HIDDEN MARKOV MODELS

Theory and applications

Contents

Markov Processes

- Markov process ➔ memoryless random process $X(t)$
- Time-domain discrete ($\{0,1,2\ldots,n\}$) or continuous ($[0,t]$)
- State-space discrete ({blue, red, green}) or continuous (temperature,...)
- Memoryless condition:

$$\Pr\left(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots X_0 = x_0\right) = \Pr\left(X_n = x_n \mid X_{n-1} = x_{n-1}\right)$$

- Stationary or homogeneous condition:

$$\Pr\left(X_n = x_n \mid X_{n-1} = x_{n-1}\right) = \Pr\left(X_1 = x_1 \mid X_0 = x_0\right) = p_{x_0 x_1}$$

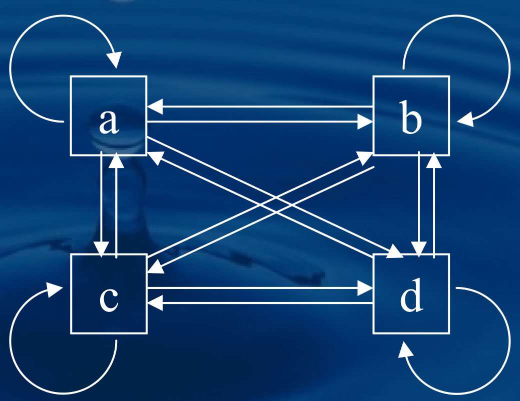
- Example: checkout line

Markov Processes

- Transition matrix:

$$\begin{array}{cc} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} p_{aa} & p_{ba} & p_{ca} & p_{da} \\ p_{ab} & p_{bb} & p_{cb} & p_{db} \\ p_{ac} & p_{bc} & p_{cc} & p_{dc} \\ p_{ad} & p_{bd} & p_{cd} & p_{dd} \end{pmatrix} \end{array}$$

- Transition graph



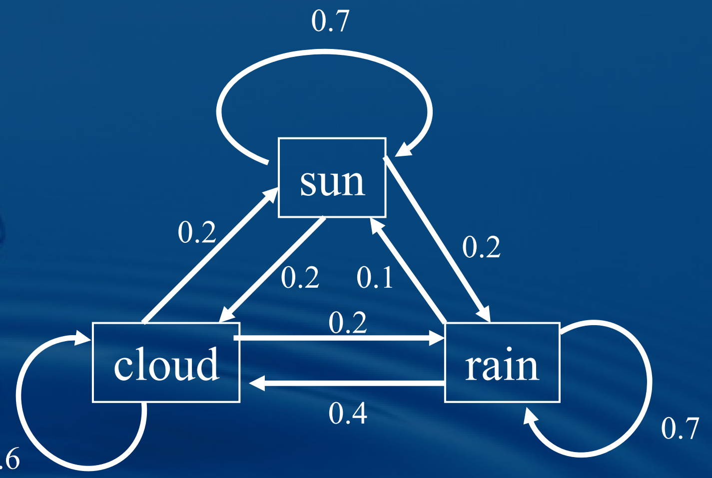
- Initial state probabilities

$$\pi = \begin{pmatrix} \pi_a \\ \pi_b \\ \pi_c \\ \pi_d \end{pmatrix}$$

Example

- Time-domain: {M,T,W,R,F,S,D}
- Space-domain: {sun, cloud, rain}
- Transition matrix:

$$\begin{pmatrix} \mathbf{0.7} & \mathbf{0.2} & \mathbf{0.1} \\ \mathbf{0.2} & \mathbf{0.6} & \mathbf{0.2} \\ \mathbf{0.2} & \mathbf{0.4} & \mathbf{0.4} \end{pmatrix}$$



- Example:

$$\begin{aligned}\Pr\left(T = sun, W = cloud \mid M = sun\right) &= \Pr\left(T = sun \mid M = sun\right) \times \Pr\left(W = cloud \mid T = sun\right) \\ &= \mathbf{0.7} \times \mathbf{0.2} \\ &= \mathbf{0.14}\end{aligned}$$

