ABSTRACT

Prior research has relied on *qualitative* analyses of customers' responses during depth interviews. This article discusses how observational monitoring, subjective encoding, and automatic encoding can be used to *quantitatively* analyze depth interview data. It compares the benefits and costs of each method, and discusses computer software packages that assist in automatically encoding verbal data. In an innovative experiment, we compared observational monitoring and automatic encoding of customers' responses to open-ended questions about telecommunication services. We conclude that automatic encoding is a feasible method of analyzing depth interviews that provides detailed quantitative information about customers' responses.

Quantitative Analyses of Depth Interviews

Ruth N. Bolton
Tina M. Bronkhorst
GTE Laboratories Incorporated

Market researchers frequently conduct depth interviews with customers to obtain customer feedback about new products, services, or advertising concepts; to stimulate new ideas about older products or services; and to generate information that is helpful in developing questionnaires or interpreting quantitative data. These interviews typically elicit unstructured or open-ended responses from customers during focus group or one-on-one (i.e., personal or telephone) interviews. Consistent with the exploratory nature of depth interviews, market researchers have almost invariably conducted qualitative analyses of interviews (Greenbaum, 1987; Wells, 1974). These qualitative analyses rely on an expert's subjective interpretation of the verbal and nonverbal data elicited from the respondents.

In contrast, researchers in other disciplines have conducted quantitative analyses of verbal data. For example, psychologists analyze verbal

Psychology & Marketing
Vol. 8(4) 275–297 (Winter 1991)
CCC 0742-6046/91/040275–23$04.00

protocols (elicited by asking respondents to "think aloud") to investigate the cognitive processes that operate when individuals are solving a problem or making a decision. Consumer researchers code and analyze individuals' verbalizations obtained in the laboratory and in retail store environments (Douglas, Craig, & Faivre, 1981). Survey researchers have used quantitative methods to examine verbal behavior elicited during personal interviews and pretests (Bercini, 1989; Weber, 1985). These studies suggest that the quantitative analysis of verbal data may provide a detailed, objective description of a depth interview.

Recently, *qualitative* analyses of verbal data have been considerably enhanced by new computer-based techniques for analyzing text files. For example, AGENDA, a microcomputer-based, free-form data-base program, performs the basic tasks of qualitative data analysis for field notes or transcribed interviews (Wolfe, 1990). CORE-AI contains artificial intelligence tools that provide qualitative-analysis capabilities and flexible access to a very large data base containing customer remarks about service satisfaction (Molina, 1991). In contrast, there are few computer-based techniques for the *quantitative* analysis of verbal data elicited during depth interviews.

This paper reviews three methods for the quantitative analysis of verbal data, focusing on computer-based techniques. It describes a software package, *Systematic Analysis of Language Transcripts* (SALT), that can be used to automatically encode verbal data. It discusses the benefits and costs of using computer-assisted coding methods to quantitatively analyze depth interviews. The article concludes with a description of an innovative experiment that compared observational monitoring with computer-assisted encoding of customers' responses to open-ended questions about telecommunication services.

BACKGROUND

A critical issue for the quantitative analysis of depth interview data is the development of a coding scheme that is appropriate to the research objectives. This section reviews three methods for coding customers' verbal behavior: Observational monitoring, subjective encoding, and automatic encoding. Then it describes the basic analysis of the coded data.

Observational Monitoring

Observational monitoring (also termed verbal interaction coding), requires trained observers to record the occurrence of certain (predetermined) verbal behaviors *as the interview takes place* (Bercini, 1989). The observers use a predefined set of codes to record on a standardized form each occurrence of specific verbal behavior (e.g., customer asks a question). Researchers develop an observational coding scheme by tape recording some preliminary interviews and noting specific verbal behaviors of the respondent and the interviewer. For example, the coding scheme of Cannell and Robison (1971) for questionnaire pretesting includes codes for the following interviewer behaviors: Correctly asked question, incorrectly asked question, partially asked question, response alternatives missing, or question omitted by mistake (Cannell & Robison, 1971; Marquis, 1971). If multiple observers code the same interview, it is possible to assess the reliability of the coding scheme. Alternatively, multiple observers can be assigned different coding schemes to capture different aspects of respondents' verbal behavior.

Subjective Encoding

In subjective encoding, trained coders read a verbalization or segment and assign a predetermined code on the basis of their understanding of the text. Subjective coding schemes have been developed for a variety of purposes. For example, Bickart, Blair, Menon, and Sudman (1990) describe a very complete coding scheme that identifies the retrieval strategies respondents use to form responses to questions about the frequency of certain behaviors. Part of their coding scheme includes seven codes for counting strategies: General recall and count, counting with adjustment for uncertainty, counting with expression of uncertainty (no adjustment), counting by domain, counting by domain with adjustment for uncertainty, counting by domain with expression of uncertainty (no adjustment), and counting by observation.

Various interactive software packages are available to assist in the subjective encoding of verbal protocol data. *SHAPA* (Sanderson, James, & Seidler, 1989) is an interactive software in which the researcher controls the assignment of codes by using predicates and arguments to indicate the content of speech bursts or segments. Other software packages with similar capabilities include PAS-II (Waterman & Newell, 1971; 1973), SAPA (Bhaskar & Simon, 1977) and *Protocol Analyzer's Workbench* (Fisher, 1988). These packages are particularly useful when subjects are assigned a problem or a decision-making task for which predicates of goal attainment can be established and reliably applied. They also perform some of the mundane tasks involved in subjective encoding (e.g., reliability checks).

