Journal of Official Statistics
Vol. 8. No. 3, 1992, pp. 377–388
© Statistics Sweden

Speech Data Entry: Results of a Test of Voice Recognition for Survey Data Collection

Richard L. Clayton[1] and Debbie L.S. Winter[1]

Abstract: The Bureau of Labor Statistics (BLS) has conducted tests of microcomputer-based voice recognition technology as an alternative to traditional mail and telephone data collection methods. Results of current research are provided covering performance measures, respondent reaction to Voice Recognition and Touchtone automated self-response methods, data quality measures, collection cost implications, hardware and software requirements, and current limitations of technology. Future developments in voice recognition technology are also discussed. The results indicate that survey designers should reevaluate traditional collection methods.

Key words: Collection methodology; touchtone data entry; CASI.

1. Introduction

The decades of the 1970s and the 1980s saw a dramatic surge of research into new data collection methods. The availability of inexpensive computers provided the means for testing a range of computer-assisted methods, primarily CATI and then CAPI for assisting interviewers and controlling and measuring the interview process. These tests (Groves and Nicholls 1986; Catlin and Ingram 1988) searched for potential mode effects negatively affecting data quality and costs. What few effects were found were overwhelmed by the potential benefits of standardized procedures, on-line edits and reconciliation, and the capture of process control information on the interview process.

This research spawned further uses for computers to streamline the data collection process. As an extension of these developments and research, the BLS has field tested a Voice Recognition (VR) system over two years.

To the general population, the use of voice recognition has been more fiction than science for decades. Only recently have techniques been sufficiently reliable and inexpensive to consider for commercial use. The VR system recognizes the human voice as input to a Computer Assisted Self-Interviewing (CASI) system. The research supports the use of VR as a viable alternative to

[1] U.S. Bureau of Labor Statistics. Contact: Richard L. Clayton, U.S. Bureau of Labor Statistics, Room 4860, 2 Massachusetts Avenue, N.E., Washington D.C. 20212, U.S.A.

Acknowledgements: The authors would like to acknowledge the contributions of Tony Gomes and Louis J. Harrell, Jr., 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 systems used in testing and for their reliable assistance throughout the tests.

data collection methods, such as mail, Computer Assisted Telephone Interviewing (CATI), and Touchtone Data Editing (TDE). The Voice Recognition system is currently in use at BLS to collect economic data in a monthly business survey.

2. The Current Employment Statistics Survey

The Current Employment Statistics (CES) program provides one of the earliest indicators of the United States' 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 U.S. monthly survey, the CES is based on a sample of over 370,000 voluntarily 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. Thus, there are generally about two weeks to collect, key enter, and edit the data, and one week for estimation, analysis, and publication of preliminary national estimates.

Traditionally, the survey has been conducted by mail, whereby an annual questionnaire is mailed back and forth to the respondent each month. The respondent completes the questionnaire with payroll information and returns the questionnaire through the mail. Mail collection typically yields only about a 50% response rate by the deadline for preliminary estimates.

In order to improve the response rates for preliminary estimates, BLS has been testing telephone data collection methods for several years in the CES. These include CATI collection since 1986, TDE since 1987, and now Voice Recognition (VR) since 1989.

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.

The testing of CATI methods yielded improvements of up to 30 percentage points in the response rates for the preliminary estimates (Werking, Clayton, Rosen, and Winter 1988). However, implementing ongoing CATI over such a large sample survey would be too expensive, therefore, research then focused on developing an automated self-response method called TDE. TDE's leading advantages include the ability to retain the same high response rates as CATI, operational ease (Werking, Tupek and Clayton 1988; Phipps and Tupek 1990), and much lower costs (Clayton and Harrell 1989). The disadvantage of TDE is that only 75–85% of our respondents have touchtone telephones. 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% had touchtone phones. In California and Alabama, over 90% of the

businesses had touchtone phones. Thus, the need to provide inexpensive, convenient self-reporting by all respondents spurred BLS to investigate voice recognition technology.

3. Voice Recognition Systems Requirements

There were four basic functionality requirements. First, the system must be compatible with the public telephone system. Second, the system need only recognize the digits "0" through "9" to capture the numeric data, and "yes" and "no" to allow respondents to indicate whether the repeated data were correct. Third, it must be speaker-independent, meaning that any respondent would be recognized without prior training of the system. Some systems, usually those designed for office dictation, required users to speak a standard vocabulary for up to 30 minutes prior to actual use. The system should also recognize continuous speech, so that a delimiter, such as a defined pause or tone, does not have to separate each spoken digit. Continuous speech recognition makes data entry more "conversational" and natural. Both speaker-independence and continuous speech are important features for a system to be used across a broad spectrum of users (Brooks 1989). Fourth, it must be sufficiently low in cost to allow testing.

