2009c] and at the First International Workshop on Spoken Dialogue Systems [Liu et al., 2009].

# 1 Introduction

The Text-To-Speech (TTS) and Natural Language Generation (NLG) modules of the CLASSIC end-to-end statistical architecture need to be trained on data specific to the targeted task or domain. This document describes the data that were collected towards the training of these modules: a domain-limited corpus of expressive speech for the TTS module and some Wizard-of-Oz data for the NLG module.

The TTS corpus consists of two expressive databases of the “Julie” voice, corresponding to the “gentle” and “expressive” speaking styles. They are to be added to the already existing “standard” database.

We also collected two Wizard-Of-Oz corpora [Fraser and Gilbert, 1991, Dahlbäck et al., 1993] for training the NLG modules: we conducted a NLG Wizard-of-Oz data collection for learning Referring Expression Generation in the SelfHelp domain, as described in section 3, and one for learning Information Presentation Strategies in the TownInfo domain, as described in section 4.

## 2 Domain-limited TTS corpus for expressive speech synthesis

### 2.1 Existing “Standard” database

Among the various voices already available in the France Telecom Baratinoo speech synthesizer, the “Julie” voice was specially designed for IVR (Interactive Voice Response) services.

The acoustic inventory of the “standard” Julie voice contains 3207 utterances, most of them being in the IVR domain, chosen to cover the most common phonetic units.

During the recording sessions, the speaker was asked to use an IVR-like speaking style. The speaking style is more expressive than most of the regular TTS databases, which are usually designed for reading. As a joint result of the chosen utterances and the speaker speaking style, the voice is really tailored for IVR services, like Self-Help or Appointment Scheduling applications.

The Julie voice is already used in several operational France Telecom IVR services. It was also used in the CLASSIC project to build a TTS re-ranker in the deliverable 4.1.

### 2.2 “Gentle” and “expressive” databases collection

Two other expressive speech databases of the same speaker Julie were collected for the project.

First, we asked the same speaker (Julie) to record 2800 additional utterances, using a gentle/soft speaking style. As for the “standard” database, the recording scripts were taken in the IVR domain, although they were chosen different from the scripts of the “standard” database. The scripts were selected with a greedy algorithm, in order to cover the most common phonetic units (diphones, triphones). As the recording scripts were taken in the IVR domain, the synthesis quality is optimized for the utterances in the IVR domain or close to it.

Additionally, we gathered 3192 utterances from a real France Telecom IVR service that uses Julie’s voice. These utterances were not originally dedicated to speech synthesis, and the speaker used a really expressive speaking style, putting a lot of emphasis on some words. Therefore this database is called the “expressive” database.

However, the phonetic content of the database is not optimised and some common phonetic sequences are missing, possibly leading to local artefacts during synthesis.

### 2.3 F0 features

Table 1 describes the three databases in terms of F0. It can be observed that all the F0 mean, median, range and standard deviation increase with the expressiveness of the database. Indeed, the expressive databases contain more and stronger pitch accents, which lead to a global higher pitch and more variability. The three labels “gentle”, “standard” and “expressive” are therefore coherent with the F0 features.

	gentle	standard	expressive
F0 mean	195 Hz	203 Hz	229 Hz
F0 median	188 Hz	198 Hz	231 Hz
F0 min (1st decile)	146 Hz	146 Hz	158 Hz
F0 max (last decile)	280 Hz	294 Hz	334 Hz
F0 range (max - min)	134 Hz = 11.3 st	149 Hz = 12.2 st	176 Hz = 13.0 st
F0 standard deviation	55 Hz = 4.3 st	62 Hz = 4.6 st	75 Hz = 4.9 st

Table 1: F0 features of the 3 databases (st: semi-tones)

The complete TTS dataset therefore consists of 9199 wav files, divided into three distinct databases.

### 3 A NLG Wizard-of-Oz data collection for learning Referring Expression Generation in the SelfHelp domain

A self-help Wizard-of-Oz (WoZ) dialogue corpus was collected for studying the effect of different kinds of referring expressions on users with different domain knowledge. Users were engaged in a technical broadband installation task and their dialogue interaction with the system was recorded.

