Component of Statistics Canada Catalogue no. 11-522-X
Statistics Canada's International Symposium Series: Proceedings

Article

Symposium 2008:
Data Collection: Challenges, Achievements and New Directions

Incorporating Audio File Information in Survey Data Collection

by Charles Loftis and Shane Trahan

2009

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2008
Data Collection: Challenges, Achievements and New Directions

Incorporating Audio File Information in Survey Data Collection

Charles Loftis and Shane Trahan[1]

Abstract

Collecting data using audio recordings for interviewing can be an effective and versatile data collection tool. These recordings however can lead to large files which are cumbersome to manage. Technological developments including better audio software development tools and increased adoption of broadband connections has eased the burden in the collection of audio data. This paper focuses on technologies and techniques used to record and manage audio collected surveys using laptops, telephones and internet connections. The process outlined involves devices connecting directly to the phone receiver which streams conversations directly to the laptop for storage and transmission.

Key Words: Audio data collection, Audio transmission

1. Introduction

1.1. Introduction

Data collection can take many forms ranging from paper and pencil tabulation to Computer Assisted Interviewing (CAI). Many of us have been exposed to these and other forms of data collection techniques; however there is one form of data collection that has eluded survey specialists for many years and that is the inclusion of audio recordings for data collection. There are many reasons for not collecting data this way, one of which is the difficulty in moving data from where data is collected to a centralized location where files can be transcribed or listened to. Expanded adoption of broadband internet connections (Horrigan, 2008) has made this task easier but managing large audio files is still cumbersome and can take significant time to transfer.

1.2. Data collection

Data collection has taken on many forms over the years and the tools range from simple paper and pencil data collection to Computer Assisted Web Interviewing (CAWI). Each one of these data collection techniques has their pros and cons and survey specialists must evaluate each type of data collection instrument to fit their specific needs. Interviews are primarily done face to face, over the telephone or done using a Computer Assisted Personal Interviewing (CAPI) system. Historically, face to face interviewing achieves some of the highest response rates (De Vaus, 2002) but it is also the most time consuming and expensive. On the other hand telephone interviews get much higher response rates versus mail based surveys but bias can be observed since only the segment of the population with phones will be interviewed. (It should also be noted that a high number of individuals only subscribe to cell phone service whose contact information can be difficult to obtain.) The CAPI and CAWI based interview techniques enable data to be directly entered into a computer where time can be saved in data processing and validation however these interviewing techniques are limited by respondent computer literacy, internet connectivity, interviewer training, hardware acquisition and software development startup costs. Paper pencil questionnaires can be mailed to a large amount of people and respondents are likely to be more truthful while responding to controversial questions (Muijs, 2004) however many people who receive these type of questionnaires do not respond. There is no doubt that data collection techniques using

[1] Charles Loftis, RTI International 3040 Cornwallis Rd, Durham NC 27709 USA; Shane Trahan RTI International 3040 Cornwallis Rd, Durham NC 27709 USA

handheld devices will soon become more prevalent as more and more people become accustomed to this type of device but those will face many of the same hurdles detailed above.

1.3. NSDUH project

The National Survey on Drug Use and Health (NSDUH) is a nationwide survey conducted annually by Research Triangle Institute (RTI) International for the Substance Abuse and Mental Health Services Administration (SAMHSA). It collects screening data from approximately 160,000 households and interview data from 67,500 individuals in all 50 states and the District of Columbia. The NSDUH is the Federal Government's primary source of national data on the use of alcohol, tobacco, and illicit substances in the general U.S. civilian non institutionalized population, age 12 and older.

