

The soundness of documentation: towards an epistemology for audio in documentary linguistics

David Nathan, *Endangered Languages Archive, SOAS*

Developed from a paper presented at the IASA International Conference, Sydney, Australia, September, 2008

Introduction

When I began working with multimedia as a member of a team developing curriculum and teaching materials for Australian Indigenous languages during the mid 1990s, problems in linguistic approaches to audio started to become apparent. It was a time when computers came into general use for research and teaching; the most salient development being the rapid growth of the World Wide Web, but it was also when typical desktop computers began to have seamless multimedia capabilities, no longer needing specialised add-ons and settings to play sound¹. In creating some simple interactive multimedia games for language teaching programmes, I collaborated with linguists who supplied linguistic expertise and audio material, typically from their field recordings. Often, however, field recordings were poor in quality as a result of equipment choices (such as using inbuilt microphones of recorders), recording methodology (microphones far from speakers, or not aimed at the speaker), and an elicitation genre that might be useful for recovering lexical or grammatical information but otherwise containing limited content useful for teaching or attractive to listen to. I drew the conclusion that linguists make field recordings to serve as *evidence*, not *performance*. For an anecdote about how a recording provided evidence of a traditional narrative, even though the narrative was presented in a different form in a grammar (i.e. a notional narrative was constructed), see Nathan 2006b. Even as evidence, audio was an auxiliary; the principal fieldwork products were field notes and the language knowledge absorbed by the researcher. It was as if the primary role of recorded tapes was to provide evidence that the fieldwork had actually taken place.

Following the emergence of the field of documentary linguistics in the late 1990s, audio issues have become harder to ignore. Documentary linguistics, as a response to language endangerment throughout the world, emphasises the collection and representation of a range of language events, where the resulting resources can be drawn on by various disciplines (Himmelfmann 1998). Naturally, audio would appear to be its principal medium. The new field attracted many who were already working on minority and endangered languages, and also caught the imagination of many young scholars, as well as the press, the public at large and funding agencies from which language documentation has attracted funding on a scale virtually unknown in academic linguistics. In the UK, for example, SOAS received a commitment of 20 million pounds from the charity Arcadia to set up the Hans Rausing Endangered Languages Project (HRELP), which has a documentation funding component (ELDP), a teaching component (ELAP) and a digital archive (ELAR; see www.hrelp.org for further information). As the archivist at ELAR, I have been privileged to meet and work with a wide range of language fieldworkers and documenters, especially through training workshops for new grant recipients that we run at ELAR in collaboration with ELAP and ELDP. The audio component of this training has steadily evolved across about 10 workshops, with accruing experience drawn from the participants, from applying a variety of teaching approaches, developments in equipment, and a changing outlook on the role of audio.

¹ Macintosh computers had these capabilities earlier, and were favoured by many linguists, often on the basis of having media capabilities (in fact I "cut my teeth" on Apple's Hypercard). Curiously, despite more or less the same cohort still retaining their loyalty to Apple, most have yet to integrate audio into their research methodology.

The event that crystallised the need for an investigation into audio goals was a one-day workshop run at ELAR in February 2006 by Dr Dietrich Schüller of the Vienna Phonogrammarchiv. Dr Schüller characterised linguists' recording methods – and by implication the quality of the resultant “data” – as unscientific, comparing the typical practice using randomly positioned and inappropriate microphones with the conduct of crucial medical research using cheap room thermometers. Since a microphone is the sensor by which we capture acoustic information about an event, then the validity and quality of the resultant data depends on choosing the right sensor and deploying it properly. I realised that although our starting point had been practical recording issues, specifications of equipment, methodology, and evaluative criteria could be formulated only in the service of goals. There was nothing in the documentation literature to even tell us if we should record in stereo or mono.

Recall the abovementioned contrast between “evidence” and “performance”, where fieldworkers were described as collecting evidence not performances. Actually, audio materials are rarely evinced as evidence for linguistic arguments anyway (except in some phonetic studies). Although Bird and Simons (2003), and Thieberger (2004) have proposed linking audio to examples sentences in grammars and similar texts (and Thieberger published software to do so), such links tend to provide direct evidence of an example's provenance rather than for the linguistic claims made in the texts. There remains a kind of unspoken and unscrutinised methodological space between audio and the written representations based on it. Exceptions exist, such as in the work of Stephen Muecke, who has been credited with innovating writing that “imitated the spoken word” through “joint authorship” between an Aboriginal story teller, Paddy Roe, and Muecke as the transcriber (Zierott 2005:36, Benterrak et al 1984).

In this sense, audio recordings cannot truly be regarded as “data”, despite the frequency with which we hear the expression “my audio data ...”. Data in the sciences refers to measurements or records of phenomena within the terms of a model or domain, where these measurements or records can be applied to reasoning and prediction within those models or domains. For language documentation, recordings are supposed to be multi-purpose, and project goals are rarely those that could be directly evidenced by audio signals².

