IDIAP
Martigny - Valais - Suisse

HYBRID ASR WITH GMM BASED EXPERTS

Andrew Morris, morris@idiap.ch

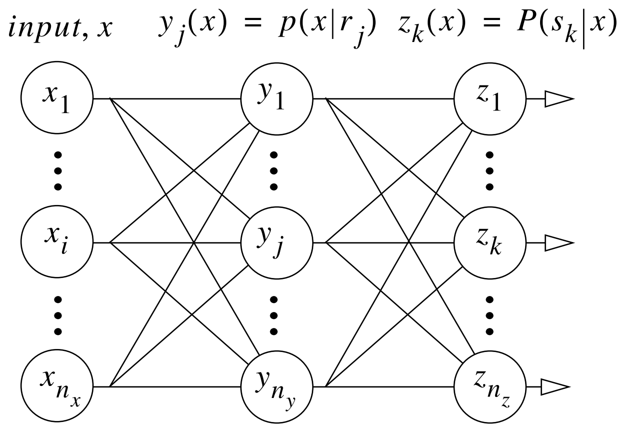
ABSTRACT

In HMM/ANN hybrid ASR there are some situations where MLP experts cannot be used because they do not provide a model of the data likelihood. GMM based experts (GME) can model both posterior probabilities and data likelihoods.

We discuss several applications of the HMM/GME hybrid in ASR: recognition with missing data (position of MD known/ unknown), and adaptive multi-stream expert weighting (ML and clean likelihood based).

Ref. "A neural network for classification with incomplete data: application to robust ASR", Andrew Morris, Ljubomir Josifovski, Hervé Bourlard, Martin Cooke, Phil Green, Proc. ICSLP 2000.

GME ARCHITECTURE



A GME is form of ANN with one input, one hidden and one output layer. Hidden unit (j) uses (diag cov) Gaussian to model probability that input vector x is from Gaussian (j).

$$y_j(x) = p(x|r_j) = N(x, \mu_j, v_j) \tag{1}$$

Output unit (k) uses Bayes' rule to transform pooled mixture likelihoods from hidden layer into a-posteriori probability that x is from class (k).

$$z_k(x) = P(s_k|x) = \frac{p(x, s_k)}{p(x)} = \frac{net_k}{p(x)} \tag{2}$$

where

$$net_k = \sum_j P(r_j, s_k) p(x|r_j) = \sum_j w_{jk} y_j \tag{3}$$

$$p(x) = \sum_{j,k} p(x, r_j, s_k) = \sum_k net_k \tag{4}$$

GME TRAINING

GME parameters are Gaussian mean and variance vectors, and output layer weights, w.

Tried training all parameters together discriminatively using EBP gradient descent procedure, as with MLPs. But found very sensitive to initialisation and local minima.

Best results by first training Gaussians as pooled mixture using BW embedded reestimation within HTK.

Subsequent training by EBP gave significant, though small (and very expensive) performance improvement.

Of three EBP objectives tested: MSE, cross-entropy & correlation., cross entropy gave slightly better results.

Used softmax function during EBP training to constrain probability weights to lie in $[0, 1]$, and sum to one.

$$w_{jk} = e^{\alpha_{jk}} / \sum_{l,m} e^{\alpha_{lm}} \tag{1}$$

Auxiliary parameters α initialised as $\alpha_{jk} = \log(w_{jk})$.

GME POSTERIORS EVALUATION WITH MISSING DATA

If data vector incomplete, optimal estimate for each class probability is expected value of classifier output. For a GME these can easily be evaluated in terms of the original function parameters, using the usual MD formulae, i.e.

