Sound Signal Statistics

Leslie S Smith, Department of Computing Science and Maths, University of Stirling, Stirling, Scotland. lss@cs.stir.ac.uk

UNIVERSITY OF STIRLING

Introduction: How might one analyse sound?

Follow the biology: Modelling brainstem responses, midbrain, auditory cortex?
– Useful, informative, but begs the question of why the auditory pathway is like it is
Follow the ecology: Need to "instrument" behaving animals
– Difficult to do
Follow the signal: Time to try this
– Particularly since many have studied this in the visual area for more than a decade! E.g. Lewicki 2002, Klein et al 2003.

Alternative introduction: Why is auditory processing like it is?

Because of sound statistics, ecological requirements, what is biologically possible sound has shaped auditory processing: over evolutionary timescale *and* over lifetime of animal

We therefore become interested in the statistics of sound

For some animals, specific sounds are all that matters
- Crickets and detecting females, Cricket parasite

But for other animals, sound has a more general utility: What and where tasks:
- auditory scene analysis.

Sound signal statistics

Information about the world comes from statistical deviations from pure randomness.
Where do we look for statistics in the sound field?
Take the hint from image analysis:
PCA and ICA on patches of images provide structures which seem to reflect image structures: edges, corners, etc.
Analysis uses patches (small circular (or square) solid angles of (usually static) image)
They also seem to provide 'receptive fields' similar to cortical neurons
Suggests applying PCA and ICA to sound.
But how? What is a 'patch of sound'?
Note: we deal here with monaural sound. Binaural sound can provide further material of interest, particularly in sound source localisation.

"Patches" of sound

What are the candidates?
A sequence of samples: 1-d, simple to produce, easy to work with, and results are sounds
Single FFT vectors: 1-d again, results are spectral analysis of section of sound
Sequence of FFT vector: 2-d: Output is spectrogram of a piece of sound
Coded filterbank output: 2-di: Output is filterbank output over a period.

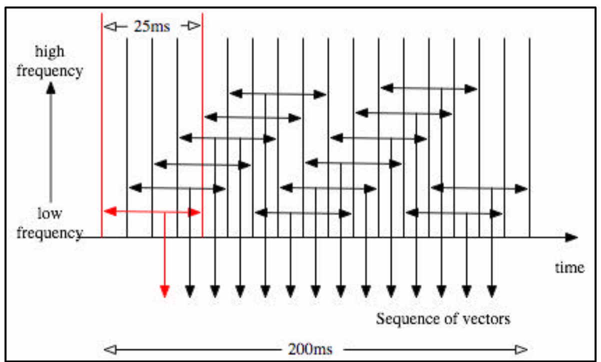
Many other possibilities as well.

1-Dimensional results

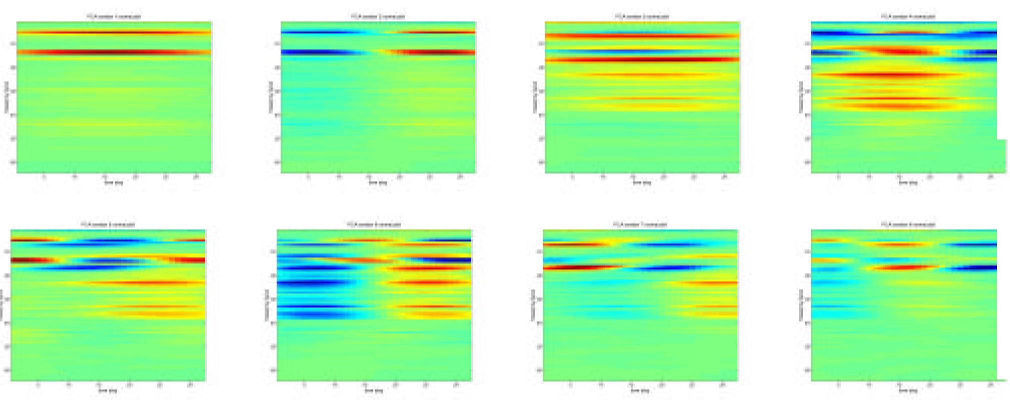
Results from sequences of samples: TIMIT dataset and Music (Coltrane's Equinox)
(Listen to them on Laptop)
- PCAs seem to average sounds together: "chorus-like" effect
- ICAs seem to pick out specific features.

2-dimensional patches of sound:
Overlapping FFTs.

