Sound signal statistics

Leslie S Smith,

Department of Computing Science and Maths,

University of Stirling,

Stirling, Scotland.

lss@cs.stir.ac.uk

(Seminar given to IMNS, School of Electrical and Electronic Engineering, University of Edinburgh May 4 2004.

Copyright 2003-4 © Leslie Smith

How might one analyse sound ?

- Follow the biology
  - Analyse the responses of cells in the auditory pathway of different animals
- Follow the ecology
  - Analyse what sound is useful for, and what affordances it may provide
- Follow the signal
  - Analyse the statistical irregularities (non-randomness) in the signal

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

Copyright 2003-4 © Leslie Smith

An Alternative Introduction

- Question: Why is auditory processing like it is?
- Answer: because of
  - Sound statistics
  - Ecological requirements
  - What is biologically possible
- Sound has shaped auditory processing
  - Over evolutionary timescale
  - Over lifetime of animal
- We therefore become interested in the statistics of sound
- We note that 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 is gleaned from statistical deviations from pure randomness.
- So what can analysis of the sound field tell us?
- Where do we look for statistics?

- Take the hint from image analysis:
  - PCA and ICA on patches of images provide structures which seem to reflect image structures: edges, corners, etc.
    - Patches were 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

Some Candidates:

- A sequence of samples
  - 1-dimensional: straightforward to work with
  - Output from PCA/.ICA is a 1-dimensional sequence
    - I.e. a short piece of sound
- Single FFT vectors
  - 1-dimensional again
  - Output from PCA/ICA is a 1-dimensional sequence
    - I.e. spectral analysis of brief section of sound
- Sequence of FFT vectors
  - 2-dimensional
  - Output is spectrogram of a piece of sound
- Coded filterbank output
  - 2-dimensional
    - Output is filterbank output over a period.
- There are other possibilities too
  - Random samples: like Bledsoe and Browning's N-tuple approach to image analysis.

1-dimensional patches of sound

- Make up vectors from sequences of samples
  - Loses all frequency information from sound
  - Bears no resemblance to biological auditory processing
    - But results can be played back (useful!)
  - Worth trying.
- Make up vectors from single FFTs
  - Tells us only about overall energy distribution during each FFT sample
  - (Could be energy and phase, if complex values were used)
  - Loses all time information from sound
  - Not worth trying: provides only information about instantaneous sound statistics

Results from simple 1-d sample analysis

- From some speech data: mixed TIMIT signals.
- PCA's, ICAs (see adjacent page)

Copyright 2003-4 © Leslie Smith

From some music (Equinox, John Coltrane)

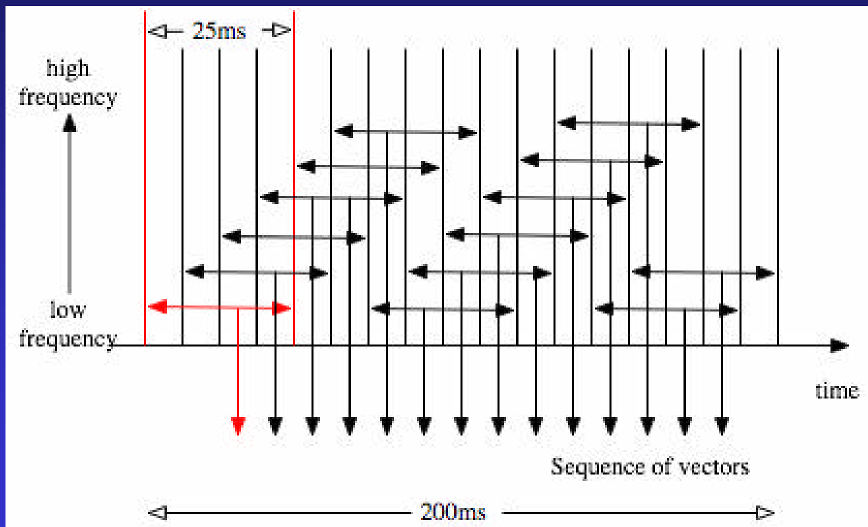
- PCAs
- ICAs (see adjacent page)

- Conclusions: PCA's seem to average sounds together: "chorus" like effect on speech, whereas ICA's seem to pick out specific features.

Copyright 2003-4 © Leslie Smith

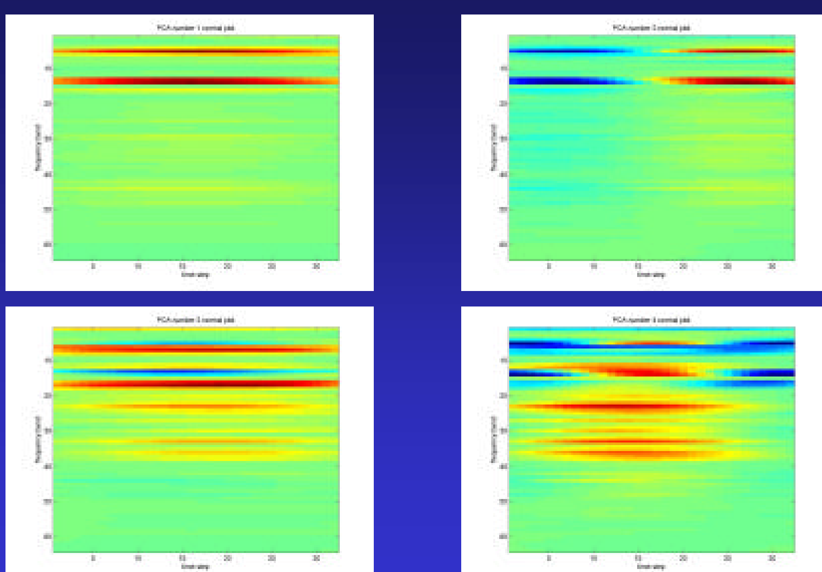
2-dimensional patches of sound (1)