Automatic Encoding

Whereas observational monitoring and subjective encoding rely on trained coders for complete and accurate results, automatic encoding utilizes the computer. Automatic encoding is a form of content analysis (Stone, Dunphy, & Smith, 1966). In automatic encoding, a computer finds and marks all the words or word phrases predetermined by the researcher. Early examples of automatic encoding entailed the content analysis of textual data (e.g., Brunner, 1983). More recently, Bolton (1991) used a coding scheme based on Tourangeau's (1984, 1987; Tourangeau & Rasinski, 1988) model of how respondents answer an attitude question to automatically encode pretest interviews.

An example of a category used to identify respondents' comprehension difficulties is shown in Table 1. It lists key words and word strings that respondents used to ask for further clarification of a question. With respect to word strings, the researcher must carefully specify the sequence of the words. Consider how the following two segments would be coded using Table 1.

R Could you *say that again?*
R *Again*, I would *say that* the answer is yes.

Although the two segments contain the same key words, the first segment would be marked with a code to indicate comprehension difficulty, but the second segment would not (because the sequence is incorrect). The list of key words/strings for an automated coding category can be very lengthy.

Software packages with automatic encoding capabilities include *Systematic Analysis of Language Transcripts* (Miller & Chapman, 1986), the *Oxford Concordance Program* (Hockey & Marriott, 1982) and the *General Enquirer System* (Kelly & Stone, 1975). These software programs usually have some counting abilities. For example, they can count and sort the number of codes for any one speaker or topic.

Quantitative Analyses of the Coded Data

All three coding methods generate a sequence of codes that describe each interview. Human beings or computers can count the number of occurrences of each coding category for each interview (or interview section) discussed by each respondent. Hence, the data base contains frequencies of codes organized by interview sections and respondents. Quantitative analyses of these data are relatively straightforward. The frequencies can be divided by the total number of respondents to yield the percentage of respondents associated with a coding category for a particular interview section. These percentages can be compared across interview sections or coding categories.

The usefulness of such analyses depends on the research objectives of the study and the specification and measurement of theoretically justified interview characteristics. A simple example illustrates the basic approach.

TABLE 1
Sample Coding Category: Comprehension Difficulty[a]

Description: Any segment requesting additional interpretation or clarification of a survey question.

repeat	ask:that:question
interpret	say&one:more:time
define	one:more:time&the:question
explain	I'm:lost
comment	I:lost:you
how&rate	listen:to:that:question
do:you:mean	read:the:question
ask:that	state:the:question
asking	hear:the:question
I&misunderstand	listen:to:the:question
I&misunderstood	ask:the:question
I:don't:think:I:got:you	read:that:question
I:thought:you:said	state:that:question
say:that:again	hear:that:question
are&talking:about	
in:terms:of	
so:it:says	
in:other:words	
about:the:question	
that:word	
don't:understand	
listen:to&again	
are:you:talking	
hear:that:again	
is:that:what:you're:asking	
problem&question	
what&looking:for	
what:was&again	
what&mean	
I:take:it:that	
I'm:sorry	
pardon:me	
wondering	
confused	
you:mean	

[a] Colons and ampersands are used to characterize word strings. The colon indicates that word 1 has to immediately precede word 2. The ampersand signifies that word 1 and word 2 have to appear in the same segment, but that they may appear in any order within the segment. This usage of colons and ampersands is found in SALT.

Suppose a researcher conducts 25 personal interviews in which business customers are asked a series of open-ended questions. If 75% of respondents mention "reliability" in the opening 10 minutes of a depth interview and 25% of respondents mention "courtesy," the researcher might infer that reliability is a more important "top of mind" attribute to these customers.

Applications of observational monitoring and subjective encoding are based on relatively well-developed theories and procedures. In the survey research tradition, there are established observational monitoring guidelines (Bercini, 1989). In the consumer research tradition, there are well-known applications of subjective encoding to investigate customers' verbalized information-processing strategies under laboratory conditions (Douglas, Craig, & Faivre, 1981). There are also well-known applications of subjective encoding that investigate the content of advertisements (Belk & Pollay, 1985; Tse, Belk, & Zhou, 1989) and the content of customers' responses to open-ended survey questions (Collins & Kalton, 1980). For example, Morton-Williams and Young (1987) tape recorded doorstep introductions of survey interviewers and respondents' answers. They subjectively encoded these protocols, and conducted a content analysis that identified the processes/response leading to agreement/refusal to cooperate.

Weber (1985) argues that a content analysis of responses to open-ended questions yields more unobtrusive measures than the analysis of close-ended survey questions (i.e., with fixed alternative responses) because the act of measurement does not confound the data. His argument supports the notion that computer-assisted content analysis is appropriate for investigating customers' responses to open-ended questions and depth interviews. Researchers in human factors, clinical psychology, and psycholinguistics have conducted computer-assisted content analyses of depth interviews (e.g., Dobroth, Zeigler, & Karis, 1989). However, with the exception of the aforementioned study by Bolton (1991), market researchers have not automatically encoded the verbal data obtained during depth interviews with customers.

A major obstacle to automatically encoding depth interviews with customers is the difficulty in developing a useful coding scheme. The researcher should have some a priori knowledge or theory about the underlying characteristics of the interviews. Plus, he/she must have some understanding of how to identify key words or strings that can be used to create meaningful coding categories that correspond to the interview characteristics. This development process can be time consuming and labor intensive.

Computer-based coding methods can reduce the time and effort required to code and analyze depth interviews. The following sections describe how to develop an automated coding scheme, apply it to a set of interview transcripts, and analyze the resultant codes. Most software packages for the analysis of textual data are "special purpose" packages that are not intended to handle depth-interview transcripts. Hence, the discussion focuses on SALT, a flexible package for the analysis of conversational language. It describes the systems that support SALT and the preparation of depth-interview transcripts for input to SALT.

