

Spectrotemporal Gabor filters for feature detection

L.S. Smith, A.K Abel

Computing Science and Mathematics

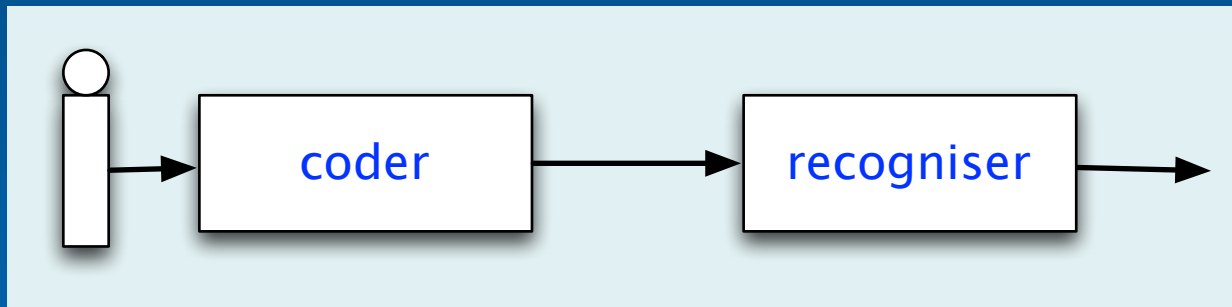
University of Stirling, Stirling UK

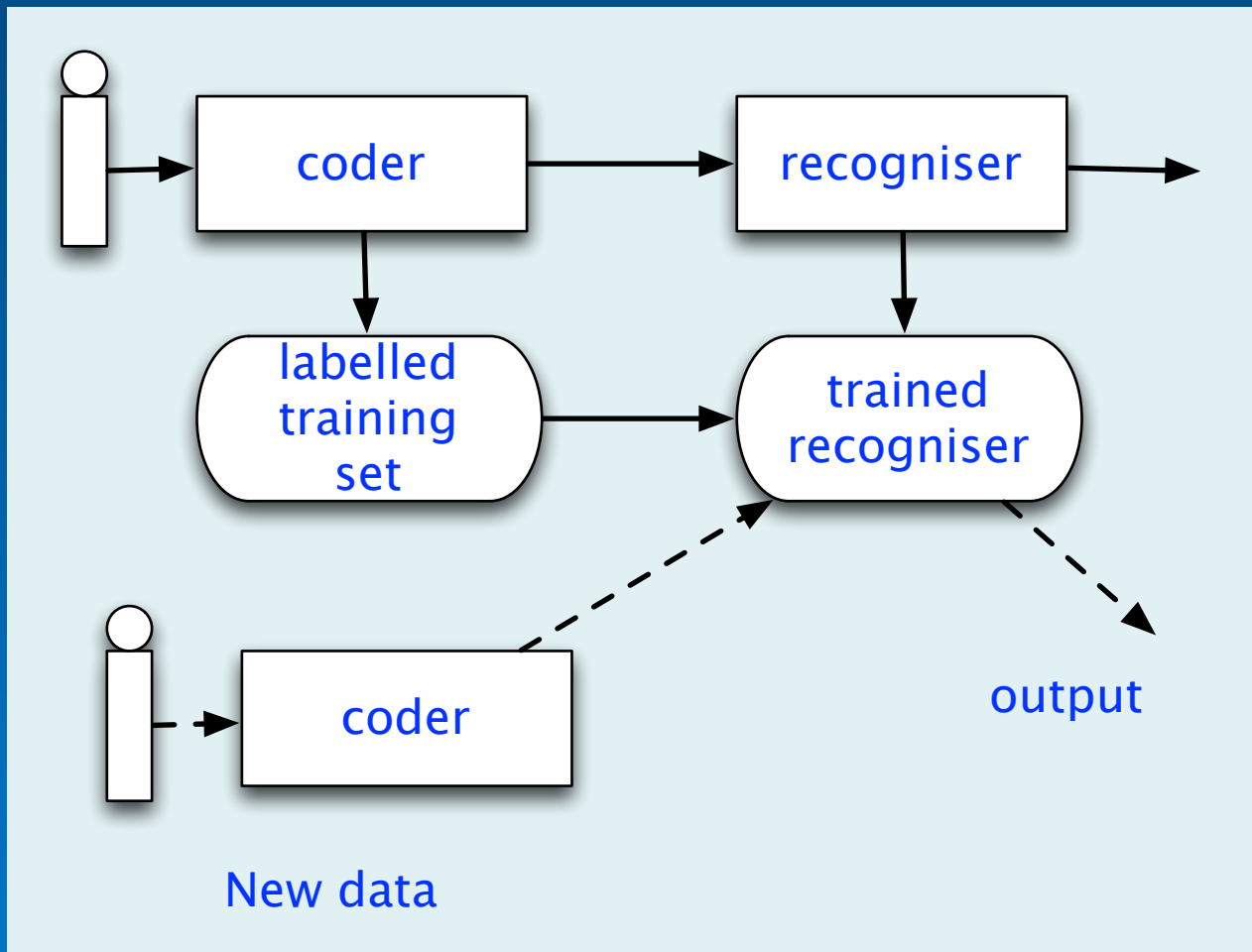


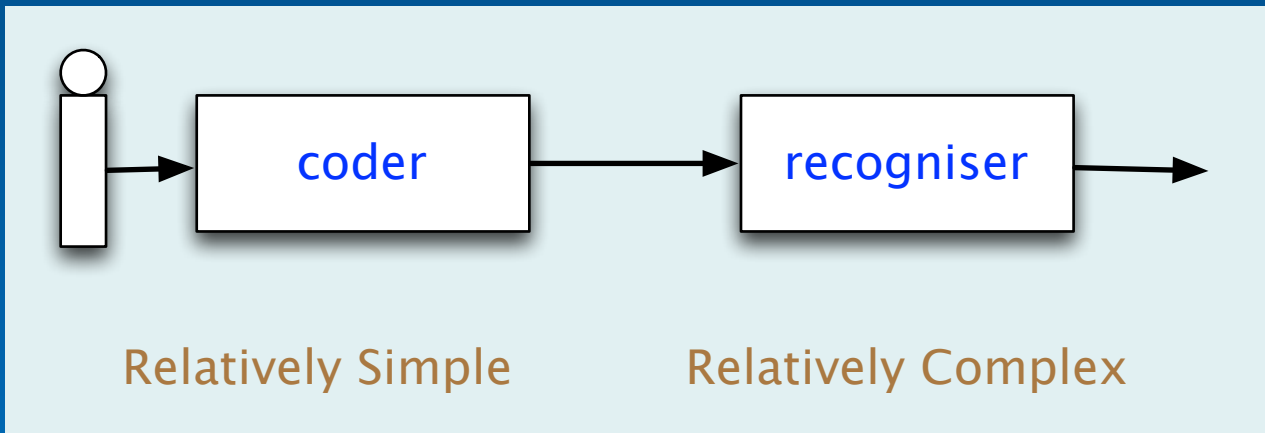
**UNIVERSITY OF
STIRLING**

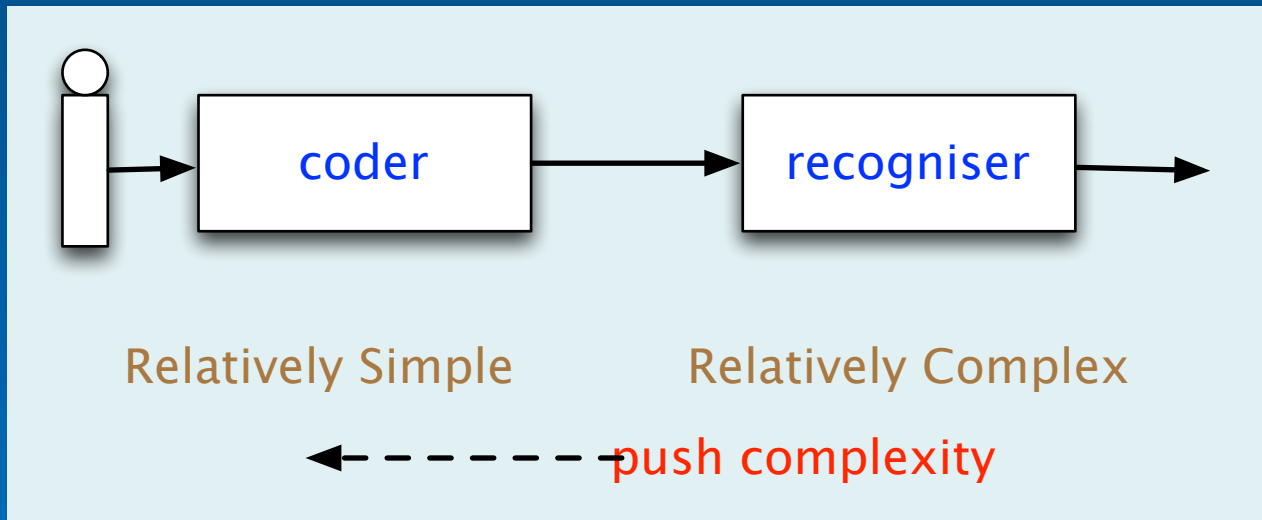
Overview

- Motivation:
 - pushing the complexity in sound/speech interpretation back to the preprocessing
- Preprocessing technique:
 - AN, Onset, Gabor filters
 - Self-organised clustering of segment vectors
- Some initial results using the above.