The ethical dimension

Fieldworkers enjoy an almost unprecedented access to language speakers, and consistently report the generosity of community members³. Simply interposing in the community raises enough ethical issues; seeking to record naturalistic, spontaneous conversation for use by arbitrary others raises far more (Thieberger & Musgrave 2007, Dobrin 2005)⁴.

Not only as beneficiaries but also as participants in recorded events, fieldworkers have unique opportunities to select equipment, its locations, and deployment, and to influence performances – all of which have a major influence on the nature of recordings. Because consultants, their community, their descendants and others have a stake in the products and outcomes, an ethical stance would oblige fieldworkers to mobilise their resources (especially those that the community might not have access to) to create quality recordings. At a minimum, this means using the right equipment and applying the skills to use it well. Sometimes, this may make the work harder, for example carrying more equipment, or holding a

² Goals of ELDP projects include - examining the influence of contact languages, “salvage” of language and culture, dictionaries and grammars, sociolinguistics surveys, and many others. For a comprehensive list, see <http://www.hrelp.org/grants/projects/>

³ I have only ever heard one fieldworker complain that community members were unfriendly and inhospitable.

⁴ Some go to the heart of language endangerment, e.g. diverting elders' time away from using the language with their community.

microphone in a particular position for an extended length of time. But, as filmmaker colleague Simon Atkins challenged our trainee fieldworkers: when either you or the consultant has to suffer discomfort in order to achieve a good recording, it had better be you!

Fortunately, there is something inherently ethical about audio. Compared to the flight to text, capturing and using audio is humanistic and transformational. The original speaker is directly represented and can be recognised as an individual by virtue of the sum of information that is captured – spoken content, distinctiveness of voice, and audio cues about the place and other participants who are present. Audio represents community members as social agents who address listeners directly, rather than as consultants filtered through the research apparatus. It provides an unbroken path between the information provider and the final user, with their performances not transformed to writing or mediated by analysis. As a result, multimedia can provide many connections – social, emotional, intellectual, and pedagogical – between the actors and their listeners (Nathan 2006a).

Text, on the other hand, transforms the language and its relationship to speakers:

Something strange happens when a language is written down. Somehow it no longer belongs to you. It is separated from you. Now what happens when that separate thing seems more real, more important, and more 'correct' than you, the speaker? Do you own the language any more, or has it turned into something which is outside your grasp? (von Sturmer 2009)

This dispossession is compounded by linguistic genres that extract and treat utterances as decontextualised instances of the language system, rather than as socially embedded performances of individuals.

Documenters' audio responsibilities begin before fieldwork, when they acquire and learn how to use equipment. And then responsibilities continue, embedded in the process of negotiating and conducting documentation, not only to ensure that speakers and their community have an input into what is (and is not) recorded, but also to ensure that recordings are made with all the skill required to capture the optimal audio information. In contrast, documentary linguistics' ethical components of audio practice are currently located at the "output" of projects, for example, in "giving back" copies of recordings as "adjuncts or by-products of a 'contract of exchange' between researcher and community" (Dobrin et al 2007). This unfortunately characterises recording as peripheral to the research activity, and locates researchers' ethical responses in the somewhat trivial process of producing and distributing cassettes and CDs.

It is understandable that a previous generation of linguists had low expectations of audio recording. The analogue (tape) equipment they used was vastly inferior to even moderate priced digital recorders that are available today. The enormous weight and battery consumption of reel-to-reel and even some cassette recorders must have made remote fieldwork feel like torture. Recordings on tape clearly had ambiguous value as audio resources; for example linguists undertaking AIATSIS-funded fieldwork (then AIAS) were told to re-use tapes after transcribing them, and not to record narratives⁵. It is understandable that participants at IASA's 2008 Annual Conference received miniature bouquets (made out of a loop of cassette tape) to celebrate the demise of analogue tape!

⁵ pc Luise Hercus.

The continual appearance on the market of new, better, and smaller digital recorders is a boon to documenters. But it will be a loss to future language documentation if only their compactness and convenience are exploited. Instead, they provide an opportunity to review goals and techniques, for example, by taking advantage of weight savings to professionalise kit with better microphones, cables and stands. In the widest sense, the recently-completed transition to born-digital audio workflow means that a raft of obstacles to producing good audio have been removed, thereby increasing the onus on documenters to formulate more ambitious roles for audio in the preservation of languages.

Towards an epistemology

Sometimes we start audio training sessions by asking “who has recorded audio?” Of course, most participants indicate that they have. But to the next question “who has published audio”, few people put up their hands; some even appear quizzical about the nature of the question. The products of linguistic research and documentation remain focused on text; audio is rarely published or mobilised for any linguistic purpose (except for the occasional online sample, or “giving back” CDs to the consultants and the language community). Sometimes fieldworkers say that they make recordings for the purpose of archiving, which merely begs the question of usages resulting from preservation and dissemination by archives. Put simply, audio is an inconvenience on the way to transcription, annotation, selection or analysis⁶.

