

**Respondents' Attitudes
Towards a
U. S. Census Voice-Recognition Questionnaire**

Cleo R. Jenkins

and

Marty A. Appel

U.S. Bureau of the Census

May, 1995

This paper was prepared for presentation at the International Field Directors' and Technologies Conference in Fort Lauderdale, Florida. Opinions expressed in this paper are those of the authors and should not be construed as representing those of the U.S. Bureau of the Census.

Respondents' Attitudes Towards a U. S. Census Voice-Recognition Questionnaire

Cleo R. Jenkins and Marty A. Appel
U.S. Bureau of the Census

The United States Bureau of the Census is the largest data collection agency in the United States, with the Census of Population and Housing (better known as the decennial census) amounting to the largest data collection effort. Since 1970, the primary means of collecting decennial data has been through the mail. Since personal visit interviews are more expensive, they are reserved for mail non-respondents. Obviously, it would be advantageous if an intermediate method could be employed-- if, for instance, respondents could be persuaded to answer the census by telephone in the event they either refuse or do not get around to responding by mail. Furthermore, this method would be made even more cost-efficient if a computer could be substituted for a live interviewer in the telephone interview. For this reason, the Bureau of the Census contracted with the Center for Spoken Language Understanding of the Oregon Graduate Institute of Science and Technology (OGI) to develop a voice-recognition decennial questionnaire (see Appel and Cole, 1994 for an in-depth description of this program of research).

To a large extent, being able to substitute a computer for a live interviewer depends on a capable technology. The OGI continues to conduct research in this area, i.e., to improve a computer's ability to simulate speech (e.g., Cole et al., 1994; Cole et al., 1992). However, having a capable technology is not the only criterion to consider, for the success of a voice-recognition questionnaire also depends on how well or poorly the technology is received by the general

public. Asking respondents to answer a voice-recognition questionnaire will not benefit the Census Bureau (or any survey organization for that matter) if respondents are against the new technology, and as a result of this negative attitude, indiscriminately refuse to be enumerated at all. This paper describes research undertaken to evaluate respondents' attitudes towards the voice-recognition Census questionnaire developed by OGI.

1. Laboratory Interviews

1.1. Recruitment of Respondents

We contracted with a marketing research company to recruit 40 subjects for a small-scale experiment. All federal and military employees were excluded from participation. Respondents were recruited from three income categories (less than or equal to \$19,999; \$20,000 to 99,999; and greater than or equal to \$100,000), three race categories (White, Black, and Other), and three age categories (less than 25, 25 to 55, and greater than 55 years). Also, they were distributed by gender (female and male). Table 1 provides a description of the respondents by demographic characteristics. It should be noted that this is a non-probability sample.

All interviews were held in the Center for Survey Method Research's (CSMR) Response Research Laboratory and (with the consent of the respondent) audiotaped. CSMR researchers conducted the interviews the last week in June and the third week in July. Half of the interviews were

conducted late in the day (between 4:00 and 6:00 p.m.) as a way of safeguarding against getting only nonworking individuals for participation.

1.2. Methodology

Respondents were assigned to one of two treatments. In the first treatment, respondents were given the voice-recognition questionnaire and, as a control, respondents in the second treatment answered an interviewer-administered questionnaire. The control served as a necessary comparison, a standard against which to measure the results of the voice-recognition questionnaire. Also, the control treatment was used to disentangle respondents' attitudes towards the mode of administration from other aspects of the survey experience.

The recruitment agency provided us with a list of respondents by appointment times. We reasoned that conducting all of the voice-recognition interviews in a concentrated block of time would lessen the technical support staff's burden, since it meant they needed to monitor the computer hardware and software for less time. We also thought it would be better to focus on one type of interview at a time. Therefore, we conducted all of the voice-recognition interviews first, followed by the interviewer-administered ones. For these reasons, we did not attempt to randomly assign respondents to the two treatments.