CASE 1: Nothing known about missing data.

$$\hat{z}_k(x) = E[P(s_k|x)|x_p] \propto \sum_j w_{jk} y_j(x_p) \tag{1}$$

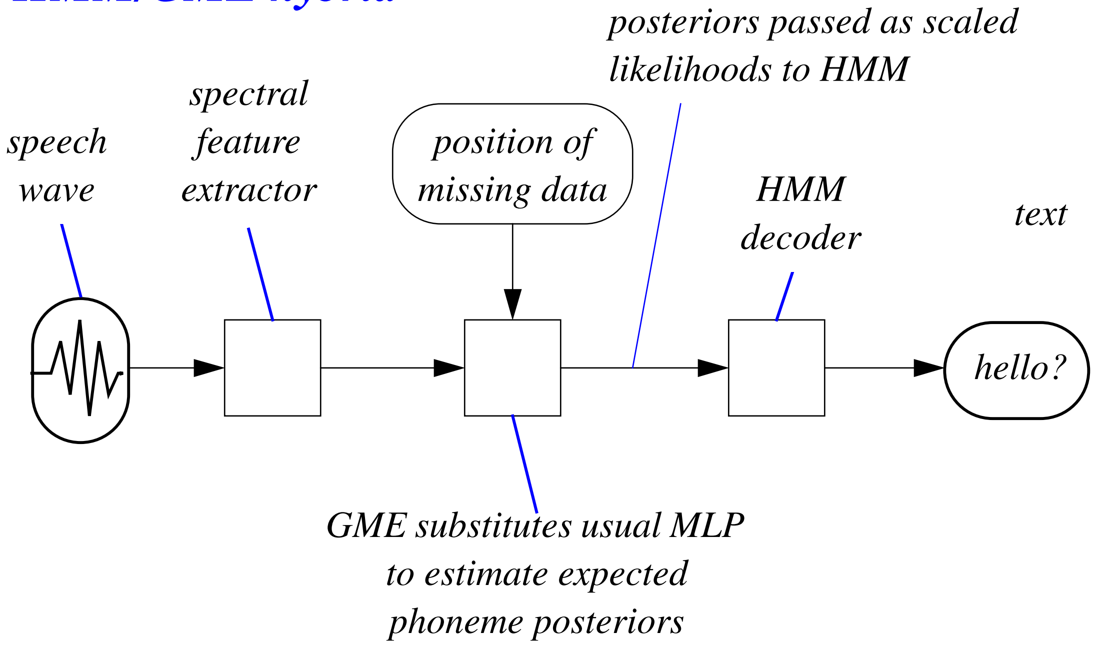
CASE 2: Missing data has limited range of possible values (as is the case for filterbank features, which are bounded below by zero and above by their observed value).

$$\hat{z}_k(x) = E[P(s_k|x)|x_p, x_m \in r_m]$$

$$\propto \sum_j w_{jk} y_j(x_p) \int_{r_m} y_j(x_m) dx_m \tag{2}$$

HMM/GME FOR ASR WITH MISSING DATA (position of MD known)

