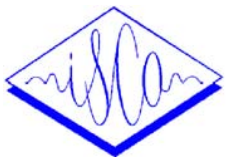


EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

RECALL MEMORY FOR EARCONS

Dawn Dutton, Candace Kamm, Susan Boyce

AT&T
2K-330, 101 Crawfords Corner Road, Holmdel, NJ 07733
dldutton@att.com cak@research.att.com sjboyce@att.com

Abstract

Our voice enabled telecommunications service, Annie, is a prototyping system that gives users the ability to access a variety of telephone-based services by voice. The user interface of Annie uses an anthropomorphic "personal assistant" metaphor. The user can maintain a "conversation-like" dialog with Annie, but user input is limited by the grammar-constrained automatic speech recognition (ASR) technology used in the service. Because the grammars change depending on the state the user is in, the system must provide clear recognition feedback and orienting information throughout the dialog. Verbal recognition feedback is tedious and time-consuming for the frequent, expert user. This paper describes an experiment that explores the feasibility of providing non-verbal recognition feedback and orienting information through the use of earcons, or auditory icons. Users of Annie were exposed to five earcons presented in parallel with verbal recognition feedback for a minimum of five days. Subsequently, users were asked to recall the identity of each of the five earcons alone. Subjects were able to reliably recall each of the earcons. Since users could recall the earcons, it is feasible that the non-verbal earcons could replace the lengthier verbal recognition feedback.

10.21437/Eurospeech.1997-490

1. Introduction

1.1 Service Description

Annie is a prototyping platform that provides a variety of telephone-based services. It is currently being used by 36 "friendly" subscribers (employees) and their families. The goal for this project is to explore which voice-based information/communication features telephone customers want to use, and how the underlying speech technologies must be improved and user interface must be designed to create compelling, easy-to-use speech interfaces.

The platform is a multi-channel, client-server system that supports mixed-initiative grammar-constrained dialog for call control and messaging. Annie is available 24 hours a day, and users can access the system both from home and away from home. Each user's home and office number are known to Annie, so when the user calls in from either location, the system automatically detects the incoming number, retrieves the user's personal directory, and greets the user. Annie currently handles an average of 105 calls per day, from an average of 22 different accounts.

The current functionality includes repertory dialing (from a personal directory), dialing by spoken numbers, access to employee directories, and voice-based and web-based label administration of the entries in the user's personal directory.

The user interface to Annie is based on an anthropomorphic "personal assistant'" metaphor, aimed at achieving a "conversation-like" interaction, within the limitations of the grammar-constrained automatic speech recognition (ASR) technology. The current interface was designed with assumption that the user is an "expert" with the system. This user model assumes that:

- the user knows the basic functionality of Annie and how to navigate the system, and
- the user knows and generally remembers some of generic commands that Annie understands.

The underlying speech recognition technology uses speaker-independent context-dependent phone models for telephone speech for most grammars. The grammars, and thus what users can say, in Annie are changed based on the state the user is in. At various points in the dialog, different grammars are active, but the system is not strictly hierarchical. Some commands and the user's personal labels are present in some, but not all, grammars.

There are a number of modules in the application. If the user attempts to access Annie from a number not recognized by the system, the user must provide and identification number in the Remote Access module. Once Annie has recognized the user, the user enters the system at the Top Level. To accomplish any task the user must move between modules of the system. Generally, when the user finishes a task, the user is returned to the Top Level. To place a telephone call using a personal label (a name in their personal directory) or telephone number, the user enters the Outbound Calling module. To administer personal labels, the user must enter the Label Administration module. To use repertory dialing for an employee, the user must enter the Employee Directory.

1.2 Recognition Feedback and Orientation

Orienting the user to their location in Annie is particularly important because the recognition grammars change based on the state the user is in. That is, the user can say a wide range of utterances at most points in the dialog, but they can not say anything, anytime. In addition, because the recognition performance of ASR systems is imperfect, the behavior of Annie must clearly indicate to the user what the system has recognized. This can be particularly troublesome when the recognition outcome results in the system moving from one module to another, where new tasks and different grammars are available to

the user. Without orienting information, the user might not realize that a recognition error has occurred. For example,

System:	Annie, here.
User:	Call Mabel
System:	**What's next?**

In the above example, Annie misrecognized the user's utterance "Call Mabel" as "Get my labels." The system response "What's next?" does not tell the user what the recognition result was. The following example shows how verbal orienting information allows the user to quickly diagnose that a recognition error has occurred and to correct the error.