- Overlapping sequence of FFT vectors: 1 element per 25ms



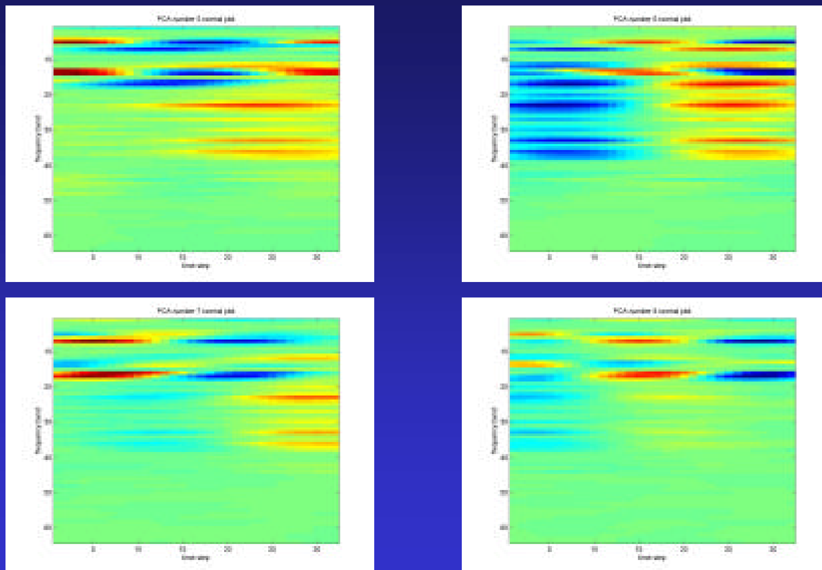
Copyright 2003-4 © Leslie Smith

FFT-based PCAs, Female speech

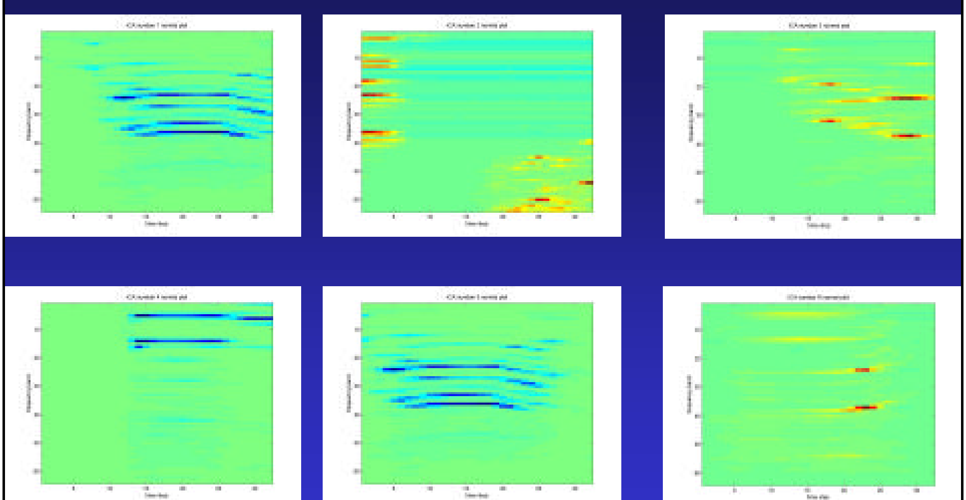


Copyright 2003-4 © Leslie Smith

FFT-based PCAs, Female speech ctd



FFT-based ICAs, Female speech



Problems with FFT

- Sound level and frequency banding is linear
- Each vector (of spectral intensities) provides one value per frequency
  - If low frequencies are included , sound duration needs to be long
  - Single duration implies low temporal resolution
    - Only partly overcome by using overlapping FFTs
- Although the linearity issue could be overcome (by repackaging the values), the temporal resolution is fixed

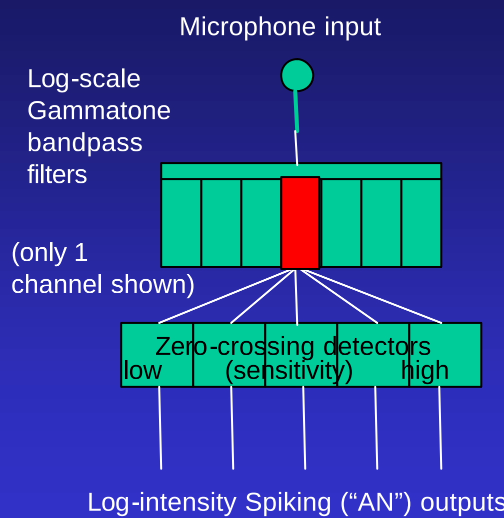
Desirable features for sound patches

- Logarithmic intensity sensitivity
  - Perceptual system is roughly logarithmic
- Logarithmic frequency sensitivity
  - Perceptual system has a roughly log sensitivity
- Duration of patch corresponds to duration of some features
  - But which features?
  - Syllables?
    - About 200ms (see also Klein et al, 2003)
  - Envelope amplitude modulation?
    - About 20ms

Filterbank approach(1)

Filtering sound signal into multiple bands (roughly log-scale) provides frequency information.

But how should the output from each filter be coded so as to provide suitable “patches of sound”?

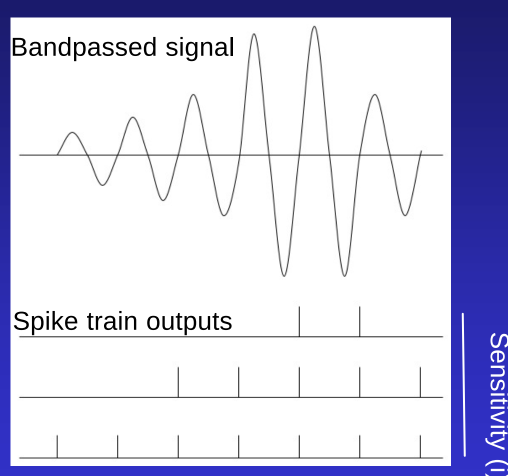


Copyright 2003-4 © Leslie Smith

Filterbank approach (2)

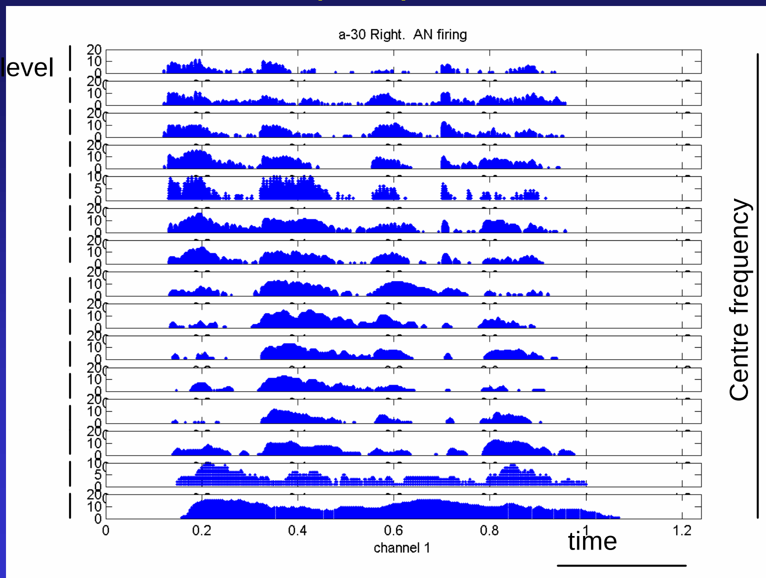
- We want to code the power output of each band
  - Logarithmically
  - Over some period
    - (and preferably in such a way that using low frequencies does not compromise temporal resolution at higher frequencies)

We use a representation based on the auditory nerve:
Multiple spike trains are produced for each band,
Each codes positive-going zero-crossings



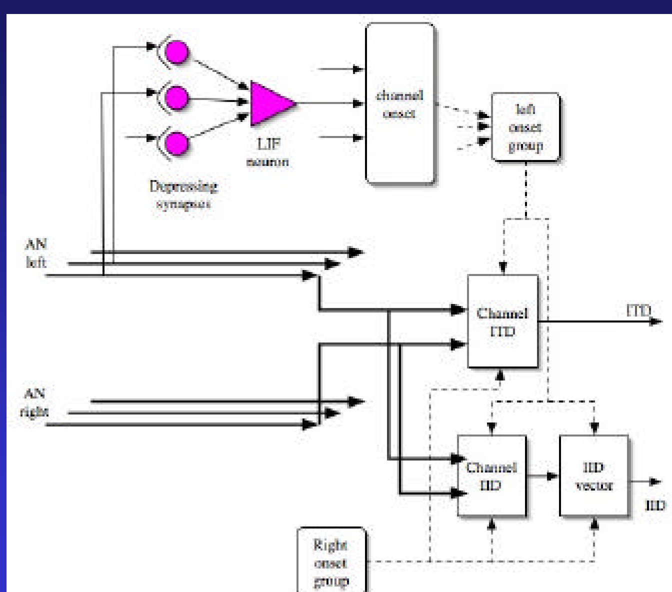
Copyright 2003-4 © Leslie Smith

Example spike train



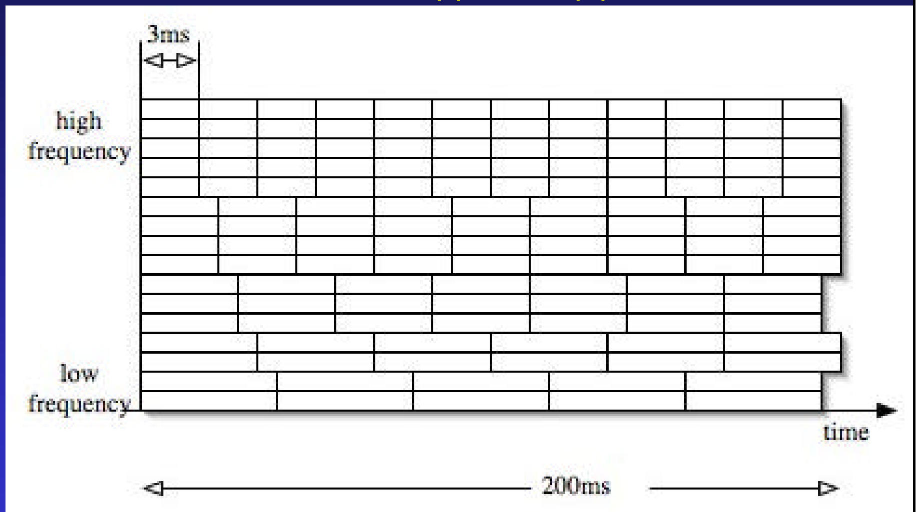
Copyright 2003-4 © Leslie Smith

An aside: the ‘where’ task



Copyright 2003-4 © Leslie Smith

Filterbank approach (3)



