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

The 1st International Conference on Language Documentation and Conservation
University of Hawai‘i Manoa
Sat March 14, 2pm

David Nathan
Endangered Languages Archive
SOAS, London
djn@soas.ac.uk
Aims

- to share some ideas about audio within language documentation
- to seek feedback/correction etc
- to promote critical thought about practices
- to stimulate experimentation and change
- with apologies to the acoustically and philosophically better informed
Past influences

- 1990s and multimedia: linguistic audio as evidence, not performance
- documentary linguistics
- Dietrich Schüller
- training documenters
From evidence to performance
Linguists and audio

- little or no training
- poor recording techniques
- little actual *usage* of audio
  - as data
  - as performance
  
  *ie*

- *audio is just an inconvenience* on the way to transcriptions, description, analysis
- *there was no methodology or epistemology for audio*
Language documentation?

- an offshoot within linguistics, starting 10 years ago as a response to global language endangerment
- a multipurpose and comprehensive record of the linguistic practices characteristic of a speech community .. the emphasis is on *the collection and representation of primary data* rather than theory and analysis

(Himmelmann 1998)
Audio is not data

- real events
- recordings
- representations
- data, abstractions
Audio is an event or a resource

- making it is both art and science
- a critical and ethical responsibility
- strongest relationship to communities
- it’s not necessary to record *everything*
Linguists and audio information

- information theory - it’s necessary to lose some data to get information/knowledge
- for reduction to writing a lot of data is lost
- recording can involve a massive and uncontrolled loss of data
- but loss of audio information is not consequential!
How are we losing audio information?

- non-optimal equipment and techniques
- lack of training
- noise
- spatial information (how can we arbitrarily ignore this?)

- we have worried about
  - *resolution* not *signal to noise ratio* (SNR)
  - *compression* not *spatial information*
What we’ve been doing about it
What we’ve been doing about it

- training
- field trials
- challenging assumptions
  - recording needs to be done unprepared
  - equipment is intrusive
Training themes

- signal and noise
  - need to define and control signal and noise

- metadata: information that provides for the discovery, usage and *understanding* of data
Training themes

- monitoring
- evaluating

44.1 KHz, 24 bit

- psychoacoustics
Training themes

- room acoustics
- shotgun vs omni with pub noise
- ORTF

- equipment:
  - frame with sleeping bag
  - headphones, speakers, amplifiers/distribution
  - mics and cables
  - recording of pub noise, portable player
  - “performers”
Stereo

- interview in noisy environment
Stereo

- interview in noisy environment
Binaural/ORTF
ORTF

17 cm

110°
Preliminary results from ORTF

- listeners agree about localisation
- participants can be “separated”
- more knowledge about environment
- can be distracting (> “information masking”)
- some environments result in very disturbing recordings

- preliminary experiment: evaluation and information extraction, comparing degraded ORTF with uncompressed mono recordings
How do we perceive spatial information?

- two ears – generate a difference
- spatial information from:
  - phase/delay
  - frequency falloff
  - frequency colouration
  - intensity variation
  - (head orientation)

and combinations of all of these!
What happens in 10ms? (1/100 sec)

- a few hundred distinguishable amplitude readings, just enough to identify most speech-relevant frequencies
- sound has travelled about 3 metres
- so our potential discrimination for environmental/spatial acoustic information is similar in scale to our ability to detect frequencies
- (compare to vision; even in 50ms you won’t notice any change at all)
Arguments from video enthusiasts

- arguments for use of video - many apply to audio!
  - disambiguating participants
  - paralinguistic expressions
  - emotions
  - capturing locations, environments/settings
  - attraction of multimedia products

- are those pro-video points mainly about spatiality?
- has video been used/advocated to make up for the inadequacies in audio practice?
- and more questions about non-spatial information ...
Some implications

- Hearing impaired people have trouble with “cocktail party/cafeteria effect”
- If we said that recording is “for a human listener”:
  - A recording that does not appropriately distinguish the focal speaker from background talk is making a recording “as heard by a hearing impaired person”
“Energetic vs Information masking”

- What are the implications for:
  - listeners? “native speakers”? linguists?
Glimpse effect

- for “native speakers”, the gaps together with the redundancy in language may make this intelligible or acceptable
- what about for linguists?
Implications

- who/what is audio for?
  - if “native speakers” then glimpse effect can make audio acceptable, but for linguist unacceptable
  - if audio is to be listened to for long periods, comfort will be important
- audio data collection should also include metadata about sources, environment, locations, orientations
- who notes the content of stereo channels?
Implications

- signal
  - deciding what is the signal
  - quality criteria (human arbiter, lip smacking?)
  - understanding its other properties
Implications

- noise:
  - deciding what is noise
  - what is noise may change
  - understanding its properties
  - how to record, vis-a-vis noise
  (all of these will have vastly greater influence than selection of recorder or resolution etc)
Implications

- we should record audio relative to goals and usages
  - teaching materials (low SNR will reduce effectiveness)
  - songs/stories/performances
  - talking dictionaries
  - transcription!
Implications

- don’t view audio only through “linguistic tools”:
  - songs interactive player – linguistic tools don’t acknowledge verses!
- different equipment and setups for different languages, or events that have different acoustic properties
- listening is important!
  - listenability (comfort etc) - actual usages
  - listening environment/method
Intelligibility vs listenability

Performance / Evaluation

SNR Signal to noise ratio

intelligibility

quality
Implications

- as participants in events that are recorded, we have a huge range of opportunities to choose equipment, physical layouts, manipulation of equipment and environment, and influence on performances – all of which overwhelmingly influence audio recording quality

- if we don’t collect good audio we may as well do fieldwork by phone or Skype!
Audio processing?

- Not treating audio recording as data-gathering, e.g. “remove the noise later”:

  when noise reduction algorithms are applied, intelligibility goes *down*
Possible desiderata

- accuracy
- intelligibility /information accessibility
- listenability /comfort/aesthetics
- localisation of performers
- representation of environment
- separation of environment
- separation of noise sources
- editability /repurposeability
- content: performance, identity, uniqueness, coverage
The future

- who can we turn to?
  - speech processing
  - phoneticians
  - music
  - radio
  - film
  - various users of audio
  - documenters who take up the challenge
Epistemology for audio in language documentation

- an audio recording represents what a human listener would experience at a particular location and orientation in a setting
- an audio recording is to convey an audio experience to a human listener
- the context, goal and methodology define the audio information to be captured
- relevant spatial and configuration metadata should be recorded
- a multipurpose record should capture the maximum spatial information