The treatments had a pretest-posttest design. The pretest session was designed to elicit respondents' preconceived notions about a computerized questionnaire: what did they think a

computerized questionnaire would be like, and specifically, were they willing to answer one based on their preconceived notions? We reasoned that if respondents had negative objections at this juncture, the pretest might provide us with valuable insight into how to best present the computerized questionnaire during an actual census to minimize respondents' negative reactions.

Following the pretest, we asked respondents to answer the voice-recognition questionnaire. We connected the only available phone in the interviewing room, the wall phone, to a portable tape recorder and we lengthened the cord attached to the telephone receiver so that it reached the respondent's seat. We did not want respondents to divert their attention from the interview by having to deal with the tape recorder. Therefore, the CSMR researchers dialed the 1-800 number, handed the telephone receiver to the respondent, and placed the tape recorder on the desk. While the respondent answered the questions, the researcher left the room to guard against her presence affecting the respondent's behavior. She listened to the respondent's half of the conversation on a video monitor from an adjacent room.

When the respondent was done, the researcher returned to the room and asked a series of probes designed to elicit respondents' reactions to the questionnaire. The strategy was to capture as much information as possible using open-ended probes like, "Please tell me your reactions to the system." and "Any other reactions?". The researchers delivered their probes slowly to give respondents ample time to formulate and voice their opinions. Only after they had exhausted respondents' spontaneous opinions on the subject, did they turn to more specific probes in an attempt to find out respondents' opinions on elements of the interview respondents either had not

thought about or had not expressed an opinion about. Also, we wanted some closed-ended responses for comparison between the two treatments.

The second treatment followed a similar format, except that the respondent spoke with a human being rather than the computer. Interviewers from the Census Bureau's telephone facility in Hagerstown administered the live interviews.

The test questionnaires in the two treatments asked the same number of questions and for the same information: the 100-percent self-reported person items (i.e., the respondent's name, gender, marital status, date of birth, Spanish or Hispanic origin, and race). However, the wording of the questions differed to accommodate differences in technology. For example, the computer asked the date of birth question as follows:

We will now ask about your date of birth. What month were you born? What day of the month? What year?

In contrast, the live interviewer asked:

What is your date of birth? Please give me the month, day, and year.

The computerized questionnaire was developed by the OGI for use specifically with the computer. The interviewer-administered questionnaire was the same as that used for the Mail and Telephone Mode Test.

Table 1 presents the demographic characteristics of the respondents by treatment. Treatment 1's sessions averaged 24 minutes, whereas Treatment 2's sessions lasted 18 minutes.

2. Limitations of the Research

Caution should be exercised in interpreting the results presented in this paper for the following five reasons. First, the U.S. Decennial Census probably has greater name recognition than any other census or survey, so care needs to be taken when generalizing the results of this study to any other census or survey.

Second, we introduced the questionnaires by saying that the Census Bureau was considering giving people the option of answering a computerized questionnaire by telephone (in the voice-recognition treatment) and the Census by telephone (in the interviewer-administered treatment). Respondents' reactions may very well have been dependent upon these introductions.

Third, due to time and resource constraints, the research only included 20 respondents per treatment. Care needs to be taken when drawing conclusions about differences between the two treatments.

Fourth, the respondents we interviewed may differ from the general population. Although we attempted to recruit a diverse group of respondents, ultimately we were dependent on respondents in the Washington, DC area who were willing and able to come to our laboratory at a specified time for an interview. Also, it should be noted that the respondents were paid \$25 for participating and they were drawn from a marketing research firm's database. In terms of the final minority composition, the participants included 13 Blacks and 4 "Others," distributed between the two treatments. Only 4 of the Blacks and 2 of the "Others" were tested with the computerized questionnaire.

Fifth, laboratory research of the kind described in this paper can suffer from demand characteristics; that is, subjects feel obliged to give the experimenter the types of behavior and results the experimenter wants (Lindzey and Aronson, 1969; Rosenthal, 1963; Orne, 1962).