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

FFT vs Filter based

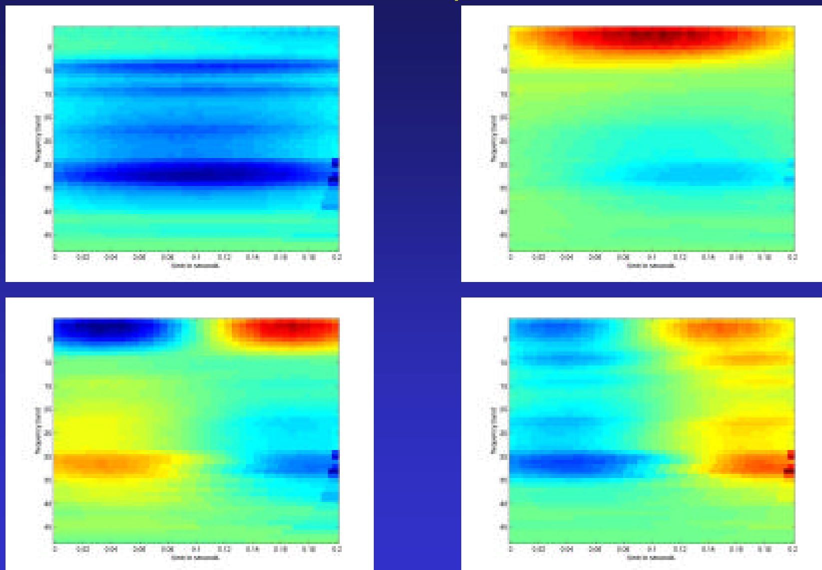
FFT based

- Linear in intensity
  - But can be modified
- Linear in frequency
  - But can be repackaged
- Each spectrally analysed segment has to be at least 1/minimum frequency long
- OK for 200ms syllable level analysis
- Not OK for 20 ms analysis

Filter based

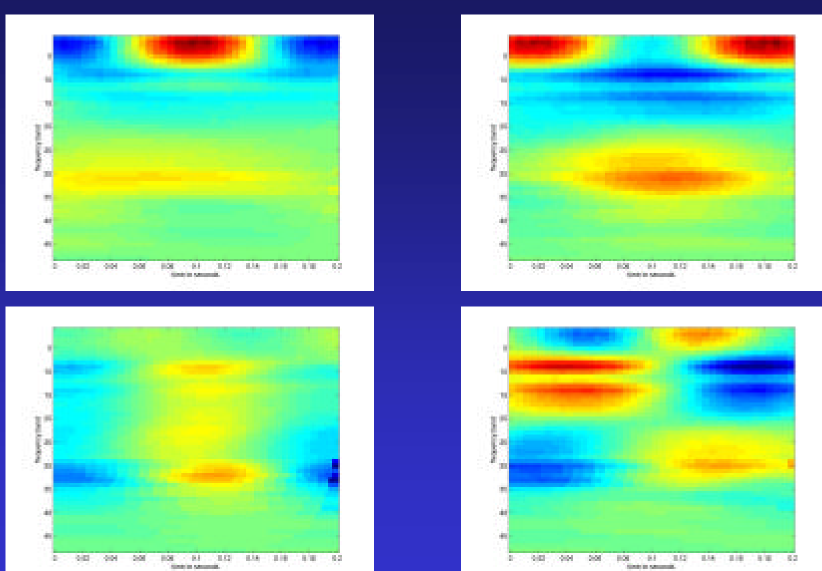
- Intensity sensitivity depends on how intensity measured
- Frequency linear or log, depending on filters used
- Can have different segment lengths for different spectral bands
- OK for both 200 ms and 20ms analysis

Applying these techniques to longer stretches of sound: TIMIT female, 50 samples. 200ms PCAs