The characterisation of audio as simply an inconvenience on the way to text is a claim that documentary linguistics is missing a crucial component in its conceptualisation. While this paper is not proposed as a philosophical contribution, ‘epistemology’ gives a close match to its aims. Alex Barber’s *Epistemology of Language* describes an epistemology as a framework that would:

- help to “make decisions on how to investigate the phenomena of language”
- tell us how communication is possible “in the same sense that an interest in, say, our capacity to know facts about the physical objects around us through perception is epistemological” (Barber 2003a: 1)

While most of the papers in Barber’s book take a mentalistic perspective and do not consider sound at all, I use ‘epistemology’ as a placeholder for this missing component, the role of audio phenomena in documentation. It may eventually help to understand how the snippets of acoustic realities that we record can inform the other concerns of documentation.

An epistemology for audio is not proposed just as an extension in thinking about documentation. The *absence* of an epistemology has had detrimental and sinister effects on our practice and its outcomes, arguably heightened by contexts of language endangerment where the recordable linguistic events are in decline, and less likely to be observed again. Without desiderata for what makes a relevant and effective audio, any methodological discussion, advice, or training remains ungrounded.

⁶ Of course the need to publish for employment and career reasons sets priorities for many linguists, and the narrow range of publications recognised by academia is part of the problem. But not all of it: if linguists do not challenge this narrow view of language, who will? In addition, linguists are increasingly funded to, or choose to, pursue language documentation, where such “traditional” priorities do not necessarily hold. So we might have expected that the new genres for expressing knowledge about languages would arise from the field of documentary linguistics, although so far this is not occurring.

And indeed it seems that “almost anything goes.” Sometimes, an uninformed opinion will do, such as the belief of one linguist that a \$2 microphone was appropriate because his recording environment was so noisy anyway. Leaders in the field advance arguments based on simple pragmatism (e.g. that video should supplant audio now that it has become affordable), or sweeping statements that because technology exists, linguists would be “stupid” not to use it (Himmelmann 2009). For many, cursory knowledge about technical parameters of digital audio have become hallmarks of “best practice”, but are really trivial proxies for proper training, skills and experience (I called this narrow and semi-religious devotion to numbers and rules ‘archivism’; see Dobrin et al 2007). “Best practice” guidelines have made fieldworkers worry about digital *resolution* (ultimately just a matter of recorder settings) instead of *signal to noise ratio*, which has a far greater influence on the properties and value of a recording and is far more complex to understand and manage. The same guidelines counsel – wisely – against data compression, but only of the digital type (e.g. MP3), without warning of the far greater information loss incurred in capturing only a fraction of the available acoustic information. Such technological diversions have suppressed understanding audio recording as both a science and an art requiring appropriate training, experience, and talent. Overall, this sad misunderstanding and neglect of audio amounts to a dangerous de-intellectualisation (for more on dubious assertions about audio see the following section).

Taking up the challenges

At the same time as addressing the need for an audio framework for documentary linguistics, there are also widely-held assumptions that warrant challenge.

For example, we often hear it protested that there is not enough time to set up equipment in the recommended ways because events are too transitory and must be recorded without delay. But in most cases this amounts to a claim that the fieldworker feels no obligation to be properly trained and prepared. Documentary filmmakers, for example, are trained to reconnoitre a situation and prepare for it so that they can begin recording with minimum delay. Many of these cases could be dealt with by simply asking speakers to wait or to tell a story later – when the roosters have stopped crowing, for example – a strategy that depends not on technical abilities but on human skills and the quality of the relationships built up between the fieldworker and consultants. In any case, that “unmissable” event was only unmissable because the fieldworker was present (it may have occurred a week before, or a week after the fieldtrip, for example). It seems that what is at stake is not the event itself but the opportunity to record it, and an inadequate recording may count equally as a lost opportunity. Is there some kind of inverted observer effect here, an over-valuing of the significance of the documenter’s presence?

Another frequently heard claim is that equipment is intrusive and distracting. This is often invoked as an argument for using a recorder’s internal microphones, i.e. for avoiding the use of well-positioned external microphones. Here again is an odd twist on the observer effect: is the presence of a microphone enough to tip the methodological situation into difficult territory, without considering *anything* about the activities of the documenter? Some researchers, in fact, have argued that the tangible presence of media equipment adds to the theatricality of events and can be of assistance in eliciting several kinds of performances⁷.

Video, being a visual medium, captures location in a more concrete way than audio does⁸. Nevertheless, as embodied listeners with two ears, audio can also provide us with spatial

⁷ pc. Anthony Jukes.

