Factor Graph Decoding for Speech Presence Probability Estimation

Thomas Glarner, Mohammad Mahdi Momenzadeh, Lukas Drude, Reinhold Häb-Umbach

Department of Communications Engineering
Paderborn University
5th October, 2016

Speech Presence Probability Estimation

- Goal: Find regions of speech activity in time frequency domain, e.g. for Speech Enhancement
- Assumption: Speech only present in some time frequency slots
- Signal model in time frequency domain:

$$X(m,k) = \begin{cases} N(m,k), & Z(m,k) = 0 \\ S(m,k) + N(m,k), & Z(m,k) = 1 \end{cases}$$

- Latent variable Z models speech activity for each slot, SPP equal to posterior for Z:

$$\gamma(m,k) = \Pr(Z(m,k)|X(1{:}M, 1{:}K))$$

IID Decoding

- Observation likelihood:
  Use exponential model
  with λ_0=1 for noise-only case

$$p(\xi|Z(m,k)=i) = \lambda_i \mathrm{e}^{\lambda_i \xi}$$

$$\xi = |X(m,k)|^2 / \hat{\Phi}(m,k)$$

- IID SPP through Bayes Theorem:
  With prior for Z(m,k) given

$$\Pr(Z(m,k)|X(m,k))$$

- **But:** Only local observation incorporated
  No utilization of adjacent T-F-slots!

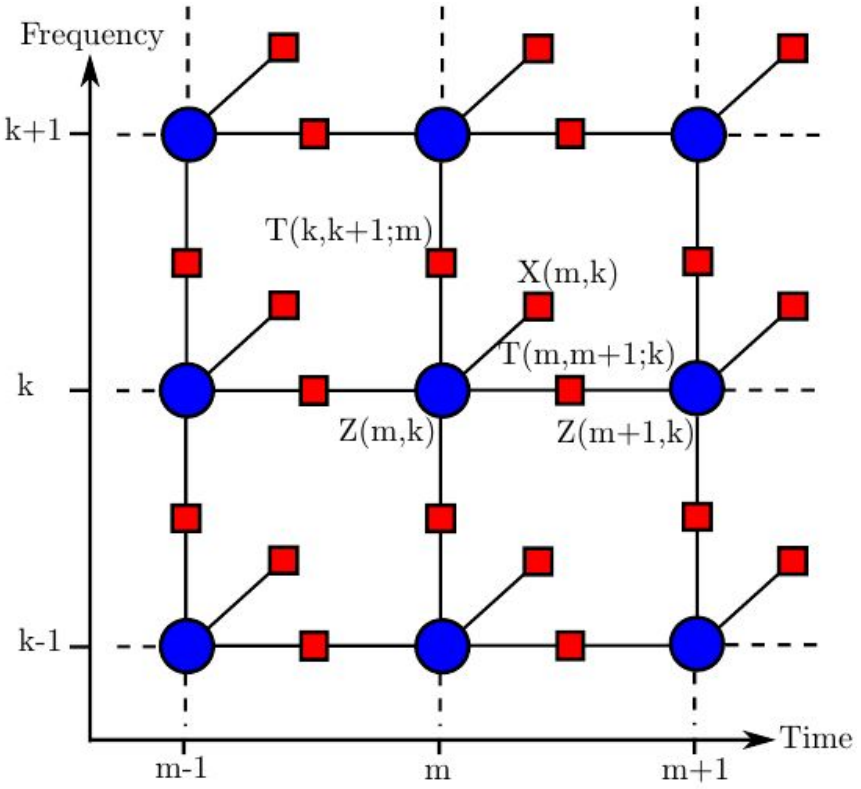
Solution: Model correlations with a 2D Graph

Undirected Graphical Model

- Symmetric Transition Factors, e.g.:

$$T(m, m+1; k) = \frac{\Pr(Z(m,k), Z(m+1,k))}{\Pr(Z(m,k) Pr(m+1,k)}$$

- Can be seen as Markov Random Field
- Problem similar to Ising Model, but with continuous observations
- Many tight loops: No exact inference through Sum Product Algorithm!
  - ➢ Use approximate inference scheme!