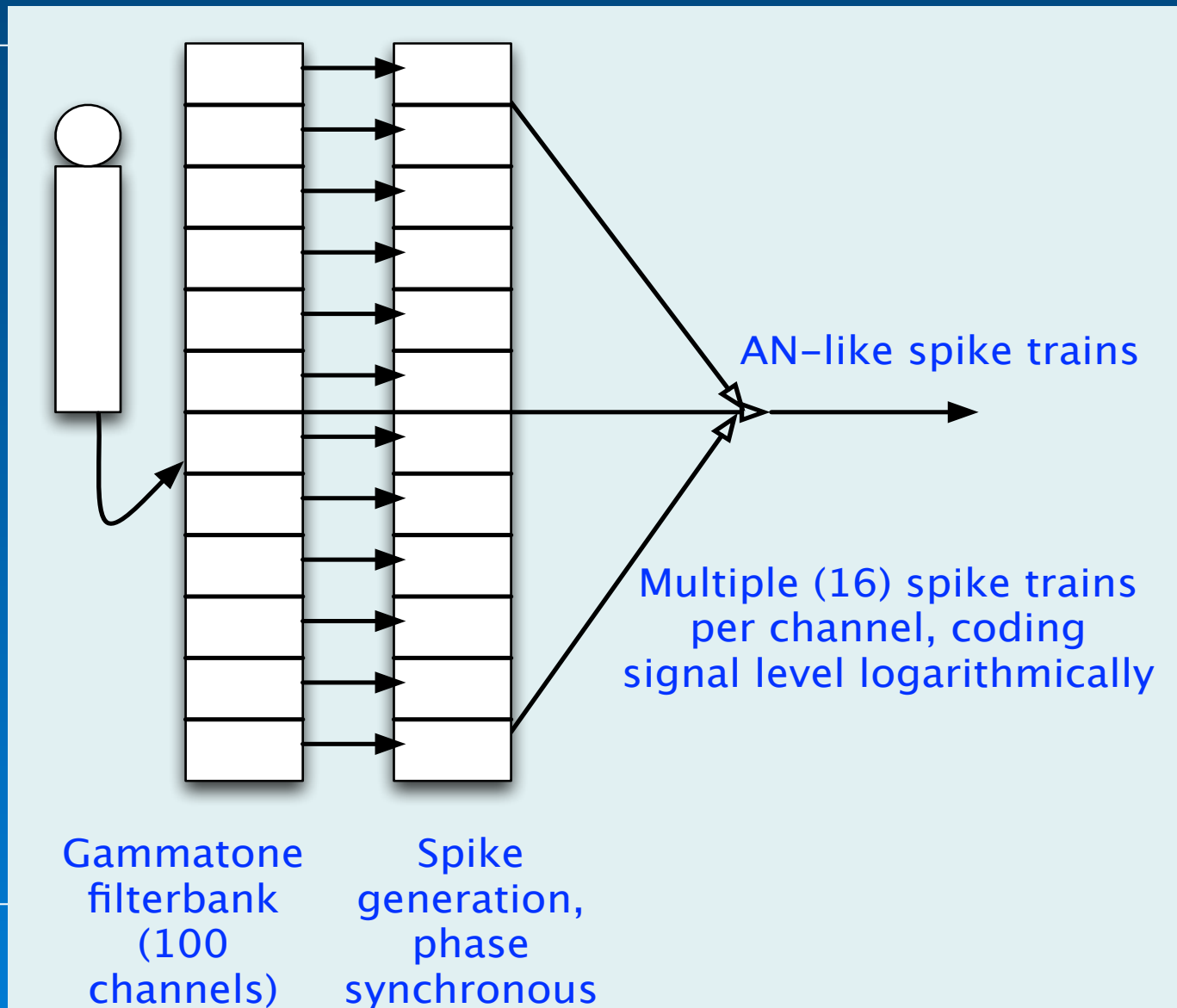




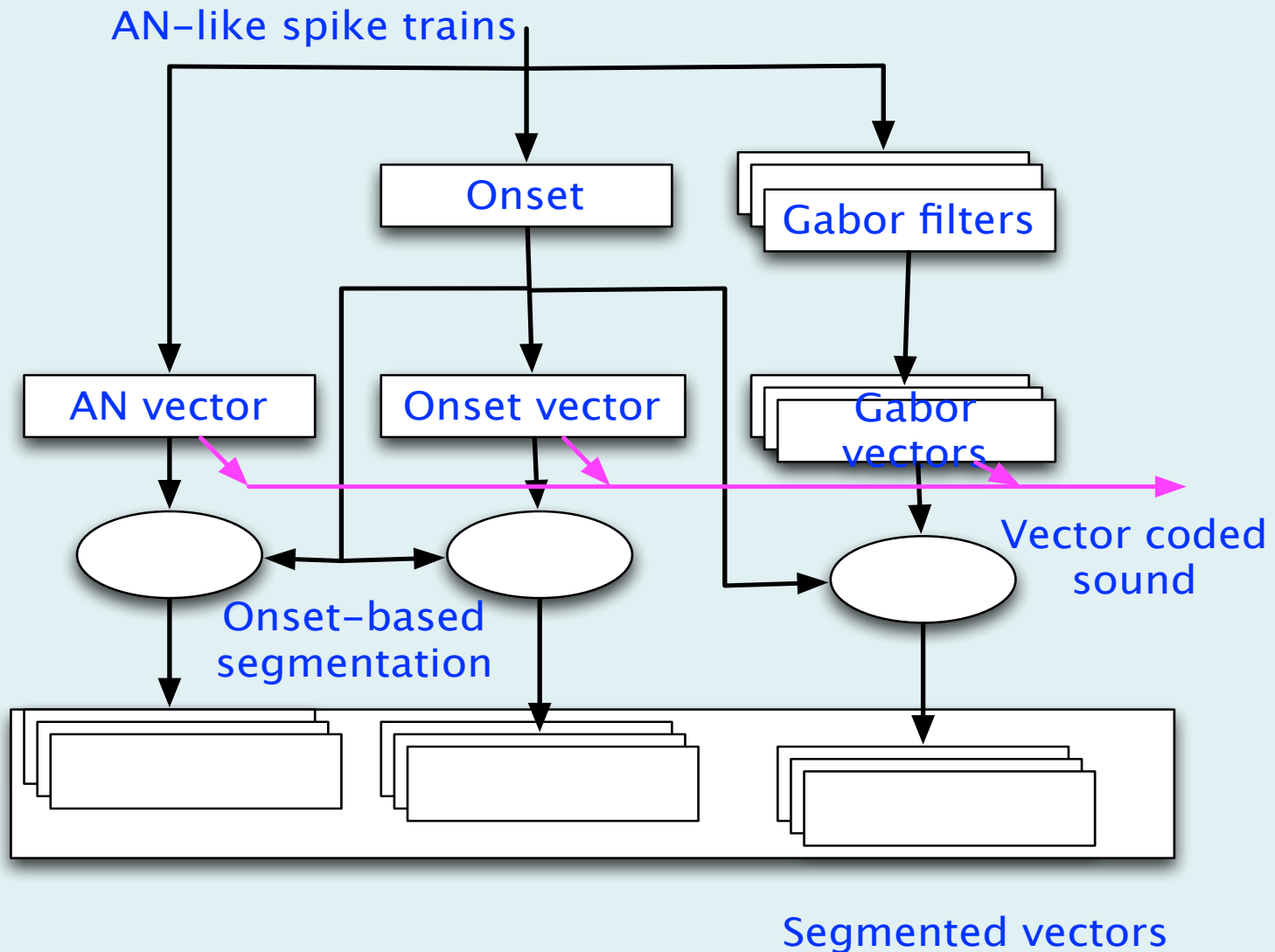




Bio-inspired initial processing



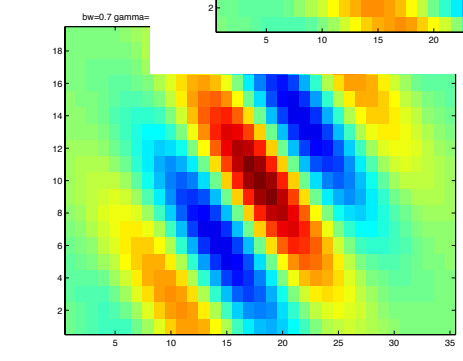
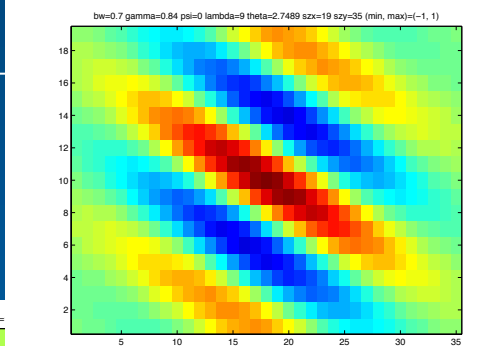
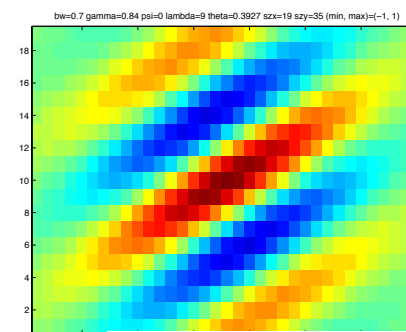
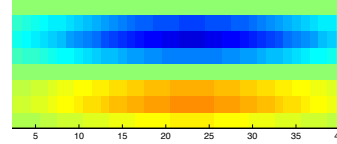
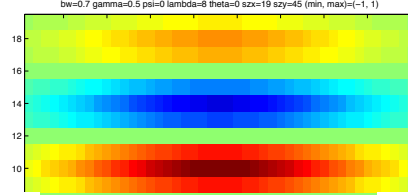
From spikes to a vector coding



Why these features?

- AN-like vectors
