GMM-based classification from noisy features

Alexey Ozerov [(1)], Mathieu Lagrange [(2)] and Emmanuel Vincent [(1)]

1st September 2011

(1) INRIA, Centre de Rennes - Bretagne Atlantique,
(2) STMS Lab IRCAM - CNRS – UPMC

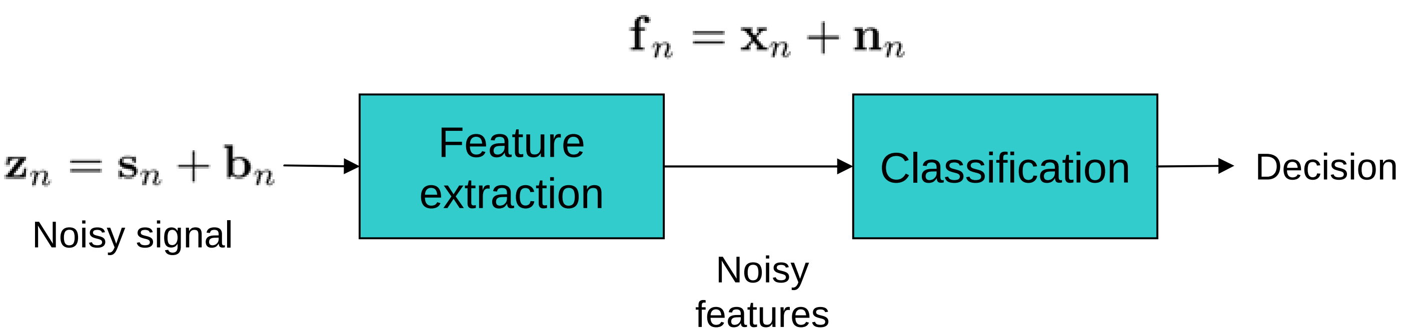
International Workshop on Machine Listening in Multisource Environments (CHiME 2011), Florence, Italy

Outline

- Introduction
- GMM decoding from noisy data
- GMM learning from noisy data
- Experiments
- Conclusions and further work

Introduction

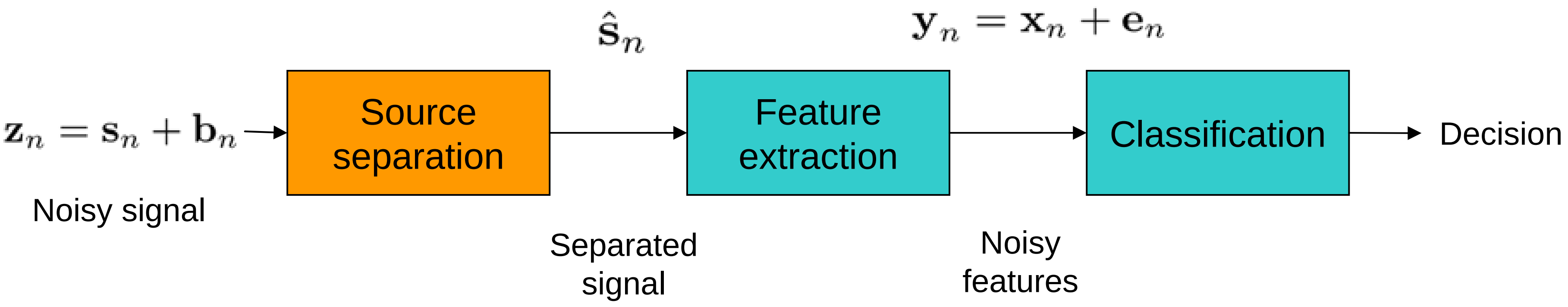
- Classification from noisy data
  - Classification from noisy or multi-source audio



