What makes a successful speech enabled call routing application?

Diana M. Binnenpoorte, Dorota J. Iskra

Customer Contact Solutions, LogicaCMG, the Netherlands
{diana.binnenpoorte,dorota.iskra}@logicacmg.com

Abstract

The key function of a speech-enabled call routing application is to connect customers who call a contact centre with the service that they want. In principle, a speech-based call routing application has the same functionality as existing and well-known touch-tone based routing applications. A touch-tone application allows the calling customer to select a service from a predefined set by pressing keys, i.e. by making choices in a multi-layered structured menu. A speech-based call routing application allows customers to formulate their question by using natural speech.
The success of a call routing application can be expressed in various ways. First, there is the technical performance; for instance, the percentage of calls that are correctly routed, but also measures that express the impact of incorrectly routed calls on a call centre organisation. Second, there is the level of customer experience; to what extent do the calling customers appreciate the speech enabled routing application? And third, there is the level of involvement within the service-providing organisation; do call centre agents understand and use the advantages of speech enabled routing in their contact with customers?
For each of these measures there are important factors to pay attention to during the design, build and implementation of a speech enabled call routing application.

1. Introduction

1.1. Why speech enabled routing?

"How may I help you?" – this question that customers hear when calling their bank, insurance company, or another service provider is invariably preceded by a number of choices they have to make in an automatic structured dialogue menu (Interactive Voice Reponse, IVR). The aim of this, always one-sided dialogue, where the computer speaks and the customers press the keys on the keypad of their telephone, is to route the customers to an appropriate agent group in the call center that will be capable of answering their question.
The ubiquitous touch-tone IVR menus are typically structured in a way remote from the customer's intuition, as a result of which a large percentage of calls end in time-outs, wrong menu choices or transfer to the operator. Complicated multi-layered touch-tone IVR menus, which are a nuisance from the customer's point of view [1], are indispensable from the point of view of an organisation. For the sake of transparency most service organisations limit the number of contact numbers. As a result various issues can be resolved under a single telephone number by various agent groups. Agent groups in a call center specialise in a restricted area of topics and are frequently physically situated in different call centre locations. Therefore, call routing is necessary.
Speech technology offers a solution satisfying both the service provider and the customer. It facilitates call routing, but in a far more customer-friendly manner. Instead of having to choose from a limited set of menu options the customers are free to speak their question and formulate it in a natural way.

1.2. How does a speech enabled call routing application work?

A speech enabled routing application actually consists of several sub-systems: a. a dialogue manager, b. an automatic speech recognizer (ASR) and c. a classification algorithm.
The dialogue manager regulates the interaction with the customer, and, in an open speech enabled routing application, typically starts by stating the question "How may I help you?" [2]. This open question evokes the customer to utter the intention of his call (see Figure 1).

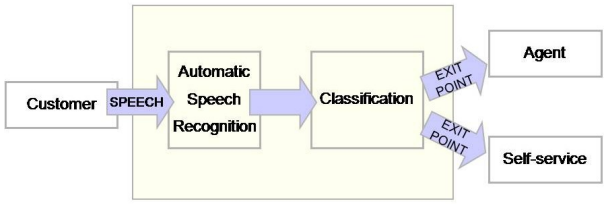

Figure 1 *Schematic diagram of speech enabled call routing*

The ASR system then recognises the spoken words using a lexicon and a grammar. The string of words, including recognition confidence scores, is subsequently given to the classification algorithm. The purpose of this algorithm is to determine the meaning of the string of words. This is accomplished by comparing the incoming word sequence with a database that is filled with word sequences that are already tagged, the so-called training material. A tag in this context is a category, or exit point (or routing goal), indicating a group of agents with specific skills (i.e. who are able to answer certain types of questions) in a call center. The classification algorithm yields one or more categories, and a classification confidence score indicating the degree of certainty for those categories. Dependent on the outcome of the classification algorithm, the dialogue manager determines the next step in the interaction with the customer: either setting up a connection to a human agent that is associated with the solely found exit point, or asking the customer for more (or different) information to find an exit point with high enough certainty. It may be clear that the more exit points, or categories, are defined, the more complex the calculations of the classification algorithm will be. Next to this, the lexicon, the grammar, the inventory of tagged sentences, and the dialogue strategies influence the performance of the application.
In short, a speech enabled routing application allows customers to just say the reason for their call without any restrictions in their choice of words. This can be accomplished by deploying speech technology which allows an open dialogue structure, as opposed to the very restricted form-filling type of dialogue. Since the customer is allowed

to phrase his intention in a natural way, it is no longer the customer who has to make a choice in a structured multi-layered menu by pressing keys, but it is the speech application that has to interpret the customer's intention, and subsequently, has to route him to the right agent or service.

