Sound and unsound practices in documentary linguistics: towards an epistemology for audio

David Nathan

1. Introduction[1]

I first noticed problems in linguistic approaches to audio 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. It was at a time when computers came into general use for research and teaching; the most important development being the explosive influence 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 sound.[2] In the process of creating simple interactive multimedia games for language teaching programmes, I collaborated with linguists who supplied audio materials, typically excerpts from their field recordings. Often, however, these field recordings were poor in quality as a result of three factors:

(a) equipment choices (such as using inbuilt microphones of recorders);
(b) recording methodology (microphones placed far from language speakers, or not suitably aimed);
(c) an elicitation genre neither attractive to listen to nor containing much content suitable for using in teaching.

I drew the conclusion that linguists make field recordings to serve as **evidence**, not **performance** (for an anecdote about how a field recording provided evidence for a traditional narrative, even though the published written narrative did not correspond to the actual recording, see Nathan 2006b). Even as evidence, audio was auxiliary, a kind of side-effect; the principal fieldwork products being field notes and the language knowledge absorbed by the researcher. It was as if the main 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, such audio issues have become harder to ignore. Documentary linguistics, as a response to language endangerment throughout the world, emphasises the collection (i.e. recording) and representation of a range of language events, where the resulting data can be drawn on by various disciplines (Himmelmann 1998, Austin 2010a). 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 from the Arcadia Fund to set up the Hans Rausing Endangered Languages Project (HRELP), which has a documentation

[1] An earlier version of this paper appeared as Nathan 2009. I am grateful to Tjeerd de Graaf, Lise Dobrin and David Nash for helpful discussions, and in particular to Tom Castle for his help in developing some of these ideas.

[2] Macintosh computers had these capabilities earlier, and were favoured by many linguists, often on the basis of having (multi-)media capabilities (in fact I ‘cut my teeth’ on Apple’s Hypercard). Curiously, despite many of the same cohort retaining a loyalty to Apple computer today, they have still not integrated audio into their methodology in any significant way!

funding component (ELDP), a teaching and research 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 ELDP grant recipients that we run at ELAR in collaboration with ELAP. The audio component of this training has steadily evolved across about 10 workshops, with accruing experience drawn from applying a variety of teaching approaches, developments in equipment, the participants' feedback and experiences, and from a changing outlook on the role of audio.

The event that firmly crystallised in my mind a 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. 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 conducting crucial medical research using cheap room thermometers. Since microphones are the sensors by which we capture acoustic information from an event, then both the quality and the validity of the resultant data depend on choosing the right sensor and deploying it properly. I realised that although our training courses included topics such as audio equipment, methods, digitisation and evaluation, all of these actually needed to be understood in the context of clearly-articulated goals. But there was nothing in the documentation literature to even tell us if we should record in stereo or mono, let alone to help us to choose equipment, learn methodologies, or formulate evaluation criteria. These issues, no matter how practical, could not be addressed without clearly stated goals for recordings and the role(s) of the resultant 'data'.

Recall the abovementioned contrast between 'evidence' and 'performance', where typical fieldwork was described as the collection of evidence, not performances. As Schüller showed, even as evidence, the typical audio results were pretty poor. And even worse - and paradoxically - 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 example sentences in grammars and texts (and Thieberger has published software to do so[3]), such links provide direct evidence only of those examples' provenance, not for the linguistic claims made about the examples.

There remains, then, an unscrutinised methodological space between audio and the written representations based on it.[4] Audio recordings cannot truly be regarded as 'data' in the normal scientific sense, 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. But it turns out that language documentation projects rarely have goals for which audio signals serve as evidence.[5]

[3] See http://www.linguistics.unimelb.edu.au/thieberger/audiamus.htm.

[4] 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).

[5] For example, goals of ELDP-funded 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/.