- Poor performance because of high noise variability

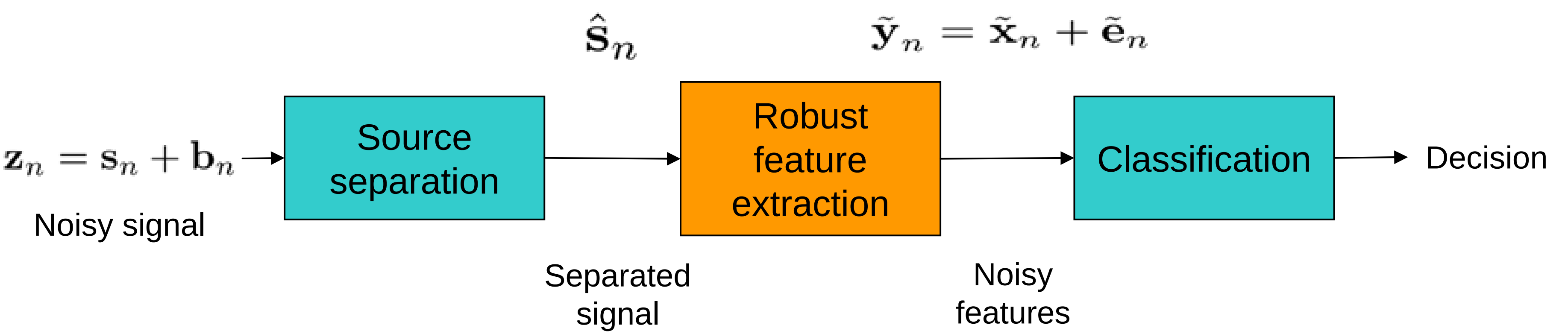
State of the art

- **Signal level:** Noise suppression or source separation



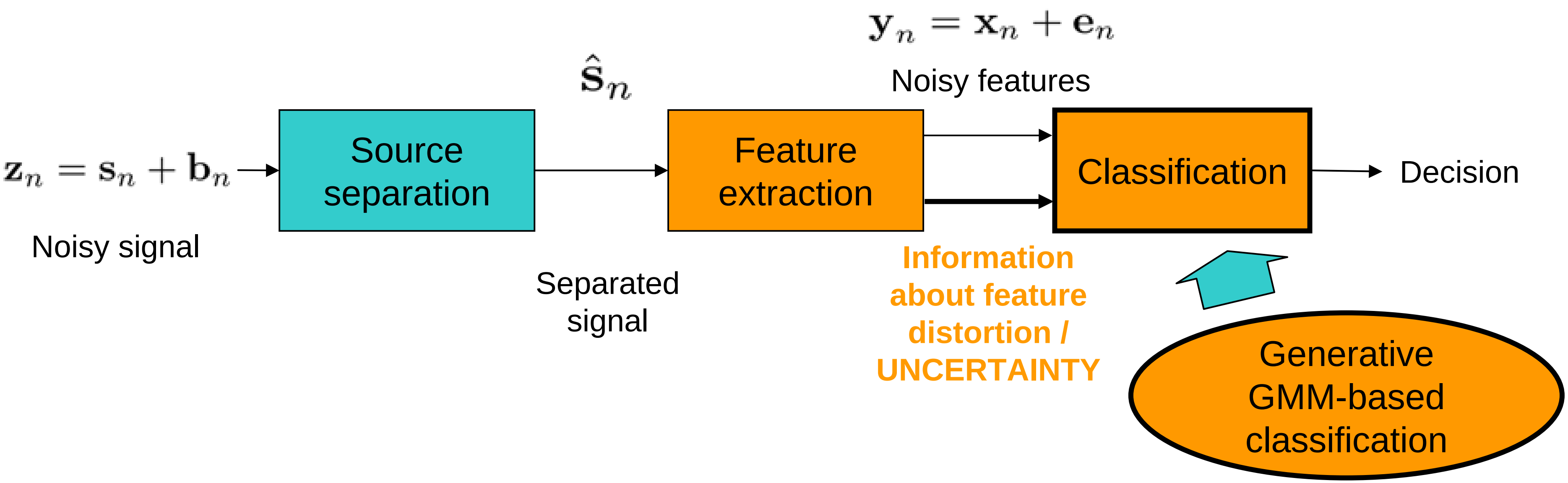
State of the art

- Feature level: Features robust to
  - additive or convolute noise
  - errors produced by source separation



State of the art

- Classifier level: Classification that accounts for possible distortion of the features, given some information about this distortion [Cooke01, Barker05, Deng05, Kolossa10]



State of the art limits and our contributions

- Limit 1: It is assumed that the clean data underlying the noisy observations have been generated by the GMMs.

