SPEECH DATA ENTRY: RESULTS OF THE FIRST TEST OF VOICE RECOGNITION FOR DATA COLLECTION

Debbie L.S. Winter and Richard L. Clayton, U.S. Bureau of Labor Statistics
Richard L. Clayton, U.S. Bureau of Labor Statistics
441 G Street N.W., Washington D.C. 20212

Key Words: collection methodology, touchtone data entry, voice recognition, CASI

I. Introduction:

Over the past year, the Bureau of Labor Statistics has field tested a Voice Recognition system. This paper documents the results of this test. The system, which recognizes the human voice as input to a Computer Assisted Self-Interviewing system (CASI), has proven to be a viable alternative to traditional data collection methods, such as mail. The Voice Recognition system is currently in use at BLS to collect economic data in a monthly business survey.

II. Background:

The Current Employment Statistics (CES) program provides one of the earliest indicators of the Nation's current economic activity. The primary products of the CES are monthly estimates of total employment, women workers, and production worker employment, hours worked, and earnings. The CES is a business establishment survey covering all nonagricultural industries. The largest federal monthly survey, the CES is based on a sample of over 340,000 voluntary reporting establishments collected by cooperating State agencies

Timeliness is of the utmost importance in our data collection. The CES requests information for the reference week including the 12th of the month. Results from the CES are published the first Friday of each month for the preceding reference month. This means that there are generally about two weeks to collect, key enter, edit, and tabulate data for the release of preliminary National estimates.

Traditionally, the survey has been conducted by mail, whereby a "shuttle" questionnaire is mailed to the respondent each month. The respondent completes the questionnaire with his/her payroll information and returns the questionnaire through the mail. Mail collection typically yields only about a 50 percent response rate by the early deadline for preliminary estimates.

In order to improve the response rates for first publication, BLS has been testing telephone data collection methods for several years in the CES. These include computer assisted telephone interviewing (CATI) collection since 1986,[1] touchtone data entry (TDE) since 1987,[2] and now voice recognition (VR) since 1989.[3] The testing of CATI methods yielded improvements of up to 30 percentage points in the response rates for the preliminary publication deadline. However, implementing ongoing CATI over such a large sample survey would be too expensive, therefore, research then focused on TDE. TDE's leading advantages are much lower costs, and operational ease. The disadvantage of TDE is that only 75-85 percent of our respondents have touchtone telephones. Thus, the need to provide inexpensive, convenient self-reporting by all respondents spurred BLS to investigate voice recognition technology.

Several characteristics of the CES survey make automated data collection a feasible alternative. The CES is a monthly longitudinal survey, where the same respondents participate in the survey each month. Few data items are collected from respondents--the number of all employees, women workers, and production or nonsupervisory workers, as well as the payroll and hours for the production or nonsupervisory workers. Most respondents obtain the information directly from their payroll records, a hard data source. Because the CES is a survey of businesses, respondent contact is relatively simple; regular business hours provide a well-defined contact period. Respondents are usually payroll clerks or heads of payroll departments, so they are usually familiar with the information requested on the questionnaire.

III. CASI Methods--TDE and VR:

Touchtone Data Entry and Voice Recognition can both be categorized as Computer Assisted Self Interviewing methods, also known as CASI.[4] Both of the CASI systems, TDE and VR, in use at BLS are microcomputer-based. There are several features common to both of these CASI applications. In both self-response environments, respondents initiate the call to report the data at their own convenience. A unique identification number controls respondent access to the system, and provides a questionnaire specific to the respondent's industry. Both systems provide verification of each data item for the respondent so that after a respondent enters data, either by touchtone or voice, the system repeats the entry for verification. Management information, such as date and time of call, and call length, is stored as a by-product of the system. The primary benefits of TDE and Voice systems are their convenience and ease of use for respondents and the potential cost savings they offer over mail data collection.

The CES TDE system is similar in design to the touchtone applications currently proliferating in the banking industry and for telephone call routing. Respondents enter numeric data from their completed questionnaires using the number keys on their touchtone phones. Not all respondents have access to touchtone phones, however, even in business environments. The availability of touchtone service and touchtone phones varies widely across the states participating in our tests. For example, of the businesses in Maine and Vermont, about 55 percent have touchtone phones. In California and Alabama, over 90 percent of the businesses had touchtone phones. Voice Recognition addresses the collection needs of respondents without touchtone phones. Rather than key-entering information, respondents enter data by simply speaking strings of digits, such as "one two three one five" to enter "12315" as a response. The CES Voice system recognizes the digits "0" (both "zero" and "oh") through "9", and the words "yes" and "no". While limited, this vocabulary is well suited for the collection and verification of data from the CES questionnaire.

