Int J Speech Technol
DOI 10.1007/s10772-008-9007-3

VoiceMarks: restructuring hierarchical voice menus for improving navigation

Pourang Irani · Peer Shajahan · Christel Kemke

Received: 12 June 2006 / Accepted: 9 October 2008

Abstract Interactive Voice Response (IVR) systems, or touch-tone telephony interfaces, are nowadays a common medium of interaction between organizations or companies and their customers, allowing users to access or enter specific company-based information. These telephony interfaces typically involve the use of hierarchically structured voice menus, through which a user has to navigate in order to locate a specific desired menu item. This navigation process is often inefficient and time-consuming, leaving users at times frustrated and annoyed. In this paper, we describe the foundation of VoiceMarks, a system designed to improve the ease and efficiency of navigation in menu-based voice interfaces. The system features personalized menus through the use of voicemarks, in a process similar to bookmarking, but adapted to voice interfaces. VoiceMarks are essentially bookmarked nodes in the voice menu hierarchy, which are stored for the respective user in a directly accessible, personal menu. We developed and tested VoiceMarks interfaces for two applications: a bus schedule information system and a cinema ticket purchase system. A comparative study of VoiceMarks and traditional interfaces of these applications showed that VoiceMarks can significantly improve the interaction between users and systems, in terms of time and number of keystrokes needed to locate a menu item, as well as regarding user satisfaction. In general, users responded very positively to the VoiceMarks interface. In addition, the study pointed to some useful modifications of VoiceMarks, which should be considered before employing the system in a commercial setting.

P. Irani (✉) · P. Shajahan · C. Kemke
University of Manitoba, Winnipeg, Canada R3T 2N2
e-mail: irani@cs.umanitoba.ca

Keywords Voice user interfaces · Personalized touch-tone menus · Telephony bookmarks · Touch-tone interface navigation

1 Introduction

Many organizations deploy Interactive Voice Response (IVR) or touch-tone telephony interfaces to allow end-users or customers access to various types of information. With telephony interfaces, end-users are able to perform a variety of tasks such as their daily banking transactions, paying their bills and retrieving ticketing information, city directions, and cinema listings at their convenience. In general, touch-tone interfaces are used in instances where access to high information bandwidth (i.e., visual displays) is impractical or impossible, and the task of querying and delivering information is complex. The popularity of touch-tone applications grew rapidly in the late 1990's with the introduction of automated call-centers. However, due to reduced levels of usability their popularity declined over the years. It is possible that with more usable touch-tone interfaces users will continue accessing information with them.

Unlike the presentation of visual information, which can be accessed in parallel, touch-tone interfaces deliver information serially. The narrow interaction bandwidth available to touch-tone interfaces results typically in degraded usability performance (Gong and Lai 2001; McInnes et al. 1999; Resnick and Virzi 1992; Tatchell 1996; Yankelovich et al. 1995). In order to optimize the use of this limited channel capacity, the backbone structure for guiding a user in accessing the information is primarily a hierarchy (Fig. 1). Options for executing a given command or accessing a specific feature are presented in a hierarchy or tree structure, with the user entering the system at the root node, and then navigating downwards, until the requested option is found. By listening to the appropriate prompts for menu items and making appropriate selections, the user can navigate through the hierarchy, branching into sub-hierarchies, until the final goal item has been reached. The use of sub-hierarchies for directing the user has effects on the navigation efficiency within the system. In systems with large hierarchical structures and many levels of options, users can easily lose context, and make errors when trying to locate the sought after option or information. Consequently, a major obstacle in promoting touch-tone interfaces is the high level of user frustration, which can arise when navigating through the hierarchical menus (Roberts and Engelbeck 1989).

There has been a considerable amount of work on exploring methods for reducing the time a user needs for navigating a hierarchical menu structure. Major solutions that have been provided so far involve optimizing the potential paths a user can take in a system (Balentine 1999), finding a balance between the length of menus and their associated prompts (Pu and Faltings 2002), inserting additional location cues in the menus (Brewster 1998; Shajahan and Irani 2004), enhancing the system with features such as

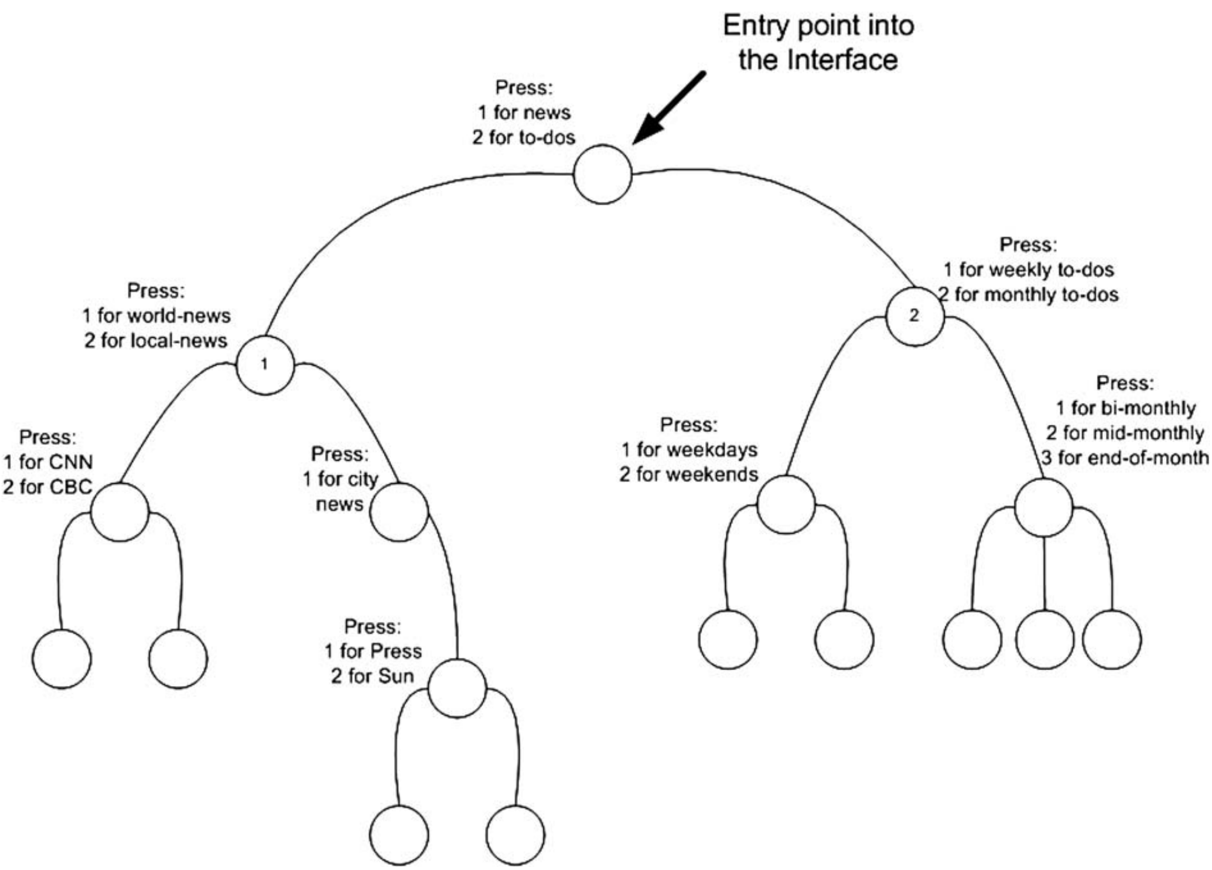

Fig. 1 Example of the hierarchical menu structure of a touch-tone interface for news and to-dos

barging-in (Karat et al. 1999), and designing the graph structure to obtain the most efficient balance between the width and the depth of the hierarchy (Gardner-Bonneau 1999). Recently, there has been renewed interest in enhancing hierarchical touch-tone interfaces with visual cues (Yin and Zhai 2006). In this paper, we describe a new method, which reduces the navigation time in hierarchical menu structure and the access time necessary to retrieve information from hierarchical touch-tone interfaces. The core concept, called ***VoiceMarks***, relies on providing users with the option to personalize the menu structure, by allowing them direct access to nodes of interest in the hierarchy without going through the various layers of prompts through a bookmarking feature.

In Sect. 2, we give an overview of the work that inspired the development of VoiceMarks. In Sects. 3 and 4, we provide a detailed description of VoiceMarks. We then present an evaluative study of this new interaction technique with potential users, and discuss the outcome and results in terms of user performance and satisfaction.

2 Related work

Two main bodies of research are relevant to the development of the VoiceMarks concept. The first set consists of various techniques and studies designed to improve navigation in hierarchical voice menu interfaces. The second set concerns the development and use of personalization in the context of graphical menu structures.

2.1 Hierarchical voice menus

The main mode of interaction with voice interfaces is through menus, which are structured in the form of a hierarchy. A number of guidelines have been recommended to limit the size of the hierarchy structure or the size of the items in the menus. Roberts and Engelbeck (1989) suggest to restrict the depth of menus to four levels, whereas Balentine and Morgan (1999) recommend to limit the number of levels in a menu structure to at most nine. Suhm et al. (2001) compared long menu labels containing items with well-defined functions to shorter menus consisting of prompts that compress several functions into one menu item. Their results indicate that long menus with specific and clearly defined categories can route users more efficiently through the system than short menus with items consisting of broad categories. Since long menus consisting of items with succinct functionalities result in fewer layers than short menus, Suhm et al. (2001) suggest to use long menus with fewer levels whenever possible.

