morris,bourlard@idiap.ch
http://www.idiap.ch/

Missing-data methods with cepstral data

❈

Summary of work done at IDIAP

❈

Main achievements

●

DUMA - Data Utility Maps from MLPs

- Can't use noise estimation to generate data utility maps for multi-condition models - noisy data may be "clean".
- Can't assume that mismatching data are outliers. (see Fig.)

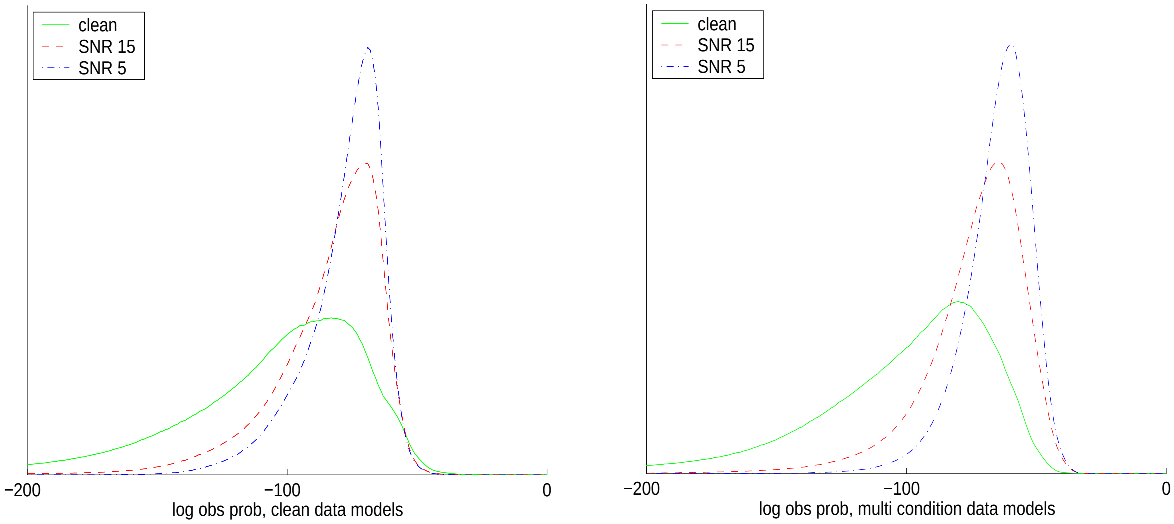
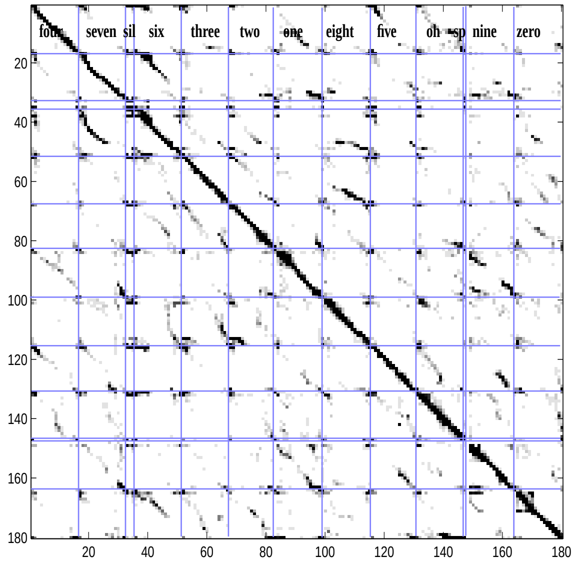