2. Success factors

The performance of dialogue systems is commonly measured as the percentage of calls where the task has successfully been completed (in our case: where the call has correctly been routed), also known as Dialogue Success Rate (DSR) [3,4,5]. Besides the DSR, recognition accuracy expressed as Word Error Rate (WER) is also often referred to as a performance measure [6]. Such measures indicate the success of the application from a technical point of view. A technically successful application is, however, not necessarily one that is widely used and appreciated by a large group of customers [7]. Neither does it imply that the service providing organisation makes the most use of what the application has to offer. In this paper we also want to address other, non-technical, factors that determine the success of a speech enabled routing application, i.e. customer experience (in the broadest sense) and level of involvement within a service providing organisation. These factors, which often represent conflicting interests, play an important role in the different stages of a project that is aimed at the deployment of a speech application. In the remaining of this paper we will list the key success factors and show the way they are intertwined according to the different phases, i.e. design, build, and deployment.

3. Design

During the design of a speech enabled routing application, various issues play an important role. First, one has to establish the strict functionalities of the application; what it should do, given a list of preconditions, from both technical and organisational sources. We will not further outline the functionalities here. Second, the dialogue of the application is designed; how the application should interact with the end user, given preconditions from the supposed end user's point of view as well as the organisation's point of view. The third issue concerns determining the inventory of exit points, or categories. This issue will be presented separately from the dialogue design issue, since it plays an important role in all success factors.
It can be stated that in the design phase the foundation is laid for the following success factors: technical performance, customer experience and involvement of the organisation.

3.1. Dialogue design

When it comes to dialogue design, various, sometimes contradicting interests have to be taken into account:

Technical performance: Optimal technical performance is reached when the dialogue is continued until the application is entirely certain about the chosen exit point. In case of ambiguities, a new dialogue step is introduced according to the following conditions:

- When a single exit point is chosen, but the confidence value is low, or when the confidence is high, but the exit point is a critical self-service, a confirmation question is posed. When answered positively, the customer is transferred to an agent or a self-service. In case of a negative answer, the dialogue has to go back to a general question.
- When two or three exit points are chosen with similar confidence levels, a menu consisting of these exit points is asked and the customer is asked to choose one of the given options.
- When the confidence level of any of the exit points does not exceed the minimum threshold, the customer is asked to reformulate his request. As an extra help, an example sentence may be given at this stage.

Customer experience: A long dialogue seldom contributes to positive customer experience. In all usability tests we have conducted it has been confirmed that the customer wants to speak to a live agent and wants to reduce the interaction with a computer system to a minimum. The customer experience may remain positive with multiple dialogue steps when the customer has a feeling that he is progressing in the right direction. Especially annoying are, however, recognition and classification errors when a confirmation or a menu question containing only wrong exit points is asked. From the point of view of customer satisfaction the dialogue length may not exceed a certain maximum defined in terms of dialogue steps or wrong guesses on the side of the application.

Organisational issues: The service providing organisation usually also has its particular interests that need to be attended to at the design stage. An example might be the pricing model for the service: if the telephone costs are paid by the organisation, it is in the interest of the organisation to keep the dialogue short and force a transfer even in case of ambiguities. Alternatively, some exit points may be more costly than others. Exit point A receives 50% of all the calls whereas exit point B only 2%. If X calls are wrongly transferred to exit point A, the impact on this particular agent group may still be negligible. If the same number of calls are wrongly transferred to exit point B, the impact on this agent group may be severe and cause substantial delays and congestion in the call centre. In order to resolve this, the organisation may, for instance, require an additional confirmation question when exit point B is chosen.
The dialogue design is a trade-off between all these often conflicting aspects of the service.

3.2. Selection of exit points

When a customer calls the speech enabled routing application, his question is first recognised and then categorized, i.e. the meaning or intention of the call is determined. Once it is clear to which category the call belongs, the customer is connected to the right agent or service that can actually answer the question. Establishing the categories, or exit points is not a straightforward task.
Often, the way call centers of large organisations are structured in terms of departments, services, agent skills, and locations, is the result of an organic process of growth, change of needs, economical aspects, and so on. Still, from an organisation's point of view the structure within a call center seems as well-founded as it is. And often this structure can be traced back from menu options in a touch-tone IVR application. And this is exactly what customers complain about: a menu that is hard to navigate through because of complex terms or categories [1]. It can be stated that there is a mismatch between the customer's mental model of the structure in an organisation and the actual operational structure. For instance, a customer is convinced that a question about his income insurance can be answered by the department of life insurances, while for the organisation it is absolutely clear the question should be answered by the department of indemnity insurances.