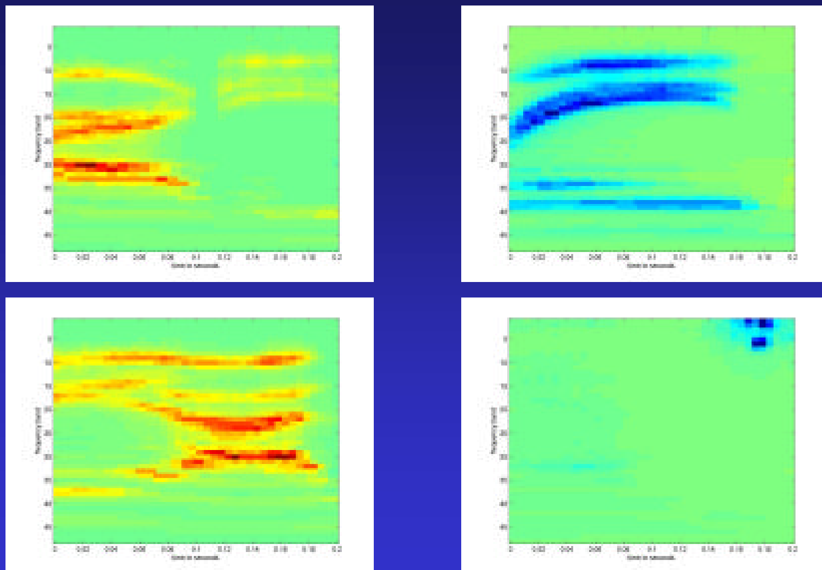
Copyright 2003-4 © Leslie Smith

PCAs ct'd



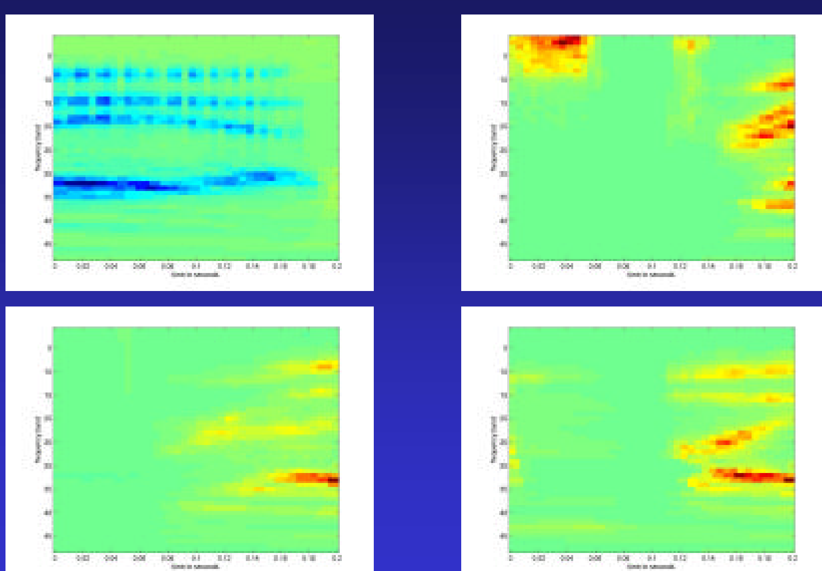
Copyright 2003-4 © Leslie Smith

200ms ICAs



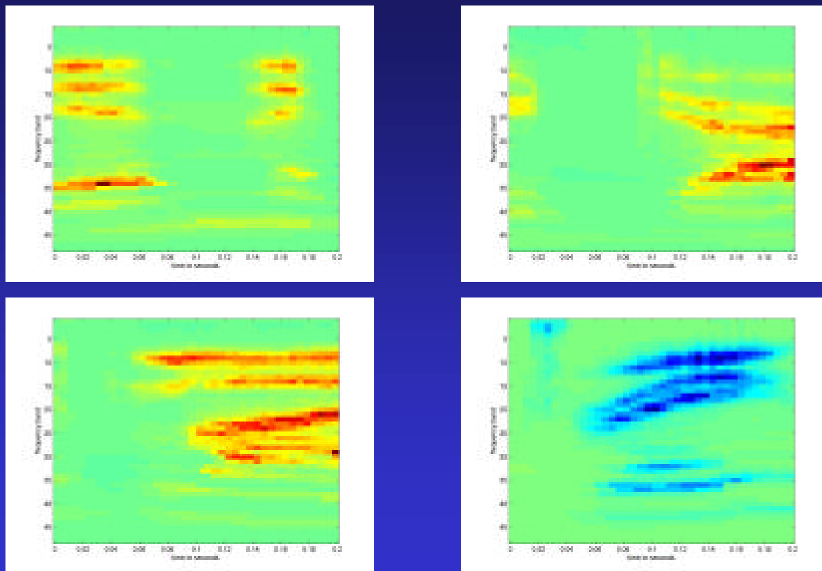
Copyright 2003-4 © Leslie Smith

ICAs ct'd



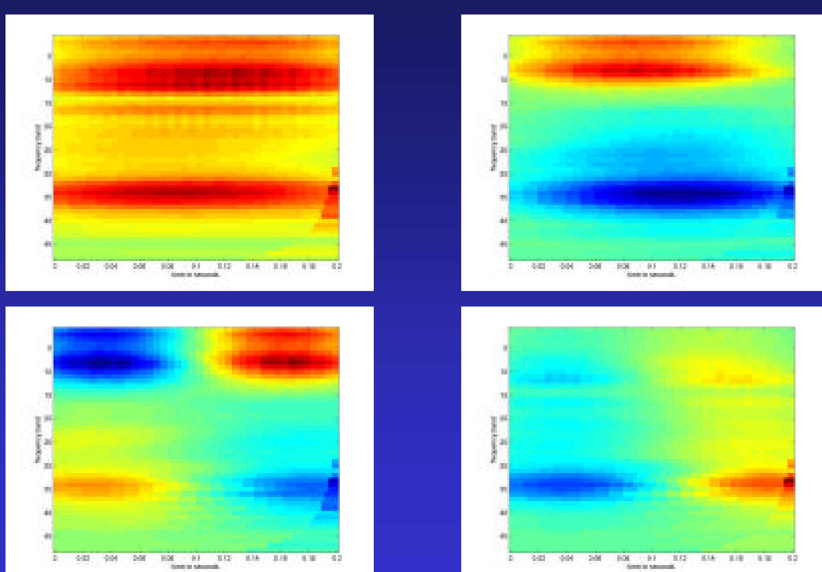
Copyright 2003-4 © Leslie Smith

...and more



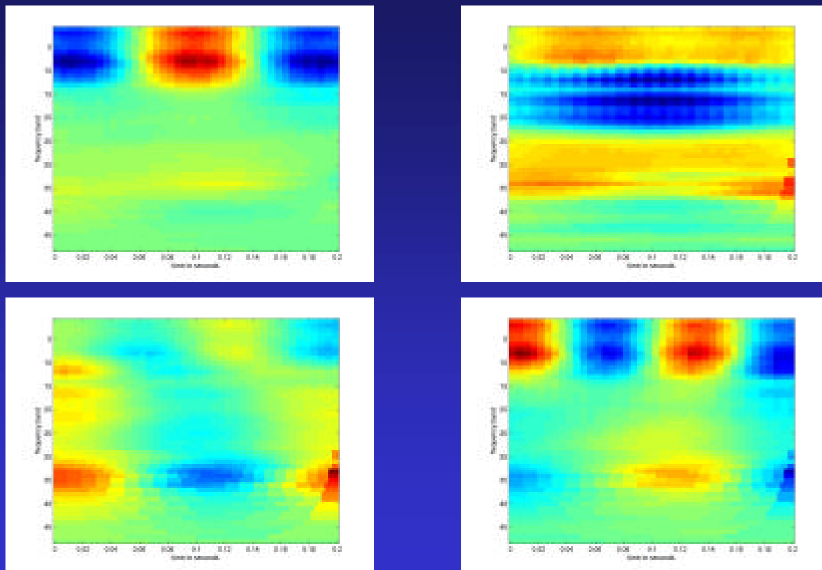
Copyright 2003-4 © Leslie Smith

50 Male samples: 200ms PCAs 1-4



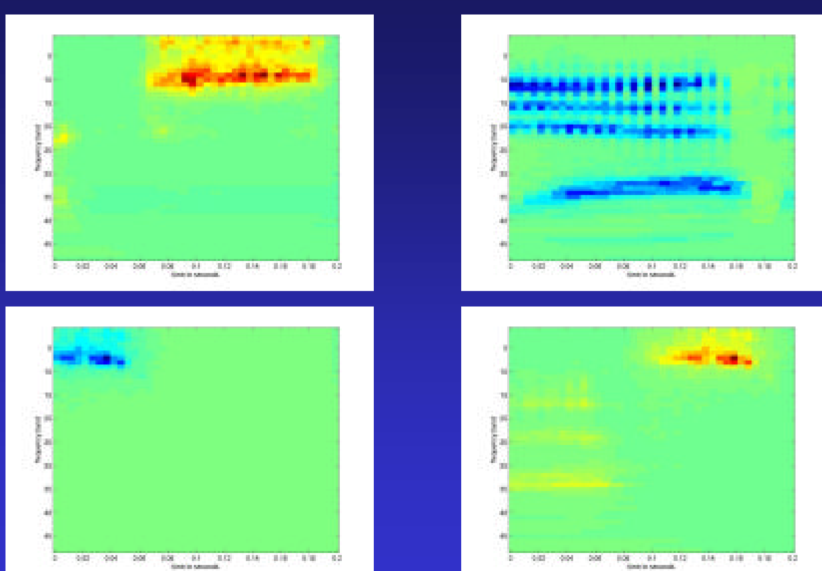
Copyright 2003-4 © Leslie Smith

PCAs 5-8



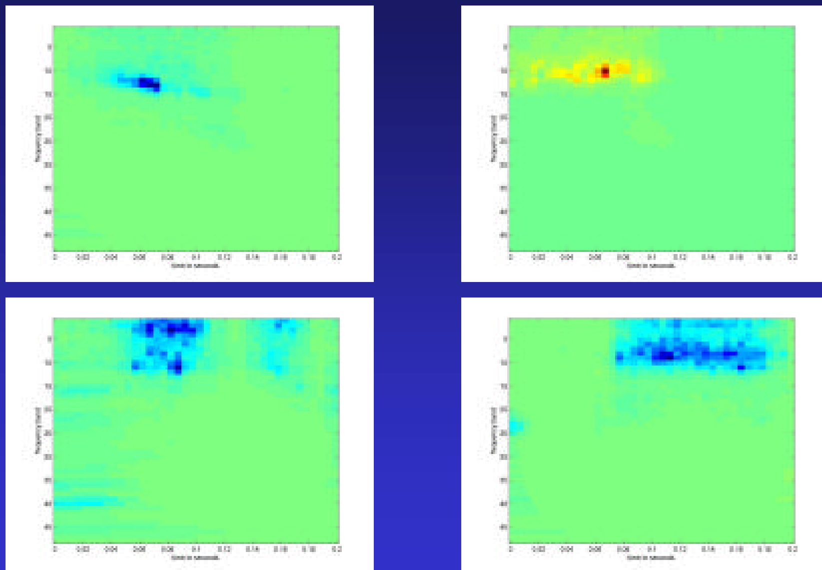
Copyright 2003-4 © Leslie Smith

200ms ICAs

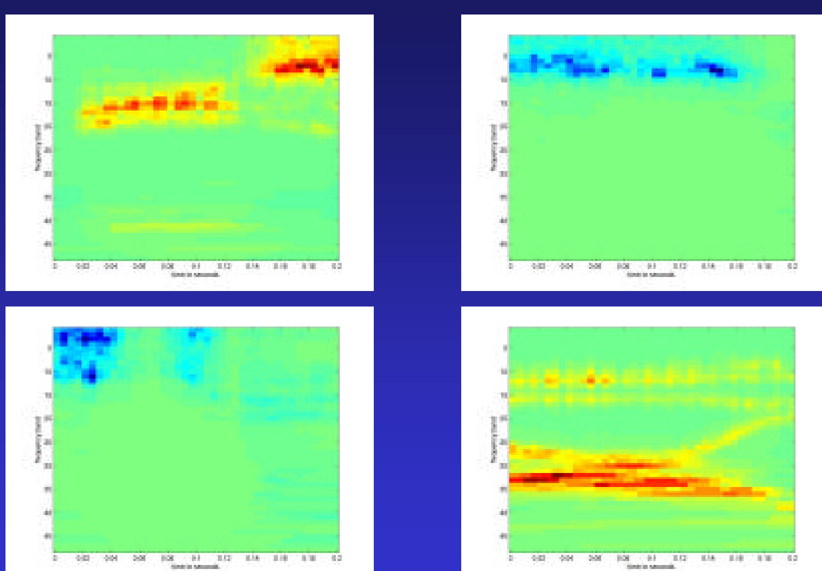


Copyright 2003-4 © Leslie Smith

More ICAs



...and more



Comparing FFT and Filterbank results

- PCAs provide differential-like operators
- ICAs provide feature detectors