System:	Annie, here.
User:	Call Mabel
System:	**Labels. What's next?**
User:	Cancel
System:	Annie, here.
User:	Call Mabel
System:	Calling Mabel…

In the above example, , the user must say "Cancel" to correct a recognition error. When the user says "Cancel," or one of its synonyms, Annie undoes the action taken by the system as a result of the incorrect recognition result, and returns the user to the point in the dialog where the recognition error occurred. In the above example, when the user says "Cancel," the user leaves Label Administration, returns to the Top Level, and is reprompted with "Annie, here."

In other words, at almost every point in the dialog, Annie's turn is in fact an implicit confirmation of the most recent recognition result. This confirmation strategy was developed because the more explicit confirmation strategies commonly used in other services using speech recognition were insufficient for this service. Annie is largely used by expert users who use the service multiple times per day. The use of direct confirmations (e.g. "Did you say…") after each recognition result would be odious to the frequent, expert user. An implicit confirmation in which the recognition result is echoed back to the user followed by a pause to allow the user to indicate if the recognition result was in error would likewise be tedious for the frequent, expert user.

It was important to develop a confirmation strategy that would satisfy all of the following constraints:

- Allow frequent, expert users to quickly detect recognition errors.
- Allow users to quickly correct recognition errors.
- Allow the dialog to progress naturally in the great majority of cases when the recognition result was correct.

Consequently the implicit confirmation strategy shown in the above example was developed.

1.3 Why Use Earcons?

While it is important to give feedback to users so that they can quickly detect and correct speech recognition errors, even the minimal feedback provided by Annie is time-consuming and tedious to the experienced user when the system has correctly interpreted the user's speech. In the example shown above, the prompt "Labels. What's next?" provided both feedback and a question to carry the dialog forward. A solution to this problem might be, in some situations, to provide non-verbal recognition feedback and orienting information to the user in parallel with verbal prompts that move the dialog along and maintain the conversational quality of the interaction. Unlike the verbal feedback, the non-verbal feedback would not initially be associated with the recognition result. Such an association would have to be learned. Initially, then, the verbal and non-verbal feedback would necessarily have to be paired until the association was learned. For example:

System:	Annie, here.
User:	Get my labels.
System:	**~Non-verbal Recognition Feedback~**[1]
	Labels. What's next?
User:	Add a label
System:	Who should I add?

Non-verbal auditory feedback is present in many visual interfaces. Some experimental and non-commercial graphical interfaces have successfully incorporated sound to convey redundant and supplementary information about system events (Brewster, et al, 1994, Brewster, et al, 1995, Gaver, 1989, 1991). Gaver (1989, 1991) found that auditory icons increase a user's sense of engagement and satisfaction with an interface.

Non-speech sound, or earcons, could effectively complement the spoken prompts in a telephony interface. A major benefit to using earcons in an auditory-only interface is that earcons could be used to provide feedback and orienting information to the user in parallel with verbal prompts that move the dialog with the user along. That is, the earcons and prompts could be mixed and presented at the same time. Later, if users learned the association between the earcons and the recognition result, the verbal recognition feedback could be removed. In that way the user would receive feedback about what Annie recognized concurrent with the system playing a prompt that moves the dialog along. In the following example the orienting word 'Labels' is replaced with a 'Label Administration Earcon':

User:	Get my labels
System:	**~Label Administration Earcon~**
	What's next?

In Annie, earcons are used to orient users when they enter a new module or to notify users of certain system events. In order for earcons to replace verbal information in the dialog, users must learn the association between the module or system event and the earcon representing that module or system event. The more an earcon's structure depends upon its meaning, and once the dependency has been pointed out, the easier it should be to learn and remember the earcon. However, once a form is well-learned, the degree of relatedness of a form to its meaning is probably irrelevant. For example, the ring of a telephone or the shape of a

[1] The tilde (~) before and after a non-verbal sound indicates that the sound is played in parallel with the verbal information that follows.

stop sign are both very well-known and remembered, but their forms are not strongly related to their meaning.