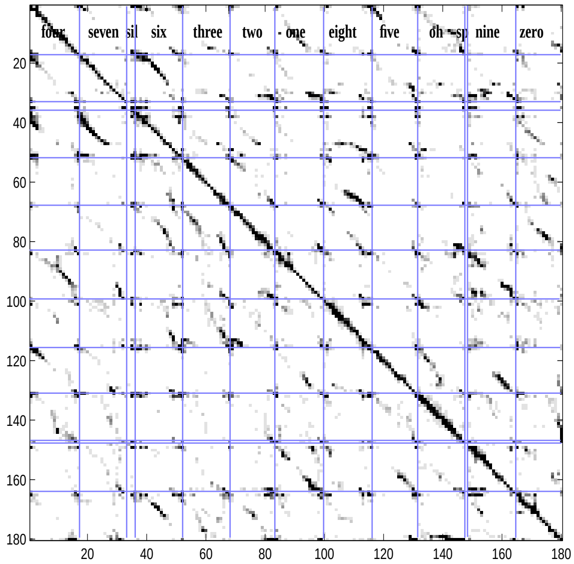
Fig shows log prob histograms for clean, SNR 15 and SNR 5 dB N1 (subway) MFCC_E_D_A data, left for clean models, right for multicondition models. Probabilities increase with noise level, rather than decrease as might be expected.

- Could the entropy of an MLP classifier trained on a small window about a data point tell us something about its utility?

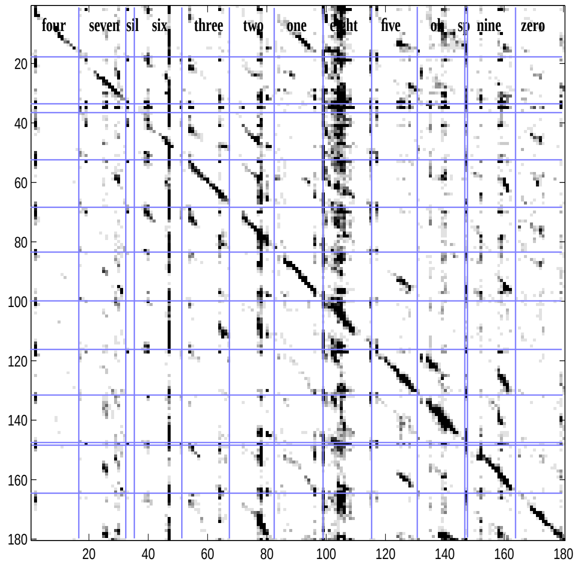
MLP state confusion characteristics



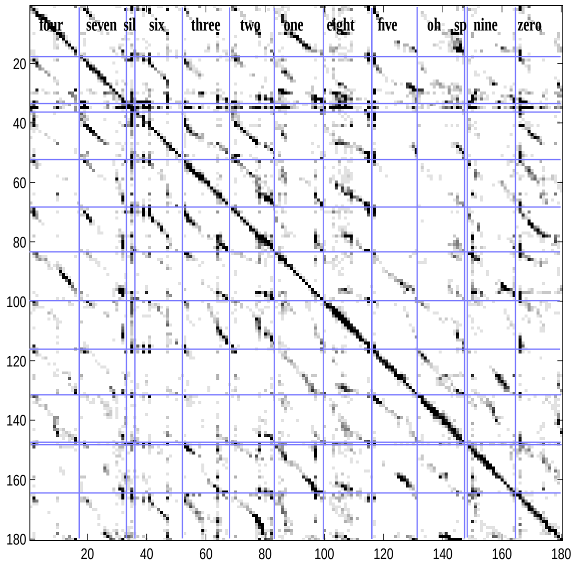
data clean, models clean
frame error rate 62.6%



data clean, models
multicondition



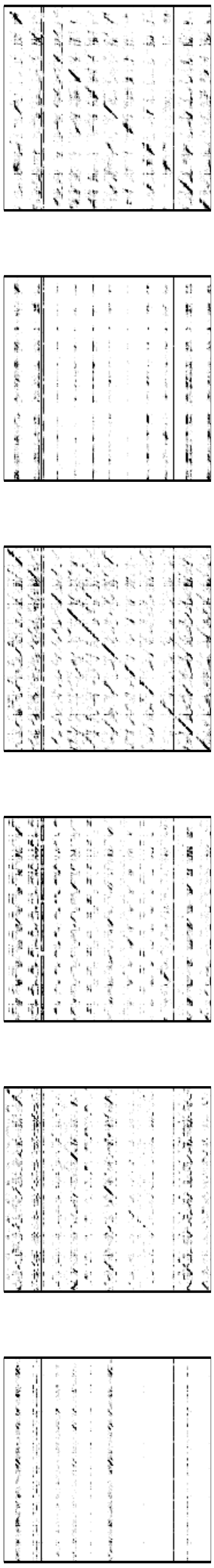
data SNR 5, models clean
frame error rate 88.7%

