Language Use, Personality and True Conversational Interfaces

Abstract

Demand for conversational interfaces is great and growing, especially for online commercial applications. Creating such interfaces so that the information provided is accurate, believable and trusted is a complex engineering task. Various studies into how users react socially to an interface have shown that it is possible to improve the usability and general attraction of an interface by carefully implementing an appropriate personality. These findings are taken and applied to a true, reactive, real world, conversational interface. The personality of the interface was manipulated by altering the linguistic style of the system's dialogue. It was found that personality effects, in such an interface, are more complex and subtle than previous research has suggested. The primary conclusion drawn from experimentation was that to produce "real" synthetic personalities the interaction as a whole must be considered. The linguistic style of different user personality groups was also found to be somewhat inconsistent with findings from written and spoken language.

Contents

Chapter 1

Background and Related Work

1.1 Introduction

> " ... an identical spoken and written language would be practically intolerable; if we spoke as we write, we should find no one to listen; and if we wrote as we speak, we should find no one to read. The spoken and written language must not be too near together, as they must not be too far apart. " T.S. Eliot

Investigations into the nature and structure of human dialogue have been a topic of research in Artificial Intelligence for decades. One of the aims of this research is to enable the implementation of reliable and believable natural language, and conversational, interfaces. A system that allowed users and computers to interactively converse, reliably and realistically, using natural language would revolutionise human-computer interaction and the role that computers play in everyday life. These interfaces would allow information to be available to users anywhere at anytime, incorporating the flexibility and efficiency of natural language. Despite the enormous technical and theoretical difficulties involved in creating these interfaces significant advances have been made over the last decade (see Zue 1997 for a review). Commercial demand for conversational interfaces is great, even though the performance of the majority of these systems is fairly weak. One relatively cheap method for improving the performance of these interfaces is to improve their social performance, thus making them more believable and attractive to use.

Reeves and Nass (1992) have consistently shown that people respond to computer interfaces in a social way; they treat the interface, in some respects, as if it were a real person interacting withthem. Reeves and Nass suggest that users reactions to such systems could be greatly improved if more attention was paid to the personality of the interface rather than the engineering problems of complex natural language processing. Reeves and Nass call this study of computer personalities as "real" personalities CASA (computers are social actors). The CASA hypothesis has been applied to many different areas including affective computing and intelligent user interfaces as well as more traditional HCI.

To test the applicability of CASA effects in conversational interfaces a prototype interface was created that projected different personality types, one was intended to act as an extravert, and one an introvert. Users reactions to the different personalities were tested using this interface; investigating how the personality of the interface affects its perceived usability as well as its general attractiveness to the user. It was found that manipulating personality by altering the linguistic style of system utterances was insufficient to show users reacting differently to either similar or different personalities. Expected CASA effects were not detected, the conclusion drawn from this is that the personality of such a dynamic and reactive system is the product of the interaction as a whole. This implies that the implementation of consistent synthetic personalities in conversational interfaces is more subtle and complex than simply altering linguistic style. Analysis of the language used by participants also shows that introverted and extraverted linguistic style in this medium may be inconsistent with personality and language use correlates in other mediums, both written and spoken.

In chapter one some relevant background ideas and theories are discussed including: related work in conversational interfaces psychological and linguistic studies into interpersonal communication; trait theories of personality; and correlations between language use and personality as well as previous studies into the relevance of these theories in human computer interaction. In chapter two some justification is given for choosing to use AIML and ProgramD as the basis for a simple model of a commercial interface, as well as some appraisal of this technology. In chapter three the processes involved in the design and implementation of the interface are discussed. In chapter four the experiments and the results of these experiments are presented along with some discusion of the results. Chapter five consists of general conclusions and discussion as well as the implications of the results for any future work in this field.

1.2 “True” Conversational Interfaces

Many natural language style interfaces can be called conversational, in that they carry out some form of interactive dialogue. These interfaces can be split into three separate categories (as suggested by Zue, 1997) depending on whether the system, user or a combination of the two plays the active role in the dialogue. Where the interface plays the dominant role, the system takes complete control of the dialogue by requiring the user to answer a set of prescribed questions (a script), these are the oldest forms of conversational interface with the dialogue taking the familiar form of question [system] - option selection [user] - answer/action [system]. There exist many different examples of systems of this type, from the sophisticated - such as voice activated menu navigation on phone lines - to the simple - such as command line interfaces. These systems can be frustrating for the user, due to their inherent functional limitations.

In an alternative model it is the user that takes control of the interaction, with the system playing a far more passive role. Systems of this type usually wait for the user to ask a question, for example: “What time is the next departure for Tahiti please?” - “Check -in for the next flight to Tahiti is 5.00 pm tomorrow”. The disadvantage to this type of system is that the user is often unaware of the capabilities of the interface. It is also easy for the user to stray “off topic” to subjects beyond the systems expertise where they will be frustrated with “Sorry, I don't understand” style defaults.

There are also systems that allow the conversational initiative to be spread between the user and the system. In systems of this type both the computer and the user actively participate in some task-oriented interaction to achieve some objective. Interactions of this type are much closer in style to human-human conversations and human-human conversations often form the inspiration for the dialogue models on which these systems are based. These interfaces can be described as “true”, due their close resemblance to natural conversation. This project concentrates on true conversational interfaces throughout, although many of the findings can be applied to system or user led interfaces.

This project also concentrates on conversational interfaces that use typed text as input and output, not on interfaces that use speech recognition and synthesis as input and output. The interface implemented and tested could, however, form a rudimentary dialogue model for a more sophisticated spoken language dialogue system.

1.3 Conversational Interfaces Online

Most commercial web sites use either a menu driven or a keyword search style interface to allow customers access to the products or information that the site is trying to sell. In most cases (all but the smallest and simplest) these methods of navigation are generally inadequate. Navigating through a site using menus can be a time consuming and frustrating task, the categorisation of products into different sections and subsections can seem unintuitive to users and can increase the time wasted by a customer trying to find a product. Menu driven navigation can also increase the interaction path a user has to follow to find a target product.The length of an interaction path - measured in mouse button clicks - is inversely related to users continued interest in a particular site. Huberman et al. (1998) showed that users interest in a site decreases exponentially with an increase in the length of the interaction path. In other words, the more mouse clicks in an interaction path, the more bored users got using a site. It is therefore important to keep the interaction path as short as possible to maintain users interest in the site, this can be difficult to achieve using menu style navigation especially where the site offers a wide range of products.

Keyword search, usually combined with some kind of menu interaction, is the navigational method used by most large commercial sites. This method also has many disadvantages. The main disadvantage is that keyword searches often fail to recognise what a customer has in mind even when the customer is looking for a specific target product. This is primarily due to assumptions made about the users level of expertise or experience in the area of the product they are searching for, users must know domain specific jargon to have a hope of finding the product they want, or are most likely to buy. Keywords are unable to precisely describe the users intention and keywords used by the user might not appear in the catalogue entry for a product that a user would be more inclined to buy. For example, a keyword search system does not understand that a search for “summer dress” should list womens clothing but a search for “dress shirt” should list men's clothing (Chai et al. 2001). There is also the possibility in keyword searches of returning no suggestions to a query, effectively shutting the door in the customer's face even when there are appropriate products in the site.

Natural language style interfaces could be used to side step a number of the navigational problems outlined above, especially in a commercial environment. NL interfaces have been used, to varying degrees of success in applications for call-centres (supplying dynamic scripts for the call center staff, for example, increasing reliability and saving time), e-mail routing, information retrieval/database access, and telephone banking, but surprisingly little in areas of e-commerce. Where natural language interfaces have been used to access information on the web they are rarely complete dialog or conversational systems (for example the user led dialogue style of sites such as www.ask.co.uk). A complete dialog with the user can be used to refine their search and discover exactly the intentions of the customer, also allowing the system to make suggestions in a conversational manner more like a helpful shop assistant guiding a customer through any decisions they may have to make before purchasing an item.

In a recent study (Chai et al. 2001) it was shown how useful a conversational style navigational interface could be in an e-commerce environment. They built a proof of concept prototype ("The Happy Assistant") that directed customers towards appropriate webpages that sold the product or service they had requested in natural language. The system identifies key concepts, rather than keywords, from the users input and either initiates more dialog to clarify a request, or displays the appropriate page. They found that twice as many users preferred the conversational NL system over menu based navigation and that this preference was stronger in less experienced users. They also showed that the length of the interaction path was reduced by over 60% in the NL interface and the total real time taken to find a suitable product was reduced by over 30%.

Chai et al. also found that the prevalence of keyword search navigation on the web has created a "search culture" amongst internet users. Users are used to typing key-words and short phrases in order to find what they are looking for, this effected how users carried out dialog with the interface, using short and linguistically simple statements (the average length of user input was 5.3 words and 85% of user input was in the form of noun phrases). They concluded that the use of shallow parsing techniques was adequate for processing user input and that more attention should be paid to the dialog management side of such a system rather than concentrating on the systems ability to understand and handle complex natural language structures.

A NL style interface is not without its disadvantages. It is harder for users to browse through a site when a NL interface is the only navigational method available, this could dissuade visitors from making purchases. The addition of a menu based alternative is a feasible method of meeting the different requirements of different users. NL systems are also difficult to create and any information contained within them, or the knowledge base they operate on, can be difficult to update or change.

> "Virtually *all* interfaces have a personality. This literally applies to anything that presents words to a user, from toaster ovens and televisions to word processors and workstations." (Reeves and Nass 1996)

Personality is relevant to all interface design, but perhaps especially so in

NL interface design where the systems language use plays such an important role on a users appraisal of the interface. This implies the NL interface designer should have some understanding of personality theories and interpersonal communication.

1.4 Personality Traits and Trait Theories

> "In everyday life, no one, not even a psychologist, doubts that underlying the conduct of a mature person there are characteristic dispositions or traits" (Allport 1937)

Personality is that which makes us individual, however it has long been argued that all people share the same fundamental dimensions of personality. The scientific study of these traits began at the start of the twentieth century when modern statistical techniques (specifically factor analysis) could be used to form a complete taxonomy of personality. Contemporary views of personality traits start with Raymond Cattell who proposed that personality is made up of sixteen "primary factors" (PFs) such as: calm, venturesome, shrewd, radical etc... (Cattell 1991). Cattell's 16 PF model has lost favour with most personality theorists and has generally been replaced with personality models using orthogonal, high level, secondary traits. Of these higher level models the two most widely accepted use three and five independent traits to describe personality.