Developing an Automatic Coding Scheme

To develop a detailed automatic encoding scheme, the researcher must be very familiar with potential interview characteristics and the types of verbal and nonverbal cues that identify these characteristics. A coding category is a list of key words and strings (e.g., "don't understand") that correspond to an underlying characteristic of the interview (e.g., lack of respondent comprehension). The list should accurately reflect the language of the respondents that participated in the interviews. For example, local colloquialisms and business jargon may be appropriate. To refine the list of key words and strings, the proposed coding category is applied to a subsample of the interview transcripts. The researcher reviews the coded transcripts and identifies any segments that have been inappropriately coded (e.g., a key word does not correspond to the relevant interview characteristic). Then, the coding category is modified and the process is repeated until an acceptable list of key words and strings is developed.

Because an iterative process is used to develop and refine coding categories, the development of a complete automatic coding scheme can be time consuming. However, a user-friendly computer software package can dramatically decrease the time required to develop the coding scheme. For example, some software packages (such as SALT) allow the researcher to examine the words and word strings from a given category in context. This feature makes it is relatively straightforward to revise and test alternative coding schemes.

Computer-Assisted Encoding with SALT

SALT is a computer software program with an automated encoding operation (Miller & Chapman, 1986). It was developed at the language Analysis Lab of the University of Wisconsin to code children's speech. It can be used to scan an interview transcript, physically mark verbal (e.g., word or word string) or nonverbal (e.g., pauses) cues with distinct codes, and count these codes for further analyses. SALT places a code at the end of any segment containing the key words from a coding category. Multiple codes can be assigned to a single segment when it contains cues for more than one category, but segments are not assigned the same code twice if more than one word/string from the same coding category appears in one segment.

There are two versions of SALT: One for the PC and the other for a VAX/VMS system. Both versions have two modules: One module codes the data and the other module produces descriptive statistics. The PC version requires 256K of memory and DOS 2.0 or higher. The VAX/VMS version requires VMS 5.1 and 521 available blocks. Verbal data are entered into the program as text files in ASCII format. There seems to be no limit to the number or size of the coding categories. SALT can run interactively or in a batch mode. In interactive mode, the programming and coding time is dependent on the size of the coding problem (especially the number of coding categories) and the skill of the programmer/analyst. In batch mode, the programming and coding times are much shorter.

The programming and coding time can be illustrated as follows. Suppose a researcher has developed 12 coding categories (similar in size to the category shown in Table 1) to be used in analyzing depth interviews. Given some basic familiarity with SALT, it would take about 10 minutes to input the coding categories and define the domain of the coding task. If SALT is installed on a DEC MicroVax 3900, the researcher should be able to code 30 interviews of about 22 blocks each (i.e., about 12,000 bytes each) using the 12 categories in about two minutes. In our experience, there are file size restrictions when SALT runs on a personal computer. Fewer interviews can be coded in a single batch run and the coding time depends on the capacity of the computer. (As a rule of thumb, the upper limit is six interviews of about 10,000 bytes each.) Using an IBM-AT 286, six interviews of about 12,000 bytes can be coded in about 5 minutes. Hence, 30 interviews would take about 25 minutes, plus the additional time required to execute the SALT program five times.

Transcript Preparation

The SALT software has certain syntactical requirements for interview transcripts. Consequently, interviews must be transcribed according to a strict set of guidelines. SALT can be used to check each transcript for errors in transcription (e.g., illegal use of punctuation, no end-of-segment punctuation, etc). It gives error messages that contain the line numbers of any errors, an identification of the type of error, and the text containing the error. A transcript must be error free before any coding or analysis can be conducted. Since punctuation markers are assigned alternative meanings in SALT, they cannot be utilized in a normal manner. For example, colons (which are normally used to preface a list or to combine complex sentences) are used within SALT to mark pauses. Their inclusion in a transcript as a normal grammatical marker would skew any SALT analyses of that marker.

Segmentation and Precoding

With automatic encoding, the unit of analysis may be a complete response to a question, termed a "verbalization." Alternatively, the researcher may decompose the verbalization into smaller "segments" in order to study the components of respondents' cognitive processes/responses. A segment may be a single word or a string of words. A human coder segments a verbalization by listening to the audiotaped interviews for cues that indicate breaks between segments. These cues include short pauses, intonation, and syntactical markers (for complete phrases, clauses, or sentences). In SALT, the unit of analysis can be an entire verbalization or a segment. A new verbalization or segment is denoted by beginning a new line in the ASCII transcript file.

In the analysis of depth interviews, it is often useful to mark the incidence of certain verbal and nonverbal cues. For example, a researcher may wish to mark the occurrence of a pause, where one thought is dropped for another, and a variety of other cues (e.g., "um," "er") that are not evident from the text. The SALT software has a system of markers that enable this information to be recorded. Because the initial intent of the SALT software is to code and analyze children's speech, the developers of SALT have assigned meanings to the markers that may not be appropriate for the analysis of depth interviews. However, the markers can be assigned any meaning the researcher chooses.

For example, our study used certain markers to denote nonverbal cues in the text file in the following way.

The colon is used to indicate a pause. Generally speaking, a pause of greater than three seconds indicates some form of nonverbal information processing that should be coded. In SALT, this symbol (:) allows the actual time to be inserted [e.g., a six-second pause would be recorded as (00:06)]. Often, the indication that a pause exists is sufficient.