Graph Factorization

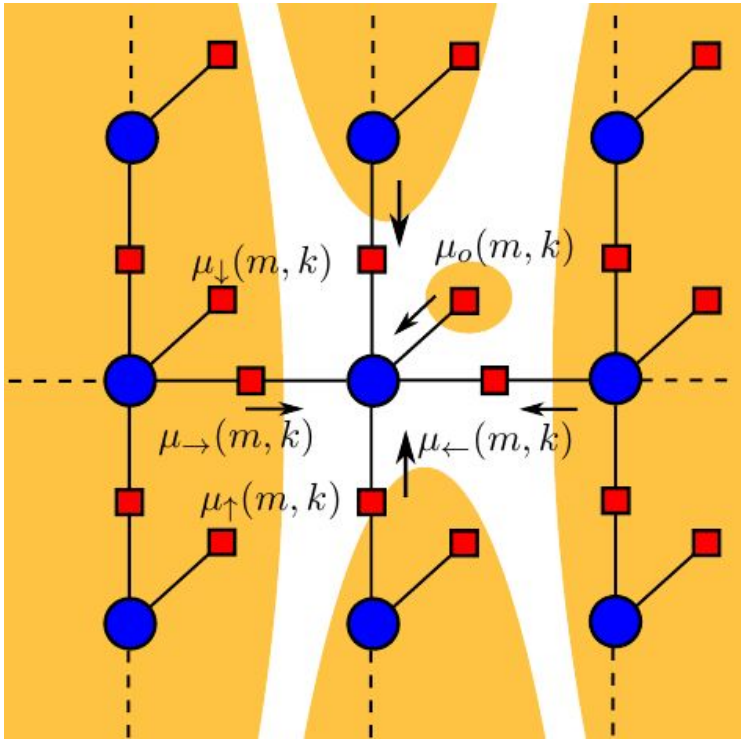
- Horizontal factorization:
  neglect horizontal links except current chain

- Factorization into messages:

$$\gamma^{\mathcal{H}}(m,k) \propto \prod_{d\in\{\leftarrow,\rightarrow,\uparrow,\downarrow,o\}} \mu_d^{\mathcal{H}}(m,k)$$