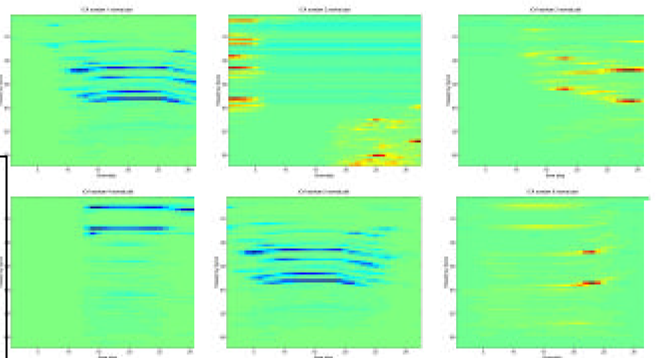
(Unfortunately, cannot resynthesize these sounds: only pictures)



PCAs, for female speech



ICAs for female speech



Issues:
Sound level and frequency banding are linear
Each vector provides 1 value per frequency
for low frequencies, implies long sound duration
Single duration (for all frequencies) implies low temporal resolution

Linearity could easily be overcome, but temporal resolution is fixed

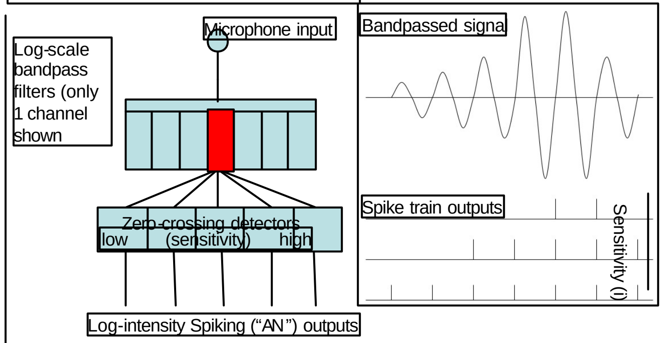
Can a biologically-inspired technique help here?

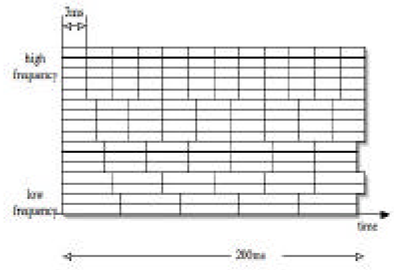
Produced for Gordon Research Conference, "Sensory Coding and the Natural Environment", 5-10 Sept 2004

Filterbank approach
Gammatone filterbank, logarithmic (approx) distribution of bands

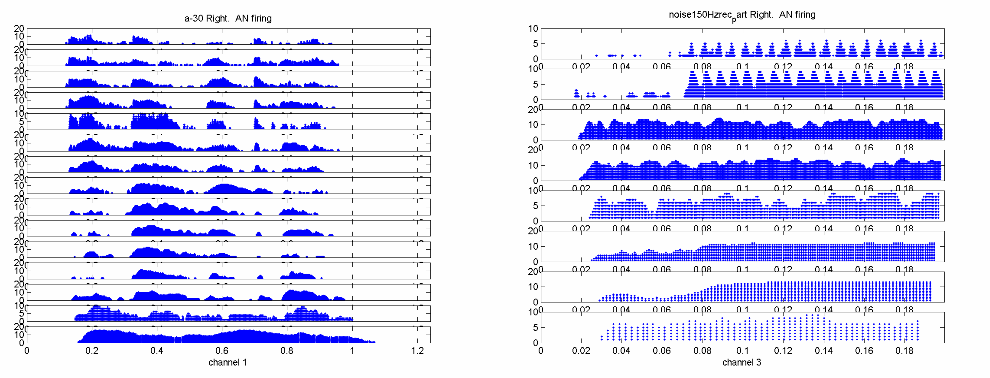
Auditory-nerve-like spike-train based coding: Logarithmic coding of intensity levels

For each time segment in each band we produce a single number from the spike coding. We can use different time segment lengths in different bands





