Phoneme discrimination from MEG data

Tuomas Lukka, Bernd Schoner, and Alec Marantz
a Harvard Society of Fellows, Harvard University, 78 Mount Auburn St., Cambridge, MA 02138, USA
b Physics and Media Group, MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA
c Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

NEUROCOMPUTING : 10 March 2000.

Keywords: MEG data; Signal detection; Phoneme discrimination

Scope

(i) The study of Subject's internal discrimination between consonants, as it is reflected in the MEG data.

(ii) *"The most common approach to the analysis of stimulus evoked responses with MEG is to record 100 or more time-locked responses, average these responses, and then perform single dipole source analysis on the averaged waves. This kind of analysis is interesting from a clinical point of view, when locating a particular function in the brain is important. However, while averaging serves to reduce noise and to remove ` background a activity unrelated to the stimulus, dipole modeling loses the statistics of the averaging and proves a data-wasteful method of reducing the dimensionality of MEG data"*

To build a classifier system to discriminate between different stimuli from the unaveraged data

Outline

MEG data in a signal detection framework :
discrimination between different phonemes heard by a test subject.

Data: responses evoked by the voiced syllables /bæ/ and /dæ/
and the corresponding voiceless syllables /pæ/ and /tæ/.

***Methods*:**

***Dimensionality Reduction* :**

➲ principal component analysis (PCA) / ICA

Feature Selection*:*

➲ (i) matched filtering or (ii) wavelet packet decomposition

Classification:

➲ use a mixture-of-experts model to classify different stimuli

Signal Detection:

➲ contrast "stimulus event" to "zero event"

Results:

Voiced/voiceless consonant discrimination : e.g. /bæ/ - /pæ/ ☑

estimate the onset time of a stimulus from a continuous data stream ☑

Discrimination between voiced (voiceless) consonants e.g. /bæ/ - /dæ/ ☒

DATA DESCRIPTION

<u>The stimuli</u> : four 300 ms syllables, /bæ/, /pæ/, /dæ/, and /tæ/.
voiced-voiceless pairs /bæ/-/pæ/ and /dæ/-/tæ/ , of acoustic difference only in the "voicing onset time": 20 vs 80 ms of aspiration, a prior to the onset of the (voiced) vocalic portion of the syllable

<u>The MEG system</u>: 37-channel - with "1st-order gradiometer" sensors.
Sensor array was centered over the left auditory cortex

<u>The recording parameters:</u> 4 stimuli were presented to the right ear 100 times each, in pseudo-random order at a variable ISI of 1-1.5 s.
400 epochs of 600 ms were recorded, time-locked to stimulus onset, with a 100 ms pre-stimulus interval**. SR** was 1041.7 Hz with a **BW** of 400 Hz.

SCHEMATIC OUTLINE

Input Data :
Single Epoch mode
37 channels x 600 msec

Dimensionality Reduction:
PCA , ICA
3 "channels"

Feature Selection / extraction:
Matched filtering
Wavelets

Classification:
Gaussian – weighted local experts
trained by Expectation Maximization algorithm