Markov Process Applications

- Finance
- Telecommunication networks
- Game theory
- Decision making

Contents

- Markov Processes
- **Hidden Markov Models**
- Applications in music retrieval
  - Folk music classification
  - Chord segmentation and recognition
  - Score following
  - Melody spotting
  - Optical music recognition
- Other applications

Hidden Markov Models

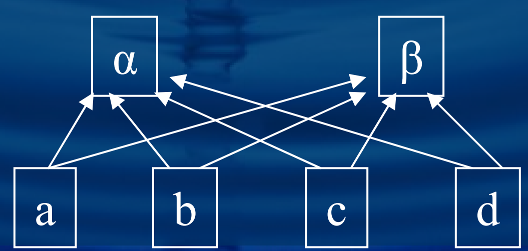
- Now, we can't observe the states
- We know an information related to the states: the observable space
- We know the observation symbol probability:

$$\Pr\left(\alpha \mid X = a\right) = q_{a\alpha}$$

- Confusion matrix:

$$\begin{array}{cc} & \begin{array}{cc} \alpha & \beta \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} q_{a\alpha} & q_{a\beta} \\ q_{b\alpha} & q_{b\beta} \\ q_{c\alpha} & q_{c\beta} \\ q_{d\alpha} & q_{d\beta} \end{pmatrix} \end{array}$$

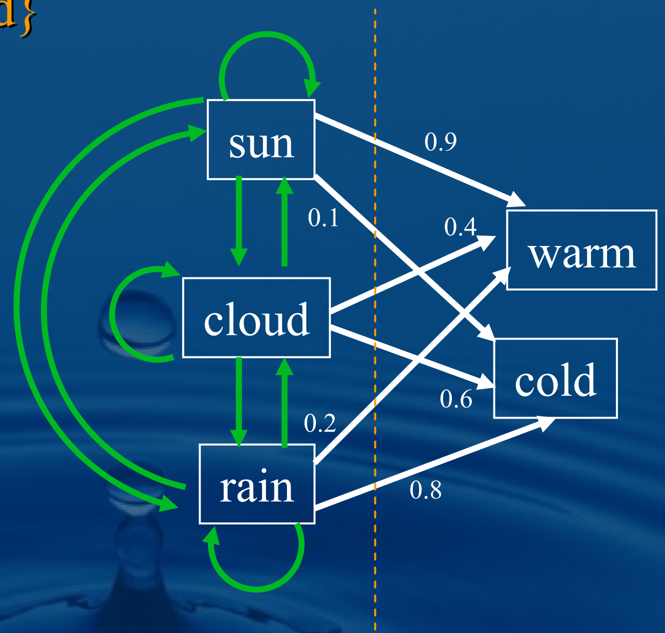
- Confusion graph:



Example

- Observable space: {warm, cold}
- Confusion matrix:

$$\begin{pmatrix} \mathbf{0.9} & \mathbf{0.1} \\ \mathbf{0.4} & \mathbf{0.6} \\ \mathbf{0.2} & \mathbf{0.8} \end{pmatrix}$$



- Example:

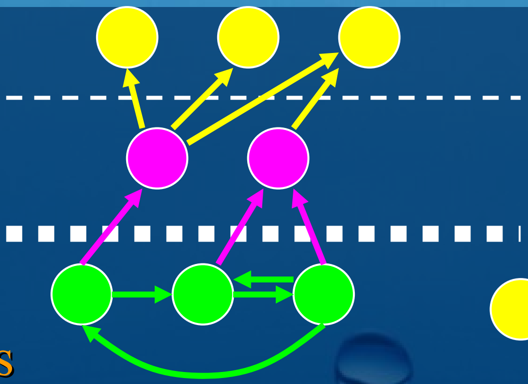
$$\begin{aligned}(T = cold \mid M = sun) &= \Pr(T = cold \mid T = sun) \times \Pr(T = sun \mid M = sun) \\ &+ \Pr(T = cold \mid T = cloud) \times \Pr(T = cloud \mid M = sun) \\ &+ \Pr(T = cold \mid T = rain) \times \Pr(T = rain \mid M = sun) \\ &= \mathbf{0.27}\end{aligned}$$