To improve caller navigation and routing in voice menu interfaces, Balentine (1999) provides a range of suggestions for re-engineering the speech menu. These suggestions include ensuring that the user successfully completes their navigation path in a minimal amount of time and that superfluous information is not presented to users. The study also suggests that providing confirmation dialogs informing the user about their route during the traversal process will reduce error rates and improve navigation time. McInnes et al. (1999) redesigned the navigation process in hierarchical menu based voice interfaces by adding confirmations to users' actions at each level. In their study, they compared various styles of confirmation, with the aim to ensure the detection and correction of speech recognition errors. In one case, the recognized details are simply spoken back to the user, who must then contradict any detail that is wrong. An alternative option is to present an explicit question after the read-back, which requires a response from the user ("yes" or "no"). The results of their study show that it is important for confirmation messages to have a question intonation. In addition, their results show that navigation is improved, since confirmation gives feedback to the caller about their previous action, which enhances progress in their task.

Resnick and Virzi (1992) have suggested a method to reduce navigation time in hierarchical voice menus, called *skip and scan*. Their assumption is that one of the main problems with hierarchical voice menus is that users are forced to listen to all the prompts in a menu before they can make a selection. In some interfaces, the size of the single prompts can be very long, which leads to user frustration and subsequently a higher error rate. In order to address this problem, a new technique, "skip and scan", was introduced. This technique allows users to scan menus and skip those items, in which they are not interested. Figure 2 shows an example of a skip and scan interface. Users can press "9" in the telephone keypad to skip the current prompt or select "7" to listen to the previous one. Resnick and Virzi (1992) compared the skip and scan

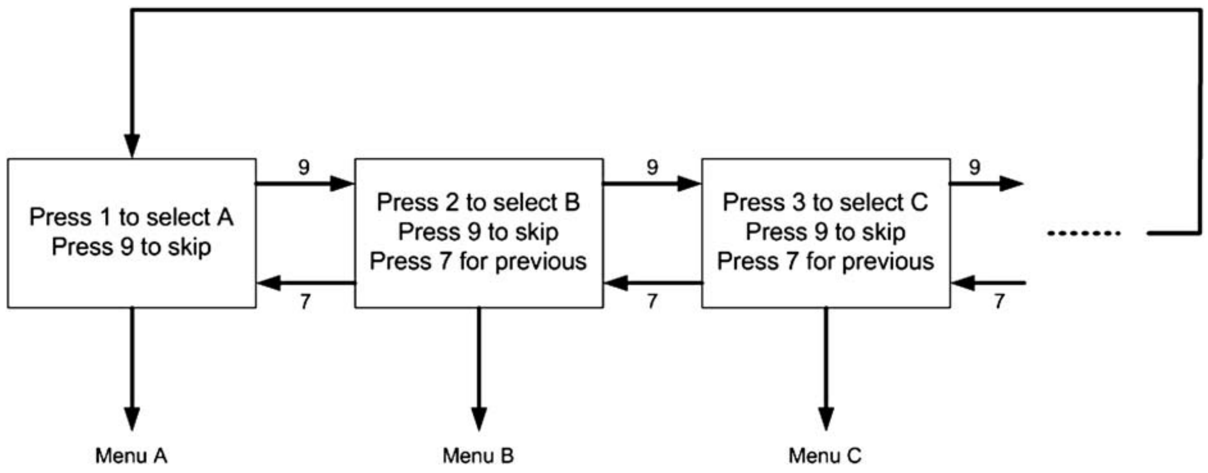

Fig. 2 "Skip and scan" interface

interface with the traditional telephony interface. Their study shows that users find a particular option in the hierarchy faster with a skip and scan interface than with a traditional interface.

While the described studies focus primarily on redesigning menu prompts and restructuring the flow of a caller's session, Leplatre and Brewster (2000) provide a framework for using non-speech audio to support navigation in menu-based interfaces. Their approach puts emphasis on mapping earcons (Blattner et al. 1989) to nodes in a hierarchical menu structure. Earcons are non-speech sounds that are created using sound parameters such as timbre, frequency, and intensity. They can be created in such a way that their parameters reflect the hierarchical structure of the interface. In their study, earcons were assigned to nodes according to the depth and the left-to-right position of a node in the tree. The earcon parameters such as frequency and timbre were modified to convey the structure of the hierarchy. Results show that users employ fewer key presses to complete navigation tasks, and that they are more often successful in finding their goal item in the hierarchy. A disadvantage of this approach is that the set of earcons has to be created specifically to match the structure of a touch-tone interface, and, each time the touch-tone interface is restructured, the set of earcons and their mapping to the hierarchy has to be modified.

Shajahan and Irani (2004) explored in an approach similar to Brewster (1998), whether items in hierarchical voice menu structures can be recognized if they are associated with multiple synthetic voices. Shajahan and Irani designed a number of prompts by manipulating synthetic voice parameters according to a set of design principles outlined in Sumikawa (1985). Results of their study show that subjects perform the tasks significantly better in interfaces with multiple synthetic voices than in interfaces with a single synthetic voice; participants recall 84% of the positions of items in the structure more accurately with multiple synthetic voices. Their results also show that users are, with very little training, able to recognize elements in the hierarchical structure using the voice cues.

Recently there has been interest in augmenting the touch-tone interface with visual representations. In a series of qualitative and quantitative studies, Yin and Zhai (2006) examined the benefits of augmenting telephone voice menus with coordinated visual displays and keyword search. Their results show that participants navigate voice menus on average 36% faster with 75% fewer errors. According to their study, participants consistently liked the visual augmentation of voice menus.

The major drawback of the suggested modifications and extensions to touch-tone interfaces is that users are still forced to navigate through all the menu layers downwards, in the vertical dimension of the hierarchy, in order to reach their final destination, i.e. the node of interest in the hierarchy. In reality, frequent users of a telephony system often memorize the sequence of actions they have to take in the touch-tone interface to reach their desired destination (Smyth et al. 2005). Nevertheless, with the current designs and suggested improvements of touch-tone interfaces described above, these users still have to follow the complete menu hierarchy in depth, in order to reach the requested node. In this way the user's navigation is highly contingent upon the design of the system, and does not take into account the users' preferences or typical behaviors. The work described below is aimed at improving this situation. The main idea is to allow users to store and later access specific paths of interest or paths that are more frequently used in the telephony interface through a process similar to bookmarking.

2.2 Personalizing visual menu structures

The work described in this paper is primarily inspired by work on personalization or customization in visual interactive interfaces. *Customization* or *personalization* is defined as the changes that are made by the user to the default system or interface with the goal to more efficiently access some of the available options of the system (McGrenere et al. 2002). Traditional software applications offer all the functionalities to users, regardless of their intended tasks and their experience with the system. However, users typically use only a few options in an interface (Carroll and Carrithers 1984; Linton et al. 2000; McGrenere and Moore 2000). Therefore, personalizing the interface allows users to access their preferred options more easily and quickly. An example of personalization or customization is the use of bookmarks in web browsers, which allow users to save the address ("node" on the Internet) of a favorite page and then directly refer to that page later. A similar concept is used in hiding menu items in Windows™ applications, like MSWord™. Menu items in these applications can be displayed or hidden, and users can select which items (functions of the application) they want to see or hide, depending on their preferences and typical behaviors or needs. This allows a faster access to frequently used tools, based on user preferences and choices. Personalization allows users to highlight often used and desired items and at the same time to place in the background and ignore irrelevant items. Personalization thus reduces the amount of time and effort spent by the user to access required features or tools significantly.

The idea of personalizing menus in hierarchical voice interfaces is related to McGrenere's solution to personalizing complex software systems (McGrenere et al. 2002). McGrenere et al. (2002), McGrenere and Moore (2000) created a system for allowing users to customize favorite options in MSWord™. In this system, users can save their favorite options in the customized interface. The system also allows users to toggle between the customized interface, called *MSWord Personal*, and the default MSWord™ interface. In a study, McGrenere evaluated MSWord Personal in comparison with the default MSWord™ interface (McGrenere and Moore 2000). Initially, the personal interface contained very few functions. Through the so-called "modify function", users are able to add more functions to the interface. The experimental evaluation showed that the users' ability to learn functions of the system and to navigate between different menus in the software improved and that users' satisfaction increased. Overall, the participants preferred to use the new customized system over the default interface. Participants said in particular that they were able to access their favorite options more easily, without the need to navigate through the entire set of menus. These results, along with the common practice of bookmarking nodes of interest in web browsers have inspired the development of VoiceMarks.

3 *VoiceMarks*: personalizing the touch-tone interface

Personalization at the interface is motivated by minimizing the gap between what users know and need to know and by reducing the amount of information that interfere with the users' workflow (Baecker et al. 2000). The idea of personalizing voice menus was inspired by the notion of bookmarks used in graphical applications and browsers. We demonstrate the concept of personalizing voice menus using telephony applications. Telephony applications provide a good environment for testing the concept of personalizing voice menus; they are highly structured, their elements are arranged hierarchically, and users typically spend a significant amount of time traversing telephony menus. We first describe the basic functionality of a personalized menu, the interface that is produced as a result of personalizing hierarchical menus, and the architecture of the prototypical applications that we developed. We then describe the type of interaction dialogs used in personalized menus. In the remainder of the document, the terms "menu" and "hierarchy structure" are used interchangeably. Similarly, the term "node in the hierarchy" is also used for referring to a menu prompt in the voice menu.