The three factor model proposed by H.J.Eysenck describes personality as being made up of three broad dimensions: neuroticism, extroversion/introversion (or surgency), and psychoticism (Eysenck and Eysenck 1991). The Eysenckian model of personality has generally been over taken by five factor models, mostly due to findings suggesting that the psychotisicm dimension is too wide and should be split to form a five factor model. The standard five factor model of Costa and McCrea (1993) is the most widely accredited in personality psychology. The trait facets associated with their five factor model are: neuroticism, extraversion, openness (intellect), agreeableness and conscientiousness. This five factor model is sometimes known as the OCEAN of personality. Costa and McCrae (1993) have it that:

> "The five factor model has provided a unified framework for trait research; it is the Christmas tree on which the findings of stability, heritability, consensual validation, cross- cultural invariance and predictive utility are hung like ornaments."

Five factor models are widely agreed upon because similar conclusions have been made by several different studies in several different areas of research. Findings from lexical studies showed that adjectives used to describe personalities can be clustered into five categories (using factor analysis techniques), while results from questionnaire studies also pointed towards a five factor model. Furthermore, the robustness of the model is supported by cross cultural studies as well as cross gender, generation and occupation.

Most trait psychology researchers are beginning to agree on certain aspects of the field, although there are still controversies over the three or five factor models and their exact dimensions. However, it is generally agreed that extraversion and neuroticism should be factors in any trait theory of personality (Mathews and Deary 1998).

1.5 Interpersonal Communication: The Law of Attraction

The principle that people with similar personalities will prefer interacting with each other has such standing amongst psychologists that it has been called the law of similarity attraction (Byrne and Nelson 1965). Studies involving the use of personality trait theory to judge interpersonal similarity and therefore attraction have shown that, for example, roommates liked each other more when they shared personality traits (Deutsch, Sullivan, Sage and Basile, 1991) and people are more attracted to strangers that share similar personality traits (Byrne and Griffit 1969).

Along with the similarity attraction principle there are many other theories of interpersonal communication: The principle of complementarity (Leary 1957) shows that people tend to try and balance the power relations in an interaction, trying to balance dominance and friendliness throughout. This idea is similar to accommodation theory (Giles and Powesland, 1975) which is concerned with the divergence and convergence of personality cues (such as gesture, gaze and verbal style) during interpersonal interactions. Although these studies show that people may indeed attempt to alter their style of communication depending upon the social context, they do not contradict the similarity attraction principle, indeed they may show why people prefer those of a similar personality, demonstrating that it requires less mutual cognitive effort to communicate effectively with similar personalities.

A generally accepted idea in personality and interpersonal communication research is the importance of personality consistency. People appreciate interacting with consistent personalities (Isbister and Nass 2000). This is not only a guiding factor in character design (for traditional media such as film, television and written fiction), but also in standard GUI design. The need for personality consistency in character design has also been supported by psychological findings. For example, it has been shown that discrepancies between gesture and speech are used by people as a clue to possible deception (Ekman 1974), sequential consistency in behaviour is just as important, allowing people to predict what will happen over the course of an interaction thus decreasing cognitive effort.

1.6 HCI and Personality: The Media Equation

> "The similarity attraction principle is so powerful that interface designers can increase positive evaluation of a product and company by matching personality and user" (Moon and Nass, 1996)

Reeves and Nass, in their book "The Media Equation: How People Treat Computers, Television and New Media like Real People and Places" (1996) show how people behave in social ways when interacting with media (especially computer interfaces). Their "computers are social actors" (CASA) theory showed that people exhibited social norms and conventions when dealing with computers including displaying Gricean politeness maxims to interfaces and reacting in different ways depending on factors such as the gender and the personality type the interface was exhibiting. They showed this by applying standard psychological theories from human-human interactions - such as the similarity attraction principle (see section 1.5) - to human-computer interactions. People would react to an interface's personality in a similar way to when reacting with a person of that personality. Through applying the law of similarity attraction they showed that extraverts preferred an extraverted interface and introverts preferred an introverted interface. Users would evaluate interfaces with similar personalities positively as regards social acceptability, friendliness and utility. They concluded that an interface that could match the personality of a user would be evaluated in a better light than a interface with a contrasting personality.

1.7 Inferring and Imitating Personality from Speech and Text

People are adept at recognising personality in others, people need to understand others well enough to manage and ease social encounters, this is known as "Attribution Theory" Heider(1958). People identify the personality of others using many different clues which can be broadly categorised into three types: non-verbal, co-verbal, and verbal. Non-verbal clues include: gender, animation, facial maturity and attractiveness. Co-verbal clues are concerned with the gestures that accompany normal conversational communication such as hand movement, change in gaze direction, facial expression and other non-verbal information conveyed through speech such as the acoustic properties of speech such as pitch, volume and rate.

In trying to artificially imitate distinct personalities through text, there are two broad approaches. Firstly one can take a less linguistic, behavioural approach as demonstrated by Reeves and Nass (1992). In their experiment they changed the personality of the text interface in four ways: the extraverted system would initiate the interaction (i.e. take the first turn) theintroverted one would wait for the user to make the first move. They also displayed a measure of how "confident" the response was; extraverted statements had high levels of confidence (above seven) and introverted statements low (below five). They also named the interfaces, one was called Linus (a supposedly submissive name) and one was called Max (a dominant name). They also changed the phrasing of the systems ourput, in a fairly arbitrary manner, with the extraverted system mainly using commands and the introverted system using suggestions to put their respective points across to the user. With these relatively minor changes to an interface they produced their convincing results.

Alternatively one could use more empirically derived linguistic tactics to get an interface to imitate a personality. Several studies have tried to identify the personalities that listeners infer from the language that speakers use (Bradac 1990, see Krauss and Chiu for review). In the surgency dimension of personality the main finding is that perceptions of extraversion increase with utterance length and that those that initiate conversations tend to be more extraverted (Palmer 1990). There have also been many studies trying to infer personality from written language style using word frequency analysis and other statistical techniques. Extraverts have been found to use more positive emotion words (Pennenbaker and King 1999) and introverts more hedge phrases and tag questions (Berry et al. 1997). Most of these studies are limited by their lack of consideration for context and the difficulty in recognising the complex stylistic protocols of written and spoken language.

A recent study of personality and language use in emails (Gill and Oberlander 2002) used bigram analysis to extract features of extraverted and introverted language use. The study showed that extraverts tended to start emails with "hi" instead of the apparently introverted "hello"; used less quantifiers; had a more relaxed and informal style; used less self reference (specifically first person singular), and were more positive in their style (less negations and more optimistic language such as "... looking forward ..." and "... a good ..."). Extraverts were also more confident, both in their ability (using "... want to ...", "... need to ...", and "... able to ..." as opposed to the more tentative "... trying to ..." and "... going to ...") and their certainty (using "... will be ..." rather than "... should be ..."). Introvert and extravert also differed in their grammatical style, with introverts tending to use the conjunctions "... , and ..." and "... , but ..." and extraverts tending to use the subordinating conjunction "... , which ...".

There are major stylistic differences between written bodies of text (including emails) and actual electronic conversations. A computer mediated conversational interaction (also known as an electronic conversation and colloquially as "chat") is a fairly novel method of communication but one that is growing in popularity (examples include: Microsoft Instant Messenger, chat rooms, IRC and teleworking and remote education applications). A computer mediated interaction is a typed conversation between two or more people where the interactees can instantly see each others response (i.e. there is a minimal delay between responses, unlike email). The interaction is more closely matched to spoken dialogue than written texts, such as letters and emails but findings from research in both these areas could feasibly be used to project extraversion in a natural language interface. A variety of studies have been carried out on electronic conversation, they have highlighted the fact that language is shaped by the medium in which it is used as well as the grammar and cognitive processing of the interactees. Written language is formally quite different from spoken language and electronic conversations or "visible conversation" [Brennan, S.E. 2000] due, on the surface at least, to their messy, disfluent, interactive style. However, it is justifiable to use results taken from written corpora as the electronic conversation is still written language, and, aspects of language use and personality seen in written language should still be in evidence in visible conversations.

1.8 Projecting Personality

Their are many choices to make in deciding how to project personality, below is a summary of these choices along with some justifications of the feasibility and appropriateness of incorporating them into the interface.

Lexical and grammatical

- Intensifiers and emotion words (negative and positive): Wide use of emotion words and intensifiers such as very and really may be an indication of introverted personality (Pennebaker and King 1999).

- Quantifiers: Extraverts use less quantifiers, introverts use quantifiers such as "a lot", "a few", "lots of" etc... with higher frequency.

- Negations: Extraverts tend to use more optimistic language, using less negations.

- Confidence: Extraverts tend to express themselves in a more confident manner using "want to", "need to", "able to" and "will be". Introverts use less confident structures such as "trying to", "going to" and "should be".

All the above characteristics of extraverted and introverted language use can be used, with some ease, to manipulate personality.

Turn

An extraverted personality will tend to initiate a dialog (Palmer 1989). Due to the nature of the background technology used to implement the interface, this potentially useful fact could not be implemented. In a real world system who takes the first turn may be dependent on the function of the interaction and thus not available for manipulation.

Confidence Scale

A confidence scale (Reeves and Nass 1992) could be used to indicate how confident the interface is in its response to a users query: the extraverted interface would be more confident in its responses than the introverted interface. However, a truly conversational interface would not practically be able to use this method. A confidence scale could confuse the user and would probably not be used in a real world implementation of a natural language interface. A confident lexical style can be used by an extraverted interface by employing other, linguistic, techniques, as described below.

Name

Giving the interface a name could again confuse the user in a real implementation of a conversational interface. No name should be given to the interface, the user should not be under any illusion that they are talking to anything other than a system.

TTR

Type-Token Ratio (TTR) is a method of linguistic analysis based on the ratio of different words - types - to the total number of words - tokens. Although TTR measurements have some personality correlates there is some dispute as to the measurement technique and its use (Berry et al. 1997, Bradac 1990), it would be very difficult to implement a method controlling TTR in an interface.

Utterance Length

Extraverts are inclined to talk for longer than introverts (Palmer 1989). Utterance length could easily be used to project either extraverted or introverted personalities: the extravert talking for longer than the introvert.