 - Contains basic information about the spectro-temporal distribution of the sound energy: like MFCC
- Onset vectors
 - Particularly describes how the sound start up
 - Also useful for segmentation
- Gabor feature vectors
 - Highly parameterised set of features
 - Localised in spectrum and time

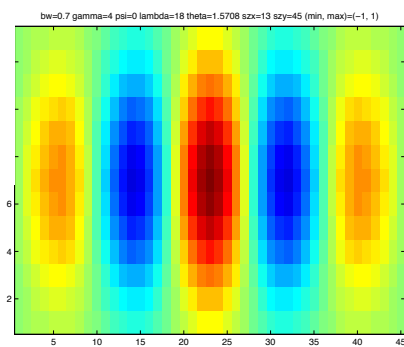
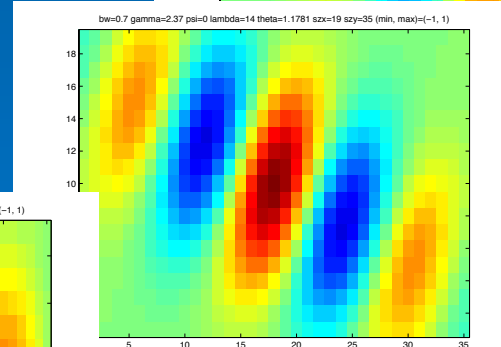
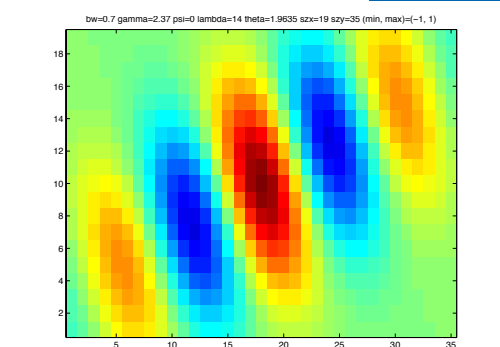
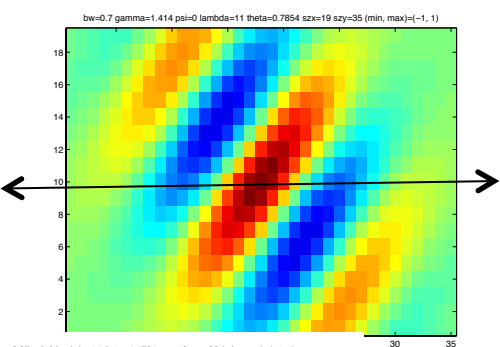
Set of 8 Gabor filters