PCA : average defined vs single epoch defined covariance

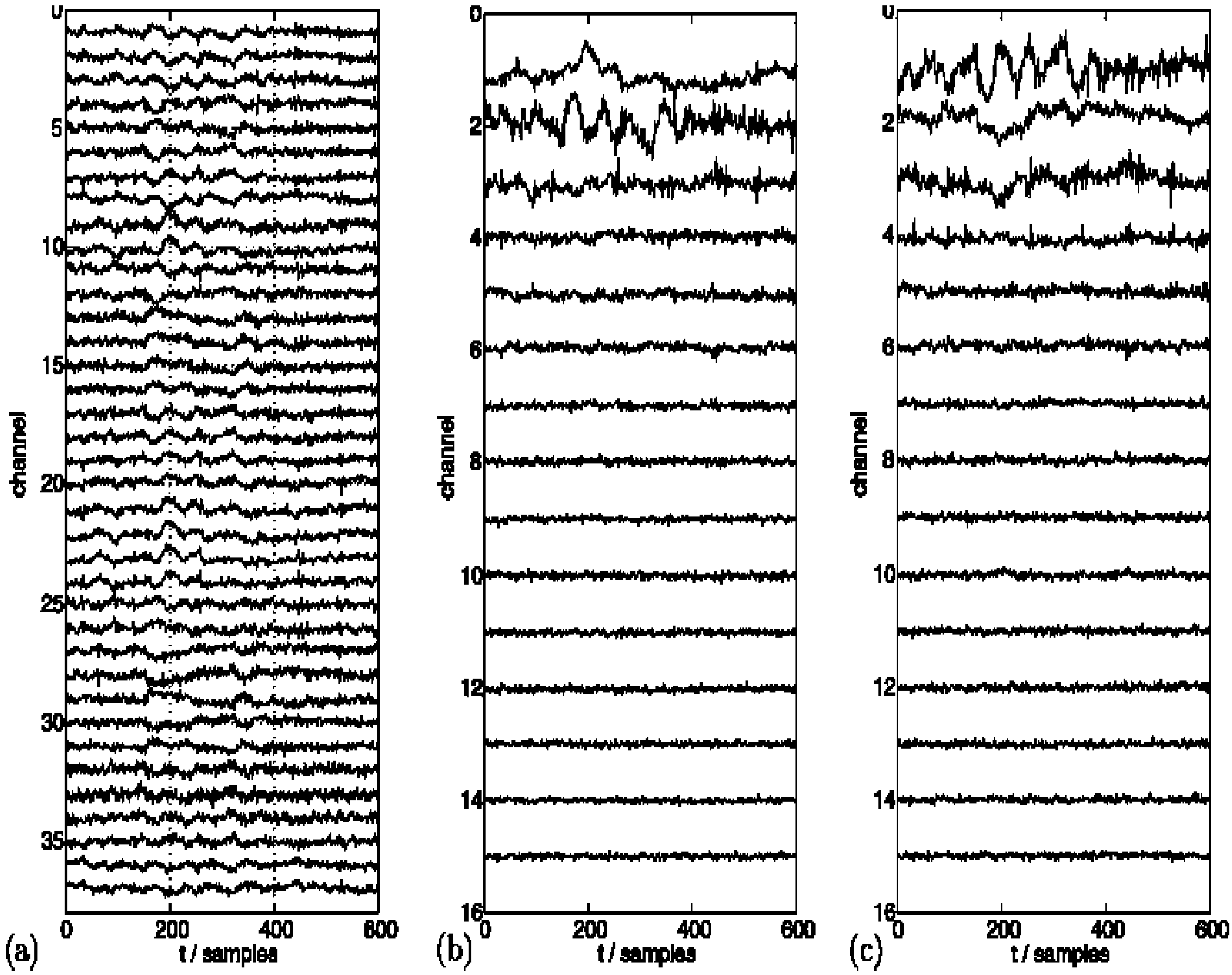


Fig. 1. (a) All channels of one raw epoch; (b) average (c) SE defined PCA

➲ “single epochs defined PCA splits the response between channels 2 and 3 whereas the average defined PCA reduces the amount of noise by concentrating the response in the "1st channels, and so seems preferable”.

ICA : some events comes out clearer than using PCA

➲ “However, ICA can also increase the effect of noise and make classification of signals difficult”.