3.1 Bookmarking menu nodes

VoiceMarks refers to a concept that allows users of a voice-based hierarchical interface to bookmark or save their favorite nodes in the menu hierarchy. The bookmarks later appear as options higher up in the menu tree. Users can then bypass the various layers of prompts and access their bookmarked information directly. Figure 3a shows a sample hierarchy before the creation and insertion of personalized bookmarks. In this example, the user chooses to save two nodes in the hierarchy, called Node 1 and Node 2, respectively. The routing information in the hierarchy for these two nodes is then stored associated with this specific user in the system database. Figure 3b depicts the new hierarchy, once the user has selected and saved those nodes. A new top level is created in the hierarchy, which allows the user to either choose from the default menu options (sub-tree labeled "Main") or to select the saved options, stored in the personalized list (sub-tree labeled "Personal"). By selecting an item in the list of personal options, the user is directly routed to the corresponding node in the original hierarchy. This is analogous to saving bookmarks in web browsers.

3.2 Interface

At the interface level, each touch-tone menu item consists primarily of two elements, a function label and an associated action (Balentine 1999). For example, in the menu item "To review your appointments, press one", the function label is "review your appointments" and the associated user action is "press one". Storing such function nodes, which are located deeper in the hierarchy, in a personalized menu, is problematic, since the function may require or involve specific information that has been collected by the system while the user traversed the hierarchy downwards to the respective node. Higher level nodes in the hierarchy serve in general the purpose of filtering the flow of the user's choices and thus possible actions. For example, the node representing the function of providing information about a user's appointments may have been reached through a route that implies specific choices determining this function in detail. In the

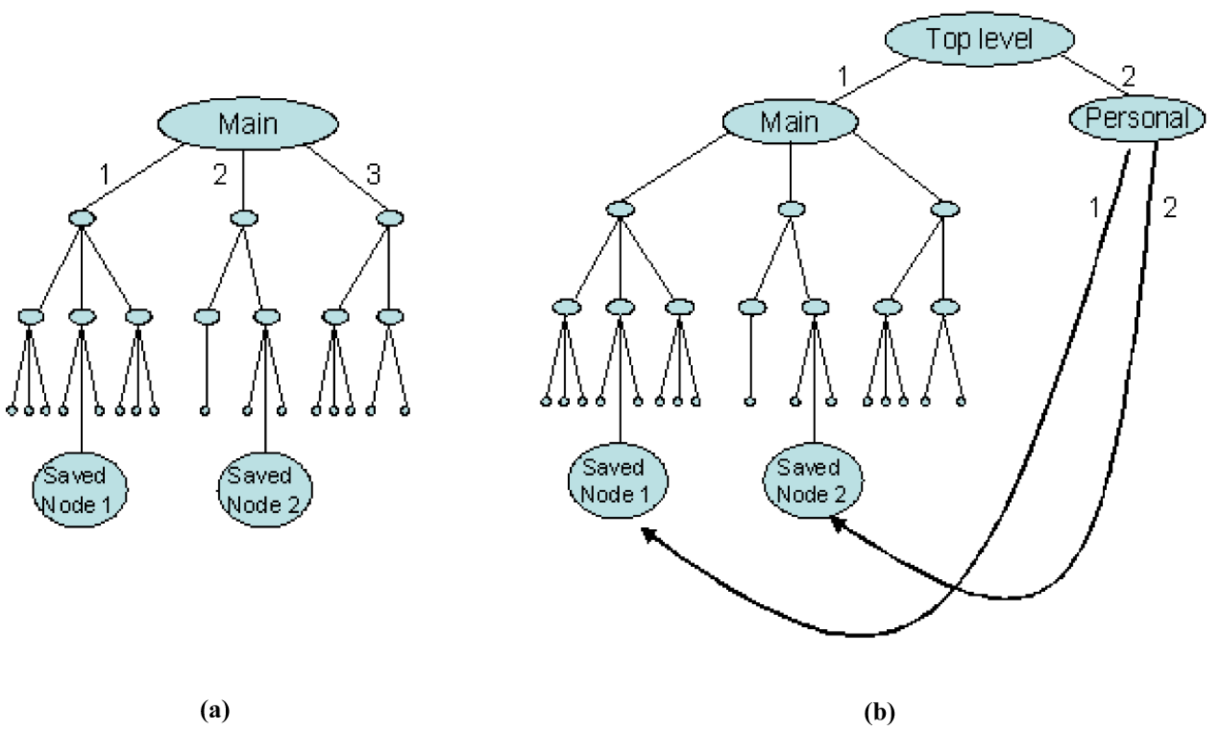

Fig. 3 Diagram showing the linking of bookmarked nodes in the list of personal options (**b**) to the nodes in the original tree (**a**). The user can avoid traversing the higher level sub-trees by selecting a node in their personal options

example, the user may have traversed a level, at which a choice would have to be made between "business related tasks" and "personal tasks". On a level further down in the hierarchy, the function of "reviewing the user's appointments" would then be to list the appointments based on this choice. Thus, simply saving the function name associated with the node in the personalized item list is insufficient to uniquely identify and specify the respective menu item. In saving menu nodes, it is necessary to distinguish the label of the node that has been saved from the function of the node in the original hierarchy.

To overcome ambiguities that may arise from simply associating the menu item's label in the list of personalized options, we allow the user to record a message using their own voice to identify the node that is being bookmarked in the hierarchy. For instance, if users are interested in bookmarking the node that accesses information about their personal appointments, they can, for example, save a recorded message by saying "personal appointments", which identifies the level of the node in the hierarchy related to this information. The recorded message is then saved with the bookmark. If the user wants to access the saved personal options later, the system replays these recorded messages. The duration of a user's recording is set in the current prototype version of VoiceMarks to a maximum of ten seconds but this limit is configurable.

To save nodes at any point in the hierarchy, the user presses a specific key—we used "9" in the prototype system. The system will then request the user to record a message serving as label for the prompt in the personal menu item list. The system stores the bookmarks and plays the list of personal prompts corresponding to the sequence, in which they have been saved, i.e. the node saved first in the list of personal menu items will be the first prompt to be replayed.

3.3 System architecture

The VoiceMarks prototype system was built by extending the VoiceXML (VXML) standard interface for touch-tone applications. We chose VoiceXML since it is a common standard for designing touch-tone applications. The test environment was set up with the Voicegenie™ (Voicegenie 2008) telephony platform.

Figure 4 depicts the major architectural components of VoiceMarks. The core of the system is a JSP (Java Server Pages) application that dynamically creates the VXML pages for the user. The JSP application creates the list of menu prompts as specified in the menu database. Each prompt is associated with a corresponding node in the original hierarchy. As users browse the system, bookmarked menus are saved in the database with associated pointers to the respective nodes in the hierarchy. Each bookmarked menu item is also connected to a voice recording to uniquely identify the original menu item. Informational entries and choices being made by a user during a session with the system are stored as state variables. These state variables can include entries like bank account number, login ID, etc. as well as values of choices made by the user, e.g. personal versus business related menus. All state variables that have been created during an individual session are stored in the menu database and are linked to the node that is being bookmarked, in order to describe its complete functionality. Upon entering a new session, the JSP application creates the hierarchy with the entire personal list of prompts including those that are saved as bookmarks.

4 User interaction dialogs

To test the concept of personalized touch-tone interfaces, we implemented prototypes for two applications: a system to retrieve bus transit information, and a second system to retrieve cinema ticketing information. The bus transit application was designed based on the real City of Winnipeg's bus transit telephone information system. We illustrate the various levels of user interaction with VoiceMarks through scenarios with both the bus transit information system as well as the cinema ticket purchase system. Figure 5 shows the basic menu structure of the bus transit information system. The user is able to retrieve the bus schedules for any given day. Additionally, the user is given the option to retrieve the current bus schedule. In either case the system requires that the user enter the bus stop number.

