Human Factors in Operetta

DISC partner: MIP
Authors: Laila Dybkjær and Niels Ole Bernsen

1. Introduction

This paper presents an analysis of human factors in the Operetta system. The analysis is presented in the form of (a) a 'grid' which describes the system's properties with particular emphasis on human factors and evaluation results, (b) a life-cycle model which provides a structured description of the system's development and evaluation process, and (c) supporting material, such as system architecture, example screen shots, dialogues and scenarios.

The presented information will be cross-checked with the developers of Operetta as well as with the complementary descriptions of other aspects of Operetta provided by the DISC partners. These other descriptions address speech recognition, done by KTH, language understanding and generation, done by IMS, and system integration, done by LIMSI.

Demonstrator: Available in Cambridge
Developer: Vocalis
Contact: David Williams

Human factors of the Operetta system

The Operetta system is a commercial call answering and routing system (an "Operator's assistant"). Another part of Operetta is a touch-tone driven voice mail system (ignored in what follows). In what follows, we shall also ignore the system's interface to the call recipients. The system was developed by Vocalis, UK.
URL: http://www.vocalis.com/pages/products/operetta.htm

System performance	
Cooperativity	**System prompts were orginally based an a typical receptionist. However, the greeting message, i.e. Welcome to the company X automatic swicthboard, has been changed on a beta-site by beta-site basis. DW has been involved in proof reading ideas from clients but there has been no formal evaluation. The latest incarnation of a short prompt can be hear on the Vocalis Operetta now. In general prompts are suggested by DW and then changed on an ad hoc basis as a function of the client's feedback.**
Initiative	**Domain communication:** system-directed. **The user is asked for a name. No further automated help is offered if the caller stays silent, they are simply connected to an available operator.** **Meta-communication: No meta-communication. Operetta only supports routing operations, i.e. names or DTMF, to a person or the operator.**
Influencing users	Explicit and implicit user instructions; walk-up-and-use system. The system's introduction presents the system and instructs the user to clearly say the first and last name of a person, offering operator back-up as an alternative. **There are a variety of messages: operator busy: please hold the line or leave a voice mail; operator no answer: leave a voice mail; out of hours (business hours of company defined in the Operetta set up): 'Sorry the company is closed etc' - say a name (someone you know may still be in the building) or leave a voice mail** The system's introduction is optional and is de-selected by keying in the recipient's extension number. Users are clearly instructed to **say the first and last name of the person they want to talk to. The opening prompt would be too long if all commands were elucidated - this is a major consideration. It was supposed people would learn of the 'star' to speak option through word of mouth, in training etc. Around 70% of people say the right thing. This was evaluated formally when Operetta was first put on the Vocalis switch.**
Real-time	The system responds in real time.
Transaction success	**Transaction success=callers getting the person they wanted with the Operetta/Operator hybrid (Operetta is marketed as the Operator assistant, not replacement) is 100%!**
General evaluation	Confirmative feedback is used (ISO 9241-10)
Speech input	
Nature	Continuous; speaker-independent; accent, age and gender independent; English.

Device(s)	1. Telephone. 2. Tape recorder for recording the names input by the caller. The caller's own name is replayed to the recipient of the call. And so is the recipient's name for use if recognition has difficulties. The tape recorder also takes messages.
Phone server	**Has a phone server been deployed? Please describe it. Any important limitations here?**
Acoustic models	Strings of phonemes constituting phrases which make up the names of the persons in the organisation.
Search	**Describe the type of search used by the recogniser.**
Vocabulary	The system currently has a vocabulary of 100 names + the word "yes". Transcriptions for these can be added and deleted at any time. **The same vocabulary is active at all points, e.g. if a recognised name has been disconfirmed by the caller it still plays a part in the next recognition.**
Barge-in	Only via DTMF by keying in the extension number.
Word hypotheses	The recogniser produces a best recognition with attached confidence score.
Grammar	There is no grammar in the speech recogniser.
Prosody	The system does not process input prosody.
Speech output	
Device(s)	Telephone.
Language(s)	English.
Device	Telephone.
Coded/parametric	Coded. **Quality measured/evaluated? How? Results?**
Prosody	No prosody processing has been included.
Voice character	Normal human voice; English; male or female (chosen by the system administrator). **The voice has changed from mail to female due to client demand.**
Lexicon	Approx. 100 pre-recorded names + the company roles of these people. No resources are shared between the speech synthesis module and other modules in the system.
User utterances	
Lexicon	Approx. 100 words (person names + "yes"). Lexical coverage is sufficient.
Grammar	None needed.
Parsing	None needed.
Semantics	No semantic representation is being built (or needed).
Style	Extremely terse: a name or "yes".
System utterances	