with e.g. $\mu_{\rightarrow}^{\mathcal{H}}(m,k) = p(X(1{:}m{-}1,1{:}K)|Z(m,k))$

$\mu_o(m,k) = p(X(m,k)|Z(m,k))$

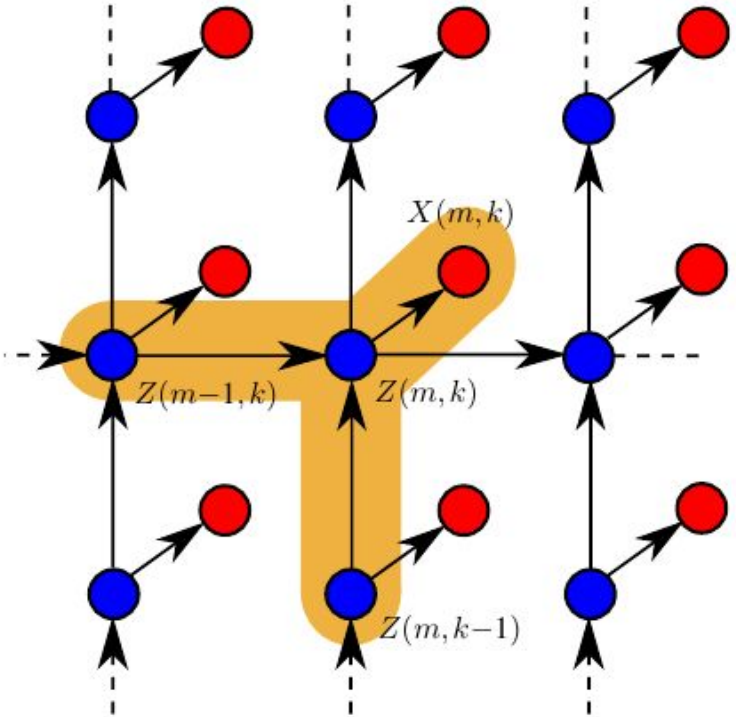


- Messages computable by forward backward algorithm

- Analogous vertical factorization

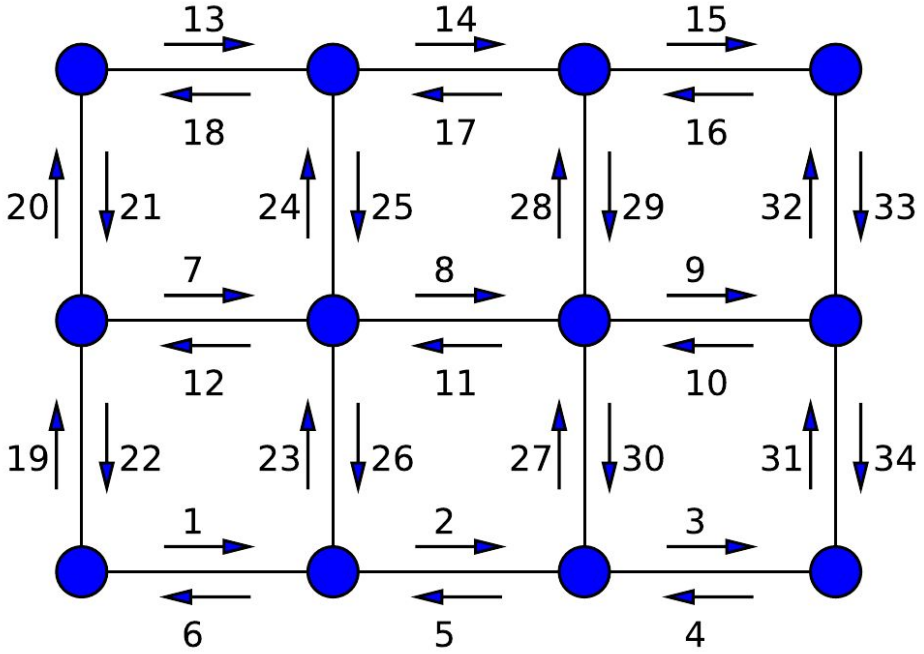
Why Undirected? Drawbacks of 2D-HMM

- Artificial causality assumption along frequency direction
- Factorization not justified due to Head-to-Head relation
- Markov chain inhomogeneous along frequency direction
  - ➢ No stationary distribution!
  - ➢ Prior not consistent between time and frequency direction



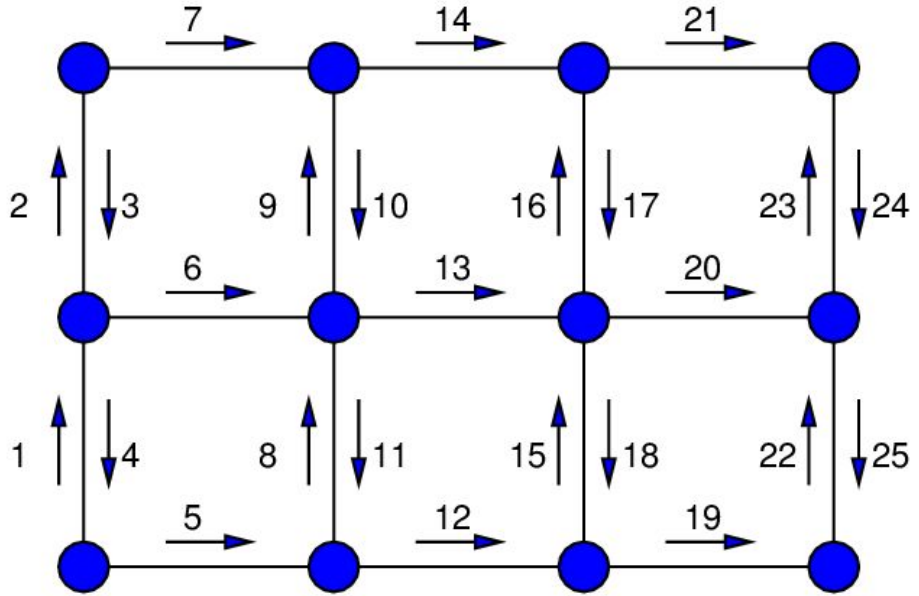
Turbo Schedule

- Alternate between horizontal and vertical factorization
- Only full messages are used
- Comparable to inference in 2D-HMM