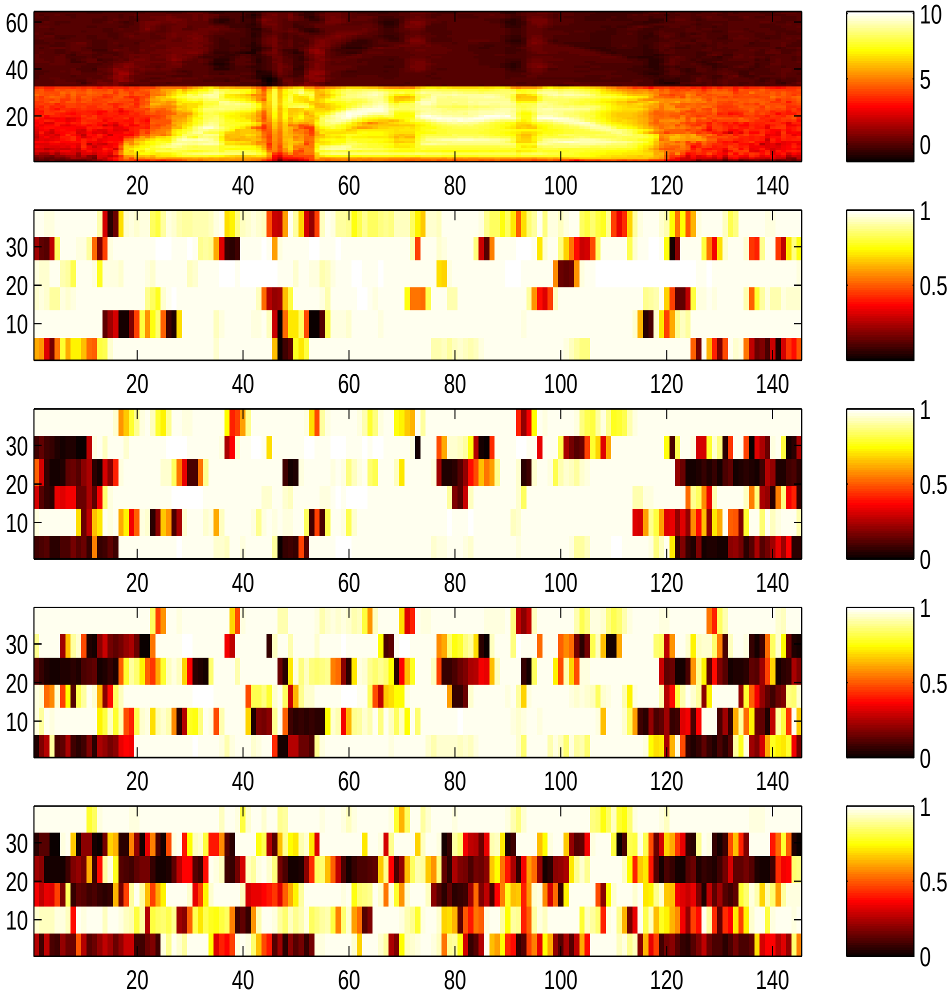


data SNR 5, models
multicondition

For clean models, “eight” is attractor. For multicond. models ‘sil’ and ‘sp’ act as attractor noise models.

Local MLP state confusion characteristics





Top = FBANK_D coeffs. Down from top are DU masks for clean, SNR 20, 10 & 0 dB. Utterance is MAH_139OA. Masks based on confidence-matrix corrected MLP output entropies. Max and median observed corrected entropy values mapped to 0 and 1.