#### 3.1 WoZ setup for data collection

Wizard-of-Oz environments are widely used in dialogue systems development for initial data collection [Fraser and Gilbert, 1991, Dahlbäck et al., 1993]. In a WoZ study a human operator, also known as “wizard”, performs some of the tasks of a SDS, e.g. Spoken Language Understanding and Dialogue Management, while the user is left in the belief that he is interacting with a real system.

Our WoZ environment was built to observe the user’s dialogue behaviour and build user simulation models [Janarthanam and Lemon, 2009b, Janarthanam and Lemon, 2009a]. In our setup, the human wizard replaces the speech recognition and understanding modules of the system and served to understand the user’s utterances. The wizard annotates the user’s speech act, using the wizard interaction tool, see Figure 1 (next page), and sends it to the dialogue manager. The built-in dialogue management module decides what instruction to give the user based on the user’s dialogue act using a fixed, hand-coded policy, since for this data collection we are only interested in NLG decisions. The task for the user is to follow the system instructions to use the set of components on the table to set up a broadband connection. He is provided with a modem, broadband filters, cables, etc that are needed for the task.

The NLG module of the dialogue system can modify the system utterances to contain three different ways to refer to the domain objects - descriptive expressions, technical terms and tutorial.

1. **Technical:** Choose technical terms for every reference to the domain objects. e.g. “*Connect the broadband cable to the broadband filter.*”
2. **Descriptive:** Choose descriptive terms for every reference to the domain objects. e.g. “*Connect the thin cable with grey ends to the small white box.*”
3. **Tutorial:** Use technical terms, but also augment the description for every reference. e.g. “*Connect the broadband cable to the broadband filter. The broadband cable is the thin cable with grey ends. The broadband filter is the small white box.*”

