AUTOMATING OPERATOR-ASSISTED CALLS USING SPEECH RECOGNITION

DR. THOMAS B. SCHALK

VOICE CONTROL SYSTEMS

ABSTRACT

A speech processing system has been developed that is capable of automating a number of operator services over the telephone network. This system, called the VoiceGateway System (VG), is equipped with the following capabilities: speech recognition, speech output, speech storage, DTMF (touchtone) tone detection, remote database access, and telephone switching system communication. The specific application addressed in this paper is the automation of operator-assisted telephone calls. Operator-assisted calls are those that require telephone operator intervention to complete. These include collect calls, calls billed to a third number, credit-card calls from nontouchtone phones, and person-to-person calls. A detailed description of the way in which the VG interacts with humans to process these calls will be presented with an emphasis on the speech recognition requirements for this application.

INTRODUCTION

For over a century, the operator has acted as a primary interface between customers and telephone companies. However, with the advent of commercially viable telephone network speech recognition capabilities, many operator-assisted services can be automated in a user-friendly fashion. There is a worldwide market for speech processing systems that are capable of automatic speech recognition over the telephone network. Although DTMF-based telephone services that do not require live operator assistance are becoming more popular, approximately 35% of the telephones in the United States still do not have DTMF capabilities [1]. This problem is even more acute in Europe and elsewhere. Even for telephones that are equipped with DTMF dialing capabilities, an operator is required to process person-to-person and collect phone calls. Speech recognition over the telephone network can be used to automate both of these types of calls, even for phones without DTMF dialing capabilities.

Perhaps the most significant disadvantages of DTMF input relate to "human factors." In general, it is easier for people to speak than it is to use a telephone keypad. Given the choice, for example, people prefer saying "yes" or "no" over pressing keys with these designations. Transactions involving multiple choices are usually handled by asking the user to press digit keys that correspond to descriptive words rather than allowing input of the actual words by voice. This cumbersome technique replaces a natural mode of communication with one that is perceived as awkward, unnatural, and unfriendly. Thus, automating telephone services with machines that can not only speak and detect DTMF tones, but also understand speech, would yield a tremendous increase in speech-based applications and potential users. This paper describes a system that has been developed at VCS in which speech recognition can be used in place of live operators to process calling-card and collect phone calls.

THE VCS VOICEGATEWAY SYSTEM (VG)

A voice processing system equipped with both a voice and DTMF input interface has been developed. This system, the VG, is also capable of remote database access and telephone switching system communication. The VG queries callers for information required to institute transactions, then retrieves and relays information to the caller using digitized voice response. The caller may interact with the VG by using either DTMF or speech input. The VG is also capable of speech storage, and it can provide database access using many standard data communication protocols. The heart of the VG is implemented on a Micro Channel card shown in block-diagram form in figure 1.

In addition to operator-assisted calls, application areas for the VG include telemarketing, audiotext, call screening, remote information access, financial transactions, and telephone polling. The speech input, output, and storage capability of the VG permits complex caller question and answer sessions that previously could have been handled only by a human operator. The ability to handle multiple calls on a single machine reduces equipment and personnel costs of manned positions. User-friendly application construction software permits quick and efficient customization to a customer's unique needs. The variety and complexity of interfaces supported by the

EUROSPeECH '89, Paris, France, September 1989
VG permits it to be used as a flexible link between callers from any type of telephone to any computer or database.

The recognition technology utilized by the VG has the following characteristics:
- * speaker-independent
- * high accuracy
- * 60-word speech recognition vocabulary
- * discrete and connected speech input
- * 50-millisecond response time

The most important vocabulary words required for telephone applications are: "yes," "no," the digits "zero" through "nine," and "oh." In addition to these words, approximately 50 carefully selected vocabulary words are available in order to automate a wide variety of telephone services. The alphabet is currently under development.

The VG recognition software completes the recognition decision and is ready for additional speech input within 50 milliseconds after each input word is captured by the recognizer. This allows for rapid data entry since only a brief pause is required between words. For digit entry, a special recognition technique is employed that allows the caller to speak digits in a connected fashion, i.e., with no perceptual pause between successive digits.

OPERATOR-ASSISTED CALL AUTOMATION

Operator-assisted calls are those which require telephone operator intervention to complete, including collect calls, bill to third-number calls, credit-card calls from nontouchtone phones, and person-to-person calls. The automation of operator-assisted calls is accomplished through interaction of the VG with telephone switching systems. Figure 2 illustrates the VG general configuration for automating operator-assisted calls. The originating party's call is routed through the local CO to the long distance network switch. After a call is received by the switch, a voice path is established between the calling party and the VG. The VG has access to proprietary verification data bases for processing the various operator-assisted calls.

Figure 3 represents a simplified protocol for automating operator-assisted calls using voice recognition. For example, the VG would handle a credit-card call in the following manner:

1. The caller dials a 0+ number to access a particular switching network that routes the call to the VG.
2. After detecting the call, the VG responds with an introductory message and then asks the caller to say one of the following words: "credit-card," "calling-card," "person-to-person," "collect," "third-party," or "operator."
3. The caller responds by saying "credit-card" and the VG uses voice response to indicate which word was recognized.
4. The VG then prompts the caller to say one of the following words: "American Express," "MasterCard," "Visa," "Optima," "Discover," or "Operator."
5. The caller responds by saying "American Express" and the VG uses voice response to indicate which word was recognized.
6. The VG then asks the caller for the first group of digits of the credit-card number.
7. The caller responds by saying "4-3-2-9" and the VG uses voice response to indicate the digits that were recognized.
8. The VG then asks the caller for the next group of digits, and so forth, until all digit information is obtained.
9. The VG accesses a remote credit verification database to validate the credit-card number. If credit is confirmed, the call is completed and billed to the credit-card. If credit is denied, the caller is informed and transferred to an operator.