  [Cooke01, Barker05, Deng05, Kolossa10]

- Contribution 1: Introduction and investigation of a new data-driven criterion for GMM learning and decoding as an alternative to the model-driven criterion.

State of the art limits and our contributions

- Limit 2: Uncertainty is taken into account only at the decoding stage, assuming that the GMMs were trained from some clean data. **[Cooke01, Barker05, Deng05, Kolossa10]**

- Contribution 2: Deriving two new Expectation Maximization (EM) algorithms allowing learning GMMs from noisy data with Gaussian uncertainty for the both criteria considered.

Outline

- Introduction
- GMM decoding from noisy data
- GMM learning from noisy data
- Experiments
- Conclusions and further work

GMM decoding from noisy data

- GMM $\theta = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \omega_i\}_{i=1}^{I}$

$$p(\mathbf{x}_n|\theta) = \sum_{i=1}^{I} \omega_i N(\mathbf{x}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

- Uncertainties
  - Binary (either observed or missing) **[Cooke01, Barker05]**
  - Gaussian (“*asymptotically*” more general) **[Deng05, Kolossa10]**

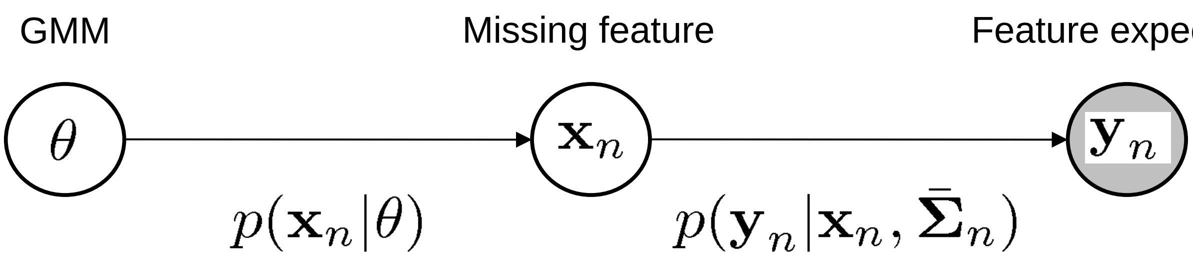
$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{e}_n \qquad \mathbf{x}_n \sim \mathcal{N}(\mathbf{y}_n, \bar{\boldsymbol{\Sigma}}_n)$$

known unknown unknown known

Criteria

- Criterion 1: Model-driven criterion (*likelihood integration*) [state of the art]

[Deng05, Kolossa10]



$$f_{\mathrm{LI}}(\mathbf{y}, \bar{\mathbf{\Sigma}}|\theta) = \boxed{\int_{\mathbb{R}^{M\times N}} p(\mathbf{y}|\mathbf{x}, \bar{\mathbf{\Sigma}})p(\mathbf{x}|\theta)d\mathbf{x}}$$

$$= \prod_{n=1}^{N}\sum_{i=1}^{I}\omega_i N(\mathbf{y}_n|\boldsymbol{\mu}_i, \mathbf{\Sigma}_i + \bar{\mathbf{\Sigma}}_n)$$

Criteria

- Criterion 2: Data-driven criterion (*log-likelihood integration*) [proposed]

$$f_{\mathrm{LLI}}(\mathbf{y}, \bar{\mathbf{\Sigma}}|\theta) = \mathbb{E}_{\mathbf{x}}\left[\log p(\mathbf{x}|\theta)|\mathbf{y}, \bar{\mathbf{\Sigma}}\right]$$

$$= \boxed{\int_{\mathbb{R}^{M \times N}} p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{\Sigma}}) \log p(\mathbf{x}|\theta) d\mathbf{x}}$$

$$= \sum_{n=1}^{N} \int_{\mathbb{R}^{M}} p(\mathbf{x}_n|\mathbf{y}_n, \bar{\mathbf{\Sigma}}_n) \log \sum_{i=1}^{I} \omega_i N(\mathbf{x}_n|\boldsymbol{\mu}_i, \mathbf{\Sigma}_i)$$

Outline

- Introduction
- GMM decoding from noisy data
- GMM learning from noisy data
- Experiments
- Conclusions and further work

GMM learning from noisy data

- Binary uncertainty
  - EM algorithm [Ghahramani&Jordan94]
- Gaussian uncertainty
  - We derived two new EM algorithms for the both criteria considered

GMM learning from noisy data

Algorithm 1 One iteration of the EM algorithm for the likelihood integration-based GMM learning from noisy data.

