Biologically inspired robust onset detection

Leslie Smith, Department of Computing Science and mathematics, University of Stirling, Stirling FK9 4LA, Scotland, UK

lss@cs.stir.ac.uk

- Why onsets
  - and robust against what?
- How to detect onsets robustly

Why onsets?

- All sound sources start up
  - and often contain internal onsets as well
- Useful grouping cue
  - most sounds are wide-band
  - the energy in different parts of the spectrum generally starts at the same time
  - so onsets that occur at about the same time in different parts of the spectrum can be used to group sound energies together
    - co-occurring onsets are like co-incident collinear edge segments
- First sound arrival is from direct path
  - so sound energy at onset arrives direct from sound source
  - usable for source direction finding (both ITD and IID)
- Robust onset detection provides an ecologically useful cue for source grouping and direction finding.

Robustness and onsets

- Detecting a signal onset from a zero base is straightforward
  - but noise can add false onsets
- However:
  - there is normally a noise level from which the signal rises
    - particularly when there are concurrent sound sources
  - direct path onsets are often followed by reflection mediated increases in level
- A robust onset detector would
  - find onsets in noise, and
  - not be confused by reflections

Onset detection techniques

- Simple “first difference techniques”
  - often used in music research
  - Bilmes 93, Goto 95, 96, Schreirer 96
- Optimal filter techniques
  - extension (improvement) to first difference
  - Smith 96, Tait 97, Majolt 2001
- Expectation based techniques
  - different extension to first difference
  - Huang et al 97

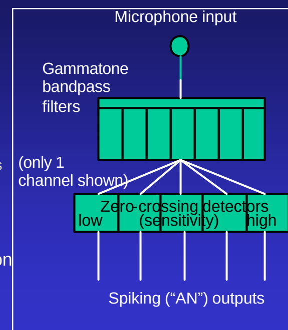
Problems:
  - what exactly constitutes an onset?
    - we know what we mean psychophysically, hard to define in a signal processing sense!
  - real onsets are very variable in form
  - can be short (1-2ms) or longer (up to 50ms)

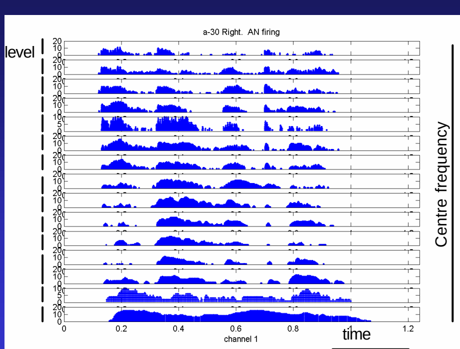
Biologically inspired techniques

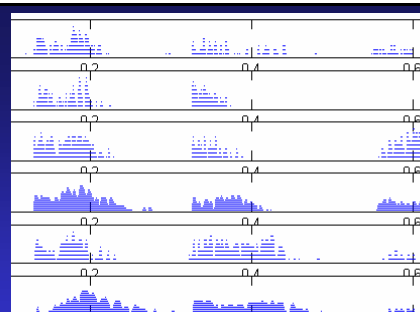
- The early auditory system is sensitive to onsets
  - AN type 1 fibers
  - CN onset cells (of various forms)
- Precedence effect and our sensitivity to onsets illustrates robust capability
- What underlies this?
  - wide dynamic range of the middle ear/inner ear/auditory nerve system
  - spectrum based filtering, allowing onsets in one part of the spectrum not to be swamped
  - characteristics of the neurons of the CN
    - e.g. non-constant leakiness
    - aspects of cell morphology
- We model the wide dynamic range, the filtering, and the high (but not non-constant) leakiness

Modelling the wide dynamic range

- Each side input starts with a transducer (microphone)
- followed by a gammatone filterbank
  - multichannel, cochlea-like response (static)
- followed by spike generation
  - on positive-going zero-crossings
  - geometric range of sensitivities
    - spike generation predicated on pre-zero-crossing level
    - levels increase geometrically
- result is AN-like representation of sound signal
  - with a form of log compression
  - similar to Ghitza 1986



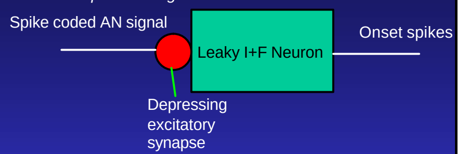




Although signal is "log-compressed", representation is made up from sequences of spikes (horizontal lines), which can be processed independently

Detecting onsets from these AN signals (1)

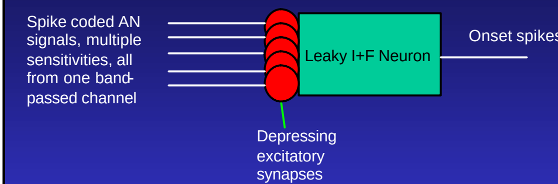
First attempt: 1 onset cell per "AN" signal



Generates multiple onset signals, in phase with AN signal, low latency, but rather sensitive to noise.