Table 1 shows the sequence of a direct dialog interaction with the user. In this scenario, the user saves the times of the current schedule for bus stop 60613. Since the user calls the system at 12:30 pm, all scheduled times after 12:30 pm are presented. The user then decides to save these timings for the selected bus stop for future reference. When the user bookmarks the menu

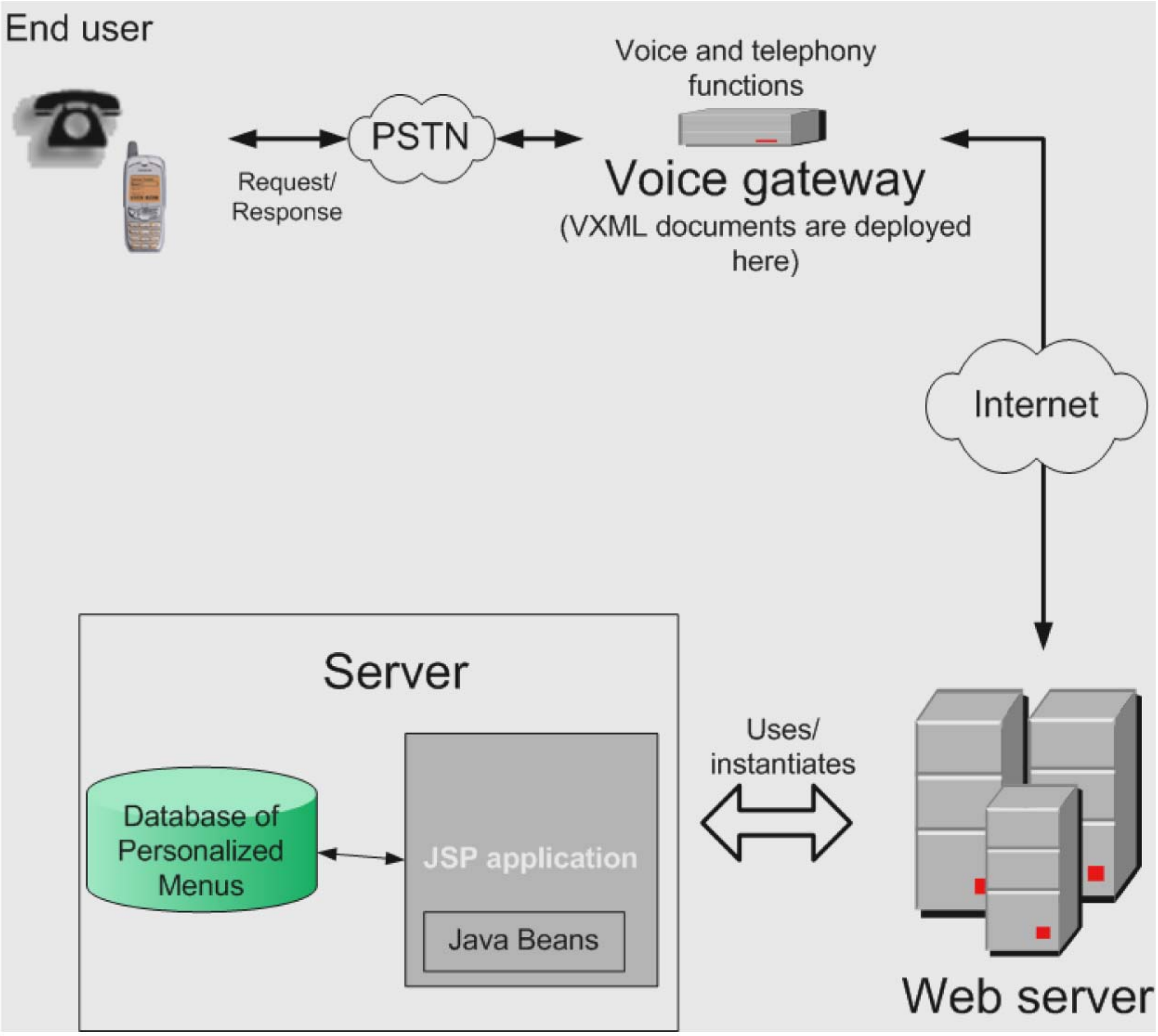

Fig. 4 Diagram showing the various components of the system

node, the system requests a recording from the user that will be stored with the new menu node in the personal item list and later be replayed, when the user accesses the personalized list.

Table 2 shows the sequence of prompts once a node has been saved in the system. The caller accesses the system at 3:30 pm and selects the previously saved node for the current timings. The system is routed to the node in the original menu hierarchy that retrieves the information from the current bus schedule for times after 3:30 pm. In this example, the user is able to avoid traversing one layer in the menu hierarchy by directly using the bookmarked node.

The number of dialog interactions between the user and the system is significantly reduced when more complex queries are involved. Table 3 illustrates the sequence of actions for a user requesting timings for a future bus schedule. In this scenario, the user goes to work at 1:00 pm on Thursdays and therefore wishes to save the schedule for 12:00 pm.

Table 4 demonstrates the reduced amount of interaction necessary for accessing information that requires multiple layers of user input. Upon entering a new session, the user can directly access the saved information. The system will therefore replay the bus schedules for the previously saved date and time and

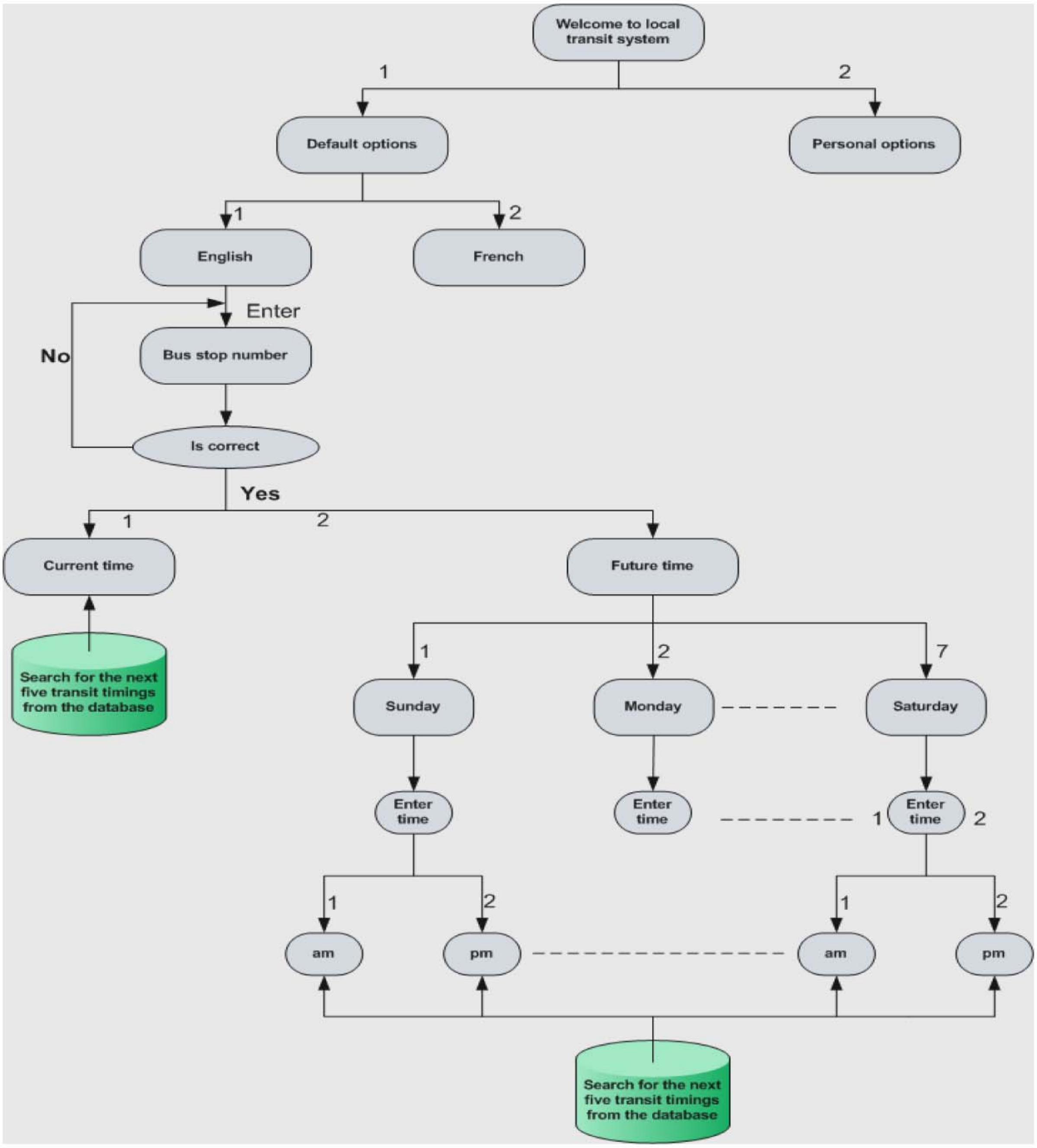

Fig. 5 Menu structure before saving bookmarks to the personal list

allow the user to bypass four layers of prompts. By minimizing the number of selections and required inputs to the system when information is being retrieved, user mistakes can be avoided and the number of errors can be reduced.

The second application, the cinema ticketing system, is in its essence very similar to the bus transit information system, but the structure of the hierarchy is in this case more complex, with a greater width of the hierarchy, i.e. more nodes per level. The structure of

Table 1 Interaction dialog for retrieving information and bookmarking a node in the bus schedule information system

System	Welcome to the local transit system. To save information at any level please press "9". Press 1 for English or 2 for French
Caller	Presses 1
System	Please enter the bus stop number
Caller	Enters 60613
System	1 for current schedule or 2 for future schedule
Caller	Presses 1
System	The current schedule is:
	Bus 75 at 1:00 pm, 1:35 pm and 2:20 pm
	Bus 60 at 1:40 pm, 2:15 pm and 2:45 pm
Caller	Presses 9
System	If you wish to save current times please record an associated prompt for this function
Caller	"Current schedule at stop number 60613, in English"
System	Item saved

Table 2 Interaction dialog for retrieving information about current bus schedules using voicemarks

System	Welcome to the local transit system. To save information at any level please press "9". Press 1 for default options or 2 for personal options
Caller	Presses 2
System	Press 1 for "Current schedule at stop number 60613, in English"
Caller	Presses 1
System	The current schedule is:
	Bus 75 at 3:40 pm, 4:15 pm and 5:10 pm
	Bus 60 at 4:10 pm, 4:55 pm and 5:35 pm

Table 3 Interaction dialog for accessing bus schedules for Thursdays at 12:00 pm and for saving the node of interest in the list of bookmarks