After an extensive search in 1987 and 1988, we could find only one system meeting these requirements, one which was still in a prototype stage of development. In a joint research project, BLS joined with Voice Processing Corporation (VPC) of Cambridge, Massachusetts to test the system in its first live use. The results have benefitted both organizations, allowing us to evaluate the accuracy and public acceptance of this technology. The VPC Voice system recognizes the digits "0" (both "zero" and "oh") through "9," and the words "yes" and "no." While limited, this vocabulary was well suited for the collection and verification of data from the CES questionnaire. The recognition patterns were developed using people from the American Midwest.

The VPC prototype system used a 386 microcomputer, speech recognizing software, the questionnaire software, and a speech synthesizer for providing instructions, asking the questions, and repeating the entries. Other systems using minicomputers were available, but at seven to ten times the price.

The current VPC technology includes the "recognizers" on a board inserted into a microcomputer, and digitized human voice phrases for a more natural interface.

4. CASI Methods – TDE and VR

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 the specific questionnaire for 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. Three attempts were allowed before the system provided a "help" number to call. Management information, such as date and time of call, and call length, was 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 offered 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. Following verbal prompts provided by the computer, respondents enter numeric data from their completed questionnaires using the number keys on their touchtone phones. Under VR, 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. Both the written instructions and the system provided explicit instructions to limit speech input to the vocabulary of the system. For example, "if your employment is four hundred and twelve, say '4 1 2'." Tests without such instructions encountered problems when respondents did not know to stay within the system's defined vocabulary (Mikkilineni and Perdue 1989).

Neither VR nor TDE perform any editing during the interview for these tests. As described below, we rely on post-collection edits and reconciliation to insure that clean data enter the estimation stage.

5. Research Goals and Approach

These feasibility tests of VR were designed to evaluate whether this method could maintain the same high response rates, user acceptance, and data quality achieved under CATI and TDE.

The following results were drawn from tests in 1989 and 1990 using existing CES respondents from the state of Maine. We chose Maine for two reasons. First, the staff of the Maine Department of Labor has effectively participated in all of the CES research efforts for almost a decade, and secondly, we wished to test the system in part of the U.S. with one of the two most difficult accents for such systems to recognize (the other one is in Louisiana). Virtually all of the units selected had weekly payroll systems. The length of the payroll period is a major determinant of the timeliness of reporting for a time-critical survey like the CES; respondents with the shorter pay periods generally have their data available earlier and can thus report earlier (Werking, Clayton, Rosen, and Winter 1988). The first 70 units were selected from those using the TDE system. Prior to that, these same units had been collected by mail and then CATI. By selecting units already using one CASI system, the conversion to VR was simplified. Also the characteristics of the two systems could be directly compared by knowledgeable users; this also minimized any potential novelty effects. The test was expanded in May 1990 when another 55 units were added for a total of 125 units. These units were converted directly to VR from CATI.

6. 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 received a respondent instruction package which includes a toll-free number and instructions. They also received a practice identification number that allowed them to access and try out the system *before* they reported live data.

After the respondent's first use of the system, a follow-up interview was conducted. The purpose of this interview was to find out whether the respondent encountered any problem with the system and to elicit any comments on the system or CASI procedures. This type of telephone follow-up was first developed to assess the respondent reaction to TDE reporting. Respondents were asked about each step in the reporting process to identify and resolve problems immediately. For the VR tests, similar questions were

Table 1. Average response rates for preliminary estimates: for respondents with weekly payrolls in Maine, 1989 and 1990

Collection method	1989		1990	
	Response rate (%)	Sample size	Response rate (%)	Sample size
Voice	92	70	93	113
Touchtone	94	296	95	640
CATI	93	630	94	350
Mail	59	290	64	253

asked seeking to identify issues involving the VR human-machine interface. These interviews did not include a quality control component because only one attempt to enter data had been made. A subsequent analysis of reported data was conducted at the end of the year, as described in Section 10, to address the accuracy of the reported data to the Voice Recognition system.

During the subsequent months, contact with respondents was quite limited. Respondents received an "Advance Notice" postcard in the mail about the time that their data were usually available and well in advance of the CES collection deadline. If respondents had not reported as the deadline for the preliminary estimates approached, survey staff placed short nonresponse prompting phone calls reminding the respondents to report.