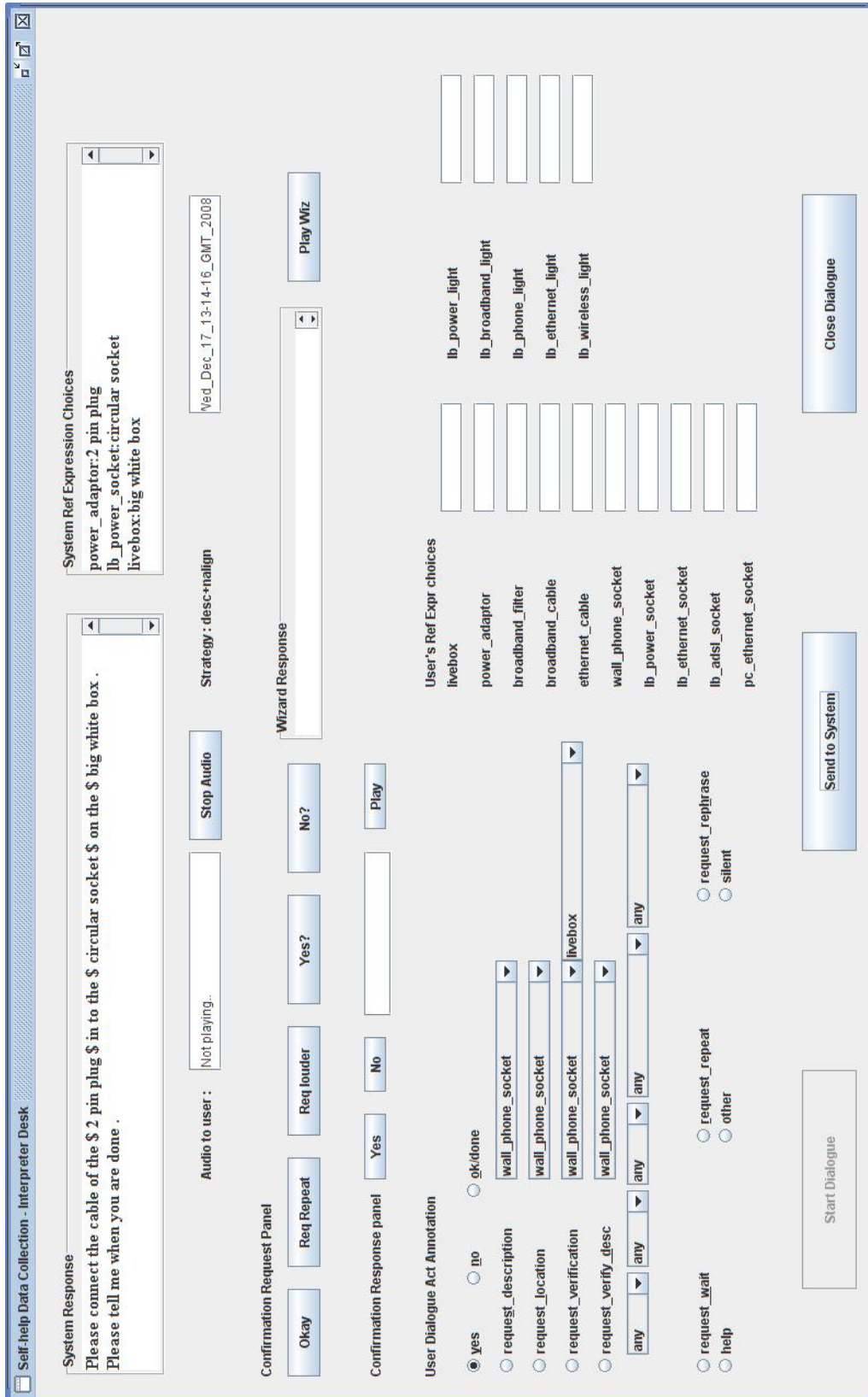


Figure 1: Wizard Interaction Tool for Referring Expression Generation in the SelfHelp domain

<b>Strategy</b>	<b>Technical</b>	<b>Descriptive</b>	<b>Tutorial</b>
No. of Dialogues	6	6	5
No. of system turns	169	155	126
Avg. no. of system turns	28.17	25.83	25.2
Avg. dialogue time	462s	411.83s	703.2s
Avg. system utterance length (in words)	470.5	471.66	945.6
Avg. clarification request count	3.16	0	0

Table 2: Data collected for 3 referring strategies

The utterances from the NLG module are played back to the user, using the Cereproc speech synthesiser.<sup>1</sup>

### 3.2 Corpus

At the beginning of each session the participants were asked to fill in a background questionnaire and were tested for their knowledge of the components on the table using a pre-task test. They were asked to point out the component that is being mentioned using the technical term. This was done to record their domain knowledge, which is an essential component in our user simulation models. After the pre-task test, the participants interacted with the system over a headset.

The NLG strategy (descriptive, technical or tutorial) is randomly chosen and for the whole session the same strategy is used. For instance, if the system chooses a descriptive strategy, all the system utterances contain descriptive references to the domain objects. During the session, the system provided the participants with instructions to put the broadband setup together. The participant's speech was recorded as an audio file in WAV format. Both the system's and the participant's moves (as annotated by the human wizard) were logged. The system's choice of referring expressions (according to the NLG strategy) were also recorded in the logs. At the end of the session, the broadband setup on the participant's table was manually observed and task completion was logged. The participants were then asked to take a post-task test, similar to the pre-test, to record their learning gain. Finally, the participants filled in a user satisfaction questionnaire, in which they were asked to rate the system based on the dialogue length, use of appropriate referring expressions and so on.

The WoZ Self-help corpus consists of data from interactions with 17 participants (900 turns in total, 450 system turns, 450 user turns). More information on the nature of the dialogues are given in Table 2.

<sup>1</sup><http://www.cereproc.com/>

## 4 A NLG Wizard-of-Oz data collection for learning Information Presentation Strategies in the TownInfo domain

We collected a second NLG WoZ corpus in the TownInfo domain, where we are interested in strategy and attribute selection for Information Presentation Strategies (IP) in Spoken Dialogue Systems (SDS).

We follow the overall framework of NLG as planning under uncertainty [Rieser and Lemon, 2009, Lemon, 2008], where each NLG action is a sequential decision problem, based on the current (uncertain) dialogue state and the expected utility/“reward” of the interaction. Other recent approaches describe this task as planning, e.g. [Koller and Stone, 2007, Koller and Petrick, 2008], or as contextual decision making according to a utility/ cost function [van Deemter, 2009].

We apply this framework to Information Presentation strategies in SDS, where the example task is to present a set of restaurants to the user.

However, we believe that the presented framework can be applied to many other domains which require complex information to be conveyed to the user, e.g. instruction giving and tutorial dialogue, or other types of Information Presentation tasks, such as sales agents, tourist or health information systems.