It should be noted that the first three limitations apply equally to both treatments, which is to say that these limitations might introduce error into the absolute numbers, but there is no reason to believe these errors would be differential. However, this may not be the case for the fourth and fifth limitations. Less minorities received the voice-recognition questionnaire than the interviewer questionnaire, and to the extent that respondents' attitudes are related to their race, differences may result. Also, it is always possible that the researchers unwittingly projected different expectations in the two treatments, thereby leading to uncontrolled differences, although we do not believe this to be the case.

3. Results and Discussion

3.1. Respondents' Reactions

Overall: Pre- and Post-test

Willingness. Respondents demonstrated a great deal of willingness to answer the census by telephone, both before and after doing so. Overall, 90 percent voiced a willingness before and 93 percent afterwards. Respondents' reasons for being willing reflect their own self-interests. Overwhelmingly, they claimed that answering the census by phone would be more convenient, easier, and faster for them. However, a number of respondents also demonstrated an awareness that the Census Bureau would benefit from using a computerized system, saying that a computerized questionnaire would be more cost efficient and easier for the census.

Respondents clearly recognized that we were asking them to call us, rather than us calling them. They immediately began to make comparisons between this and the norm and began to explain the two things they disliked most about the telemarketing calls they generally receive. First of all, they dislike picking up the phone to hear a pre-recorded message telling them to hang on for some upcoming message. In the second place, they dislike being called at an inconvenient time. The thought of being given control over when the interview takes place, thus alleviating being "bugged" by someone else, was seen as a very positive element by participants. A few

respondents wisely pointed out, however, that this also provided procrastinators and less motivated people the opportunity to put off responding.

Preference. After we had completed 10 voice-recognition interviews, we adjusted our protocol. At this point, we realized that although we were capturing respondents' willingness to answer the census by phone, we were not learning if they preferred this method, unless they happened to volunteer this information (fortunately, 8 of the 10 did spontaneously tell us their preferences). So we decided to include a question about this. Since research demonstrates that hypothetical answers are not necessarily indicative of behavior, we asked respondents to first describe for us how the census is taken. If respondents provided this information (and 27 of the remaining 30 respondents did), we then asked them if they would prefer answering the census the way they thought it was taken or by phone. Altogether, we learned the preferences of 35 respondents.

Respondents' knowledge about how the census is taken varied greatly. A number of respondents had very serious misconceptions about this process. About one-third of the respondents thought that the census was taken, as they put it, door-to-door. Generally, this was because they were remembering earlier times when the census was entirely collected by personal visit interview. There was one person, however, who thought filling out the census was like getting a driver's license. She thought she needed to go to a center, where she would be given a questionnaire and an accompanying answer sheet in which to fill in circles. The rest of the respondents (nearly 60 percent) understood that the census was taken by mail. Of these, a few described the census

questionnaire with remarkable accuracy, and understood that unreturned mail questionnaires were followed up with personal visit interviews.

In the pretest interview, only 57 percent (20 out of 35) of the respondents said that they preferred to answer the census by phone. In other words, although people were willing to answer the census by phone, this was not necessarily their preference. Nearly 34 percent of the respondents (12 out of 35) expressed a preference for either mail or personal visit interviewing, and the remainder did not have a preference.

In the posttest interview, respondents' preferences changed slightly, if at all. At this point, nearly 64 percent (21 out of 33) preferred the phone, in comparison to the 27 percent (9 out of 33) who preferred either mail or personal visit interview. The remainder did not have a preference.

The overriding conclusion to be drawn from these results is that respondents appear very willing to answer the census by phone, probably too willing, given what we know about response rates. More than likely this is an indication of demand characteristics in operation--that is, respondents are telling us what they think we want to hear. Also, respondents were selected for cooperation. These are people who were willing to come into the laboratory to answer questions in the first place. Still, the fact that fewer respondents actually prefer answering the census by phone than are willing to do so is probably a reliable finding.

By Treatment.