> The greater than symbol is used to indicate that a respondent has suddenly shifted his/her thoughts, termed a broken utterance.

() Enclosed parentheses (referred to by SALT as a "maze") can be placed around any key word or words that a researcher wishes to count. In this study, parentheses were placed around all unintelligible utterances ("uh," "er," and "um") because they may indicate incompletely verbalized information processing.

When preparing group interview data for analysis, it may be useful to mark the text to indicate different speakers or participants. When preparing group or personal interview data, it may also be useful to distinguish between respondents' answers to certain questions or topics. In SALT, the researcher can specify the speaker and/or topic by using a letter of the alphabet as an indicator. The number of speakers and/or topics that can be examined within one transcript is limited only by the number of letters in the alphabet.

The following example depicts what a transcribed and precoded interview would look like. In this example, the interviewer's question is identified by the letter I. The respondent's answer to question one is identified by the letter A at the beginning of each segment and his/her answer to question two by the letter B at the beginning of each segment. These letters are initialized in the first line of the transcript by the dollar sign.

$	Interviewer, A question one B question two	
+	John Doe, questionnaire #1	
I	When was the last time you ate in the cafeteria?	
A	Well, :.	Question #1
A	I'm trying to think of the last time.>	
A	Wait!	
A	it was yesterday.	
I	What did you have for lunch?	
B	that's easy.	
B	I had a sandwich for lunch.	Question #2
B	(um).	
B	and a soda.	

CRITERIA FOR ASSESSING ALTERNATIVE CODING METHODS

This section discusses the strengths and weaknesses of observational monitoring, subjective encoding, and automatic encoding for the quantitative analysis of verbal data. The three coding methods are evaluated according to the following criteria: ability to test propositions, flexibility, and efficiency (Becker, Gordon, & LeBailley, 1984).

Ability to Test Propositions

The researcher's choice of a coding method will primarily depend on whether it will facilitate testing propositions about interview characteristics. Observational monitoring codes tend to identify general interview characteristics. A coder is usually only capable of handling about seven codes. In contrast, subjective and automatic coding schemes use many codes to make fine discriminations about interview characteristics. For example, an observational monitoring form might include a code for whether or not the customer was satisfied. However, a subjective or automatic encoding scheme might code the underlying attributes or causes of customer satisfaction using four or five codes.

Subjective and automatic encoding can be used with segmented and precoded interview transcripts. Thus, these methods can examine highly specific interview characteristics, such as nonverbal cues. A sudden shift in topic, termed a *broken utterance*, may indicate a change from top-of-mind to in-depth processing. A *pause* (e.g., greater than three seconds) may indicate retrieval or evaluation problems, or other information-processing problems. *Unintelligible utterances* may indicate nonverbalized processes that may be associated with information processing problems. Hence, the investigator may decide to precode the interview transcripts to record these and other nonverbal cues. Because there is virtually no limit to the number of coding categories, the researcher can make subtle distinctions in the coding scheme.

If the investigator is interested in *general* interview characteristics, observational monitoring may be sufficient. If more detailed information is required about customers' responses and processes, researchers should consider subjective or automatic encoding of verbal protocols. Automatic encoding is somewhat more appropriate when the research focus is the *content* of customers' responses, whereas subjective encoding is somewhat more appropriate when the research focus is customers' cognitive *processes*.

Market researchers are likely to be interested in both the *content* of customers' responses and the *processes* that lead to these responses. For example, Goldman and McDonald (1987) suggest that the following should be considered in the analysis of depth interviews: The order in which respondents raise issues, the time they spend on them, the intensity of people's reactions, and the reasons they give for acceptance or rejection of new concepts. Hence, the selection of a coding method ultimately depends on the nature of the responses and processes to be measured.

Flexibility

In general, subjective encoding is more flexible than automatic encoding—albeit slower. However, the use of human coders introduces the potential for bias. Coders may be influenced by prior knowledge of the hypothesis; they may make assumptions about the respondent's thought processes; or they may be affected by the sequencing of verbalizations. For this reason, researchers have advocated that observers/coders should not be informed about the research hypotheses, and that verbalizations should be coded in random order to prevent inferences from context (Ericsson & Simon, 1984). Computer-assisted coding can address the latter problem. For example, SHAPA can present segments randomly to the coder. Multiple coders are typically employed to code the same categories, so that reliability of the results can be assessed. However, Hughes and

Garrett (1990) argue that the current methods used to determine intercoder reliability are not adequate because they do not account for chance agreement, systematic coding errors and intraclass correlations.

Automatic encoding requires all of the underlying vocabulary and inference rules of the coding scheme to be explicitly defined so that a computer can identify and mark specific verbal and nonverbal cues in the interview transcripts. The coding scheme is always applied consistently and its robustness to changes in vocabulary and rules can be tested (Ericsson & Simon, 1984). Automatic encoding software can very quickly depict the content of an interview. For example, the software can calculate the incidence of references to products/service/attributes and display these key words in context. More importantly, it is relatively straightforward to revise and test alternative coding schemes.

Efficiency

Table 2 summarizes the relative time and cost requirements of observational monitoring, subjective coding, and automatic coding. It examines the following activities: transcription and precoding, development

TABLE 2
Comparative Benefits of Three Coding methodologies

	Observational Monitoring	Subjective Encoding	Automatic Encoding
Transcription and precoding	No	Yes	Yes
Time to develop coding scheme	Medium	High	High
Development of categories	' ow (subjectively developed)	High (subjectively developed	High (subjectively developed but iterative testing possible)
Coding time	Low	High	Medium (includes programming time)
Level of Expertise Required of Coder	Moderate/High	High	None
Reliability	Medium (checks required)	Medium (checks required)	High
Cost	Low	High	Medium

of coding scheme, coding time, level of expertise required of coder, reliability, and cost.