HMM/GME hybrid



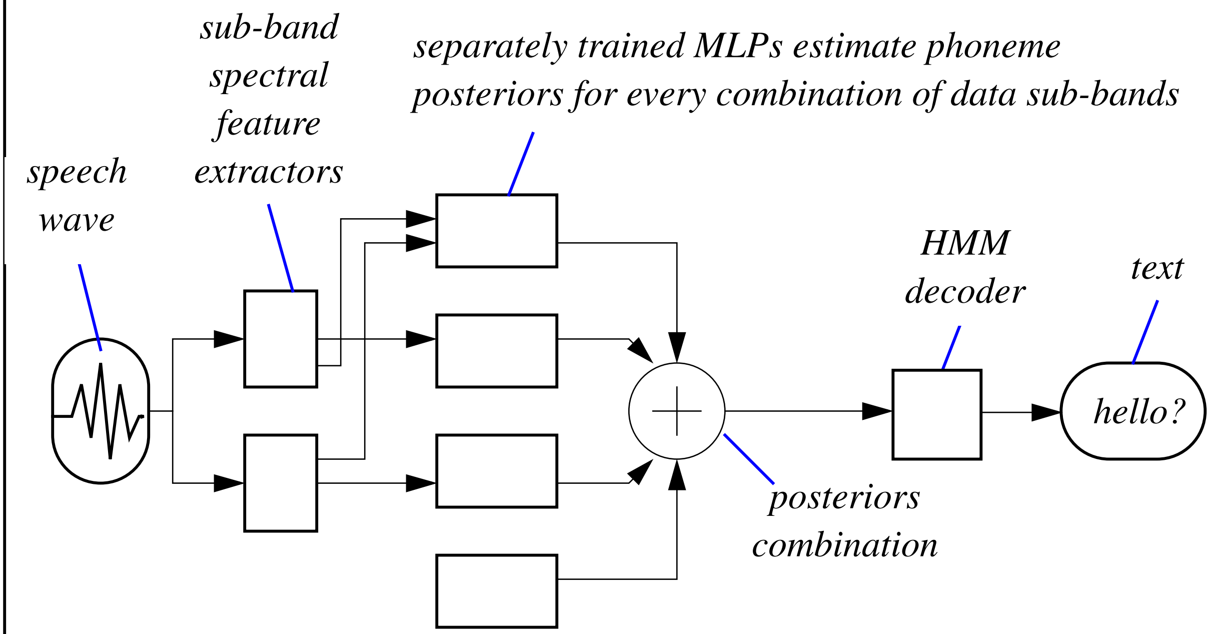
Here only theoretical advantage of HMM/GME over HMM using simple evaluation of expected likelihoods is that the GME introduces the possibility for discriminative training.

In tests with TIDigits, HMM/GME and HMM systems had almost same performance under all noise conditions tested.

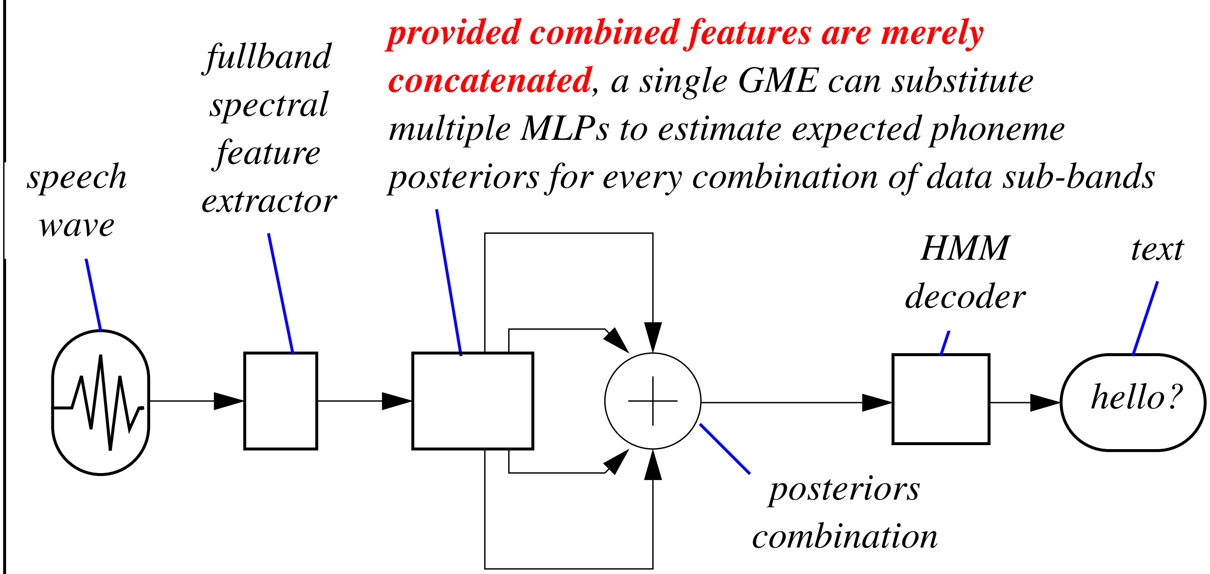
Both HMM/GME and HMM performance would likely be improved using discriminative MCE training.