- Very similar results form FFT and filterbank systems
  – Though order is different

- Actual order depends on exactly which bands we use
  • and exact dataset

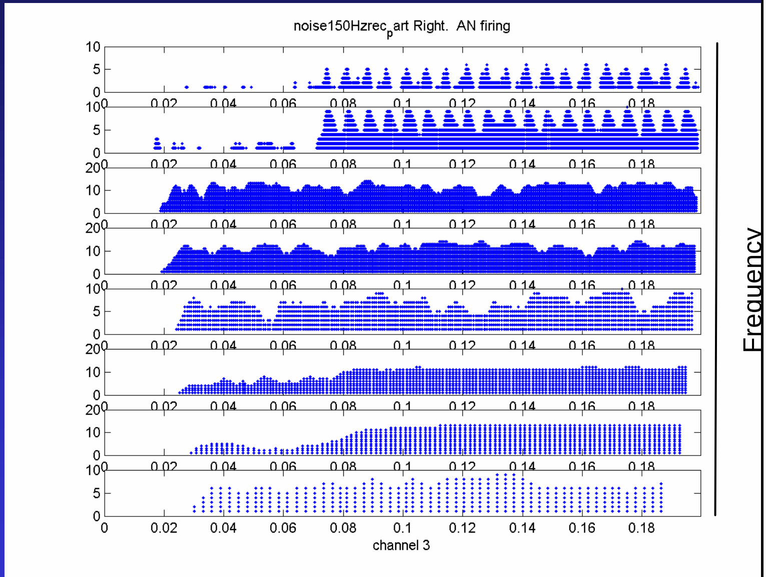
- Can experiment with PCAs ICAs and exact form of filterbank
  – Ganmmatone
  – Other more biologically plausible
  – Other more Engineering oriented (e.g. OTA based implementation, Chia and Collins at Oxford: ISCAS 2004.)

Shorter timescale components

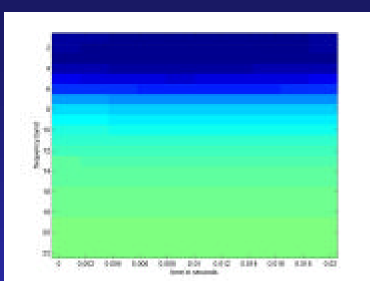
- Using the filterbank technique, we can find PCAs and ICAs for shorter timescales.
  – Which is harder with FFTs
- Do these give us insight into the short-term time structure of sounds?

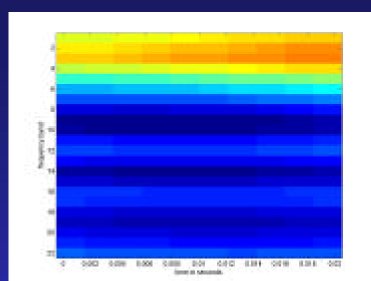
'AN' output showing short-term structure

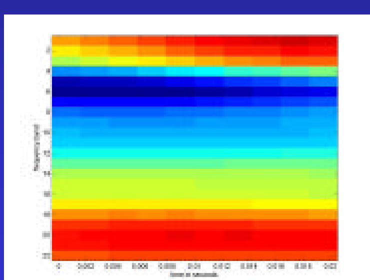
5 harmonics of 150 hz (3-7), plus bandlimited white noise (4KHz) SNR 0dB.