7. Response Rates

The primary research goal in assessing the feasibility of Voice collection was to measure its ability to retain the very high response rates already established under CATI and TDE collection. Over the two year test, Voice did maintain the high response rates experienced under CATI and TDE. Table 1 shows the average response rates for the preliminary estimates for the period from August 1989 to December 1990. Because the length of the pay period is a major determinant of reporter timeliness, this table compares response rates of weekly payroll reporters in Maine across all methods. Thus the response rates compare units with comparable payroll systems in Maine, not with the overall CES sample.

Voice recognition has consistently exceeded a 90% response rate over the period of the tests. This consistency suggests that the high response rates are not due to a novelty effect. The BLS TDE system, differing from the VR system only in the "input" procedure, has sustained high response rates over a five-year period (Werking and Clayton 1991).

8. Respondent Acceptance

The ongoing high response rates achieved over the test periods also indicated a high level of respondent acceptance. Summaries of the first month follow-up interviews reinforced the evidence offered by high response rates. During the 1989 test, all of the 70 respondents using Voice for the first time were asked to comment on Voice reporting and compare it to Touchtone reporting. About 60% of respondents preferred Voice to Touchtone, with most citing that Voice is easier to use and more natural. About 32%

preferred TDE; 10% 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" compared to VR. 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 when, in fact, Voice calls were about 20 seconds longer than the average two minute Touchtone interview, because of lengthier instructions and prompts.

9. Problems Encountered

During the first month, 85% of the respondents converted in 1989 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 useful; the other half did not seem to think it was necessary.

During ongoing collection, after the first month on Voice, 98% of all interviews were successfully completed. Only 2% 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 two years of the tests, 6 respondents out of 125 were returned to touchtone collection due to these persistent recognition problems with the prototype system. The system could not recognize the voices of these 6 individuals over the course of several months. From our follow-up conversations, we know that most of these respondents spoke with the Maine "Down East" accent, and that the system had trouble with twos and fours, consistent with the pronunciation differences heard in this accent compared to that of the Midwest.

By the end of 1990, 113 of the 125 units were using Voice. Besides the 6 units returned to TDE, 3 were out of business, and 3 had refused further participation in the survey.

10. Data Quality: Mode Error

Two major areas of concern that would also affect the successful use of VR were recognition accuracy and respondent acceptance; indeed these two are linked as no respondent would accept the system if the system would not allow them to successfully complete reporting their data. With regard to accuracy, the system would have to be able to recognize the speech patterns of a wide range of respondents, and any limits of the recognition would have to be acceptable to respondents. Respondent acceptance was also a matter of interest – How well would respondents accept "conversing" with a computer system? Were the procedures used and instructions provided satisfactory for accurate reporting? Records of problems encountered and solutions would provide rough measures of recognition capabilities and respondent acceptance.

The use of the VR system entailed evaluating recognition errors and respondent data entry errors that might introduce an additional source of nonsampling error called "mode error." While every collection method carries the potential for introducing some error, the choice of methods entails a comparison of the levels and characteristics of the errors.

To measure the effect of voice collection on data quality, voice respondents were asked to return copies of the CES questionnaires they had maintained during the calendar year. The completed questionnaires were then compared to the data entered into the voice collection system as a measure of the respondent's ability to enter data correctly, and the system's ability to interpret the incoming speech correctly. For the purpose of this study, the data on the questionnaires were assumed to be correct. For 1989, 47 of the 70 units returned the forms, and 100 of the remaining 113 units returned the forms for the calendar year 1990.

Each record of data entered into the voice system contains a number of input data items: the unique identification number, the reference 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. After each data item is entered, it is repeated for verification. The respondent was allowed three attempts if necessary. Each incorrect data item counted as one error.

The results of the comparison yielded a very small overall incidence of error. In the original 1989 study from a total of 1164 observations, 27 observations contained an error, for an overall error rate of 2.3%. The 1990 study found an annual error rate of 1.3% based on 6463 observations.

These errors can be broken into two categories: data entry or recognition errors and procedural errors. The cause and treatments for each type are different.

10.1. Data entry or recognition error

The analysis originally focused on data entry or recognition errors since it was expected that these types of errors would be most prevalent in such a new technology. 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. Since we could not distinguish between errors in recognition versus data entry, or misspeaking, these are defined together. Only 7 such errors (involving five respondents) were found in the 1989 data (0.6%), and 35 errors in 1990 (0.5%). In other words, the entries were accurately entered and recognized in 99.4% and 99.5% of the observations, respectively. See Table 2.