In a speech enabled routing application the problem for the customer to 'file' his question under one of the given menu options such as in a touch-tone IVR is solved: the application determines to which category the question belongs. However, feedback to the customer can sometimes be necessary by means of a confirmation question. In this type of questions, the name of a found category is often mentioned ('Is it true that you are calling for X?'). Therefore, the names of the categories must be chosen with great care as to fit in as well as possible with the customer's idea of an organisation.
Determining the names of the categories is just one side of the medal, determining the number and type of categories is another. When making the inventory of exit points, the organisation of the call center, how agents are grouped based on certain skills, is leading. Before determining the type and number of exit points the organisation needs to carefully examine what type of questions can be answered by which skilled agent. Even within one subject range it is sometimes necessary to define more than one category since agent skills can be differentiated in, for instance, answering more complex questions versus more common questions.
So on the one side, the organisation requires a rather detailed division, while from the technical point of view it holds that the more exit points are defined, the harder it is to find the right one. There is no golden standard regarding the optimal number of exit points, [8] reports on 40 exit points, while [6] reports on only three.
The number and the inventory of exit points influence both customer experience as well as the technical performance.

4. Build

The quality and size of sentence material which is used to train the grammar is of major importance for the technical performance of the system. In terms of size, the rule "the more the better" applies only to a certain extent. At a given point, which must be established experimentally for each application, a saturation point is reached [9]. Above that point, which indicates that an adequate representation of customer questions has been reached, adding more sentences does not contribute to better performance, but has no or even a slightly negative effect. By adding more sentences there is also a risk of overlapping sentences between exit points which makes them harder to be uniquely identified when similar questions are asked.
In this context *quality* refers to the degree in which training material reflects customer questions with regard to the following aspects:

- *Terminology*: do sentences contain terminology that is used by customers rather than the one internally used by the service providing organisation?
- *Formulation*: are sentences formulated in terms of style and syntax the way customers would formulate them? In practice many spoken sentences contain hesitations, repetitions, etc.
- *Variance*: do training sentences cover the wide range of different subjects (and thus exit points) and speaking styles that are used by customers?

Training material can be collected in several ways. The below list of these approaches is ordered according to the quality with the best quality at the top:

Live sentence material: the best technical performance can be achieved if the training material is collected through a dummy application. The dialogue of such a dummy application ideally reflects the design of the future speech enabled routing application so that customers behave similarly to a large degree. The difference is in the handling of the question. The customer sentences are only recorded for the purpose of training and the customers are transferred to broadly skilled operators who can answer their questions or transfer them to another agent.

Simulated live material can be collected in two ways. Firstly, the customers who have just spoken to an agent can be asked to say their question one more time to an especially for this purpose designed application. This approach results in a wide range of sentences, but they are less spontaneous in terms of formulation since the customer has been warned beforehand. Also, not all questions are transferred to agents and a large proportion is handled by self-services. These sentences will be missing from the collection.
Alternatively, customers or employees of the service providing organisation can be requested to make calls to a special application whose sole purpose is recording training sentences. However, the response from customers in this kind of actions is usually very low. The disadvantage of using employees, even agents, is that their way of formulating questions can only be an approximation of what the customers say. The success of this approach is very dependent on the acting skills of the participants involved. In terms of planning it can coincide with the trial period preceding the development of the actual application which is used to test the dialogue design, collect sentences and give an indication of the expected performance.

Written sentence material: if it is impossible to collect spoken sentences, the second best is text. Written sentences can, for instance, be derived from a website application if the latter allows for natural language input. This kind of sentence material, although containing the vocabulary used by customers, costs a lot of time and effort to be made ready for training purposes. Internet sentences contain numerous spelling mistakes, spelling variations, textual signs such as, e.g. "+", and abbreviations. All of these have to be repaired before the sentences can be used for training. Moreover, it has to be kept in mind that the way of formulating questions in a website application is very different from speech. In writing, customers tend to use short expressions, often one or two words, and restrict their enquiries to keywords.

