

# Using Computer Audio Recorded Interviewing (CARI)

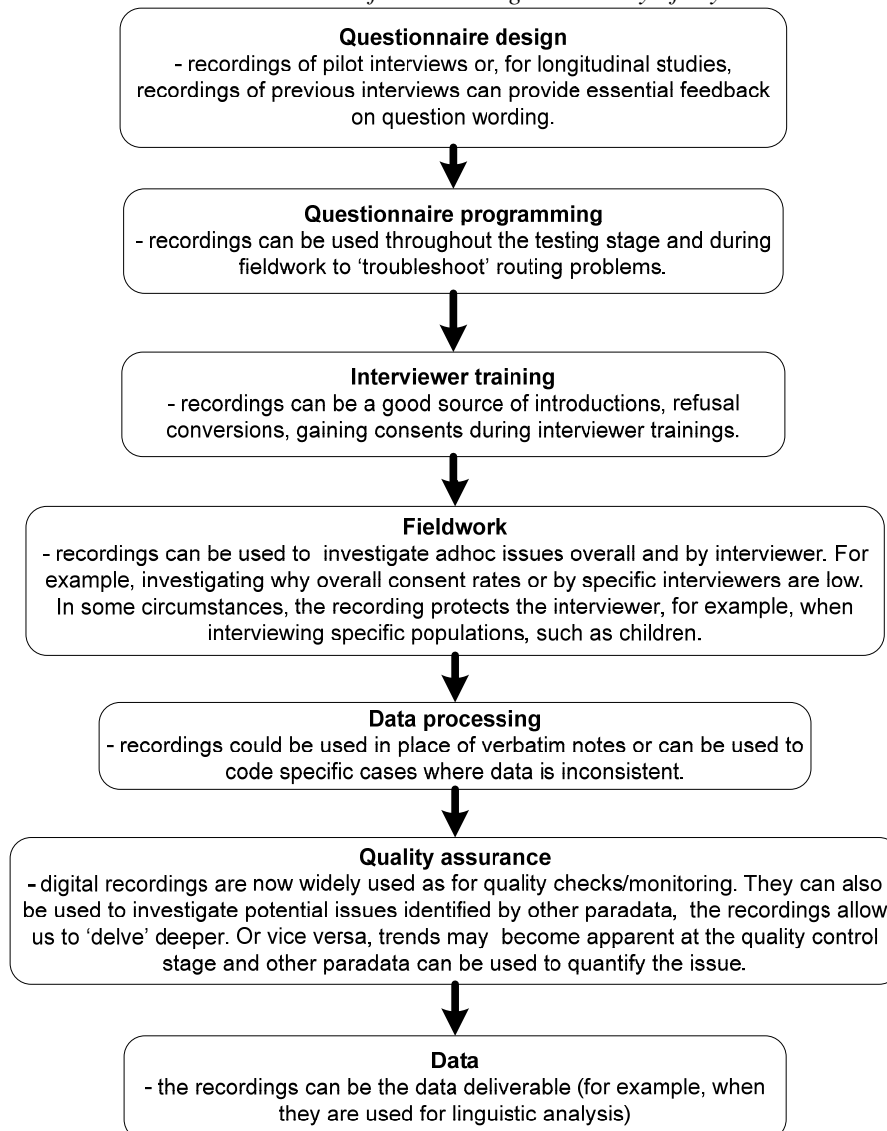
Rebecca Gatward, Gina-Qian Cheung and Patty Maher,  
Survey Research Center, University of Michigan

## 1 Introduction

Computer Audio Recorded Interviewing (CARI) is now among the standard set of tools used during survey data collection. Prior to digital recording monitoring systems in centralized telephone interviewing facilities provided valuable and instant feedback on interviewer performance. This tool also provided researchers with the ability to collect timely feedback on the design of questions. However, until CARI, there was no equivalent systematic system for face to face or decentralized telephone interviewing – other than using portable recording devices.

CARI is now used to provide a valuable insight into factors that affect data quality throughout the survey lifecycle, from question design to responding to queries following data delivery. The following illustration describes some of the ways CARI is used during the survey process.

*Illustration 1. Use of CARI through the survey life cycle*



Although CARI has clearly developed into an important (and almost standard) tool in the survey data collection process, little has been shared across the industry about best practices or challenges implementing this technology and utilizing the resulting data. University of Michigan, Survey Research Center, Survey Research Operations organized a two day meeting to bring together five survey research organizations. The key objective of the meeting was to share best practices and challenges encountered whilst implementing Computer Audio-Recorded Interviews (CARI) and utilizing the data. The event was held on 14-15 November 2011. The following organizations participated:

- Research Triangle Institute (RTI),
- Statistics Canada,
- Statistics Netherlands,
- University of Michigan-Survey Research Center
- US Census Bureau and
- Westat.