2. An ethical dimension

Fieldworkers enjoy almost unprecedented access to language speakers, and consistently report the generosity of community members.[6] Their mere presence in a community raises enough ethical and methodological issues (see Austin 2010b); seeking to record naturalistic, spontaneous conversation for use by arbitrary others raises far more (Thieberger & Musgrave 2007; Dobrin 2005).[7]

Fieldworkers are not only beneficiaries of the events they record, they are also participants in them. But unlike the other participants, they typically have opportunities to obtain good equipment and determine its deployment. Since good equipment and techniques are a major influence on recording outcomes, it could be argued that ethical researchers must at the very least mobilise their advantages and opportunities by acquiring and using the right equipment and skills in order to create quality records of the language for a variety of purposes. Choices must made in pursuit of excellence of recordings, not the researcher's convenience; e.g. more or heavier equipment may need to be carried, or discomfort endured in holding a microphone in the best recording position for an extended length of time. As filmmaker colleague Simon Atkins[8] challenged our ELDP trainee fieldworkers: if it is either you or your consultant who has to suffer to achieve a good recording, it had better be you!

Ultimately, using advantage, skills and effort is simply a way of paying respect to speakers, their knowledge and their contributions. Fortunately, there is a return on making these investments because there is actually something inherently ethical about audio. Compared to the linguist's typical flight to text, capturing and using audio is humanistic and transformational. The original speakers are directly represented; their identities are preserved through the totality of information captured, not only through spoken content, but also through their distinctive voice quality and audio cues about the location and other participants who are present. Audio thus establishes community members as social agents who address listeners directly, rather than as consultants who supply 'data' filtered through the research apparatus. Audio provides an unbroken path between the information provider and the final users, without speaker performances reduced to writing or mediated by analysis. As a result, multimedia resources 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, 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)

[6] I have only ever heard one fieldworker report that community members were unfriendly and inhospitable.

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

[8] See http://www.simonatkins.com/.

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

More broadly, then, documenters' audio responsibilities begin well **before** fieldwork, when they need to acquire equipment and learn how to use it skilfully. And the responsibilities continue, embedded in the process of negotiating and conducting documentation, not only to ensure that speakers and their community have a say about 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 (what counts as 'optimal' is discussed below). However, current discussions of ethical conduct in documentary linguistics usually describe it as located at the **output** end of the research, for example, by 'giving back' copies of recordings as 'adjuncts or by-products of a 'contract of exchange' between researcher and community' (Dobrin et al 2007). This makes ethical action peripheral to documentation rather than central to it, and has consequences such as constraining ethical responses to 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 the moderately 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 were sometimes regarded as having transitory value only; for example linguists undertaking AIATSIS-funded fieldwork (then AIAS) were instructed to re-use tapes after transcribing them, and not to record narratives.[9] It is understandable why participants at IASA's 2008 Annual Conference wore miniature bouquets made out of a loop of cassette tape to celebrate the demise of analogue tape!

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, e.g. by taking advantage of weight savings to professionalise equipment 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.

3. Towards an epistemology

In our training courses at SOAS, we sometimes start 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 meaning of the question. The products of linguistic research and documentation remain focused on text; audio is rarely published or disseminated for any linguistic purpose (except for the occasional online sample, or 'giving back' CDs or cassettes 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 what usages result from people accessing what is preserved in archives.

[9] p.c. Luise Hercus.

Put simply, audio is presently seen as an **inconvenience** on the way to transcription, annotation, selection or analysis.[10]

This characterisation of audio as simply an inconvenience on the way to text is a way of identifying a missing component of the theory and practice of language documentation, a component that could be called an **epistemology** for audio. Barber (2003a:1) describes a language epistemology as a framework that would help to 'make decisions on how to investigate the phenomena of language', which captures quite well the spirit of the investigations that Schüller sparked.[11] Whatever the merits or otherwise of using the term 'epistemology', it is used here as a placeholder for the missing component: the role of audio phenomena in documentation. It may eventually help us to understand how the selective acoustic realities that we record contribute to a more complete characterisation of language usage and language knowledge.

The **absence** of such an epistemology has had detrimental and sinister effects on documentation practice and outcomes. Lacking desiderata for what makes relevant and effective audio, much previous work may prove to be inadequate. And if it does, it will be unforgivable in the context of language endangerment where recordable linguistic events are ever less likely to occur again.

And indeed it seems that presently 'anything goes.' Sometimes, a completely uninformed opinion will do, such as the claim of one linguist that a $2 microphone was appropriate for his project because his recording environment was so noisy anyway. Even leaders in the field advance arguments based simply on pragmatism (e.g. that video should supplant audio now that it has become affordable), or sweeping statements that just because particular technologies exist, linguists would be 'stupid' not to use them (Himmelmann 2009). For many, cursory knowledge about technical parameters of digital audio have become hallmarks of so-called 'best practice', while they are actually just trivial proxies for proper training, skills and experience (I have 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 value of a recording but takes far more knowledge, skill and effort to manage. The same guidelines counsel - wisely enough - 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 led to neglect of audio as both art and science requiring appropriate training and experience.