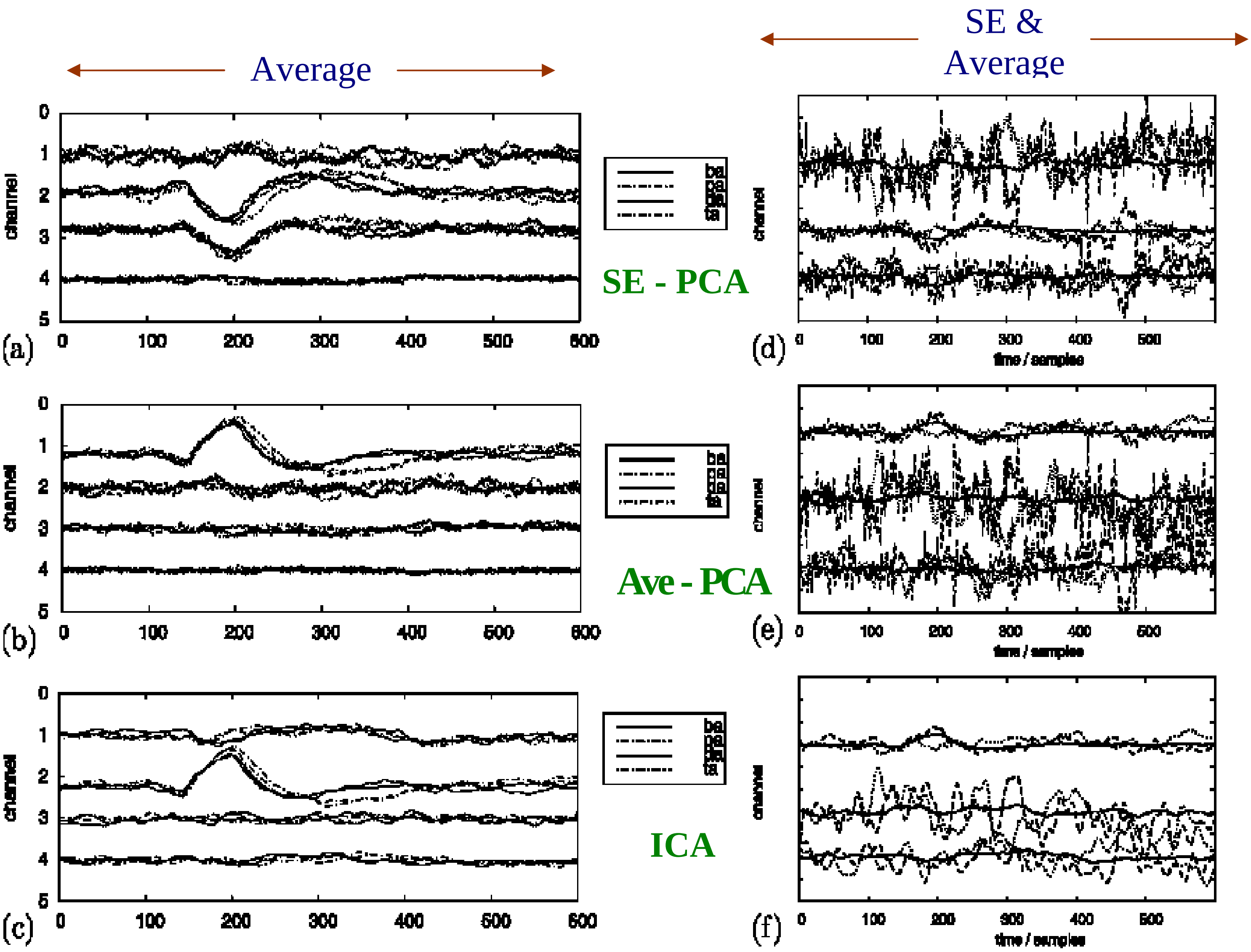