Two experiments are described below. In the first, the Earcon Preference study, a large set of earcons was developed. Subjects chose a preferred earcon for each of nine earcon categories. Data from the first experiment was used by the experimenters to select the first five earcons to be placed in Annie. For orienting earcons, the earcons were mixed with verbal prompts that gave users ASR feedback and orienting information concerning where they were in Annie. For the system event earcons, the earcons were mixed with an informative prompt. The second experiment, the Earcon Recall Study, investigates the recall memory for five earcons placed in Annie.

2. Earcon Preference Study

A large set of earcons was developed by the experimenters and professional musical consultants. Subjects in this experiment chose a preferred earcon for each of nine earcon categories. Data from this experiment were used by the experimenters to select the first five earcons to be placed in Annie and tested in the Earcon Recall Study.

2.1 Method

2.1.1 Participants

Twenty-three subjects that began this experiment were all employees of AT&T. Eighteen subjects completed the experiment. Five subjects stated a preference for some, but not all, of the earcons. Subjects were recruited via email.

2.1.2 Earcons, Apparatus, Design and Procedure

A total of 53 earcons were tested in this experiment. Earcons were constructed by creative music professionals in consultation with the experimenters at a professional music studio. The experimenters specified the modules or events each earcon would be used to represent and general rules for earcon design (see also Sumikawa, 1985):

Calls were made to a TFLX program running on a Macintosh Powerbook 165c. Subjects were asked to select one earcon that they preferred for each of eight individual earcon categories. The eight categories, presented in the following order, were: Top Level, Goodbye, Label Administration, Employee Directory, Travel Directory, Community Directory, Error Condition, and Message Delimiter. Subject responses were given by pressing keys on their touch-tone telephone keypad.

Because the Messaging Earcons (Message Retrieval, Outbound Messaging, and New Message) were constructed as three families of earcons that sound similar within a family, subjects were asked to select one of the three sets of Messaging Earcons, rather than selecting a separate earcon for each individual category.

2.2 Results

The dependent measure in this study was the earcon preferred by subjects for each earcon category and the messaging set of earcons. In five of the nine earcon categories tested (Employee Directory, Travel Directory, Error Condition, Message Delimiter, and Messaging Set), subjects demonstrated a strong preference for one earcon or earcon set over all the others. That is, there was one choice that was preferred by 50 percent or more of the subjects. In the other four earcon categories (Top Level, Goodbye, Label Administration, and Community Directory) no strong preference was demonstrated by subjects.

3. Earcon Recall Study

At the time this experiment was run, the system had the following modules: Top Level, Label Administration, Employee Directory, Outbound Calling, and Remote Access. System events that required earcons were: Error Conditions and Goodbye (system disconnects from the user). The experimenters had concerns about how well users would be able to remember the earcons, so an effort was made to minimize the number of earcons used. The experimenters decided to use a total of five earcons for the following modules and system events: Top Level, Label Administration, Employee Directory, Error Condition, and Goodbye.

3.1 Method

3.1.1 Participants

Fourteen participants in this experiment were all employees of AT&T and users of Annie. All employees who were users were requested to complete five sets of five scenarios. Completing each set of scenarios exposed the user to all five of the earcons. Fifteen of the users of Annie completed the five sets of scenarios. Fourteen of those fifteen users were available to be subjects in this experiment.

3.1.2 Earcons

Five earcons from the set of earcons presented to subjects in the Earcon Preference Study were chosen to be used in Earcon Recall Study. The experimenters chose the earcons to be used in this recall experiment based on data in the Earcon Preference Study and expert judgment. Earcons for the categories that were strongly preferred in the Earcon Preference Study, were selected first. Subject preference data, subject comments, and the professional judgment of the experimenters were used to help choose the other three earcons to be placed in Annie for this study. Table 1 shows the earcons and the context in which they were used for the Earcon Recall Experiment.

3.1.3 Design, Apparatus and Procedure

Users were advised via email that the earcons had been added to the system and were asked to complete five sets of scenarios, each of the five sets on a different day. Each scenario was a user-system script that specified the prompts played by Annie and the utterances the users should say. By completing each set of scenarios, users were exposed to each of the five earcons in Annie service at least once. Subjects were run using a TFLX program running on a Macintosh Powerbook 165c. The five earcons put into the systerm were presented randomly and

without any associated prompts. Subjects were asked to identify each of the five earcons.

