

Voice Recording: Tips for Non-Audiophiles

Timothy Keirnan

On behalf of Tec-Ed, Inc.

tim.keirnan@jnli.com

ABSTRACT

The author relates using portable audio recorders in field studies to provide clients with vivid impressions of their users through use of audio highlights, as well as his experiences using audio recorders as back-ups for lab test equipment. The author reviews principles of sound propagation and its impact on recording speech, and profiles portable audio recorders available today.

INTRODUCTION

Human speech is a powerful communications medium. Compared with handwriting, speech often evokes more of a usability study participant's meaning and state of mind. Variations in tone and volume during speech—including its ability to communicate through silences between words—enable speech to convey more revealing information about the speaker than the written word. These advantages of speech can be harnessed to enhance the effectiveness of usability reports for their audience, as well as provide effective back-ups to other session recording methods.

The human voice may well have been the first musical instrument. And in the thousands of years before widely available books—or radio, movies, and television—speech was the primary storytelling medium. In 2004, storytelling around a campfire is still appreciated by many. The power of the storyteller to hold a group of listeners transfixed, whether in the case of a wandering minstrel 600 years ago or on radio programs in the 21st century, is still very strong. Radio shows such as *This American Life* out of Chicago or *The Prairie Home Companion* from Minnesota attract audiences nationwide. The myriad ways in which visual images saturate 21st century cultures have augmented but not replaced oral storytelling.

During their work, usability practitioners often record interactions with study participants. Sometimes these recordings must include video—for instance, during lab tests when recording how participants performed tasks is required. However, for other usability methods such as contextual inquiry and ethnographic interviews, audio recording has specific advantages for capturing the experiences related by participants. Also, some usability professionals use a portable audio recorder during lab test sessions to record users thinking aloud in the event video equipment failure prevents some or all of the video from being used. Obtaining high-quality audio recordings of participants is of great advantage to the usability practitioner, and many excellent portable audio recorders are available to do so.

USABILITY AS STORYTELLING: RECORDING THE STORIES

In the usability profession, and especially in field studies, practitioners work amidst oral storytelling. In conversing with participants or observing them during think-aloud protocols, practitioners hear stories created in participants' present moments or hear stories told based on participants' pasts.

In field research and in lab studies, usability practitioners keenly observe users and interview them in great depth for clients who want to design products for real-world use or learn how well existing products support their intended users. From the non-fiction stories users tell, usability professionals write and edit what might be termed "clinical non-fiction" for delivery to their clients. Therefore, in effect, usability practitioners are "clinical storytellers of non-fiction."

As part of their research, many usability practitioners may audio-record study participants' non-fiction storytelling for one or more of the following reasons:

1. To supplement written notes made during ethnographic interviews or contextual inquiries.
2. To back up lab test sessions in the event of trouble with a lab's video recording equipment.
3. To provide raw material for creating sound clips of participants for inclusion in a results report.

The first two reasons above are basic, professional responses to managing risks in a usability project. By preparing for the sudden, unwelcome appearance of Murphy's Law in the form of a memory lapse or video equipment failure, usability practitioners can avoid losing critical data they need to analyze for clients. In the third instance, audio clips of participants expressing key points of the research, in their *own* words, can enhance the results report to reinforce specific conclusions drawn by the usability practitioner.

Whatever the purpose in recording speech, professional-sounding recordings are extremely useful to have. Until the latter part of the 1990s, most voice recording in field studies was done with the traditional cassette tape recorder. However, digital recording devices have matched and, in many cases, surpassed cassette tape recorders in sound quality, reliability, and usability in the last several years. This paper explains basic principles of recording speech and describes recording options available to the usability practitioner. Because selecting the most appropriate recording hardware depends entirely upon numerous factors known only to each usability practitioner, this paper does not recommend specific hardware; rather, its purpose is to arm readers with knowledge that empowers them to make their own informed decisions when purchasing and using portable audio recorders.

WHERE AUDIO OUTWEIGHS VIDEO

With the arrival of lower-priced digital video camcorders and affordable personal computers that enable digital video editing, it is easier than ever to shoot video at usability study sessions and create video deliverables for clients. Some field studies and many lab studies are designed with videotaping the sessions a requirement, and the results reports include a CD or DVD containing video clips of participants making key points.

However, the mere availability of a tool should not be the sole determining factor for using it on a project. The author has found that incorporating video recording into a field study introduces three factors that may be undesirable:

1. Participants may appear more awkward or shy in front of a video camera than they would near a portable audio recorder, which is often less noticeable or conspicuous than a video camera.