Hedges

Hedge phrases are of the form: I may be wrong but, I don't know if this is interesting/relevant but, etc Hedge statements are words such as maybe, might, possibly, etc Hedges are a sign of tentativity and lack of confidence in what is being said. An introverted interface would use more hedges in the dialog whereas an extraverted interface would use less or possibly none at all (Berry et al. 1997).

Tag Questions

Tag questions are forms of hedging and consist of questions added to the end of statements such as "ok?", "yeah?", "really?" or "don't you think?". Tag questions are, again, a sign of tentativity and lack of confidence in the speaker. An extraverted interface would use few tag questions, an introverted interface many.

1.9 Summary of Research

Reliable conversational interfaces represent a very important step forward for computer interface design, and conversational interfaces are also essential to the predicted spread of ubiquitous computing. The personality projected by these interfaces will play an important role in users appraisal, and acceptance of these systems. An undersatanding of personality theories, principles of interpersonal communication and how personality can be projected is thus an important aspect of conversational, as well as NL, interface design.

Chapter 2

ALICE, ProgramD and AIML

ALICE is an implementation of a web based “chat-bot”; a system designed to converse with the user, on any topic, in as real and believable a manner as possible. ALICE is implemented in the non-standard, evolving AIML (“Artificial Intelligence Markup Language”) XML specified markup language. AIML is interpreted by ProgramD, which also acts as an http server (built around the open source Jetty Java http and servlet server). ProgramD itself is one of the very few open source “AI” projects in development, it has a wide developer community and a code base that stretches back five years, it is also very popular, being downloaded on average 5,000 times a month. Various high profile “bots” have been created using AIML and ProgramD including: Maddy, an avatar which appeared on the BBC’s “Tomorrow’s World” (March 2002); the interface that appeared on the “AI:Artificial Intelligence” film (2001) web site; ALICE itself which won the Loebner prize in 2000 and 2001 (the Loebner prize is an annual Turing test style competition) and generated international press attention. Along with these high profile “bots” are a number of commercially developed interfaces that use the same technology, these interfaces are mainly used as general information providers (sometimes known as FAQbots), sales assistants and even automated helpdesks and tutors. In addition to these commercial uses there is much amateur enthusiasm for the project with hundreds of chat-bots online, this popularity is due to the simplicity of AIML and the availability of a PHP/mySQL version of ProgramD; ProgramE.

2.1 History

ALICE forms part of a long line of performance driven natural language “engines” that stretches back to Weizenbaum’s ELIZA (Weizenbaum 1966) via Colby’s less well known but much more successfull PARRY (Colby 1971). PARRY acted the part of a paranoid hospital patient, it was built around a large set (about 6000) of patterns which would match any input. PARRY was robust and always had something to say and its status as “insane” probably helped the user forgive a lot of its problems. PARRY and ELIZA’s success, and subsequent fame, has been attributed to an effect known as the “Weizenbaum illusion”;

users are very willing to attribute high levels of intelligence to even the simplest of machines, as long as the output is in some way "believable". ELIZA also represents the start of a split in natural language interface research between theoretically motivated models, such as Winograd's block's world conversation generator (Winograd 1972) and performance led systems such as PARRY. This dichotomy exists today. Theoretically motivated models that use (mostly) symbolic reasoning and deep understanding forms the majority of academic research in this field. Performance led models, often incorporating no syntactic analysis and very simple pattern matching tactics, form most of the commercial, and rapidly growing amateur, interest. The performance of these models is rapidly improving, mostly due to commercial interest and general implementational simplicity, as well as the lack of any need to justify the cognitive viability of their models.

ALICE is an instance of an ALICEbot. ALICEbots form a family of "bots" that share the same simple underlying technology. The driving force behind this ALICEbot technology, is one of minimalism. ALICE shares this "big data, small program" (Wilkes and Catzione) philosophy with many other similar performance led interfaces including CONVERSE (Wilkes and Catzione) and CARTMAN (Boone 1999). ALICE is, however, even simpler than most other interfaces in this category being driven purely by top-down pattern matching with no syntactic parsing whatsoever.

The suprising performance of ALICE and ALICEbots can be attributed to three main factors: Firstly, because of the evolutionary and open source nature of the project, the developers are acutely aware of the environment in which the ALICEbot exists - the unfriendly user (De Angeli et al 2001). Because of the relative maturity of the project, as well as its popularity, many features (including the Herculean task of programming the AIML) have been added to enable ALICEbots to interact relatively robustly with users (as a measure of this the average number of turns per user has grown from three to eight in the past five years). Secondly, because the AIML responses are written by the designers as responses to what they themselves would say (given the presence or absence of some context), the language used by the ALICEbot in a turn is very convincing, even if the dialogue as a whole is not. Thirdly, the developers make no claim to be following any cognitive theory in their system, this means that any design decisions made by the ProgramD developers is purely one of engineering, in an attempt to improve the performance of the system.

Because of the server like nature of ProgramD (see fig.1) ALICEbots can be implemented either as stand alone local applications, or as a web based applications, accessed via a browser. Because the ALICEbot has its own dedicated server it is also generally fast and the user (client) does not need to install any additional software or browser plug-ins to talk with the interface. ALICEbots can handle multiple users and are generally reliable (again thanks to its open source development). The AIML used to store the patterns can be updated dynamically, and ProgramD outputs useful information for the ALICEbot's developers use, such as the user inputs it had to resort to defaults for, allowing the operator to "evolve" the AIML continuously. AIML is also extensible and customisable. Javascript and Unix system commands can be embedded in the AIML, meaning that the ALICEbot can form the interface to a variety of applications, not just act as the front end to a knowledge base.

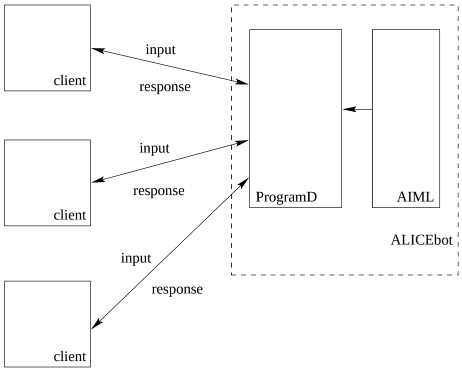

Fig.1 An ALICEbot's client/server structure.

2.2 Writing AIML

This basic top-down pattern matching principle of ALICEbots is extended by including:

- The ability to tailor the responses according to the previous (one or two) responses adding some context to the dialogue.
- The ability to use "topics" again adding context to the dialogue.
- Memory, the system can remember certain key utterances of the user

The best way to demonstrate how AIML incorporates these features, as well as to highlight its simplicity, is to give an example dialogue and show how AIML can be used to implement the dialogue:

- `user> knock knock.` [user initiates interaction]
- `agent> who's there?` [simple pattern matching, system sets topic to <joke>]
- `user> ivana.` [system stores response]
- `agent> ivana who?` [response due to preceding dialogue, system retrieves user's last words]
- `user> ivana half a billion dollars.`
- `agent> that's not very funny.`[response due to preceding dialogue]
- `user> why don't you tell me a joke.`
- `agent> do you want to hear a joke?` [default response for <joke> topic]

This fairly complex interaction could be implemented using the following AIML:

```
<category>
<pattern>KNOCK KNOCK</pattern>
<template>
  Who's there?
     <think>
       <settopic>joke</settopic>
     </think>
</template>
</category>

<!-- knock knock joke recognised so topic set to joke.
think tags used to hide template side AIML from user
-->

<topic name=''joke''>
<!-- start of joke topic -->

<category>
<pattern>*</pattern>
<that>WHO IS THERE</that>
<template><star/> who?</template>
</category>

<!-- * wildcard picks up users response, must be in a
knock knock interaction due to joke topic and ''who is
there'' as previous response.  <star/> is an
abbreviation of <star></star>, whatever pattern the
wildcard matched to will be placed here -->

<category>
<pattern>*</pattern>
<that>* WHO</that>
<template>That's not very funny</template>
</category>

<category>
<pattern>*</pattern>
<template>Do you want to hear a joke?</template>
</category>

<!-- default for joke topic -->
<!-- etc etc etc -->

</topic>
```

The above categories use a small sub-set of the tags available to AIML programmers, but sufficiently show the general structural outline of AIML.

As it is this AIML set is not very robust; it can only handle a very limited interaction. It would not be a complex (though perhaps a lengthy) task to extend the AIML into a more reliable "comedian". This small example shows the simplicity of AIML but also its potential to give surprisingly good feedback to the user. The obvious drawbacks to AIML is the enormous number of categories needed to get anything like an apparently natural conversation (ALICE currently runs with 40,000 categories), and the possibilities of complex discourse planning faced by the designer. These impracticalities, however, can be lessened by reducing the space of the dialogue by using AIML as a tool to write very specialised applications, and by designing the discourse so that the user is encouraged to stay 'on topic' as much as possible (by, for example, insightful use of defaults etc...).

2.3 Implications of Using an ALICE Style Interface

Implementing a prototype conversational interface using ProgramD and AIML has several advantages:

- ALICE style interfaces have proven to be usefull and successfull, they have also been adapted for commercial purposes.
- AIML itself is simple, yet powerful enough to make a convincing beta interface that would allow for experimentation.
- ProgramD's server like structure means it will be simple to implement a browser based user test.
- ProgramD is one of the most popular amateur, purportedly AI, systems available, it is also open source.
- ProgramD is Java based and platform independent.
- ProgramD is robust and reliable.
- ProgramD has many log keeping facilities built in.

The main Disadvantage to using ProgramD and AIML is:

- The success of the interface depends on the size of the AIML, this is potentially very large, Programming the AIML is a time consuming and laborious task. The size of the AIML could be minimised by keeping the area of dialogue as small as possible and by careful dialogue design.

Chapter 3

Implementation

It was decided to implement a conversational interface, using AIML and ProgramD, to act as a testbed for the following experimentation. The actual implementation of the interface was a non-trivial task in itself and an outline of the methods used to create the interface are detailed below.