System	Welcome to the local transit system. To save information at any level please press "9". Press 1 for English or 2 for French
Caller	Presses 1
System	Please enter the bus stop number
Caller	Enters 60613
System	1 for current schedule or 2 for future schedule
Caller	Presses 2
System	Enter the day, 1 for Sunday and 7 for Saturday
Caller	Presses 5 (Thursday)
System	Enter the time
Caller	Enters 12:00
System	The bus timings are:
	Bus 75 at 12:15 pm, 12:50 pm and 1:35 pm
	Bus 60 at 12:20 pm, 1:40 pm and 2:15 pm
Caller	Presses 9
System	If you wish to save this schedule please record an associated prompt for this function
Caller	Says "Thursday schedule after 12:00 pm"
System	Item saved

Table 4 Dialog for retrieving future timings using voicemarks

System	Welcome to the local transit system. To save information at any level please press "9". Press 1 for default options or 2 for personal options
Caller	Presses 2
System	Press 1 for "Current schedule at stop number 60613, in English". Press 2 for "Thursday schedule after 12:00 pm"
Caller	Presses 2
System	The schedule is:
	Bus 75 at 12:15 pm, 12:50 pm and 1:35 pm
	Bus 60 at 12:20 pm, 1:40 pm and 2:15 pm

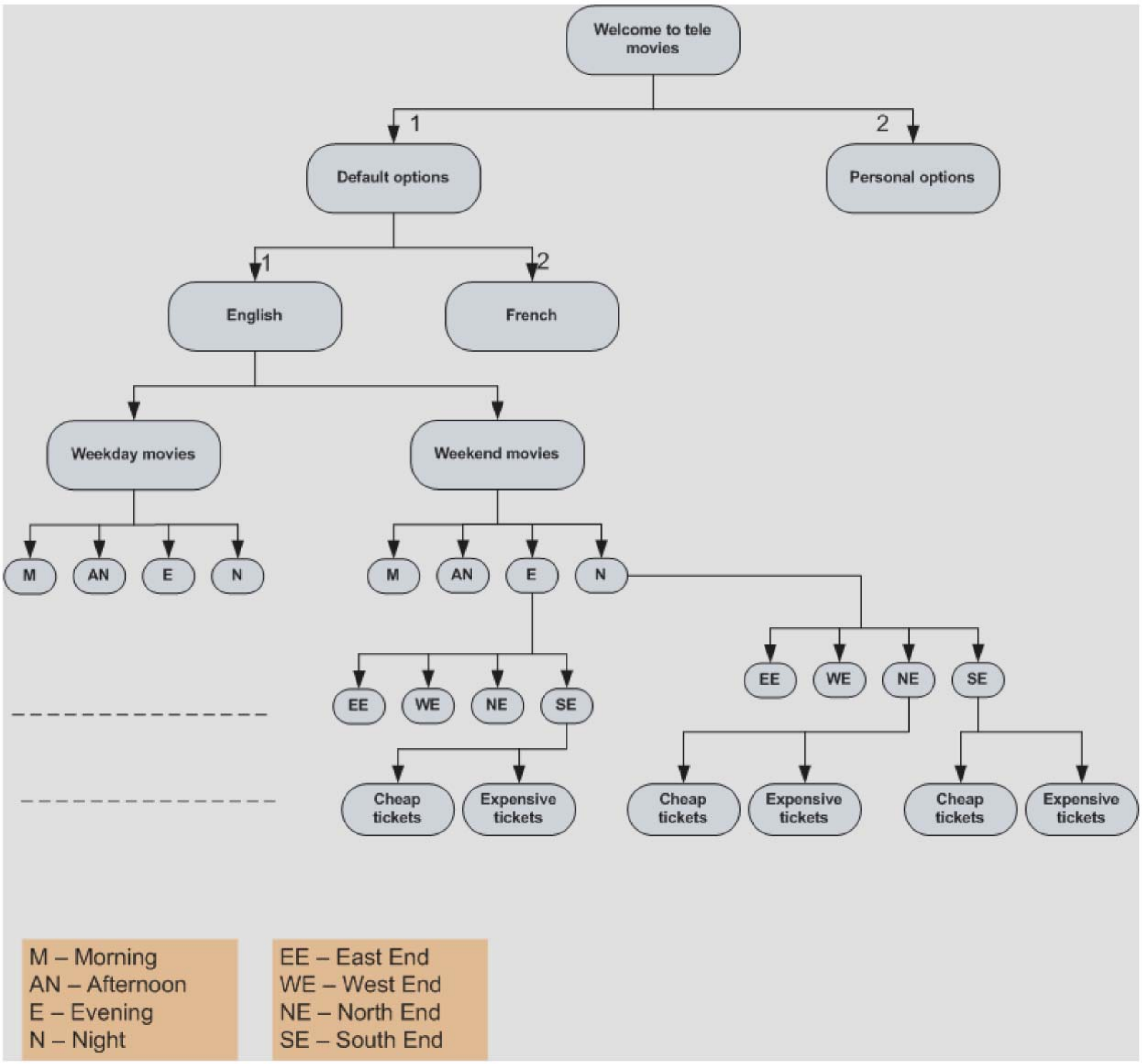

Fig. 6 Structure of the cinema ticketing interface before personal choices have been saved

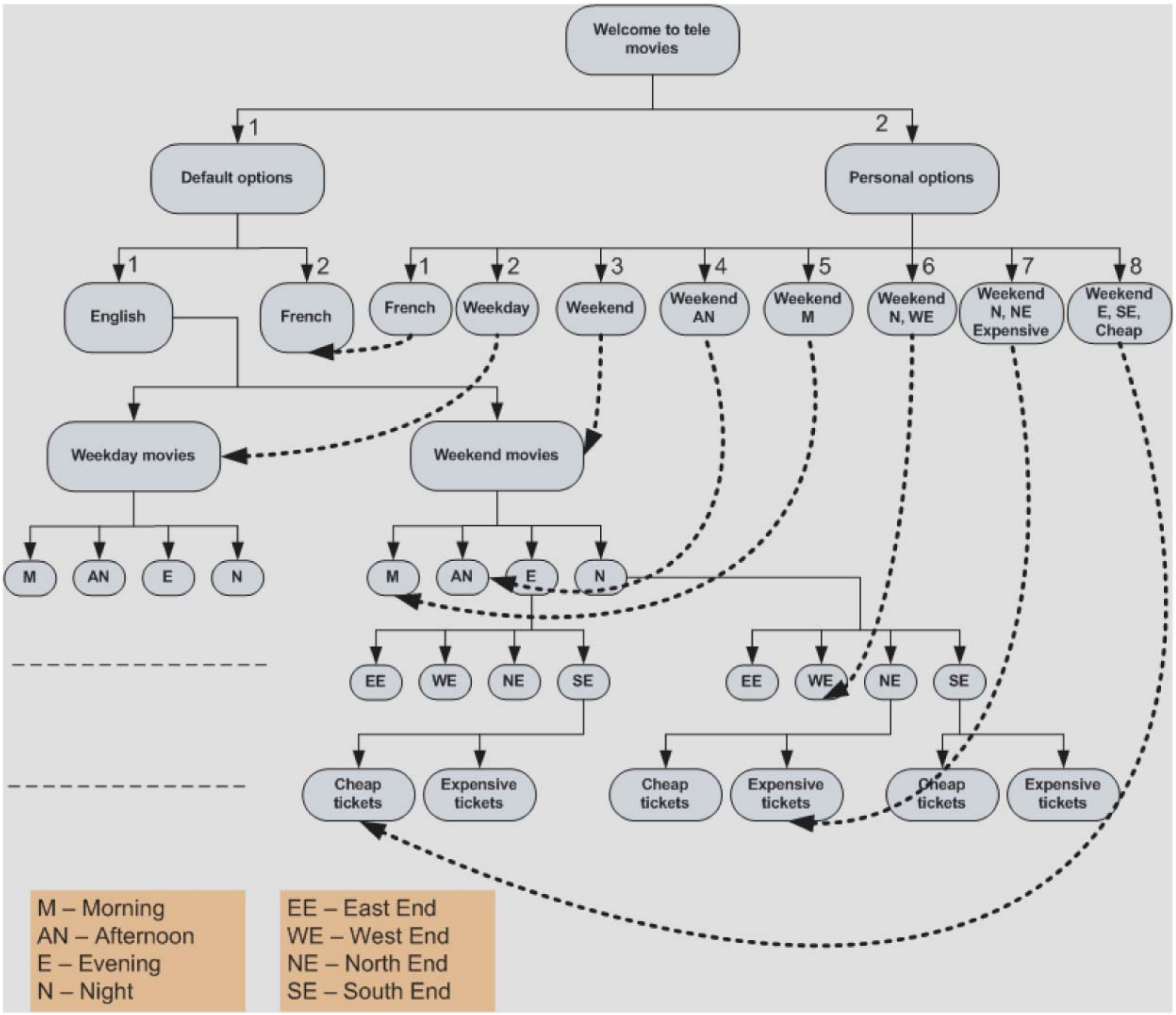

Fig. 7 Structure of the interface after personalization. The *dashed lines* show the connections between bookmarks and their original nodes

the application, before saving any options in the personalized list, is shown in Fig. 6.

Figure 7 shows the structure of the interface after some options have been saved in the personalized list.