- This “smile, you’re on candid camera” aspect of video recording—its intrusiveness in the study environment—may negatively impact the data collected during a session. If the goal of a study is to capture participants’ behavior as closely as possible to their non-observed state, then making them acutely aware through videotaping of being “on stage” rather than “at home” or “at work” may not elicit behavior as natural as desired.
- The author has witnessed multiple occasions where participants—forewarned that videotaping was involved in their participation in a field study—demonstrated through comments and body language that they were not comfortable under the “eye” of the camcorder on its tripod until well into a session. In contrast, the author has not noticed a single instance where the presence of a portable audio recorder has apparently caused the same unease. Research involving human participants demands an obligation to respect their comfort.
- Some environments do not permit use of video recording due to proprietary or security reasons. In such environments, it may be possible to obtain permission for recording audio.

2. The delivery and playback of digital video may be more complex, both for the usability practitioner and the client, than delivery and playback of digital audio.

- Unlike digital audio files, digital video files are still in a developing stage of immature complexity. Digital camcorders require a learning curve to operate successfully; as with photography, videography is a skill that must be learned to consistently produce good-looking results. When video is transferred to a computer, more layers of complexity appear:
 - Which file format to import the video?
 - Which video compression/decompression scheme to use to achieve acceptable-looking video that doesn’t require too many gigabytes of space?
 - Which output file type to use?
 - Which delivery medium to use, video files on a CD or a DVD? If DVD, which DVD format to use, the “plus” or “minus” format?
- Producing the digital video files is time-consuming on “middle class” personal computers. Rendering video on a 2GHz PC with 512MB of RAM and a 7200 RPM hard drive can often take 4-5 hours to complete a full-screen video of less than an hour’s duration. Using less than full-screen size runs the risk of losing subtle details in the picture that may be important for viewers to see.
- Clients also may have difficulty viewing video deliverables. For example, some computers in corporate environments lack DVD-ROM drives. Video formats that can fit onto CDs require special software to play, as is the case with Video CDs or Super Video CDs. A client’s computer may lack a critical piece of software called a codec to play a certain video file, even if the computer has the correct video player application. Not all recordable DVDs, both the plus and minus variety, are guaranteed to play back on all DVD players. The DVD format itself is not consistently designed and implemented.

- Meanwhile, production of digital audio files has matured to a state that, in the author's experience, is easier to learn and produce than has video thus far. The increasing popularity of the .mp3 file format for audio files provides a common, space-saving, multi-platform solution for audio delivery, assuming high-quality .mp3 files are created.

3. The cost in terms of equipment resources, learning curve, and production time for producing digital audio clips is less than the cost for producing digital video.

- In the author's experience, shooting, editing, and producing video requires more time to learn (on initial projects with video) and produce (on all projects with video) than does audio. The reason is simple: video recordings contain both sound and video to process, whereas audio recordings contain only sound. Naturally, these costs vary with each practitioner's skill set and office resources. Some field studies may have budgets to cover video's extra cost in learning curve and resources (maybe even providing funds for using help whose job is only to produce video), but field studies on tighter budgets may benefit from restricting highlight clips to audio.
- While video clip processing requires at least a "middle class" computer due to the huge amount of information being processed, audio production on middle class or low-end systems is quite satisfactory.

The author does not mean to discourage the use of video for usability research. For some projects, clients may insist on video recording, or the goals of the study and the study design may require video. However, for other projects, audio recording may be a more effective option.

ACOUSTICS

Understanding how voices are heard is the first step in learning to record them well. The following section offers a very brief introduction to acoustics.

The Nature of Sound

Imagine a sound source suspended in the sky on a still day. Sound would emanate spherically from the source in all directions simultaneously and, were people able to approach this source suspended in the air, they would hear what is called "direct" sound coming from the source. This direct sound would be pure—that is, with no acoustic interference affecting it. The sound waves would emanate from the source and move continually away from it, without ever being reflected back.

In reality, sound becomes more complex after it emanates from its source. Imagine two people—call them Pam and Quincy—standing at opposite ends of a conference room. When Pam says "Hi, Quincy," Quincy hears the sound waves not only directly from Pam's mouth to Quincy's ears, but also sound waves from Pam's mouth that were delayed because they bounced off walls, tables, furniture, floor, and ceiling before entering Quincy's ears. Envision reflected sound waves of "Hi, Quincy" that zoom around the room as a "stew" of sound that is prepared and served to Quincy's ears.