Other studies, using different systems yielded similar results (Wilpon 1985; Wilpon, Mikkilineni, Roe, and Gokcen 1990).

Determining whether any error rate is acceptable may be accomplished by comparing the error rate with that of alternative modes. In both Voice and TDE, each data item is repeated to the respondent for verification, providing 100% verification of each entry. The error rates described above were comparable to key entry error rates found in one other study using fully verified key entry (Lockerby 1989).

We also looked at the timing of the errors. Table 3 shows the incidence of data entry/recognition errors by the number of months using VR. For the units added during 1989, there appears to be a learning curve after which recognition errors seem to diminish. We worked with each respondent that alerted us to problems, or that we could identify as

Table 2. Incidence and error rates by type of error in Voice data collected

Error type	1989		1990	
	Incidence	Error rate (%)	Incidence	Error rate (%)
Data entry or recognition	7	0.6	35	0.5
Procedural	20	1.7	49	0.8
Totals	27	2.3	84	1.3

having problems. One of the respondents with problems, accounting for two of the errors, was returned to TDE collection.

The data for 1990 reflect the newly added respondents, May through December. The data do not necessarily show the same pattern, although a review of the data collected during 1991 may show a decline in the incidence of this type of error.

10.2. Procedure error

A second type of error was 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." In 1989, the error rate for this type was 1.7%, but only 0.8% in 1990. It was discovered that the vast majority of these errors in 1989, 17 of 20, were committed by one respondent over several months. After follow-up with this respondent, the error rate declined in the following year.

It is important to note that these errors would usually be identified by post-processing system edits. By comparing the original VR data file with the file used in the actual estimates, we found that all were corrected prior to estimation.

Table 3. Incidence of data entry/recognition errors by length of reporting

Number of months using Voice	Incidence of data entry/recognition errors	
	1989	1990
First	4	1
Second	1	3
Third	1	2
Fourth	0	5
Fifth	1	0
Sixth	0	4
Seventh	0	1
Eighth	–	0
Respondents	70	113

10.3. Comparison of error rates: VR versus TDE

To further evaluate the errors, we compared VR results to a similar study of TDE reported data (Phipps and Tupek 1990). Table 4 shows the incidence of errors by data type along with the potential field length. For both systems, higher error rates were seen for the data items requiring the most digits. For example, the payroll item typically has the largest number of digits, and potential for error. However, we found the greatest number of procedural errors with the entries for payroll and hours. For example, for entries of payroll, the CES asks for dollars only. Eleven of the 49 procedural errors found in 1990 were respondents entering cents. Similarly, 25 of the 49 procedural errors entailed

Table 4. Comparison of error rates: VR versus TDE

Data type	Maximum field length	VR	TDE
Reference month	2	0.7%	–
All employees	6	0.5%	1.2%
Women workers	5	0.1%	1.5%
Production workers	5	0.4%	1.5%
Payroll	8	3.1%	2.5%
Hours	7	1.8%	2.5%
Total	–	1.3%	1.8%
N (for each item)		6463	1930

entering a zero when a blank was appropriate.

11. Solutions

Since most first month respondents reported without incident, and problems in subsequent months were negligible, the respondent materials providing instructions seemed 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 the initiative to solve their own problems with the system before calling survey staff for assistance. About half of all respondents 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 capabilities of 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 call survey staff for assistance, and would often succeed on the next report.

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

12. Costs of Voice Recognition Relative to Other Collection Methods

Given respondent acceptance and recognition accuracy, costs are the next major determinant of implementation feasibility. As in other automation efforts, machines are used to replace labor intensive tasks. Both TDE and VR replace many of the manual activities in a mail survey. Figure 1 lists the major activities involved in data collection. For each collection method, arrows indicate the incidence of costs and the direction of recent cost changes for each activity. For example, mail collection involves mail-out and mail-return activities each month followed by data entry and verification. Under TDE and VR, these monthly activities are largely eliminated. Respondents retain the questionnaire for a year, computer generated "advance notice" postcards remind respondents to call in each month, and respondents enter and verify their own data.