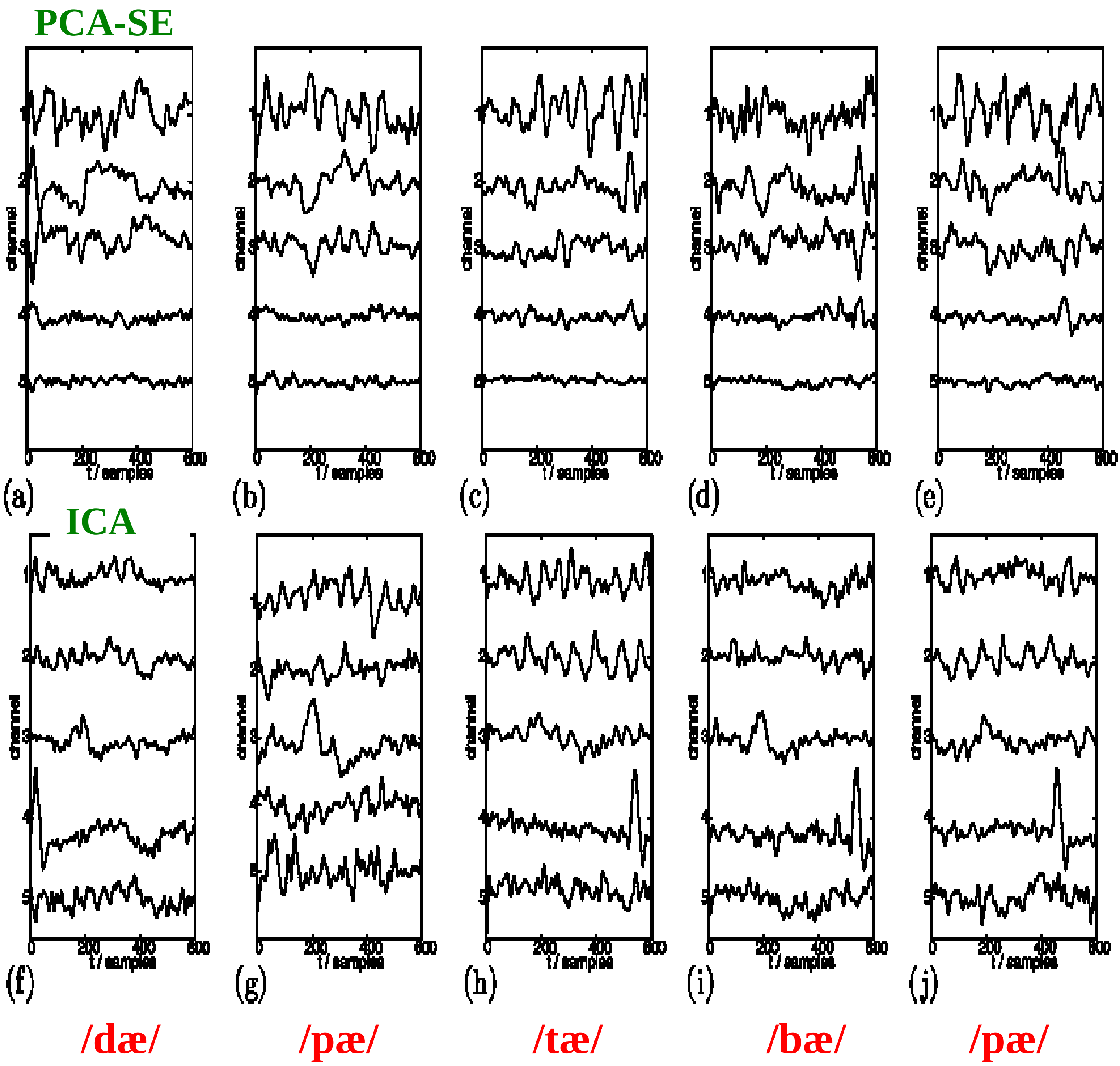