4. Confronting the challenges

Before describing the shape of an epistemology for audio, I will take a short excursion in this section to challenge some widely-held assumptions about recording.

[10] Of course the need to publish for 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 'products', 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 new genres for expressing knowledge about languages would arise from the practice of documentary linguistics, and indeed there are some tentative indications that this may happen in the future. At the 2010 annual meeting of the Linguistic Society of America a resolution was passed calling on linguistic departments to recognise the creation of corpora, multimedia and other documentary products for the purposes of promotion and tenure decisions. How this will play out and whether it will be adopted more widely, along with the impact it will have on linguists' practices, is yet to be seen

[11] Note that most of the papers in Barber's book take a mentalistic perspective and none consider audio.

We often hear documenters protest that there is not enough time to set up equipment such as microphones, stands, and windshields because the events of interest are too transitory and must be recorded without delay. But in most cases this amounts to an admission by the fieldworker that he or she feels no obligation to be properly trained or prepared. Documentary filmmakers, by contrast, are trained to prepare their equipment and reconnoitre situations so that they can begin recording with minimum delay. And many such cases could be addressed 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 equipment but on the fieldworker's human skills and the quality of relationships built up with consultants. In any case, that 'unmissable' event was often only unmissable because the fieldworker was present; it may have otherwise occurred a week before or a week after the fieldtrip, for example. What seems to be at stake is not the event itself but the opportunity to record it, and an inadequate recording may equally count as a lost opportunity. Is there some kind of inverted observer effect here (cf. Schembri 2010), where the fieldworker over-values the significance of his or her own presence?

Another frequently heard claim is that quality equipment is large and therefore intrusive and distracting. This is invoked, for example, as an argument for using a recorder's inbuilt microphone, i.e. for avoiding the use of well-positioned external microphones. Here again is an odd twist on the observer paradox: a claim that the presence of a microphone is enough to tip methodology into difficult territory, without consideration of its relation to the presence and activities of the documenter. Some researchers, in fact, have argued the opposite: that the tangible presence of media equipment adds to the theatricality of events and can be of assistance in eliciting several kinds of performances.[12]

Video includes a visual component that captures location in a more concrete way than audio does.[13] However, audio can also provide us - as embodied listeners with two ears - with spatial information, and indeed some of this audio information is that which does not appear in video images, such as the location of sources out of frame, or the subtle audio cues that convey the nature of a recording environment. In a recent debate[14] about the role of video in language documentation, I challenged the increasing trend among documentary linguists for shooting video, arguing that much of it seemed to be of dubious value while being very 'expensive' in terms of cost, equipment, power requirements, methodological issues, processing, and storage. Some researchers countered by offering several well-motivated arguments for the value of video. Yet, looking back at those arguments in the context of the present chapter, it appears that many actually were arguments for the usefulness of spatial information, **not video per se**. Examples include help with identifying the speaker in multi-person conversations, capturing emotions and paralinguistic meanings, and portraying the setting - all of which can be supplied, to a greater or lesser extent, by well-recorded audio. Despite the undeniable potential of video for language documentation, could it be, however, that video has been enthusiastically adopted in order to compensate for our historical ineptness at audio recording?

[12] p.c. Anthony Jukes.

[13] 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 chapter.

[14] Some of which appears in *Language Archives News* - see Nathan (2007) and replies from McConvell (2007) and Wittenburg (2007).

5. Audio and events

Audio can be thought of as a package of acoustic information that is increasingly lost and/or compressed as it moves along a five-part chain:

event > recording > representations > data > abstractions

Here, only the first two sections of this chain will be discussed.[15]

Because the task of documentary linguistics is to collect and represent 'primary data' on 'linguistic practices' (Himmelmann 1998:166), the primary data can be taken to be audio records[16] of spoken utterances which are, in turn, the originating real-world events.