Preconceptions. Before asking respondents if they would be willing to answer the census by computer, we asked them to describe for us what they thought a computerized questionnaire was. It was perfectly clear that they expected the computerized questionnaire to be a touch-tone system with a pre-recorded voice. They repeatedly said that they expected to have to "push buttons." Respondents of all age groups and income levels were quite familiar with touch-tone systems. Younger respondents cited registering for classes through these systems or using AT&T's voice mail, and older respondents and respondents of all income levels cited familiarity with electronic banking systems. One respondent probably summed up this situation for most people when he said that he was "getting used to machines."

Quite a few respondents expected a pre-recorded voice. They described their expectation like this: "a taped voice, a tape recording conversation, questions are read off a tape." But the most creative description came from the respondent who expected "a digital man."

A fair number of respondents understood that their answers would be input directly into a computer. As mentioned earlier, a few respondents also recognized that this direct link between the respondent and the computer was beneficial to the Census Bureau.

Post Reactions. It is interesting to note that, in general, respondents were willing to answer the census using a touch-tone system. No one seemed to anticipate a voice-recognition system. As a result, most were impressed and relieved when they discovered that they did not need to push buttons and that the interview was more natural than they had anticipated. For instance, one

savvy respondent said, "It was pretty user friendly. It was at least somewhat conversational. It actually understood 'yes' or 'no' and it translated everything." Another said, "No numbers. I definitely prefer spoken than push-button." Another said, "It did not sound like a Computer Willie." In fact, it was quite interesting to hear respondents talk about the computer as if it was a person. Since it had a male voice, they sometimes referred to the computer as "he" rather than "it."

The conclusion to be drawn from respondents' remarks is that increasing their awareness of the conversational nature of the system before asking them if they would like to use it would probably enhance their already favorable dispositions.

Objections. However, in all fairness, a number of respondents were not so kind. To begin with, two "willing" respondents were dissatisfied with the system. According to one of these respondents, the recorder was antiquated. He thought the question-response task was slow and methodical and suggested that it be quickened. He also thought the chimes and beeps in the conversation were redundant. He said that he could understand the first beep, but not the second one. The other respondent said he thought the system "stunk." He thought the system was too slow. He said it was rather irritating to have to go through it piece by piece.

Also, two "unwilling" respondents objected to the computerized questionnaire. Beforehand, one of these said that he did not like talking on the phone. He said that he would prefer having more time to fill out a mail questionnaire. Afterwards, he said that he thought the computerized voice

was just too impersonal. The other respondent talked at great length about the fact that he could not see the benefit to him. He said that if we wanted to computerize to save money and manpower that we would need to make it clear to people why they should answer the census. He said that simple explanations, such as the government will provide better services, were not adequate.

However, two respondents also changed their minds after going through the computer interview. One was initially willing, but unwilling afterwards. She decided that the system would take too long if there were more questions than just the few we asked and that it would be better to just fill out a questionnaire. On the other hand, another respondent was converted from a respondent with mixed feelings to a willing respondent for just the opposite reason. He decided that it would be simpler to answer the computerized questionnaire than having to walk to the mailbox.

Likewise, people were also converted in the live interviewer treatment. One went from unwilling to willing, while the other went from unwilling to mixed feelings. The unwilling respondent was initially unwilling because she believed that everything should be in writing. Afterwards, she decided that she would be willing to do it by telephone if it was compulsory, although she still preferred to do it by mail. The other respondent was suspicious. Beforehand he said that "things on the telephone can be manipulated" and that "they can slip so many hidden messages on the phone." Afterwards, he said that the confidentiality statement made him nervous.

Willing with More Questions. Sometimes respondents mentioned that they expected more questions. When this occurred, we took advantage of the opportunity and asked them if they would still be willing to answer the questionnaire if it contained more questions. The majority said they would in both treatments.