⁸ The relationship between audio and video, and the role of audio in video, are important topics for documentation but are beyond the scope of this paper.

information that could never appear in video images, such as the location of sources out of frame, and the subtle audio cues that convey the nature of the recording space. Recently there was a debate in several forums about the role of video in language documentation (some of the discussion appears in *Language Archives News*)⁹. I challenged the increasing trend among documentary linguists for using video, much of which seemed to be of dubious value (while very “expensive” in terms of equipment, methodological issues, processing and storage). Others responded with good reasons for the potential value of video, and yet, looking back at them in the context of the present paper, many were arguments for spatial information, not video *per se*. Examples include identifying the speaker in multi-person conversations, capturing emotions and other paralinguistic meanings, and portraying the setting; all of which can be achieved in audio (to a greater or lesser extent than moving image). Despite the undeniable value of video in language documentation, could it be that video has been so enthusiastically proposed and adopted in order to make up for inadequacies in audio recording?

What is most interesting about these various assertions is that they are often put forward as self-evidently true, without evidence, thus providing further examples of the de-intellectualisation of audio.

Audio and events

Audio can be thought of as acoustic information that is increasingly compressed as its traces move along a five-part chain (of which only the first two will be discussed¹⁰):

event > recording > representations > data > abstractions

Documentary audio results from real-world events – not, for example, from artificially generated signals, or edited together from unconnected sounds. It might be a safe bet to take those events to be spoken utterances¹¹ because documentary linguistics aims to create “a multipurpose and comprehensive record of the *linguistic practices* characteristic of a speech community ... [where] the emphasis is on the collection and representation of *primary data* rather than theory and analysis” (Himmelman 1998; emphasis mine).

The relation between an event and a recording is mediated by the physical capabilities and location of the equipment, especially the microphone/sensor(s). But it is not as simple as this. Firstly, those physical factors are considerably modulated by the recordist through his/her selection and deployment of the equipment, and his/her explicit or implicit influence on the acoustic sources¹². Secondly, “linguistic practices” are generally characterised as instances of genres (Johnson & Dwyer 2002). Although some genres, such as song, may have specific acoustic characteristics, in general genres are *not* properties of the recording but the result of listener interpretation.

Thirdly, “other” sounds in the environment might be relevant too, such as clapping or noises whose sources are topics of conversation. This issue regularly arises in our training sessions,

⁹ This author’s initial article “Digital Video in Documentation and Archiving” (http://www.mpi.nl/LAN/issues/lan_09.pdf), and replies from Patrick McConvell and Peter Wittenburg in the following issue (http://www.mpi.nl/LAN/issues/lan_10.pdf).

¹⁰ The other levels are of less interest here – *representations*, typically symbolic, in the form of phonetic or orthographic representations of instances of the linguistic system understood to be associated with the audio; and *data* and *abstractions* which depend on theories and formalisms which give significance to the symbolic representations.

¹¹ Assuming the case of spoken, rather than signed, languages.

¹² The common label of “observer effect”, referring to the influence on the performers resulting from their awareness of being observed or recorded, is an oversimplification of what really happens in fieldwork situations.

where participants ask how to record in situations where there are constantly chickens clucking, insects buzzing, or craftsmen hammering. While we can show how to optimise recordings under these conditions, this is not really what is at stake. It is as if the question of what counts as focal to a recording and what counts as extraneous is simply left to be dealt with under the rubric of recording technique – as a mere technical matter – due to the absence of linguistic, methodological, and philosophical principles. Whether a child's crying, a bird's tweeting, or the grating of a saw is relevant depends on a large number of factors, each non-trivial, including the documentation goals, the social setting, topics of conversation, and personal viewpoints; and most of these may also change moment-to-moment.

Audio training at ELAR

We have tried to address some of these issues in our training courses for newly ELDP-funded language documenters, where we allocate a little over one day to audio. Although clearly not enough time to do justice to the area, it is the best we can do within the constraints of a six-day programme; nevertheless we have heard from almost all participants that it far exceeds any audio training they have ever previously received.

The training sessions have evolved, in particular with decreasing attention to digital audio topics (signal encoding, digitisation etc) in favour of more perceptual and skill- and practical-oriented activities. This change has corresponded to the rise of the concerns discussed in this paper, at the same time as solid state recorders have become equipment of choice and fostered basic literacy in digital encoding issues¹³.

A major theme is developing critical listening skills (Alten 2005: 9). We examine *signal* (what you want to capture in a recording) and *noise* (what you don't want) from several perspectives, which provides a holistic integration of:

- equipment issues (e.g. selection, compatibility);
- moment-to-moment and situation-to-situation skills of managing equipment, settings, participants, and the physical space to maximise signal to noise ratio (e.g. how to capture a speaker's voice against background noise);
- quality: what counts as a good recording?
- wider linguistic and ethnographic issues that decide what constitutes a soundscape containing all elements crucial to understanding the event and its linguistic content (e.g. did that voice come from another room? is the sound of that crying child "signal" or "noise"?).