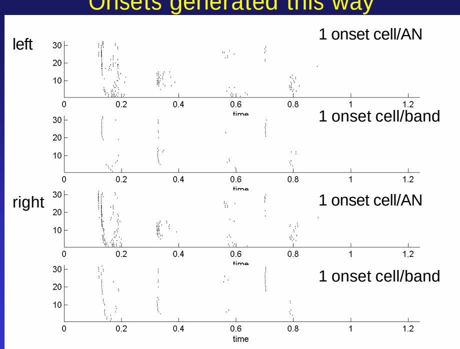
Detecting onsets from these AN signals (2)

Second attempt: 1 onset cell per band- passed channel

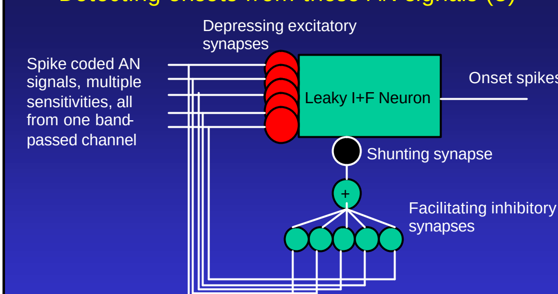


Generates a single onset signal, except at low frequencies, where multiple onsets are sometimes generated. Tends to miss onsets that start from a medium level.

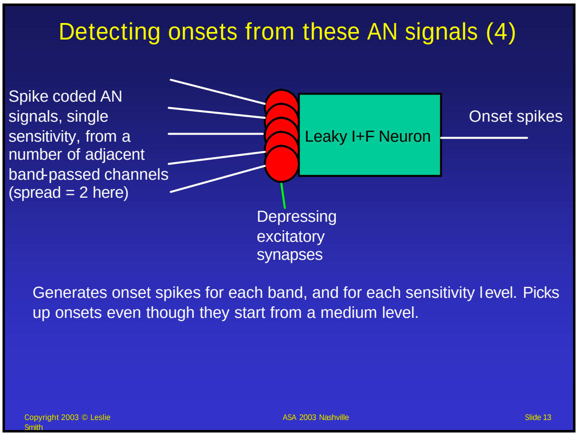
Onsets generated this way

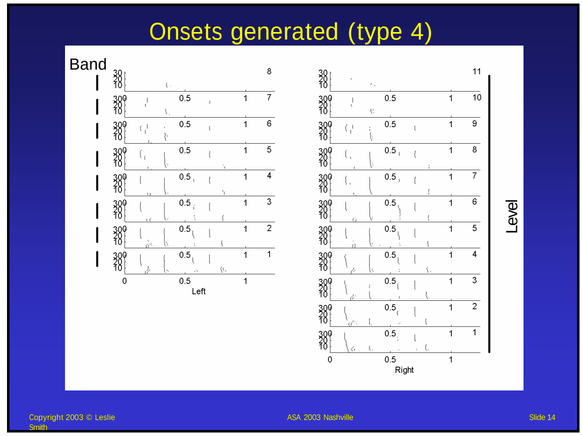


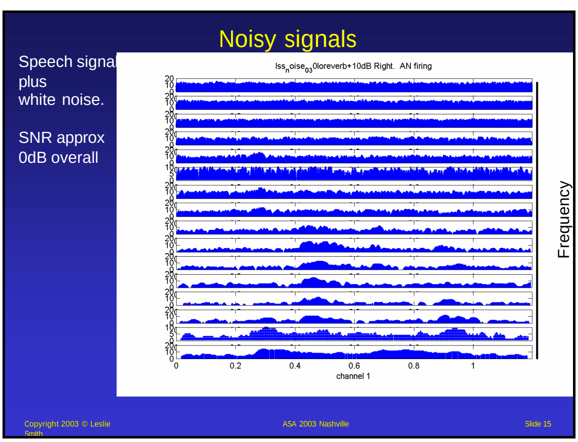
Detecting onsets from these AN signals (3)