Overall Assessment. Respondents' reactions in both treatments were similar, with minor differences favoring the interviewer-administered questionnaire. Two additional respondents liked the live interviewer "very much" compared to the computer. Also, the most negative thing respondents said about the live interview was that they "neither liked nor disliked it," whereas three respondents actually went so far as to say that they disliked the computerized questionnaire "somewhat" and one of the more intractably unwilling respondents actually disliked it very much. The reason cited most often for disliking the computerized questionnaire was its slowness.

Errors. No one had difficulty understanding either the computer or the live interviewer. Respondents often said that the computer was very clear and distinct. However, the opposite was not true. The computer often had difficulty understanding respondents. A little over half the time it either got the person's name wrong or their date of birth wrong. A couple of times it got both the name and birthdate wrong and once, it misunderstood a respondent's race. The majority of people saw this as a minor frustration. They realized that the system was in its infancy and simply assumed that we would iron out the bugs eventually. However, a few people were annoyed and suspicious of the errors. In fact, one person thought we purposely planted the errors to test his tolerance level.

Confusing. We asked respondents if anything about the questionnaires was confusing. Again, the interviewer-administered questionnaire fared better by a small amount. Four people in the computer treatment compared to one person in the live-interviewer treatment reported that something was confusing. As is often the case with the census form, respondents mostly had contextual problems with the Hispanic origin and race questions. They did not understand why we asked a question about Hispanic origin first and why this ethnic group was singled out. In addition, two respondents who answered the computerized questionnaire did not understand why we only provided three race categories (White, Black, Other), and one respondent had a problem understanding the "now married" category. He suggested that we simply ask if a person is "married, single, or divorced" rather than if they are "now married, widowed, divorced, or separated."

Frustrating. After we had conducted the first nine interviews, we adjusted the protocol a little and began to ask respondents if anything about the questionnaires was frustrating. Here the results were about equal, although it is essential to realize that we did not ask this question of as many respondents in the computer treatment. About a quarter of the respondents reported frustrations in both treatments. In the computer treatment, respondents were concerned about the computer inputting the wrong data. The frustrations in the live interviewer treatment had to do with either too short or too long a processing time in some way. One respondent complained that the interview was too short, while another complained that the list of race categories was too long, and a third complained about having to spell everything out. However, one respondent had a different concern. He was sensitive about being asked personal information.

Natural Sounding. One major concern about using a computerized questionnaire was that people would not want to answer it because it would not sound natural enough. In order to test this, we asked respondents in both treatments how natural the interview sounded to them. Again, the results were similar, with the interviewer-administered questionnaire demonstrating another slight advantage. Nineteen respondents thought the live interviewer sounded either very or somewhat natural sounding, compared to 16 respondents in the computer treatment. No one thought the live interviewer sounded unnatural, while 4 thought the computer sounded either somewhat or very unnatural.

It is interesting to note that altogether a third of the respondents rated both interviews as only "somewhat natural." Respondents in the live interviewer treatment astutely pointed out that the interviewer was reading from a script. In other words, respondents understood that even the interviewer-administered questionnaire was comprised of constrained speech as opposed to free dialogue.

Speed and Pace of Interview. The largest difference between the two treatments was respondents' reactions to the speed or pace of the interviews. All of the respondents thought the live interviewer spoke at just the right speed, whereas six respondents thought that the computer spoke too slowly. Along with this, nearly all of the respondents said that they had just the right amount of time to provide an answer in the live interview, but half of the respondents said that they had too much time to answer the computer. They did not appreciate the fact that the computer needed time to determine when they were finished speaking. In comparison to human

speech, a noticeable gap existed between the time they finished answering one question and the start of the next one.

Correcting Wrong Information. At the end of the interview, the computer summarizes what the respondent told it by repeating the information as recorded by the computer. It then asks the respondent to list the questions it got wrong. We asked respondents if they had any difficulty with this part of the interview. Four people reported having difficulty. Three of these suggested that being read the entire list of information was problematic. For instance, one person said that she got stuck listening to the computer spell back her name, and so she missed the rest of the list. Another explained that if the computer had gotten more than three items wrong, he was sure that he would not be able to tell the computer which items were recorded improperly. Another pointed out that if an operator had come on, she would have asked the operator to read the list again.