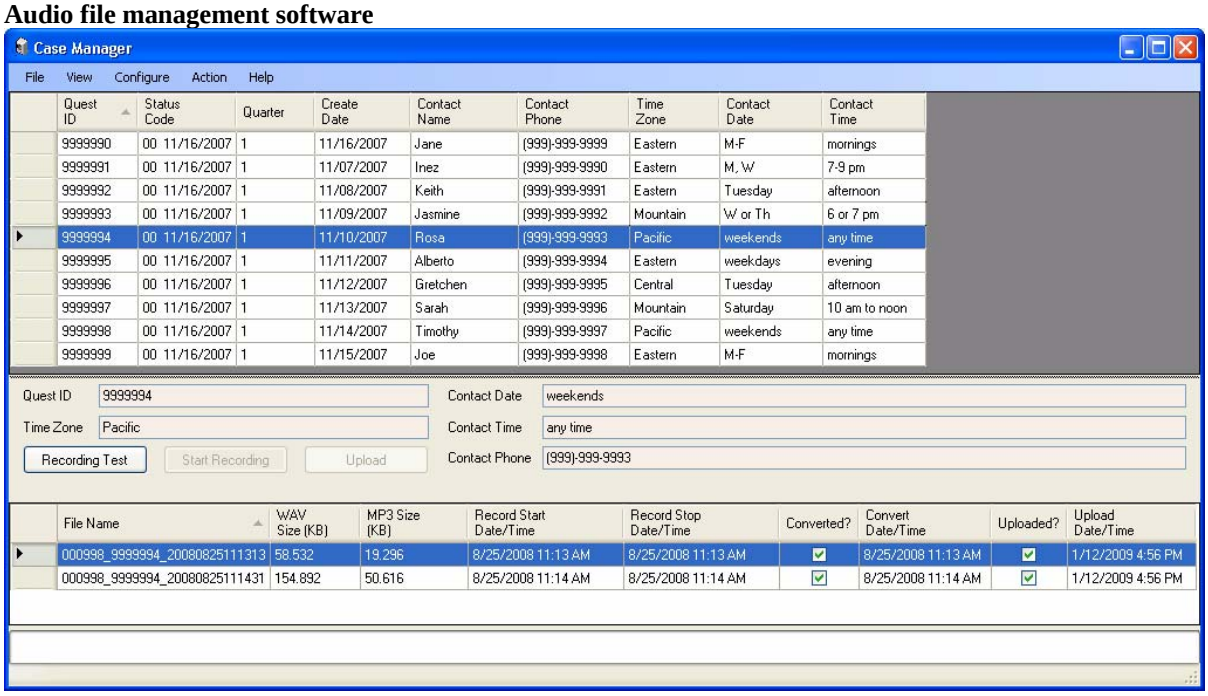
From time to time, SAMHSA requests that the NSDUH conduct sub-studies for assessing survey improvements, optimizations and cost saving measures. For the survey year 2008, SAMHSA requested that NSDUH conduct a telephone based sub-study to estimate adult serious illness and determine the performance of survey questions modules for retention in future NSDUH interview instruments. The telephone based survey required specially trained clinical interviewers to record all responses using paper and pencil all the while consensually having the entire conversation's audio recorded. The audio files were stored on a project provided laptop and transmitted back to RTI to use for validation and quality control purposes. The remainder of this paper outlines the solution developed by RTI to record, manage, transmit and archive audio data for use on NSDUH's mental health sub-study.

2. Methodology/Technical approach

2.1. Model

The goal to enable the sub-study was to gather these audio files to RTI for further review by project related staff. In order to do this it was required to pull or transmit these files to a centralized location. Since it was assumed that audio files could be very large it was decided that an approach similar to how data packets are transmitted through the internet would be an ideal model for the design of our transmission system. The transmission system breaks up the audio files into more manageable packets or chunks and sends them to a web based service that would put these files back together into a single file. Each chunk once received would be saved to the server and a size validation is performed to validate transmission success. If the chunk size validation fails the chunk is deleted and the service issues a request to the client to resend it. Once all of the chunks have been received and validated the service puts the file back together into a single base64 text string and converts it to a binary (audio) file for review by users of the secure website.

Figure 2.1-1
Audio file management software



To successfully transmit this data back to the centralized server there was a need to not only manage the transmissions but to also manage the actual interviews themselves. The audio file management software (Figure 2.1-1) that was developed incorporates both the capture and conversion of audio files as well as necessary functions to communicate effectively with web based services. The interface developed allows optimal management for the acquisition of audio recordings while keeping the complexities of the transmission and file conversions behind the scenes. The case manager interface allows interview staff the ability to review status of recordings including a simple to use scheduling application to manage workload.