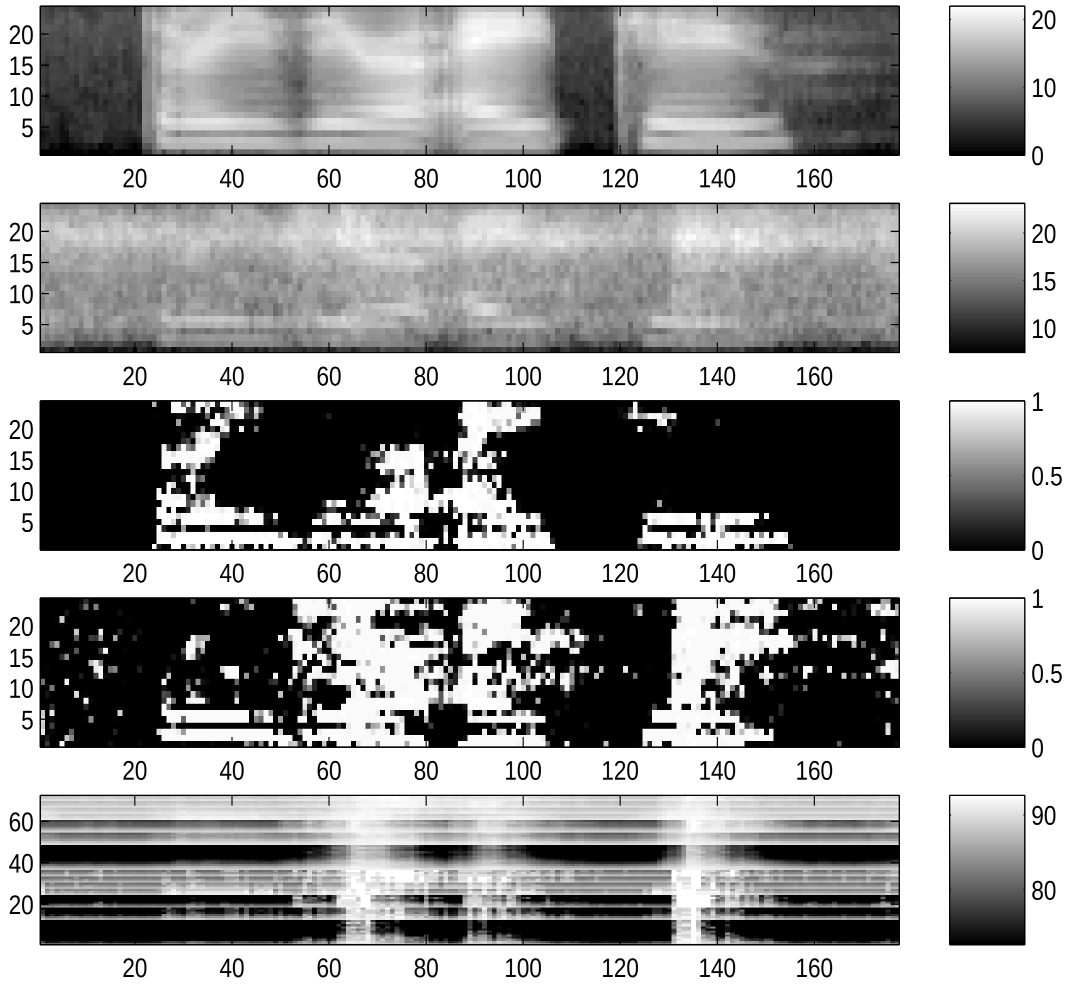
Left are conf. mats for the 6 subband MLPs.

Missing-data methods with cepstral data

When log spectral data have evidence pdf

$$u(x_i) = \varphi_i \delta(x_i - x_i^{obs}) + (1 - \varphi_i) unif(0, x_i^{obs})$$

the evidence pdf for any linear function of this data can be obtained, and has the same form.



Top=clean fbank (power), 2=SNR0 fbank, 3=oracle MD mask, 4=simple MD mask, 5= **multi-cepstral intervals** for hard MD mask. Signal = FAK_3Z82A, noise = N1 (subway).

CDPP - Clean Data PDF Propagation

Intervals of uncertainty for cepstral coeffs are extremely wide everywhere except where almost whole spectral frame clean.

Can reduce problem by appending subband cepstral features.

Recognition still bad unless intervals somehow scaled down.