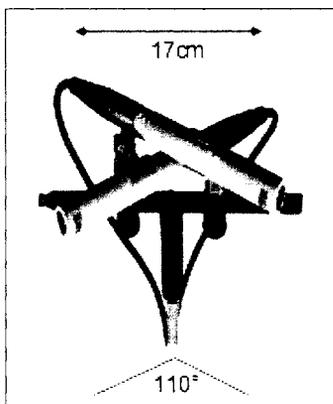
The *goals* of recording ought to be the source of criteria for classifying what is signal and what is noise, which in turn enables the mobilisation of skills to achieve a good recording. Without recording goals, or the corresponding recording skills, results can only be hit and miss. At a workshop conducted by the author and Peter Austin at the Tokyo University of Foreign Studies in 2008, participants were invited to give feedback about the audio sessions. Some offered the honest and revealing response that until the workshop, they had never even *considered the possibility* of managing the recording process to attain better results. They had thought that all they could do was switch on the recorder and hope for the best. Why? Because they had never been exposed to any audio goals or criteria. For them, the workshop had been a happy revelation.

¹³ Or rendered them simple and accessible as menu settings of recorders. I estimate that the proportion of fieldworkers using solid state recorders has increased from around 10% (5 years ago) to 100% today.

The apparatus we use for training includes a set of chained amplifiers and headphones that allows all participants to listen simultaneously, a portable stand to hold dampening material (which is a sleeping bag!), CDs of pub noises, pre-recorded audio of various types, and a range of recorders, cables and microphones. Sometimes, groups of three or four are sent out to various locations to make short recordings. Later, the whole class listens to them, trying to correlate their strengths and weaknesses with the equipment, techniques, sources and locations that the group had used. More recently, we try out various configurations of equipment, props and volunteer speakers “live” in the classroom while participants listen using the headphone system. This has proved extremely effective, as participants are more likely to be convinced by the incontrovertible evidence in front of their eyes as they hear the effect of, for example, changing between a lavalier microphone and a shotgun microphone when monitoring the sound of a speaker standing in front of a window onto a busy street. Participants can make suggestions and can be set problem-solving tasks, with immediate feedback of results. More generally, the method drills the crucial importance of monitoring when recording.

We recently added the topic of capturing spatial information, covering basic psycho-acoustics, and listening to audio from stereo and ORTF microphones. Encouraging groups to experiment with spatial recording provides a range of learning opportunities. For example, one group recorded an interview in a noisy environment using a stereo microphone (RØDE NT4, XY type). When asked later which way they had aligned the stereo axis, they admitted they had not thought about it at the time (it would have made a good item of metadata – see below). Actually, they had aligned the two speakers (interviewer and interviewee) off the stereo axis, i.e. achieving no separation between them. Nevertheless, we found this to be a useful strategy. The competing background noise could be separated from the interview, which makes for a more comfortable-to-listen-to and easier-to-transcribe recording than one using the other axis to separate the two participants from each other but not from the background noise.

To achieve a full 3-dimensional “spatial illusion” when listening back using headphones, a specifically configured pair of microphones can be used (Alten 2005: 24). The configuration shown in the diagram is known as ORTF¹⁴.



¹⁴ From ‘Office de Radiodiffusion Télévision Française’, who invented it.

Participants hear and discuss several ORTF examples: pre-recorded conversations, fieldwork examples made by the author, and “live” monitoring as described above. We also performed an informal experiment, which involved recording an interview against a background of multiple conversations. We compared two versions of the recording: a full-resolution mono version, and several degraded ORTF versions (degraded by applying various levels of MP3 compression). We found that even significantly compressed ORTF-recorded versions were preferable to listen to, because they still provide enough separation of the sources to allow a listener to engage with the content. The mono sound-stage, despite its *prima facie* higher quality, collapses all the conversations into a single space and leaves the listener continually confused and unable to focus on the interview.

Preliminary results are that our training participants agree that, by using ORTF:

- separation and localisation can be achieved;
- more knowledge about the recording environment is captured;
- on the other hand, the richness of captured information can sometimes be distracting¹⁵ and recordings made in some environments are very difficult to listen to¹⁶.

We are not at this stage advocating that fieldworkers use ORTF, since more work needs to be done on understanding its properties, and the setup is somewhat unwieldy; nevertheless it has proved a good way to illustrate the potential for spatial audio and how much information is lost if it is ignored.

