Proceedings of the Institute of Acoustics

BUILDING AN AUTOMATIC RECEPTIONIST – A FRAMEWORK FOR SPOKEN LANGUAGE CALL STEERING

Mike Edgington & David Attwater

(mike.edgington@bt-sys.bt.co.uk, dave.attwater@legend.bt.co.uk)
BT Laboratories, Martlesham Heath, Ipswich IP5 3RE

1. INTRODUCTION

Since the very earliest days of the telephone there has been considerable interest in automatic call switching. In 1891 Almon B Strowger, a Kansas undertaker, patented the 2 motion electromechanical selector for automatic telephone switching. His motives were purely commercial - the local manual exchange was operated by the wife of his main rival, and he feared that much potential custom was being diverted to his competitor. The technology of call switching advanced steadily, until in 1958 Subscriber Trunk Dialling (STD) was introduced throughout the UK, enabling direct dialling of any other telephone number in the country. This was followed in 1973 by International Direct Dialling (IDD). In subsequent decades many improvements have been made to transmission and switching technology, but conceptually very little has changed – the customer still needs to enter an arbitrary numeric sequence to connect to another customer or service. In fact the range of services enjoyed by customers was actually reduced by automation, from the voice operated, intelligent services offered by human operators with local knowledge such as call transfer, follow me and call duration and charge advice. With the introduction of modern digital switches, these facilities are now available, but the customer is still presented with a user unfriendly interface to the network.

In the past decade there has been a huge growth in call centres as companies increasingly rely on the telephone as the primary customer contact channel. This has resulted in the use of a single contact number for access to multiple services. In large organisations this has meant that telephone numbers alone do not uniquely define the final destination of a customer support call, and has led to a proliferation of touch-tone menu-based call steering systems, which operate as a front end filter to the call centre. Unfortunately, these systems are difficult to use and can cause confusion.

This paper examines the potential of current spoken language recognition technology to tackle the call steering problem. Section 2 considers customer behaviour with human agents, and describes a general task based model of call handling. Section 3 examines calls to BT's operator assistance service, describes characteristics of these calls, and identifies several types of initial customer request, which can be defined in terms of the general call handling model. A number of current technologies for call steering are compared in Section 4, ranging from touch tone menu systems to the word spotting systems. It is shown that these approaches work for some request types, but do not deal effectively with calls where the customer can only provide a general description of their problem, and will not or can not map this to a particular service. To provide a broadly applicable call steering system, more general topic spotting approachs must be developed and employed.

2. CUSTOMER CALL HANDLING MODEL

2.1 Four-Layer Call Handling Model

There are typically four phases during a call to a customer service centre:

- *Problem Specification* – in which the customer describes the problem to the agent
- *Task Identification* – in which the customer intent is identified within the framework of available services
- *Information Gathering* – in which all details necessary to achieve the task are passed to the agent
- *Task Completion* – in which the customer actually receives the service or information they require

In practice, when a customer calls a human agent there is often significant overlap between these various phases. For example there may be several stages of negotiation in order to discover the actual problem experienced by the customer, during which several potential services may be offered to the customer. Some phases may also be implicitly satisfied, for example where the customer offers extra information with their request without actually making an explicit service request. Table 1 shows a real call to a BT international operator, annotated into these four phases.

Operator Services, Jane Speaking *yes please i would like to make phonecall to Iraq Baghdad* **phonecall to Iraq** *Baghdad*	*Problem Specification.*
OK er what you need to know the code do you *no i want - i want you to dial to ...aq <Iraq> Baghdad* **you want me to dial for you** *ya because from here it difficult do dial dial it*	*Task Identification*
right yes even for us it's difficult to (laughing) dialling - right what's the number including country code please *erh country code i think - 0 0 9 4* **9 6 4** *Bagdad N N N* **N N N** *N N* **N N** *N N* **N N** *N* **N thank you**	*Information Gathering*
dialling that for you	*Task Completion*

Table 1 : Straightforward call to an international operator annotated as four-layer call handling model

2.2 Scope of Deployed and Prototype Call Automation Systems

The four-layer model provides a framework for positioning various forms of automation:

- Touch tone *call steering* systems, for example the BT residential customer service system '150', automate the task identification phase, but provide no mechanism for dealing with the problem specification. Subsequent phases are performed by human operators.
- In contrast, the BT directory enquiries service on '192' uses a human operator to achieve information gathering, and then automates the task completion phase by use of recorded number announcement. Since the directory enquiry service is very well known, the first two phases are implicitly fulfilled when the customer dials the '192' access number.
- The majority of prototype and laboratory-based systems automate the final two phases without addressing either of the first two, for example corporate directory systems, such as BT's Brimstone [1]. Some systems do address the task identification phase to a limited extent, for example BT's Freedom prototype system [2] for voice control of several network services. However, for successful use, the customer must be very clear about what they want from the service, and what the service can provide, which is not evident when such a system is exposed to a representative range of real customers.