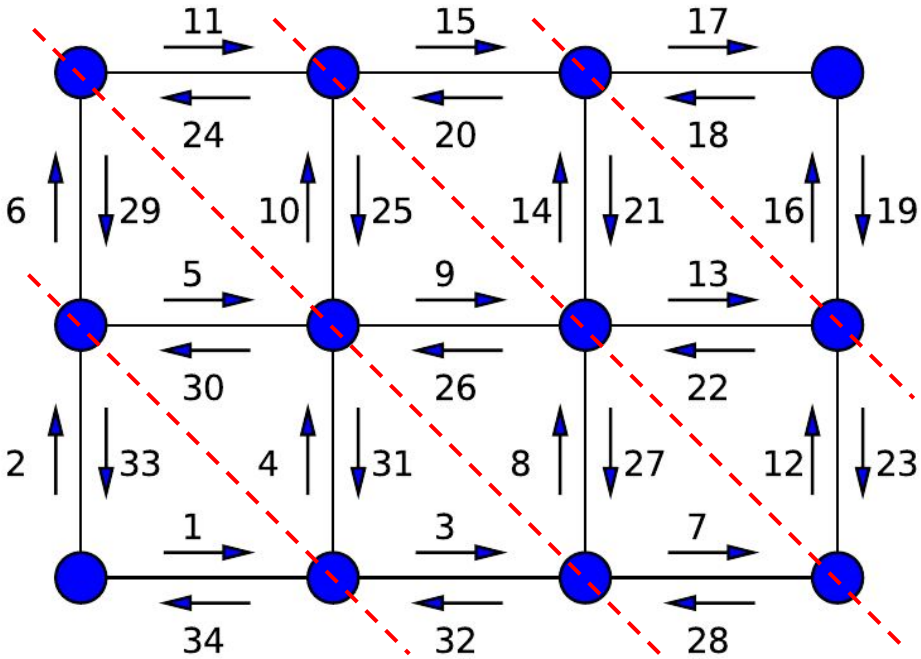
Vertical Dominant Schedule

- Full forward backward in dominant vertical direction
- Only forward or backward in horizontal direction
- Investigates whether favoring one direction is advantageous
- Analogous: horizontal dominant schedule



Plane Wave Schedule

- Diagonal scheme
- Messages flow through graph from one corner to opposing one (like a wave front)
- Investigates whether rectangular flow is really the best approach



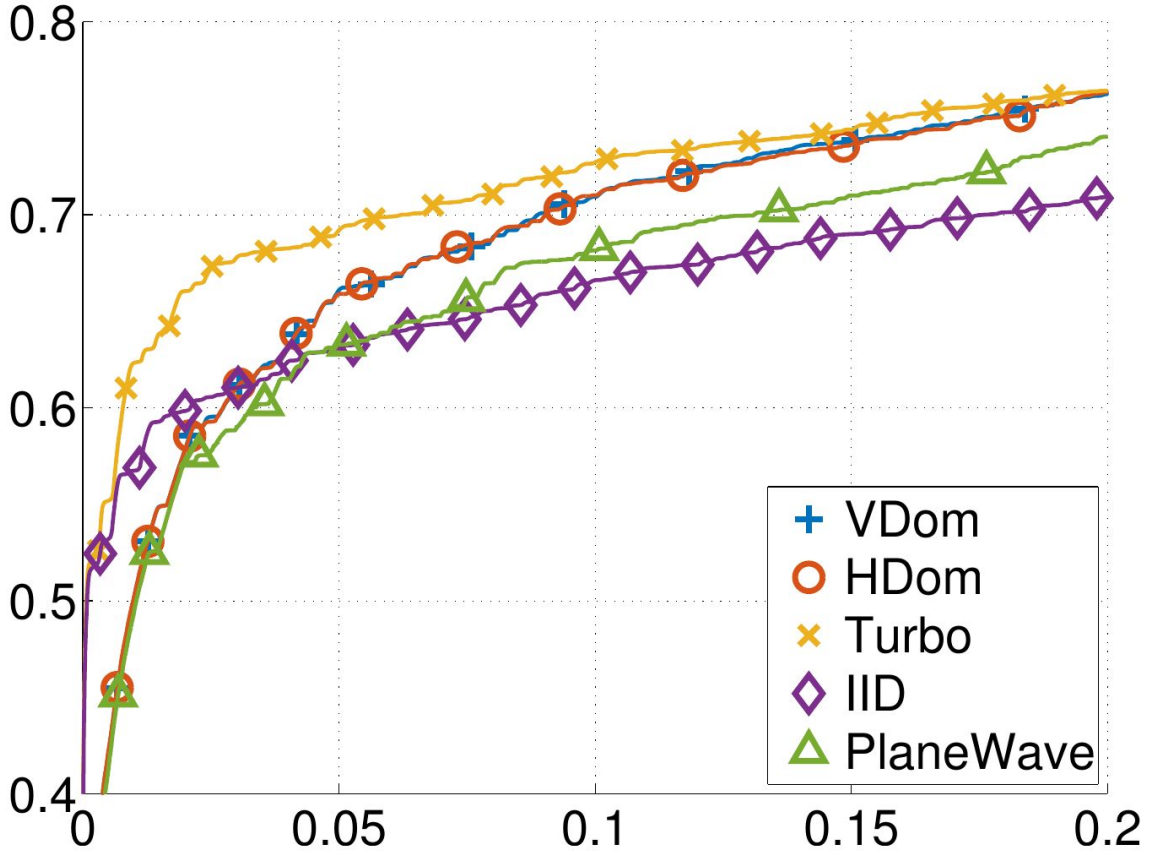
Experimental Setup

- TIMIT Database, all male speakers
  - 3260 utterances in train
  - 1120 utterances in test
- Noise: Additive white gaussian, different SNRs
- Ground truth: Ideal Binary Masks
  - created with 0.9-energy quantile criterion on clean speech STFT
- Model training: Counting on IBMs from TIMIT male train set