Time →
Band ↑

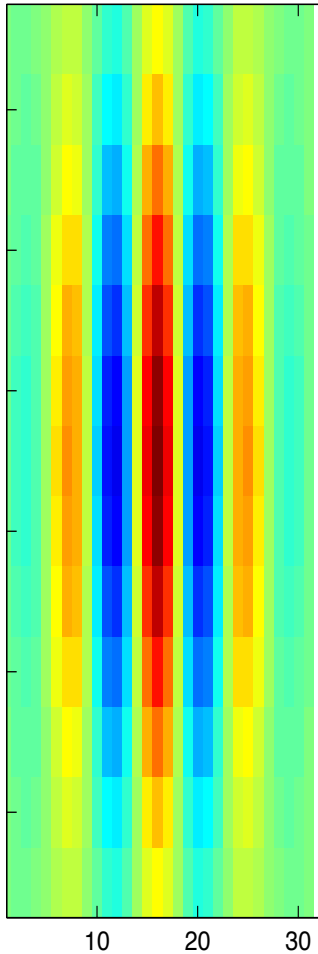


40 ms

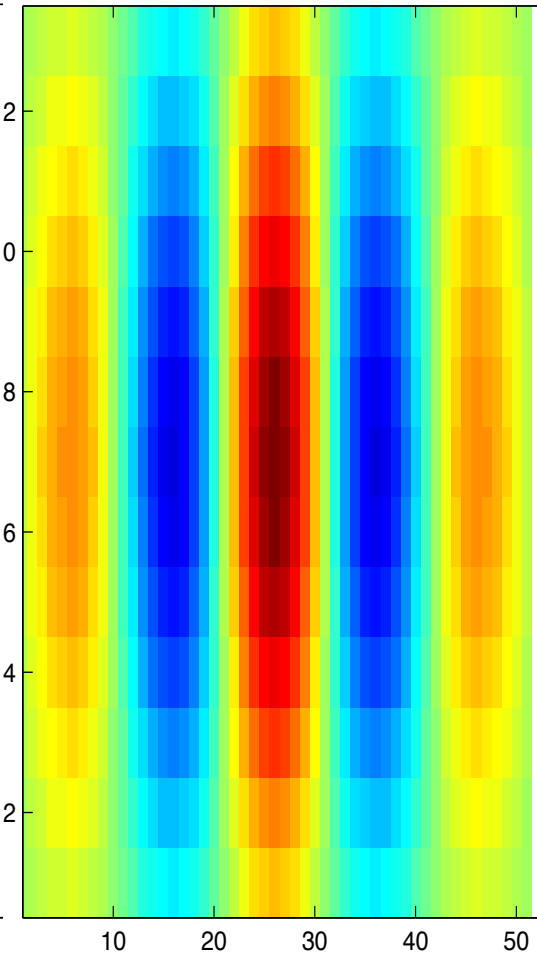


Gabors: same orientation

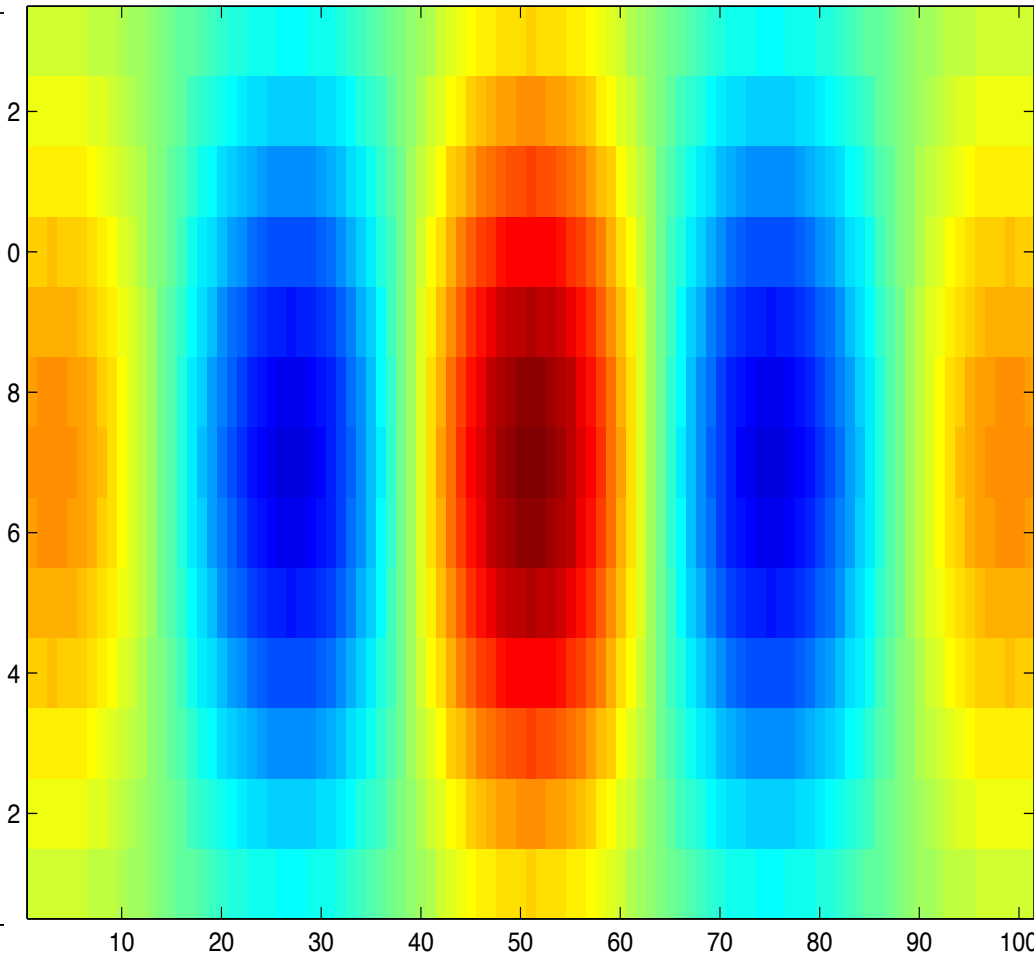
bw=0.7 gamma=2 psi=l



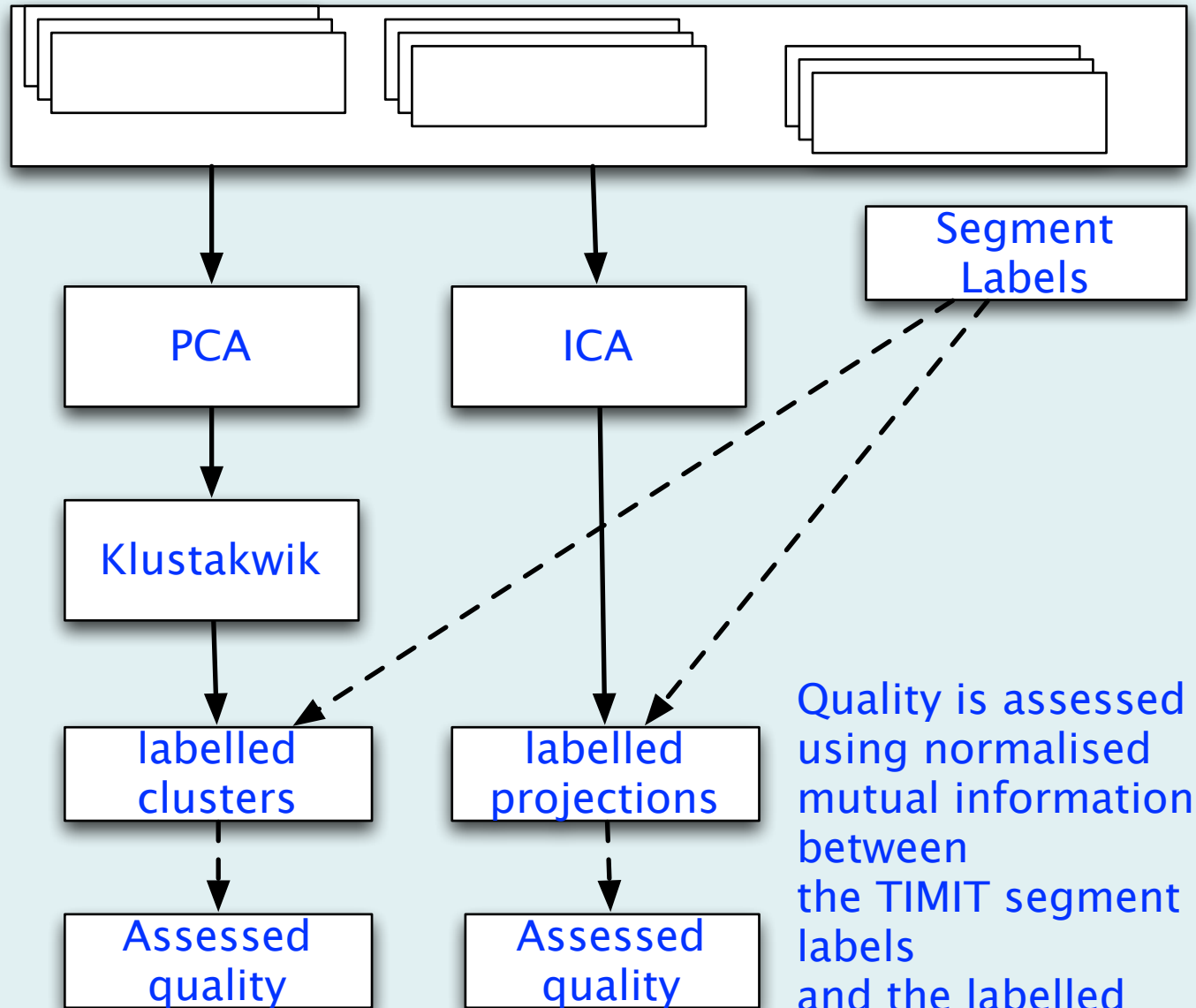
bw=0.7 gamma=4 psi=0 lambda=21 theta=1



bw=0.7 gamma=10 psi=0 lambda=50 theta=1.5708 szx=13 szy=101 (min, max)=(-1, 1)



Segmented vectors



Quality is assessed using normalised mutual information between the TIMIT segment labels and the labelled segments in the clusters

Un-supervised clustering

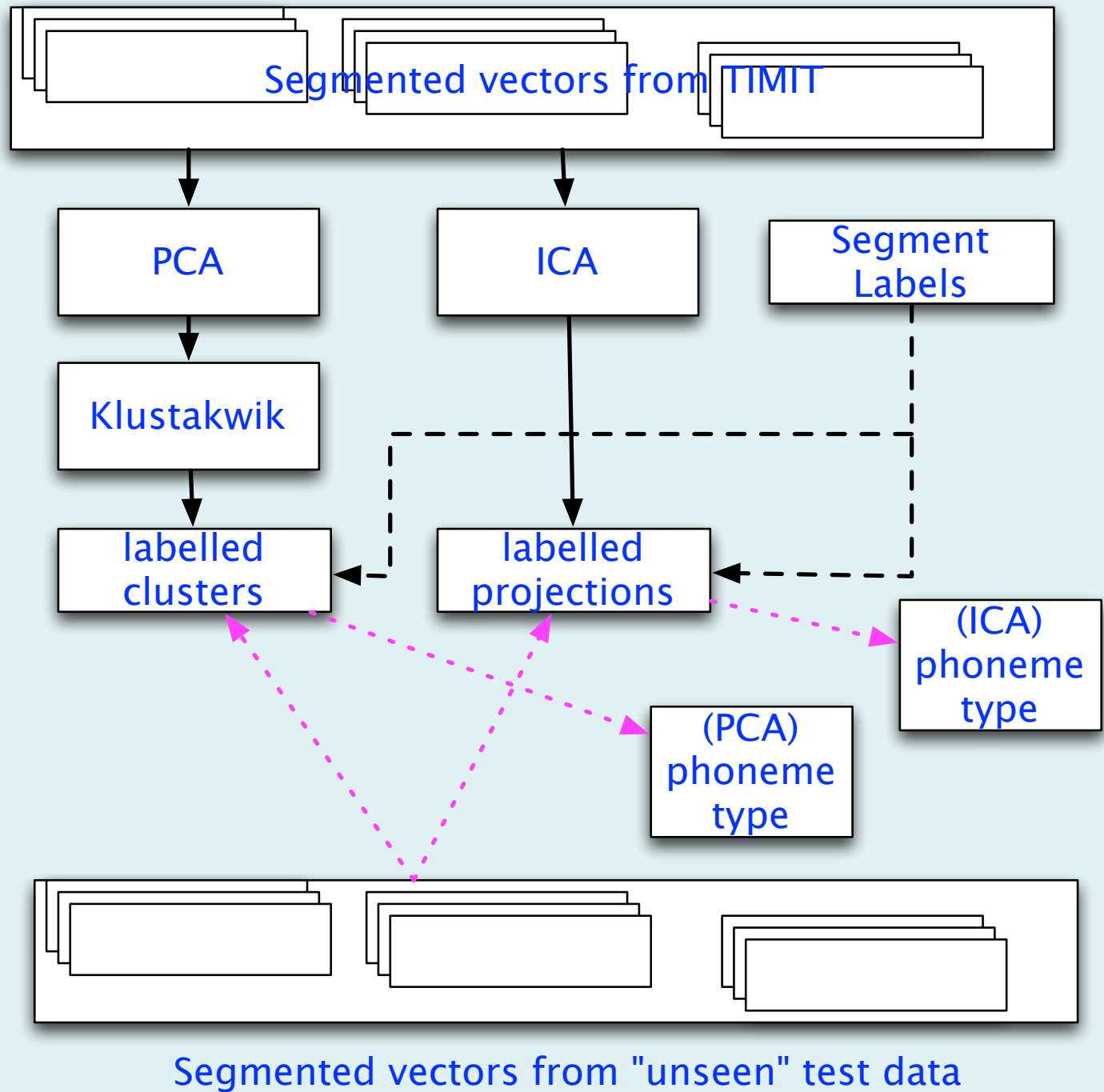
Preliminary results: TIMIT

- Whole TIMIT dataset. Segments are 75 or 100ms long.
- Assessment: Normalised mutual Information between clusters and TIMIT labels
- Coding:
 - AN code, + Onset code + 8 Gabors
 - AN code + Onset Code + 3 Gabors

Normalised Mutual Information

	AN+ Onset +8 Gabors		AN + Onset + 3 Gabors	
	male	female	male	female
PCA	0.431	0.499	0.479	0.544
ICA	0.325	0.368	0.376	0.433

Classification of new data



Why this shouldn't work well...

- Training is with TIMIT, single gender.
- Test is unseen data, same gender:
- Training is unsupervised: simple K means clustering (PCA), or ICA projections as centres
- No attempt to use context.