The system was implemented to act as an interface for users to solve some task oriented problem in a small but real domain. The domain chosen was a sub-set of information describing the various talk plans and products available from One2one mobile. The subject of mobile phones was chosen because the area contains a good balance between the amount of information users may want and the complexity of that information. The style of the information allows for rich dialogue on limited data allowing the interface to take the role of an adviser rather than a as a simple "information point". Using real world data and real world problems also allows for some insight into what commercial interfaces must be capable of. Participants in the experiment may also be more inclined to give an honest critique of a system they could actually see themselves encountering. One2one was chosen because the information on its website was well ordered and the talk plans available were limited. The One2one website also contained its own conversational style interface which allowed an appraisal of the standard of performance expected from a real world commercial interface of this type - unfortunately the interface has since been decommisioned.

In its role as an automated advisor the system will act similarly to the Desert Survival Problem advisor used for a number of studies into interpersonal communication, for example, Moon and Nass (1996). In that experiment the interface advised users as to the advantages and disadvantages of certain pieces of desert survival equipment. This is undoubtably an artificial scenario; it is more interesting to engineer an interface that users may well actually encounter (or at least have some previous experience of the problem at hand). In the desert survival problem the user is likely to be well behaved, that is, if the user is told that a bottle of water is more useful than a bottle of gin he will have no option than to believe the interface (there seems to be very little scope for argument and therefore richer dialogue). Users will also be more honestly critical of the system if they are told they are performing a usability test of an actual commercial interface. Moon and Nass's interface did not act as true conversational interface, indeed system output was not in any way contingent on user input.

It is hoped that by making the interface seem more like an impartial advisor than a salesman, some contextual personality effects will be reduced. For example, in a sales-customer discourse users may lean towards an extraverted personality style in making judgements about the utility and believability of the agent, assuming that extraverts make better sales people, whereas in a collaborative task such as recommendation this contextual bias should be reduced.

3.1 Dialogue Modelling

The main design consideration during implementation was how to plan the dialogue. It was decided to put as much effort as resources would allow into the design of the dialogue. A good robust design would mean less time testing and re-implementing the interface, as well as providing an excellent blueprint to write the AIML files. The main source of the choices made whilst planning this dialogue was the structure of the information on the One2one website (the website is no longer online, a similar information structure can be found at http://www.tmobile.com - One2one is now T-Mobile). The menu based navigational structure of the website was fairly easy to map onto a dialogue plan, for example, when a customer wants to browse the various talk plans available, the first menu he comes across is a choice between pay as you go and pay monthly talk plans. Thus in the conversational interface; the user is first asked if he would prefer a pay monthly or pay as you go setup. The conversational interface then asks the user to consider this choice by presenting them with the pros and cons of each setup, a feature missing from the website but one which a conversational interface lends itself to very neatly. The FAQ section of the website was also modeled to provide more general information for users. The design of the system was also influenced, inevitably, by the designers own experience and intuitions.

The completed model can be seen in figs 1 and 2. The model is graphically presented as a state chart with user input - which change the state of the interface - represented on the edges, and the responses of the system represented at the nodes. Both user input and system responses are paraphrased in the graph. A user can say "yes" in many ways, this is not represented in the graph, however the system itself will match many variations on the "ideal" users input. In this way the system is able to come to some conclusion of the users intentions and change state accordingly. The "*" symbol in the graph is a wildcard symbol and represents either that the next utterance of the user is ignored by the system, or that the system does not recognise the users input - as can be seen when the system has to resort to defaults.

The graphical representation of the dialogue model represents the complete specification for the interface and the AIML. As well as providing a basis for structuring and planning the laborious implementation of the AIML. Formally specifying the interface in this way cut development time and provided a basis for white box testing of the system (see section 3.2).

There are four "topics" implemented in the interface - shown as dashed boxes - one for each of the talk plans and one for the pay as you go sub dialogue. A topic gives the interface some added information about the context of the users input. For example the question "How much does it cost to call a land line off-peak?" when in the Everyone topic, will produce a different answer than the same question asked in the Anytime topic. Answers can be given if their are no matches on topic (as with the pay monthly defaults). Topics can be nested, a powerful feature, and can even be implemented in a stack. It was not necessary to use either of these features in the interface.

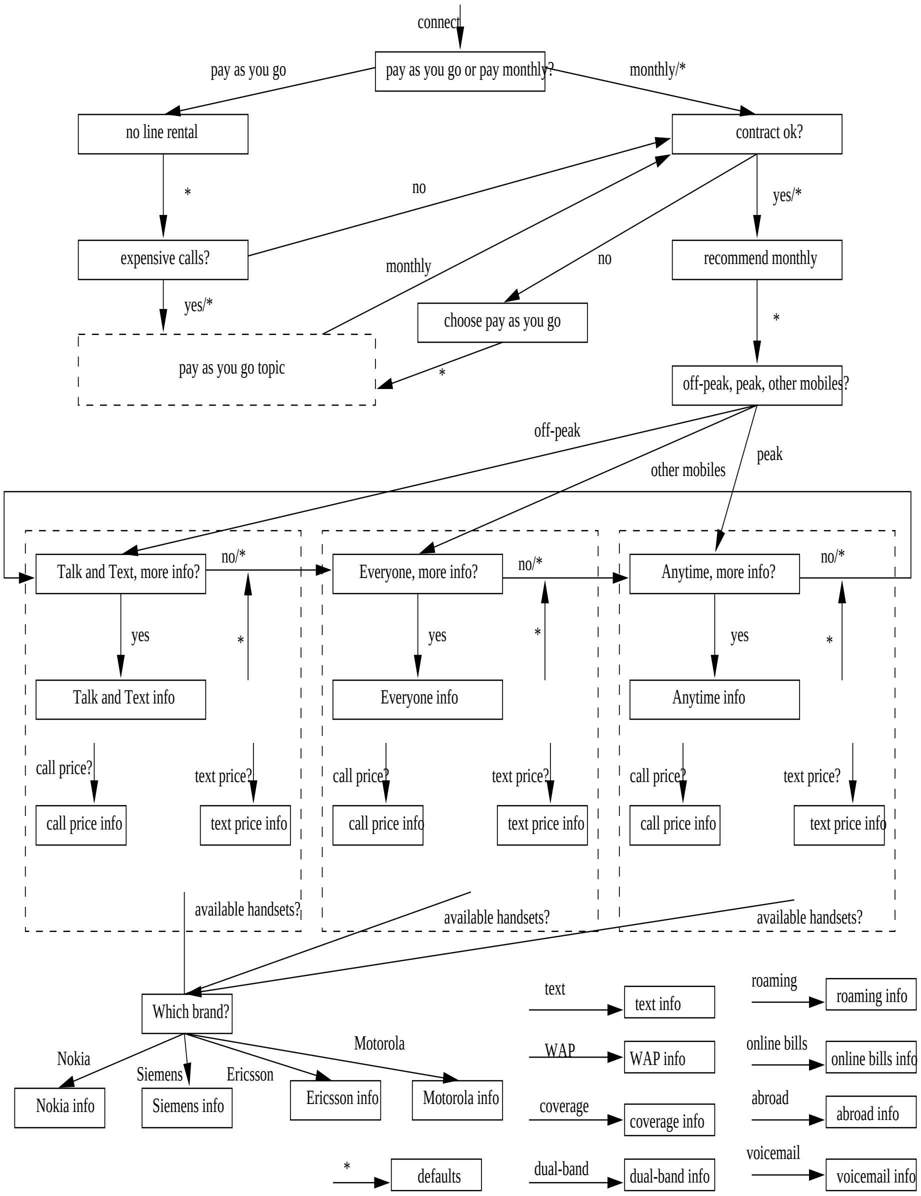

Fig. 2 State chart representing the main features of the dialogue model implemented in the interface.

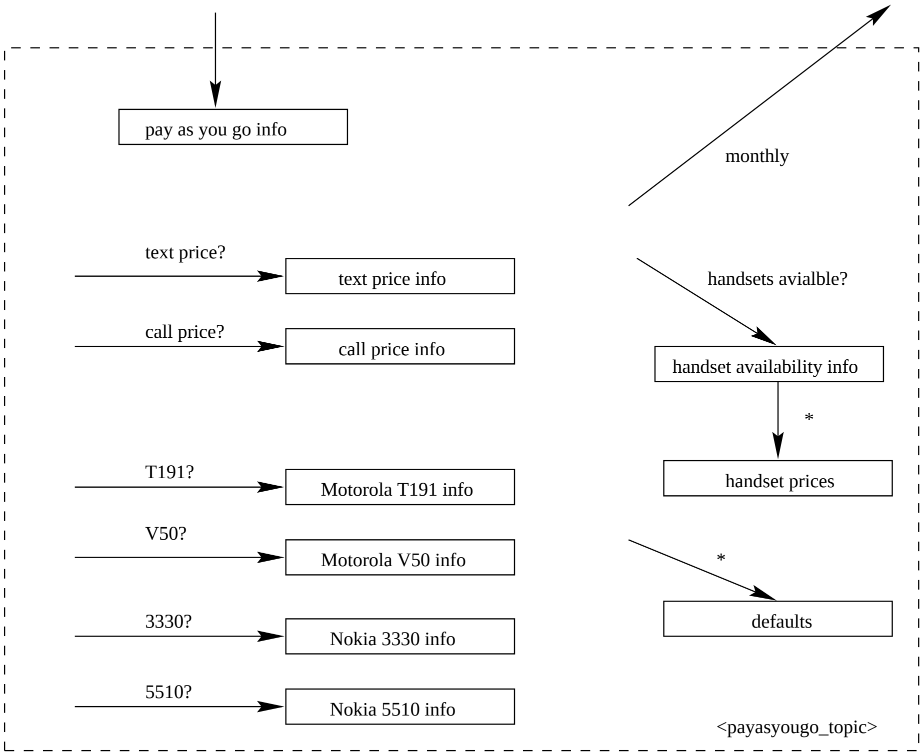

Fig3. State chart representing the "pay as you go" sub-dialogue, implemented as a topic.

The dialogue models represent a true conversational interface because the dialogue initiative is spread evenly between user and system. Although the system takes control during the early stages of the interface, the user can at any time ask, for example "What is WAP" and be presented with an informative reply.