### 4.1 WoZ data collection



Figure 2: Overview of the WoZ setup

Our WoZ setup involves 4 stages (see Figure 2): First, the user’s utterance is tagged with attribute values by the wizard. For some of the cases artificial ASR noise is introduced (see Section 4.2.2). In the noise condition it is the experimenter who listens to the user’s utterance and does the tagging. For the no-noise condition the wizard directly listens to the user’s utterance and translates into attribute values. For example, “I am looking for Indian restaurants in the Old Town” gets tagged as `cuisine=Indian`, `location=Old Town`. The wizard then queries the database, where we use a real database of Edinburgh restaurants provided by TheList,<sup>2</sup> and selects an NLG strategy.

<sup>2</sup><http://www.list.co.uk/>

There were no general time constraints on the wizards. However, wizards were encouraged to act as quickly as possible.

The strategy then gets generated by a surface realiser (see Section 4.2.1). The final utterance is played back to the user via TTS, using the Cereproc speech synthesiser.<sup>3</sup> We manually transcribed restaurant names (especially those with international names) in order to improve the TTS quality for proper names.

## 4.2 WoZ environment

Figure 3 shows the web-based interface for the wizard. The experimenter, the user, and the experimenter have similar interfaces, which communicate with the wizard's page using a web-based server-client architecture. The audio is transmitted and recorded using VOIP<sup>4</sup>. The wizard GUI contains 5 main panels (see Figure 3):

- A:** The wizard receives the user's query as noisy attribute values from the noise model. The experimenter has a similar input panel. There are 5 searchable attributes in total, which can also be negative ("not expensive").
- B:** The retrieved database items are presented in an ordered list. We use a User Modelling approach for ranking the restaurants, see e.g. [Polifroni and Walker, 2008], where we assume that a default user cares about cheap food with high quality and good service.
- C:** The wizard then chooses which strategy and which attributes to generate next, by clicking radio buttons. The attribute/s specified in the last user query are pre-selected by default.
- D:** An utterance is automatically generated by the NLG surface realiser every time the wizard selects a strategy, and is displayed in an intermediate text panel.
- E:** The wizard can decide to add the generated utterance to the final output panel. The text in the final panel is sent to the user via TTS, once the wizard decides to stop generating.

(Please see Figure 3 on the next page.)

---

<sup>3</sup><http://www.cereproc.com/>

<sup>4</sup>Skype, <http://www.skype.com/>

**CLASSIC WoZ -- TownInfo NLG**  
*University of Edinburgh (12.2008)*

**A** **User Hangup**  
Select User Input below

Cuisine:  Not  
 International  Blisro  Chinese  Fish  
 French  Indian  Italian  Mexican  
 Scottish  Spanish  Thai  Vegetarian  
 Unknown

Price:  Not  
 Cheap  Moderate  Expensive  Unknown

Location:  Not  
 Old Town  New Town  West End  City Centre  
 Both  Southside  Tollcross  Unknown

Food Quality:  Not  
 Poor  Good  Excellent  Don'tCare

Service Quality:  Not  
 Poor  Good  excellent  Don'tCare

**B** **Task 1 of 12**

Number of Results = 21

name	cuisine%	price%	priceType%	cityArea%	food%	service%
Amber	Scottish	20	moderate	Old Town	excellent	excellent
The Grain Store	Scottish	21	moderate	Old Town	excellent	excellent
New Walk Bar	Scottish	20	moderate	Southside	excellent	excellent
The Jazz Bar	Scottish	20	moderate	Old Town	excellent	excellent
manseilas bakery	Scottish	21	moderate	Tollcross	excellent	excellent
shang high	Scottish	21	moderate	Meadowbank	excellent	excellent
Restaurant Botham	Scottish	21	moderate	West End	good	excellent
A Room in the West End	Scottish	19	moderate	West End	good	excellent
number 3	Scottish	19	moderate	Canon Hill	good	excellent
honeyeels	Scottish	19	moderate	Bruntsfield	good	good

Page: 1 of 3

**C** **Select the Strategy and Number:**  
 Cuisine -- Price -- Location -- Quality -- Service

SummarizeAll:  no UM  With UM  
 Compare:  byItem  byAttribute  
 Recommendation:  Recommendation

**D** **Modify / Add to Final Utterance Box on the right:**  
 The restaurant called Amber and the restaurant called The Grain Store are both in the moderate price range, are both Scottish restaurants, are both located in Old town.