Suggestions. Respondents had suggestions, and some excellent ones at that, for improving the computerized questionnaire. Their suggestions centered around the following three concerns: correcting the errors the computer made, quickening the pace of the interview, and improving the wording of some questions.

1. Correcting Errors

Respondents cared that the computer got information about them wrong. Furthermore, being asked to listen to a long list of information at the end of the interview taxed their cognitive abilities. As a result, they requested that the burden be more evenly distributed. Seven of the twelve respondents who were misunderstood by the computer suggested that the computer repeat the information back to them after each question, giving them the opportunity to immediately correct it. This is something that should be very seriously considered, since it was offered as a solution by so many respondents.

2. Quickening the Pace

The next most discussed problem was the pace of the interview. Here again, a couple of respondents had a similar suggestion. They recognized that the pace may need to be different for different types of people, so they suggested that they be given control over the pace. To accomplish this, they suggested that they be allowed to say something like "next" to signal to the computer that they were ready to move on.

3. Question Wording

Not surprisingly, recommendations for re-wording the questions centered around the Hispanic origin and race questions. One person suggested that we not single out Hispanics in the Hispanic origin question and another suggested that we provide more

than three race categories in the computer questionnaire. Of course, the live interview questionnaire offered all of the race categories, and one person complained about this, too.

4. Conclusion

The development of a voice-recognition Census short form represents the Census Bureau's first attempt to simulate the verbal transaction between a live interviewer and a respondent. The goal of our research was to learn how respondents felt about this simulation. Relative to the interviewer-administered questionnaire, respondents in this study reacted quite favorably towards the voice-recognition questionnaire. They ranked the voice-recognition questionnaire just slightly below a comparable interviewer-administered questionnaire in terms of most characteristics. They anticipated a touch-tone system and were clearly pleased with the voice-recognition system from that perspective. Despite the system's achievements, respondents identified two areas in need of improvement. They were: 1) the error rate and 2) the pace of the interview. Respondents also made some excellent suggestions for decreasing the error rate and increasing the pace of the interview.

Although we found nothing in our research to suggest that, in general, respondents would refuse to be enumerated simply because we offered them the option of answering the census by computer, the limitations of this research must be kept in mind. Experimental research with a

representative probability sample is necessary before we can confidently conclude what the effect of this new technology would be on response rates.

References

- Appel, M. V. and Cole, R. (1994). "Spoken Language Recognition of the Year 2000 Census Questionnaire," Proceedings of the American Association for Public Opinion Research.
- Cole, R. Novick, D.G., Fanty, M., Vermeulen, P. Sutton, S. (1994). "A Prototype Automated Spoken-Language Questionnaire for the Year 2000 Census," *Speech Communication, Special Issue*, December, 1994.
- Cole, R., Fanty, M., Roginski, K. (1992). "Recognizing Spelled Names with Telephone Speech," Proceedings of Speech Tech/Voice Systems Worldwide 1992.
- Lindzey, G. and Aronson, E., eds. (1969). The Handbook of Social Psychology IV, Reading, Mass: Addison-Wesley Publishing Co.
- Orne, M.T. (1962), "On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications," *American Psychologist*, 17, 776-783.
- Rosenthal, R. (1963), "On the social psychology of the psychological experiment," *American Psychologist*, 51, 268-283.

Table 1. Respondent Characteristics, Overall and By Treatment

Characteristics	Overall	Treatment	
		Computer	Interviewer
Income			
< \$20,000	7	2	5
\$20,000-99,999	24	15	9
> \$99,999	9	3	6
Race			
White	23	14	9
Black	13	4	9
Other	4	2	2
Age			
< 25	6	3	3
25-55	21	9	12
> 55	13	8	5
Sex			
Male	16	9	7
Females	24	11	13
Interview Time			
8:00am-4:00pm	19	9	10
4:00pm-6:00pm	21	11	10