Voice Processing Corporation of Cambridge, Massachusetts developed the prototype Voice Recognition system for the CES test. Their system features speaker-independent recognition, meaning that the system does not need to be "trained" to recognize a particular voice. The system also recognizes continuous speech, so that a delimiter does not have to be entered by the user at the end of each spoken digit. Continuous speech recognition makes data entry more "conversational." Both speaker-independence and continuous speech are important features for a system to be used across a broad spectrum of users.

IV. Collection Procedures:

CASI procedures in the CES are fairly simple. CES respondents already have a questionnaire and other survey information in a special survey folder. When first converting to CASI collection, they receive a respondent instruction package which includes a toll-free number and an instruction sheet. They also receive a practice identification number that allows them to access and try out the system *before* they report live data.

After the respondent's first use of the system, a follow-up interview is conducted. The purpose of this interview is to find out whether the respondent encountered any problems with the system and to elicit any comments on the system or CASI procedures.

During the subsequent months, contact with respondents is quite limited. Respondents receive an "Advance Notice" postcard in the mail about the time that their data are usually available and well in advance of the CES collection deadline. If respondents have not reported as the earliest survey processing deadline approaches, survey staff place short non-response prompting phone calls reminding the respondents to report.

V. Research Issues:

Several research goals were defined in order to assess the feasibility of Voice

collection. The response rates under Voice would be compared to the very high response rates established under CATI and TDE collection to evaluate the fundamental effectiveness of Voice Recognition as a data collection technique.

Two major areas of concern that would also effect response rates were recognition and respondent acceptance. In the recognition area, the system would have to be able to recognize the speech patterns of a wide range of respondents. Respondent acceptance was also a matter of interest--how well would respondents accept "conversing" with a computer system? Records of problems encountered and solutions would provide rough measures of recognition capabilities and respondent acceptance.

Data quality was also an area of concern. Recognition errors and respondent data entry errors might introduce an additional source of nonsampling error called "mode error." Mode error describes nonsampling error that is introduced by the mode of data collection. Mode error would exist if the respondent data entry errors exceed existing key entry error rates, or if recognition errors went uncorrected.

VI. Response Rates:

Voice maintains the high response rates experienced under CATI and TDE. Table 1 below shows the average first publication response rates for the period from August 1989 to June 1990. Voice Recognition has consistently exceeded an 85 percent response rate over the period of the test. This would indicate that the response rates are not due to a novelty effect. TDE, similar to Voice in many ways as a CASI technique, has sustained high response rates over a three year time period.

The response rates shown in Table 1 suggest that those for Voice Recognition exceed those of CATI and TDE, however, the difference in average response rates is may be due to the small size of the voice sample and the different compositions of the voice, TDE and CATI samples by size of business establishment, and previous reporting behavior.

Table 1. Average First Publication Response Rates, August 1989 through July 1990

Collection Method	Response Rate
Voice	92 %
Touchtone	83 %
CATI	85 %
Mail	51 %

Most Voice respondents had previously reported their data using TDE, except for a group that were converted from CATI collection. Use of both self-response systems allowed respondents to compare their features. (Former CATI respondents who had not used touchtone were excluded from the comparative analysis.)

VII. Respondent Acceptance:

High response rates indicate a high level of respondent acceptance. Summaries of the first month follow-up interviews reinforced the evidence offered by high response rates. Respondents using Voice for the first time were asked to compare voice reporting to Touchtone reporting. About 60 percent of respondents preferred Voice to Touchtone, with most of the 60 percent citing that Voice is easier to use. About 32 percent preferred TDE, 10 percent had no preference. Many new Voice respondents felt the interview had a "more natural" flow, since they did not have to press keys on their phones to complete the interview. One respondent commented that pressing keys under touchtone was "work". Respondents who preferred TDE were likely to have experienced some sort of recognition or procedural problem during their first use of the Voice system.

Interestingly, most respondents reported that Voice required less time to report than the touchtone system. In fact, Voice calls are about 20 seconds longer than the average two minute touchtone interview, because of lengthier instructions and prompts.