Categories of problems

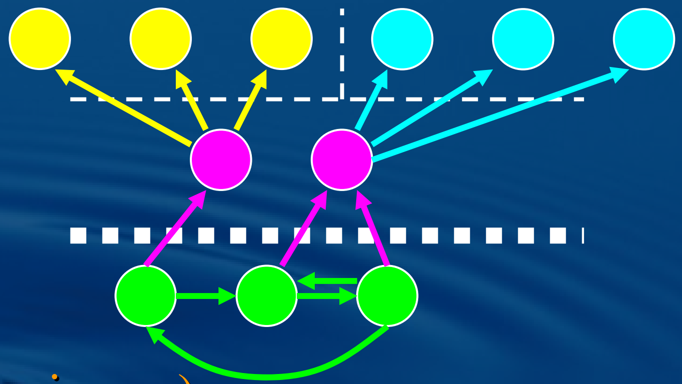
- Evaluation: given a sequence of observations (e.g. {cold, warm, warm}), what is the probability the model produced it?
  - ➔Forward algorithm
  - ➔Does the model fits observations?
- Decoding: given a sequence of observations, what is the most probable sequence of events of the model that produced it?
  - ➔Viterbi algorithm
  - ➔What is the actual sequence of events?
- Learning: given a sequence of observations, what model would best fit it?
  - ➔Forward-backward algorithm
  - ➔What is the actual model?

Improved HMMs

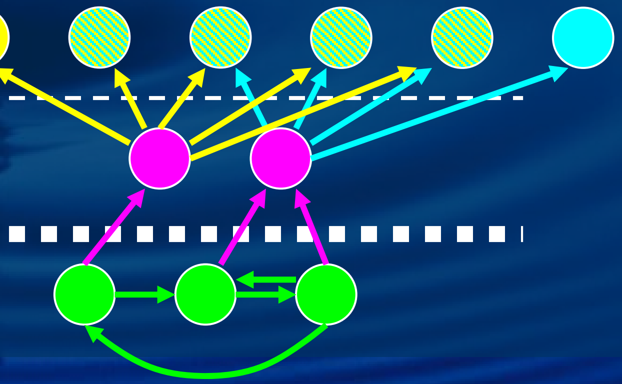
- Layered HMMs



- Hierarchical HMMs



- HMMs with observation distribution (Gaussian,…)



General HMM-based recognition

- Definition of HMM characteristics
  - Complexity of HMM implementation
  - Observable space(s) – Number of layers, layer dependence
  - State space(s) - Number of states, number of layers, structural properties
- Training of HMMs ➔ Forward-backward algorithm
- Recognition ➔ Viterbi algorithm

Contents

- Markov Processes
- Hidden Markov Models
- **Applications in music information retrieval**
  - Folk music classification
  - Chord segmentation and recognition
  - Score following
  - Melody spotting
  - Optical music recognition
- Other applications

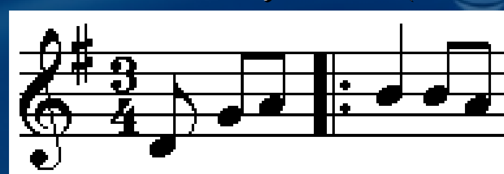
Folk Music Classification

(Chai & Vercoe, MIT, 2001)

- Data set: Austrian (104), German (200) and Irish (187) folk music

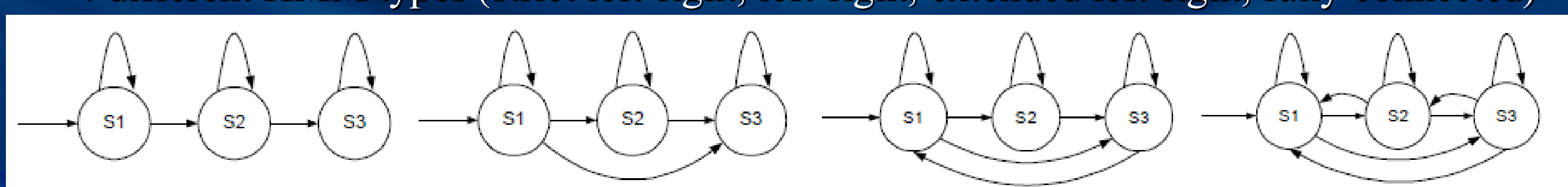
  ➔Monophonic melodies

- Four different observable representations
  - Absolute pitch on one octave – 12 symbols
  - Absolute pitch with duration representation (½ beat repetition) – 12 symbols
  - Interval representation from -13 to +13 semi-tones – 27 symbols
  - Contour representation – 5 symbols (No change: 0; 1 or 2 semi-tones: +/-; >3 semi-tones: ++/--)