2.2. Audio file acquisition

After properly connecting the telephone to the audio input of the laptop using an analog or digital dongle connector, the clinical interviewer begins the interview and starts recording audio. Each audio recording is stored as raw data (also known as the "WAV" format). While WAV files provide the highest quality recording it also requires considerable storage space. A two (2) minute recording at typical phone transmission quality (mono, 4-8 MHz) consumes approximately 1 MB of disk space. Freely available conversion utilities allow lossless WAV audio data to be converted to MP3 format. Using a free "lame encoder" and configuring different sampling rates (number of times audio data is stored per second in Hz), channels (mono, stereo) and bits per sample (1, 2 bytes, etc.) the quality of the conversion can be controlled. We empirically determined conversion settings that most closely reproduced the original audio in the compressed MP3 format while still significantly reducing the file size. The compression rate was substantial - nearly 3 to 1 in size for WAV to MP3. Thus, after capturing the audio data in a lossless format our first action is to convert the WAV file to a much smaller sized MP3 file. An overview of this process is outlined in Figure 2.2-1.

Figure 2.2-1
Convert the WAV file

2.3. Client side audio preparation

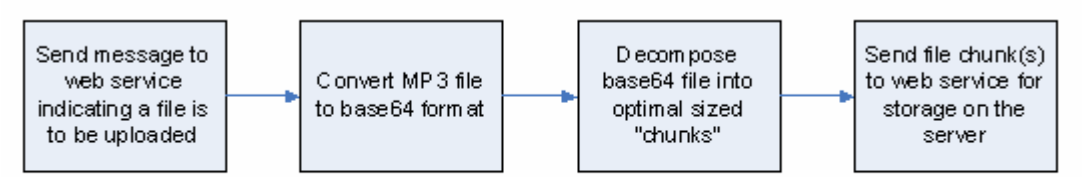
Once a recording has been completed and converted from WAV format to MP3 it undergoes another format conversion to base64. Base64 is the set of all upper and lower case characters, numbers 0-9 and the dash character ("-"). The conversion of the binary, MP3 file to a base64 text file facilitates the transmission of the audio data over standard hypertext transfer protocol (HTTP) connection without interference from intranets and/or firewalls.

Once the file is in text format it is ready to be transmitted for review and storage. The Audio File Management software then notifies the web based service that an audio file is ready for upload. The message is sent to the web service as an XML formatted message providing the size of the final MP3 file, the name of the file and other important parameters. Once the service has acknowledged receipt of this information the web based service then sends an XML acknowledgement back to the Audio File Management software letting the client know that the service is ready to receive the file chunks. Based on the internet connection speed, the Audio File Management software then divides the base64 file into optimally sized file chunks and begins transmitting them back to a central repository for merging. The chunks are only considered transmitted when a successful acknowledgement has been received from the transmission web service. The Audio File Management software sends the following parameters to the web based service with each base64 data file chunk:

ID – Distinct Interviewer ID
Offset – Text file offset
LengthVal – Chunk length
ClientFileName – Filename comprised of the completed file
Base64Chunk – The base64 text chunk

An overview of this process is outlined in Figure 2.3-1.

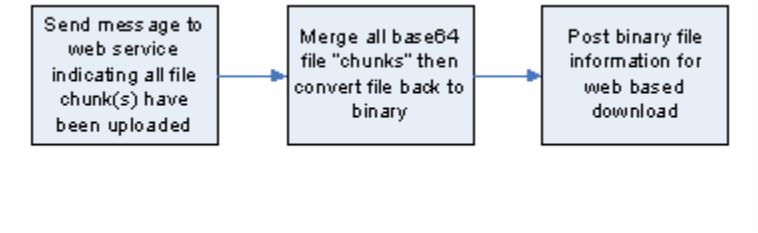
Figure 2.3-1
File transmission



2.4. Server side audio manipulation

All verified chunks are saved to a folder with corresponding parameter information saved to a database. Once the last file chunk is received the Audio File Management software will send a message to the service indicating that all chunks have been received and verified indicating that the service can move forward with merging the text files together and convert the file back into binary format. The service then queries the database of all file chunks received for a specific file in descending order of the offset parameter and builds a complete base64 text file according to the information in the database. The service will then convert the base64 text file to a binary file and compare the size of the binary file to the information originally sent. If file verification is successful, the new file is posted to the management web site where it can be reviewed. An overview of this process is outlined in Figure 2.4-1.