VIII. Problems Encountered:

During the first month, 85 percent of the respondents experienced no difficulty using the Voice system. The remainder

experienced mostly "procedural" problems, such as calling the touchtone system by mistake, losing the toll-free voice system number, or reporting live data while using the practice identification number. About half of the respondents used the practice identification number to try the system and found it to be useful. The other half did not seem to think it was necessary.

During ongoing collection (after the first month on Voice), 98 percent of all interviews were successfully completed. Only 2 percent of the interviews required some type of assistance from survey staff. In contrast to the problems encountered during the first month of Voice response, most of the ongoing problems were due to persistent recognition problems. During the 12 months of the test, 3 respondents out of 120 (2.5 percent) were returned to touchtone collection due to these persistent recognition problems with the prototype system. (Early tests of the production version of the system suggest that this small recognition problem is somewhat reduced).

IX. Data Quality:

The testing of any new collection methodology carries the potential for introducing new sources of error. With voice, the greatest concerns were the accuracy of the recognition, and respondent reporting behavior. Data quality is concerned with system's ability to accurately recognize respondents' data entry.

To measure the impact of Voice collection on data quality, Voice respondents were asked to return copies of the CES questionnaires they had maintained during the 1989 calendar year. The completed questionnaires were then compared to the data entered into the voice collection system. For the purpose of this study, the data on the questionnaires were assumed to be correct. The analysis is somewhat limited by the small number of observations: 47 respondents over a 6 month time period, June through December, 1989.

Each record of data entered into the voice system contains a number of input data items: the unique identification number, month, all employees, women workers, production/nonsupervisory workers, payroll, hours, and overtime or commissions. Each input data item, except the identification number, was treated as one observation. The identification number was not treated as an observation because correct entry of this item is required before the interview proceeds any further; by definition, it must be correct for all input records. Each incorrect data item counted as one error.

The results of the comparison yielded a very small overall incidence of error. Out of a total of 1164 observations, 27 observations contained an error, for an overall error rate of 2.3 percent. These errors can be broken into two categories: data entry or recognition errors and procedural errors. The cause and treatments for each type is different.

A. Data Entry or Recognition Error: The analysis originally focused on data entry or recognition errors since it was expected that these type of errors would be the most prevalent in such a new technology. However, only 7 errors met this description, at a rate of 0.6 percent.

Table 2. Incidence and Error Rates by Type of Error in Voice Data Collected: June-December 1989.

Error type	Incidence	Error Rate
Data entry or recognition	7	0.6 %
Procedural	20	1.7 %
Total errors	27	2.3 %

This small error rate is comparable to data entry error rates found in other studies.[5] With self-response data collection, such as voice and TDE, each data item is repeated to the respondent for verification, providing, in effect, 100 percent verification. To the extent that this error rate may have exceeded data entry error rates under other collection methods, mode error would exist. Since the rate is quite small, we have not found evidence of mode error related to respondent data entry errors or recognition errors.

Comparisons of the entered data to that on the questionnaire were inconclusive as to whether the error was truly a data entry error or a recognition error. Recognition errors

might be characterized by a predominance of errors with a particular digit, or combination of digits. Too few errors were observed to determine whether one number was more likely to be in error than others.

B. Procedure Error: Errors were also discovered that seem to be related to specification errors to the respondent. These "procedural" errors include incidences of respondents reporting "zero" for a data item when a "blank" (no answer, no data available) is the correct entry, an important distinction for the estimation process. Procedural errors also include instances when legitimate zero values were left blank, or "not reported". The vast majority of these errors (17) were committed by one respondent over several months. These types of errors are usually flagged by post-processing system edits for review. In this case, they were corrected prior to use in estimation; however, we will modify respondent instructions to help prevent these errors before they occur.

X. Solutions:

Since most first month respondents report without incident, and problems in subsequent months are negligible, the respondent materials providing instructions seem to work well. Among those respondents who experienced some type of problem during the first month of Voice reporting, the solution seemed to lie with the respondents themselves. These respondents took initiative to solve their own problems with the system before calling survey staff for assistance. About half of them used the practice identification number prior to live use of the system.

When encountering a recognition problem, respondents "trained" themselves; that is, they changed their speech patterns and pace of speech, adapting to the recognition system. The system allows three "tries" to correctly recognize voice phrases for a particular input item. After three incorrectly recognized entries, the system refers respondents to a "help" number. Respondents often would call the system to try again rather than calling survey staff for assistance, and would often succeed on the next report.