Generation	Is part of dialogue management.
Lexicon	Pre-defined prerecorded names and phrases.
Grammar	None needed.
Semantics	None needed.
Style	Friendly.
Multimodal aspects	
Device	Telephone keyboard.
Non-speech input	In addition to speech, the system accepts DTMF input.
Non-speech output	None.
Role(s)	**0 for operator, 'star' to skip greeting, 'hash' to access voicemail. Voice mail is totally DTMF driven.**
Evaluation	**No evaluation.**
Attentional state	
Focus, prior	None. The entire vocabulary is in active memory at any time.
Sub-task id.	The system does not do sub-task identification but requests the sub-task to be done by the user at any time.
Intentional structure	
Tasks	Operetta picks up a ringing phone, greets the caller, offers operator fall-back **[always or only during daytime?]**, asks who they want to talk with and can **[doesn't it always do that?]** ask for the caller's name. It then rings the called party, and when they answer, it announces the call and puts the caller through - just like a human operator. If the switchboard supports music on hold, the caller will hear music. If the called party does not answer their phone, or the line is engaged, Operetta offers to take a message or to put the caller through to somebody else. The called party can access their messages whenever, and from whereever they choose. **If a person has left it is up to the system admin. person to update the name database. If all goes well the caller will not be recognised (OOV) and will be routed to the operator to find out the person has left (again the 100% accurate hybrid system!)**
Task complexity	Simple tasks; well-structured. **No complexity measures have been applied.**
Communication	Domain communication: system directed only. Highly constrained: only single phrases (names) and "yes" are allowed from the users. No explicit or echo system (domain) feedback. System-initiated meta-communication**: If the recogniser returns a medium confidence, the caller is asked to confirm the recognised name, i.e. "was that <David Williams> " <>denote the name message recorded by David Williams.**

	User-initiated meta-communication: **None.** **Problems: Saying names is tricky. What are the cultural norms, e.g. surname first name, titles, etc. This is a problem for the opening prompts. It is not sufficent just to say 'full-name' as this is ambiguous.** **Other forms of system communication than domain and meta-communication: Tones signifying the 'Recogniser is listening' event.**
Interaction level	Only one level is involved in the system initiated meta-communication. No graceful degradation is being used.
Dialogue structure	**The dialogue is a flow chart with recognition calls. Designers used a GUI dialogue tool eventually, though initially a form based tool (awful).**
Linguistic structure	
Speech acts	The system does not need to identify speech (or dialogue) acts in the users' input.
Discourse particles	The system does not need to identify discourse particles in the users' input.
Co-reference	The system does not need to do co-reference resolution.
Ellipses	The system does not need to do any particular processing of ellipses.
Segmentation	The system does not need to do user turn segmentation.
Interaction history	
Linguistic	The system does not maintain a record of the surface language of the users' utterances.
Topic	The system does not maintain a record of the order in which topics have been addressed through the interaction.
Task	The system does not maintain a record of the task-relevant information which has been exchanged.
Performance	The system does not maintain a record of the user's performance during interaction.
Domain model	
Data	Internal phone book including names, staff positions and extension numbers (a maximum of 100). Operetta comes with a system administration program which allows one to update the internal phone book.
Rules	There are no rules operating on the domain data.
User model	
Goals	Assumed to be to talk to a staff member of the company.
Beliefs	No user beliefs are handled on-line.
Preferences	No user preferences discovered through earlier interactions are being

	handled.
User group	No novice/expert distinction has been made.
Cognition	No specific cognitive characteristics of users have been taken into accoount, such as task load, limited memory, natural “response packages” or limited attention span.
Architecture	
Platform	Processor: Pentium 120MHz. Memory: 64 MByte.OS: SCO Unix 3.2v4.2. Disk: 550 MByte. Telephony: Card: Dialogic D41D. Hardware customisation: Telephony card Dialogic D41D. Depending on how many ports are installed, Operetta can answer up to 4 or 8 calls at the same time. The system consists of a stand-alone unit with its own keyboard and monitor. Operetta simply plugs into four (or eight) spare analogue extensions. There are several possible configurations of Operetta, depending on the switch available, **cf. Figure 7**.
Tools and methods	**Describe the tools and methods used.**
Generic	**A runtime system (CAGE) provided the dialogue flow framework, steps specified procedurally, flow specified by goto's - distributed (dialogue, telephony, recognition).**
No. components	**Executable calls telephony and recog. libraries.**
Flow	**Recogniser returns result, telephony card returns DTMF, switch gives telephony events, e.g. flash hook, engaged tone etc., database returns numbers and extension, global settings, that's it really.**
Processing times	**N/A**