- {2,7,9,11,11,9}
- {2,7,9,**11**,**11**,11,9}
- {5,2,2,0,-2}
- {++, +, +, 0,-}

- One HMM by country trained by Baumed-Welsh method
  - 4 different numbers of hidden states (2, 3, 4, 6)
  - 4 different HMM types (strict left-right, left-right, extended left-right, fully connected)



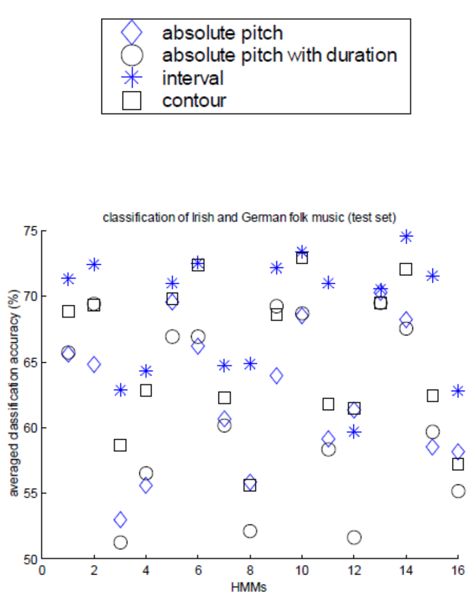
Folk Music Classification

(Chai & Vercoe, MIT, 2001)

- Training set: 30% of the data set chosen randomly
- Results:
  - Little influence of the number of hidden states
  - Strict left-right and left-right types outperformed the others
  - Interval representation generally performs better
  - Quantitative rating of music style similarity (classification accuracy, distance of 2 HMMs…)

Table 2: *Classification performances of 6-state left-right HMM using different representations. The first three rows correspond to 2-way classifications. The last row corresponds to the 3-way classification. I: Irish music; G: German music; A: Austrian music.*

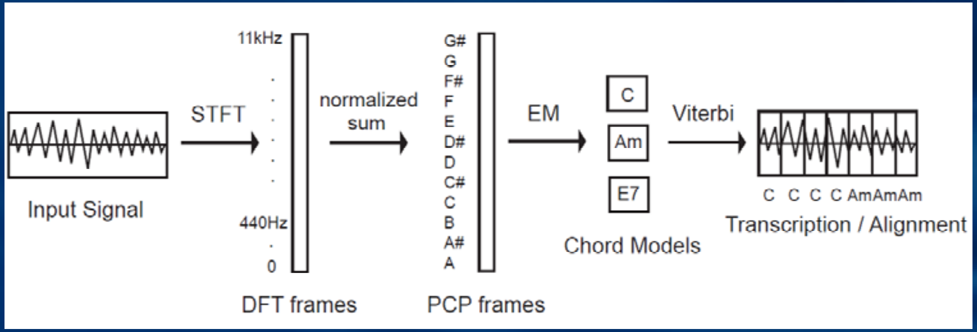
Classes	rep. A	rep. B	rep. C	rep. D
I-G	68%	68%	75%	72%
I-A	75%	74%	77%	70%
G-A	63%	58%	66%	58%
I-G-A	56%	54%	63%	59%



Chord Segmentation and Recognition

(Sheh and Ellis, Columbia, 2003)

- Input: unstructured, polyphonic, and multi-timbre audio from popular music
- HMMs trained with the expectation-maximization algorithm
- 1 chord = 1 process state = 1 distribution of Pitch Class Profile vectors
- 2-level HMMs: 1 state → 1 distribution → Several observed frames
- 2 tests
  - Segmentation (chord sequence known)
  - Unconstrained recognition
- Weighted averaging of rotated PCP vectors improvement