Example results (3 Gabors)

Male speaker: /Kiz/

- PCA based: /k, eh/ /t, ix, kcl/ /f/
- ICA based: /s/ iy/ /s/ /Xans/ (sounds: Ans/)
- PCA based /iy/ /ae/ /f/ /sh/
- ICA based /t, eh/ /ae/ /s/ /m, s/

Female speaker /ak/

- PCA based: /ay/ /k,aa/ /ix, kcl, k/
- ICA based: /b,ay/ g,r,iy/ /w,aa/

Female speaker /od/

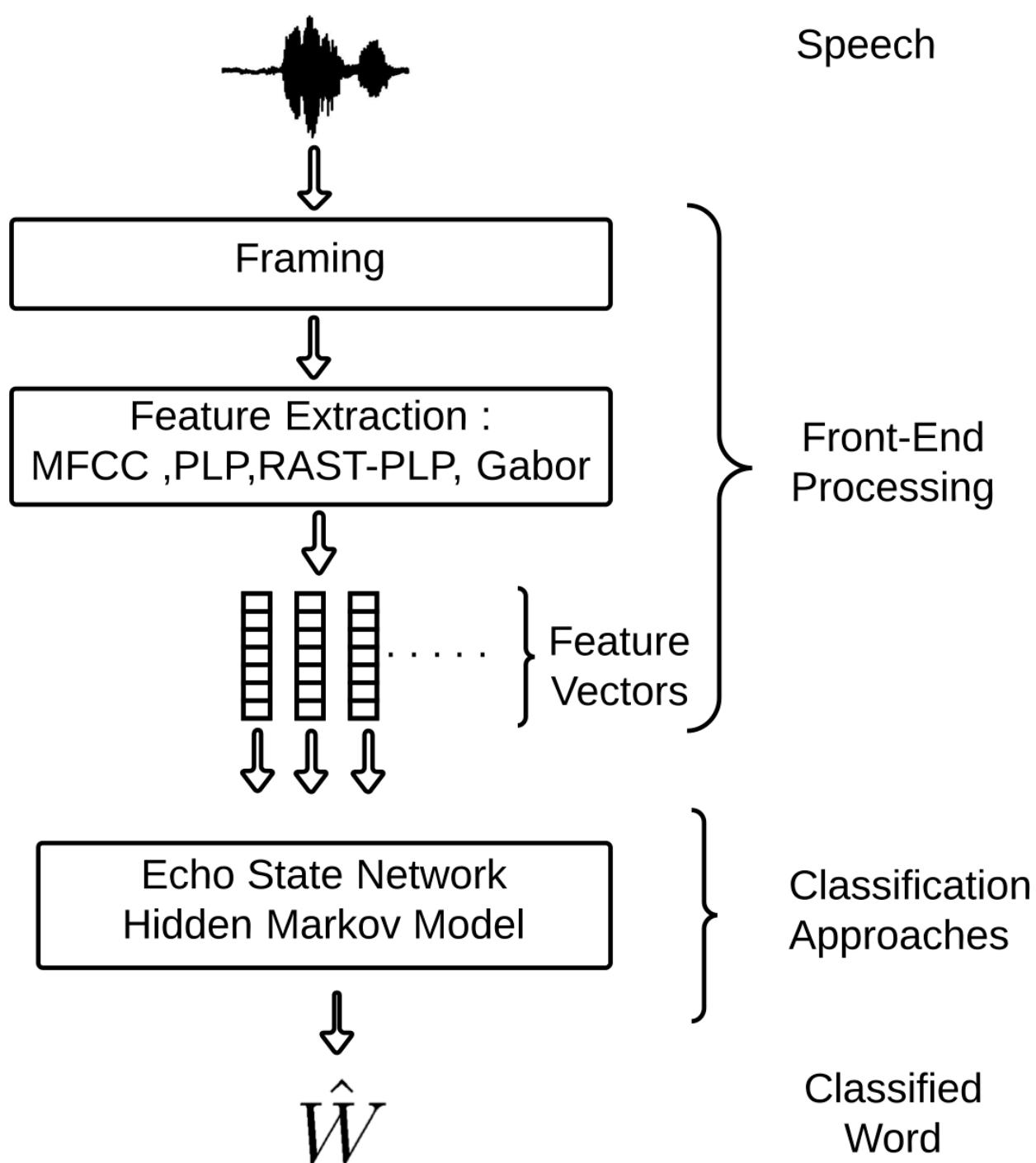
- PCA based: /dh,ix,dcl/ /t/
- ICA based: //d,ow/ /sh/

Initial tests using a recogniser

Dataset: arabic isolated words, + noise (white, babble)

(poster: 1pSC6. A noise robust Arabic speech recognition system based on the echo state network, A. Alalshakmubabarak, L Smith)

ASA 2014



Initial results: Gabors AN, plus Onsets, plus a single Gabor filter

Dataset is Arabic single words, Babble noise added.

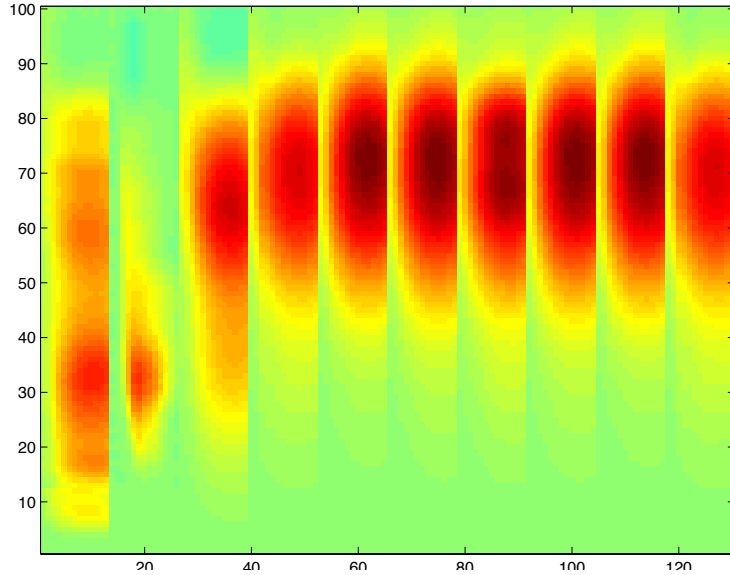
	SNR	Preprocessing	HMM	ESN (10 tests)
Babble Noise	30 db	MFCCs	95.85 %	97.23 %(0.29)
		PLP	97.05 %	97.87 %(0.36)
		RASTA-PLP	98.65 %	99.22 %(0.19)
		Gabor	————	99.46% (0.14)
	20 db	MFCCs	78.49 %	89.72 %(0.87)
		PLP	86.64 %	89.47 % (2.43)
		RASTA-PLP	96.75 %	97.18 % (0.42)
		Gabor	————	97.27 % (0.41)
	10 db	MFCCs	31.77 %	64.12 % (2.31)
		PLP	54.23 %	56.23 % (4.82)
		RASTA-PLP	85.14 %	85.45 % (8.6)
		Gabor	————	66.56 %(1.15)

Conclusions, Further work

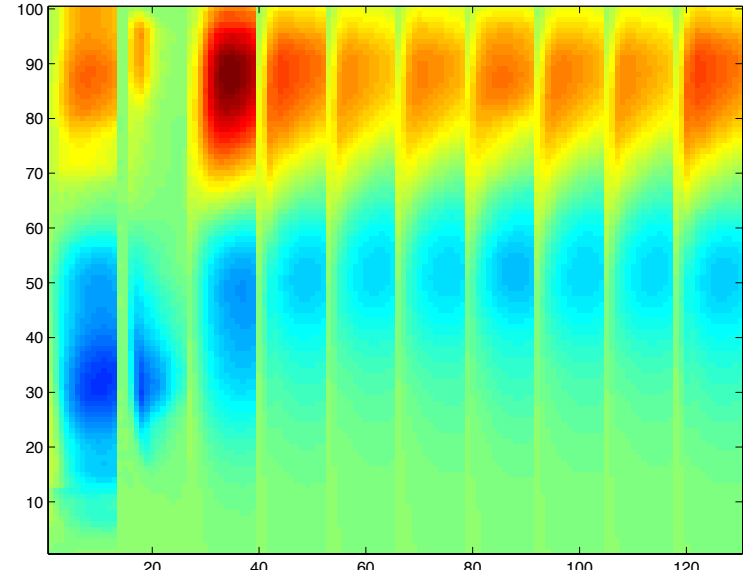
- Initial results suggest that the Gabor spectrotemporal feature is useful:
 - But this needs proven
- A much more sophisticated recogniser needs to be used to find out if this really does work
 - And we need to extend this to continuous speech
- We want to consider other related problem from auditory scene analysis
 - For example, interpreting other, non-speech sounds.