Figure 1. High-level description of the Operetta System. Boldface is used for comments, new text and entries, and unresolved questions. Underlining is used for evaluation.

System/component architecture

Operetta dialogues

- Good morning. You are speaking to the Vocalis automatic switchboard. After the tone please clearly say the first and last name of the person you want and I will connect you. If you prefer to speak to the operator please stay on the line. **1.**
- Clyde Cox. **2. 3. 4.**
- Who is calling please?
- David.
- Please hold on whilst I transfer you.
- I am sorry, there is no answer for that extension. If you would like to leave a message for Clyde Cox after the tone please say "yes"; to speak to someone else please stay on the line.
- No.
- After the tone please clearly say the first and last name of the person you want, and I will connect you. If you prefer to speak to the operator please stay on the line.
- David Williams. **5.**
- Please hold on whilst I transfer you.
- I am sorry that extension is busy. If you would like to leave a message for David Williams, Human Factors Consultant, after the tone please say "yes"; to speak to someone else please stay on the line. **6. 7.**
- Yes.
- Hey! This is David Williams. You are through to my voice mailbox. I am unable to take the phone right now so you can leave your message and I will get back to you when I can. Thanks very much.
- Hey there, David, it's David here just leaving a test message to demonstrate the system.
- Thank you for calling Vocalis: We will get back to you as soon as possible. **8.** Good bye.
- **••••••••••••••••••••••••••**

1. **What can the user do if s/he didn't get what the system just said?** Hold for an operator - there is no repeat command.

2. What happens if the system does not recognise the name - repair meta-communication? Medium confidence=confirm "was that X?" Low Conf. (OOV) = transparent to Operator - one option allows a WOZ type transfer where the name recording is played to the operator and they connect manually. The caller never knows!

3. Apparently there is no feedback from the system to the user to allow the user to judge whether the system did correct recognition? If high conf. there is simply a 'Transferring message' after the callers name is captured.

4. What if two employees are called Clyde Cox? An option can be set to route to the Operator on this name being recognised.

5. What happens if the user again asks for Clyde Cox, and again and again? I guess that the system just carries on with the same task again and again? There are a variety of options for an Operetta

1- drop to operator on first disconfirm

2- drop on second

3- transparent redirect (see above) on first
4 - as 3 on second

6. The system does not mention that the recipient may have left a message for callers. This might be useful to know. It might tell me, e.g., that NN is away for some longer duration, or that I could call secretary XY who is present. Any comments on this point? The voicemail message is recorded by the owner - it is up to them to keep it up-to-date.

7. Callers are not offered the possibility to wait on the line until the recipient's line is no longer busy. Any comments on this point? Callers can just say the name again when asking for 'someone else'.

8. Is this sentence (a) always played, (b) only played if people have said "yes" to the question about leaving a message? Only on a message being left.

- ~~Good morning. Please answer your mailbox number now or press # to leave a message in another mailbox. Please enter your password now. You have one new message. To listen to your new message press 1, to send the message press 3, for advanced features press 4, to go to another extension press 5, to quit voice mail press 0.~~
- ~~Thursday December 11th, 9.34 am. Hey there, David it's David here just leaving a test message to demonstrate the system.~~
- ~~To listen to your new message press 1, to remove it press 2, to forward it press 6, to archive it press 7, to return to the main (mail) menu press #, to quit voice press 0.~~
- ~~Are you sure you want to remove this message? Either press 1 for NO or 2 for YES.~~
- ~~The Vocalis message has been removed. You now have no new messages. To send a message press 3, for advanced features press 4, to go to another extension press 5, to quit voice press 0. Leaving with no new messages in your mailbox. Good bye.~~
- **••••••••••••••••••••••••••**
-
- A call for David Williams from Jenny Hanks.
- Good morning. You are speaking to the Vocalis switchboard.
- Dave Williams.
- Who is calling please?