# Telephone Voice Interfaces on the Cheap

Thomas Hornstein

UBILAB, Union Bank of Switzerland
Bahnhofstr. 45, CH-8021 Zurich

e-mail: hornstein@ubilab.ubs.ch

*Traditional interactive voice response applications are based on well-known menu-like structured dialogues using DTMF. This navigation technique is application-dependent and has limitations. It cannot be improved by simply switching from DTMF to voice input. Rather, we propose an application-independent navigation method called Zap & Zoom in combination with voice and key input. Users can Zap over a list of items (subjects) and Zoom into items of interest (content of subject). A set of application-independent commands was defined for this type of navigation and trained for voice input in three languages. Design recommendations have been set up to employ the Zap & Zoom navigation in telephone information systems and to achieve an open, easy-to-use and consistent voice interface. Two different information services based on the Zap & Zoom navigation were built.*

## 1 Introduction

Telephone-based information services have been introduced in the last decade. The interaction between the machine and the user was based on the telephone and its key pad using touch-tone (DTMF[1]) signals. This technique is fairly efficient since it is simple and people are used to seeing a similar interface on other devices like automatic seller machines etc. Unfortunately the distribution of DTMF based telephones is still not homogeneous in most countries. In some areas DTMF dialling is not yet supported. In many countries - particularly in Europe - people often keep using their old pulse-dialling telephone. Additionally ISDN has been introduced as a third standard. One way of building interactive voice applications is to support the various communication standards. Another way is to be independent of the communication standard by supporting voice input.

Supporting various communication standards raises several problems in the case of pulse detection. The response time of the application depends on the entered digit. Pulse detection is also error-prone. Additionally phones with a dial do not have a star or a hash key. These limitations could be overcome by the so-called pocket dialler (DTMF-generators) but this is an additional device and is often not accepted by, or not available to customers.

On the other hand, telephone-based information services supporting voice input are independent of the type of telephone equipment and the underlying communication standard. Another reason for supporting voice input is that speech interfaces are often more time-effective and subjectively easier to handle for novice users than DTMF interfaces [FRM93]. Since the beginning of the nineties the deployment of applications with voice input has grown slowly. There are two main reasons: first, the computing

1 DTMF=Dual-Tone Multi-Frequency

power needed for voice recognition is expensive; and second, the development of voice recognition vocabularies for particular applications is very time-consuming. The latter is still true while the former has become less important. This paper describes how we avoided setting up the development of expensive and application-dependent vocabularies.

In section 2 we first describe the requirements for voice navigation and summarise the traditional navigation methods and their limitations. In section 3 we introduce the domain- and application-independent navigation technique Zap & Zoom - an alternative to the traditional menu-based navigation method. In section 4 we then describe our design model for IVR applications which leads to generic telephone information services and faster development. Finally we show an example of a phone banking information service using the ideas proposed.

## 2 Voice Navigation

### 2.1 Principles

In early 1992, we started to investigate voice technology with the aim of building telephone-based information systems. After studying basic aspects, we realised that human computer interaction and the ergonomics of voice interfaces are the essential factors in the success of information services for occasional users. This leads to a user oriented system development where usability tests play an important role.

Many papers describe various kinds of DTMF-based interfaces [HAL89], [DET90], [PEL93], design criteria for telephone based applications [KLO94] and style guides [FRT91]. The described techniques do not help in the design of voice recognition interfaces. They often describe only basic interaction types such as the traditional menu navigation. We define the following guidelines which help to extend the user interfaces of telephone information systems successfully to include voice input:

- Application-independent navigation
- Suitable for selections from a large number of choices
- Easy for novices and fast for experts
- Active users instead of passive users

The following sections describe why traditional navigation techniques do not satisfy all these guidelines.

### 2.2 Simple Menu Approach

Traditional menu navigation is tree-based. It uses the digits zero to nine, yes and no as keywords. These twelve keywords are simply a mapping of the telephone key pad. The computer first plays a message containing all possible options of the menu. The users are then asked to select one option by speaking a digit specifying the number of the option. The advantage of this approach is that only one small vocabulary is needed for both navigating through the service and entering data (integer values). This helps to provide a high recognition accuracy and keeps the training effort for vocabularies in different languages to a minimum.