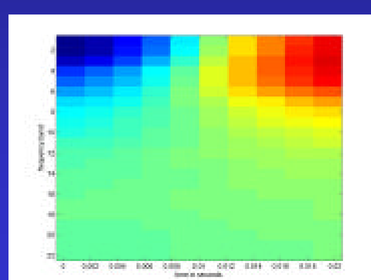


20ms PCAs: 1.5KHz to 7Khz, Female TIMIT data, 1-4

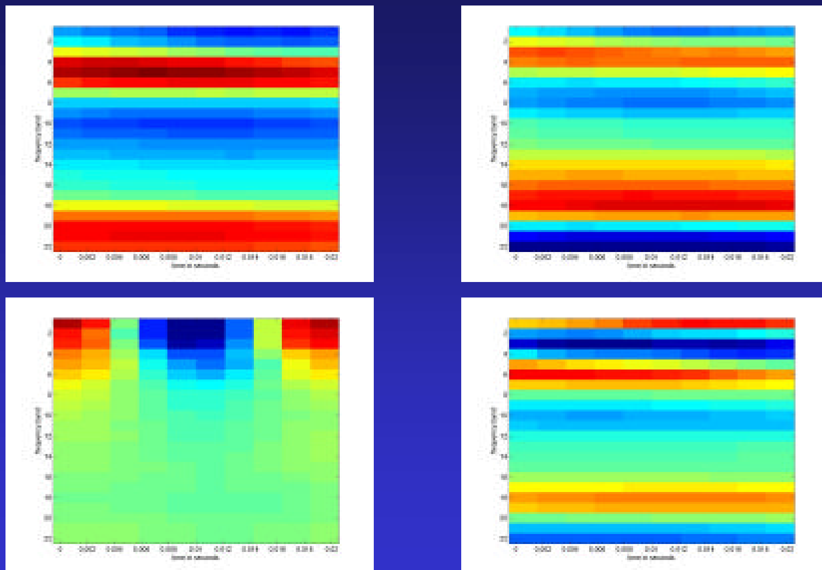






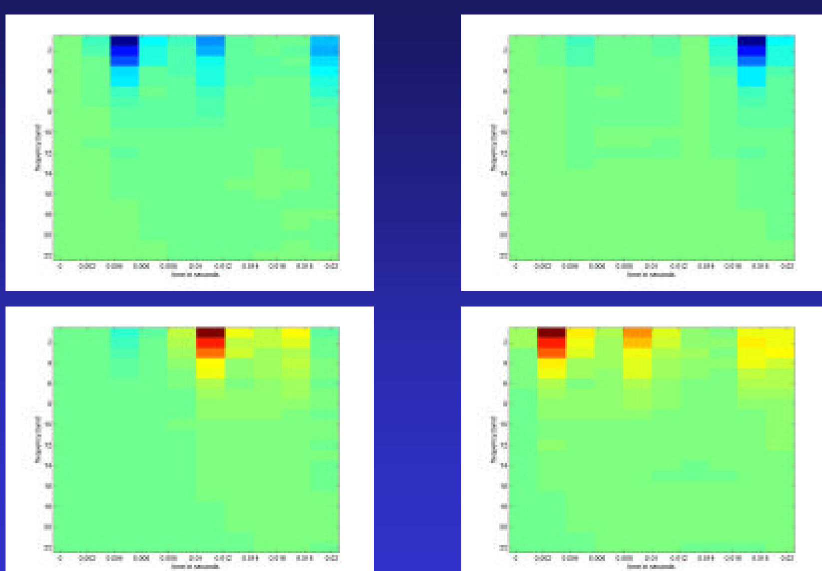


PCAs: 1.5KHz to 7Khz, 5-8



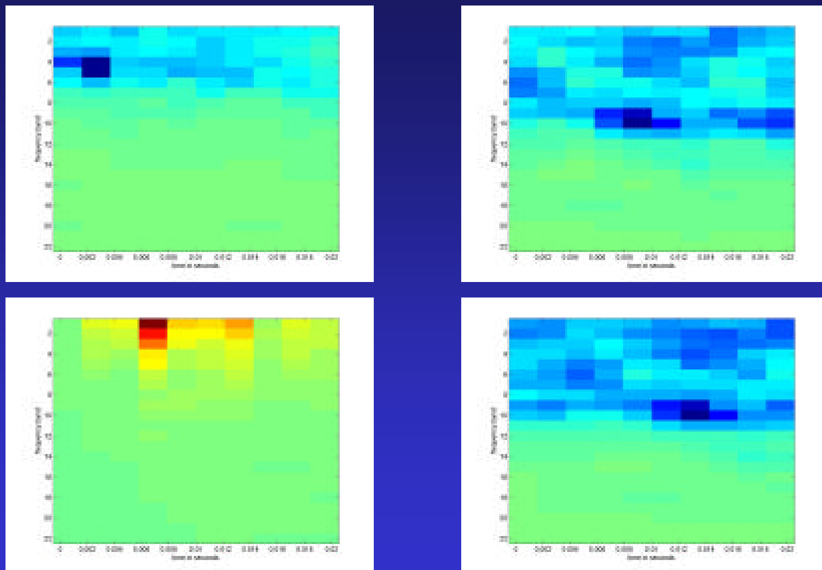
Copyright 2003-4 © Leslie Smith

20ms ICAs: 1.5Khz-7Khz



Copyright 2003-4 © Leslie Smith

ICAs: 1.5Khz-7Khz



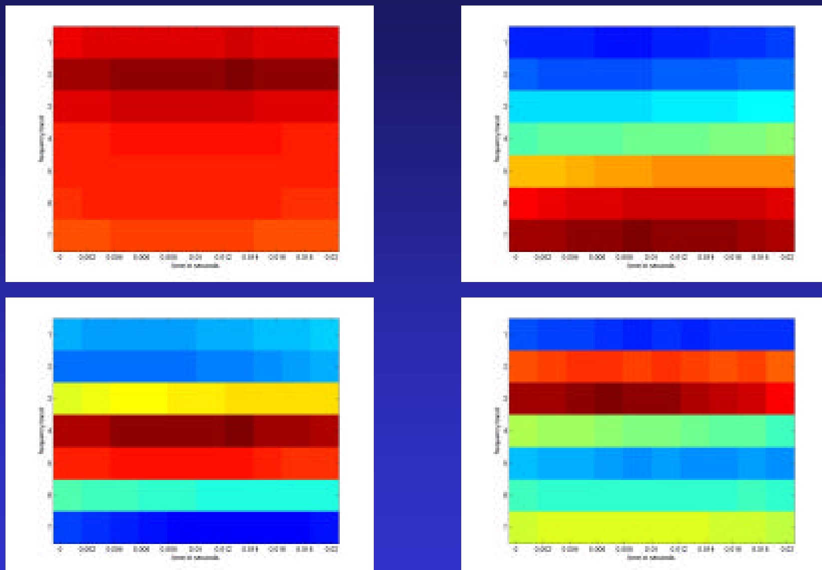
Copyright 2003-4 © Leslie Smith

Discussion

- These PCAs suggest
  – The interesting structure is at the high frequencies
  – (though later PCAs may contain lower frequency structure)
- The ICAs suggest the same
  – And that there is some envelope structure there as well
- We therefore try the same with a smaller range of frequency bands.