PCAs 8 Gabors)

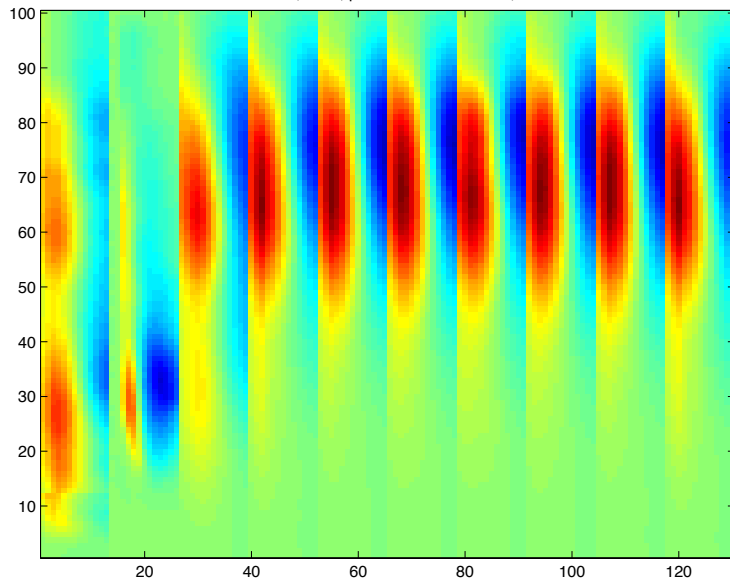
Info = Gabor 8, male, pca1 CLim=-0.023315,0.023315



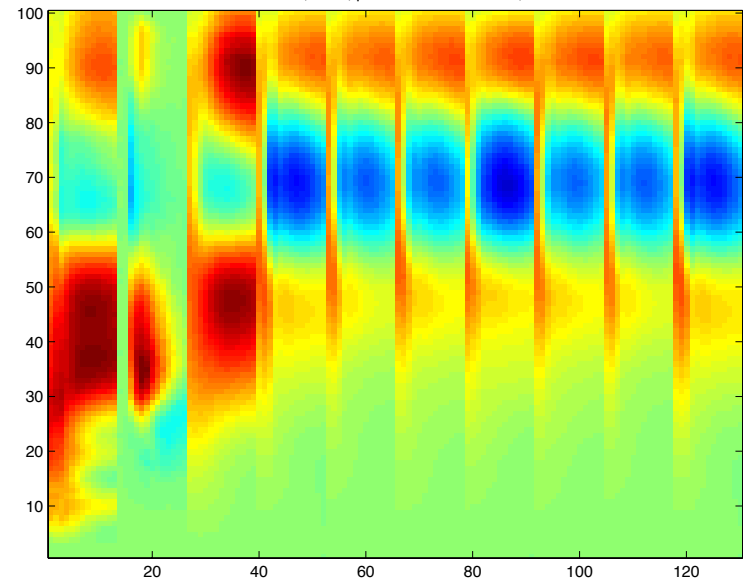
Info = Gabor 8, male, pca2 CLim=-0.033635,0.033635



Info = Gabor 8, male, pca3 CLim=-0.026632,0.026632

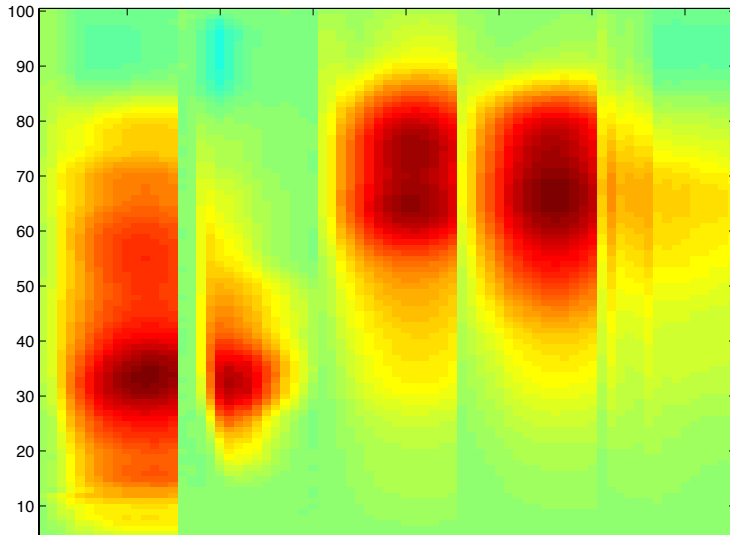


Info = Gabor 8, male, pca4 CLim=-0.025847,0.025847

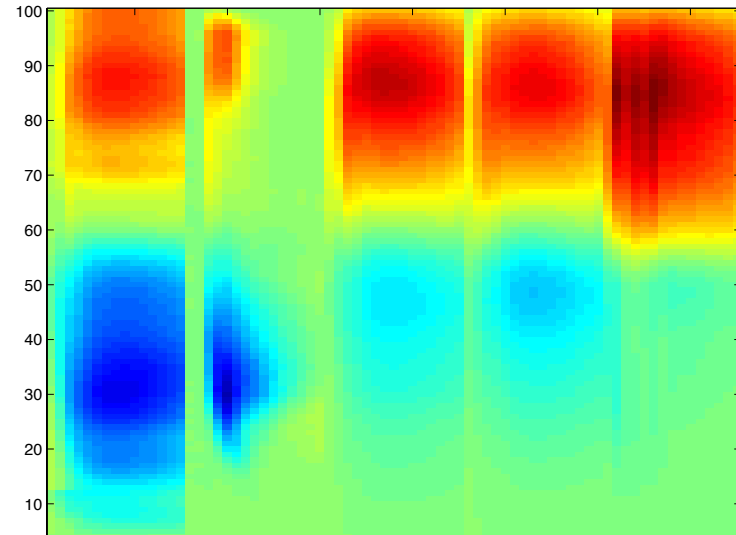


PCAs (3 Gabors)

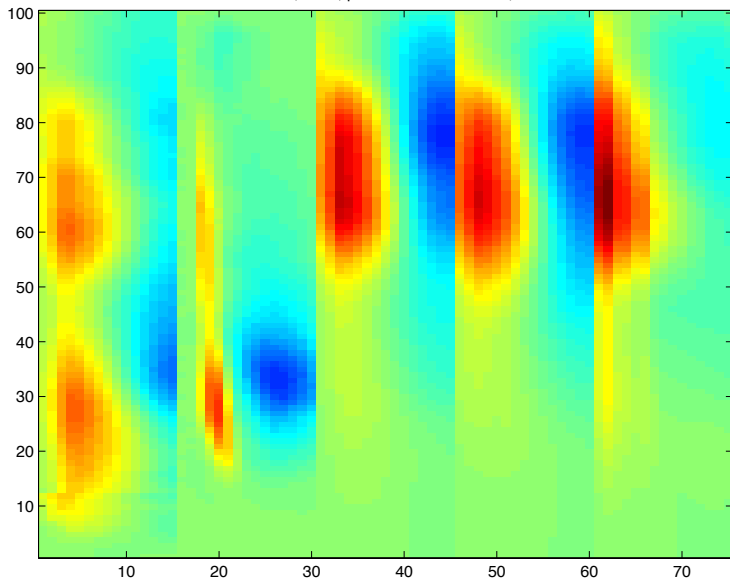
Info = Gabor 3, male, pca1 CLim=-0.032573,0.032573