Auditory-nerve-like spike output

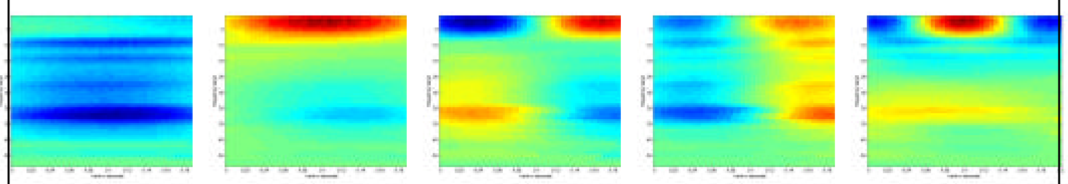


1.2 seconds speech

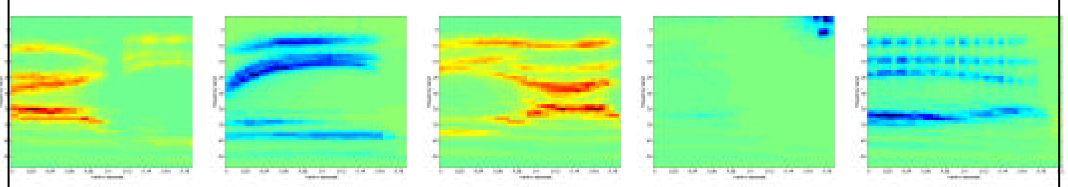
200 ms speech

Results

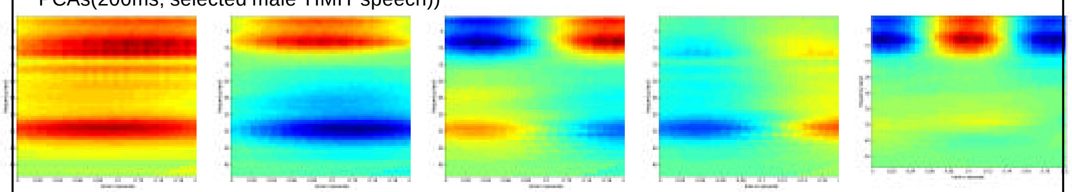
PCAs(200ms, selected female TIMIT speech))



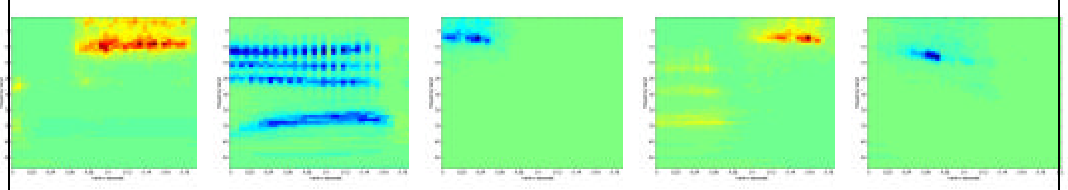
ICAs (200ms selected female TIMIT speech)



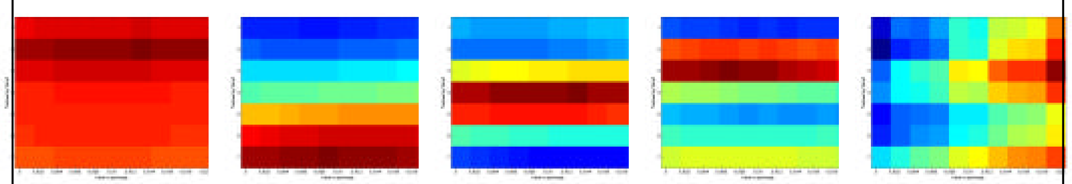
PCAs(200ms, selected male TIMIT speech))



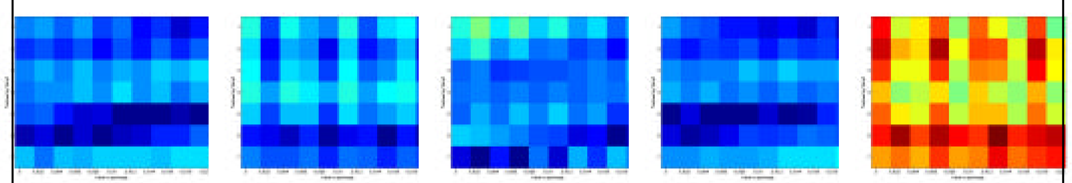
ICAs(200ms, selected male TIMIT speech))



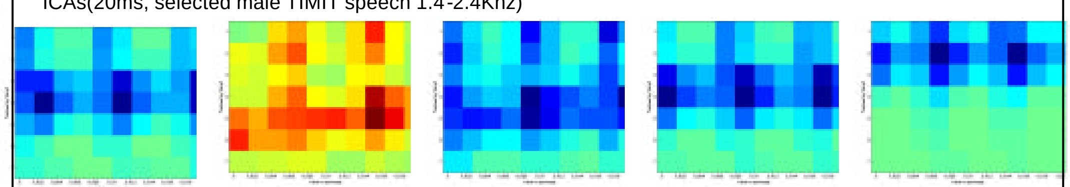
PCAs(20ms, selected female TIMIT speech, 2.07-3.46Khz)



ICAs(20ms, selected female TIMIT speech 2.07-3.46Khz) Note presence of amplitude modulation at F0



ICAs(20ms, selected male TIMIT speech 1.4-2.4Khz)



Conclusions
ICAs look more interesting than PCAs. The appear to pick out characteristic features of the sound set.
This is the case for all three techniques investigated here.
ICA applied to 200ms stretches picks out notes or particular sounds in music, and characteristic sounds
in speech
ICA applied to short-time-scale signals can pick out amplitude modulation

Further Work
Try techniques here on other sound sets (e.g. birdsong, babies cries, other animal vocalisations)
Look at binaural sound.