The relationship between an event and an audio recording of it is mediated by the physical properties and the locations of the equipment that is used, most particularly the microphone(s), which is the transducer responsible for the singular task of converting the energy of moving air into an electrical signal. But things are not quite as simple as this. Firstly, those physical factors are considerably modulated by the documenter through his or her selection and deployment of the equipment. Secondly, the documenter is generally present during the recording and has an implicit or explicit influence on the events and thus the sounds that result from them.[17]

Thirdly, other non-linguistic sounds in the vicinity of the event might have to be taken into account too. Some, such as applause or noises whose sources are topics of conversation, might be relevant to the content of the communicative event, while others are deprecated as 'noise'. The question of how to deal with such sounds regularly arises in our training sessions, where participants ask how to record in situations where there are constantly insects buzzing, chickens clucking, and craftsmen hammering. While we can show techniques for optimising the capture of human speech under these conditions, such as minimising the loudness of the chickens etc. in relation to the voices, this is not really what is at stake. To treat the issue as one of suppression or relative loudness is to trivialise the important methodological question of what belongs in a recording, i.e. what the 'primary data' is. We cannot answer that question merely with stock recording techniques, but by applying linguistic, methodological, or philosophical principles. Having applied these principles we can decide which techniques should best fulfil them. Whether a bird's tweeting, the rasping of a saw, or a child's crying is relevant to a communicative event depends on a large number of factors, each non-trivial and possibly transitory, including the documentation goals, the social setting, topics of conversation, and personal viewpoints.

Finally, it is worth noting that 'linguistic practices' are often characterised as instances of genres (Johnson & Dwyer 2002). Although genres such as song may have

[15] 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.

[16] Assuming spoken, rather than signed, languages; for signed languages, video is the default method for recording (see Schembri 2010).

[17] The label 'observer effect', referring to the influence on performers of their awareness of being observed or recorded (see Schembri 2010), is a gross oversimplification of what really happens in fieldwork situations.

specific acoustic characteristics, in general genres are **not** properties of the recording but the result of listener interpretation.

6. Audio training at ELAR: listening with both ears

As part of our five-day training courses for ELDP-funded language documenters, we devote about one day to exploring several of the audio issues raised in this paper, in the form of discussion, practical and evaluative activities. Although it would be preferable to have more time to spend on the activities, almost every participant has told us that this training has far exceeded any audio training they have previously received.

Over the past five years the content of the audio sessions has evolved. Notably, we have gradually jettisoned topics in digital audio. This change is a result of documenters' growing familiarity with digital recorders,[18] together with our increasing attention to identifying and serving documentation goals through recording, and an increasing understanding that the best way to approach recording skills is through the development of listening skills. Therefore, a major theme is developing critical listening skills (Alten 2005: 9). We examine **signal** (what you want to be heard in a recording) and **noise** (what you don't want to be heard) from several perspectives, providing a holistic integration of:

- equipment issues (e.g. attributes, selection, compatibility)
- the moment-to-moment and situation-to-situation management of equipment, settings, participants, and the physical environment to capture all of the desired sound (e.g. the various voices in a multi-party conversation), and 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'?)

Following 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. Several offered the honest and revealing response that until the workshop, they had **never considered the possibility** of managing the recording process to attain better results. They had previously thought that all they could do was switch on the recorder and hope for the best; that they were hostages to the physical setup. Why? Because (like generations of linguists before them) they had never been exposed to any audio goals or criteria. For them, the workshop had delivered the happy revelation that the **goals** of recording provide criteria for deciding audio requirements (such as what is signal and what is noise), which in turn enable the application of learned skills to achieve good recordings. Without goals, and the corresponding skills for achieving them, results can only be hit or miss.

We generally use a training setup that includes a set of chained amplifier/splitter units that feed a pair of closed headphones for each participant,

[18] I estimate that the proportion of fieldworkers using solid state recorders has increased from around 10% (6 years ago) to 100% today. What were formerly mysterious concepts such as sampling rate are now simply a matter of making standard selections from recorder menus.

allowing all participants to listen to the same sounds simultaneously. We have other miscellaneous apparatus such as a portable stand to hold dampening materials of various types (including a sleeping bag!), CDs of recorded sounds such as chickens or pubs, recorded audio of various types and qualities, and a range of recorders, cables, connectors, stands, and microphones. In some courses, we send out groups of three or four to external locations to make short recordings. Later, the whole class listens to each recording, evaluating its quality and attempting to correlate its strengths and weaknesses with that group's equipment, techniques, sources and locations.