The purpose of this paper is to share details of the information and discussion that was exchanged during the two day meeting. All omissions or interpretations are the responsibility of the authors. The paper should provide a useful status report of how CARI is currently being used within the participating organizations and an insight into some planned developments and new ways they would like to use CARI in the future.

## **2.1 Methods: Operations and Sampling**

The focus of the first part of the two day meeting was on sharing experiences of administering and implementing CARI within each organization. To facilitate this, participating organizations were asked to prepare presentations around two sets of questions. The first set is based on administration and are listed below;

### **Sharing survey administration experiences with this technology –**

- Recording across samples. What percentage of sample is recorded? Does this percentage change during the production period? What are the reasons for the change?
- Recording across data collectors. How is this controlled? Are there attributes of the interviewers that are taken into account when recording? Does this change during the production period? What are some of the reasons for the change?
- Selection. What are the recording variables? Full interview, selected questions, or random questions?
- Reviewing the recording. How are the recorded interviews reviewed? Who does this and do the reviewers go through the full recorded interviews or segments?
- Evaluations. What are the main criteria for reviewing the recording?
- Feedback. How are the results of the review communicated to interviewers and researchers?
- Implementation. How does this technology affect your organization's quality of data collection? Has there been feedback from users about the technology implementation? (i.e., does recording slow down the survey instrument during the interview?)

Common themes emerged as each organization shared details around the administration of CARI and specifically, CARI as a tool to facilitate quality control. Some of the key themes were discussed in more detail during the meeting and will be covered later in this paper, the following is a brief summary of all the themes.

- **Rate of recording vs. rate of evaluation**

A key decision is setting the rate of recording i.e. how many interviews are recorded (n=completes) and then of these, what proportion are evaluated (% of completes). Factors that influence these decisions included practical concerns, such as, the size of the audio files and possible effect on transmission or interviewing speed. The rate of recording varied across

organizations; ranging from, recording all interviews and evaluating a proportion of these recordings, record at a predetermined rate and then evaluating a percentage of these recordings to only recording what will be evaluated.

Frequency of interviewer monitoring/evaluation also varied by organization – this variation seemed to be based on how field interviewers were organized – if they work across different projects, then the cycle of evaluations is based on time – for example, they are evaluated so many times a month. If they typically work on one project then evaluations are based around the cycle of a project – for example, the first two interviews at the start of fieldwork and then a set number or interval (unless there are specific quality concerns).

- **Selection of cases to be evaluated**

Following on from determining the rate of recording, the selection of cases to be evaluated is also a key decision. The typical rate across organizations represented at the meeting was ten per cent, however, criteria used to select the cases and how/whether this rate varies by interviewer was not consistent. Issues around recording and evaluation across samples and data collectors were discussed in some detail during the meeting. Key points from the discussion are summarized below:

- Evaluation rates can vary by type of project or by type of interview. Some clients may also request a higher level of monitoring on some projects. Generally, 10 per cent was the norm.
- Selection of cases for evaluation can also be based on answers to survey questions (rural cases, high/low income, and other demographic variables). Although there was strong support for selection of cases based on questionnaire content for evaluating the questionnaire, whereas selection of cases for evaluation should be focus on interviewer attributes or sample distribution.
- Is it sensible to preselect cases to be evaluated at the interviewer level based on workload? For example if the rate of evaluation is set at 15 per cent of assigned cases – which means you actually get less than 10 per cent because they do not result in completed interviews. Should the sample be selected by recording 15 per cent and then evaluate them all?
- An alternative is to record a set number of an interviewer’s initial interviews completed and then evaluate a proportion of subsequent cases completed. Using this method results in all interviewers having the same number of cases evaluated early on in the field period.
- Some organizations also set a target of a certain number of interviews per interviewer to be evaluated per week.
- There was some discussion about adjusting evaluation rates by experience of interviewer (for example, new to survey organization, experienced interviewer or new to a project). It was felt that this is useful to an extent, but some methods literature shows experience is not always correlated with better quality and fewer errors.
- Selecting a small percentage of non-interviews who were then followed up by telephone was a procedure followed by one organization – others thought this was good practice.
- It is possible to identify those interviewers who are not following procedures to save time by looking at data. For example, higher levels of missing respondent telephone numbers, missing data, and higher refusal rates for CARI, item level timings – faster timings could indicate falsification or skipping more complicated or lengthy sections of the interview.