HMM/GME FOR ASR WITH MISSING DATA (position of MD unknown)

Full-combination multi-band HMM/MLP hybrid



Full-combination multi-band HMM/GME hybrid



MAXIMUM LIKELIHOOD BASED MULTI-STREAM WEIGHTING

Phoneme dependent expert weights $P(b_i|q_k)$ for posteriors combination:

$$P(q_k|x) = \sum_i P(b_i|q_k)P(q_k|x_{c_i})$$ (how justified?)

can be obtained by first initialising weights all equal, then maximising data likelihood over N local data frames, using trained data likelihoods, $p(x_n|b_i, q_k)$, with EM iteration:

$$P_{new}(b_i|q_k) = \frac{1}{N}\sum_n \frac{p(x_n|b_i, q_k)P_{old}(b_i|q_k)}{\sum_j p(x_n|b_j, q_k)P_{old}(b_j|q_k)}$$

to obtain locally estimated weights, which are then used to update a travelling weights estimate:

$$P(b_i|q_k) = (\alpha)P(b_i|q_k) + (1-\alpha)P_{new}(b_i|q_k)$$

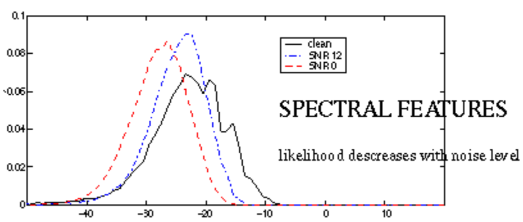
CLEAN DATA LIKELIHOOD BASED MULTI-STREAM WEIGHTING

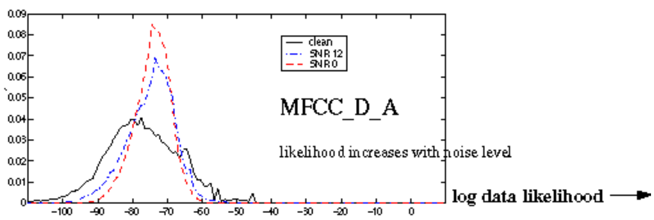
Local subbands with high data mismatch should (intuitively) have a low clean data likelihood.

On this basis, clean data likelihood models, for each subband, could provide a probability $P(clean_i|x_i)$ that observed local data x_i in each subband is clean.

From this could obtain probability that any given subband combination was clean.

However, we have found that for all commonly used acoustic features, apart from spectral, data likelihood actually increases rather than decreases with noise level.





As spectral features do not give competitive performance in clean speech, this form of reliability weighting is unlikely to be useful until we can find suitable acoustic features. Such features should also improve mismatch robustness of any kind of ML based adaptation.

GME vs MLP EXPERTS

advantages

- GME gives posteriors for every combination of subbands
- GME also gives data likelihood, which permits:
    - recognition with missing data
    - ML based adaptation for, e.g. expert weighting and speaker adaptation
    - direct measurement of data mismatch
- training unsupervised and permits use of hidden states

disadvantages

- classification performance inferior to MLP
- optimum number of Gaussians can be very high
- GD discriminant training is very slow and ineffective

CONCLUSION

GMEs are useful for classification with missing data and ML adaptation, but likelihood based experts will not be a satisfactory substitute for MLP experts until their classification performance can be made more competitive.