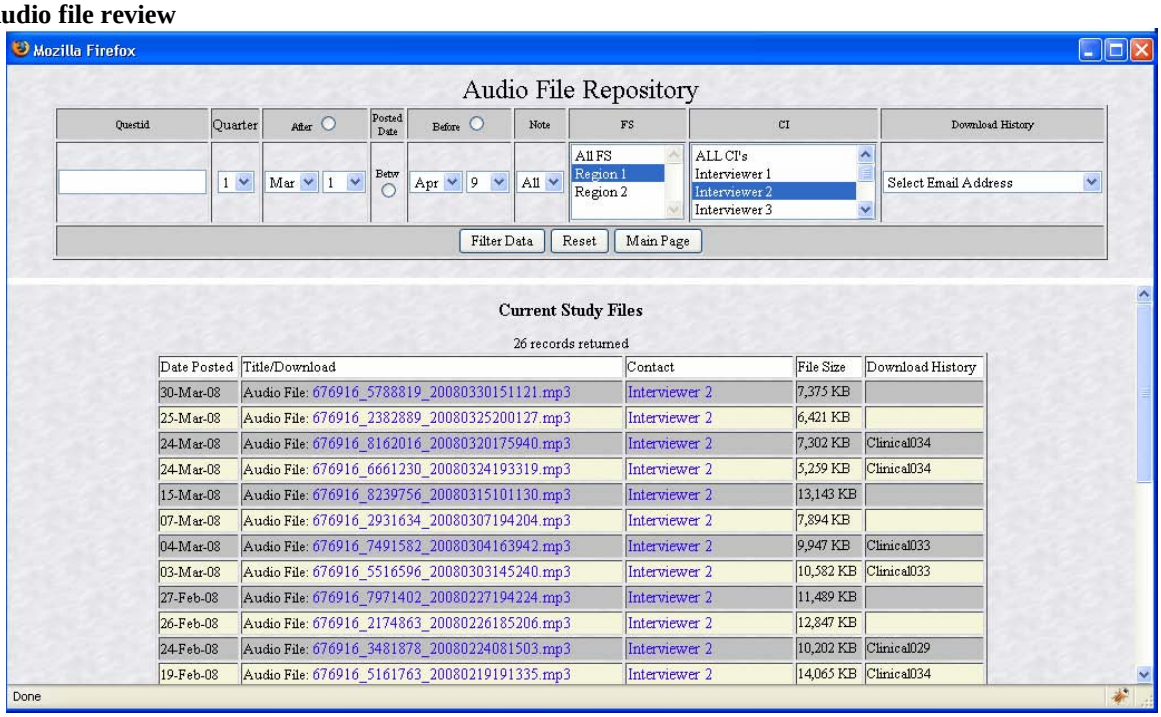
Figure 2.4-1
Post binary file information



2.5. Audio file review

Once files have been uploaded to the centralized server, audio files can then be reviewed via secure website access. The web based review system enables users to filter for audio files by upload/receive date download history, questionnaire identifier (Questid) or interviewer. Audio files are in standard MP3 format and can be played using a variety of browsers and operating systems.

Figure 2.5-1
Audio file review



The website is also narrowcasted allowing only supervisors or clinicians access to only their specified regions of the file repository. For example, field supervisors of region one can only access their interviewer's audio files but project clinicians can access all uploaded audio. In addition to the narrowcasting feature each audio file reviewed by a clinician or any other user of the site has their action logged to a database, which is used for tracking of potentially sensitive recordings. With all files centrally located and accessible via a simple web based interface audio recordings are easily reviewed by anyone with adequate permissions.