Sound reflected off objects in a room and mixing with itself when reflected off other objects is often less clear than direct sound. Rooms always add their own acoustic signature to sounds generated inside them, sometimes with unpleasant results.

The Nature of Hearing

Getting sound from the source to the eardrum is only half the process of hearing. Hearing occurs in the conscious mind only after the inner ear and brain have processed the incoming sound. This processing of sound before conscious awareness occurs is known as psychoacoustics. It is beyond the scope of this paper to delve into psychoacoustics, except for one key point: the way early sound reflections as well as later reflections can affect the intelligibility of voice recordings.

Remember the sound stew in the conference room when Pam said “Hi, Quincy” and the sound waves zoomed all around to mix, slightly later, with the direct sound at Quincy’s ears? Quincy did not hear each sound wave separately. He heard one “Hi, Quincy” sound made from all sound waves—direct and delayed—reaching his ears in the first 50 milliseconds after Pam greets him. After 50 milliseconds, any sound waves left in the room from “Hello, Quincy” sounded like reverberation to Quincy (the “concert hall” or “tiled restroom” effect) or, if they arrive much later than 50 milliseconds behind the direct sound, as distinct echoes of “Hi, Quincy” (the “echoing canyon” effect). These “after-effects” can sound pleasant (as with reverberation in a well-designed concert hall) or confusing (as with too much reverberation that “smears” the sound, or too many echoes that muddy the sound).

In sum, hearing is as much about mental processing of auditory stimuli as it is about those stimuli reaching the eardrum initially. By reducing as much as possible the amount of sound reflections entering a portable audio recorder’s microphone, the usability practitioner can achieve better recordings.

The Nature of Microphones and Good Recordings

Usability practitioners should, ideally, record speech in rooms containing few acoustically reflective surfaces. Annoyingly reflective surfaces are often more “hard” than “soft.” Ideally, a room with lots of carpet and upholstered furniture is better for recording speech than a room with bare wood floors, hard walls, glass table tops, and high ceilings. The fewer sound reflections entering the mic, the cleaner-sounding speech recordings will be. Professional recording studios spend lots of money on acoustically isolated rooms that make it easy to record mostly direct sound from vocalists.

Microphones do not have human brains combining the first 50 milliseconds’ worth of sound wave reflections of one noise into one sound. Reflections of sound waves entering the mic are dutifully sent down the mic cable to the recording device. Too many reflected sound waves in those first 50 milliseconds can blur the sounds of words in listener’s minds, and too much room sound can also degrade intelligibility. Thus, both early and late sound reflections and microphones are a bad mix in most situations.

Heating and air conditioning ducts can ruin an otherwise excellent room for recording. The solution to this problem, and to recording in noisy environments like shop floors or outdoors, is to use lapel mics that attach to participants’ clothing near their throats. Admittedly, clipping mics to participants may make them feel the “candid camera” syndrome mentioned earlier in regard to video recording, but the usability practitioner must balance the desire for intelligible audio recordings with the desire to capture participant behavior as naturally as possible.

EDITING SOFTWARE AND DELIVERY METHODS

Audio editing software is available for all computer platforms. Try the software before buying because the more usable the editing software, the more time saved in the editing and production phases. Many audio editing applications have free trial versions available that allow recording and editing of clips, but prevent saving or exporting anything. Some applications are available as shareware that last for weeks in fully operational condition, so one can do an entire project with the tool before deciding whether to buy it for future projects.

In the Windows world, CD-quality audio is saved as .wav files. In the Macintosh world, CD-quality audio is saved in the .aiff format. To ensure that clients can play audio highlights easily, usability practitioners should learn which platform the client uses. Even better, the finished audio clips can be converted to the universal .mp3 compressed file format. Software for .mp3 playback is available on each platform, and the smaller file sizes of .mp3 are an added bonus.

Regardless of the editing software used, usability practitioners who create audio highlights must decide how to deliver the audio clips to clients. Compact disc is a universal standard and has room aplenty for almost any project's audio clips. If only a few audio clips reinforce the results report, email is another delivery method, but make sure clients have broadband Internet access before emailing large audio files to them.