SPHEAR/RESPITE RELATED WORK AT IDIAP

BAYESIAN NETWORKS FOR ASR

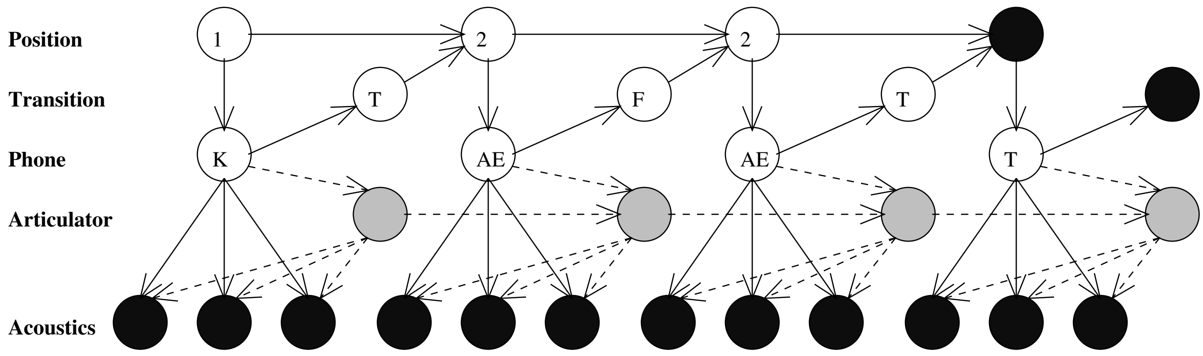


Fig. shows discrete DBN for /cat/. Black (acoustic) nodes always observed. Grey (artic.) nodes observed only in training. Normal HMM obtained by removing all grey nodes.

Initial IWR tests with discrete DBN combining articulatory with acoustic features improved WER from 9.8% to 7.7%. But discrete DBN suffers acute problems with computational complexity. Now extending DBN system to permit mixing discrete with continuous variables.

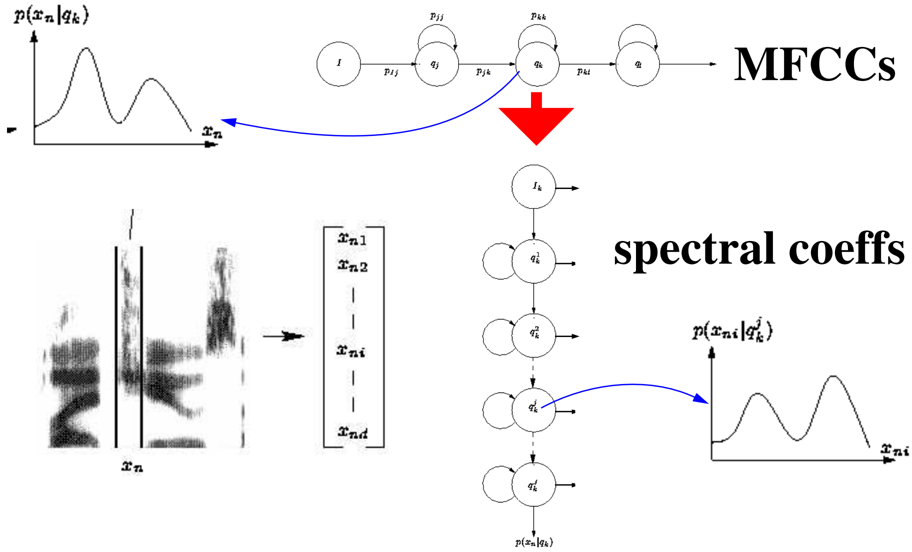
Planned experiments with continuous DBN include:

- combining acoustic with pitch/phonetic/other data
- combining spectral acoustic with mismatch data (missing data being a special case of 0/1 mismatch) - if meaningfull -

Ref. "Automatic speech recognition using dynamic bayesian networks with both acoustic and articulatory variables", T. Stephenson, H. Bourlard, S. Bengio & A. Morris, Proc. ICSLP 2000.