5 User evaluation

The main goal of the evaluation was to determine whether personalization improves navigation in auditory interfaces. In order to evaluate the effect of personalization, the VoiceMarks interface was compared to a conventional touch-tone interface for the two applications described above, i.e. a bus schedule information system with a simple menu hierarchy, and a cinema ticketing system with a more complex hierarchical menu structure.

5.1 Hypothesis

Based on previous studies related to personalizing menus (McGrenere et al. 2002; McGrenere and Moore 2000), we anticipated that participants will perform search and navigation tasks in a voice based menu hierarchy on average faster with the personalized interface than with a traditional interface. In addition, we anticipated higher levels of satisfaction with the personalized interface than without.

5.2 Method

5.2.1 Participants

Ten students from a local university volunteered to participate in the study. None of the participants reported a history of auditory disorders. All the participants reported that they had heard about or used touch-tone interfaces in the past. The participants also stated that they had previous experience listening to synthetic voices. They all were fluent in English.

5.2.2 Materials

The experiment was performed on the Voicegenie™ platform (Voicegenie 2008), which simulates the effects of using a telephony system to access information. The participants interacted with the bus transit information system (a simple menu hierarchy) and the cinema ticketing system (a complex menu hierarchy).

5.2.3 Design

A within-subjects experiment was conducted to test the hypothesis that users perform search and navigation tasks faster with the personalized interface. A fully balanced Latin-square design (i.e. half the participants performed on the Conventional menu first then VoiceMarks, the other half did the reverse; we presented simple menu style first then complex) was used to reduce learning effects. The participants performed the experiment with both applications starting with the bus transit system and then the cinema ticketing system. The experiment focused on measuring objective data like the time taken to locate a requested node and the number of clicks to reach a destination node. We also obtained information about the user satisfaction through questionnaires.

5.2.4 Procedure

Participants conducted the experiment one at a time. Participants were given a brief introduction to touch-tone systems and pertinent information about VoiceMarks. The information about VoiceMarks consisted of introducing the concept of the personalized menu interface, without any bias regarding the interaction technique. Next, we explained the tasks the participant needed to perform in the experiment, followed by a demo of the conventional touch-tone interface as well as of the VoiceMarks interface. In particular, participants were shown what commands they have to use to bookmark a node in the personalized interface, and how to access the bookmarked node later. The experiment started, when the users indicated that they were comfortable with both interfaces. Before answering the questions listed in the set of tasks given in Table 5 and Table 6, the participants were given a booklet with a list of items they had to bookmark in the VoiceMarks interface prior to doing the requested tasks. The participants were asked to bookmark several nodes for each application.

Once the experiment began, the subjects had to locate the items listed in Table 5 and Table 6 in the hierarchy. The tasks were performed by all participants with both interfaces. The participants were assigned to two groups: one group used the VoiceMarks system first to perform the tasks, and then switched to the traditional interface; the other group used the traditional interface first, and then the VoiceMarks system.

After completing the tasks with both interfaces, the users were asked to provide a subjective evaluation regarding the use of the conventional touch-tone interface and of VoiceMarks (see Table 7).

6 Results and discussion

To gain a quantitative evaluation of the difference between the conventional touch-tone interface and VoiceMarks, objective data were obtained through two measures (see Sects. 5.2.3 and 5.2.4). Measured variables include the time each participant needed to complete each of the tasks shown in Tables 5 and 6 (Sect. 6.1), as well as the number of keystrokes used for each task (Sect. 6.2). For each participant, these measurements were obtained for both interfaces, according to a fully balanced Latin-square design. After the task session, the users were asked to complete a questionnaire, which served to obtain information about the user satisfaction with both types of interface (see Sect. 6.3).

6.1 Task completion time

The results of the experiment for the first application (the bus transit information system) suggest that

Table 5 List of tasks for the bus transit application performed by the participants

Task number	Task description (bus transit information application)
T1	Find the current bus timings for the bus stop number "60613"
T2	Find the bus timings for Sunday after 12:00 pm, for the bus stop number "60613"
T3	Find transit timings for Monday after 22:25 for the bus stop number "23422"

Table 6 List of tasks for the cinema ticketing application performed by the participants

Task number	Task description (cinema ticketing application)
T1	Find the list of English language expensive Hollywood movies that are played at nights on weekends in the north end
T2	Find the list of English language cheap Hollywood movies that are played in the evenings on weekends in south end
T3	Find the list of English language cheap Hollywood movies that are played at nights on weekends in the north end
T4	Locate English language, weekends, night, west end
T5	Locate English language, weekends, night, west end
T6	Locate English language, weekdays
T7	Locate English language, weekends, afternoon
T8	Locate French language

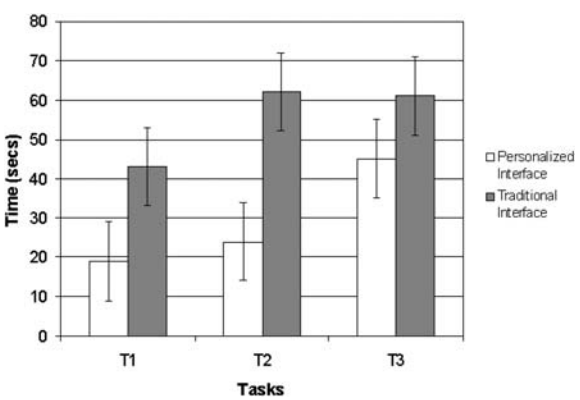

Fig. 8 Comparison of average task completion times of tasks T1–T3 with the bus transit information system, for both interfaces

the subjects performed all tasks from Table 5 comparatively better (faster) with the personalized interface than with the traditional interface. The average task completion times for the three tasks are shown in Fig. 8.

A t-Test yields that there is a significant difference regarding the task completion time between the personalized interface and the traditional interface, $T(9, 0.05) = 7.781$, $p < .001$. The improvement in performance using the personalized touch-tone interface compared to the traditional interface is attributed to a shorter search path for a requested menu item through the personalized item list. The results confirm the hypothesis that subjects perform tasks faster with the personalized interface than with the traditional interface.

In the second application (the cinema ticketing system), the subjects were required to complete the eight tasks listed in Table 6. The results with the second application showed that subjects performed the tasks faster with the default interface than with the personalized interface.

For an overall comparison of the performance regarding time used, we averaged the completion times across all tasks for each application and each of the two interfaces, i.e. VoiceMarks and the conventional interface (see Fig. 9). A t-Test shows a significant difference in performance between the personalized interface and the conventional interface, $T(9, 0.05) = 4.949$, $p < .001$. The overall results show that the subjects performed the tasks 1.15 times faster with the personalized interface than with the conventional interface (Fig. 9). This supports our hypothesis, that users work faster with the personalized interface of VoiceMarks compared to a standard interface. We observed, however, also an inferior performance with VoiceMarks for the second application. In this case, users performed worse, i.e. took longer, to complete the tasks with VoiceMarks than with the traditional interface. We assume that the reason for this inferior performance of VoiceMarks is due to the larger number (8) of voicemarks that had to be saved by the users prior to the performing the tasks. The results highlight an effect which we had not anticipated prior to the experiments: the use of a personalized menu can lead to a long list of personal bookmarks, and thus can have a negative impact on search and retrieval times. In contrast to visual interfaces, the user of telephony interfaces has to search sequentially through the sequence of voicemarks, in order to find a specific item. For example, if the user has to find and select the sixth voicemark in a personalized list, s/he has to listen to the previous five voice prompts first. Thus, the access time to bookmarked nodes can increase in VoiceMarks considerably with the length of the list of voicemarks and the length of the voicemarks themselves, in a form which is not comparable to visually organized lists of bookmarks.

For example, in the cinema application, in order to perform task T2, the subjects had to pass seven options in the personalized list, and to listen to the attached voicemarks, before they could select the eighth option, representing the requested item (see Fig. 11). In the standard interface, on the contrary, the subjects had to traverse only five levels in the hierarchy, in order to locate the node for solving T2 (see Fig. 11).

Another factor that facilitated the navigation in the traditional interface is training through the introductory session. Since the users had some previous experience with the applications through the introduction, they were able to select some menu items more quickly with the traditional interface. For example, in the cinema ticketing application, the system prompts at level four "Press 1 for weekdays. Press 2 for weekends". As soon as the users heard "Press 1 for weekdays", they interrupted the system by pressing "2" in the keypad, because they remembered this option and thus did not wait for the second prompt to play. In such telephony systems, with easy to remember options, and accompanying significant prompts, users seem to be able to save a significant amount of navigation time

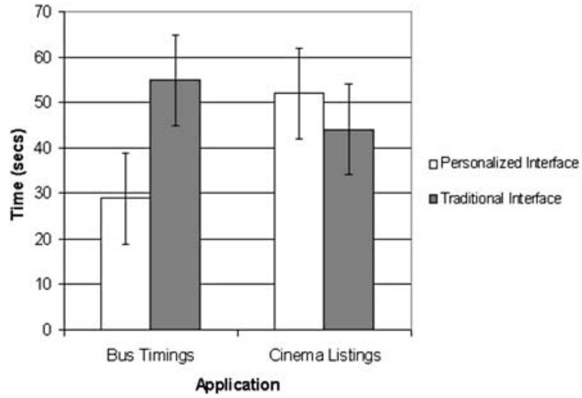

Fig. 9 Average task completion times for the two applications and both interface types

in the traditional interface through actions like barging in and skipping menu options.