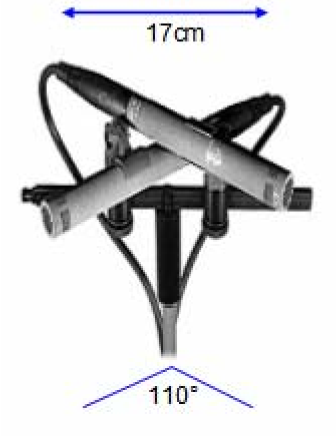
For more advanced classes, we developed a pedagogical approach whereby we exhibit the use of various configurations of equipment, props and audio sources (including, but not limited to, human speakers) 'live' in the classroom with participants listening using the headphone system. This has proved extremely effective; participants are more likely to be convinced by the incontrovertible evidence in front of their eyes as they hear the effects of, for example, swapping between a lavalier microphone and a shotgun microphone while 'listening' to a speaker standing in front of a window onto a busy street. And with this setup, participants can make suggestions that can be tried out immediately, and problem-solving tasks can be set and solved; more generally, this 'listening-centred' approach reinforces the importance of audio awareness and monitoring when making field recordings.

More recently, we added a focus on capturing spatial information. This topic covers basic psycho-acoustics, and its practical component includes stereo recording and listening to various audio outputs from stereo and ORTF microphones.[19] While stereo/binaural/spatial recording is an area that has been entirely neglected in documentary linguistics, encouraging participants to experiment with it has provided a range of useful learning opportunities. For example, one group of trainees recorded an interview in a noisy environment (in a park, next to a road) using a stereo microphone (RØDE NT4, XY type). When we later asked them which way they had aligned the microphone's stereo axis while recording, they admitted they had not thought about it at the time (it would have made a good item of metadata - see below). Actually, it was discernable from the recording that they had aligned the two speakers (interviewer and interviewee) perpendicular to the stereo axis, thereby achieving no separation between them. Nevertheless, we discovered that this could turn out to be a very useful strategy in the right context. In their recording, a listener can separate out **the competing background noise** from the **interview**, which makes for a more comfortable-to-listen-to and easier-to-transcribe recording than would have been the case if the recording had been made in the default manner, which would have separated the two participants from each other but not from the background noise.

To achieve a fully 3-dimensional spatial 'illusion' when listening back using headphones, a specifically configured pair of microphones, known as ORTF, can be used (Alten 2005: 24; see Figure 1). Training participants hear, evaluate and discuss several ORTF examples: pre-recorded conversations, fieldwork examples made by the author, and 'live' monitoring as described above.

[19] From 'Office de Radiodiffusion Télévision Française', who invented it. We think of it as 'stereo on steroids'. Actually, it is only one example of the broader category of **binaural** recording.

Figure 1. ORTF setup



Because spatiality in linguistic recordings is an unexplored area, we have also performed some practical but informal experiments. The first involved pre-recording an interview against a very noisy background of multiple conversations. We then compared several versions, each of which was derived from the ORTF original but in a very different way: the first was a full-resolution (16 bit, 44.1 KHz) **mono** version; the others were **degraded** but still two-channel ORTF versions (they were degraded by applying various levels of MP3 compression). The results were that even significantly degraded ORTF-recorded versions remained preferable to the high-resolution mono version, because they provide enough separation of the sources to both make listening tolerable and to allow the listener to engage with the focal spoken content. The mono sound-stage, despite its **prima facie** higher technical quality, collapses all the conversations into a single space and leaves the listener disoriented and unable to focus on the interview. The second experiment took place in our 3L Summer school training in 2009, and involved using the ORTF array (see Figure 1) and monitoring it live, except that it was placed, out of sight, in an adjacent room to where the training was taking place. But in that adjacent room, another training event was taking place - software training involving various conversations amongst pairs of people sitting at computers arranged around the room, many of whom were typing and mouse-clicking at the computers. The critical question that we addressed was: was the spatial audio experience strong enough such that our participants could feel psychologically present in the other classroom? After the experiment, participants were asked simply to state whether or not they had been ‘teleported’ into the next classroom, and over 70% of them agreed that they had been.[20]

Overall, the preliminary results of these explorations into spatial recording and listening using ORTF[21] are that:

- separation and localisation of sound sources can be achieved
- much more knowledge about the recording **environment** is captured
- on the other hand, the richness of captured information can sometimes be distracting to listen to[22] and recordings made in some environments are quite disorienting[23]

[20] I therefore claim that this is the world’s first teleportation of a whole (or nearly whole) class into another location, although I do not know how to verify this claim.