BUILDING AN AUTOMATIC RECEPTIONIST

The authors are not aware of any currently existing systems that address the problem specification phase, although this is crucial to many customer handling calls. The remainder of this paper investigates what happens in human-human interactions, and some considers approaches to automation.

3. CASE STUDY: BT OPERATOR ASSISTANCE

3.1 Call Centres within BT

BT has the largest call centre capability in Europe with around 115 call centres dealing with BT customer calls alone. These are staffed by the equivalent of nearly twenty thousand full time operators, taking over one billion calls per year. The most general contact point for BT is the operator assistance (OA) service, accessed through the very well known '100' code. Calls to OA operators cover an extremely wide range of topics, including simple requests for information, malicious or inappropriate calls, explicit requests for non-OA services, requests for various BT services, and 'other' which includes confusing requests, confused customers, and the plain odd (see Table 2). Since the OA service has such a broad functionality and customer profile it was seen as a very challenging case study for advanced call steering techniques.

operator service Lyndsey speaking
this is satan *<laughing>*
hello this is the operator can I help you
this is satan, I'll be seeing you soon

a)

operator service Tina speaking
yes, I would like you to put me through to the customer customer services
customer service, one moment please
<transferred>

b)

operator service Steve speaking
er good evening I wonder if you can help me I want to take off, I keyed in erm, a service where someone calls back and and tells me how much a call costs
er yes that's a an A D C call
an A D C ca... erm how do I take that off please
how do you take that off
yeah
I'm not quite sure did you do it through the operator or was it something you've done through your own telephone
we did it erm ourselves... I think
right bear with me one moment, see if I can find out for you
<call continues>

c)

operator service Lyndsey speaking
yes I've been more onto you about five times already it's not your fault at all err but it's a complicated thing I have to contact an Italian office like you know a Tax Office or something like that, (**right**) they given me three numbers to contact and they keep putting me on different numbers and dial this and then in the end after making about millions of phonecalls they tell me that this number is non-existent it's typical of Italy I'll tell you that I have to contact then I don't know what all else to do they give a message in Italian and then in English, and the last one I get after making all these phonecalls through you and the following they say that this number is non-existent so I don't know what to do I just received this letter and it's a kind of a Tax Office I have to contact and God knows what I am going to do
right I see, the problem is that making with making a call to Italy, I can't help you with that call
yeah but what do I have to do because I mean this number they give me is an English one is 0800
it is an English number
it is 0 800 N N N N N, when you dial that you get all sorts of information and instructions, when you get to the end of the line you are you are just told that this number is non-existent I don't know what to do
right I see can you repeat the number again slightly more slowly
<call continues>

d)

Table 2 : Transcripts of example OA calls a) nuisance b) referral c) network information d) confused

3.2 Database Collection

In order to get an understanding of the language and dialogue behaviour of customers on the OA service, a data collection exercise was undertaken. The first phase collected almost 1000 calls to the Cambridge OA centre over the course of a typical week. Calls were collected using a stereo DAT recorder connected to the operator headset, which captured an analogue signal of all parties on one channel, and the operator microphone signal in the other channel.

All calls were fully transcribed, including hesitations and restarts, and classified into detailed semantic classes. This database has been used in the pilot study to investigate general issues of dialogue structure, trends in language use and simple classification strategies. A second database is currently in preparation containing around 25,000 calls to the OA service over a one month period. This has been collected digitally, with full separation of operator and customer speech onto distinct channels. This larger database is therefore suitable for recognition experimentation, and training of fine-grained statistical language models.

3.3 Pilot Database Characteristics

3.3.1 Call Volumes. Each operator in an OA call centre can handle several different classes of calls, depending on their training and experience. For example, many operators are trained to handle International Operator Assistance calls, and most OA operators also handle '999' emergency service calls (where there is absolutely no intention of introducing automation). After excluding other classes, 752 were identified as incoming inland OA calls. The origin of the OA calls was also investigated, and approximately half (49.7%) originated from payphones, the implications of which will be described later.