Fig. 2. Average responses to the four different stimuli after
(a) single-epoch-defined PCA, (b) average-response-defined PCA
(c) ICA transform (lp at 60 Hz).
A single epoch and the average superimposed,
(d) single-epoch-defined PCA, (e) average-response-defined PCA and
(f) ICA transformed data.

Recorded MEG epochs stimulated by /dæ/, /pæ/, /tæ/, /bæ/ and /pæ/.
(a)-(e) PCA (Single Epoch defined) transformed responses.
(f)-(j) Same epochs ICA transformed



Some events come out clearly, such as the heart beat in channel 4 and the stimulus response in channel 3; however, the whitening required by the algorithm has increased the noise levels.

➲ The "channel 3" is proposed for signal detection purposes

Matched filtering

noise-free response are constructed by averaging over the training epochs and correlation of incoming signal with these “true responses” is used for

(i) response onset time detection (maximum peak detection)

(ii) discrimination between different stimuli (best template match)

➲ The use of only the 1st PC is suggested

Wavelet packets : Daubechies

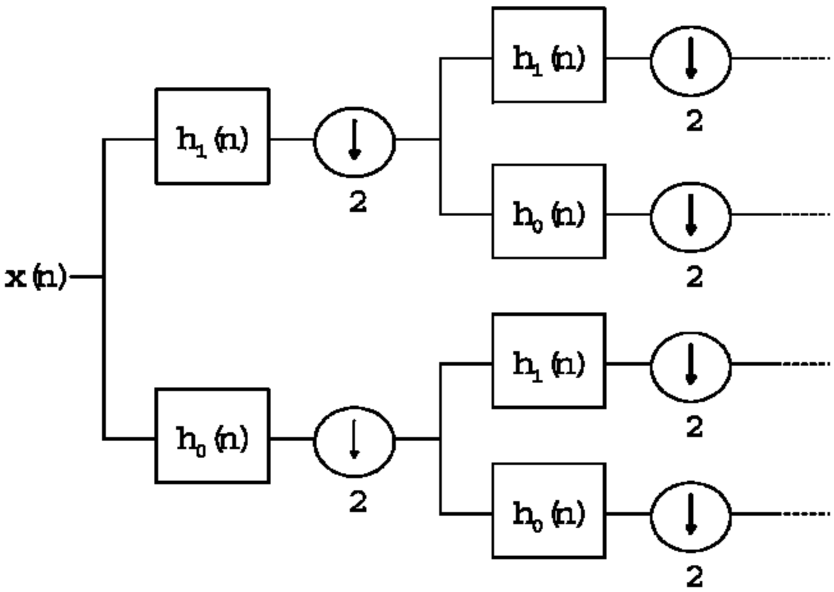


Fig. 4. Filtering steps in the discrete wavelet packet transform. h0(n) and h1(n) are the half-band low-pass and high-pass filters, 2↓ stands for down-sampling by a factor 2.

*** selection of a reasonable number of coefficients to form the feature vector:**

(i) a subset is chosen to maximize the square distance discrimination measure :

$$D_{SD} = (\overline{w}_{i1} - \overline{w}_{i2})^2 \big/ (s_{W_{i1}} s_{W_{i2}})$$

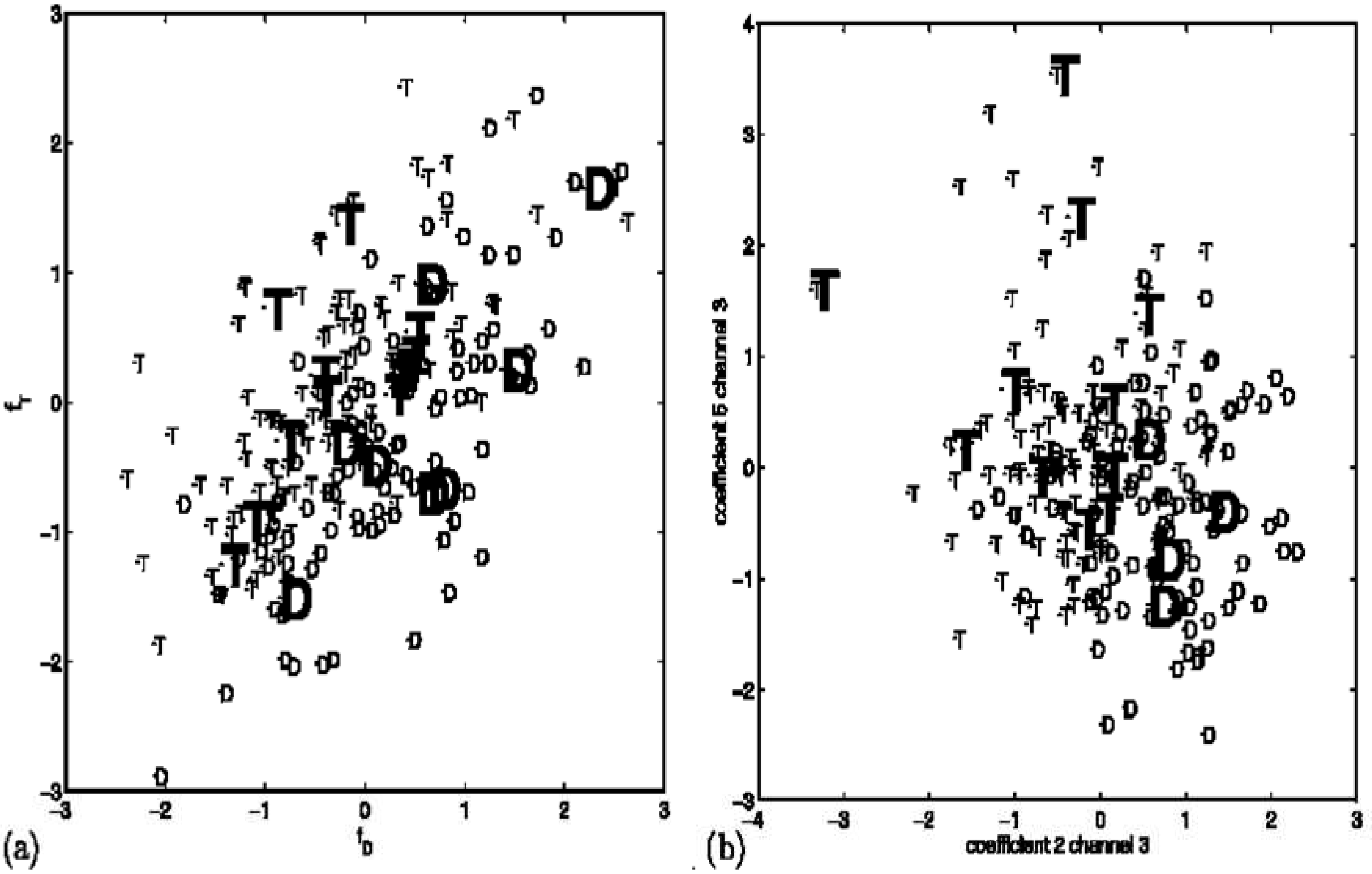
(ii) or the symmetrized relative entropy (Kullback-Leibler distance)