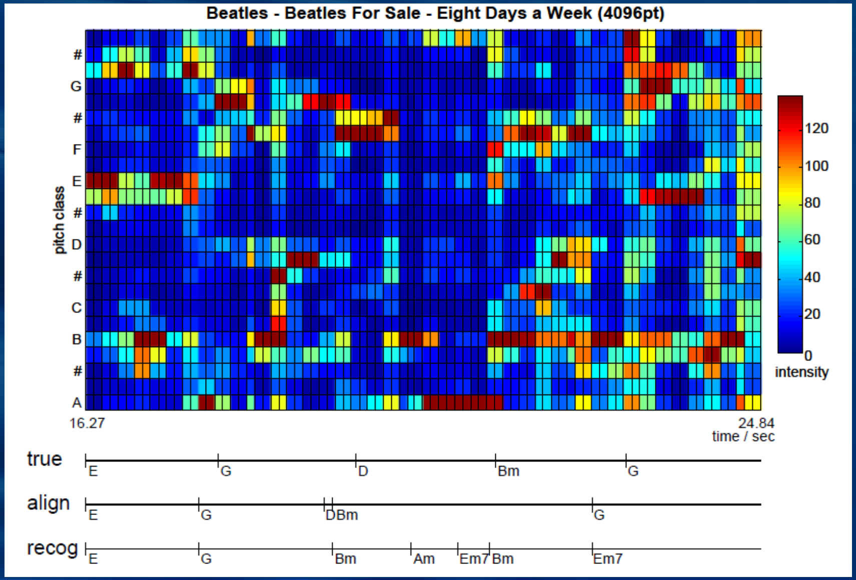
Chord families	maj, min, maj7, min7, dom7, aug, dim
Roots	A♭, B♭, C♭, D♭, E♭, F♭, G♭, A, B, C, D, E, F, G, A♯, B♯, C♯, D♯, E♯, F♯, G♯,
Examples	Amaj, C♯min7, G♭dom7

Chord Segmentation and Recognition

(Sheh and Ellis, Columbia, 2003)

- Test set:
  - 20 songs from three early Beatles albums – mono files at 11025 Hz – 10 PDP frames per second
  - Chord sequences from a standard book of Beatles transcriptions
  - Training: 17 songs / Test: 3 songs
- Improvements:
  - More datas and parameters (Gaussian mixture models)
  - Frequency resolution (minor/major confusion)
  - Adaptive tuning

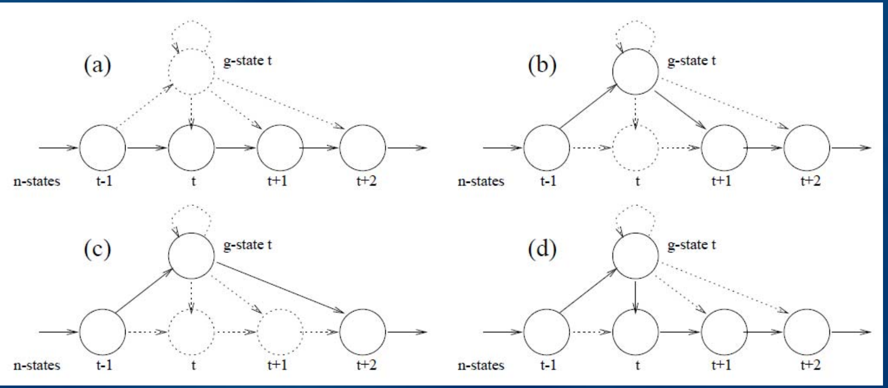
Feature	Align train18	Align train20	Recog train18	Recog train20
MFCC	27.0	20.9	5.9	16.7
	14.5	23.0	7.7	19.6
MFCC_D	24.1	13.1	15.8	7.6
	19.9	19.7	1.5	6.9
MFCC_0_D_A	13.9	11.0	2.2	3.8
	9.2	12.3	1.3	2.5
PCP	26.3	41.0	10.0	23.6
	46.2	53.7	18.2	26.4
PCP_ROT	68.8	68.3	23.3	23.1
	83.3	83.8	20.1	13.1