3.2 Testing

Ideally the interface should have been fully tested with a wide group of users before the experimentation began. A commercial interface of this type would normally go through many implement-test-revise iterations before going "live". The system implemented here was only supposed to act as a prototype interface, whose real purpose was to portray personality and "extract" dialogue from the users. There was however, some limited testing where three volunteers where asked to "try and break it" (which could be construed as a form of stress testing), the interface generally behaved well under these conditions and the testers gave mostly positive feedback. The designer also tested all aspects of the system's ability to traverse the dialogue model. The scenarios used in the experiment (see section 4.1.3) were designed to reflect both the ease and difficulty of using the interface to navigate through the dialogue model.

3.3 Adding Personality

Personality in the interface was projected in two ways, using either general linguistic style and or lexical choices to manipulate personality. As for linguistic style the introverted interface used more tag questions, hedge phrases and had a shorter average utterance length to reliably project its personality. The extraverted interface used no tag questions or hedge phrases and also used lengthier utterances to deliver the same information (justifications for these decisions are discussed in section 1.8). Brennan and Ohaeri (1999) found that during a typed discourse between collaborators hedge phrases were used approximately once per hundred words, equivalent to approximately once every ten turns. Therefore, to model a reliably introverted personality (while still being believable) a hedge phrase every four or five turns was deemed sufficient. Tag questions were added at a similar frequency.

In the same research into electronic conversations Brennan and Ohaeri found that the average number of words per turn was 8.3 (these were human- human electronic conversations, unlike Chai et al. 2001 where users were found to use on average 5.3 words when talking to an electronic agent). In practise it was found to be very difficult to keep responses down to eight words. In the actual system the extraverted interface used, on average, 20 words per turn, the introverted interface 15 words per turn.

Lexically the two interface styles differed in the amount of intensifiers and emotion words used in the interaction. The introverted style used more intensifiers and emotion words than the extraverted interface. The exact amounts were somewhat subjectively chosen, this was to ensure that the interfaces, although exhibiting consistent personalities, acted as believably as possible; too many intensifiers and emotion words would reduce the believability of the interface but too few may mean the personality differences are too subtle to be noticed by the user.

The interaction between the interface and the user should be as long as possible so that the user has sufficient time to recognise the personality of the interface. Ideally the interaction should consist of no less than 30 turns, however, if all the questions of the user are answered and repetitions in the interface start to frustrate the user the judgement of the users on the interface could be biased. Therefore repetitions of responses and therefore user boredom were kept, as much as possible, to a minimum. It was hoped that the interface worked well enough to avoid too much user frustration, however this may be an inevitable consequence of using the simple stimulus response structure of ALICE to base the interface around.

Below are the actual phrases used by the extraverted and introverted interfaces at the beginning of each interaction, along with a comparison of the differences between each utterance. These excerpts form a good sample of the general differences between the two interface styles.

extravert> "Hi, to help you choose the right phone, I need to know if you would prefer to pay as you go or pay monthly."

introvert> "Hello, could you tell me if you would prefer to pay as you go or pay monthly?"

Comparison: extravert uses "hi", introvert "hello"; extravert uses more words; extravert uses "need to", introvert uses "could you".

extravert> "Paying monthly is cheaper if you use your phone regularly, but you must sign a 12 month contract. Do you mind being in a contract?"

introvert> "You may have to sign a 12 month contract, would you mind being in a contract?"

Comparison: extravert uses more words; extravert uses confident "must", introvert uses "may"; introvert says "would you", extravert says "do you".

extravert> "Because the calls will be cheaper, I recommend choosing one of One2one's pay monthly talk plans."

introvert> "Because the calls should be cheaper, I think I would choose one of One2one's pay monthly talk plans."

Comparison: extravert says "will be", introvert says "should be"; extravert says "I recommend" introvert says "I think I would" - more self reference.

extravert>You must sign a contract if you want cheaper calls. If you don't want to be in a contract choose a pay as you go plan.

introvert>If you don't want to be in a contract I think you should choose a pay as you go plan.

Comparison: extravert uses "must", introvert uses "should"; introvert uses first person.

extravert>Calls will be more expensive with pay as you go, so are the handsets. Is this a problem?

introvert>Calls might be more expensive with pay as you go. Would this be a problem?

Comparison: extravert says "will", introvert "might"; introvert asks if it "would" be a problem.

extravert>There is no line rental if you pay as you go, which means it will be cheaper if you rarely use your phone.

introvert>There is no line rental if you pay as you go, which may mean it will be cheaper if you rarely use your phone.

Comparison: introvert - "may".

Chapter 4

The Experiment

4.1 General Experimental Methodology

4.1.1 Participants

Participants were 32 University of Edinburgh undergraduates from the division of informatics (studying Computer Science and/or Artificial Intelligence), average age of participants was 19. Undergraduates were approached as they sat in the undergraduate terminal room. They were asked if they would mind taking part in the experiment, no incentives were given to the participants who were, on the whole, happy to take part. Two interfaces were running on different machines at once, one extraverted and one introverted, allowing two tests to be run simultaneously. It was presumed there would be a fairly even distribution of personalities among the participants (this was indeed the case). Each test lasted approximately twenty minutes.

Users were told they would be participating in a usability study of an interface that used natural language, and were given a "test pack" (see appendix I). They were told that the study was in three parts, a personality assessment, actually using the system and then a short questionnaire.

4.1.2 Personality Test

The personality assessment consisted of a browser based questionnaire (see appendix II) adapted from the International Personality Item Pool[1] . It consisted of twenty questions designed to give a measure of the individuals surgency. On completion of the questionnaire an I.D. number was output, this consisted of a number representing the surgency of the individual (calculated by a small Javascript function) appended to a six digit random number. Participants were asked to write this on the front of their test pack.

4.1.3 The Interaction

Participants were then asked to carefully read the following instructions on the test pack:

[1]http://www.ipip.ori.org/ipip/

"The system you are about to help test is a prototype for an automated online sales agent that uses conversation to interact with its users. The agent will guide you through some decisions you may need to make when buying a mobile phone, as well as answering any queries you have about mobile phone services. It uses One2one's services and service information.

To test the system you will have a brief conversation with it. Please talk with the system in as natural a way as possible by typing into the dialogue box on the screen. The system will respond to each of your inputs. The test should take no more than ten minutes.

Below are three different scenarios, each scenario asks you to play the role of a customer looking for information about mobile phone tariffs and handsets. In each scenario there are three questions that you have to try and answer by talking with the system. If you get stuck on a question please move on to the next scenario by typing reset in the system's window, please write "stuck" as an answer.

After you have attempted each question in a scenario please type reset in the dialog box, this will reset the system.

If you have any questions or problems please ask the test supervisor."

Note that the participants are told they are testing a real prototype for an automated online sales agent. The participants were presented with a very simple interface contained within the familiar setting of a web browser. Users were able to see their previous input, the last response from the interface, and a highlighted text box within which they input their dialogue.

Below are the three scenarios and corresponding tasks:

" First Scenario

Imagine that:

- You are looking for a *pay monthly* phone.

- You *don't* mind being in a twelve month contract.

- You will mostly use your phone to *call other mobiles*.

Bearing this in mind, please try and use the system to answer the following questions:

1. Find out which specific One2one talk plan would be most suited to you (i.e. "Everyone", "Talk and Text" or "Anytime").

2. Find out how much off-peak calls to land-lines cost on this talk plan.

3. Find out the prices of the Samsung A300 and the Nokia 6210 handsets with this talk plan.

Now type reset.

Second Scenario

Imagine that:

- You are looking for a *pay monthly* phone.

- You *don't* mind being in a twelve month contract.

- You will mostly use your phone in the *evenings*.

Bearing this in mind, please try and use the system to answer the following questions:

1. Find out which specific One2one talk plan would be most suited to you (i.e. "Everyone", "Talk and Text" or "Anytime").

2. Find out how much off-peak text messages cost on this talk plan.

3. Find out if this is cheaper than on other talk plans.

Now type reset.

Third Scenario

Imagine that:

- You are looking for a *pay as you go* phone.

- You *don't mind* paying extra for calls.

Bearing this in mind, please try and use the system to answer the following questions:

1. Find out if you have to pay to receive voicemail with a pay as you go talk plan.

2. Find out if you will be able to use your phone abroad."

Each scenario was designed to test a separate area of the system dialogue, and to give the participants maximum exposure to the language of the interface. The tasks were designed so that within each scenario the user would have to find: details based on the different information given at the start of each scenario (the user is being prompted so that the system can give recommendation style details); general factual information, either on or off topic; Information to allow for comparisons (an operation that is generally considered difficult for natural language interfaces to achieve). The scenarios where intended to increase in difficulty (the first being the easiest, the third the hardest). Task three of scenario three was removed after three tests as the experiment as a whole was taking an impractical amount of time to complete.

Below are brief excerpts from three interactions:

Excerpt 1: An extraverted user talking to the introverted style interface, the user is working through scenario one, part one.

```
//
response> Hello, could you tell me if you would prefer
to pay as you go or pay monthly?
input> pay monthly
```

response> You may have to sign a 12 month contract,
would you mind being in a contract?
input> I don't mind
response> Because the calls should be cheaper, I think
I would choose one of One2one's pay monthly talk plans.
input> which one
response> Could you please tell me if you are going to
be mostly using your phone at all times of the day, at
off-peak times, or for calling other mobiles?
input> calling other mobiles
//

Excerpt two, an introverted user talking to the extraverted style interface, the user is successfully (after some initial problems), working through the last part of scenario three.

//
input> can i use my phone abroad
response> With pay as you go you can choose from a
selection of 4 handsets: The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input> can I use the phone abroad?
response> With pay as you go you can choose from a
selection of 4 handsets: The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input> tell me about using the phone abroad
response> With pay as you go you can choose from a
selection of 4 handsets: The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input> can I go abroad
response> If you have a dual-band phone you will be
able to use your phone while you are abroad.
input> which phones are dual-band?
response> All handsets available on One2one have
dual-band capabilities, which means using your phone
abroad is no trouble.
//

Excerpt three, an introverted user talking to the extraverted interface. This interaction is fairly "off topic" and occurred after all the tasks had been completed. Note how the interface's defaults ensure the user is directed towards a topic (also how the random "what?" default luckily appears at just the right moment).

//
input> hi there how are you
response> Pay as you go phones make good gifts.
input> what do you want to talk about
response> Pay as you go phones make good gifts.
input> what sort of gifts

```
response> The range of handsets available with pay as
you go is limited.
input> how limite
response> What?
input> sorry, how limited is the range of handsets
response> With pay as you go you can choose from a
selection of 4 handsets: The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
//
```

A full appraisal of the general performance of the interface can be found in chapter 5.

