

Lawrence A. Malakhoff, Martin V. Appel¹
U.S. Bureau of the Census, Washington, D.C. 20233

Key Words: speech recognition, voice recognition, voice recognition entry, global positioning system

1.0 Background

"Area segment listing" is a field operation performed for the Decennial Census and continuing surveys². According to the Field Representative's (FR) manual, an area segment is a land area with well-defined boundaries, such as streets, rivers, mountain ridge lines, and railroad tracks. Listing is the recording of address information, either on paper or electronically. Area segments for most surveys are defined in rural areas where a large percentage of housing units have incomplete postal addresses. Area segments may also be defined for some urban areas, where complete addresses are not available, such as new residential construction.

The area segment listing application is comprised of two main operations. First, the area is pre-canvassed to estimate the number of housing units within the segment. Then FRs canvass the area and list all residential units. Group quarters (GQ) and regular housing units are listed separately.

Each month at the Census Bureau, about 200 FRs perform area segment listing operations. These FRs record addresses on about 2400 paper listing forms. These forms are mailed to the FRs assigned Regional Office (RO) where these forms are checked for errors and completeness. Following the ROs edit, copies of the forms are mailed to the Census Bureau's processing office for keying. The keyed data is transmitted to Bureau headquarters for updating our address files. Since this is a continuous operation, no additional processing is done for the Decennial Census. The result is an improved address list.

The cost to mail and key the information is almost \$5 per listing sheet (\$12,000/month, 600 segments/month) (Newman 1997). Overhead costs for printing, verifying, editing, and copying the listing forms are not available.

2.0 Motivation

Speech recognition technology now permits a user to enter text and data directly into a computer by voice. This technology already has applications for data entry in the medical and legal professions (Padilla 1997). Affordable software is now available that lets a user dictate numbers and text into MS-Windows applications. Has this technology evolved to the point where it is viable for use by Census Bureau FRs? Could these listing operations benefit from speech recognition technology? Could an FR drive around a neighborhood and, using speech only,

accurately enter street names and house numbers into a notebook PC? Sometimes walking is necessary for listing in urban areas. FRs have complained about the weight and bulkiness of PCs in this environment. Could an FR use Voice Recognition Entry (VRE) and a notebook PC while walking? Are there viable alternatives to a conventional notebook PC?

3.0 Research Plan

The Voice Recognition Field Listing project will test the feasibility of integrating speech recognition technology and listing procedures. The project has six phases: (1) Selection of Voice Recognition Entry (VRE) software; (2) Selection of the hardware; (3) Determination of the vocabulary and dialogue for a prototype; (4) Development of a prototype; (5) A small-scale usability study; (6) A field test.

4.0 Selection of VRE Software

Based on a literature review, Dragon System's DragonDictate dictation software was selected. It has a 60,000 word vocabulary for dictation. The software monitors and improves its recognition rate by using statistics about which words the speaker most frequently uses. Currently, the software uses discrete speech for dictation, which means a user must pause briefly between each word. Soon, 1 to 2 years, developer kits will be available for building applications with the more natural, continuous speech. The dictation software³ is compatible with the Visual Basic programming language which allows development of windows applications driven by voice.

5.0 Selection of Hardware

FRs currently use notebook PCs for Computer-Assisted personnel interviews and other field work. Because FRs have complained about the weight and bulkiness of notebook PCs, an alternative is being investigated to address this problem - a mobile computer. The mobile computer is worn on a belt, rather than carried by hand.

Feasibility tests for the prototype will be performed on two types of computers: a notebook PC⁴ and a mobile computer. Researchers recently witnessed a demonstration of a mobile computer⁵ at the Census Bureau. The computer consists of three components: the processing unit (2 pounds); battery pack (1.5 pounds) (these items are separately enclosed in waterproof plastic housings which are attached to a belt around the waist); a headset equipped with a microphone, earphone and a small head-mounted VGA display (15 ounces) connects to the processing unit.

Initial responses to this device indicated that a head mounted display was not preferred. Instead, a purely audio interface consisting of earphones and a microphone was proposed.

6.0 Dialogue and Vocabulary Development

The first step in the construction of the prototype was to obtain the FR's procedures manual, and a sample of completed area segment and GQ listing forms (Bureau of the Census 1996). These materials were used to identify words and phrases meaningful to the listing operation.

The dialogue for the prototype was created by reenacting the completion of the example set of responses for each field. A specific vocabulary was designed for a particular field. If an open-ended response is valid for a field, e.g. the "other" category, then that field uses the system's 60,000 word dictation vocabulary (*DragonDictate User's Guide* 1996).

Table 1 describes the vocabularies used in this prototype. Words in CAPS are voice commands defined for this application; words in *italics* are a subset of the voice commands provided with the dictation software. The system's vocabulary allows the user to switch between dictating and spelling in dictation fields. We hope spelling will not be necessary because the current state of speech recognition technology cannot differentiate between spoken letters with high confidence; a user must speak the code word corresponding to the letter. It is anticipated that in future releases of VRE software 'A' will stand for 'A', rather than *alpha* representing 'A'.