Regardless of whether voice or key input is used this method has several limitations and drawbacks which prevent the building of more sophisticated services:

- Users are forced to listen to long menu prompts before selecting one option. That means most of the time users are inactive.
- The receptivity of users is limited when listening to prompts and this often makes it impossible to play more than four to five options at the time [ENG90], [PAA86].
- As a consequence of using digits as input tokens, the semantics of these tokens differ from menu to menu and application to application. This makes it difficult for the users to learn the menu inputs.
- It is difficult to compose menus dynamically at run time

### 2.3 Enhanced Menu Approach

A common method of enhancing the usability of traditional menu navigation is to extend the vocabulary. Instead of speaking a digit to select a menu item, users can speak a keyword for that item. This means that each menu item has its dedicated keyword. The navigation principle remains the same as the one described in 2.2 with all its limitations.

On the other hand, this technique makes interaction more intuitive. Dedicated keywords for menu items also allow direct access from a given menu to any other menu in an information service. But as a consequence the vocabulary becomes highly application-dependent. Once the application changes, new keywords have to be trained for a new vocabulary.

The enhanced menu navigation approach is application- and domain-dependent. This is probably the most significant obstacle for service providers employing this type of interface. Collecting the speech training data for a given vocabulary is a significant cost factor when developing an IVR service. To reduce costs and to make the vocabulary reusable, the underlying navigation technique of an information service should be application- and domain-independent.

The use of a large vocabulary can be seen as a CISC approach as it is for microprocessors. We believe that small reusable vocabularies, combined with an application-independent navigation method, are more effective. This is more like a RISC approach.[2]

## 3 The Zap & Zoom Navigation Approach

Zap & Zoom (Z&Z) navigation is list-based. It is designed for easy-to-use telephone applications. In addition, the method is suited for general purpose information systems. The principle of a list-based telephone interface using only key input has been introduced by [RES92] as an extension to conventional menu navigation. Experiments have also been done by Apple to store voice notes on personal devices in a list-based fashion [STI93]. We adapted the idea of the list-based principle for voice recognition and added more features in order to make it more flexible.

### 3.1 Principle

A Z&Z list consists of single interaction elements. These interaction elements are basic building blocks and are called Z&Z elements. In each Z&Z element the user is prompted to select that item. Users can zap over the Z&Z elements with the *next* or *back* command

---

2 CISC=Complex-Instruction-Set Computer, RISC=Reduced-Instruction-Set Computer

and they can zoom into an item of their interest with the *yes* command. It is a little like watching TV without a program listing. You *zap* through the channels until you find the program you are interested in. Figure 1 shows a Z&Z element on the left side and a traditional menu selection on the right side.