Can obtain much tighter cepstral pdf by deriving clean speech log spec. energy pdf directly from noise spectral energy pdf.

In this case the "oracle" would 100% restore clean data.

General formula for pdf of function y=f(x) of rand variable x. If g(x) = f^{-1}(x) is monotonic, then $p_y(y) = p_x(g(y))|g'(y)|$

For noise energy pdf $p_n(x)$ the resulting evidence pdf for clean log speech energy is

$$u(x) = e^{x_{clean}} p_n\left(e^{x_{obs}} - e^{x_{clean}}\right)$$

This has strong squashing effect on noise pdf.

- Results so far do not better clean cepstral baseline - except for 0.5% on clean speech (though at 98.83% acc, this is still a 60.7% decrease in WER, or 66.0% decrease in WIL).
- May (or may not!) improve over "max assumption" for MD with spectral data.

Summary of work done at IDIAP

1999

FCMB Full Combination Multi-Band
IDCN Incomplete Data Classifier Network

2000

MLCW ML (etc.) Combination Weighting techniques
FCMS Full-Combination Multi-Stream

2001

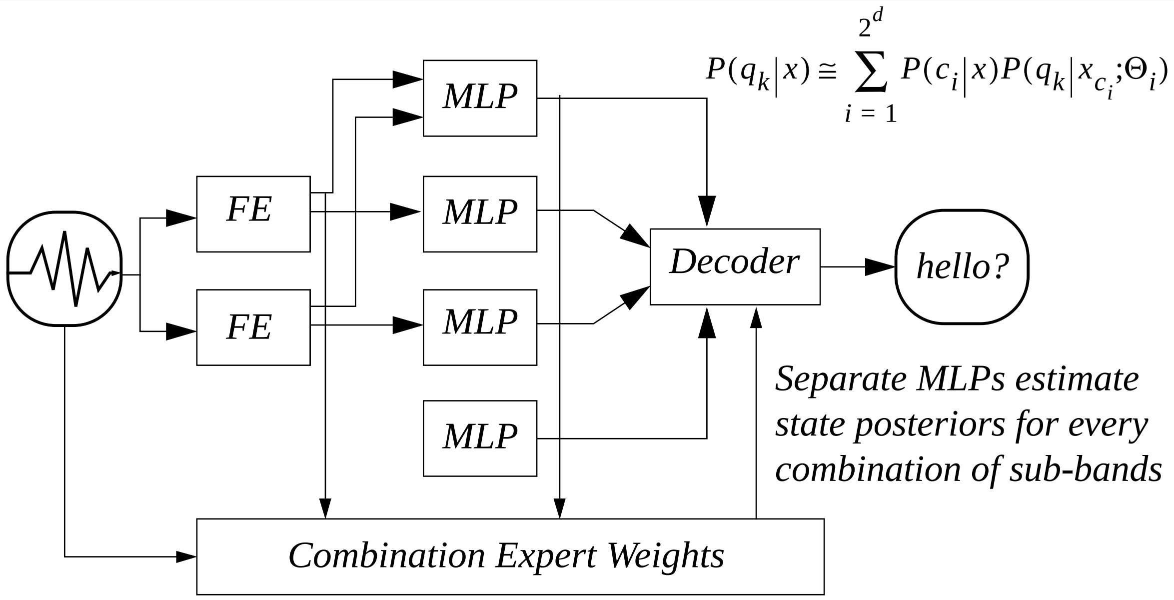
MPFC MAP Full Combination
TRUD Theory for Recognition with Uncertain Data