Marketers will be particularly interested in the time required for each method. Because exploratory studies eliciting verbal protocols usually contain small sample sizes, the time differences tend to be rather small. Consider the scenario in which 15 1-hour personal interviews are conducted. Observational monitoring would require little time because the interviews are coded as they take place and the analysis of the codes is relatively quick. After the interviews have been administered and transcribed, a researcher might spend about 15 hours reviewing the transcripts in preparation for a conventional qualitative analysis. The same 15 1-hour interviews could be segmented and precoded in preparation for subjective or automatic encoding in approximately 20 hours. However, additional time is required to create and apply a subjective or automatic coding scheme.

A researcher that has never used automated encoding will require considerable time to develop a simple coding scheme. For example, the time required to develop a simple automatic coding scheme (say, six categories) could be about 40–50 hours. Subjective encoding schemes require about the same development time, but much more time for coding, reliability checks, and so forth. In contrast, although it is laborious to develop an automated coding scheme, the software codes the transcripts very quickly.

THE FIELD STUDY

An exploratory study was conducted to assess the advantages and disadvantages of computer-assisted coding and quantitative analysis of depth interviews. It compared observational monitoring and automatic encoding schemes. These two methods highlight the difference between human and computer-assisted coding methods. Observational monitoring is a form of subjective encoding that occurs in real time and (consequently) is less detailed; whereas automatic encoding works with transcripts and is extremely detailed. Subjective encoding can be considered the "middle ground" between these two methods. It is not examined here for two reasons. First, Morton-Williams and Young (1987) have documented subjective encoding with personal interview data. Second, subjective encoding is highly labor intensive and (consequently) costly.

In this study, 30 residential telephone company customers were recruited to pretest new telecommunications survey questions. The interviews were conducted at a focus group facility in Tampa, Florida in July of 1989. Prior to the interview, customers were informed that their responses were being recorded for research purposes. The respondent's answers were monitored from behind a one-way mirror, audiotaped, and videotaped.

The respondents were asked a variety of questions, including the four open-ended questions shown in Table 3. Because information about telephone service experiences are likely to be stored in long-term memory, respondents are likely to engage in in-depth processing. Hence, detailed responses were obtained by asking customers to think aloud while forming answers to the questions.

TABLE 3
Comparison of Subjective and Automatic Encoding[a]

Questions	Automatic Encoding (AE) Questions	Repeat[b]	Pauses	Observational Monitoring (OM) Comprehension	Retrieval
1. Compared to 12 months ago, would you say that the overall quality of telephone service is better/same/worse? (n = 30)	12.7%	6.7%	36.7%	13.3%	20.0%
	3.3	*0.3*	*4.2*		
2. If you were moving to a new area and could choose your own telephone company, how likely would you be to continue using GTE? (n = 24)	25.0	8.3	16.7	16.7	12.5
	3.2	*1.0*	*1.9*		
3. Considering all aspects of the service provided by GTE, would you rate overall quality of services provided to you as poor, below average, average, good, or excellent? (n = 21)	42.9	19.0	42.9	28.6	19.0
	7.6	*2.3*	*5.9*		
4. How would you rate the reliability of the telephone service GTE is providing? (n = 22)	27.3	18.2	22.7	31.3	9.1
	6.6	*1.8*	*3.1*		

[a] Italics denote intensity measures.
[b] This category appears in its entirety in Table 1.

The interviewers were trained in the elicitation of concurrent verbal protocols. They read task instructions to the respondents that explained the think-aloud processes, asked the questions, and back channeled (when appropriate). Back channeling involves giving positive feedback to continue speaking without encouraging or leading respondents to generate a particular response.

The telephone company was interested in testing a variety of propositions about the interview. This paper focuses on a limited number of interview characteristics to demonstrate the quantitative analysis of depth interview data. The observational monitoring and automatic encoding schemes were designed to measure cognitive processes, rather than measuring cognitive content. Both coding schemes measure comparable cognitive processes. Specifically, we examine information processing difficulties that the customers experienced during the depth interviews. The remainder of this section will describe how the data were coded, present quantitative data obtained from both coding methods, and compare the results.

Observational Monitoring Scheme

To develop the observational coding scheme, some practice interviews were conducted over the telephone. These interviews were based on the same questionnaire as the (subsequent) personal interviews. By listening to the telephone interviews, the investigators developed a six-item coding scheme for observational monitoring (OM). Four codes were developed to identify problems the respondent had replying to the question. These codes indicated different types of information processing problems: Comprehension of the question, retrieval of information from memory, evaluation/integration of information, and use of the rating scale. Two codes were developed to identify difficulties that the interviewer had in asking a question.

This paper focuses on two of the six codes: Comprehension and retrieval. The "comprehension" category was coded if the respondent asked to hear the question again, asked for clarification, repeated the question aloud to him/herself, or paraphrased the question. The "retrieval" category was coded if the respondent paused longer than three seconds, changed his or her answer in midstream, or expressed some indication of difficulty (e.g., um, er, I don't know, that's tough).

Three different interviewers were used so that the interviews could be completed within a three-day period. Because interviews could occur simultaneously, a single observer could not code all interviews as they occurred. Thus, 14 of the 30 interviews were coded by listening to audiotapes while reading the transcripts. Ultimately, an ASCII file was created with a record for each customer's response to each question. Each record described whether or not the respondent experienced difficulty answering the question by coding each category with a one (yes) or zero (no). Summary statistics were calculated in the following way. We counted the number of respondents that experienced a particular category of difficulty in answering a particular question by summing across respondents. Dividing this number by the total number of respondents, we obtained the percentage of respondents that experienced a particular category of difficulty for a question. The right side of Table 3 shows the results for the comprehension and retrieval coding categories. For example, the OM results indicate that 13.3% of the respondents experienced comprehension problems when they answered the first question.