While VR reduces many labor intensive

(Arrows show direction of recent price change)

Cost Category	Mail	Automated Data Collection (TDE & VR)
LABOR		
Mail out	↗	
Mail return	↗	
Data entry	↗	
Edit and edit reconciliation	↗	↗
Nonresponse followup		↗
NON-LABOR		
Postage	↗	↗
Telephones		↘
Microcomputers		↘

Recent Annual Price Change Factors

Labor	+4.1%	ECI, State and Local Government
Postage	+5.0%	U.S. Postal Service
Telephone	-2.0%	CPI-U, Intrastate toll calls
Microcomputers	-24.9%	Producer Price Index for 16 bit computers

Fig. 1. Data collection costs

costs, new capital costs are incurred, particularly for telephone charges, the amortized microcomputer technology and systems design, and ongoing maintenance. As with most new technological products, initial costs are high, then as start up costs are recouped and mass production efficiencies are achieved, prices drop rapidly. Since the BLS studies began, the costs of the most expensive part of the VR system, the recognizer, had dropped to 20% of its original cost.

The overall cost comparison for the CES is very favorable (Clayton and Harrell 1989). TDE costs were about 30% less than current mail costs, with VR costs somewhat higher than TDE. This cost advantage is likely to grow over time. The costs of labor have risen at about a 5% rate annually, whereas the costs of telephone charges and microcomputer technology have declined in recent years.

13. Conclusions

Response rates. Voice collection meets the high response rates attained under other automated collection methods (CATI and TDE) and has consistently exceeded the 85% 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 exacerbate 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.

Costs. The costs of automated methods, even when accounting for amortized research and development costs, are substantially lower than mail operations, and this cost advantage will increase over time.

Implementation. Based on the results of these tests, the ongoing collection performance, and the known limitations of the availability of touchtone phones, Voice Recognition is an essential part of the overall implementation plan for automating data collection in the CES survey.

14. 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 test data become 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 turn-around 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. 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 and user specified vocabulary can be accommodated. In addition, speaker-independent systems can already recognize numeric phrases such as "sixty-four" and exclude non-numeric utterances such as "the number of employees is . . . ." These new features are likely to increase respondent acceptance and recognition accuracy. VR systems are beginning to appear in the general market place through the telephone systems, and for other applications including office dictation and factory floor data entry (Byford 1990). Many providers of familiar touchtone applications are including Voice features to augment the existing systems.

15. References

Brooks, C.M. (1989). User Interface Design for Speech Recognition Telephone Applications. Speech Technology, 4, 58–60.

Byford, R.G. (1990). Advances in Voice Data Collection. Transactions of the 44th Annual Quality Congress, American Society for Quality Control, 592–600.

Catlin, G., and Ingram, S. (1988). The Effects of CATI on Costs and Data Quality: A Comparison of CATI and Paper Methods in Centralized Interviewing. In Telephone Survey Methodology, eds. R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, New York: John Wiley, 437–450.

Clayton, R.L. and Harrell, Jr., L.J. (1989).

Developing a Cost Model for Alternative Data Collection Methods: Mail, CATI and TDE. Proceedings of the Section on Survey Research Methods, American Statistical Association, 264–269.

Groves, R.M. and Nicholls II, W.L. (1986). The Status of Computer-Assisted Telephone Interviewing: Part II – Data Quality Issues. Journal of Official Statistics, 2, 117–113.

Lockerby, W.D. (1989). Final C & S Double Keying Process Report. Internal Report, Bureau of Labor Statistics, Washington, D.C.

Mikkilineni, R.P. and Perdue, R.J. (1989). Experiences in Implementing Automatic Speech Recognition in the Public Telephone Network. Proceedings AVIOS '89, 27–32.

Phipps, P.A. and Tupek, A.R. (1990). Assessing Measurement Errors in a Touchtone Recognition Survey. Paper presented at the International Conference on Measurement Errors in Surveys, Tucson, Arizona.

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. Proceedings of the Section on Survey Research Methods, American Statistical Association, 431–436.

Werking, G.S., Tupek, A.R., and Clayton, R.L. (1988). CATI and Touchtone Self-Response Applications for Establishment Surveys. Journal of Official Statistics, 4, 348–362.

Werking, G.S. and Clayton, R.L. (1991). Enhancing Data Quality Through the Use of Mixed Mode Collection. Survey Methodology, 17, 3–14.

Wilpon, J.G. (1985). A Study of the Ability to Automatically Recognize Telephone Quality Speech for Large Customer Applications. AT&T Technical Journal, 64, 423–451.

Wilpon, J.G., Mikkilineni, P., Roe, D.B., and Gokcen, S. (1990). Speech Recognition: From Research to the Real World. Proceedings AVIOS '90, 103–108.

Received March 1991
Revised February 1992