- **Selection of content to be recorded**

The selection of variables to be recorded also varies by organization – the various methods are listed below;

- Questions or groups of up to three questions are selected from across the entire interview – totaling between around 8 to 12 questions. If more than 12 are recorded then probabilities are set so only 8-10 of the items are scored.
- Pre-defined set of questions are recorded, the questions are selected based on key analytic variables, such as, new items, challenging questions or sets of questions, key routing variables, consent questions and incentive payment scripts.
- Variables are chosen to be recorded but specific variables are used for each component of an evaluation, for example, different variables are used for quality assurance than for behavior coding.
- One organization keeps track of which variables are being evaluated, from this it is possible to identify those variables which are being missed.

Again, the discussion around how we select what is recorded generated some topics for further discussion by the group after the meeting. These include;

- Stratifying for selection of segments – how can paradata inform the selection of segments to be recorded?
- Best practices for selection of items (including, electing across the entire interview, optimum length of recording).
- Difference in CATI vs. CAPI.

- **Evaluation Criteria**

As might be expected, each organization uses a standard set of evaluation criteria, which are, broadly, consistent. Interviewers are evaluated according to adherence to standard interviewing techniques (skills and behavior) and knowledge and implementation of survey specific procedures. Generally, interviewer errors are rated according to the type and severity of an error – rating errors as critical, major or minor is common. An overall rating is then based on the number and severity of errors. Feedback is provided to interviewers via a standard report – and shared via an interviewer supervisor. It was noted that, in some organizations, aggregate performance data are produced by project, supervisor and across projects.

The following two ideas emerged during the discussion on evaluation criteria which were noted for future discussion (post-meeting) by the group.

- Could a common framework for evaluation be conceived?
- How to best use paradata attached to CARI files?

- **Training – monitors and interviewers to collect and evaluate recordings**

Training provided to interviewers in the quality control process ranged in detail. The lowest level of training was simply informing interviewers that they are required to ask the respondents the consent to record question at the beginning of an interview. More detailed training included interviewers being trained on the standard QC approach – i.e. outlining the evaluation criteria, how interviews will be evaluated, what they should expect from the feedback process and their role in the process. Interviewers are also trained, as necessary in the practical aspects of CARI for example, setting up an external microphone (if used) or setting up equipment for recording when interviewers are carrying out decentralized telephone interviewing.

The training provided to those completing the evaluations was generally, more extensive. These staff is generally very experienced interviewers so they have knowledge of interviewing skills but need to be trained in the evaluation criteria and the evaluation tool or

system. Inter-rater reliability is used by some organizations to monitor and achieve consistency of scoring, although this is not the norm.

- **Legal Issues**

Legal issues is a topic that was not discussed in detail but was identified as one that should be included in further post-meeting discussions. Items for future discussion include:

- best practices to share;
- file destruction practices;
- are there appropriate ways to train interviewers to respond to R questions regarding consent and legal issues regarding recordings; and
- Long-term storage of recordings.

- **Affects of CARI**

The overall consensus amongst the group was that CARI has facilitated an improvement in data quality – primarily through enhanced feedback on interviewer performance but it also encouraged ease and frequency of monitoring by clients and project teams which assisted the questionnaire development process which contributes to better survey measures and data quality.

## **2.2 Methods: Operations and implementation of CARI**

Organizations were also asked to share their experience around implementing CARI - the following questions were designed to facilitate this.

### **Sharing the implementation experiences with CARI**

- Tool. What tool did you use? Why did you select it?
- Technology. What is the recording technology used? Audio only, audio and video?
- Storage. What is the recording storage medium? Are the recordings sortable by interviewer ID or questionnaire number?
- Quality of Sound. What problems or challenges have you faced on the quality of the sound file?

A summary of the discussion of the key issues around implementation of CARI is provided below.

- **Audio or Audio and Screen (or video)**

Blaise CARI is used by all organizations who participated in the meeting – the output from CARI is an audio file and, if requested, a screen shot of the corresponding screen. University of Michigan also extensively uses a system that is controlled from outside the Blaise questionnaire via a capture list. The system records a video of the sections of an interview as determined by the capture list (using, Camtasia®, the video capturing software).