3.3.2 Call Duration. The OA calls have been divided into three sections: the initial recorded operator greeting, the first response from the caller and the remainder of the call. Each operator pre-records their personal greeting based on an approved wording, and varies in duration from operator to operator, on average lasting 1.3 ± 0.4 seconds. The duration of initial customer utterance (defined in Section 3.3.3 below) is 5.8 seconds with a wide variation ranging from ~0.6 (e.g. "Er, reverse charges") to ~48 seconds. The initial utterance duration found in this work is very similar to that found previously in a US study [3], in which the average utterance duration was 5.9 seconds, with a similar distribution.

3.3.3 Language Use in Initial Customer Utterance. The transcription of the initial customer response was analysed for language use. The initial customer response is defined as the customer's utterance following the operator greeting, up until the point of the operator's next productive response. Any operator utterances which act merely as a confirmation that they are listening without interrupting the customer, are ignored (e.g. 'uh-ha', 'yea', etc.) Figure 1 shows the distribution of initial customer utterance by the number of words. The average utterance length is 17.7 words, with a median of 13 words. The useful range is from 1 word ('hello'), through to 163 words (the Italian tax office enquiry). There were a small number of calls where the customer never spoke - presumably the customer dialled '100' in error. Again, these results show strong similarity to the US study [3].

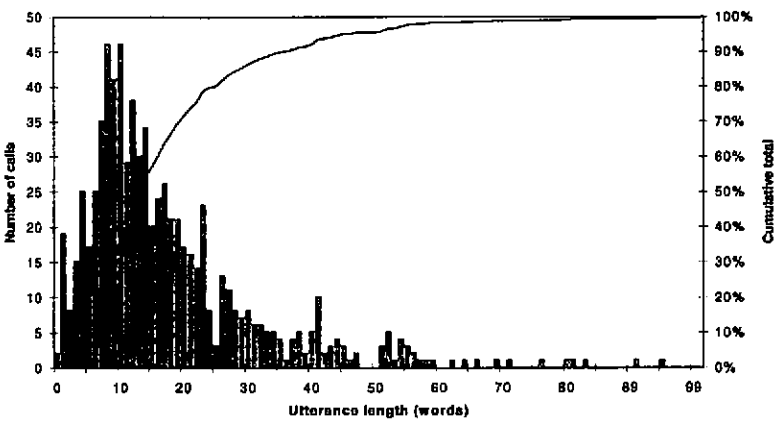

Figure 1 : Distribution and cumulative total of initial customer utterance length in words

3.3.4 Vocabulary Growth. The transcription of the initial customer utterance can be analysed to give information about growth of the customer vocabulary. Figure 2 shows how the vocabulary (i.e. number of distinct words) grows with the number of calls observed. Since the detailed shape of the vocabulary growth curve is determined by the actual order of call arrival, the graph shows a smoothed plot of the average vocabulary size over several permutations of call order. A total of 1228 distinct words were observed across all initial customer utterances, and as can be seen from the gradient, after all 752 calls vocabulary growth is still around 0.8 new words/call.

It has been shown in [4] that Zipf's Law can be used to model vocabulary growth in this class of data. This mathematical model predicts that the vocabulary size can be expressed in the form: $v = A\ c^{\ m}$ where v is the vocabulary size and c is the number of observed calls. Standard linear regression methods can be used to estimate the values of the two parameters A and m, and using the pilot study data, the model predicts a vocabulary size of around 4500 words after 10,000 calls, growing at around 0.25 new words/call.

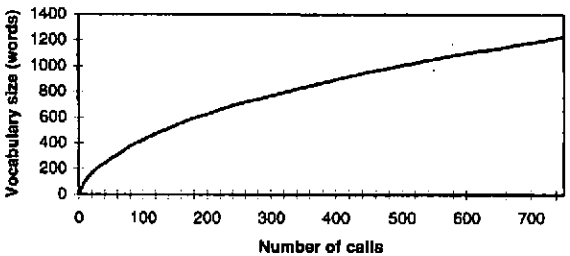

Figure 2: Vocabulary growth of the pilot study OA calls

Analysis has been performed of the differences between calls originating from payphones and non-payphones. In particular the first customer utterance is shorter in payphone calls, averaging 14.2 words, against 20.4 words for non-payphone calls. Vocabulary growth is also slower in payphone calls, primarily due to the limited range of requests (mainly 'lost coins' and 'reverse charge' calls), and thus fewer topics being discussed.

3.4 Types of Request.
By considering the detailed semantic classes, it was possible to classify the customer's initial response into one of four broad request types, which are summarised and exemplified in Table 3. This can be compared to the simpler taxonomy of short initial user utterances in a banking application described in [5].