4.1.4 The Questionnaire

After the user has completed the personality test and discourse they were asked to complete a questionnaire, on paper, detailing their reactions to the interface. This questionnaire was not completed on the computer to counter Reeves and Nass (1992) findings that users tend to be more generous when typing assessments of an application they just used on the same machine.

The questionnaire consisted of 22 items. Participants were asked to "assess the system you have just used by giving your response to statements describing the system. Please respond to each statement. All answers will be kept confidential. Please take your time and be honest in your assessment". Participants rated each statement on a seven point Likert scale, score 1 being "I strongly disagree", score 7 being "I strongly agree". Score 4 was marked with "I neither agree nor disagree".

Each statement was designed to measure one of four things: Four statements measured the users assessment of the personality of the interface. Seven statements (taken from the Simple Usability Scale) measured the users assessment of the usability of the system. Two statements measured the social attraction of the interface. Four statements measured the intellectual attraction of the interface. Four statements measured the emotional satisfaction of the interaction. Details of the specific statements, their justification and sources can be found below.

4.1.5 Data

Logs of all the interactions were kept, marked with the participants ID number. After all the experiments were finished the logs were normalised and analysed, see below. The scores for the four factors of the questionnaire were also counted, and analysed, see below.

4.2 Hypothesis 1

- H1: Users will detect the surgency of the interface, using linguistic cues to personality .

Moon and Nass (1996) in there study on similarity attraction between subjects and interfaces found that users detected whether their interface was extraverted or introverted with a high degree of accuracy. It seems intuitive that personality can be detected purely by linguistic means. This hypothesis, to an extent, is intended to replicate Moon and Nass's findings. However, there are many differences between this experiment and that of Moon and Nass, specifically:

- In this interface all personality cues are purely linguistic (that is, personality is manipulated by lexical choice and general utterance style alone), not by turn, name or confidence rating.
- This interface is conversational; the user actually inputs natural language into the system which replies in natural language. Each task needs some coherent discussion or conversation (more than two turns) to be completed successfully.
- The users are testing, what they believe to be, a prototype interface of an actual commercial interface.
- The system's responses were contingent on the users input.

4.2.1 Measures

The dependent variables were measured as part of the final "usability" questionnaire. The personality measure was an index of five items:

- "I thought the system was friendly"
- "I thought the system was assertive"
- "I thought the system was cheerful"
- "I thought the systemwas talkative"
- "I thought the system was sociable"

These statements contain adjectives taken from the International Personality Item Pool as well as Moon and Nass's experiment, where they are accredited with being reliable descriptors of personality. The index maximum was 30 (indicating strong extraversion), the minimum was 0 (indicating, weak extraversion or intraversion)

4.2.2 Results

I or X interface	N	Mean	Std.Deviation	Std. Error Mean
Introvert	16	17.25	4.71	1.18
Extravert	16	18.81	3.99	1

t	df	Sig	Mean Dif	Std Error Dif
-1.013	30	.319	-1.56	1.54

As can be seen from the table above no main effect was observed between the means of the personality index for the two separate interfaces with $t(30)=-1.013$, $p>0.3$. The hypothesis that H1, that users will detect the projected personality of the interface, using purely linguistic cues to personality is, therefore, false.

4.2.3 Discussion

Participants did not, as expected, reliably detect the projected personalities of the interface. Possible interpretations of the above results can be split into three categories: those concerning the method of projecting personalities in the interface, those concerning the effects of the medium of communication itself, and those concerning specific aspects of the experimental design.

Linguistic style was the only way in which personality was projected in the interface. There is the possibility that the interfaces simply did not project extraverted or introverted personalities, either the personalities projected were too subtle to be noticed by the users, or the linguistic choices made to implement the two separate personalities were wrong. Reeves and Nass (1996) maintain that even very slight differences in personalities in media generated characters can be detected. If their theory is correct, then there is no reason to think that the personalities of the interfaces were too similar to show any effect. Every effort was made to make the two linguistic styles of the interface appear as extraverted or introverted as possible, all decisions on how exactly to project the two different personalities were taken from empirical evidence as to how different personality types use different styles of language so it is reasonable to believe that the two interfaces were displaying aspects of two different personalities, but that these personalities were not identified by the participants, at least not in the same way as was found in similar experiments by Moon and Nass (1996) in their investigations into the "reality" of computer personalities.

The main difference between this experiment and that of Moon and Nass (1996) with respect to projecting personality was that as well as changing the linguistic style of the interface, they changed certain other aspects of the interface. They gave the two interfaces different mean "confidence scores" associated with the replies they gave, they also gave the two interfaces different names and the extraverted interface initiated the dialogue. The justifications for not incorporating these elements in the interface boil down to an attempt to keep this interface as true to a real, viable commercial interface (more details are given in section 1.6). It is possible then that language alone is not sufficient for communicatees to reliably detect the personality of the communicator. So, although language use may reflect the personality of the communicator it can not, of itself, be used reliably by people to make complex personality judgements. However, as language use is clearly effected by the personality of the communicator, and those effects are in themselves fairly reliable, it seems unintuitive to assume that language use has no effect on personality judgements.

The unusual medium of communication, "visible conversation" (Brennan 1999), exhibited in this experiment is different in style and structure to both written and spoken language. It could be that results from studies of personality and language use in these more traditional medium do not necessarily follow to this specific medium. In which case the language used to project the different personalities would be inconsistent and therefore have no reliable effect on the users personality judgements of the interface. However, as a visible conversation combines styles, structures and conventions from both written and spoken language, personality effects could reasonably be assumed to be at least similar in this medium. It is also reasonable to assume this because written and spoken language are themselves very different and yet there are common findings of personality effects in both these mediums (the increased use of hedges and tag phrases by the introvert, for example).

Any obvious aspects of the interface that would encourage anthropomorphism, such as giving it a name, face or referring to it as he/she, were avoided in the system (although the interface did refer to itself as I during the interactions, the introverted interface did this more than the extraverted interface). The interface was also referred to as a system throughout the instructions given to the participants and the assessment questionnaire, it was hoped that this would reduce any anthropomorphical effects of the interface (which would perhaps have encouraged the users to assume the interface was more "real" than it actually was). It could well be that users were reluctant to describe such an anthropomorphically sterile system with adjectives such as "friendly", "cheerful" and "sociable". In which case, although the participants were not willing to report on the interface having qualities of a personality per se. the different language style used by the interfaces could still have an effect the users reactions to the interface. The reluctance to ascribe a personality to the interface could also be due to the technical nature of most of the participants in the experiment, who are probably less willing than most to treat a computer personality as a "real" personality.

In conclusion, the lack of expected effect could be due to a combination of using language alone to differentiate the personalities of the interface, and the effect of minimising any anthropomorphic effects. The language use of the interfaces could still effect the users behaviour and assessments of the interface, even though the users could not reliably, consciously identify the personality of the interface, this means that hypothesis 2 (below) is not automatically disproved by the above findings.

4.3 Hypothesis 2

- H2: Users will be more attracted and find more usable an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along this dimension.

Again this hypothesis is based on the work of Reeves, Nass, Moon et al. (see section 1.6)They have shown the principle of similarity attraction, from interpersonal communication, to hold in human computer interaction. They proved this hypothesis to be true in many replicated studies. However, they have not proved there hypothesis to be correct in a true conversational interface, such as the one being tested in this experiment.

4.3.1 Measures

4.3.1.1 Usability

The usability index measured how usable the interface was perceived to be by the users. The seven measures were taken from the widely used Simple Usability Scale. The measures have been shown to be robust in many HCI usability tests. These seven were selected from the ten most reliable measures due to there easy application to a conversational interface. The seven were:

- "I think I would use this system frequently"
- "I found the system unnecessarily complex" (reverse scale)
- "I thought the system was easy to use"
- "I think that I would need the support of a technical person to be able to use this system"
- "I would imagine that most people would learn to use this system very quickly"
- "I found the system very cumbersome to use" (reverse scale)
- "I felt very confident using the system"

The index was also measured on a seven point Likert scale, the scores were summed and scaled to produce a score out of 100.

4.3.1.2 Social Attraction

Social attraction was an index of two statements, adapted from Moon and Nass (1996), who found them to be reliable indicators:

- "I liked using this system"
- "I liked working with the system"

These statements are in turn adapted from interpersonal studies of attraction, where the questions are usually of the form: "How much did you like this person?" and "How much did you like working with this person?". The index was measured on a seven point Likert scale. The scores were summed and scaled to produce a score out of 100.

4.3.1.3 Intellectual Attraction

Intellectual attraction measured how intelligent the interface was perceived to be by the users. It was measured as an index of four statements, again adapted from Moon and Nass(1996), who found them reliable indicators:

- "I thought the system was competent"
- "I found the system intelligent"
- "I thought the system was clever"
- "I felt the system was insightful"

The index was also measured on a seven point Likert scale, the results were summed and scaled to produce a score out of 100.

4.3.1.4 Emotional Satisfaction

Emotional satisfaction of the interface was measured as an index of four statements, again adapted from Moon and Nass (1996), who found them to be reliable indicators:

- "I thought the system was boring to use" (reverse scale)
- "I enjoyed using the system"
- "I found the system interesting"
- "I found the system engaging"

The index was also measured on a seven point scale, summed and scaled to produce a score out of 100.

4.3.2 Results

4.3.2.1 Usability

user/interface personality	N	Mean	Std. Deviation	Std. Error Mean
similar	14	49.85	13.69	3.66
not similar	18	54.74	10.97	2.59

t	df	Sig	Mean Dif.	Standard Error Dif.
-1.12	30	.27	-4.89	4.35

Hypothesis H2 can be altered to form hypothesis H2.1:

> H2.1: Users will find more usable (rate an interface higher in terms of usability) an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along this dimension.

It can be seen from the above results, and difference of means analysis, that hypothesis H2.1 is false; participants using an interface with a similar personality did not give significantly higher usability ratings to that interface.

4.3.2.2 Social Attraction

user/interface personality	N	Mean	Std. Deviation	Std. Error Mean
similar	14	54.15	19.26	5.15
not similar	18	56.46	19.7	4.64

t	df	Sig	Mean Dif.	Standard Error Dif.
-0.33	30	0.74	-2.31	6.95

Again hypothesis H2 can be altered to form hypothesis H2.2:

> H2.2: Users will be more socially attracted to an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along these dimensions.