E step. Conditional expectations of natural statistics:

$$\gamma_{i,n} \propto \omega_i N(\mathrm{y}_n|\mu_i, \Sigma_i + \bar{\Sigma}_n), \quad \text{and} \quad \sum_i \gamma_{i,n} = 1, \quad (13)$$

$$\hat{\mathrm{x}}_{i,n} = \mathrm{W}_{i,n}(\mathrm{y}_n - \mu_i) + \mu_i, \quad (14)$$

$$\widehat{\mathrm{R}}_{\mathrm{xx},i,n} = \hat{\mathrm{x}}_{i,n}\hat{\mathrm{x}}_{i,n}^T + (\mathrm{I} - \mathrm{W}_{i,n})\Sigma_{\mathrm{x},i}, \quad (15)$$

where

$$\mathrm{W}_{i,n} = \Sigma_i \left[\Sigma_i + \bar{\Sigma}_n\right]^{-1}. \quad (16)$$

M step. Update GMM parameters:

$$\omega_i = \frac{1}{N}\sum_{n=1}^{N}\gamma_{i,n}, \quad (17)$$

$$\mu_i = \frac{1}{\sum_{n=1}^{N}\gamma_{i,n}}\sum_{n=1}^{N}\gamma_{i,n}\hat{\mathrm{x}}_{i,n}, \quad (18)$$

$$\Sigma_i = \frac{1}{\sum_{n=1}^{N}\gamma_{i,n}}\sum_{n=1}^{N}\gamma_{i,n}\widehat{\mathrm{R}}_{\mathrm{xx},i,n} - \mu_i\mu_i^T. \quad (19)$$

Algorithm 2 One iteration of the EM algorithm for the log-likelihood integration-based GMM learning from noisy data.

E step. Conditional expectations of natural statistics:

$$\gamma_{i,n} \propto \omega_i N(\mathrm{y}_n|\mu_i, \Sigma_i)e^{-\frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\bar{\Sigma}_n)}, \quad \text{and} \quad \sum_i \gamma_{i,n} = 1, \quad (20)$$

M step. Update GMM parameters:

$$\omega_i = \frac{1}{N}\sum_{n=1}^{N}\gamma_{i,n}, \quad (21)$$

$$\mu_i = \frac{1}{\sum_{n=1}^{N}\gamma_{i,n}}\sum_{n=1}^{N}\gamma_{i,n}\mathrm{y}_n, \quad (22)$$

$$\Sigma_i = \frac{\sum_{n=1}^{N}\gamma_{i,n}(\mathrm{y}_n - \mu_i)(\mathrm{y}_n - \mu_i)^T + \bar{\Sigma}_n}{\sum_{n=1}^{N}\gamma_{i,n}} \quad (23)$$

Needed some approximations

Generalizes "*asymptotically*" the binary uncertainty EM [Ghahramani&Jordan94]

Outline

- Introduction
- GMM decoding from noisy data
- GMM learning from noisy data
- Experiments
- Conclusions and further work

Artificial uncertainty

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{e}_n$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{y}_n, \bar{\mathbf{\Sigma}}_n)$$

$$\bar{\mathbf{\Sigma}}_n = \mathrm{diag}\left\{[\bar{\sigma}^2_{m,n}]_m\right\}$$

1. $\log \bar{\sigma}^2_{m,n}$ is drawn from a Gaussian
2. $\mathbf{e}_n$ is drawn from $\mathcal{N}(0, \bar{\mathbf{\Sigma}}_n)$

- Artificial uncertainty
  - gives us a possibility to control some characteristics of the uncertainty,
  - allows us leaving the study of the following situations for further work:
    - realistic feature-corrupting noise,
    - estimated uncertainty covariances.