Request Type	Description	Example
A	Explicit service request	can I er, could you put me through to directory enquiries please
		can I want a reverse charge please
B	Implicit service request	can I have the number for the er Probation Office on Dover Street, S E 1
		I'm very sorry is this 0 8 3 1 a mobile number
C	General problem specification	hi my name is XXXX XXXXXX and we have a problem here, there's someone whose trying to er call er who is calling us the whole time trying to fax us something, we haven't got a fax machine so must, must have the wrong number, it's been going on the whole day
D	Other	yeah I wanted them er er to find out the how to spell a place that I wanted to send a telegraph to in Cornwall please

Table 3 : Primary Request Types in Operator Assistance Calls

These request types can also be described with reference to the four-layer model of customer call handling in Section 2:

- A - an explicit named service request, where the customer makes a direct request for a specific service. The customer has resolved the problem specification stage, and offers their solution to the task identification phase to the operator.
- B - an implicit service request, where the customer may give details of the problem without explicitly asking for a specific service. This is similar to A, but the customer does not explicitly co-operate in the task identification phase. Instead they immediately request a specific solution. This request may contain elements of information gathering as well. In this case the operator will often explicitly complete the task identification phase by way of confirmation.
- C - general problem description, where the customer is unaware of what service they require, but know that the operator should help. The customer is at the problem specification phase and expects the operator to engage in a dialogue in order to move to the task identification stage.
- D - other. There is evident confusion within the problem specification phase, or about what the operator can do.

4. CALL AUTOMATION TECHNOLOGY

The following sections describe several approaches to automation of the call steering problem, and present results where available.

4.1. Touch Tone IVR Systems

There is a currently a proliferation of touch tone IVR systems in the telephone network. Although they are often used in the information gathering phase of an automated system, for example to enter account numbers, a major use is in call steering front-end to a call handling bureau. It is widely known that these systems can be difficult to use, and much skill is need in the design of the voice menu prompts. Even designs using best-practice have several fundamental weaknesses. In particular, the mapping system function and user action (pressing a key) is completely arbitrary and therefore difficult to remember. To alleviate this problem, menus must be kept very short, which can lead to complex hierarchical menu structures which are difficult to navigate. A closely related problem is that many users have significant difficulty in mapping their request onto one of the listed system options. Using the classification of Section 3.4, touch tone IVR systems can be effective for request type A, may sometimes work with request type B, but are completely inappropriate for request types C & D.

4.2. Spoken Menu Systems

Spoken menu systems are the natural extension of touch tone IVR systems using speech recognition technology. Their main advantages are a reduction in the prompt length, and a direct relationship between meaning and action – for example saying the word 'operator' rather than pressing an arbitrary key. However, many of the limitations of classic touch tone systems remain: the difficulty of mapping customer desire onto the menu options, and a strictly hierarchical navigation structure. There is also the added difficulty of non-perfect recognition performance, and the consequent need for error recovery strategies, such as multi-level prompting.

A trial of a spoken menu approach for OA calls was recently carried out at BT. The results appear to show that when a customer has a very specific request, using a service for which they have a strong model (request type A), then spoken menu systems can be quite successful. In contrast, where a customer either believes that their request is unclear, or does not know how to relate their problem to the menu options (request type C & D), spoken menu systems are inefficient. Since the cognitive load on the customer is generally lower in spoken menu systems than touch tone systems, type B requests are more likely to be successful. Figure 4 shows a summary of several hundred customers' behaviour when presented with a prompt of the form *'Welcome to BT. Please hold for an operator or say directory enquiries, faults, sales or reverse charge'*. This graph shows that a high percentage of customers who wanted 'reverse charge' or 'directory enquiries' replied appropriately to the menu. From focus group studies of BT

customers [6], these two operator assistance services were the most highly understood. In contrast, far fewer customers responded to the prompt then they were enquiring about a potential fault or needed customer services. It is not entirely clear whether this is just due to the customer's difficulty in mapping their individual problem onto one of the menu terms, or whether there is also an unwillingness to engage with an automated system for some types of enquiry.

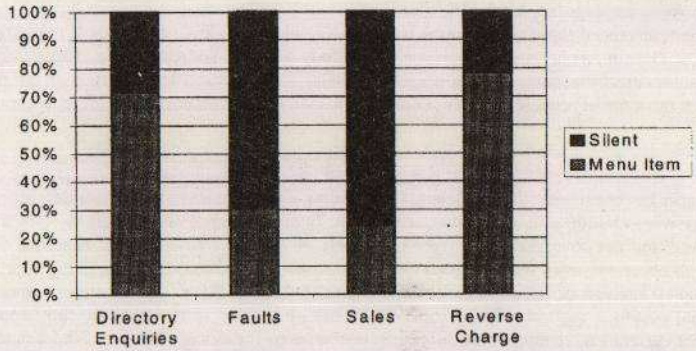

Figure 3 : Customer replies to a spoken menu (based on final destination)

4.3. Word Spotting

For some applications word spotting appears to be a viable approach to task identification [7]. Several experiments have been carried out on the pilot database using simple word spotting techniques. Figure 4 shows the maximum performance of the word spotting grammars on calls which resulted in the offer of reverse charge or line-test services. The grammars were automatically learned from the annotated database by ranking candidate word tokens according to a usefulness metric based on correlation. A range of operating points were obtained by changing the threshold for incorporating a word token into the grammar.