[21] Note that 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. However, it proves to be an excellent way to illustrate the value of spatial audio and how much information is lost if it is ignored.

7. 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 use aural information to comprehend the physical spaces we are in. We experience 'spatial or binaural localization' by using our two ears 'to localize a sound source within an acoustic space' (Huber and Runstein 2005: 62). This ability takes into account not only sounds received directly from sources, but also those reflected from 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.[24]

Aural processing involves the ears **and** the brain.[25] We interpret the space around us by comparing and analysing the following properties of sounds reaching each ear, and the differences between them:[26]

- **loudness** - each ear receives sound of different loudness due to different distances travelled, as the energy falls off according to the inverse square law
- **phase/delay** - sound reaches each ear at slightly different times due to the different distances travelled
- **frequency falloff** - higher frequencies lose energy sooner than lower frequencies, so sounds travelling different distances to reach each ear have different frequencies
- **frequency colouration** - sounds reflected off different materials have different frequency distributions (cumulatively in the case of multiple reflections) when they reach the ears via different paths

Furthermore, audio information is processed in the context of the listener's short-term and long-term knowledge. Short-term knowledge includes his or her current and transitory knowledge (gained through any of the senses) of the immediate environment (e.g. location, orientation, identification of audio sources). Long-term knowledge refers to our cumulative experience, as embodied actors in the world, of how perception is influenced by the nature of sources, materials and spaces.

We integrate aural processing and these types of knowledge both consciously and unconsciously. At a conscious level, we can direct our attention to particular sources. This underlies what is commonly called 'the cocktail party effect',[27] the ability to pick out the speech of one individual even in a crowded and noisy environment. An example of unconscious processing is our tendency to quickly lose awareness of the presence of backgrounded audio sources, such as fans, traffic, or chickens, when paying attention to speech or music. These effects showcase our

[22] 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.

[23] A recording made in the domed plaza of the British Museum was very disorienting. It seems that there is an exaggeration of some kinds of echo/reverberation.

[24] While a mono recording can also convey an impression of a space - for example echo suggests a large empty space and loudness indicates closeness - a mono listening experience does not enable localisation; the listener cannot place sources within a 3-dimensional mental sound stage.

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

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

[27] Also known under the more proletarian label 'the cafeteria effect'.

capacity to use spatial information to navigate and orientate ourselves as we move about the world, without being consciously aware of the role played by our aural processing. However they also mean that when we record audio in an environment, even if we record in stereo or ORTF, we are detaching the audio signals from most of the perceptual and short- and long-term knowledge that listeners in that environment have access to.

8. Lost in space

The preceding section described the huge amount of spatial information available to listeners, and how they use this information in everyday life. How much of this information can be recorded? With suitable equipment and techniques, much of it **can** be captured in a recording. The word 'captured' is important here because spatiality is not inherently present in a recording. Recordings can only make the two channels of information available to a listener who is capable of interpreting them in order to construct a mental 'sound stage' resembling the original recording environment.

If we make mono recordings, we do not capture - and therefore deny to all future listeners - the vast amount of information in those two channels. The remainder of this chapter argues that those two channels of information are valuable components of a language documentation.

Documenters who move quickly to transcription and from that point work only with text may view spatial information as irrelevant. Their workflow involves a massive **loss** of information. Let us roughly quantify and compare the information in a 5-second utterance represented as audio and text:[28]

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

The documenter who quickly abandons audio in favour of text eliminates over 99.99% of its information! Losing information is not necessarily a bad thing: information theory tells us that losing information is the essence of moving from data to understanding, as long as the **correct information** is discarded. For this particular documenter, it is unlikely to matter where that 99.99% of information is lost, whether at the original 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 poor reproduction for listening (e.g. listening through cheap computer speakers). None of these deficiencies is revealed in the outcomes of this documenter's work - until a community member, teacher, historian 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.

[28] Assumptions: acoustic information is quantified on the basis of the CD-audio standard; transcribed speech is at the rate of 3 syllables per second; a syllable is written as 2 characters, each of which is 1 byte in size.