between either two stimuli (for phoneme discrimination)
or a stimulus and a non-stimulus window (for phoneme onset detection)

➲ Selection was based on the use of all the “significant” PCA channels

Cluster-weighted classification

- use of Gaussian-weighted local experts in a Cluster-Weighted Modeling framework to discriminate between stimulus classes.
- each local expert represents one (of the many) distribution over one class
- the influence of each local expert is a multivariate Gaussian
- the model is trained by Expectation Maximization algorithm



Two dimensions of the feature vector for the bæ/dæ discrimination:
(a) A/MF, (b) A/WP. The small letters refer to the actual sample points; the large letters are the centers of the local experts. The letter T refers to the voiceless and D to the voiced version of the consonant.

Classification Results

- epochs of each stimulus were randomly divided into a training set of 70 and a testing set of 30 epochs / Windows of 256 samples long
- no significant differences between MF and the WP methods
- Two wavelet coefficient used for the WP feature vector *(the type of wavelet has no significant effect)*
- ***ICA*** didn't offer any improvement

Table 1
Results for discriminating voiced/voiceless syllables. The last four columns are the detection results, the numbers before/after the slash are the number of correct/incorrect classifications. C_1 and C_2 refer to the first and second stimuli used in that test run, e.g. on the first row bæ and pæ, respectively

Syllables	PCA	Detector	N_c[a]	Window offset (samples)	Classification			
					Training		Testing	
					C_1	C_2	C_1	C_2
bæ/pæ	A[b]	WP[c]	10	105	52/18	62/8	25/5	21/9
bæ/pæ	S[d]	WP	4	105	50/20	53/17	25/5	21/9
bæ/pæ	A	KL[e]	—	205	59/11	63/7	25/5	18/12
bæ/pæ	A	MF[f]	15	205	52/18	56/14	19/11	25/5
dæ/tæ	A	WP	4	205	45/25	51/19	19/11	20/10
dæ/tæ	A	WP	2	105	50/20	49/21	21/9	22/8
dæ/tæ	A	MF	15	205	57/13	65/5	21/9	25/5

[a] Number of clusters (local experts).
[b] Average-defined PCA.
[c] Wavelet packet coefficient and cluster-weighted detection.
[d] Single-epoch-defined PCA.
[e] Kullback–Leibler distance discrimination.
[f] Matched filtering discrimination and cluster-weighted detection.

- It is possible to get a statistically significant detection accuracy for voiced/voiceless discrimination.
- Discrimination between two voiced (or two voiceless) consonants was impossible

Signal Detection Results

- The average response can be used to detect the presence/onset of a stimulus in a continuous data stream:

(i) peak of signal convolution with the "reference" epoch

(ii) wavelet expansion best discriminating "stimulus" and "zero" event

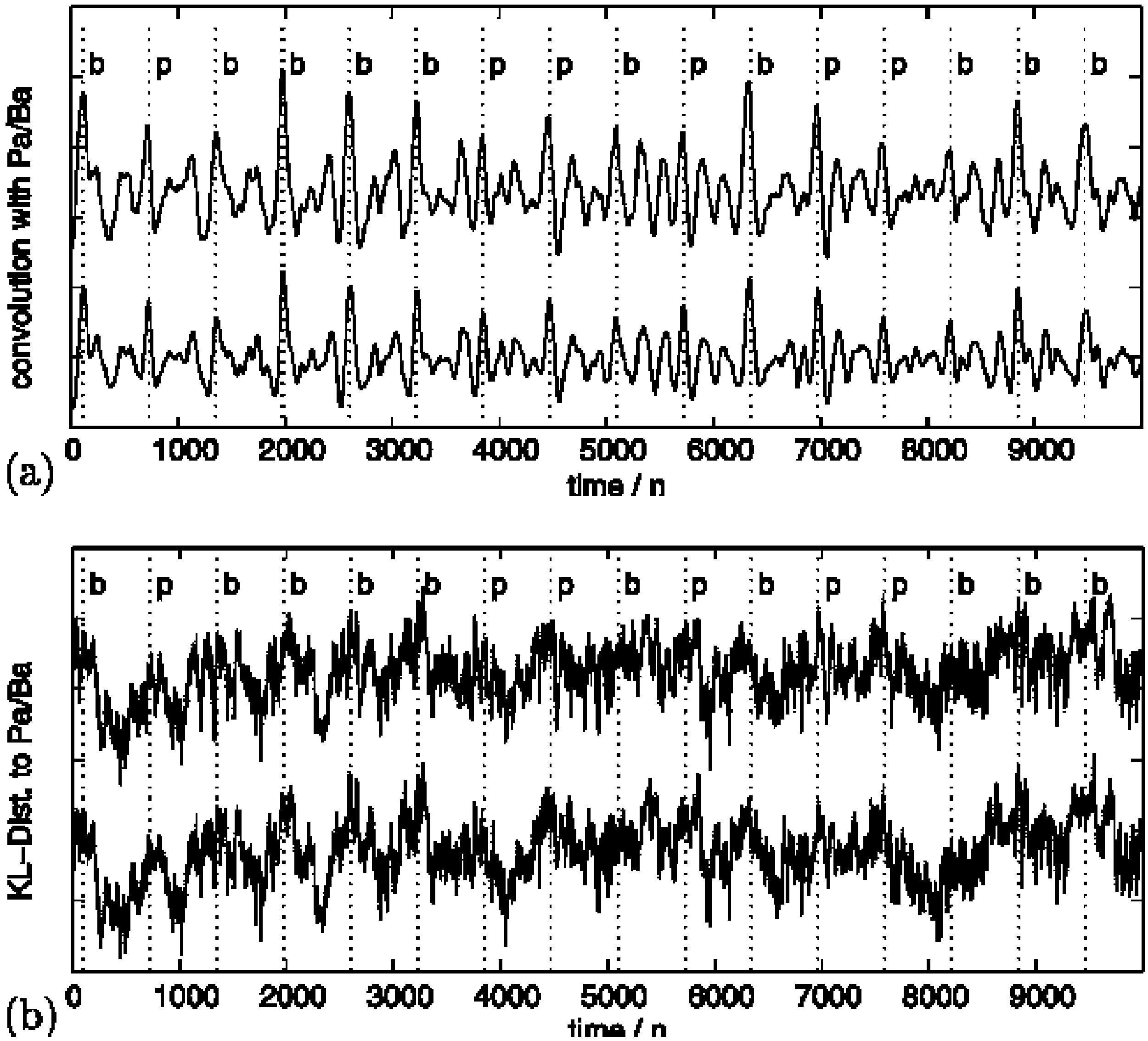


Fig. 7. Two example signals from the onset detection.
(a) Matched filtering; (b) Kullback-Leibler distance.

Conclusions and future work

- “Since MEG provides an extremely rich source of data on brain function, it is important for cognitive neuroscience to develop analysis techniques for extracting signal from noise and for identifying crucial features of evoked responses”.

- “One future possibility would be to develop an event-based maximum likelihood model for interpreting the data.

Discussion

➊ the effect of pure “auditory” response / content based difference in brain’s responses

➋ (prestimulus) state of brain

➌ superficial approach

➍ oversimplified approach for such a “cognitive” task