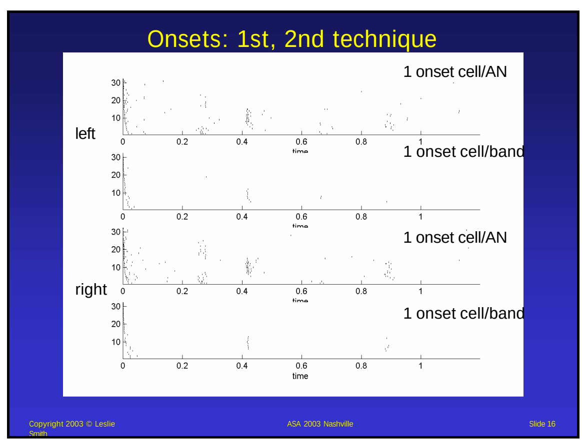


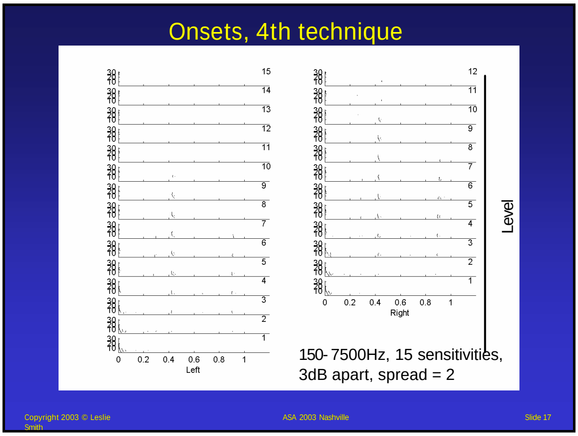
Generates a single onset signal even at low frequencies. Still tends to miss onsets that start from a medium level. Complicated as well. Precise time constants in shunting synapse matter.

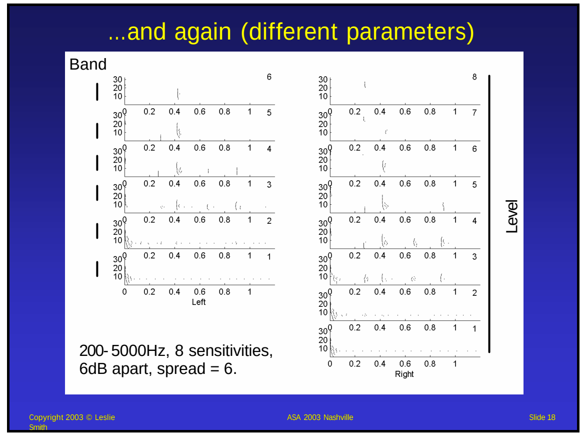

Detecting onsets from these AN signals (4)
Spike coded AN signals, single sensitivity, from a number of adjacent band-passed channels (spread = 2 here)
Leaky I+F Neuron
Onset spikes
Depressing excitatory synapses
Generates onset spikes for each band, and for each sensitivity level. Picks up onsets even though they start from a medium level.
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 13

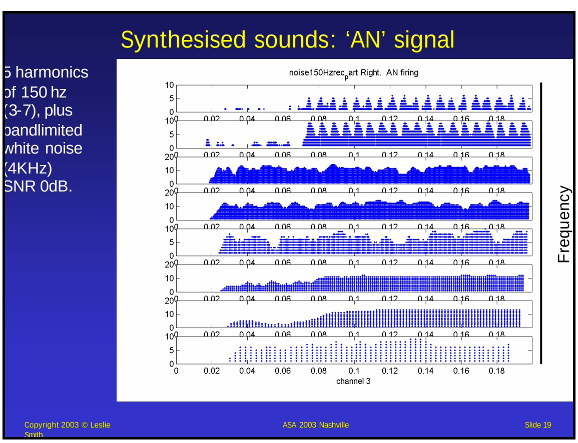

Onsets generated (type 4)
Band
Left
Right
Level
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 14

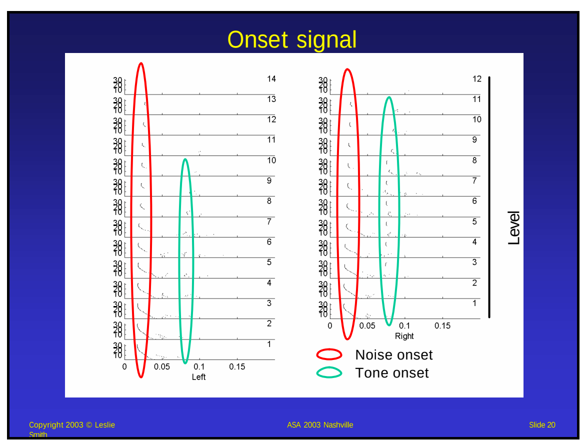

Noisy signals
Speech signal plus white noise.
SNR approx 0dB overall
channel 1
Frequency
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 15


Onsets: 1st, 2nd technique
1 onset cell/AN
1 onset cell/band
1 onset cell/AN
1 onset cell/band
left
right
time
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 16


Onsets, 4th technique
Left
Right
Level
150-7500Hz, 15 sensitivities,
3dB apart, spread = 2
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 17