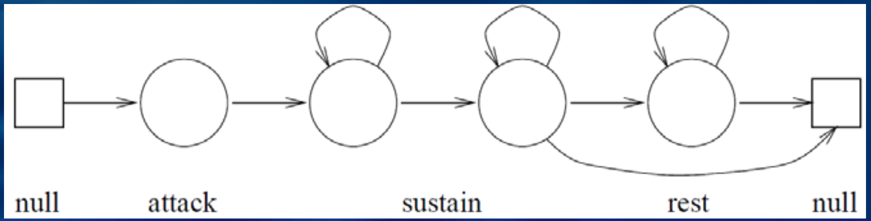


Score Following

(Orio & Déchelle, IRCAM, 2003)

- Events: notes, trills, chords,…
- Two-level left-to-right HMM ➔ Sequential representation
- High-level states: normal and ghost states
  - 1 normal and 1 ghost per event
  - High-level transition types: normal, wrong, extra and skip
- Low-level states: note shape (duration)
- Observable events: audio signal input
- Score following: decoding problem
- Alternative training method to adapt the topology ghost/normal





Melody Spotting

(Durey and Clements, GeorgiaTech, 2001)

- Problem: musical database query by melody (humming,…)
  ➔ similar to speech recognition
- System for raw audio file (wav, aif, mp3,…)
- 5-state left-to-right HMMs for each note (C4-G5) plus rest
- Observation: Pitch/FFT vectors/Scalar vectors
- Process:
  - Transform input in observation sequence
  - Build an HMM associated with the sequence (concatenation + fillings)

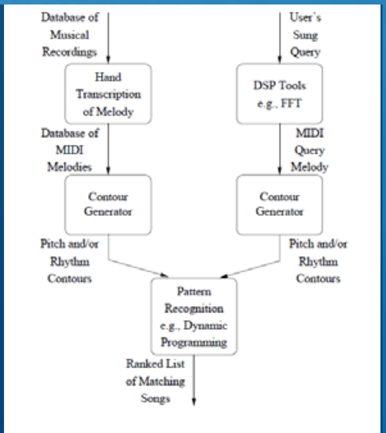
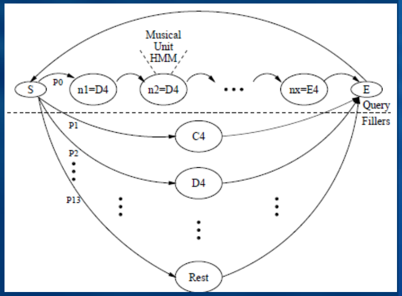
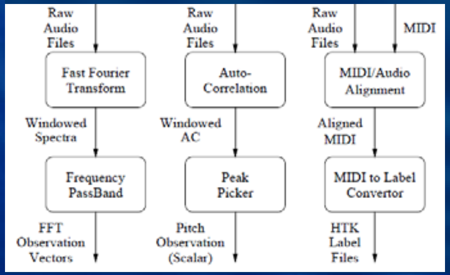
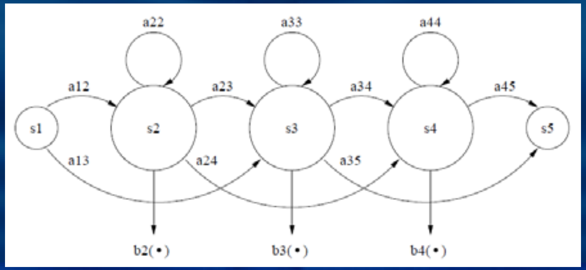
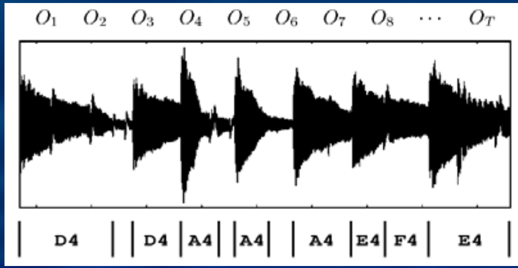


Figure 1: Generalized architecture used for melody-based music information retrieval