The vocabulary for the street name field would be prohibitively large if it contained all United State street names. We will constrain this vocabulary by using a global positioning system to acquire the local latitude and longitude coordinates which will be used to search Census Bureau geography files to retrieve a list of street names for the specific "NEIGHBORHOOD."

7.0 Prototype Development

The prototype is being designed to be used eyes and hands free. However, the development software does also generate a graphical user interface for input. Even though the screen is not being used for input, it might be used for visual cues as to which voice commands are active while walking or when the car is stopped.

Figure 1 shows the screen when the prototype is first started. The prototype is designed to begin area segment listing at the address line. An address line consists of the following fields: NEIGHBORHOOD-STREET; HOUSE NUMBER; LOCATION; YEAR BUILT; OCCUPANT and EXTRA.

There is one exception to the "eyes-free" convention in this prototype. The FR will look at the NEIGHBORHOOD text window and find the index number corresponding to the street name to be listed and

speak that number into the STREET dictation box. The street name will appear in the STREET NAME dictation box. If the street name is not in the neighborhood list, it can be dictated into the STREET NAME box.

As fields are completed, the FR will be automatically prompted for the next field. If the FR needs to enter the information in a different order from that of the listing form, it can be done by saying the name of the field. Fields using dictation are not echoed back at entry time because correcting recognition errors can be achieved faster with a visual interface and keyboard entry after the car has stopped. If the area segment information is being updated, then the street name, house number, and apartment number, if applicable, are spoken TO the FR by a speech synthesizer. Fields needing updates are accessed by speaking their name and then the update information.

Each element of the address in the segment listing will have start and stop times. When the data gathering is done for the day, the computer's modem will be connected to a telephone line, and the FR will say "transmit" to send the data to Census HQ.

8.0 Usability Study

The Center for Survey Methods Research (CSMR) at the Census Bureau will conduct research on the prototype with test subjects recruited from HQ personnel. Their evaluation will cover four phases: a test of procedures for voice training; a test on the functionality of the prototype for two kinds of hardware; gathering information on attitudes towards using voice recognition; and timing and cost studies comparing paper and electronic versions of the listing forms.

8.1 Voice Training

Training of the speech recognizer consists of saying a set word up to three times to form a voice pattern. Training software exists that will prompt the FR for words. Not all vocabulary words are trained because the software learns how other words are pronounced based on the string of phonemes generated within each syllable. Phonemes are elementary components of speech; they make up syllables which make up words. The process the computer goes through while it is training is called adaption. Adaption also occurs during dictation, so performance improves with each use. Training lasts 20-30 minutes.

A CSMR analyst will be observing test subjects to determine if the training procedures are effective (Malakhoff et al. 1995).

One limitation of voice recognition technology is that not all voices can be understood. The software selected for this project uses recognition models based on American English. Persons with speech impediments or having a heavy foreign accent will not be understood with high confidence. Research needs to be performed to determine how much of a problem this causes.

8.2 Prototype Functionality

Listing goes on continuously. One to three years after an area is listed, it is re-listed. In this process about 65% of the time is spent on new construction, 20% on GQ listing, and 15% for updating area and GQ segments due to conversions, omitted units, etc. An area segment listing averages about 60 addresses. An FR will need to be able to perform all the same listing tasks with the electronic form as they could with the paper form. Test subjects selected from HQ personnel familiar with the listing operation should be able to simulate the actions real FRs perform. Test subjects will list a hypothetical rural or urban area segment. Inside and outside the CSMR laboratory we will time each of these processes to obtain estimates of data entry tasks. These tests will compare notebook and mobile computers for portability and entry tasks.

8.3 Evaluation of attitudes about using voice recognition software

Upon completion of entering data for an area segment, and GQ listing, the test subject will be asked a series of questions about their experiences with using the prototype (Jenkins 1995). These questions will obtain information about the use of the mobile and notebook computers for listing while walking and driving, and determine the importance of audio feedback and prompting for data fields.

8.4 Timing and Cost Studies

Processing costs need to be compared between the use of electronic and paper GQ and area segment listing forms. Costs for entering address listing lines, updating procedures, and editing can be estimated. The error rates for both GQ and area segment listing forms will be computed and included as part of the costs.

9.0 Field Test

If the prototype successfully completes the CSMR evaluation, we hope to perform a small field test. What we propose is that the FR will list the same area three times; once by paper and pencil, second, by notebook PC, and third, by mobile computer. These listing trials will be used to time the data entry tasks for each approach. The listing segments selected should have the same mix of GQ and regular housing units as a typical area segment.

We also want to measure the FRs attitudes towards using VRE for area segment listing. There should also be questions on the vocabulary and dialogue used to control the application. Are they meaningful? Easy to remember? What words might work better? Do the FRs have a preference between PC and mobile computers?

10.0 Conclusions

The purpose of this research is to test the feasibility

of speech recognition technology for listing applications. The choice of mobile computer or notebook PC for listing will be based primarily on how well the FR does carrying out basic listing and updating tasks. Will the ease of carrying the mobile computer make up for the fact that it lacks a visual interface to control the application?