Characteristics of the uncertainty

- Feature to Noise Ratio (FNR) (dB)

$$\mathrm{FNR} = 10\log_{10}\frac{\sum_n \|\mathbf{x}_n\|^2}{\sum_n \|\mathbf{x}_n - \mathbf{y}_n\|^2}$$

- Noise Variation Level (NVL) (dB)

$$\mathrm{NVL} = \mathrm{stdev}\left(\left\{10\log_{10}\bar{\sigma}^2_{m,n}\right\}_{m,n}\right)$$

Evaluated setups

- All possible combinations of

$$\mathrm{FNR}_{\mathrm{train}} = \{-20, -10, 0, 10, 20\}$$

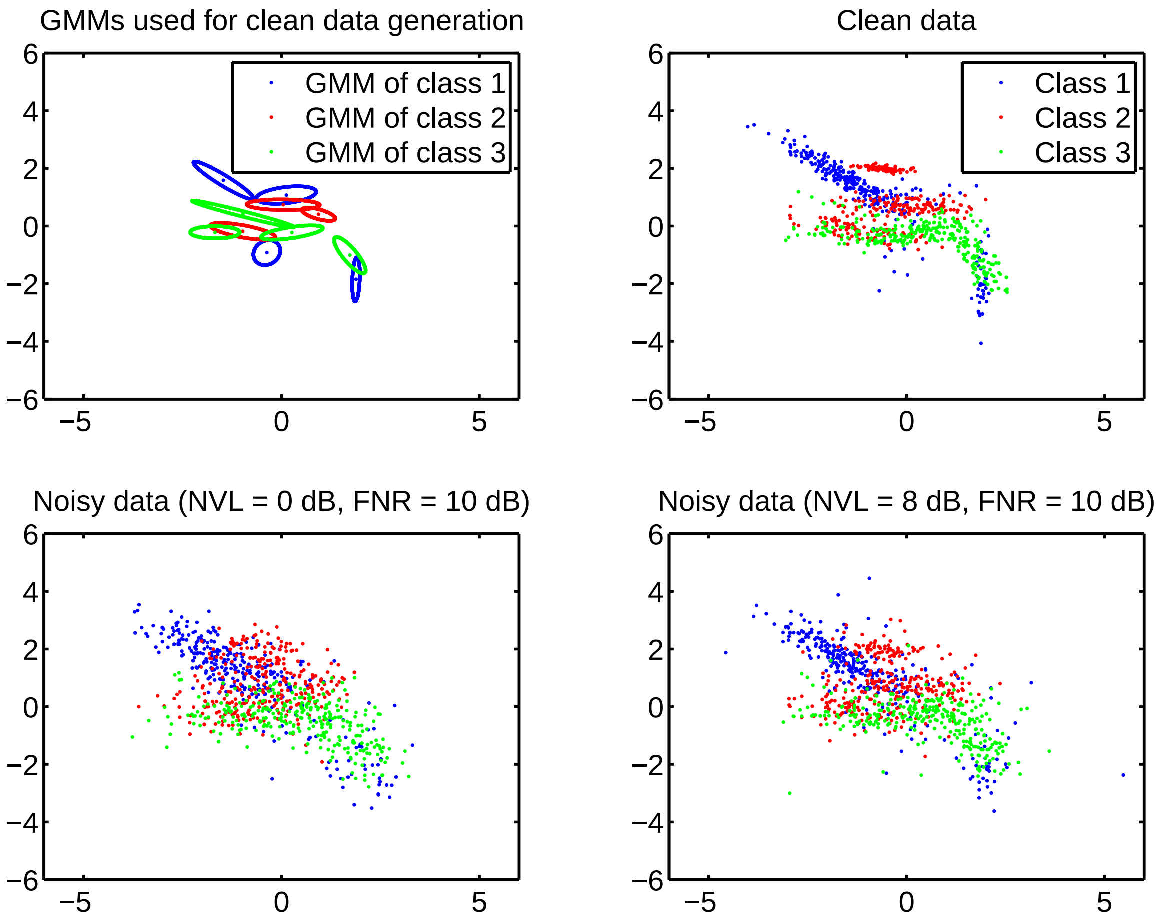
$$\mathrm{FNR}_{\mathrm{test}} = \{-20, -10, 0, 10, 20\}$$