It can be seen from the above table and the analysis of difference of means shows that H2.2 is false; participants using an interface with a similar personality did not find that interface more socially attractive than interface with a non-similar personality.

4.3.2.3 Intellectual Attraction

user/interface personality	N	Mean	Std. Deviation	Std. Error Mean
similar	14	46.47	19.62	5.24
not similar	18	49.35	17.42	4.1

t	df	Sig	Mean Dif.	Standard Error Dif.
-0.439	30	0.66	-2.88	6.56

Hypothesis H2 can also be altered to form hypothesis H2.3:

> H2.3: Users be more intellectually attracted to an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along these dimensions.

It can also be seen from the above table and the analysis of difference of means that demonstrates that H2.3 is false; participants using an interface with a similar personality did not find that interface more intellectually attractive than interface with anon-similar personality.

4.3.2.4 Emotional Satisfaction

user/interface personality	N	Mean	Std. Deviation	Std. Error Mean
similar	14	57.69	13.59	3.20
not similar	14	54.51	16.48	4.4

t	df	Sig	Mean Dif.	Standard Error Dif.
0.6	30	0.55	3.18	5.31

Lastly, hypothesis H2 can be altered to form hypothesis H2.4:

> H2.4: Users be more emotionally satisfied by an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along this dimension.

Yet again, the above table and the analysis of difference of means that shows that H2.4 is false; participants using an interface with a similar personality did not find that interface more emotionally satisfying than interface with a non-similar personality. It should be noted that participants were kept unaware as to the nature of the experiment, and most were surprised to hear of the actual reason for experimentation - users did not see any significance to the personality test administered before the interaction part of the experiment

4.3.3 Discussion

Due to the nullification of hypotheses H2.1 to H2.4, it follows that the corresponding null hypothesis of H2 is true for this interface. That is:

- H2: Users are not more attracted or find more usable an interface that projects, using natural language, a personality that is similar to themselves along the extravert/introvert dimension compared to an interface that linguistically projects a personality that is dissimilar to themselves along this dimension.

The similarity or non-similarity of an interfaces personality compared to the users personality has no significant effect on either the users attraction to that interface or usability of that interface. Their are two possible reasons that may account for the proving of the null hypothesis. Firstly, their was no significant effect because the extraverted and introverted personalities were not projected appropriately. Secondly, the similarity attraction principle does not apply to an interface of this style, because this computer is not being treated as a "real" social actor.

The first possibility is very real, especially when considered in conjunction with the above finding (section 4.2) that users did not accurately report the personality type of the interface. As mentioned above this could either be because the personalities were not consistently "strong" enough for the users to detect them, or that the linguistic styles used to project the personalities were not appropriate to this style of interaction. The resultant inconsistencies in personality would certainly reduce the reality of the interface's personality, thus reducing any effects of the projected personalities on the users behaviour.

A reasonable method for checking the appropriateness of the language used to project the personalities, would be to analyse how the participants themselves used language in the interactions, whether their personality had any significant effect on the language used, and whether or not the language used in the two interfaces is consistent with this. As the linguistic styles used by the interfaces was derived from empirical results from studies of language use in other mediums of communication, this analysis can also be used to show whether these findings are in fact applicable to this type of interaction. An investigation of this can be found in section 4.4.

Due to the reactive nature of the interaction and its similarity to human-human communication, it seems unreasonable to presume that the interface is not being treated as a social actor in any way. It also seems unreasonable to presume that the linguistic differences of the interfaces, which were derived from empirical studies, had no effect on the personality projected by each system utterance. However, personality is projected by the system over the whole interaction. This interaction closely resembles a natural conversation, and so - if computers really are social actors - effects of interpersonal communication, such as complementarity (see section 1.5) would have affected users reaction to the interface. Any inconsistencies in personality, especially in dimensions not controlled for (such as agreeableness), would also have seriously contaminated any personality effects.

4.4 Hypothesis 3

- H3:Extraverted and introverted users will use different linguistic styles while communicating with the interface.

This general hypothesis can be shown to be true by proving the following hypothesis:

- H3.1: Extraverted users used more words than introverted users while communicating during the interaction.

This hypothesis follows from the finding of, amongst others, Palmer (1989) that extraverts talk more than introverts.

- H3.2: Introverted users will use more hedges than extraverted users.

- H3.3: Extraverted users will use more confident language than introverted users.

- H3.4: Introverted users will employ the first person more than extraverted users.

4.4.1 Measures

Some of the most obvious methods used to project personality in the interfaces were:

- Number of words used during the discourse.

- Confidence of language (lexical choices)

- First Person style

- Hedges

These four items were thus analysed from the logs of the interactions. Turn length was measured as the total number of words used by the participant during the interactions. The interactions were also analysed for word frequency; as a measure of confidence words such as want, need and will would indicate a more extraverted style; first person style was measured by analysing the number of "I"'s used by either introverted or extraverted interfaces. Any effects of the personality of the interface should be negated as this variable is spread evenly through the two user groups.

4.4.2 Results

4.4.2.1 Word Frequencies

user X or I	N	Mean	Std. Deviation	Std. Error Mean
extravert	20	156.65	49.65	11.1
introvert	12	192.67	46.66	13.47

t	df	Sig	Mean Dif.	Standard Error Dif.
-2.031	30	.05	-36.02	17.74

Contrary to the hypothesis, introverts used significantly more words on average than extraverted users, with $t(30)=-2.031$, $p<0.5$. There was no significant effect of interface personality.

introverts	N	N/users(12)		extraverts	N	N/users(20)
i	130	10.83		i	129	6.45
pay	75	6.25		pay	110	5.5
no	69	5.75		how	104	5.2
the	68	5.67		the	99	4.95
to	67	5.58		no	99	4.95
how	60	5		much	95	4.75
much	52	4.33		to	85	4.25
plan	46	3.83		talk	73	3.65
talk	45	3.75		you	71	3.55
phone	45	3.75		text	71	3.55
other	45	3.75		are	70	3.5
monthly	42	3.5		yes	66	3.3
ok	41	3.42		is	62	3.1
you	40	3.33		monthly	59	2.95
yes	40	3.33		ok	56	2.8
are	40	3.33		use	55	2.75
use	38	3.17		abroad	54	2.7
text	36	3		what	52	2.6
what	35	2.92		go	52	2.6
me	21	1.75		me	35	1.75
my	25	2.08		my	23	1.15

This table shows the top twenty most frequently used words (also shown as mean words per user). Introverts, as expected, seem to use the first person more frequently than extraverts. Otherwise the frequency of words used seems to be surprisingly similar between the two user groups. The most frequently used words, such as “pay”, “how”, “much”, “talk”, and “plan”, are due to the context of the dialogue. It is also interesting to note the introverts preference for the informal “ok” rather than “yes”, the opposite of the extraverted user.

word	introvert N	introvert N/users		extravert N	extravert N/users
want	9	0.75		15	0.75
need	1	0.08		0	0
will	23	1.92		15	0.75
trying	0	0		0	0
should	11	0.92		6	0.3
going	1	0.08		3	0.15
lot	2	0.17		1	0.05
lots	1	0.08		0	0
few	0	0		0	0

This table shows the frequency of "confident" words ("want", "need", "will"), words that may show less confident phrasing ("trying", "should", "going"), and quantifiers ("lot", "lots", "few"). Introverts were expected to use less confident phrasing and less quantifiers. The frequencies of these words in both user groups are too low for any significant difference or similarity to be observed between the two. A significant lack of hedging was also observed in both user groups.

4.4.3 Discussion

The main conclusions that can be drawn from these results are that:

- introverts used more words per interaction
- introverts used more first person phrases

The first conclusion shows that, in this case, hypothesis H3.1 is false. Extraverts did not use more words during an interaction, in fact extraverts used significantly less. This result is contrary to findings from analysis of both written and spoken dialogue.

Although difficult to analytically prove, extraverted users tended to be more direct in their responses and questions to the interface. For example, an extraverted user might reply to the question "do you mind signing a contract?" with a brusque and to the point "no" whereas an introverted user may tend to use the slightly more polite phrasing "no I wouldn't mind", "no problem" or "contract is fine". Introverts may be talking more, but the extra words are accounted for by a more patient and polite style of language. This is consistent with what would be expected from an introverted personality type. The impatient and less polite one word answers given by extraverts are also consistent with a more forceful, "dominant" personality.

Another possible interpretation of these results is that introverts tend to be more considered in their responses, while extraverts are quicker to react and avoid pauses in conversation. Eysenck notes that "the introvert is more thoughtful than the extravert, taking more head of the maxim that one should be sure the brain is engaged before before putting the mouth into gear" (Eysenck 1971). Were these experiments repeated it would be interesting to time each interaction and note whether the introvert really does take more time to complete an interaction than the extravert, there was no significant difference in the mean number of turns per interaction between groups (the average of all interactions was 40.36 turns).

Observation of the interaction logs also reveals introverts using more informal language, at a "surface" level at least, their is some evidence for this in the introverts preference for "ok" as opposed to the "yes" used by extraverts. Surface formality is characterised by Heylighen and Dewaele (1999) as "attention to form for the sake of convention or form itself", as opposed to "deep" formality which is characterised as "attention to form for the sake of unequivocal understanding of the precise meaning of the expression". In a deep sense the language used by both groups is equally formal, to complete the tasks of the usability test the users tended to recognise that they should avoid obscurity and ambiguity in their language. The general low level of surface formality of the introverts, compared to the extraverts is unexpected but could be explained, again, by the introverts more considered approach to the interaction, the extravert may rush in with rapid formal expressions, the introvert may take a bit more time and add more content to their responses, producing a less formal linguistic style, on the surface at least.

The second conclusion - that introverts used significantly more self referents - shows that hypothesis H3.4 is correct. Introverted participants tended to use the first person style more than extraverts in their interactions. It is assumed that the number of self referents ("I"s) in a users interaction is a good indication of first person style. On observation of the interactions there seems to be a general tendency for introverts to say, for example "Can I use my phone abroad" rather than "can the phone be used abroad" or "how much will I have to pay" rather than "how much is it". This finding is consistent with the general introspective nature of introverts.