Melody Spotting

(Durey and Clements, GeorgiaTech, 2001)

- Result evaluation: numerical figure-of-merit
- Input data: Yamaha W7 keyboard – mono audio at 22050 Hz
- Test with zero-error queries
- Results:
  - Pitch-based: long preprocessing – little information – poor results
  - FFT-based: short preprocessing – large information – good results
  - Scalar-based: short preprocessing – medium information – good results
- Improvements:
  - Different scaling of filler penalties
  - Better result evaluation – include similar matching

Table 1: List of Songs Used in Testing

	Song Title
1	*Auld Lang Syne*
2	*Barbara Allen*
3	*Frere Jacques*
4	*Happy Birthday to You*
5	*I'm a Little Teapot*
6	*Mary Had a Little Lamb*
7	*Scarborough Fair*
8	*This Old Man*
9	*Three Blind Mice*
10	*Twinkle, Twinkle, Little Star*

Table 2: List of Instrument Voices Used in Testing

Instrument Name
Clarinet
Flute
Piano
Soprano Saxophone
Violin

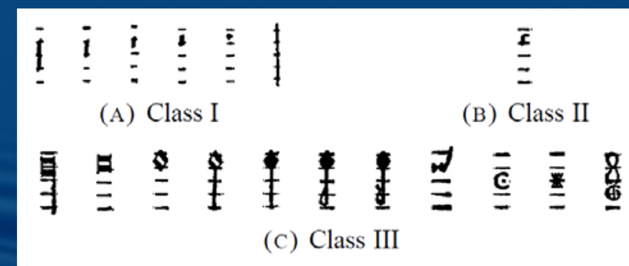
Table 4: Query Results Using Pitch-Based HMM Recognizer

Query	*# Hits*	*# FAs*	*# Actual*	*FOM*	P_n
3a.	70	2	70	95.63	1/500
3b.	19	69	50	0.00	
4.	40	0	40	100.00	1/500
5.	30	10	30	78.80	1/500
6.	40	32	40	0.00	1/50
7a.	30	19	30	42.80	1/250
7b.	8	5	20	28.00	
8a.	30	10	30	78.60	1/500
8b.	20	0	20	100.00	
8c.	18	24	20	27.50	
9a.	30	7	30	93.40	1/500
9b.	20	0	20	100.00	
9c.	8	7	20	23.20	
10a.	20	30	20	23.80	1/500
10b.	8	8	20	20.80	

Optical Music Recognition

(Pugin, McGill, 2006)

- Data set: Early music prints (16th and 17th centuries)
- No staff removal ➔ ubiquitous and complicated operation (irregular staff lines)
- Training data: 240 pages, 52178 characters
- Segmentation-free approach
- Feature extraction with a sliding window
- Observation:
  - number of connected black zones
  - black pixels repartition (gravity centers)
  - Area of the largest and smallest black element
  - Total area of black with weighting mask
- Left-right HMM
  - Number of states ➔ close to the symbol width (handwriting recognition)
  - Three topological classes
- Silence detection (speech processing)



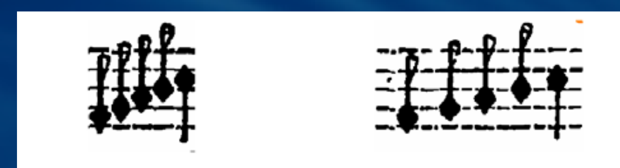


Figure 4. Space between symbols

Table 2. Recognition rates

	F_{MIX}	W_{FS}	W_{MF}
REC	96.82	97.16	95.77
MUS	97.11	97.42	96.22

Contents

- Markov Processes
- Hidden Markov Models
- Applications in music information retrieval
  - Folk music classification
  - Chord segmentation and recognition
  - Score following
  - Melody spotting
  - Optical music recognition
- **Other applications**

Other Applications

- Recognition
  - Phone (decoding phones from audio input)
  - Speech (decoding words from phones)
  - Language (decoding language from words/phones)
  - Gesture
- Classifications (Text types,…)
- Medical domain:
  - DNA sequences
  - Proteins
- Signal processing (exploit statistical dependencies in real signals)
- Fault-tolerance modeling

Conclusion