$$\mathrm{NVL}_{\mathrm{train}} = \{0, 4, 8\}$$

$$\mathrm{NVL}_{\mathrm{test}} = \{0, 2, 4, 6, 8\}$$

- 375 setups

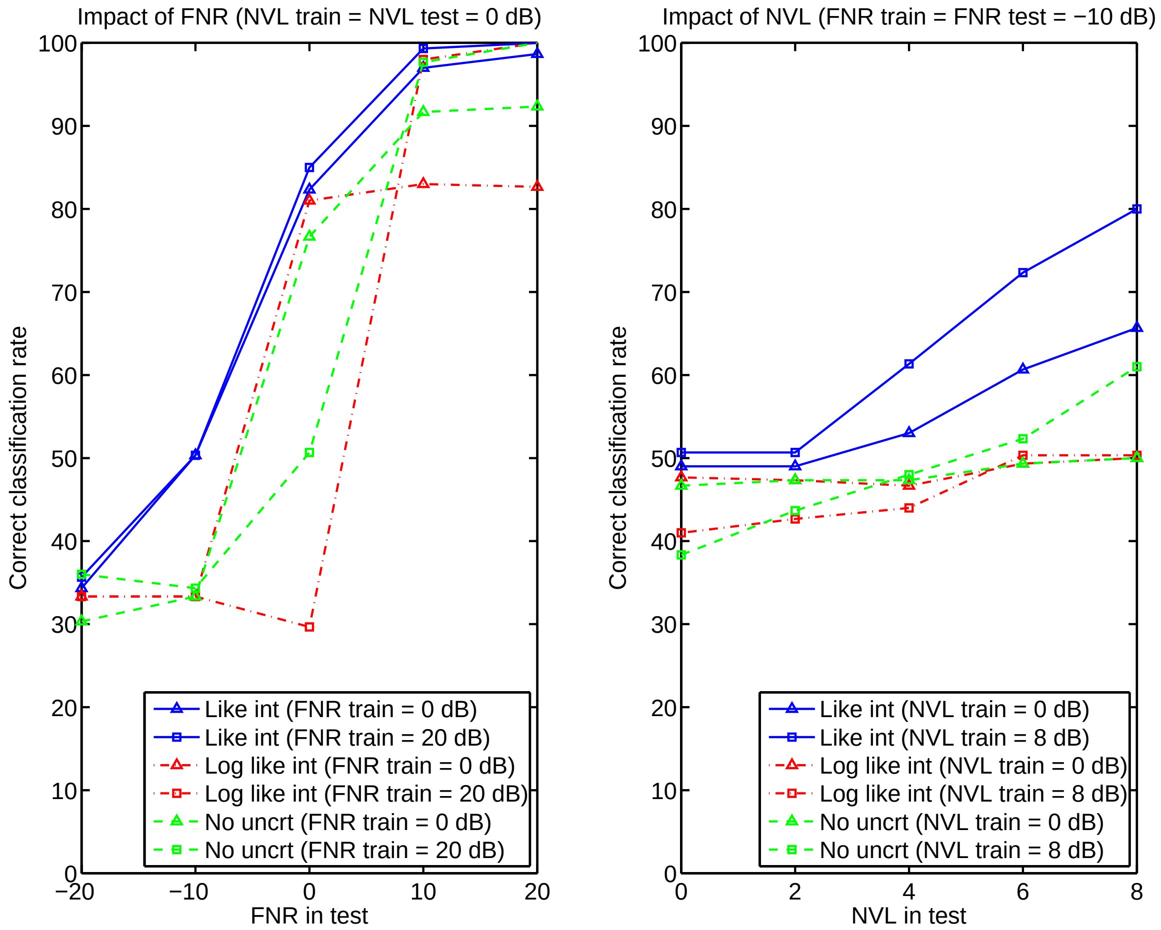
Artificial data



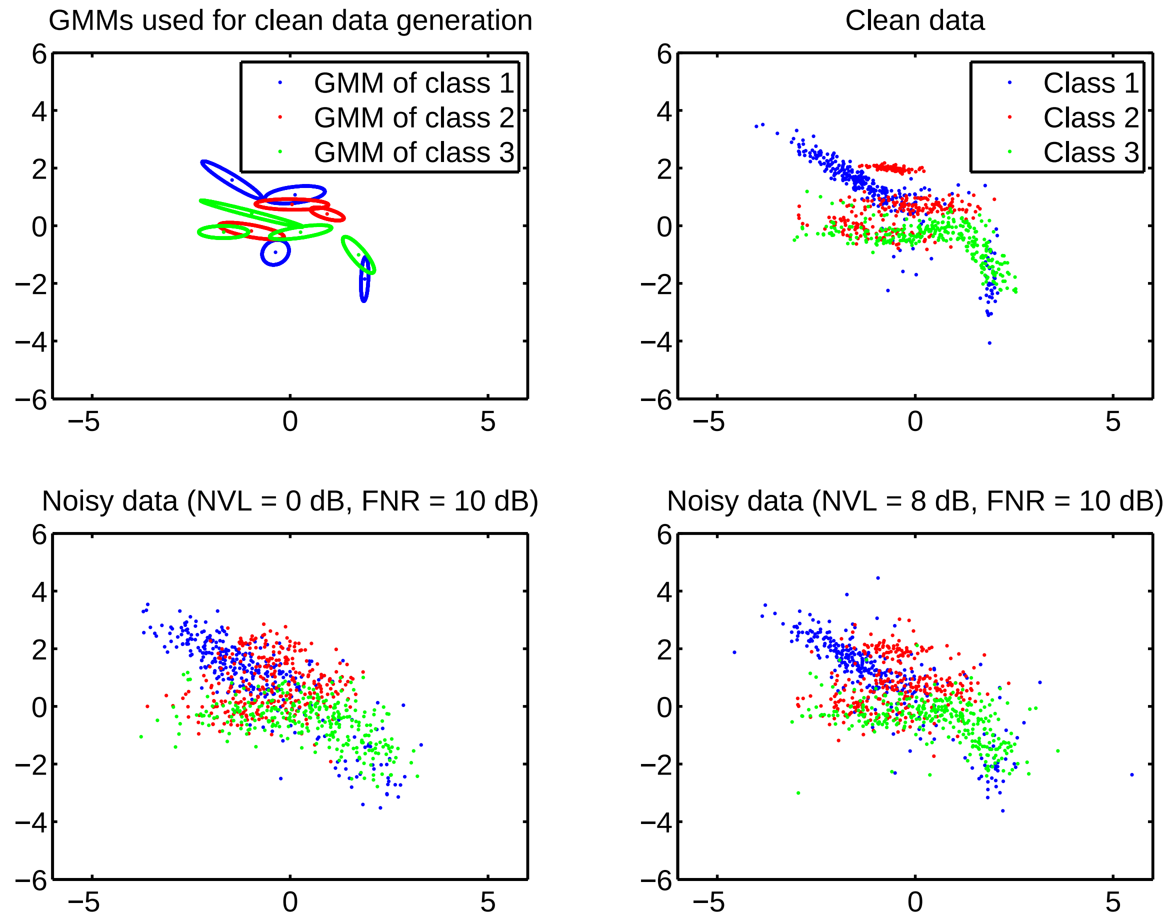
Real data

- Speaker recognition task
- Setting is quite similar to [Reynolds95]
  - TIMIT database
  - 10 male speakers
  - 16-states GMMs
  - Feature space dimension = 20
- Differences with [Reynolds95]
  - Features: Logarithms of Mel-Frequency Filter-Bank outputs (LMFFB) instead of MFCC
  - GMMs with full covariance matrices