3. Results

3.1. Results of audio capture system

The software and system developed for this paper were incorporated into the 2008 NSDUH Mental Health sub-study and found to be an invaluable resource. During 2008 we received and managed approximately 6,604 (audio) interview files across 451 folders totaling over 12 GB worth of data. The average file size was approximately 7.3 MB (equivalent to about 90 minutes of audio data) with an average upload time of approximately 3 ½ minutes at 34.06 KB/s. We did require each laptop to connect via broadband connection however our approach to the transmission of audio files is sufficient for dial up users as well. It should be expected that dial up users would be at a severe disadvantage when transmitting large files.

Both clinicians and survey managers were able to review and track the progress of the audio based survey responses from the web based interface and our approach to transmitting data and managing such files proved to be more than adequate and flexible to manage an audio response based survey. The NSDUH Mental Health Sub-Study was so successful that it was continued for the 2009 survey year.

3.2. Considerations

Some of the issues that were encountered during this audio data collection effort included recording interference and some problems related to transmissions. Although our team was able to overcome obstacles related to audio recording we did encounter problems difficult to solve because of limitations outside of our control.

Audio Recording Interference:
During audio interviewing we discovered that using the telephone hookup dongle to record phone conversations while the laptop was plugged in (using the grounded AC identifier) caused an audible humming sound to be incorporated into the recording. In many instances, the interference was loud enough to deem a recording inaudible. We learned that the static noise was caused by ground interference coming through the ground prong on the AC power unit. Therefore when using the laptop's battery as the sole power supply the static interference would be eliminated. The downside to this is that interviews could be long and battery life would become a factor during the interview. Interviewers were trained to keep their battery fully charged at all times.

Transmission issues:
If problems were encountered during transmission or conversion of the WAV file to MP3 format an alternate method for obtaining the recordings had to be created. A process was established where central, home office staff could flag a specific file on the laptop to be transmitted back to the centralized server even if it had not been converted to MP3 format (such as the case when an audio file was corrupted because of sudden loss of power). The file would still be converted to base64 format and sent in chunks back to RTI just like any other audio recording. Any files flagged by RTI staff to be pushed by the client to the centralized file store or re-transmitted would be attempted to be uploaded at each connection until the file had been confirmed to be received.

Our technical approach eliminated many of the issues commonly found with conducting audio surveys. Our methodology greatly minimized the file size needed to store the interview and our transmission system was both flexible and reliable for getting information back to the central file repository.

4. Conclusion

4.1. Conclusion

Overall the approach outlined in this paper to record, manage, and transmit audio files for survey research was found to be successful in clearing the initial hurdles encountered which include the storage and management of

potentially large files. Other opportunities we see involve an evolution of utilizing technology to better capture audio recordings. This could include the automatic creation of transcripts and the tagging of key words for each interview. Eliminating the labor intensive task of transcribing interviews could also help clinicians and survey methodologists target specific interviews for further investigation. The single largest benefit of audio interviewing is the ability to probe for more information however probing requires interviewers to establish rapport with the respondent which can include non-important dialog during an interview; such auto-transcription can help filter this chatter out. Auto transcribing audio files to text would also present the opportunity to translate the textural data into other languages. There are numerous services currently available that will translate text from one language to another. Although these services are relatively costly they are coming down and will probably be much more common place in the years to come. Translating the results of audio interviews into multiple languages will not only expand the audience of such research but will help to alleviate language barriers encountered during interviews across various populations. We also foresee the possibility of making minor changes to the Case Management software to improve field usability. One possible change discussed is to automatically "stream" data files back to the centralized server during idle computer and network usage times.

References

Horrigan, J.B. (2008). Home Broadband Adoption 2008. Pew Internet & American Life Project. <http://www.pewinternet.org/pdfs/PIP_Broadband_2008.pdf>.

De Vaus, D. (2002). *Surveys in Social Research*. 5th edition. New York: Routledge, 127

Muijs, D. (2004). *Doing Quantitative Research in Education: with SPSS*. 1st edition. London: Sage Publications, 41.