Automated Encoding Scheme

The interviews were transcribed and the verbal protocols were segmented. The average customer spoke 10 segments in response to each question. However, some customers were more verbose than others. The number of segments spoken in response to a question ranged from 1 to 38 segments. Because certain nonverbal cues cannot be determined by reading the transcripts alone, the transcripts were precoded to indicate broken utterances, pauses, and unintelligible utterances. The interview transcripts were automatically encoded (AE) in the following way. Thirty verbal and nonverbal coding categories were developed based on a preliminary examination of the transcripts. These categories reflected respondents' comprehension, retrieval evaluation, and response processes in greater detail than the OM codes.

This study focuses on three of the AE coding categories that correspond to the two OM coding categories described previously. Two AE codes measured comprehension difficulty: questions by the respondent and requests for the interviewer to repeat the question. One AE code measured general processing difficulty (including retrieval difficulty), namely, pauses of greater than three seconds by the respondent. There were many additional categories that measured cognitive processes.

There were also coding categories to measure different aspects of the *content* of the responses. For example, one content category concerned segments that referred to low frequency events (e.g., "that never happens," "seldom," "rarely," etc.) Analyses revealed a high incidence of segments about low-frequency events in response to questions about service orders/changes in the past 30 days. Because customers consider service orders/changes to be a low-frequency event, we might expect a variety of heuristics and biases leading to inaccurate responses.

Analysis

SALT (installed on an IBM-PC 286 machine) was used to code the data. By noting whether or not a customer's entire response to a question was assigned a particular code, it was possible to develop a data base of records that described the presence or absence of each of the three AE codes for each customer's response to each question. Each record described whether or not the respondent experienced difficulty answering the question by coding each category with a one (yes) or zero (no). Summary statistics were calculated in the following way. We counted the number of respondents that experienced a particular category of difficulty in answering a particular question by summing across respondents. Dividing this number by the total number of respondents, we obtained the percentage of respondents that experienced a particular category of difficulty for a question. The roman type on the left side of Table 3 shows the results for the two comprehension categories and the pause category. For example, the AE results indicate that 12.7% of the respondents asked a question when they answered the first question.

SALT was also used to count the *number* of occurrences of each AE category for each question for each respondent. This statistic was divided by the total number of segments spoken by the respondent to obtain a measure of the *intensity* of the processing problem. This division creates a statistic that lies between zero and one. For example, a respondent might pause three times during 10 segments uttered in response to a particular question, yielding an intensity measure of 0.30. Each statistic (i.e., both the number of occurrences and the intensity) was summed across respondents and divided by the total number of respondents to obtain an average. The results are displayed in italics in the left side of Table 3.

RESULTS

The percentages displayed in the upper half of Table 3 show some agreement between the OM and AE codes. Pearson product-moment correlations are displayed in Table 4. As expected, the two AE comprehension difficulty codes (i.e., questions and repeat requests) are correlated with the OM comprehension code ($p < 0.01$). The AE code for pauses is correlated with the OM retrieval category ($p < 0.001$). These results show that the OM and AE coding schemes code the interviews in a similar fashion.

The AE code for questions (usually indicative of comprehension difficulties) is correlated with the OM code for retrieval difficulties ($p < 0.01$). This result seems to be an artifact of the OM retrieval category

TABLE 4
Correlations among OM and AE Measures

	OM—Comprehension	OM—Retrieval
AE—Questions	0.72*	0.28**
AE—Repeat	0.27**	0.10
AE—Pause	0.11	0.34*

Note. The data have been pooled across all four questions. Therefore, the correlations in this table are based on 97 pairs of scores.
*$p < 0.001$ (one-tailed). **$p < 0.01$ (one tailed).

definition. The observer was instructed to code a retrieval difficulty when the respondent exhibited certain verbal behaviors (e.g., "um", "er"). Unfortunately, these verbal behaviors could be indicative of other kinds of information-processing difficulty, as well as retrieval difficulty. This result illustrates how OM coding categories tend to be more broadly defined than AE coding categories.

Examining Table 3, both OM and AE methods indicate that a high percentage of respondents experienced comprehension difficulties with the third question. OM results indicate that 28.6% of respondents experienced comprehension problems and AE results indicate that 42.9% of respondents asked questions. However, the additional coding category supplied by automatic encoding gives the researcher some insight into the type of comprehension difficulties customers experienced. Note that the AE statistics (see Table 3) reveal that there is a relatively large percentage of customers that ask to hear the question again (19.0%). This result may be due to the length (i.e., 27 words) and complexity (i.e., "considering all aspects of the service provided by GTE") of question 3.

Consider the AE statistics (shown in italics) that describe the intensity of customers' processing problems. For question 3, an average of 7.6% of the segments were questions and an average of 5.9% of the segments were pauses. Because respondents utter about 10 segments in response to a question, these percentages imply that some respondents are asking more than one question or pausing more than once. Thus, the AE results for question 3 indicate that individual respondents' comprehension difficulties are relatively severe. The AE results also show that respondents had problems understanding question 4. An average of 6.6% of the segments were questions and an average of 3.1% of the segments were pauses (see Table 3).