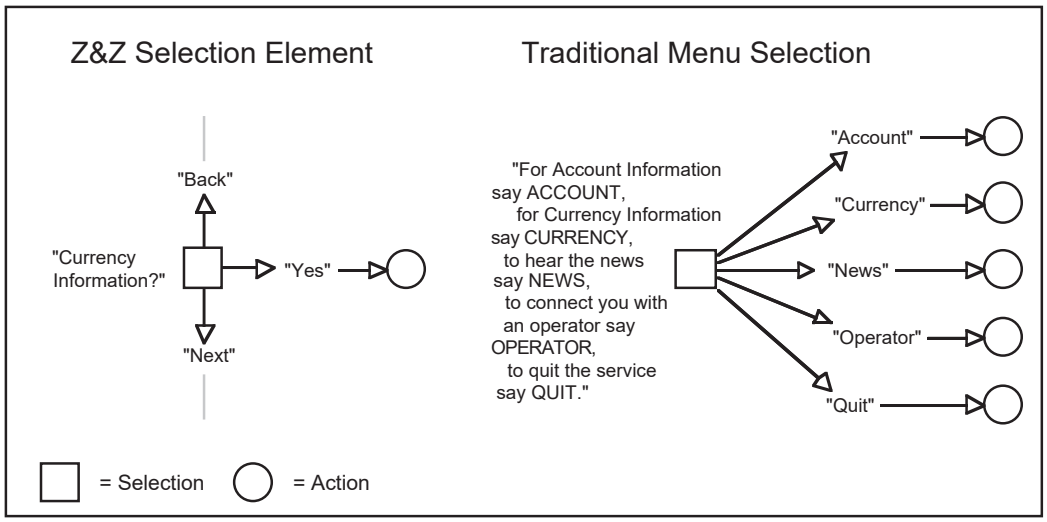


Figure 1: Difference between Z&Z navigation and traditional menu selection.

The selection elements for Z&Z and the menu in figure 1 are atomic units of their underlying navigation principle. The Z&Z element is generic and application-independent while the menu selection is not. A Z&Z list representation of the menu selection in figure 1 using connected Z&Z elements is shown in figure 2.

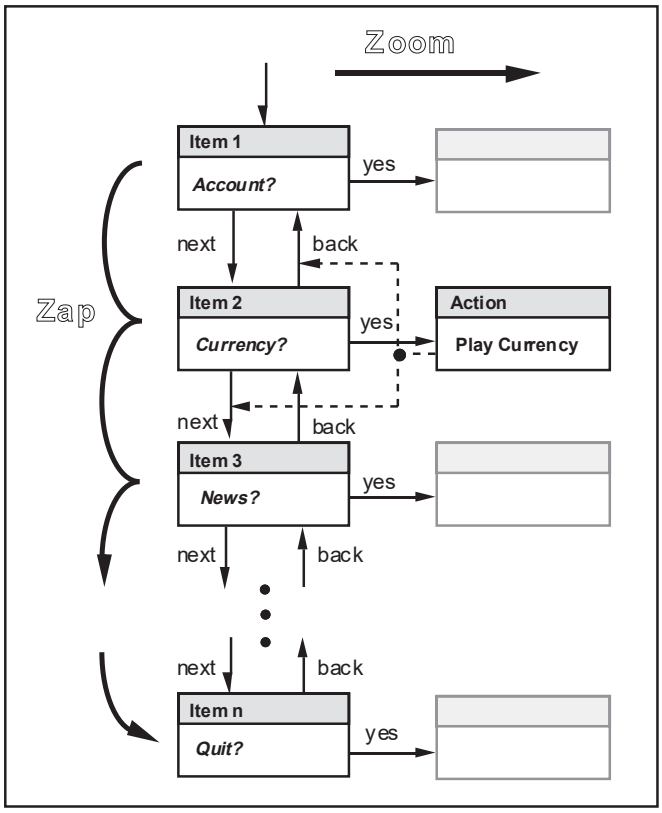


Figure 2: Principle of a Zap & Zoom list.

A selected action automatically moves to the next item when it has finished. To make the dialogue intuitive it is important that next means the next item relative to the user's navigation direction. This avoids unnecessary repetitions of prompts. When the user zaps forwards, the action moves to the "next" item. When the user zaps backwards, the action moves to the "previous" item.

Forcing users to prompt after each item gives them more initiative and allows them to explore the service on their own. Items can be easily connected together at run time.

When moving from traditional menu-based navigation to the Z&Z navigation, three main differences can be summarised which characterise a Z&Z:

- Only one item at the time should be played and users have to answer after each item
- Users can move forward and backward
- The system knows the direction the user is moving

### 3.2 Commands

The next, back and yes commands do not fulfil all requirements of Z&Z navigation. To add the missing functionality a set of twelve commands is defined. All twelve commands are application-independent and are listed below.

Commands are printed in italic letters. They represent an exactly specified meaning and are place holders for keywords of the voice recognition. A service that supports different languages has exactly the same Z&Z lists and uses simply different keywords as commands. Keywords must be chosen very carefully when training a vocabulary for any language so that they are unequivocal to users [KLO94]. Good keywords can be found only through repeated usability tests [TOG91]. Depending on the performance of the

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.