On the other hand, we noticed that participants were able to quickly select nodes, when the required options were placed closer to the beginning of the menu. Thus, the ordering of the bookmarks in the menu plays a significant role regarding users' performance and the efficiency of using personalized touch-tone interfaces.

The results of our experiments thus indicate that, under specific circumstances, the average navigation time increases with VoiceMarks, depending on the number of voicemarks or bookmarked items stored in the personal list.

6.2 Keypad selections

We also recorded the number of keypad presses required to complete the tasks. The average numbers of keypad-presses that are required to complete the tasks for each application are shown in Fig. 10.

The results suggest that the number of selections or keypad presses is significantly smaller in VoiceMarks compared to the traditional interface, for both applications. One possible explanation is that the default interface forces the user to select an option at each level in the hierarchy, before the user can move to the next lower level. For example, in order to reach T2 in the cinema listings, the user has to press at least 6 times a key, since 6 levels in the hierarchy have to be traversed to reach the requested node. However, the same destination node can be reached in VoiceMarks with just two keypad clicks (Fig. 11).

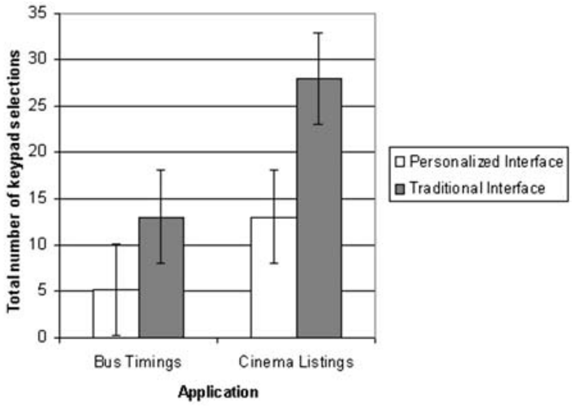

Fig. 10 Average numbers of keypad presses required to complete the tasks

Thus, compared to the results described in Sect. 5.1, that users needed more time in VoiceMarks than in the traditional interface for certain tasks, due to the length of voice-prompts and the necessity to listen to them sequentially. We do not observe degraded performance with VoiceMarks when considering the number of keystrokes.

It seems that the two effects of prolonged time needed due to having to listen to voice prompts from the personal list of voicemarks, and the reduced number of keystrokes needed to make the proper selection and retrieve the requested menu item from the hierarchy, are counter-balanced. Thus, it seems suitable to add features suggested for improving the navigation time in telephony systems with menu hierarchies (see Sect. 1), like barging-in and skip-and-scan, also to the personal list of voicemarks/bookmarks.

6.3 User experience/subjective evaluation by users

The subjective evaluation of the system by the users is summarized here. We asked the participants at the end of the study to answer a set of questions addressing both types of interfaces, i.e. with and without personalized menus. Two questionnaires were administered: the first one was designed to obtain the users' opinion on the two types of systems; the second questionnaire was designed to gather information about the users' preferences regarding specific aspects of the personalized touch-tone interface.

The first post-experiment questionnaire contained questions comparing the performance of the personalized touch-tone interface to the conventional touch-tone interface, related to the tasks the participants performed with the systems. For each task, the participants rated the interface that they felt was better suited for the task. The analysis of the first questionnaire is summarized according to efficiency, navigation, learnability, and ease-of-use.

Efficiency. 80% of the users favored the VoiceMarks interface for accessing menu items efficiently. There was general consensus that participants found specific information quicker, when they used the personalized touch-tone interface, if items were bookmarked beforehand. All the users confirmed that the personalized touch-tone interface would facilitate their daily interactions with telephony systems.

Navigation. 60% of the users favored the personalized touch-tone interface to choose the right menu in order to reach the intended destination efficiently and effectively.

Learnability. 50% of the users suggested that the personalized touch-tone interface is as easy to learn as the conventional interface. Additionally, users indicated that by recording their own voice prompt, language did not impose a barrier in accessing the bookmarked information. The other half of the participants found the personalized touch-tone interface and the use of VoiceMarks was slightly more difficult to learn, and that it initially requires more effort to use this type of interface.

Ease-of-Use. 90% of the users suggested that once they got accustomed to VoiceMarks, they found it easy to use. After repeated use, they understood the system and were comfortable using and navigating the bookmarks. 80% of the users said that they would like to use the personalized touch-tone interface on an everyday basis as they could skip levels and options that slow down their navigation speed. In general, the users highly preferred the personalized touch-tone interface for flexibility and ease of use.

The second post-experimental questionnaire was designed to acquire information about the users' perception of specific aspects of VoiceMarks. Participants responded to questions using Likert-scale replies, ranging from 1 to 5 (1 = strongly disagree, 5 = strongly agree). The averaged ratings are summarized in Table 7.

The ratings correspond to our initial intuition regarding the usability and practicality of personalized menus. From the responses to the questionnaire, it can be concluded that most participants are in favor of the

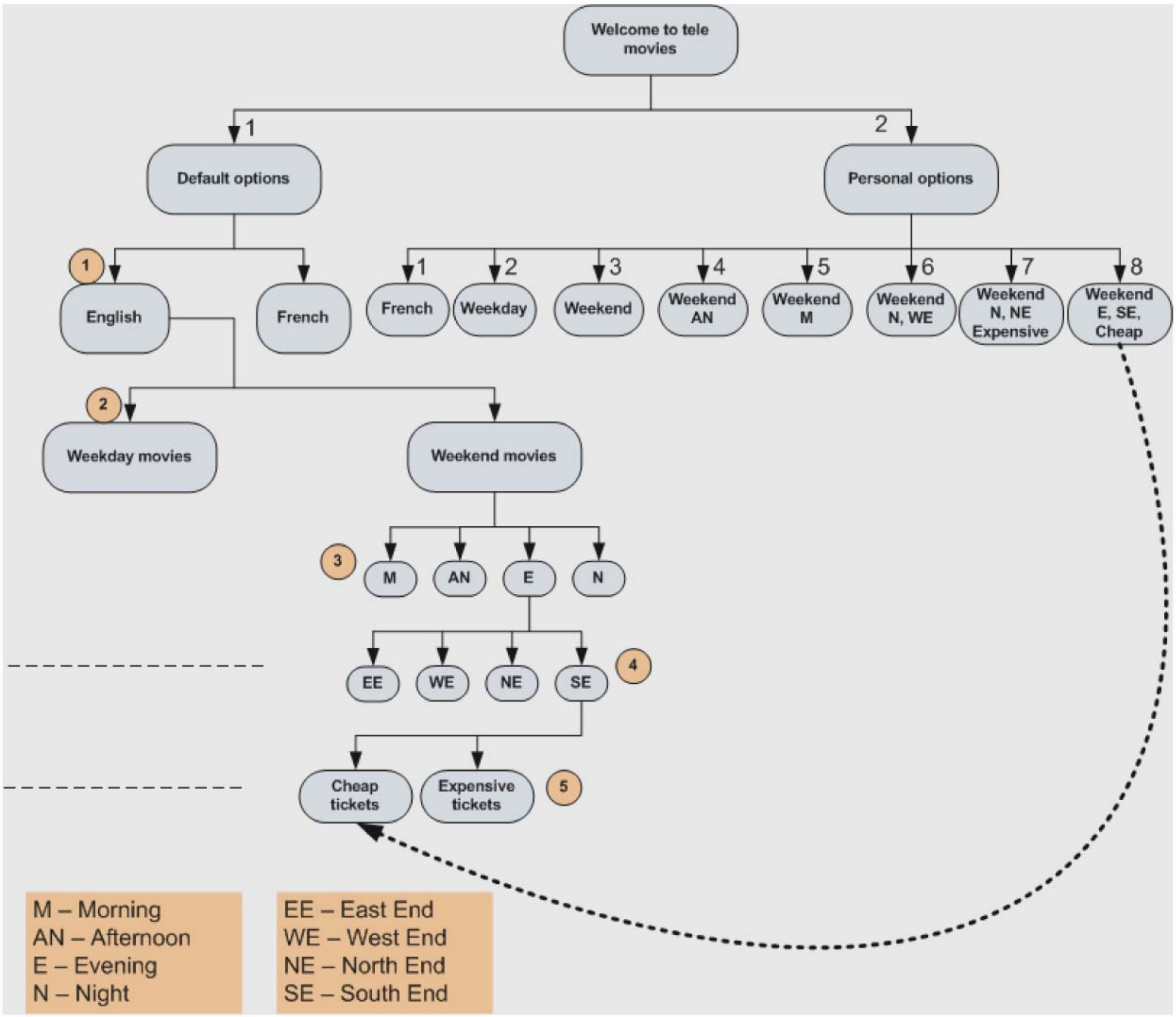

Fig. 11 Number of keypad presses required to locate a node in the hierarchy

concept of personalizing menu interfaces (Q11). The questionnaire revealed that the majority of participants would prefer to use the personalized interface for daily interactions (Q3, Q12). They also noted that it is easy to learn (Q7), and intuitive and easy to use (Q9). Users liked the idea of recording human voice for saving options in the personalized list (Q4, Q11) and found listening to their own voice prompt pleasant (Q6).

One of the major problems, according to this evaluation, was that the participants found it difficult to remember the options they have saved in their personalized lists (Q1 and Q10). One way to address this problem would be to allow users to scan and skip irrelevant menu items, which would save a significant amount of time searching through the personal list of voicemarks.