9. New roles for metadata

The preceding discussion can help to diagnose common problems in recording. For example, many documenters are surprised to find that the audio they made was spoilt by the presence of extraneous noises. All of those noises, of course, had been present in the recording environment, but had been psycho-acoustically filtered from the documenter's attention at the time of recording.[29] 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 spoilt or inferior rendition of the original event depends on a number of factors, many of them subjective and connected to the purposes for listening. But there are also objective factors based on the information that was present for a listener in the original setting and whether it is accessible to someone listening to the recording:

- the acoustic (including spatial) information in the recording environment
- the (original) 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 an important component, because the ability to separate out simultaneous events is crucial for intelligibility and for comfortable sustained listening. While a good recording can capture most of that acoustic information, a listener to a recording can never replicate the experience of an event participant, even if only for the fact that event participants have knowledge about the location and what was happening before the recorder was switched on. Thus, the extent to which listening can correspond to original experience is also dependent on who is doing the listening and on their knowledge about the participants, location, and history of the original event.

This leads us to consideration of metadata, and whether it can provide a way to convey some of this knowledge. Metadata is commonly defined as **data about data**, and its function is to enable the management, identification, retrieval, and understanding of data (OAIS 2002). In current language documentation practice, metadata for audio recordings typically consists only of information about the location of the recording, and information about the speakers - their names, sex, age etc.. Less frequently, fieldworkers note down technical details such as equipment type and settings. Virtually no fieldworker makes a spatial characterisation of the event - how the microphones were arranged, their relation to the sound sources, orientation of the sources, and the layout and nature of the recording space. Diagrams and photographs would be useful formats for some of these categories of metadata.[30] Even simple information about which speaker is heard in which stereo track is usually omitted.[31]

[29] 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 information or converse with consultants.

[30] David Nash informed me that 'sometimes I used to make a little notebook sketch of the layout around the recorder, including labelling of cardinal directions.'

[31] 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 a range of listeners who simply want to listen to the audio.

10. On listening

Until now, documenters have typically thought about recordings in terms of what **linguistic phenomena** they are assumed to contain. In contrast, this chapter has taken **the 'listener'** to be the pivotal concept. Recordings have content and significance only in terms of the experience of particular listeners. If nobody ever listens to them, their only significance is as a memento of fieldwork. By factoring listening, and listeners, into documentation, we can start to talk about the intentionality of recording; **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. We can then hope that, as a result of our efforts, listeners have a satisfying experience, without naively assuming that we are directly delivering specific content. We also understand that the act of recording constructs listeners, whether imagined or real, because, just like video, an audio recording imposes a perspective 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 though these production techniques and what kind of listener is being produced?

The idea of recording for listeners is as novel for documentary linguistics as it is central to the music industry. For us, it opens up new ways of thinking. For example, consider the 'cocktail party effect' discussed earlier, which enables us to selectively pay attention to one of many audio sources. This ability declines with age and is particularly affected by even mild hearing disability. We can now say 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'. For example, two recordings that are equally undistorted and intelligible can differ significantly according to how pleasant they are to listen to. Our experience is that people typically agree about the listenability of any particular recording. Since language documenters are likely to be the most ardent and persistent listeners to their recordings (transcribing an hour of audio can take 50 or more hours of listening), it is a valid part of a research methodology to make recordings that are comfortably and sustainably listenable over long periods using headphones.[32]

Here is another example. Recently, the documenter Carolina Aragon explained her difficulty in recording the Akuntsú people of the Amazon because their rainforest environment is perpetually full of loud bird and animal calls (and she believes it would not be safe to take people elsewhere to record them). She had tried almost every technique for overcoming these 'noises'. However, the important observation here is that since the Akuntsú people always hear their language in this soundscape, interesting linguistic questions are raised about how those speakers and listeners, and their communicative practices, deal with it. Thinking about what we seek to achieve by recording, and therefore how we record, is a relevant part of any investigation into the acoustic phenomenon we call spoken language.

[32] Documenters often ask for advice about suitable headphones for sustained listening, but not about how to record for it.