Table 1. Earcons used in Annie.

Module/ Event Name	Earcon Description	When Played
Top Level	Stylized wind chimes	With the Top Level primary prompt For example, with "Annie, here."
Goodbye	Final, resolving piano chords	With prompt before Annie disconnected the user from the service For example, with "Goodbye."
Label Administration	Sounds like several taps on a hollow bottle	With the Label Administration Top Level prompts For example, with "You have" <N> "labels." and with "Labels. What's next?".
Employee Directory	Whirlwind of violin	With the primary prompt for the employee super-directory and each of the four employee directories For example, with "Employee Directory" or "Consumer Lab"
Error Condition	Two discordant chords	With instructions to call the system service help line. For example with "Please call the help line XXX-XXX-XXXX."

3.2 Results

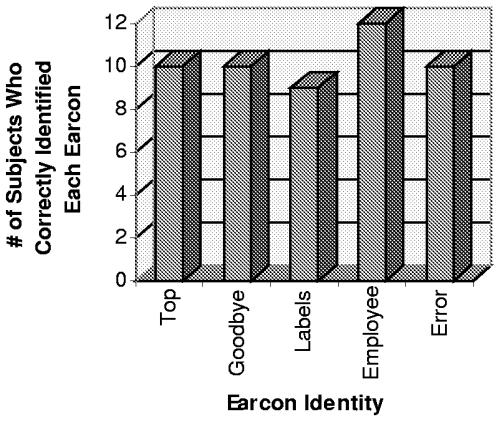
The dependent measure for earcon recall was correct earcon identification by subjects. The number of subjects who correctly identified each of the five earcons tested is shown in Figure 1. For each earcon, the null hypothesis was that users would be equally likely to identify each earcon with any of the five categories of earcons. To determine the significance of the recall of the earcons overall, a one-tailed Chi Square analysis was conducted of the observed and expected counts of all of the earcons. The analysis was significant with $p < .001$ ($X^2 = 99.5$, $df = 4$).

The Chi Square was probed with the Test for the Significance of a Proportion, or z-score. In order to compute the z-scores, the observed and expected counts above were converted to percentages. The proportion of subjects who were able to correctly recall an earcon was significant for each of the five earcons tested in this experiment ($p < .001$ in all cases).

3.3 Discussion

The main finding of this experiment was that subjects were able to remember the association between the earcons and the module or system event that each earcon was played with. In order to remove the cumbersome orienting prompts played when users enter some modules, it is necessary for subjects to remember the association between a module and the earcon representing that module. Since earcons were well-remembered in this experiment, it is likely that earcons could be used to provide feedback to users about what Annie recognized. Also, earcons could be used to alert users to system events and to orient users to changes in their location in the system.

Figure 1. Number of 14 subjects who correctly identified each earcon.



A further test of the feasibility of replacing verbal recognition feedback with earcons would be to study whether experienced users of Annie use information provided by the earcons when the verbal information is removed. If the earcons were found to be effective in cuing recognition errors to subjects, a novice and expert mode for the service would be necessary. In the novice mode, users would be presented with both verbal recognition feedback and earcons. At some point, the users would be offered the opportunity to switch to the expert mode in which the verbal recognition feedback would be removed leaving the earcons to provide recognition feedback and alert the users to system events.

References

Brewster, S. A., Wright, P. C., and Edwards, A. D. N. (1994). The design and evaluation of an auditory-enhanced scrollbar. *Proceedings of CHI '94,* Boston, 173 - 179.

Brewster, S. A., Wright, P. C., Dix, A. J., and Edwards, A. D. N. (1995). The Sonic enhancement of graphical buttons. *Proceedings of INTERACT '95*, 43 - 48.

Gaver, W. (1989). Sonic Finder: An Interface that uses auditory icons. *Human Computer Interaction*, 4, 67 - 94.

Gaver, W., Smith, R. B., and O'Shea, T. (1991). Effective sounds in complex systems. *Proceedings of CHI '91,* New Orleans, 85 - 90.

Sumikawa, D. A. (1985). *Guidelines for the integration of audio cues into computer interfaces.* Unpublished master's thesis. University of California, Davis, CA.