Audio files can be delivered as separate files with descriptions as an appendix of the written usability study report, or the audio files can be delivered as CD audio tracks with the usability practitioner narrating each clip's relevance to study results. In the narrated method of distribution, clients can even listen to part of a report in their cars during commutes (if their cars have CD players). The basic process for adding narration to audio clips is as follows:

1. Transfer session audio recordings to a computer.
2. Locate parts of the recordings that should appear in the audio highlights of the results report.
3. Edit each audio clip into a separate file with a unique label.
4. Compose short scripts for narrations and practice them before recording them.
5. Record the narrations and leave a couple seconds of silence at the end of each one.
6. Edit the narrations into separate files with labels to identify them. Keep the two seconds of silence at the end of each narration clip.
7. Copy the sound from a narration clip and paste it to the beginning of the appropriate participant clip.
8. Play the resulting file to ensure the right narrative precedes the right participant clip. Adjust any volume differences as needed.
9. Save the finished highlight clip.

PROFILE OF PORTABLE AUDIO RECORDERS

The following sections list the characteristics of portable audio recorders on the market today and their approximate costs. The author does not mean to emphasize one device as “the answer” to recording speech in field studies or as backups to lab equipment, but to provide an overview of the advantages and disadvantages of each device in both the recording phase and editing/production phases.

Types of Recorders

Due to the ever-changing nature of the consumer and professional electronics world, this list is not exhaustive. Note that a recorder that is easy to use in the field may not be as easy to use in the editing and production phases, and the author encourages readers who use recorders other than those listed here to contact him with their opinions on those devices.

Cassette Tape Recorder

Recording Pros	Recording Cons	Production Pros	Production Cons
Very easy to learn and use.	Tape hiss and motor noise, which can obscure a soft-voiced speaker’s remarks.	None.	Contents must be transferred to computer in real time.
Somewhat durable and inexpensive.	Cassette must be flipped over on many models to record on second side.		Tape hiss can be reduced with special software, which adds more steps to the production process.
Cannot lose all of a study’s recordings if one participant’s cassette tape fails.	Internal mic usually sounds poor.		Finding a specific index point is time-consuming due to tape’s linear, sequential nature.
90 minutes per cassette (tapes longer than 90 minutes use thinner tape that is more fragile).			

Conclusion: Do not expect cassette recordings to produce quality audio. However, a cassette recording is better than no recording.

MiniDisc Recorder

Recording Pros	Recording Cons	Production Pros	Production Cons
Provides virtually CD-quality audio on blank mini discs that are cheap to buy and extremely durable.	Usually need a separate mic to record live sound.	Index points are instantly accessible (and numbered), which helps users find a sound clip or mark sections of an interview that correspond with section numbers of session notes.	With the original MiniDisc format, contents must be transferred to computer in real time due to poorly conceived and implemented copyright protection schemes. This is improved in the new Hi-MD format.

Voice Recording: Tips for Non-Audiophiles

Recording Pros	Recording Cons	Production Pros	Production Cons
80 minutes stereo or 160 minutes mono, with more time available in the new Hi-MD format; excellent for soft-spoken participants because of its high resolution.		Audio quality does not require any de-hissing or polishing.	
Easily inserted index points for instant playback later.		The new Hi-MD format allows quick uploading from MiniDisc to computer.	
Cannot lose all participants' recordings if one participant's mini disc fails (the author has never seen an MD fail and hasn't heard of one failing, either).			
The lack of an internal mic encourages acquiring a good external mic to use with the unit.			

Conclusion: MiniDisc is not as popular in the USA as it is in Europe and Asia, but it is well worth considering as a portable audio recorder. The new Hi-MD format may make MD even more attractive to users.

Digital Voice Recorder

Recording Pros	Recording Cons	Production Pros	Production Cons
Very small size.	Limited recording time (typically several hours but not enough for a whole day of sessions)	Supports direct transfer of audio files to computer for editing.	Original files are in proprietary .dss format, which must be converted to .wav or .aiff or .mp3 later.
Solid-state reliability.	Generally poor audio quality due to use of audio processors like those used in grainy-sounding digital telephone answering machines. This problem should improve with time.	Some multi-platform support.	
Can record sessions into different directories to keep files organized before uploading them to computer.			

Conclusion: The author has yet to hear a digital voice recorder that sounds as good as a MiniDisc recorder or a laptop recorder. Too often, the audio processing chips in digital voice recorders are no better than those used in digital telephone answering machines. Eventually, digital voice recorders should achieve a balance of CD-quality audio with convenient recording times.