Results 1

ROC for SNR = -5 dB

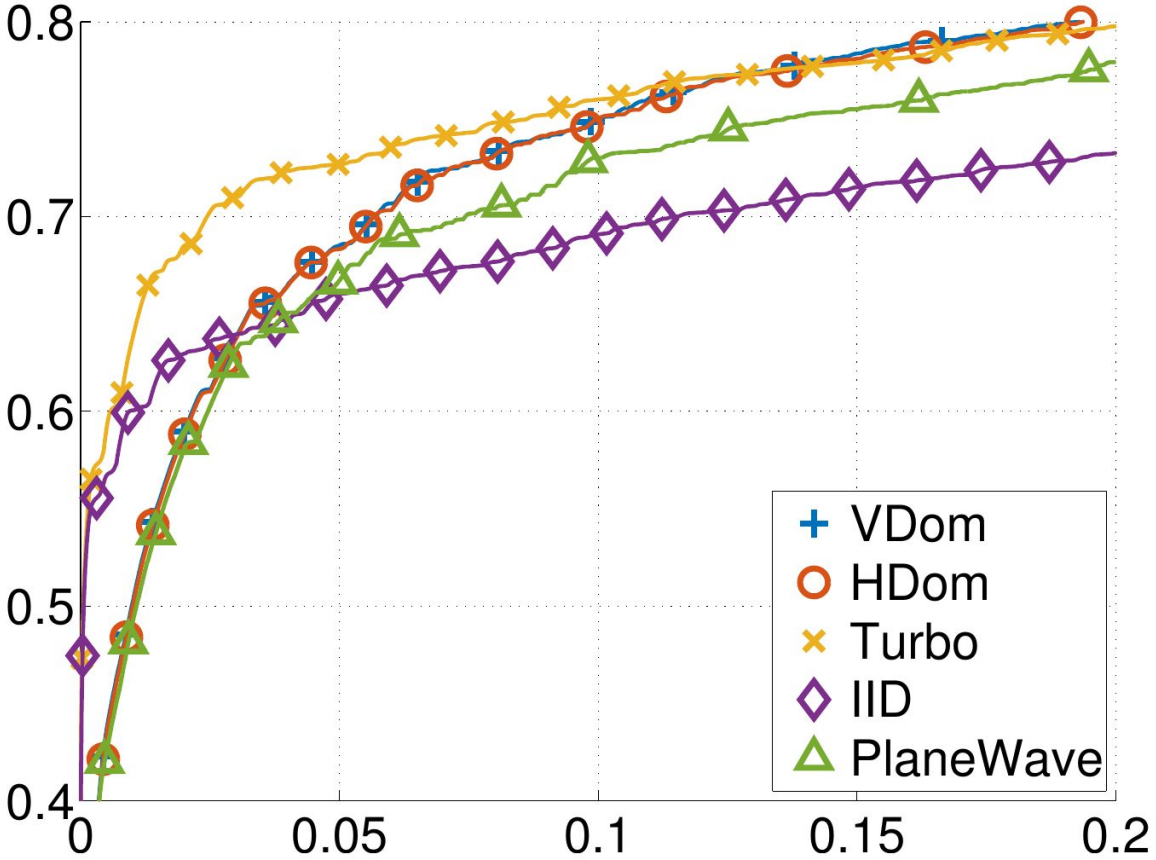
- Turbo Schedule dominates other approaches
- Most Schedules better than IID almost everywhere
- Horizontal and Vertical Dominant Schedules have equal performance
- Plane Wave Schedule not really useful



Results 2

ROC for SNR = 0 dB

- Similar picture to prev. case
- Dominant schedules close or slightly above Turbo for high false alarm rates
- Plane Wave performs better, but still weak in comparison



Conclusions

- Undirected Graphical Model is theoretically advantageous:
  - Equations simplified
  - No contradictions between
    - equations and graph
    - directions
- Results similar to 2D-HMM
- Higher flexibility allows for different message passing schedules
- Turbo Schedule still outperforms other tested approaches