**E** **Send to TTS and User:**

**FinalConfirm**

**Last System Utterance:**  
 Okay, please wait a moment while I search the database.

Figure 3: Wizard interface

### 4.2.1 NLG prompt generation

We have implemented an NLG surface realiser in order to generate IP strategies in real time. This generator is based on data from a stochastic sentence planner called SPaRKY [Stent et al., 2004]. We replicate the variation observed in SPaRKY by analyzing high-ranking example outputs (given the highest possible score by both SPaRKY judges) and implement the variance in dynamic templates. The realisations vary in sentence aggregation, aggregation operators (e.g. ‘and’, ‘full stop’, or ellipsis), contrasts (e.g. ‘however’, ‘on the other hand’) and referring expressions (e.g. ‘it’, ‘this restaurant’) used. In particular, we consider the following realisations of Information Presentation actions (see examples in Table 3):

- RECOMMEND the top-ranking restaurant (according to UM).
- COMPARE the top 2 restaurants by Item or by Attribute.
- SUMMARY of all matching restaurants with or without User Model (UM), following [Polifroni and Walker, 2008]. The approach using a UM assumes that the user has certain preferences (e.g. cheap) and only tells him about the relevant options, whereas the approach with no UM lists all the options.

strategy	example utterance
SUMMARY no UM	I found 26 restaurants, which have Indian cuisine. Furthermore, 11 of the restaurants are in the expensive price range. However, 10 of the restaurants are in the cheap price range and 5 of the restaurants are in the moderate price range.
SUMMARY UM	26 restaurants meet your query. There are 10 restaurants which serve Indian food and are in the cheap price range. There are also 16 others which are more expensive.
COMPARE by Item	The restaurant called Kebab Mahal is an Indian restaurant. It is in the cheap price range. And the restaurant called Saffrani, which is also an Indian restaurant, is also in the cheap price range.
COMPARE by Attr	The restaurant called Kebab Mahal and the restaurant called Saffrani are both in the cheap price range , and are both Indian restaurants.
RECOMMEND	The restaurant called Kebab Mahal has the best overall quality amongst the matching restaurants. It is an Indian restaurant , and it is in the cheap price range.

Table 3: Example NLG IP actions generated for user provided `cuisine=Indian`, and the wizard also selects the additional attribute `price`.

### 4.2.2 Noise modelling

We also introduce artificial noise in the WoZ setup, replicating non- and misunderstanding errors as introduced by ASR. In contrast to previous work, e.g. [Rieser et al., 2005, Stuttle et al., 2004],

we simulate noise on the semantic level, modelling deletions, substitutions, and insertions for the attribute-value pairs. Our noise model introduces noise for 36.3% of the attributes provided by the user, where 61.7% are deleted, 26.4% are replaced, and 11.8% are inserted, according to an analysis of data gathered by [Lemon et al., 2006]. A similar noise model has also been used by [Crook and Lemon, 2009].

We are interested in the kind of effects the input noise has on the wizard's decisions, and also whether the user prefers different IP strategies in the presence of noise. For example, the user might prefer shorter system prompts if there was some mis-recognition.

## 4.3 Experimental setup and Data collected

### 4.3.1 Experimental setup

We use a fixed, hand-coded dialogue management strategy, since we are only interested in IP strategies generated by our NLG module. In particular, we hand-code dialogue behaviour such as greeting the user in the beginning, triggering wait messages when the database is searched by the wizard, informing the user if no items are found, and final messages (see Example 1). The wizard can also directly answer user requests by typing messages into the final TTS panel (E) and sending them to the user.

The user can either provide more constraints, select a restaurant, ask a question, e.g. request more information, ask to repeat the last system utterance, or hang up.

- (1) **WoZ:** Welcome to the Edinburgh restaurant system. How may I help you today? (**greeting**)  
**User:** I am looking for a French restaurant which is not in the City Centre. (**provideInfo**)  
**WoZ:** Okay, please wait a moment while I search the database. (**wait message**)  
**WoZ:** I found 16 restaurants , which are not located in City Centre, and which have French cuisine. The restaurant called La Garrigue is in the cheap price range. It is located in Old Town. The restaurant called Jacques is in the cheap price range and it is located in Tollcross. Do you want to choose this one? (**IP**)  
**User:** What is the name again? (**RequestMoreInfo**)  
**WoZ:** La Garrigue (**typed message**)  
**User:** Ok, great. I'll take it. (**select**)  
**WoZ:** Please click the Finish button and then take some time to fill out the questionnaire. (**final message**)

We collect data from 2 different wizards (w1,w2) on this task. Both are native speakers of English and have some background in computer science. The wizard and the users are located in two different rooms.