Hypothesis H3.2 and H3.3 could not be shown to be either true or false due to the lack of hedges and markers of confident language in the corpus of either groups responses. The lack of hedging may be due to the context of the interaction, which was mostly made up of questions and replies (either from the user or the interface). Any hedging by the users would generally be inappropriate in this context, especially with the ambiguity that such hedges could introduce to the users utterance. Brennan and Ohaeri (1999) found that hedges in electronic conversations occurred less frequently than in spoken conversations but were still present. Their results were taken from human-human electronic conversations in a collaborative context that would give plenty of opportunities for hedging in the process of coming to an agreed conclusion. The opportunities for hedging are simply not present in this context, thus it would not be correct to assume that users are not using hedges because they are not applying human-human communicative norms to the interface (which would be counter to the CASA model). The lack of markers to confident language, and quantifiers, can also be interpreted as effects of the context of the interaction.

The average length of a user's turn was 4.27 words, this is consistent with Chai et al. (2001) (see section 1.1) who found users interacting with an interface similar in style and context would use on average 5.3 words. They also suggested that users tended to adhere to a "search culture" using key-words instead of more natural utterances in their interactions. Evidence of a "search culture" can also be found in the language used in these interactions. Consider this fairly typical interaction:

input> `Is that with Anytime20`
response> `I don't know what you mean, sorry.`

```
input> Anytime20
```

or

```
input> Can I use the phone in another country
response> With pay as you go you choose from a
selection of 4 handsets:  The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input> Can I use the phone abroad
response> With pay as you go you choose from a
selection of 4 handsets:  The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input> Phone abroad
response> With pay as you go you choose from a
selection of 4 handsets:  The Motorola T191, the
Motorola V50, the Nokia 3330 and the Nokia 5510.
input>Pay as you go abroad
response> If you have a dual-band phone you could use
your phone when you are abroad.
```

When users did not receive a response that answered their question, they tended to repeat themselves while increasing the "lexical density" (Halliday 1985) of their utterances - increasing the proportion of information dense "content" words to connecting "function" words. The user seems to drift from a conversational style to a more lexically dense written style, when confronted with unhelpful information. There is also a preference for the use of nouns over verbs as the user repeats themselves. This is also consistent with the user moving from a conversational style to a written style. The "search culture" characterised by Chai et al. could be described as a kind of hyper dense written style - due to the absence of any function words in the utterances. This hyper dense style reduces all ambiguity and obscurity from the users language. Such ambiguity and obscurity that could, as the user determines, cause the system confusion. Informal analysis did not show any marked difference in use of this tactic between the user groups.

To sum up, there were differences between the language use of the introverted and extraverted users. The introverts tended to use more words in a turn, be less formal and use the first person significantly more than extraverts. These differences in language use are subtle, complex and to an extent unexpected. Contextual effects of real world task oriented dialogue reduced the possible scope of analysis of the users input corpus (such as hedge frequency).

Chapter 5

General Discussion and Further Work

The CASA hypothesis - that users would react to the interface in a social way - did not seem to be affecting users reactions to this interface. According to the hypothesis users should have reacted to different interface personalities in different ways; users interacting with an interface with a similar personality should have found that interaction more attractive, likeable and useful than those interacting with an interface of a dissimilar personality.

This apparent inconsistency does not, however, invalidate the CASA hypothesis. During the experimentation it was presumed that altering the linguistic style of individual utterances of the interface would be enough to convincingly display either an introverted or extraverted personality. This assumption is simplistic and not necessarily applicable to true conversational interfaces where the interaction is reactive, dynamic and closer to human-human style interactions. The complex personality judgements made by users, consciously or not, depend on the interaction as a whole, not on the linguistic style of single utterances. For example, no attempt was made to implement Gricean maxims of co-operative conversation in the system. Grice's principle states that to act cooperatively in a conversation a participant should make their "conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which one is engage" (Grice 1975). An interface not adhering to these principles would appear not only uncooperative but also impolite, factors which would seriously contaminate any attempt to project a consistent personality throughout an interaction.

Other theories of interpersonal communication could also have effected the personality of the interfaces. No attempt was made to try and regulate the dominance of the interface over the user during the interaction. The extraverted and introverted interfaces both dominated the conversation at the start, as they heavily controlled the direction of the interaction. However towards the ends of the conversations the interface took a far more submissive role, answering questions it could, and falling back on apologetic defaults if it could not. These variations would not only contaminate any interface personality effects, but also introduce inconsistencies in the personality of the interface. Such personality inconsistencies have been shown to have very negative effects on peoples assessments of general social attraction (see section 1.3).

For a true conversational interface to appear believable an attempt should be made to model some of the theories related to the complexities and subtleties of human-human communication. Only with the application of principles such as Grice's co-operative maxims can true conversational interfaces be made to appear realistic, believable and thus trustworthy.

Although controlling the communicative dominance of an interface (especially in an interface that uses such simple technology) would be very hard, some attempts have been made to implement Gricean principles to conversational systems. Bernsen et al. (1996), whilst working on a Danish dialogue system, proposed a set of 24 principles that should be observed when designing dialogue models for conversational interfaces involved in co-operative interactions. These principles were based on Gricean rules regarding informativeness, truth and evidence, relevance and manner as well as principles they added due to the task oriented nature of their interaction. These extra principles concerned partner asymmetry, background knowledge and repair and classification. They found that the majority of problems with their dialogue model could be ascribed to violations of their co-operative principles.

Thus, when designing dialogue models for conversational interfaces that need to be as real and attractive to use as possible, the primary consideration should be applying basic principles of interpersonal communication, such as personality consistency and Gricean maxims, as well as concentrating on the performance of the background technology. A secondary consideration, one that follows from CASA should be to control the effects of matched and mismatched personalities between interface and user. Without considering the dialogue as a whole it is nearly impossible to implement consistent personality traits in the system.

The finding that introverted users tend to use more first person forms in their dialogue could be used to implement adaptive personalities. A system that could adapt its own personality to suit that of the user would not only benefit from the similarity attraction principle but also from social gain theory (Aronson and Linder 1965). Briefly, gain theory states that people who, in some way, adapt the state of their personality to a state similar to that of the person they are interacting with will be more favourably assessed. An interface could benefit from this favourable reaction by detecting the personality of the user and adapting their linguistic style accordingly (by matching the personality of the user). Reeves and Nass (1996) suggest that it would be relatively easy for an interface to detect the personality of the user in indirect ways such as "...the linguistic style of a user, monitoring the use of cautious claims or the propensity to interrupt." As the results of the experimentation contained within this report clearly show, detecting the personality of a user is not as intuitively easy as they presume. For example, many of the features expected of introverted and extraverted users language were either not present or not numerous enough - even after fairly long interactions - to provide a reliable method for the interface to learn the personality of the user.

Reeves and Nass also suggest that: "This learning could come from almost any exchange, everything from a registration form to a natural language help system." The results of experimentation with a model conversational interface suggest that users language may vary significantly depending on the type of interaction the user is involved in, as well as the medium of communication itself. However, the finding that introverted users reliably use more first person singular formations than extraverted users could be a useful method for the interface to learn about users personality in a true conversational interface. It would be useful to investigate if other dimensions of personality - such as openness and agreeableness - have such linguistic markers that could be used to get a fuller understanding of the users personality.

One of the main findings of the project was the generally excellent performance of AIML and ProgramD. The users generally responded well to the interface, even though it wasnot intended to be more than a fairly rough prototype. The interface was also very easy to implement, and update. AIML is undoubtably a very useful tool for rapidly developing a dialogue model into a conversational interface, the dialogue model itself being based on a the structure of a website. The method that was used to "convert" the information contained in a web site into a conversational interface was found to be generally sound and robust. The method, as outlined in fig 4, takes the structure of information in a web site (presuming the site has a menu like structure), as the basis for a normalised dialogue model that can be represented graphically as a state chart. The abstracted dialogue model can then be very easily implemented into AIML and so a conversational interface. There is a certain amount of intuition and artistry involved in creating dialogue models and system utterances - especially so if the system is intended to project some personality. However, once a dialogue model has been established, the ease with which the model can be converted into AIML means it would be feasible to automate, to a certain extent, the generation of the AIML. For example, the normalised user inputs could be extended automatically - where a user's response may be a plain "no" other alternatives could be generated to take care of users who may prefer to say "no, thanks" or even "I don't think so". This semi-automatic generation of AIML would significantly reduce its main disadvantage - the scale and laboriousness of the programming involved. For interfaces such as these to work well they need to be constantly tuned and updated. An important part of the development process of the interface is thus the analysis of user logs, and redesign and updating of the dialogue model and AIML. Developing successful conversational interfaces should be an evolutionary process.

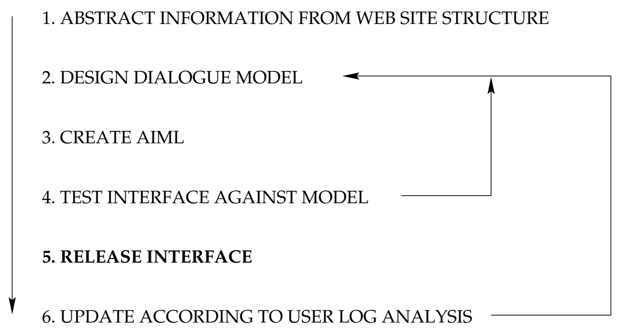

Fig4. Proposed interface development methodology

While the performance of current interfaces, be they theoretically inspired or performance led, generally remains unimpressive - at least to the general user - the commercial demand and value of these interfaces is great and growing. Analysts have predicted that: "Companies could spend $1 billion to buy virtual customer assistant software ... in 2005, up from $100 million in 2001" [Esteban Kolsky, analyst Gartner Group, from USA today 18/02/2002[1]]. Because of the intense commercial interest in natural language systems a lot of effort is being spent on the background technologies involved in implementing these interfaces. It appears, however, that careful consideration of the social role played by the interface could significantly improve the performance of that interface. This project represents an attempt to expand on the CASA hypothesis and has highlighted the non-trivial nature of synthetic personality design. The project has highlighted the performance of the simple natural language technology, involving no syntactic parsing, underlying the interface. The interface scored reasonably well in terms of usability and attractiveness, especially considering the technical sophistication of the users and the ease with which the interface was developed.

[1]http://www.usatoday.com/life/cyber/2002/02/18/web-bots.htm