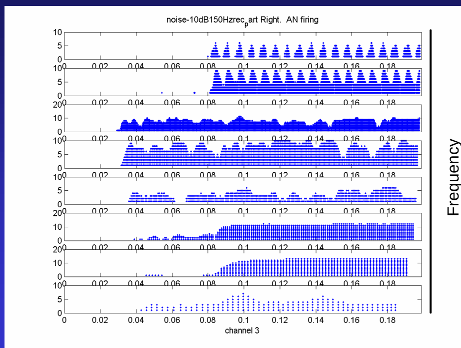

...and again (different parameters)
Band
Left
Right
Level
200-5000Hz, 8 sensitivities,
6dB apart, spread = 6.
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 18


Synthesised sounds: 'AN' signal
5 harmonics
of 150 hz
(3-7), plus
bandlimited
white noise
(4KHz)
SNR 0dB.
noise150Hzrec_part Right. AN firing
channel 3
Frequency
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 19

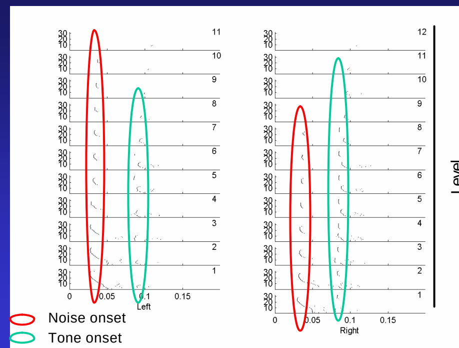

Onset signal
Left
Right
Level
Noise onset
Tone onset
Copyright 2003 © Leslie Smith
ASA 2003 Nashville
Slide 20

AN with noise 10dB lower

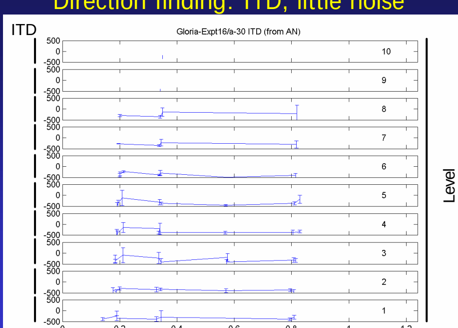
5 harmonics of 150 hz (3-7), plus bandlimited white noise (4KHz) SNR 10dB.



Onsets found

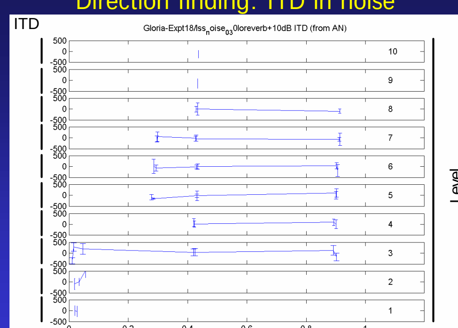


Direction finding: ITD, little noise



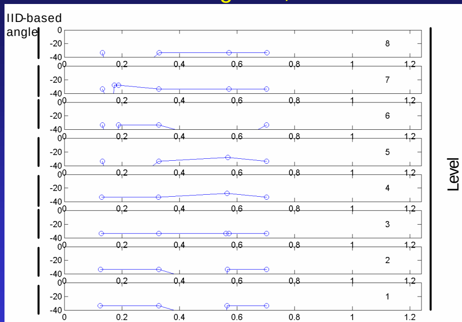
Signal is at -30 degrees = -250us

Direction finding: ITD in noise



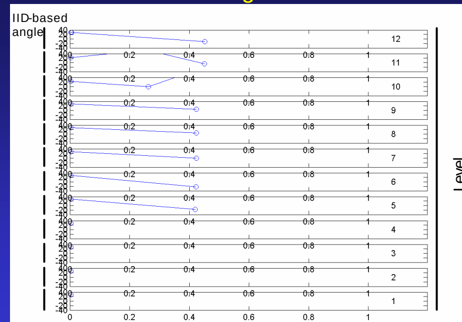
Signal is at 0 degrees (noise at 30 degrees)

Direction finding: IID, little noise



Signal is at -30 degrees

Direction finding: IID in noise



Signal is at 0 degrees (noise at 30 degrees)

Conclusions

- We have developed a biologically -inspired onset detection technique which is
  - able to deal with wide dynamic range
  - able to cope with onsets starting from a non-zero level
- This has been used to cluster onsets
  - and hence to find the directions of concurrent sound sources
- Issues: on onsets
  - Would probabilistic AN spike generation widen the dynamic range?
  - Would incorporating inhibition into type 4 onset detection assist?
  - Need to characterise the precise capabilities of this technique fully.
- On onset-feature based streaming
  - how to string grouped features together over time
    - use direction of sound?
    - Use characteristics of the features themselves: level, signal energy structure, post -onset envelope