The OM and AE results for the first question (see Table 3) show a relatively high incidence of retrieval or other processing difficulties. The AE results indicate that respondents hear the question correctly (because they rarely ask the interviewer to repeat it), but some respondents (12.7%)

ask questions and many respondents (36.7%) pause. One possible reason is that the reference period of 12 months may be too long. Comparing the AE results across all four questions, customers seem to have experienced the least information-processing problems with question 2.

DISCUSSION

Our exploratory study shows that quantitative analyses of observational monitoring codes provide a broad overview of customers' responses. Quantitative analyses of the coded data are obtained without spending substantial amounts of time or money. This is because the interviews are coded in real time and the codes are analyzed with conventional software. Quantitative analyses of observational monitoring codes are a useful adjunct to qualitative analyses of the interviews. Alternatively, they can be used to screen the interviews prior to more detailed analyses. For example, our study used observational monitoring to code customers' responses to all questions in the depth interviews. The results indicated that respondents experienced significant difficulties with about one-third of the interview questions. Consequently, only these "problem questions" were segmented and automatically encoded. Utilizing observational monitoring in this manner decreased the total cost of the analyses and the amount of time spent on automatic encoding.

In our exploratory study, we used computer-assisted coding for both qualitative analyses (i.e., to develop the coding scheme) and quantitative analyses. Qualitative reports can be generated by searching for key words in context, measuring the duration of interview segments (by inserting appropriate markers) and so forth. Concurrent with such qualitative analyses, the researcher can develop coding categories suitable for quantitative analyses. Unlike subjective encoding, it is easy to test the robustness of results to changes in the vocabulary and rules that define the coding categories. Consequently, an accurate coding scheme that makes subtle distinctions among a variety of verbal cues can be developed in a relatively short time. Subsequently, the software can count the coding categories and the counts can be used in a variety of quantitative analyses. A key benefit of automatic encoding packages such as SALT, is that they perform *detailed* quantitative analyses in an efficient manner.

In our experience, SALT worked well on both an IBM-PC and on a VAX/VMS system. Because SALT will accept any ASCII file, we were able to use transcripts prepared by a commercial transcription service rather than transcripts prepared by specially trained research assistants. The software handled multiple transcripts easily. In this study, automatic encoding and quantitative analyses of our 15 1-hour interviews required about twice as much time as qualitative analyses of the same interviews. Thus, quantitative information based on the codes was obtained with an additional 15 hours of work. Since the output files are in ASCII format, they can be imported into conventional word processors.

There are two main disadvantages associated with using SALT to analyze depth interview data. Because it was originally developed to analyze children's speech, the SALT manual does not discuss how the software can be used to examine other research objectives. The researcher must adapt the program functions accordingly. For example, this study used alphabetic markers intended to identify different speakers to identify a single speaker's responses to different questions. The second disadvantage associated with SALT is that it can only produce counts of coding categories for the entire interview transcript. Therefore, if the researcher wishes to obtain separate counts for different excerpts of the interview (e.g., different questions or topics), each excerpt must be analyzed in a separate SALT run. For example, to analyze a customer's responses to 12 questions (identified by alphabetic markers indicating different speakers), SALT must be executed 12 times.

CONCLUDING REMARKS

This paper demonstrates the benefits of computer-assisted analyses of coded interview data. Quantitative analyses of observational monitoring codes can provide a broad description of interview characteristics (Bercini, 1989). Based on our experience, quantitative analyses of observational monitoring codes require about the same (or slightly less) time than qualitative analyses. The procedure is relatively low cost because it simply requires that one or more observers code the interviews (or recordings of the interviews) in real time.

When the research objectives dictate a detailed analysis of interview characteristics, the interviews should be subjectively or automatically encoded for quantitative analysis. The selection of subject or automatic encoding will depend on the interview characteristics of interest, plus the availability of resources (e.g., trained coders, software).

Software packages with automatic encoding capabilities include *Systematic Analysis of Language Transcripts* (Miller & Chapman, 1986), the *Oxford Concordance Program* (Hockey & Marriott, 1982) and the *General Enquirer System* (Kelly & Stone, 1975). These packages usually have qualitative analysis features, as well as quantitative analysis features. They can handle large numbers of coding categories and make subtle distinctions among interview characteristics. Hence, they can provide a rich data base to test propositions. For example, we used SALT to automatically encode 30 different categories of cognitive processes and responses spoken by customers. Other coding schemes are possible. For example, a researcher might focus on the content of a group interview, such as the trial use of a new product or customer satisfaction with an existing product.

Because software packages with automatic encoding capabilities have not been designed specifically for the analysis of depth interviews, they can be somewhat inflexible. Most packages are primarily designed to count codes. Nevertheless, they are very efficient, so that the incremental benefits of automatic encoding schemes are likely to outweigh the costs. Because studies using depth interviews usually involve small sample sizes, the additional time required to conduct a quantitative analysis tends to be rather small. As the number (or length) of the interviews increases, automatic encoding becomes even more attractive because the fixed cost to develop the coding scheme is spread over more interviews. Also, as the investigator gains experience, automatic encoding becomes increasingly time and cost effective.

Although depth interviews are frequently utilized in marketing studies, the analysis of the verbal data obtained through these interviews has not been subject to rigorous quantitative analyses (Greenbaum, 1987; Wells, 1974; Wolfe, 1990). This study shows that rigorous quantitative analyses of depth interviews are feasible using computer-based methods. Such analyses are an important complement to conventional qualitative analysis. Weber (1985) argues that content analyses may yield "better" measures than analyses of closed-ended survey questions. If so, additional research will be necessary to explore how content analysis should be applied to human speech. Such research should include theoretical work that investigates customers' cognitive processes and responses to depth interview questions and empirical work that develops and refines appropriate coding schemes for different interview topics and characteristics. In addition, there is a need for more flexible software packages designed specifically for the quantitative analysis of depth interviews. In the future, computer-based methods of quantitatively analyzing depth inerviews may become an integral part of the market researcher's repertoire as qualitative analyses.