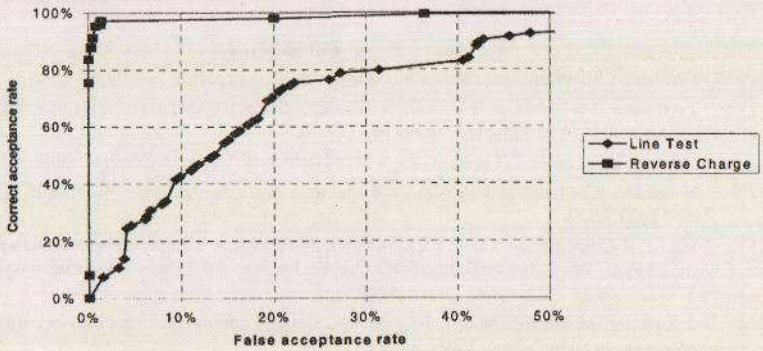

Figure 4 : Maximum Performance of Word Spotting System

It is clear that the performance of the two word spotting grammars are very different – the equal error rates are around 25% for line-test and 2% for reverse-charge. This is primarily due to the consistent language use in the reverse charge case – most requests were of type A with the remainder type B. In contrast, line tests requests were characterised by high frequencies of problem specification – mostly request type C with some type B. The language was therefore very similar to that used in other type C requests which were resolved by the offer of a different service.

4.4. Spoken Language Topic Identification

The use of advanced topic identification techniques to identify many of the type C requests is the subject of ongoing research. The authors are aware of one only one study which is addressing the identification of type C requests [3], using automatically acquired salient phrase fragments for call classification. In contrast, other studies either do not consider this type of request at all, or attempt to exclude them from automatic identification [5].

5. CONCLUSIONS

This paper has considered the problems of automating a single telephone number contact point for an organisation – in other words building an automatic receptionist. In such an application, callers often have a very wide range of enquiries, and use several distinct types of request. A general four-level call handling model has been developed based on observations of human agents, which provides a framework for describing approaches to call automation. By detailed analysis of calls to a particular customer helpdesk, BT's '100' operator assistance service, several types of initial customer request are identified, which can be defined in terms of the call handling model. A number of different current call steering technologies were compared for this application. Whilst many customers were steered successfully, a considerable proportion of customers had requests which did not appear to be dealt with effectively by existing technologies. These customers were generally describing the problem which they were experiencing, and were not relating the problem to any specific service available from the operator. Therefore our results suggest that to be generally successful, the call steering system must be able to offer solutions to general descriptions of a customer's problem. Fortunately, the call steering task operates in an interactive environment, which allows the system to enter into a clarification dialogue with the customer.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of their colleagues Peter Durston and Andy Gorham for analysis of the spoken menu results, and of all the operators who took part in the experiments and data collection.

7. REFERENCES

[1] D J Attwater, S J Whittaker 'Large vocabulary access to corporate directories' Proc. IOA **18** Part 9 (1996).

[2] D J Attwater, J S Fisher, H R Greenhow 'Towards fluency – structured dialogues with natural speech input', Proc. American Voice Input Output Society (1997).

[3] A L Gorin, G Riccardi, J H Wright, 'How may I help you?' Speech Communications **23** (1997) pp113-127.

[4] P N Garner, 'On topic identification and dialogue move recognition' Computer Speech and Language **11** (1997) pp275-306.

[5] C-H Lee, R Carpenter, W Chou, J Chu-Carroll, W Reichl, A Saad, Q Zhou 'A study on natural language call routing', Proc. Workshop on Interactive Voice Technology for Telecommunication Applications, Turin (1998) pp37-42.

[6] D J Attwater, M D Edgington, P J Durston, L Cape 'Popping the question – natural language for customer handling applications', Proc. Voice Europe (1998).

[7] F R McInnes, I A Nairn, D J Attwater, M A Jack 'Effects of prompt style on user responses to an automated banking service using word spotting'. *To be published in BT Technology Journal*