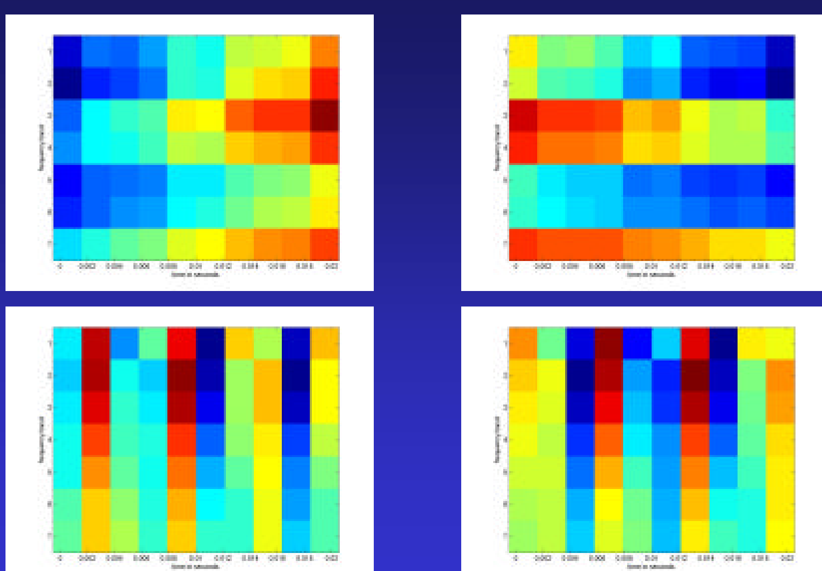
Copyright 2003-4 © Leslie Smith

Female, PCAs 2.07-3.46 KHz



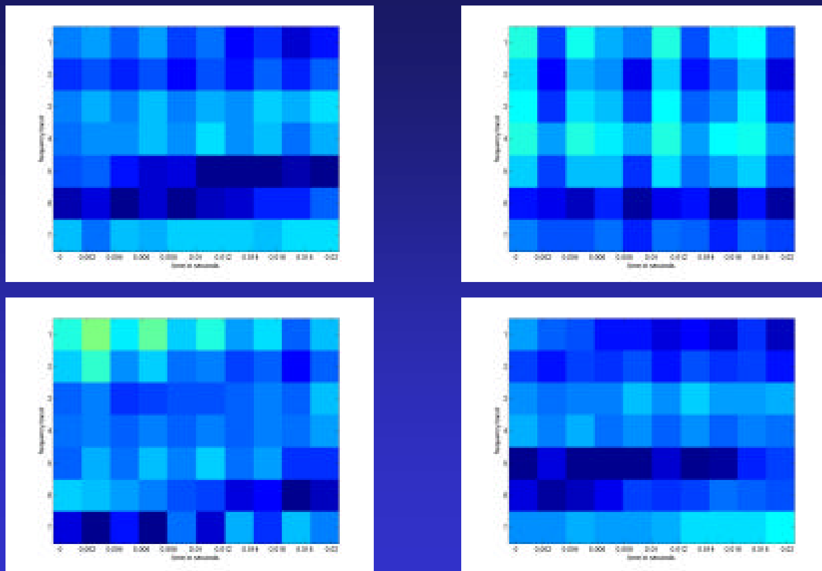
Copyright 2003-4 © Leslie Smith

Female, PCAs 2.07-3.46 KHz



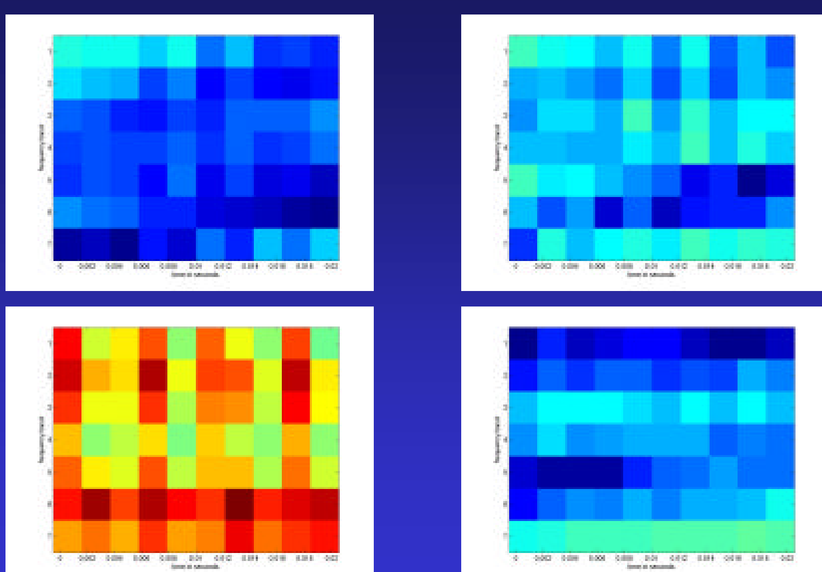
Copyright 2003-4 © Leslie Smith

Female, ICAs 2.07-3.46Khz



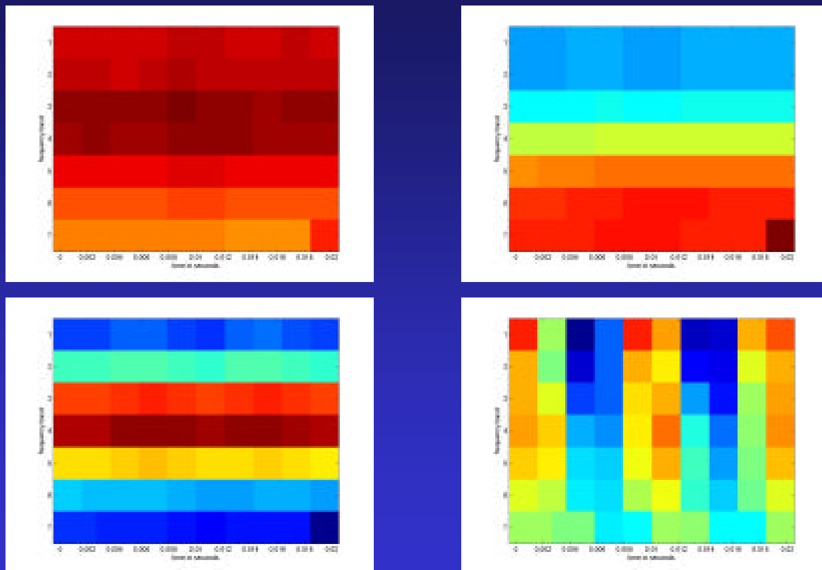
Copyright 2003-4 © Leslie Smith

Female, ICAs 2.07-3.46Khz

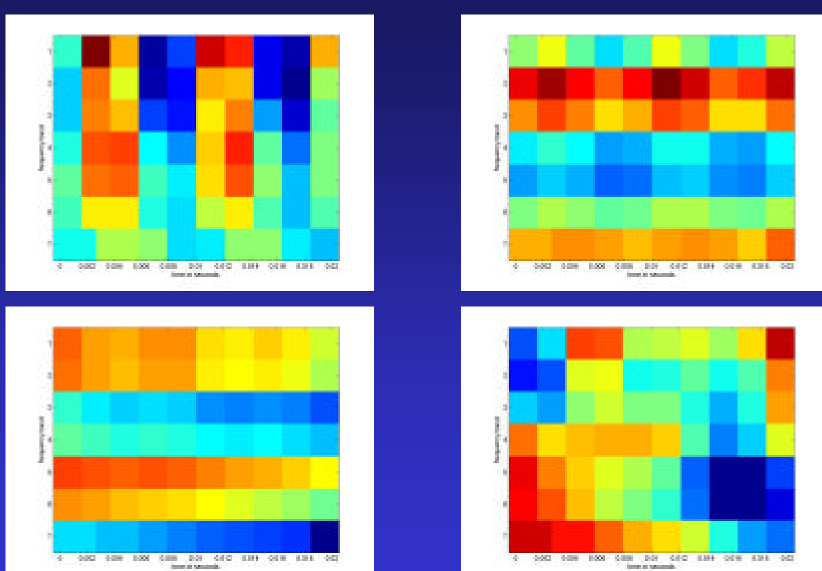


Copyright 2003-4 © Leslie Smith

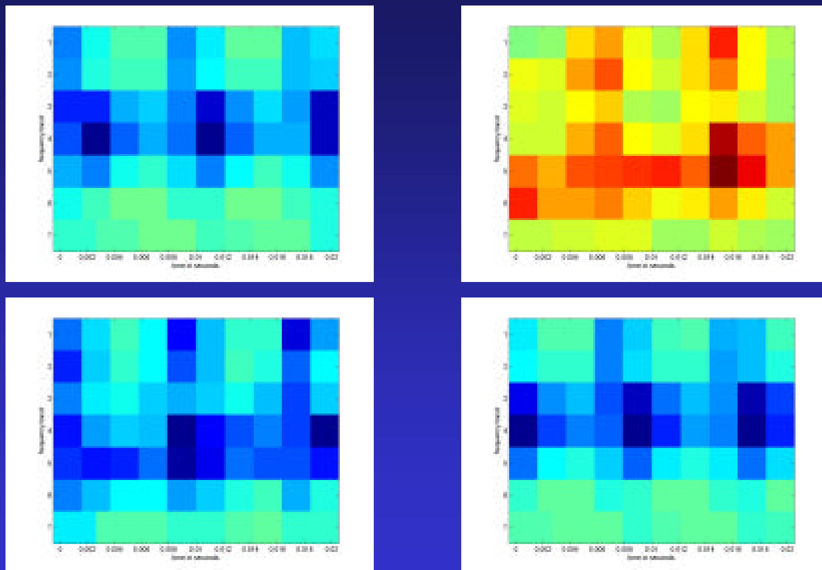
Male, PCAs, 1.4-2.4 Khz



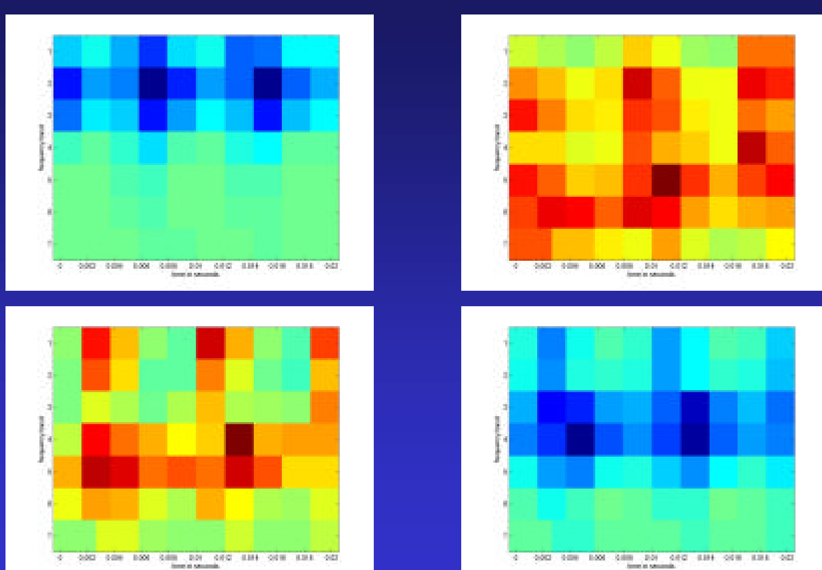
Male, PCAs,



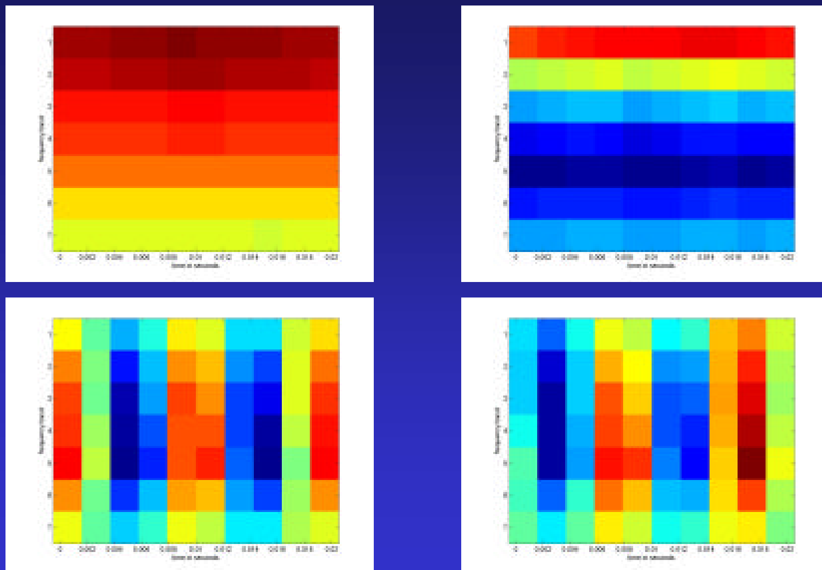
Male ICAs,



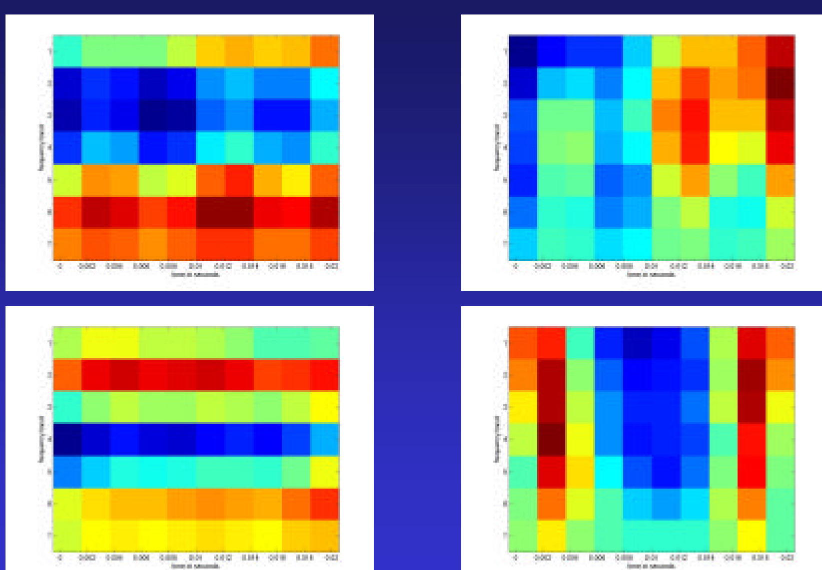
Male ICAs,



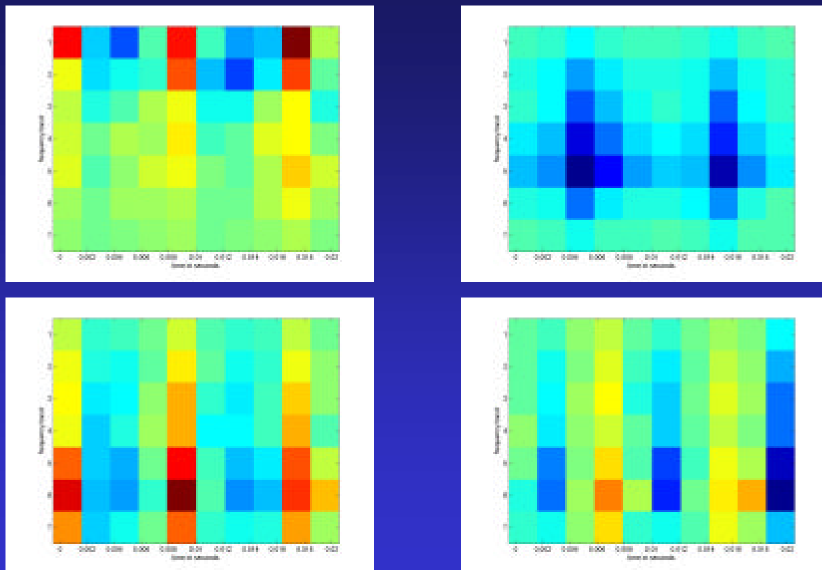
Male PCAs, 3-5Khz



Male PCAs 3-5Khz



Male ICAs 3-5Khz: showing AM

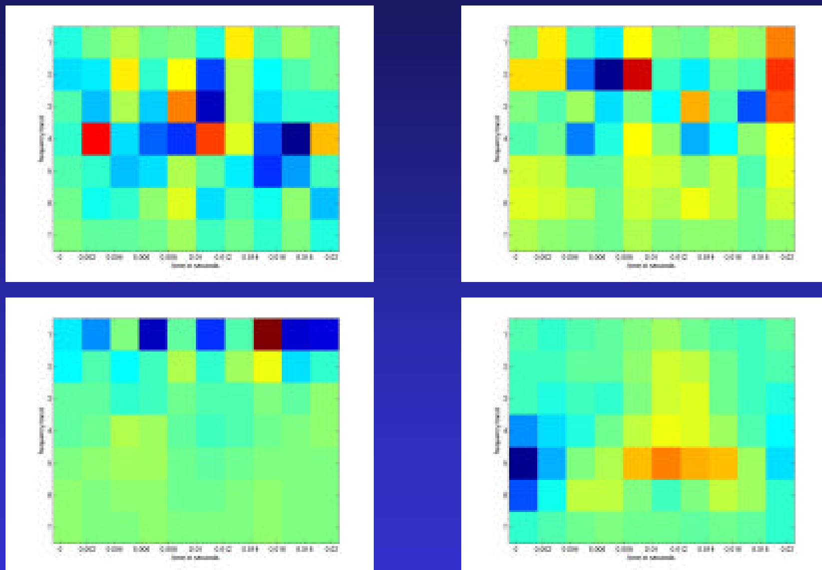


Copyright 2003-4 © Leslie Smith

Are we trying to create a new tartan?

Copyright 2003-4 © Leslie Smith

MaleICAs 3-5Khz: ?



Copyright 2003-4 © Leslie Smith

Short timescale statistics

- Amplitude modulation is very much in evidence
- Appears to be relatively uncorrelated across wide frequency bands
- But well correlated across nearby frequency bands
- Presumably from unresolved fundamental harmonics
  - About 180 Hz in female speech
  - About 100 Hz in male speech
- There are many ICAs produced
  - Same number as length of vector
  - And not ordered by variance explained

Copyright 2003-4 © Leslie Smith

Conclusions/Further work

- ICAs look more interesting than PCAs
- There’s lots of scope for further investigation of the ICAs
  - Different sounds
    - what sound would be most appropriate?
  - Different filters
  - Different frequency banding
  - Different timescales
  - Binaural sound

What do they suggest for sound perception?

- 1-dimensional results
  - PCA: shape by mean signal?
  - ICA: respond to specific features
- 2-dimensional results: 200ms
  - PCA: various derivative-like features
  - ICA: response to specific characteristic features
- 2-dimensional results: 20ms
  - PCA: derivatives again
  - ICA: amplitude modulation?

Other Applications

- Matched filters: providing a signal based feature description layer for interpretation of sound/speech
- For example:
  - Find ICAs for one particular class of signals,
    - for a specific front end
    - i.e. filterbank, etc.
      - (could be biologically inspired or not, as required)
  - Produce filters from this feature set
    - and recode the signals using them
  - Train an NN or HMM to recognise these sounds

Copyright 2003-4 © Leslie Smith