Laptop Computer

Recording Pros	Recording Cons	Production Pros	Production Cons
CD-quality audio.	May not be simple to set up pending OS and audio software used.	No time wasted transferring audio files to the computer.	None. You need to use audio editing software anyway to create deliverables—there’s “no middle man” by using a laptop.
Recording time limited only by available space on the hard drive.	Hard drive crash would potentially lose all recordings from a day’s work if files are backed up every night, and the entire project if no back-ups made.	Can start editing immediately.	
Built-in mic on many models that sounds adequate for speech recording (unless laptop cooling fans turn on).	Walking around with a laptop in field study sessions while also taking notes may be inconvenient. Battery failure or computer crash during a session would lose the data for that session.		

Conclusion: If using the laptop’s internal microphone, understand that the recording quality may suffer greatly if cooling fans in the laptop turn on during a session. Plugging an external mic into the laptop and moving the mic closer to the participant can solve this problem.

Costs of Portable Audio Recorders

Consider these two costs when buying any audio recorder: 1) initial cost (the price you pay at the store) and 2) life-cycle cost (which includes blank media, production time and expenses, etc.). How many field studies or lab studies does the usability practitioner perform in a year? Does the practitioner use speech recordings to back up handwritten notes, or to also produce audio highlights for clients? Considering these issues can help the practitioner decide on the most cost-effective portable recording solution. The following table provides costs (based on 2004 catalogs and estimates) for each device and its media.

Device	2004 Purchase Cost	2004 Media Cost
Cassette tape recorder	\$35 - \$75	\$1.25 per 90 minute tape
MiniDisc recorder	\$150-\$300 plus \$25-\$50 for a microphone.	\$1.25 per 80 minute blank MD (160 minutes in mono). The forthcoming Hi_MD format will allow much longer recording times.
Digital voice recorder	\$100-\$300	Some have a fixed amount of time such as 3 hours and others have replaceable memory modules to increase recording capacity. Those memory modules are currently more expensive per minute than cassettes or MiniDiscs.
Laptop computer	\$900-\$3000	Not really an issue—just need free hard drive space.

CONCLUSION

Each person's voice is unique and is one of the most distinguishing characteristics that differentiate one person from another. Regional accents, diction, volume, vocabulary, and other cues that communicate a speaker's personality and state of mind can give clients vivid impressions of a participant in a usability study.

Video production, while easier than ever to do on modern computers, still requires more knowledge and resources than audio recording in this author's experience. In some usability projects, video is essential to achieving a study's goals and fulfilling client needs. However, usability professionals should consider whether professional-sounding audio highlights might be more appropriate for some projects when:

- Budget or schedule constraints make video highlights unfeasible.
- A project is already low-budget, but proposing audio highlights to enliven the results report might swing the bidding process to one's favor.
- Video recording is not allowed on a site.
- A session's video is unusable due to equipment failure or videography errors (poor lighting, wrong shutter speed, bad focus, bad tape, battery failure, lens cap, etc.)

Professional-sounding audio clips are well within reach of people's technical skills if they follow these guidelines:

- Room acoustics affect sound quality at least as much as microphone and recorder quality. Learn to "hear with the eyes" by quickly analyzing rooms for their acoustic properties and testing one's predictions of the "stew" of sound reflections against the actual recordings. Acoustic negatives include lots of hard, reflective surfaces like plaster, metal, glass, and hardwood, as well as noisy ventilation systems and high ceilings. Acoustic positives include carpeted floors, low ceilings, plush furniture throughout the room and furniture along the walls.
- Microphones lack the "Cocktail Party Effect" of being able to focus on speech in an "acoustically murky" room as a person can. Any extraneous noise in a room can hamper a speech recording. Heating or air conditioner sounds, traffic noise, groundskeepers' lawn mowing, and nearby conversations are usually quite bad for a recording.
- In field study sessions where environmental noise may be unavoidable, consider "wiring" the participant with a lapel mic. Make sure during recruiting that candidates won't mind if they are asked to do this during their session.
- Buy a portable recording solution that is usable! Buy from a store that allows penalty-free merchandise returns within a reasonable period.

Whether a portable audio recorder is needed to back up usability test lab equipment or to use on field studies, audio clips are an information-rich resource that can ensure accurate data collection and increase the impact of results reports on clients. The author welcomes contact with readers who may be interested in relating their own portable audio recording experiences in usability studies, or researching the impacts of video versus audio recording on the behavior of participants in usability studies.