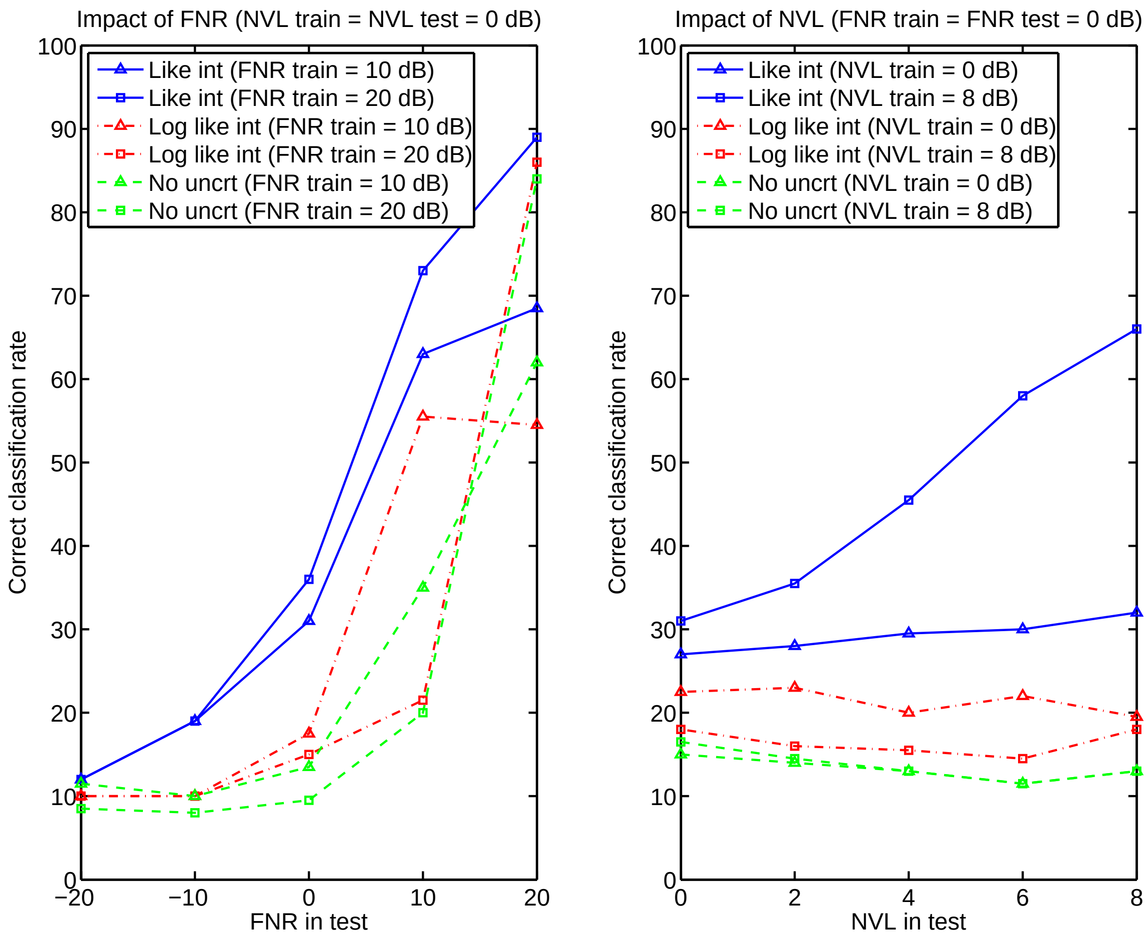
Artificial data results



Artificial data



Real data results



Outline

- Introduction
- GMM decoding from noisy data
- GMM learning from noisy data
- Experiments
- Conclusions and further work

Conclusions and further work

- Conclusions
  - We validate the model-driven uncertainty decoding approach as compared to a data-driven approach.
  - We show that considering the uncertainty allows us to
    - handle the heterogeneity of noise between the training and testing sets,
    - exploit the variability of noise for improved performance.
- Further work
  - Considering realistic feature-corrupting noise and uncertainty covariances estimation.
  - Considering the log-likelihood integration within a GMM-based classification framework with discriminative training.

References

- [Cooke01] M. Cooke, "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Communication, vol. 34, no. 3, pp. 267–285, Jun. 2001.
- [Barker05] J. Barker, M. Cooke, and D. Ellis, "Decoding speech in the presence of other sources," Speech Communication, vol. 45, no. 1, pp. 5–25, Jan. 2005.
- [Deng05] L. Deng, J. Droppo, and A. Acero, "Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 3, pp. 412–421, May 2005.
- [Kolossa10] D. Kolossa, R. Fernandez Astudillo, E. Hoffmann, and R. Orglmeister, "Independent component analysis and time-frequency masking for speech recognition in multitalker conditions," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2010, pp. 1–14, 2010.
- [Ghahramani&Jordan94] Z. Ghahramani and M. Jordan, "Supervised learning from incomplete data via an EM approach," in Advance on Neural Information Processing Systems, 1994, pp. 120–127.
- [Reynolds95] D. Reynolds, "Large population speaker identification using clean and telephone speech," IEEE Signal Processing Letters, vol. 2, no. 3, pp. 46–48, Mar. 1995.