Basic listing costs could be reduced and the process speeded up if a computer equipped with voice recognition entry and modem is used. Savings should accrue in printing, copying, keying, postage and their associated overheads. Additional savings could be achieved if FRs could edit and correct their entries in the field to eliminate editing procedures in the ROs. But will these reductions offset the increases in hardware and software costs?

11.0 References

- Bureau of the Census, *Listing and Coverage Manual for Field Representatives*, issued 12/96, see esp. Forms 11-1 and 11-5.
- DragonDictate User's Guide*, 1996
- Jenkins, C.R., Appel, M.V. (1995). "Respondents Attitudes Towards a U.S. Census Voice-Recognition Questionnaire." Paper presented at the International Field Directors and Technologies Conference
- Malakhoff, L.A., Appel, M.V., Cole, R. (1995). "Report on Spoken Language Recognition Year 2000 Census Questionnaire" *Proceedings of the American Statistical Association Joint Statistical Meetings*.
- Newman, T.; Personal Conversation, June 1997.
- Padilla, E. (1997) "Voice Article", *Senior Technician on Voice Recognition*, Volume III, January 1997, http://www.voicerecognition.com/voice_article.html

NOTES

1. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau.

2. Current Population Survey (CPS); National Crime Victimization Survey (NCVS); Survey of Income and Program Participation (SIPP); Consumer Expenditure Diary Survey (CED); Consumer Expenditure Quarterly Interview Survey (CEQ); American Housing Survey-Metropolitan Sample (AHS-MS); and National Health Interview Survey.

3. DragonDictate 2.52 dictation software; Dragon Xtools 1.0 Development Kit; <http://www.dragonsys.com/>; Visual Basic 4.0; Retki OCX Development Kit for global positioning systems.

4. Toshiba 420CDS, 100 MHZ Pentium processor; 16 bit sound chip, internal speaker, headset; Toshiba Noteworthy global positioning system pc card (manufactured by Retki), antenna.

5. Mobile Assistant II, <http://www.xybernaut.com/>, 133 MHz Pentium processor; battery pack; 16 bit sound chip, earphones, microphone; detachable keyboard;

monitor cable; belt for attaching the processing unit and battery pack to the users waist; Toshiba Noteworthy global positioning system pc card, antenna.

The screenshot shows a dark-themed mobile application interface with the following elements:

- Neighborhood:** A large rectangular text input field labeled "Text1".
- HOUSE NUMBER:** A text input field labeled "Text2".
- LOCATION:** A text input field labeled "Text3".
- Occupant:** A text input field labeled "Text4".
- Number:** A small text input field labeled "Text5".
- Extra:** A text input field labeled "Text7".
- Streename:** A text input field labeled "Text6".
- REMARKS:** A button located at the bottom left.
- TRANSMIT:** A button located below the REMARKS button.
- LAST LINE:** A button located at the bottom left, below TRANSMIT.
- YEAR BUILT:** A button located in the center.
- NEXT LINE:** A button located in the center, below YEAR BUILT.
- LOGOUT:** A button located at the bottom right.
- FOOTNOTES:** A button located below the LOGOUT button.
- HELP:** A button located at the bottom right, below FOOTNOTES.

AREA SEGMENT LISTING CONTROL SCREEN

FIGURE 1

Table 1
Prototype Vocabularies

Vocabulary Name	Contents	Function
neighborhood	nearby street names	verification of street names
numeric	1-9, oh, zero	address numbers
spelling	26 spelling words	spelling
dictation	60k words <i>SpellMode</i>	general dictation switches to spelling mode
remarks	DEMOLISHED OUTSIDE segment boundary conversion to COMMERCIAL use conversion to STORAGE MERGED added UNIT LISTED incorrectly NEW merged unit VACANT MOBILE HOME converted to GROUP QUARTERS	enters phrase or word into remarks field " " " " " " " " " " Starts GQ entry
year built	BEFORE Census Day AFTER Census Day UNKNOWN	sets year built flag " "
GQ Footnotes	verified as MILITARY base is CLOSED - list converted to INSTITUTIONAL group quarters type CHANGE complete RELISTING OTHER	enters a footnote " " " " starts dictation
Footnotes	NAME PHONE MANAGER UNINHABITED all GROUP QUARTERS housing OTHER	dictation of mobile home park name phone number dictation of manager's name enters a footnote " starts dictation
Control	STREET NAME HOUSE NUMBER LOCATION OCCUPANT EXTRA unit ADDUNIT NEXT LINE LAST LINE TRANSMIT LOGOUT OK CANCEL HELP	opens streets vocabulary opens numeric vocabulary opens dictation for apt #, etc. Opens dictation vocabulary " opens a new address line after the last address line save data, move to next listing line save data, move to previous listing line transmit data via modem save all data and log out saves data for the current field quits current field, returns to previous field announces current vocabulary
System	<i>GoToSleep</i> <i>WakeUp</i> <i>CommandMode</i> <i>SpellMode</i>	Pause the application Restart the application switch to <i>CommandMode</i> from <i>DictateMode</i> switch to <i>SpellMode</i> from <i>DictateMode</i>
RecordType	living QUARTERS PERSONS	sets record type flag "