Calls to be charged to calling cards are handled in a similar fashion. The caller responds with "calling-card" during step 2 and is then prompted for the calling-card type to which the call is to be charged. As in the credit-card example, a digit transaction occurs, validation occurs and the call is extended.

Third-party calls are handled by responding with "third-party" during step 2, and then the VG prompts for the phone number to which the call is to be billed. After validation, the call is extended. For collect and person-to-person calls, the VG records speech spoken by the caller and plays it to the destination party for identification purposes. For example, if the caller responds with the word "collect" during step (2) above, then the caller is prompted with "Would you like to make a collect call?" After responding with "yes," the prompt "Please state your name" occurs and the caller's name ("John Doe") is captured and recorded digitally. After the caller's name is captured, the call is extended. If the other party answers, then that person is prompted with "Will you accept a call from 'John Doe'? Please say 'yes' or 'no.'" If "yes" is recognized, the call is completed. Otherwise, a live operator is connected.

SPEECH RECOGNITION REQUIREMENTS

Speaker-independence is a critical requirement for automating operator-assisted services. The recognition...
Factors that affect performance, such as telephone line conditions, vocabulary size and confusability, and user experience level.

For credit-card, calling-card, and third-party calls, the most critical transactions involve digit entry. Correct entry or long digit strings require high individual digit recognition accuracy as well as convenient methods to correct recognition errors if they occur.

During digit entry, a special recognition technique is used that allows users to speak digits in a connected fashion, with virtually no perceptual pause between successive digits. Beeps occur as the digits are captured by the recognizer to provide aural feedback to the user. After each digit group, voice response is used to indicate the recognized digit sequence. The prompt "Correct?" is used for verification.

The VG achieves average substitution error rates of less than 1% for individual digits when entered in a discrete or "connected" fashion. A rejection error is said to occur when a word in the active vocabulary is spoken correctly and the recognizer fails to classify the utterance as a word in the vocabulary. When a rejection occurs, the user is immediately requested to repeat the word. For digits, this occurs less than 2% of the time.

One might argue that this application requires a "continuous" digit recognition capability (where digits can be run together). The main problem with using a true continuous capability, if it existed, is that users could not get interactive feedback during digit entry. For example, if a digit is rejected due to classification uncertainty or a low signal level, immediate aural feedback, such as "Please Repeat" or "Speak Louder," can be provided on the fly. Furthermore, the vocabulary requirements for true continuous digit recognition can go beyond recognizing the eleven digits (zero through nine, and oh). Words like "hundred," "thousand," "fifty," "fifteen" and so forth are often used during continuous digit entry.

For collect and person-to-person calls, the destination party should be made aware that "yes" and "no" are the active vocabulary words when answering the prompt questions discussed earlier. For example, when attempting to complete a collect call, an appropriate prompt might be, "Would you accept a collect call from [John Doe]? Please say 'yes' or 'no.'" Clearly, instructions are needed for those not familiar with the system. In spite of proper instructions, users will still respond with words (or sounds) that are outside of the "yes"/"no" vocabulary. Utterances such as "O.K." "Yeah," "What?" and "Am I talking to a computer?" [2] need to be anticipated. Therefore, for this transaction, it is important that the recognizer exhibit good spurious immunity while maintaining a low rejection error rate. The "yes"/"no" recognition technology is designed to recognize "Yes," "O.K.," "Yeah" and "No" and reject utterances and sounds that do not "rhyme" with one of these words. Ideally, all spurious sounds that do not correlate well with any of the active vocabularies should be rejected.

**FUTURE APPLICATIONS**

According to the Probe Research Report on Voice Recognition [3], U.S. telephone companies employ 80,000 or more operators to perform toll, assistance, directory, and intercept services. Automating these services with speech recognition can have a tremendous impact on the cost of providing such services. Future applications using speech recognition over the telephone network will benefit from vocabulary expansion with words including the alphabet. The key to success in automating telephone services is the development of user-friendly systems that operate as efficiently as live operators.

Improvements and extensions of VCS telephone network recognition technology are currently being pursued along three major directions: (1) recognition accuracy improvement through fundamental algorithm research and development, (2) recognition response time reduction, and (3) vocabulary development and data collection for foreign telephone networks. VCS takes advantage of the flexibility of its software-based technology by keeping the basic recognition algorithms under constant analysis and revision. Several significant algorithm improvements within the last year alone have resulted in major reductions in error rates. The much lower percentage of DTMF-capable phones used in foreign telephone networks makes speech recognition an even more attractive option in these markets. VCS has already developed telephone speech recognition technology for the British network and plans to develop technology for other foreign networks in Europe and Asia.

**References:**


CN CAN BE CHANGED, VIA PLUG-IN MODULES TO OTHER INTERFACES, E.G. 4-WIRE E&M, GROUND START

Figure 1 VCS VoiceGateway™ Voice Processing Card Overview

Figure 2 VCS VoiceGateway™ Configuration Overview

Figure 3 General Logical Flow for Automation of Operator Assisted Calls