2002

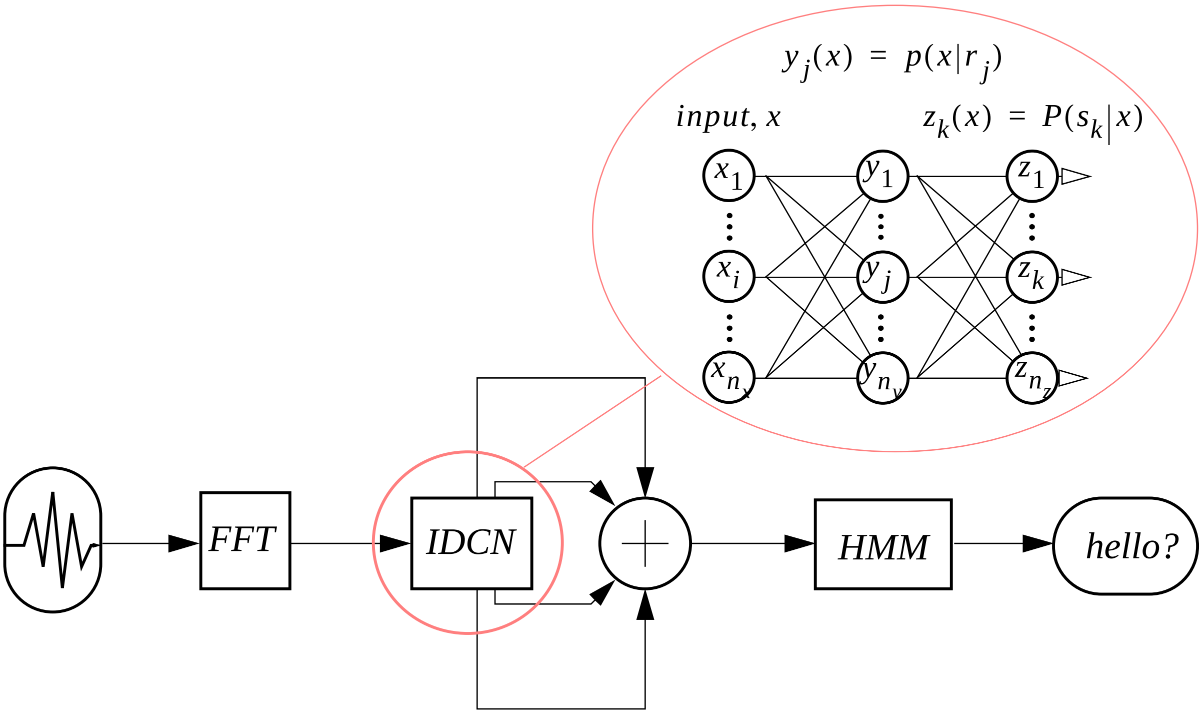
MSTK MultiStream ToolKit
DUMA Data Utility MAps
CDPP Clean data PDF Propagation