The final training theme is metadata. Metadata is commonly defined as *data about data*. Its function is to provide the keys to managing, understanding, identifying and retrieving data (OAIS 2002). It also thus “defines and constrains the audiences” for audio resources and how they can effectively use them (Nathan & Fang 2009). In current language documentation practice, metadata for audio recordings typically consists of information about the location of the recording, and information about the speakers – their names, sex, age etc. Less frequently, fieldworkers note technical details such as equipment type and settings. Virtually no fieldworker notes down the information that can be crucial for a spatial characterisation of the event – how the microphones were arranged, their relation to the sound sources, and the layout and nature of the nature recording space. Even simple information about which speaker is found in which track in stereo recordings is omitted¹⁷. Diagrams and photography would be useful tools for some of these categories of metadata.

Psycho-acoustics and spatial information

Psycho-acoustics is the study of human perception of sound. Much of it is concerned with our sophisticated ability to understand the physical space we are in through the audio information we receive via our two ears. A mono recording can also convey an impression of a space; for example echo with a certain delay will suggest a large acoustic space, and relative intensities suggest how close or far a source is from the microphone. Nevertheless, a source cannot be localised in the sense of knowing where it is within a 3-dimensional “sound stage”.

Humans in real listening situations experience “spatial or binaural localization” by using two ears “to localize a sound source within an acoustic space” (Huber and Runstein 2005: 62).

¹⁵ A minority of trainees found that the increased life brought to the recording by ORTF made it distracting for them. This may be due to the novelty of this method and may be overcome if more frequently experienced.

¹⁶ A recording made in the domed plaza of the British Museum was very disorienting. It seems that there is exaggeration of some kinds of echo.

¹⁷ This information is likely to be found in a technical transcription format such as ELAN or Transcriber, but these require special software and skills, and will not be accessible to an array of listeners who simply want to know which speaker appears in each channel.

This takes into account not only sounds received directly from sources but also those reflected from passive objects in the acoustic environment. Walls, floors, windows, plants, furniture, and human bodies all modify and reflect sound, thus contributing to the amount, quality and duration of sound reaching the ears.

Aural processing involves the ears *and* the brain¹⁸. We interpret the space around us by comparing and analysing the following properties of sounds reaching each ear, and the differences between them¹⁹:

- intensity – due to different distances travelled, and the falloff of energy according to the inverse square law, each ear receives sound of different intensity;
- phase/delay – due to different distances travelled, sound reaches each ear at slightly different times;
- frequency falloff – higher frequencies lose energy faster than lower frequencies, so sounds travelling different distances have different frequency distributions;
- frequency colouration – sounds reflected off different materials (and cumulatively from multiple reflections) will have different frequency distributions.

Audio information is processed in the context of the listener's knowledge: his/her transient knowledge (gained through any of the senses) of the current environment (e.g. location, orientation and changes in audio sources), as well as long term experience, as an embodied actor in the world, of how perception is influenced by the nature of sources, materials and spaces. The fact that processing takes place under the guidance of long and short term experience accounts for phenomena such as quickly losing awareness of the presence of fans, traffic, or even chickens, when listening to somebody speaking.

At a conscious level, we can be aware of and direct our attention to particular sources. This underlies what is commonly called "the cocktail party effect",²⁰ the ability to pick out the speech of one individual even in a crowded and noisy environment. This effect can be regarded as a showcase of human listeners' capacity to mobilise spatial information for useful functions.

Lost in space

The preceding section described the massive amount of spatial information available to a listener, and how listeners use this in everyday life. But how much of this information can be recorded? With good equipment and techniques, much of it can be captured in a recording. The word "captured" is important here because perceived spatiality is not inherently present in a recording. A recording can only make the information available to a listener capable of interpreting multiple informational cues to experience a "sound stage" resembling the recording environment. The ORTF method of achieving this was briefly described above. However, the science and art of stereo and other types of spatial recording are large and complex topics beyond the scope of this paper. My aim here is to argue that spatial information cannot arbitrarily be ignored and that it may be invaluable in language documentation.

Spatial information appears irrelevant to the documenter who aims to move quickly to transcription and to work only in text. This workflow involves a *loss* of information. How

¹⁸ It also involves transmission through the body and the head, and high level integration with other senses such as vision.

¹⁹ More spatial information is available (through triangulation) if the listener - or any other object, whether emitting, reflecting or absorbing sound - is moving.

²⁰ Also known under the more proletarian label "the cafeteria effect".

much information is lost? Let us roughly estimate the quantities over a 5-second utterance²¹:

Information type	Bytes of information in 5 sec speech
acoustic	44.1 KHz x 16bit x 2 (stereo) x 5 sec = about 900,000 bytes
transcribed	3 (syllables/sec) x 2 (bytes/syllable) x 5 sec = about 30 bytes

Losing information is not in itself a bad thing. Information theory tells us that losing information is the essence of moving from information to understanding; as long as it is the *right information* that is discarded. The documenter who quickly abandons audio in favour of text eliminates masses of information (over 99.99% of it!). It is unlikely to matter to him where that information is lost, whether at the audio sensor (e.g. through poor choice or use of microphones), the recorder (e.g. through incorrect settings or compression), subsequent processing (e.g. conversion to mono or different resolution), or, indeed, poor reproduction for listening (e.g. listening through cheap computer speakers). None of these deficiencies impacts the outcome of this documenter's work – until a teacher, community member, researcher or multimedia producer comes along with a project that requires good quality, listenable audio, or audio that accurately portrays the whole of the recorded event.