**Task sets:** We designed 10 task sets. Each task set contains 12 tasks of 3 different types. Type I tasks specified constraints so that 2-4 database hits were returned. Type II tasks define a set of 5-13 database hits, and Type III tasks 15-46. We designed different task types in order to expose the wizards to sets of different sizes to present to the user. Task types are arranged into different orders for every task set. Furthermore, the task description included some vagueness about the values, in order not to prime the user to use certain words. The following task description, for example, can be translated into attribute-value pairs as {location=Leith}, {price= moderate or cheap}.

*You're on a visit to Leith and it's time for dinner. You're open to any type of food, as long as it's not expensive.*

Each user has to perform a total of 12 tasks, where no task set is seen twice by one wizard. The users are given general instructions and an agreement form in the beginning of each session. We also ask the users to provide us some background information on age group, profession, native language, how often the users eat out in restaurants (in order to find out their general interest in the domain task), and whether they used a SDS before and his prior expectations/experience with a SDS.

After each task the users answer a questionnaire on a forced-choice, 6 point Likert scale (*strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree*). This task-based questionnaire enquires about different aspects of the Information Presentation strategy:

1. The system's voice was easy to understand. (**TTS quality**)
2. The way the system presented information was good. (**Information Presentation**)
3. The system's utterances had the right length. (**Utterance Length**)
4. The system gave me a good overview of all the available options. (**Coverage**)
5. The restaurant I chose was a good match for this task. (**Task Success**)

At the end of the session the users were asked to fill out a final questionnaire about his general experience with the "system":

1. The time delay of the system affected my ratings. (**timeDelay**)
2. I wasn't able to choose a restaurant because I could not understand the system's voice. (**TTS and TaskSuccess**)
3. I'd like to call the system in the future for restaurant information. (**Future Use**)

Each user received a £10 compensation, with £2 extra when s/he successfully completed all the tasks. We introduced this additional reward in order to encourage attention to task completion.

### 4.3.2 Data

We have collected 213 dialogues with 18 subjects and 2 wizards in this setup. The dialogues are on average  $7.03(\pm 3.0)$  turns and  $02 : 21(\pm 01 : 28)$  minutes long. Table 4 presents an overview of the data. Note that both wizards contributed to both noise conditions, however the proportions are in diametrical opposition: w1 contributed almost  $\frac{3}{4}$  of the noise condition, whereas w2 contributed almost  $\frac{3}{4}$  of the no-noise condition. Ideally these proportions would be balanced.

Most of the subjects were female (77.6%), most were students between 20 and 30 years old (63.1%), and most were native speakers of English (83.2%). They can be assumed to have a reasonable interest in restaurants: 66.8% eat out once a month, and 27.67% eat out every week. Most of the subjects have not used a SDS before (57.9%).

wizard	noise	no-noise	total
w1	5 subjects $\times$ 12 dialogues = 60 dialogues  72.43% of noise condition	3 subjects $\times$ 12 dialogues = 36 dialogues  27.9% of no-noise condition	= 96 dialogues, 8 subjects, 249 NLG actions
w2	2 subjects $\times$ 12 dialogues = 24 dialogues  28.6% of noise condition	8 subjects $\times$ (11+6+12+18+10+3 $\times$ 12) dialogues = 93 dialogues  72.1% of no-noise condition	= 117 dialogues, 10 subjects, 219 NLG actions
total	= 84 dialogues, 7 subjects, 194 NLG actions	= 129 dialogues, 11 subjects, 274 NLG actions	213 dialogues, 18 subjects, 468 NLG actions

Table 4: Data collected for Information Presentation strategies for two wizards under two noise conditions.

The data contains ca. 2236 utterances in total: 1465 system prompts and ca. 771 user prompts.<sup>5</sup> 32% of the system utterances are NLG Information Presentation strategies ( $n=468$ ). The other system prompts are greeting the user (14.7%), wait messages (33.2%), final prompts (14.2%), informing the user that no matches were found (1.43%), and other system prompts such as the wizard typing in text to send to the user (4.5%).