SPHEAR/RESPITE RELATED WORK AT IDIAP

HMM2



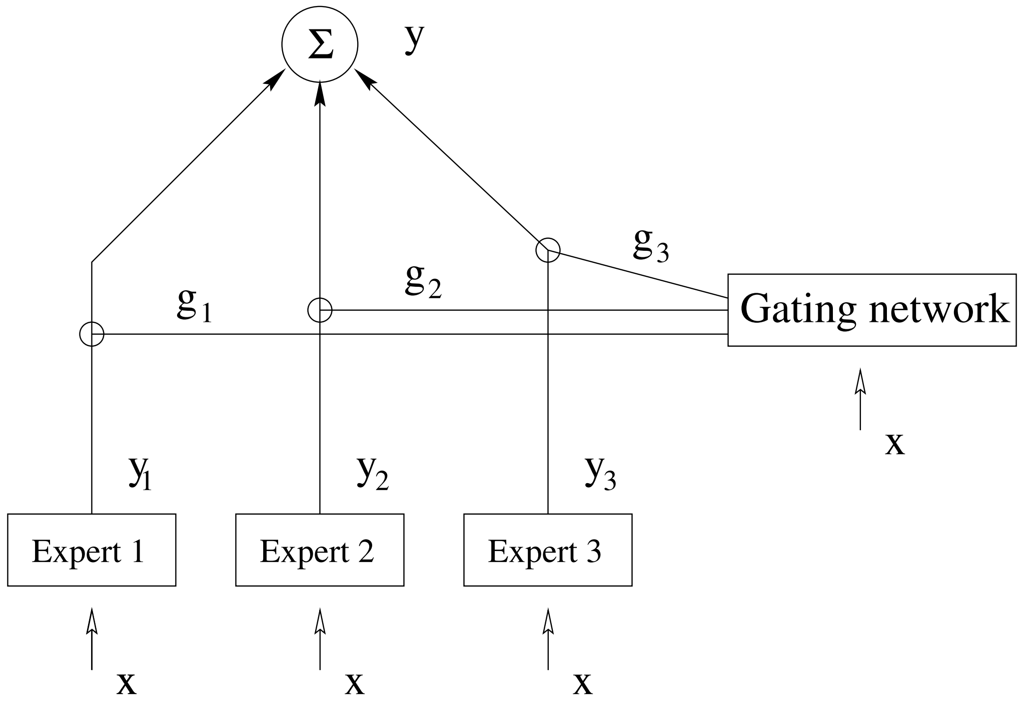
HMM2 is a new HMM architecture for ASR. Each state of each temporal HMM is modelled by frequency HMM. Different topologies have been tested with spectral and cepstral features. Initial results not competitive with conventional HMMs.

However, HMM2 has potential to extract structural information, suitable for example for formant tracking or vocal tract normalisation. Further tests planned with HMM2 variants, and integrating structural info. into other systems.

Ref. "HMM2 - A novel approach to HMM emission probability estimation", K. Weber, S. Bengio, and H. Bourlard, Proc. ICSLP 2000.

SPHEAR/RESPITE RELATED WORK AT IDIAP

MIXTURE OF EXPERTS



A new PhD student (Ronan Collobert) as about to take up previous work at IDIAP (by Perry Moerland) on the the “mixture of experts” model, with a view to applications multi-stream ASR.

This represents an approach to the problem of optimum combination of fixed experts which we have not yet studied.

Mixtures of Experts Estimate A Posteriori Probabilities, Perry Moerland, IDIAP-RR 97-07, 1997

SPHEAR/RESPITE RELATED WORK AT IDIAP: SUMMARY

Andrew Morris. Research Scientist, developing and testing different approaches to **multi-stream ASR.**

- Optimal combination of data streams; Analysis of the problem of data likelihood increasing with noise level for most commonly used acoustic features; RBF network for permitting posteriors estimation with missing data;

Astrid Hagen. PhD Student, developing and testing different approaches to **multi-stream ASR.**

- Applying multi-stream approaches to delta coefficients; FC product as alternative to usual FC sum rule for expert combination; HMM experts for ML based adaptive weighting; combining diverse data streams.

Todd Stevenson. PhD Student, on application of **Bayesian Networks** to ASR.

- Implementation of discrete DBN. This was tested on ASR incorporating articulatory information; implementation of continuous DBN.

Katrin Keller. PhD Student, A new SR approach to Exploiting Dependencies with HMM

- Various experiments with **HMM2**

Ikbal Mohamed. PhD Student, Developing extensions to HMM based ASR.

- Implementing in house code for **HMM2**

Hervé Glotin. PhD Student (co-supervised with ICP, Grenoble) developing CASA based approaches to **multi-stream ASR.**

- CASA based expert weighting.
- Binaural cues for **data enhancement** and **blind separation**.