11. Conclusion: an epistemology

This chapter has shown that audio is a necessary, complex, and rich component of the documentation of spoken languages. The practical and aesthetic aspects that have been covered can be summarised as a set of criteria for evaluating recordings:

- **accuracy**: is the audio source captured with fidelity and without distortion?
- **intelligibility/information accessibility**: can the intended content be identified?
- **signal versus noise**: is the ratio acceptable?
- **separation of noise**: can all the noise sources be separated from the focal sources?
- **localisation**: is enough spatial information captured to place the sources on the 'sound stage'?
- **listenability/comfort/aesthetics**: is it easy on the ears? will it be debilitating to listen to for an extended time?
- **representation of environment**: are the acoustic properties of the recording environment appropriately represented?
- **content** (identity, performance, uniqueness, coverage): were the right people recorded doing the right things? did they do them well?
- **editability/repurposability**: can the recording be used to create a range of appropriate resources?

The broadest aim of the chapter is to stimulate discussion about the goals and purposes of audio in our field. As an initial contribution to an epistemology for audio in language documentation, 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(s) define the recording methodology
- ethical recording respects language speakers and honours their contribution through application of skill and effort
- a recording should capture spatial information
- detailed metadata about the recording and its physical setting are required if a complete 'record' is to be made.

References

Alten, S. 2005. *Audio in media*. Belmont CA: Thomson.

Austin, Peter. 2010a. Current Issues in Language Documentation. *Language Documentation and Description*, Volume 7. London: SOAS.

Austin, Peter. 2010b. Ethics in Language Documentation. *Language Documentation and Description*, Volume 7. London: SOAS.

Barber, Alex. 2003a. Introduction. In Barber (ed). *Epistemology of languag*e. Oxford: Oxford University Press.

Barber, Alex (ed). 2003b. *Epistemology of languag*e. Oxford: Oxford University Press.

Benterrak, Krim, Stephen Muecke & Paddy Roe. 1984. *Reading the Country*. Fremantle: Fremantle Arts Centre Press.

Bird, Steven and Gary Simons. 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. *Language Documentation and Description*, Volume 3, 42-52. London: SOAS.

Dobrin, Lise, Peter Austin & David Nathan. 2007. Dying to be counted: commodification of endangered languages in documentary linguistics. In Peter K. Austin, Oliver Bond, & David Nathan (eds.) *Proceedings of the Conference on Language Documentation and Linguistic Theory*, 59-68. London: SOAS. [Online at http://www.hrelp.org/publications/ldlt/papers/ldlt_08.pdf]

Himmelmann, Nikolaus. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36, 161-195.

Himmelmann, 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.

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]

Kheshti, Roshanak. 2009. Acoustigraphy: Soundscape as Ethnographic Field. *Anthropology News*, April, p15.

McConvell, Patrick. 2007. Video - A linguist's view (A reply to David Nathan). *Language Archives Newsletter*, 10: 2-3.
[online at http://www.mpi.nl/LAN/issues/lan_10.pdf]

Nathan, David 2006a. Thick interfaces: mobilising language documentation. In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*, 363-379. Berlin: Mouton de Gruyter.

Nathan, David. 2006b. A Talking Dictionary of Paakantyi NSW. In Laurel Dyson, Max Hendriks & Stephen Grant (eds.) *Information technology and Indigenous People*, 200-204. Hershey PA: Idea Group.

Nathan, David. 2007. Digital video in documentation and archiving. *Language Archives Newsletter*, 9: 3-4.
[on line at http://www.mpi.nl/LAN/issues/lan_09.pdf]

Nathan, David. 2009. The soundness of documentation: towards an epistemology for audio in documentary linguistics. *Journal of the International Association of Sound Archives,* vol 33, June 2009.

Schembri, Adam. 2010. Documenting sign languages. *Language Documentation and Description*, Volume 7.

Thieberger, Nicholas. 2004. Documentation in practice: Developing a linked media corpus of South Efate. *Language documentation and description,* Volume 2, 169-178. London: SOAS.

Thieberger, Nicholas & Simon Musgrave. 2007. Documentary linguistics and ethical issues. *Language Documentation and Description,* Volume 4, 26-37. London: SOAS.

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.

Wittenburg, Peter. 2007. Video - A technologist’s view (A reply to David Nathan). *Language Archives Newsletter*, 10: 3-5.
[on line at http://www.mpi.nl/LAN/issues/lan_10.pdf]

Zierott, Nadja. 2005. *Aboriginal Women’s Narratives: Reclaiming Identities*. Piscataway NJ: Transaction.