Putting it together

This wider informational and psychoacoustic background helps to understand common problems in recording. Many documenters are surprised to find that the audio recordings they made were spoiled by the presence of extraneous noises. All of them, of course, had been present in the recording environment, but had been psycho-acoustically “filtered” at recording time²². This is only one of an unlimited number of ways in which a recording can fail to convey the original acoustic experience.

The extent to which a recording counts as a “spoiled” or inferior version of the original event depends on a number of factors, many of them subjective and depending on the purposes for listening. But there are also objective factors based on the types and amounts of information that were present for a listener in the original setting and whether they are accessible to someone listening to the recording:

- the acoustic (especially spatial) information in the recording environment;
- the listener's knowledge.

These have very different implications for the eventual listener. If acoustic information is missing (or distorted) the listener will experience the event differently. We have seen that spatial information can be the most important component, because the ability to separate simultaneous events is crucial for intelligibility and for sustained listening without disorientation.

²¹ Assumptions: acoustic information is quantified on the basis of CD-audio, which is extremely conservative; speech is at the rate of 3 syllables per second; a syllable can be written as 2 characters, each of which is 1 byte in size.

²² This class of problem can generally be avoided by monitoring the recording through closed headphones, which forces the fieldworker to “hear” from the perspective of the microphone(s), rather than as a human participant. But this may not be feasible if the fieldworker needs to elicit or converse with consultants.

While a good recording can capture most of the acoustic information, a listener to a recording can never exactly replicate that of an event participant, even if only for the fact that those participants knew about the location and what was happening in it before the recorder was switched on²³. Thus, the extent to which listening can correspond to original experience is dependent on who is doing the listening and on their knowledge about the participants, location, and history of the original event.

On listening

Throughout this paper, the 'listener' has been a key concept. Until now, documenters have been under the mistaken belief that they can think about recordings in terms of what they contain, instead of having significance only for a listener. Of course, if no-one ever listens to them, their only significance is as a memento of fieldwork. Once we factor listening – and listeners – into documentation, we can talk about what we record a particular event *as*, for example as a performance of a story, as evidence for a syntactic or phonetic phenomenon, as a teaching resource for children etc. In each case, we can hope that listeners can have a useful experience, without naively assuming that we are directly delivering a specific content. We also acknowledge that the act of recording constructs its listeners, whether imagined or not, because, just like video, an audio recording imposes a point of view that “constructs knowledge about its subjects as ‘others’” (Kheshti 2009:15). Kheshti notes that:

the positionality from which sound recordings are produced, and the aural perspective that recordings attempt to elicit, enables us to ask: what kind of sonorous body is being materialized through these production techniques and what kind of listener is being produced?

The idea of recording for listeners is, of course, as central to the music industry as it is novel for documentary linguistics. For us, it opens up new ways of thinking. For example, consider the capacity for the “cocktail party effect” discussed earlier. This capacity declines with age and is particularly affected by hearing disabilities. It means that a recording which insufficiently enables a listener to pick out the focal speaker from background talk could be classed as a recording “as heard by a hearing impaired person”.

And there is a qualitative property we could call “listenability”. As a result of group listening to many recordings in training sessions it is clear that people typically agree about listenability. For example, two recordings that are equally intelligible and clear can differ significantly according to how comfortable or pleasant they are to listen to, and most people concur in judgement. Given that language documenters are likely to be the greatest listeners to their recordings (since transcribing an hour of audio can take 50 or more hours of listening) it is a valid part of a research methodology to aim to make recordings that are comfortably and sustainably listenable over long periods using headphones²⁴.

Here is another example. Recently a documenter, Carolina Aragon, was explaining her difficulty in recording the Akuntsú people of the Amazon because their forest environment seems to be perpetually full of loud bird and animal calls (and she believes it would not be safe to record them elsewhere). She asked about techniques to overcome these “noises”, but seemed to have already tried almost every suggestion I could offer. However, there was a greater lesson in this: since the Akuntsú people hear their language in this noisy soundscape, it raises interesting linguistic questions about how those speakers and listeners deal with

²³ Metadata might supply some of this knowledge or “missing” information.

²⁴ Documenters do ask for advice about suitable headphones for sustained listening, but not about how to record.

it. Further, it tells us that thinking about what we seek to do in recording, and therefore how we record, is a relevant and important part of any investigation into the acoustic phenomenon we call spoken language.

Conclusion

This paper has shown that audio is a rich, complex, and rewarding component of language documentation. Much more thinking needs to be done regarding its role in documentation and related areas of linguistics.