Simulated written sentence material: if it is impossible to collect spoken sentences and the service providing organisation does not have a website allowing for natural language input, a sentence collection can be organized where the employees of a service providing organisation are asked to make up sentences on a number of given subjects. This approach is similar to *simulated live material*, except that the sentences are provided as text.

5. Deployment

5.1. Training agents

Once the application is built, tested and put in production, customers will be routed to agents or services. Assume that the customer has perfectly been helped by the application and is directly routed to the right agent. What if that human agent starts the conversation with the question "How may I help you?". The customer will at least have the feeling of having to start all over again. It is important that the agent already knows what the call is about before the actual conversation starts, at least in more detail than on department or skill level. When possible, it is desirable to display the recognised sentence on the agent's telephony desktop. The agent then first reads the sentence before

starting the conversation with the customer. In this way, the conversation is already one step ahead, giving the customer a higher appreciation of his interaction with the organisation.
However, as open speech recognition, as ASR in general, does not perform 100%, the recognised sentence on the agent's desktop deviates from what has actually been spoken by the customer. The degree of deviation determines the usefulness of the displayed sentence. When the sentence contains many recognition errors, the agent can best start the conversation by apologising and trying to retrieve the customer's question himself. The difficulty is, that it is not clear to the agents whether the spoken customer input is recognised very poorly or not. The agent has always to try to interpret the recognised word sequence and decide whether it gives enough clues to start the conversation directly.
The new conversation technique requires some training and change in the way of work of the agents. Once both interest and support from the agents towards the application is fostered, agents will not only appreciate the advantages of a speech enabled routing application, but are also inclined to act as ambassadors towards the customers, making the application successful.

5.2. Management information

The speech enabled routing application can be a valuable source of management information for the service providing organisation. With an appropriate logging and reporting mechanism the application can provide various statistics about, for instance, the number of calls, their distribution with regard to the time of day, or classified exit points.
In a more sophisticated set-up it can also measure, for instance, the effectiveness of an advertising campaign. If a new product is advertised, the service provider can calculate the number of questions posed about this product to the application. If a separate exit point is designed for this product, the calculation is based on the number of calls classified to this exit point. However, if the product is accommodated within an existing exit point, the calculation can be based on the extraction of the words related to that product from the recognition results.
The extent to which the speech enabled routing application is used to provide management information is dependent on the level of involvement of the service providing organisation. The higher this level, the more justified the deployment of a call routing application not only as a means of routing calls, but also as a source of valuable customer data.

6. Conclusions

In this paper we have shown that besides the commonly evaluated technical performance of the speech enabled routing application, a number of other aspects play an important role in determining the final success of this application. These aspects relate to customer experience and the level of involvement of the service providing organisation. We have described a number of issues during the different phases of the project such as dialogue design, selection of the exit points, collection of sentence material, and so forth, where our success factors intertwine and occasionally even represent conflicting interests.

7. References

[1] Suhm, B., and Peterson, P., "A Data-Driven Methodology for Evaluating and Optimizing Call Center IVRs", International Journal of Speech Technology 5, 2002, p 23-37.

[2] Gorin, A.L., Riccardi, G. and Wright, J.H., "How may I help you?", Speech Communication 23 (1997), p 113-127.

[3] Rahim, M., Pieraccini, R., Eckert, W., Levin, E., Di Fabbrizio, G., Riccardi, G., Kamm, C., Narayanan, S., "A Spoken Dialogue System for Conference/Workshop Services", Proceedings of ICSLP 2000.

[4] Sturm, J., den Os, E. and Boves, L. "Dialogue Management in the Dutch A Train Timetable Information System", Proceedings of Eurospeech 1999.

[5] Natarajan, P., Prasad, R., Suhm, B. and McCarthy, D., "Speech-Enabled Natural Language Call Routing: BBN Call Director", Proceedings of ICSLP 2002.

[6] Hessen van, A. and Hinke, J. "IR-based classification of customer-agent phone calls", Proceedings of Interspeech 2005, Lisbon, p 597-600.

[7] Boves, L. and den Os, E., "Applications of Speech Technology: Designing for Usability", Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding 1999.

[8] Kuo, H-K. J. and Goel, V. "A Data Visualisation and Analysis Method for Natural Language Call Routing Systems Design", Proceedings of Interspeech 2007, Antwerp, p 2729-2732.

[9] Schohn, G. and Cohn, D. "Less is more: Active learning with support vector machines", Proceedings of the ICML 2000.