REFERENCES

Becker, H. S., Gordon, A. C., & LeBailley, R. K. (1984). Field work with the computer: Criteria for assessing systems. *Qualitative Sociology, 7*, 16–33.

Bercini, D. (1989, May) Observation and monitoring of interviews. *Quirk's Marketing Research Review*, pp. 21–23, 81–84.

Belk, R. W., & Pollay, R. W. (1985). Images of ourselves: The good life in twentieth century advertising. *Journal of Consumer Research, 11*, 887–897.

Bhaskar, R., & Simon, H. A. (1977). Problem-solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science, 1*, 193–215.

Bickart, B. A., Blair, J., Menon, G., & Sudman, S. (1990). Cognitive aspects of proxy reporting of behavior. In M. Goldberg, G. Gorn, & R. W. Pollay (Eds.), *Advances In Consumer Research, 17*, 198–206.

Bolton, R. N. (1991). An exploratory investigation of verbal protocol analysis applied to questionnaire pretesting. *Advances In Consumer Research, 18*, 558–565.
Brunner, R. D. (1983). *Forecasting growth ideologies (Discussion Paper No. 7)*. Boulder: University of Colorado Center for Public Policy Research.
Cannell, C. F. & Robison, E. (1971). *Analysis of individual questions* (Working Papers on Survey Research In Poverty Areas). Ann Arbor: University of Michigan, Survey Research Center.
Collins, M., & Kalton, G. (1980). Coding verbatim answers to open questions. *Journal of the Market Research Society, 22*, 239–247.
Dobroth, K. M., Ziegler, B. L., & Karis, D. (1989). Future directions for audio interface research: Characteristics of human-human order-entry conversations. In *Proceedings of the Annual Meeting of the American Voice I/O Society* (pp. 277–282).
Douglas, S. P., Craig, C. S., & Faivre, J. (1981). Protocols in Consumer Research: Problems, Methods and Uses. *Research in Marketing, 5*, 29–58.
Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.
Fisher, C. (1988). Advancing the study of programming with computer-aided protocol analysis. In G. Olson, E. Soloway, & S. Sheppard (Eds.), *Empirical studies of programmers*, 1987 Workshop. Norwood, NJ: Ablex Publishing Corporation.
Goldman, A. E., & McDonald, S. S. (1987). *The group depth interview: Principles and practice.* Englewood Cliffs, NJ: Prentice-Hall, Inc.
Greenbaum, T. L. (1987). *The practical handbook and guide to focus group research.* Lexington, MA: D. C. Heath & Company.
Hockey, S., & Marriott, I. (1982). *Oxford concordance program version 1.0 users' manual.* Oxford: Oxford University Computing Service.
Hughes, M. A., & Garrett, D. E. (1990). Intercoder reliability estimation approaches in marketing: A generalizability theory framework for quantitative data. *Journal of Marketing Research, 17*, 185–195.
Kelly, E. F., & Stone, P. J. (1975). *Computer recognition of English word senses.* Amsterdam: North Holland.
Marquis, K. H. (1971). *Purpose and procedures of the tape recording analysis* (Working Papers on Survey Research In Poverty Areas). Ann Arbor: Survey Research Center, University of Michigan.
Miller, J. & Chapman, R. (1986). *Systematic analysis of language transcripts (SALT).* Unpublished manuscript, University of Wisconsin.
Molina, M. (1991). *CORE AI.* Paper presented at the International workshop on quality of telecommunications services, networks and products.
Morton-Williams, J., & Young, P. (1987). Obtaining the survey interview: An analyses of tape recorded doorstep introductions. *Journal of the Market Research Society, 29*, 35–54.
Sanderson, P. M., James, J. M., & Seidler, K. S. (1989). SHAPA: an interactive software environment for protocol analysis. *Ergonomics, 32*, 1271–1302.
Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis.* Cambridge, MA: MIT.
Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabins, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). Washington, DC: Naval Academy Press.
Tourangeau, R. (1987). Attitude Measurement: A Cognitive Perspective. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 149–162). New York: Springer-Verlag.
Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*, 299–314.
Tse, D. K., Belk, R. W., & Zhou, N. (1989). Becoming a consumer society: A longitudinal and cross-cultural content analysis of print ads from Hong Kong, the People's Republic of China, and Taiwan. *Journal of Consumer Research, 15*, 457–472.
Waterman, D. A., & Newell, A. (1971). Protocol analysis as a task for artificial intelligence. *Artificial Intelligence, 2*, 285–318.
Waterman, D. A. (1973). *PAS-II: An interactive task-free version of an automated protocol*

analysis system. Pittsburgh: Department of Computer Science, Carnegie-Mellon University.

Weber, R. P. (1985). *Basic content analysis* (Sage University Series on Quantitative Applications In the Social Sciences No. 07-049). Beverly Hills: Sage Publications.

Wells, W. D. (1974). Group interviewing. In Robert Ferber (Ed.), *Handbook of marketing research*, New York: McGraw-Hill.

Wolfe, T. J. (1990). Using AGENDA for qualitative data analysis. *Journal of Business and Psychology*, *5*, 261–274.

Ruth N. Bolton is Principal Member of the Technical Staff and Tina M. Bronkhorst is a Research Assistant in the Service Applications and Technology Laboratory of GTE Laboratories Incorporated.