Overall, the users supported the idea of VoiceMarks and its concept of personalizing menu-based interfaces. Users found the feature most useful for performing frequent, repeated, routine actions, since this would optimize the trade-off between personalizing the system and getting used to it, and the benefits from using the personalized menu. A feature that participants enjoyed particularly was the ability to record their own voice prompts for the personalized menu options. This feature avoids problems of ambiguous and unclear prompts present in standard touch-tone interfaces. Designers of such interfaces always have to

Table 7 Summary of the second post experiment questionnaire

	Statement	Mean (1–5)
Q1	Difficult to locate menu options using VoiceMarks	2.4
Q2	I am able to complete my tasks efficiently using VoiceMarks	4.0
Q3	I am satisfied with the speed of accessing information using VoiceMarks	3.5
Q4	I like the technique of recording my voice for providing information about personalized menus in the "Personalized Menus list"	4.1
Q5	Recorded menus in the "Personalized Menus List" improve navigation	4.1
Q6	Replay of recorded human voice is pleasant	4.1
Q7	VoiceMarks interface is easy to learn	3.5
Q8	I know how to use the system after the trial session	3.3
Q9	VoiceMarks is easy to use	4.5
Q10	I can remember the list of menus that I have saved in the "Personalized Menus List"	2.5
Q11	Recording option is a necessary element of the interface	4.3
Q12	Personalized touch-tone interface is the preferred choice for daily transactions	4.4
Q13	Overall satisfaction with VoiceMarks	4.1

be concerned about the proper wording of the voice prompts, which is not an issue here, since the user can record their own prompts in VoiceMarks.

In summary, users "liked" the feature of personalizing the menus and were in strong favor of having the feature implemented in those telephony systems, which they use on a regular basis.

7 Conclusion and outline of future work

We described a method for personalizing menus in touch-tone interfaces. This method, referred to as *VoiceMarks*, consists of allowing a user to create bookmarks for specific nodes of interest in the hierarchy. These bookmarks are stored in a personal menu list and can later be accessed by the user directly, without having to navigate and search in the hierarchy. The concept of personalized menus was implemented using two applications, a bus schedule information system and a cinema ticket purchase system. We conducted a study with 10 participants, evaluating the VoiceMarks interface for each of those applications in comparison to the respective conventional interface. The study used objective measurements to assess user performance, like the time to locate a specified menu item and the number of keystrokes used, as well as a subjective evaluation of the two interface types (conventional versus VoiceMarks) regarding user satisfaction.

The results of our study show that the time to access specific menu items can be reduced significantly with the personalized interface. The use of long lists of bookmarks, however, can cause an inferior performance of VoiceMarks compared to the traditional interface. Some modifications to the system, as outlined further below, thus seem to be useful and necessary.

Subjective evaluations by the participants were positive. They were overall pleased with the effectiveness, efficiency and ease of use of the personalized touch-tone interface. They reported that it requires some effort to learn to work with the personalized interface and thus found the feature most useful as an addition to telephony systems they use on a regular basis. Users obviously liked their personal voice as prompt. Thus, concerns about mixing human speech and synthetic speech, which can imply a lack of consistency and degraded performance (Gong and Lai 2001), do not seem to apply to our method of personalizing interfaces described in this paper.

Although the initial results are encouraging, it is clear that the current version of VoiceMarks has its limitations and drawbacks. A major issue is that the ease of accessing information using personalized menus depends heavily on the individual user, i.e. the way in which the user selects and saves VoiceMarks to their personal list. There is a risk that users create unhelpful personalized menus, e.g. through recording improper prompts or entering too many or insignificant bookmarks.

Some simple modifications of the VoiceMarks system should avoid these obvious disadvantages of the prototype version. There are essentially two different ways to implement such improvements: through user-initiated modifications, and through automatic changes, based on observing the user's behavior.

Future work will comprise of allowing users to perform maintenance operations on their personal list of menu items. For example, users should be able to attach new recordings to bookmarks and to change the order of VoiceMarks in the personal list according to their preferences. Features like scan-and-skip and barging-in should also be applicable in the list of bookmarks, so that users have the flexibility to skip irrelevant voicemarks and to make selections, before a voice prompt is finished, at any time.

Useful and significant changes to the VoiceMarks system could also be done in an automated way, based on observing the users' behavior, when they interact with the telephony system (Smyth et al. 2005). It is possible to determine often visited menu items and separate them from those, which are rarely or never used, in the menu. Based on the relative frequency of use, the system can initiate a re-ordering of bookmarks, moving frequently used ones to the beginning of the list, as well as suggest the deletion of rarely used bookmarks and the introduction of new bookmarks for frequently visited nodes. The VoiceMarks system can thus adapt to the user's specific needs and preferences by featuring those frequently visited nodes in the personalized menu.

The final conclusion is that the method to bookmark nodes in menu-based telephony interfaces, developed and tested in VoiceMarks, can, with some additional modifications, significantly improve the user performance and satisfaction in interactions with menu-based telephony systems. This is also the perception of the participants of our study, who appreciated and welcomed the VoiceMarks concept of a personalized touch-tone interface. We believe that the concept of VoiceMarks, if applied appropriately, improves the efficiency and usability of telephony interfaces, so that they regain widespread acceptance and popularity among users.

References

Baecker, R., Booth, K. S., Jovivic, S., McGrenere, J., & Moore, G. (2000). Reducing the gap between what users know and what they need to know. In *Proceedings of the conference on universal usability* (pp. 17–23). New York: ACM Press.

Balentine, B. (1999). Re-Engineering the speech menu. In D. Gardner-Bonneau (Ed.), *Human factors and voice interactive systems* (pp. 205–235). Norwell: Kluwer Academic.

Balentine, B., & Morgan, D. P. (1999). *How to build a speech recognition application*. San Ramon: Enterprise Integration Group, Inc.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, *4*(1), 11–44.

Brewster, S. A. (1998). Using nonspeech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *5*(3), 224–259.

Gardner-Bonneau, D. (Ed.) (1999). *Human factors and voice interactive systems*. Norwell: Kluwer Academic.

Carroll, J., & Carrithers, C. (1984). Blocking learner error states in a training-wheels system. *Human Factors*, *4*(26), 377–389.

Gong, L., & Lai, J. (2001). Shall we mix synthetic speech and human speech? Impact on users' performance, perception, and attitude. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 158–165). New York: ACM Press.

Karat, C. M., Halverson, M., Horn, D., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 568–575). New York: ACM Press.

Leplatre, G., & Brewster, S. A. (2000). Designing non-speech sounds to support navigation in mobile phone menus. In *International conference on auditory display*, 2000.

Linton, F., Joy, D., Schaefer, P., & Charron, A. (2000). Owl: A recommender system for organization-wide learning. *Educational Technology & Society*, *1*(3), 62–76.

McGrenere, J., & Moore, G. (2000). Are we all in the same "bloat"? In *Graphics interface*, pp. 187–186, 2000.

McGrenere, J., Baecker, R. M., & Booth, K. S. (2002). An evaluation of a multiple interface design solution for bloated software. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 164–170). New York: ACM Press.

McInnes, F. R., Nairn, I. A., Attwater, D. J., Edgington, M. D., & Jack, M. A. (1999). A comparison of confirmation strategies for fluent telephone dialogues. In *Human factors in telecommunication*, 1999.

Pu, P., & Faltings, B. (2002). Personalized navigation of heterogeneous product spaces using smart client. In *Proceedings of the 7th international conference on intelligent user interfaces* (pp. 212–213). New York: ACM Press.

Resnick, P., & Virzi, R. A. (1992). Skip and scan: Cleaning up telephone interfaces. In *Proceedings of ACM CHI'92* (pp. 419–426). New York: ACM Press.

Roberts, T. L., & Engelbeck, G. (1989). The effects of device technology on the usability of advanced telephone functions. *ACM SIGCHI Bulletin*, *20*(SI), 331–337.

Shajahan, P., & Irani, P. (2004). Representing hierarchies using multiple synthetic voices. In *8th international conference on information visualization* (pp. 885–891). Los Alamitos: IEEE Computer Society.

Smyth, B., McCarthy, K., & Reilly, J. (2005). Mobile portal personalization: tools and techniques. In B. Mobasher & S. S. Anand (Eds.), *Intelligent techniques in Web personalization* (pp. 255–271). Berlin: Springer.

Suhm, B., Freeman, B., & Getty, D. (2001). Curing the menu blues in touch-tone voice interfaces. In *Extended abstracts on human factors in computer systems* (pp. 131–132). New York: ACM Press.

Sumikawa, D. A. (1985). *Guidelines for the integration of audio cues into computer user interfaces*. Livermore: Lawrence Livermore National Laboratory.

Tatchell, G. R. (1996). Problems with the existing telephony customer interface: The pending eclipse of touch-tone and dial-tone. In *Conference companion on human factors in computing systems* (pp. 242–243). New York: ACM Press.

Voicegenie. www.voicegenie.com (2008).

Yankelovich, N., Levow, G., & Marx, M. (1995). Designing speechacts: issues in speech user interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 369–376). New York: ACM Press/Addison-Wesley.

Yin, M., & Zhai, S. (2006). The benefits of augmenting telephone voice menu navigation with visual browsing and search. In *Proceedings of ACM CHI'06*, pp. 319–328, 2006.