For persistent problems, respondents often take the initiative to call survey staff. Respondent persistence in solving their own recognition problems seems to indicate not only a high degree of commitment to survey participation, but also an acceptance of Voice as a reporting method.

XI. Conclusions:

Response rates. Voice collection maintains the high response rates attained under other automated collection methods (CATI and TDE) and has consistently exceeded the 85 percent target rate during each month of testing.

Respondent Acceptance. Respondents indicate that they like self-response methods, and there are indications of preference for Voice reporting over TDE reporting, particularly when the first exposure to Voice response is free of difficulties.

Problems Encountered. Most of the problems were experienced by first-time callers, and most of these problems were procedural. These first-time procedural difficulties can be reduced with better instructions. Few data entry/recognition errors occurred; the incidence rate is comparable to other studies of data entry error rates, suggesting that voice data entry does not contribute to overall error.

Solutions. Overall, respondents like self-response and will take initiative to learn how to use the system effectively. When a problem occurs that they cannot solve, respondents will seek assistance from survey staff, particularly if the problem prevents them from reporting their data.

XII. Further Research:

Additional research of CES data collection by Voice Recognition is planned to augment this study. The study of data collection method effects on data quality will be expanded, and voice and TDE error profiles will be compared. Data entry error rates will be monitored to ensure that they do not exceed those experienced under other data

collection methods currently in use in the CES. Improved respondent instructions will be developed to reduce procedural errors. As more CASI test data becomes available, long term response rates, attrition rates, and respondent attitudes will be studied, particularly to compare the Voice results to those obtained under TDE collection.

Because of the ease of use cited by first time respondents, CASI methods are likely tools for: 1) short repetitive surveys, 2) one time, quick turnaround surveys, and 3) mixed mode surveys, using CASI with mail, or CATI/CAPI collection. Testing CASI applications now is a solid investment in future survey data collection methods, as their cost-effectiveness and acceptance by respondents is likely to increase with time. Potential cost savings offered by automation and self-response become more feasible as the cost of microcomputers and hardware declines and labor-intensive costs such as postage increase.[6] Also, with an increasingly computer-literate business environment, user acceptance is likely to grow. Internationally, commercial Voice Recognition is already operating in multi-lingual applications. Unlike the U.S., the availability of touchtone service is very limited in Europe, providing further impetus for VR methods.

Voice recognition research is advancing rapidly. Speaker-independent vocabularies are expanding beyond the current limits. In addition, speaker-independent systems may soon recognize numeric phrases such as "sixty-four" and be able to exclude non-numeric utterances such as "employees". These new features are likely to increase respondent acceptance and recognition accuracy.

XIII. Acknowledgements:

The authors would like to acknowledge the contributions of the many members of the Division of Monthly Industry Employment Statistics, especially Tony Gomes, who participated in the development and implementation of the Voice Recognition tests. We would also like to thank the staff of the Maine Department of Labor, Bureau of Employment Security, and the BLS Regional Office in Boston and the reviewers of this article. It is also important to acknowledge the Voice Processing Corporation for their provision of the prototype system used in testing and for their reliable assistance throughout the tests.

XIV. References:

1. Werking, G.S., Clayton, R.L., Rosen, R.J., and Winter, D.L.S. (1988): Conversion from Mail to CATI in the Current Employment Statistics Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods.

2. Werking, G.S., Tupek, A. R., and Clayton, R.L. (1988): CATI and Touchtone Self-Response Applications for Establishment Surveys. Journal of Official Statistics, Volume 4, Number 4.

3. Clayton, R.L., Winter, D.L.S., and Gomes, T. M. (1990): Research in Voice Recognition and Voice Response Methods for Survey Data Collection. Proceedings of SpeechTech'90, Media Dimensions.

4. Office of Management and Budget 1990. Computer Assisted Survey Information Collection, Statistical Policy Working Paper 19. Springfield, VA. National Technical Information Service (PB90-205261).

5. Lockerby, William D., Final C & S Double Keying Process Report, (1989) internal BLS report.

6. Clayton, R.L., and Harrell, Jr., L.J. (1989) Developing a Cost Model for Alternative Data Collection Methods: Mail, CATI and TDE. America Statistical Association, Proceedings of the Section or Survey Research Methods.