All the interactions are logged as XML files and the whole conversation is recorded.

<sup>5</sup>User prompts are calculated as system prompts minus wait messages minus final messages.

## 5 Conclusion

Two speech databases were collected to increase the expressiveness of the Baratinoo Julie voice. Besides the existing “standard” database, an additional database was recorded with a “gentle” speaking style and another database of real IVR sentences was gathered, referred as the “expressive” database. The three databases are used in the deliverable 4.3 to generate TTS expressive variants.

In addition, two Wizard-Of-Oz corpora were collected for training the NLG modules: a corpus for learning optimal Referring Expression Generation in the SelfHelp domain and another corpus for learning optimal Information Presentation Strategies in the TownInfo domain. They will be used for training the NLG module in tasks 4.2 and 4.4.

## References

- [Crook and Lemon, 2009] Crook, P. A. and Lemon, O. (2009). Accurate Probability Estimation of Hypothesised User Acts for POMDP Approaches to Dialogue Management. In *12th Annual Research Colloquium of the special-interest group for computational linguistics in the UK and Ireland (CLUKI)*.
- [Dahlbäck et al., 1993] Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of Oz-studies – why and how. In *Proc. of the Workshop on Intelligent User Interfaces*.
- [Fraser and Gilbert, 1991] Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech and Language*, 5:81–99.
- [Janarthanam and Lemon, 2009a] Janarthanam, S. and Lemon, O. (2009a). A Data-driven method for Adaptive Referring Expression Generation in Automated Dialogue Systems: Maximising Expected Utility. In *Proceedings of PRE-CogSci*.
- [Janarthanam and Lemon, 2009b] Janarthanam, S. and Lemon, O. (2009b). A Two-tier User Simulation Model for Reinforcement Learning of Adaptive Referring Expression Generation Policies. In *Proceedings of SIGDial*.
- [Janarthanam and Lemon, 2009c] Janarthanam, S. and Lemon, O. (2009c). A Wizard-of-Oz Environment to study Referring Expression Generation in a Situated Spoken Dialogue Task. In *Proceedings of ENLG*.
- [Koller and Petrick, 2008] Koller, A. and Petrick, R. (2008). Experiences with planning for Natural Language Generation. In *ICAPS*.
- [Koller and Stone, 2007] Koller, A. and Stone, M. (2007). Sentence generation as planning. In *Proceedings of ACL*.
- [Lemon, 2008] Lemon, O. (2008). Adaptive Natural Language Generation in Dialogue using Reinforcement Learning. In *Proceedings of SEMdial*.
- [Lemon et al., 2006] Lemon, O., Georgila, K., and Henderson, J. (2006). Evaluating Effectiveness and Portability of Reinforcement Learned Dialogue Strategies with real users: the TALK TownInfo Evaluation. In *IEEE/ACL Spoken Language Technology*.
- [Liu et al., 2009] Liu, X., Rieser, V., and Lemon, O. (2009). A Wizard-of-Oz interface to study Information Presentation strategies for Spoken Dialogue Systems. In *Proc. of the 1st International Workshop on Spoken Dialogue Systems*.
- [Polifroni and Walker, 2008] Polifroni, J. and Walker, M. (2008). Intensional Summaries as Cooperative Responses in Dialogue Automation and Evaluation. In *Proceedings of ACL*.

- [Rieser et al., 2005] Rieser, V., Kruijff-Korbayová, I., and Lemon, O. (2005). A corpus collection and annotation framework for learning multimodal clarification strategies. In *Proc. of the 6th SIGdial Workshop on Discourse and Dialogue*.
- [Rieser and Lemon, 2009] Rieser, V. and Lemon, O. (2009). Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems. In *Proc. of EACL*.
- [Stent et al., 2004] Stent, A., Prasad, R., and Walker, M. (2004). Trainable sentence planning for complex information presentation in spoken dialog systems. In *Association for Computational Linguistics*.
- [Stuttle et al., 2004] Stuttle, M., Williams, J., and Young, S. (2004). A framework for dialogue data collection with a simulated ASR channel. In *Proc. of the International Conference of Spoken Language Processing (Interspeech/ICSLP)*.
- [van Deemter, 2009] van Deemter, K. (2009). What Game Theory can do for NLG: the case of vague language. In *12th European Workshop on Natural Language Generation (ENLG)*.