The practical and aesthetic aspects discussed could be summarised as a set of desiderata for evaluating recordings:

- accuracy: how well is the signal captured, as true to its sources and without distortion?
- intelligibility/information accessibility: can the desired content be identified?
- signal vs. noise: is the ratio acceptable? Can the focal source be separated from all sources of noise?
- listenability/comfort/aesthetics: is it easy on the ears? Will it be debilitating to listen to for an extended time?
- localisation of sources: is enough spatial information captured?
- separation of noise sources: can all the sources of noise be separated?
- representation of environment: are the acoustic properties of the recording space appropriately represented?
- content (identity, performance, uniqueness, coverage): were the right people recorded doing the right things?
- editability/repurposeability: is the recording suitable for turning to relevant purposes?

Finally, the broader aim of the paper is to stimulate discussion about the goals and purposes of audio in language documentation, and, as an initial contribution to an epistemology, I offer the following:

- an audio recording is made in order to be experienced by a human listener;
- an audio recording conveys what a human listener would experience at a particular location in an event setting;
- the documentation goal defines the audio recording methodology;
- ethical recording respects language speakers and honours their contribution through application of effort and skill;
- a recording should capture spatial information;
- metadata about the recording and the recording setting are required for full interpretation.

References

- Alten, S. 2005. *Audio in media*. Belmont CA: Thomson.
- Barber, Alex. 2003a. Introduction. In Barber (ed).
- Barber, Alex (ed). 2003b. *Epistemology of language*. Oxford: OUP.
- Benterrak, Krim, Stephen Muecke & Paddy Roe. 1984. *Reading the Country*, Fremantle: Fremantle Arts Centre Press.
- Bird, Steven and Simons, Gary. 2003. Seven Dimensions of Portability for Language Documentation and Description. *Language* 79, 557-582.
- Dobrin, Lise. 2005. When our values conflict with theirs: linguistics and community empowerment in Melanesia. In Peter Austin (ed) *Language Documentation and Description*, Volume 3. London: SOAS. 42-52.
- Dobrin, Lise, Peter Austin & David Nathan. 2007. Dying to be counted: commodification of endangered languages in documentary linguistics. In Peter Austin, Oliver Bond, & David Nathan (eds) *Proceedings of the Conference on Language Documentation and Linguistic Theory*. London: SOAS. 59-68. [Online at http://www.hrhelp.org/publications/ldlt/papers/ldlt_08.pdf]
- Himmelman, Nikolaus. 1998. Documentary and Descriptive Linguistics. In *Linguistics* 36, 161-195.
- Himmelman, Nikolaus. 2009. Linguistic data types and documentary linguistics. Plenary at the First International Conference on Language Documentation and Conservation, Manoa, Hawai'i. 12 March 2009.
- Huber, David & Runstein, Robert. 2005. *Modern Recording Techniques*. Sixth edition. Burlington MA: Elsevier.
- Kheshti, Roshanak. Acoustigraphy: Soundscape as Ethnographic Field. In *Anthropology News*, April 2009, p15.
- Johnson, Heidi & Dwyer, Arienne. 2002. Customizing the IMDI Metadata Schema for Endangered Languages. In *Proceedings of The International Conference on Language Resources and Evaluation 2002*. [Online at <http://www.mpi.nl/lrec/2002/papers/lrec-pap-05-JohnsonDwyer.pdf>]
- Nathan, David & Fang, Meili. 2009. Language Documentation and Pedagogy for Endangered Languages: a Mutual Revitalisation. In Peter Austin (ed) *Language Documentation and Description*, Volume 6. London: SOAS.
- Nathan, David 2006a. Thick interfaces: mobilising language documentation. In Joost Gippert, Nikolaus Himmelman and Ulrike Mosel (eds.), *Essentials of language documentation*. Berlin: Mouton de Gruyter. 363-379.
- Nathan, David. 2006b. A Talking Dictionary of Paakantyi NSW. In Laurel Dyson, Max Hendriks & Stephen Grant (eds) *Information technology and Indigenous People*. Hershey PA: Idea Group. 200-204.
- Thieberger, Nicholas. 2004. Documentation in practice: Developing a linked media corpus of South Efate. Peter Austin (ed). *Language documentation and description*, Volume 2. London: SOAS. 169-178.
- Thieberger, Nicholas & Simon Musgrave. 2007. Documentary linguistics and ethical issues. In Peter Austin (ed) *Language Documentation and Description*, Volume 4. London: SOAS. 26-37.
- von Sturmer, John. 2009. Language matters, in *Voice of the Land* (newsletter of the Federation of Aboriginal and Torres Strait Islander Languages), Issue 39 (2009), p 12.
- Zierott, Nadja. 2005. *Aboriginal Women's Narratives: Reclaiming Identities*. Piscataway NJ: Transaction.