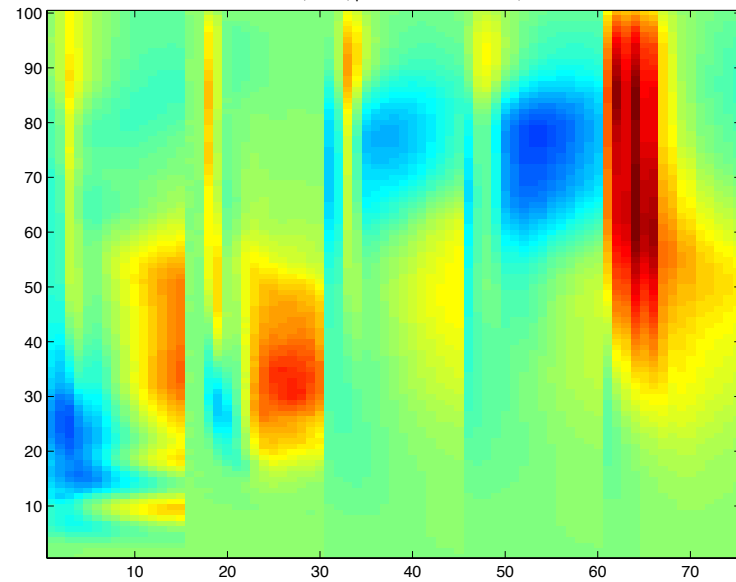
Info = Gabor 3, male, pca1 CLim=-0.031947,0.031947



Info = Gabor 3, male, pca3 CLim=-0.044844,0.044844

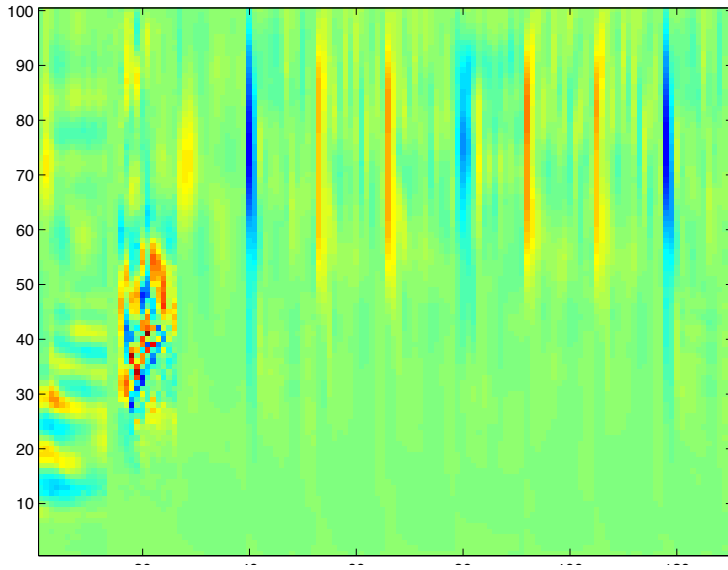


Info = Gabor 3, male, pca4 CLim=-0.045874,0.045874

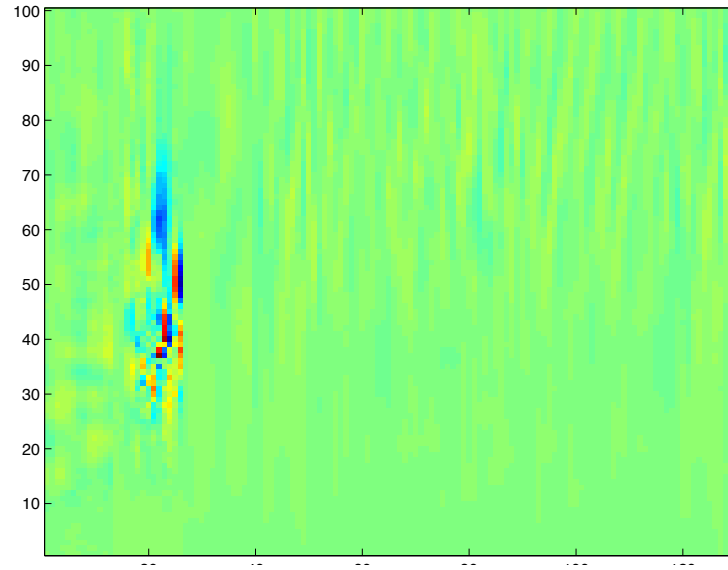


ICA 8 Gabors

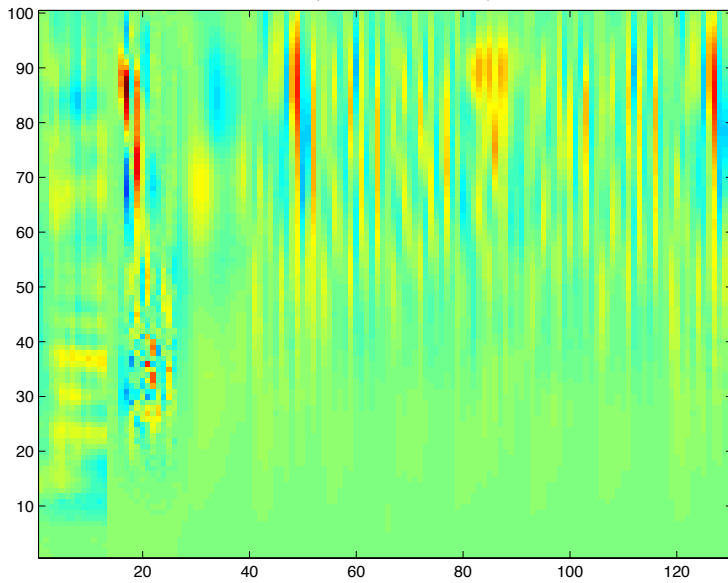
Info = Gabor 8, ica 1 CLim=-0.029857,0.029857



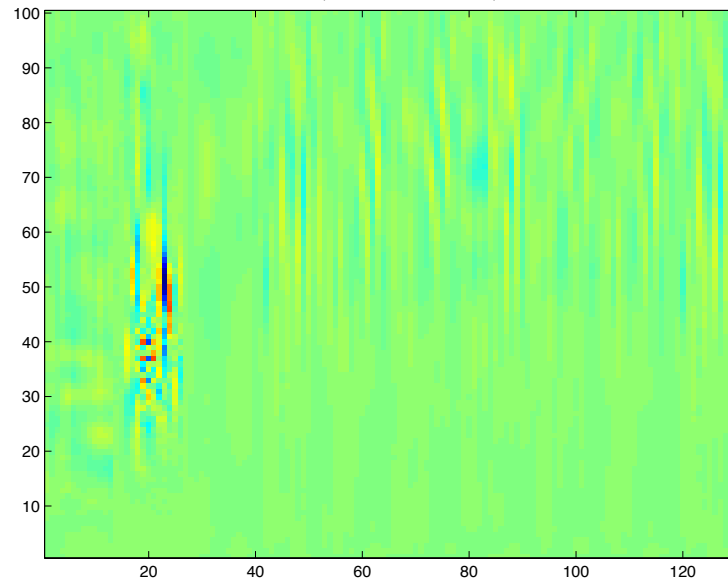
Info = Gabor 8, ica 2 CLim=-0.075694,0.075694



Info = Gabor 8, ica 3 CLim=-0.029488,0.029488



Info = Gabor 8, ica 4 CLim=-0.071893,0.071893



Preliminary results 1: birdsong

7 birdsongs,
8 Gabor filters
Note the
different
peaks and
relative
sizes for
different
birds