1999

FCMB Full Combination MultiBand

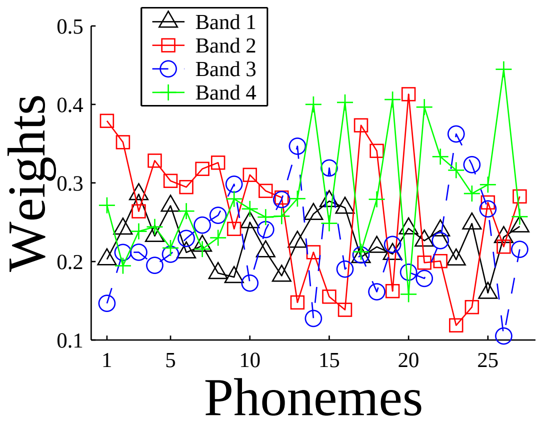


IDCN Incomplete Data Classifier Network

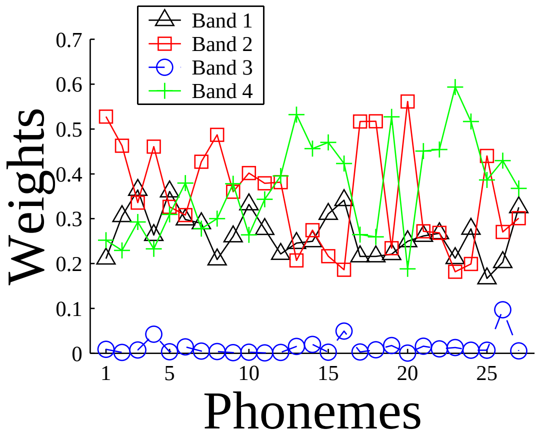


2000

MLCW ML (etc.) Combination Weighting techniques



Clean speech

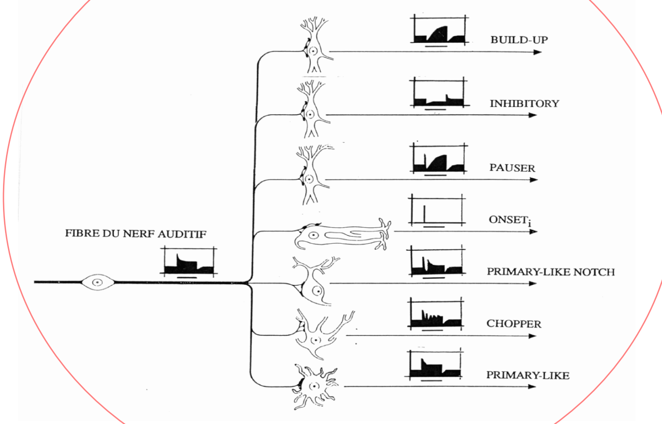


Noise in band 3

FCMS Full-Combination Multi-Stream

Combine multiple complementary sources of speech information

- short term spectrum (10 ms)
- difference features (50 ms)
- amplitude modulation spectrum (100-500 ms)
- visual features (mouth shape)
- different features at each scale (FFT, MFC, LPC, PLP)

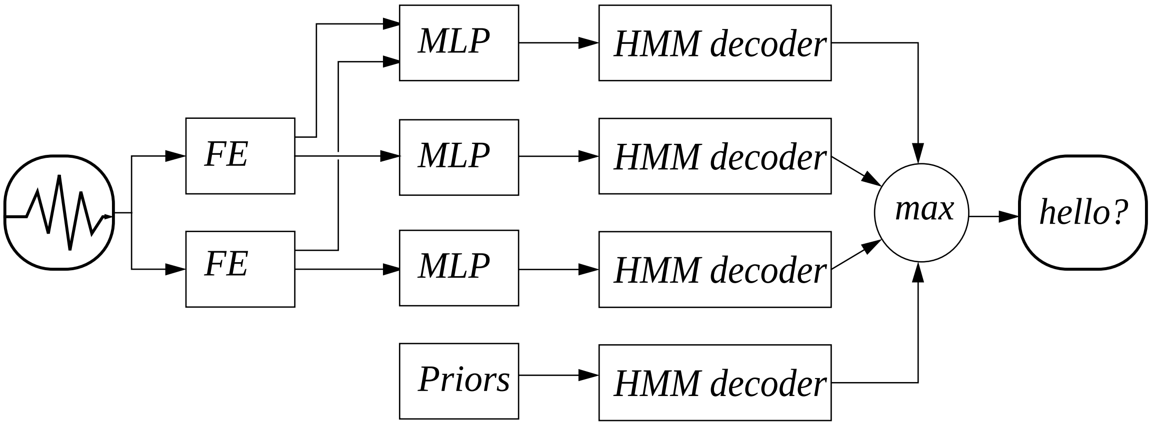


2001

MPFC MAP Full Combination

For expert weights static, MAP FC weights give weight 1 to expert with highest MAP score for each utterance.

Tests with static + diff ftrs showed strong % improvement



TRUD Theory for Recognition with Uncertain Data

$$Q_{MAP} = \mathit{argmax}_Q E[P(Q|X,\Theta)|X \sim s(X)]$$

$$= \mathrm{arg}\mathit{max}_Q P(Q|\Theta)\int p(X|Q,\Theta)u(X)(dX)$$

For soft missing-data with "max assumption"

$$u(x_i) = \varphi_i\delta(x_i - x_i^{obs}) + (1-\varphi_i)\mathit{unif}(0, x_i^{obs})$$

2002

MSTK MultiStream ToolKit

MDDM Missing-Data with Duration Models

DUMA Data Utility MAps from MLPs

CDPP Clean data PDF Propagation

Main achievements

- Not useful: IDCN, MLCW
- Maybe useful in future: TRUD, DUMA, CDPP
- Useful: FCMB / FCMS, MPFC, MSTK
- Integration of missing-data with multi-stream methods