There was some discussion during the meeting about the added value of having a video over a screenshot – and varying views on this. It was generally agreed that the major advantage of video above audio and a screen shot is that video captures interviewers' actions and interaction between questions – for example; 'off script' discussions which might highlight poor interviewer behavior, how soft checks are handled, what feedback is provided and frequency of feedback by the Interviewer. It also captures interviewer: respondent interaction when an external (i.e. non-Blaise) program is launched. For example, the completion of an event history calendar or cognitive test which is launched via a DLL from within the Blaise questionnaire can still be captured and displayed for evaluation.

Having a video image of the laptop screen allows staff to see if an interviewer has any other programs open on their laptop at the time of the interview that they should not be using. The video can also provide a useful insight – especially at the piloting stage – of any usability issues of new question design or procedures.

- **Storage/file size/performance/sound quality**

One of the main practical concerns organizations faced during the implementation of CARI was file size and the possible impact on other systems – such as interview speed and transmission of files. However, these issues and risks were controlled by saving files in formats that produced files of a smaller size, setting a minimum file size which removed empty recordings and compressing files to a lower but still usable quality.

Sound quality was not an issue, the only problems are a result of unavoidable circumstances during face to face interviews.

The discussion turned to ways organizations would like to use CARI that they are not able to currently and planned developments of current systems. The following were identified;

- **Record all interactions between a face to face interviewer and the respondent/informant**

A complete record of interactions with the respondent is available for centralized telephone interviewing, however, for face to face or decentralized telephone interviewing this stage of the interview is not yet recorded. Organizations agreed that recording the full interaction between respondent and interviewer from doorstep to the start of the Blaise questionnaire would be incredibly useful. In addition, to complete the entire process, the recordings would need to include any telephone contact they make with the respondent prior to arriving on their doorstep. Ideally, it could be presented like an audit trail of all contact attempts during the entire process to gain cooperation. There was some discussion about the type of technology that would be used for CARI outside the interview and legal and ethical issues were also noted.

- **New technologies and capturing CARI using them?**

The following new technologies were mentioned as possible ways to capture CARI, all organizations expressed an interest in investigating these new technologies for CARI.

- Capturing CARI on smart phones
- Capturing CARI on other hand-held or non-traditional devices
- Using Skype or other systems to capture video files.

- **Flexibility**

The ability to make changes ‘on the fly’ varied across the organizations and was dependent on the systems they have developed or if they are using CARI or another recording system. Organizations commented they would like to have more flexibility around making changes to what is being captured – such as changing the questions or sections that are being recorded during data collection, change the rate at which interviews are being recorded overall or vary the rate by interviewer. In Blaise CARI it is possible to change what is being recorded – but it is not possible to vary what is being recorded for different cases or by interviewer. It was agreed that the goal should be to capture recordings of all interviewers across the full questionnaire. The ability to randomize the capture list would facilitate this.

- **Other uses of CARI**

- Can or should CARI be used in lieu of verbatim recording on (some) open-ended items?
- Key word searches from database of recordings to identify cases to evaluate

### **3 Conclusion**

Recent versions of Blaise have supported CARI and Blaise CARI is now well developed and supports the current use of CARI as outlined above. Blaise CARI has been developed with the user in mind i.e. providing a tool that has many options (in the CARI settings) that address the practical issues that need to be considered when implementing CARI – such as, the ability to control the maximum and minimum file size, recording time and file format. It is also possible to operationalize many of the requirements essential to making CARI a useful tool within the CARI settings in Blaise. For example, the response to an assigned consent question will control if recording is activated and how much of an interview is recorded. Blaise CARI cannot be controlled in anyway by the interviewer and as is preferable with any CARI system, interviewers are not aware when the system is recording. Blaise CARI also produces a log file, providing a trail of when the recording starts and stops during the interview and the name of the corresponding sound and image file (a screen shot).

A limitation of the Blaise CARI system is that it only records the interaction which occurs between the interviewer and respondent when they are within the Blaise questionnaire. Blaise does not capture any interaction which occurs when external programs are used for data collection, from within Blaise.

During the meeting at the University of Michigan an idea was formulated to open up CARI in such a way that the user can specify to use another recorder than the built-in audio recorder. This would make it possible for the user to rely on their own recorder component while having the full benefits of Blaise CARI. This idea, also known as ‘the CARI hook’, has by now been implemented in a beta version of Blaise 4.8. The CARI hook can for instance be used to record screen video (using for instance Camtasia®) but it can also be used to record VOIP. Note that the user will need to create an ActiveX® component to interface between Blaise CARI and their recorder component.

Participating organizations agreed to continue to meet and share updates periodically.