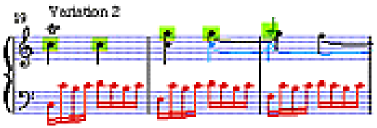


ISMIR 2003 Oct. 27th – 30th 2003
Baltimore (USA)

Automatic Labelling of tabla signals

Olivier K. GILLET , Gaël RICHARD

TELECOM PARIS
école nationale supérieure des télécommunications

get

Introduction

- **Exponential growth of available digital information**
  ⇨ need for Indexing and Retrieval technique

- **For musical signals, a transcription would include:**
  - Descriptors such as genre, style, instruments of a piece
  - Descriptors such as beat, note, chords, nuances, etc…

  – Many efforts in instrument recognition (Kaminskyj2001, Martin 1999, Marques & al. 1999 Brown 1999, Brown & al.2001, Herrera & al.2000, Eronen2001)

  – Less efforts in percussive instrument recognition (Herrera & al. 2003, Paulus&al.2003, McDonald&al.1997)

  – Most effort on isolated sounds

  – Almost no effort on non-Western instrument recognition

- **OBJECTIVE :Automatic transcription of real performances of an Indian instrument: the tabla**

Outline

- Introduction
- **Presentation of the tabla**
- **Transcription of tabla phrases**
  - Architecture of the system
  - Features extraction
  - Learning and classification
- **Experimental results**
  - Database and evaluation protocols
  - Results
- **Tablascope: a fully integrated environment**
  - Description & applications
  - Demonstration
- **Conclusion**

Presentation of the tabla

- **The tabla:** an percussive instrument played in Indian classical and semi-classical music

The Dayan: wooden treble drum played by the right hand

The Bayan: metallic bass drum played by the left hand

Presentation of the tabla (2)

- **Musical tradition in India is mostly oral**
  - ➡ Use of mnemonic syllables (or ***bol***) for each stroke

- **Common bols:**
  - – **Ge**, **Ke** (bayan bols), **Na**, **Tin**, **Tun**, **Ti**, **Te** (dayan bols)
  - – **Dha** (Na+Ge), **Dhin** (Tin + Ge), **Dhun** (Tun + Ge)

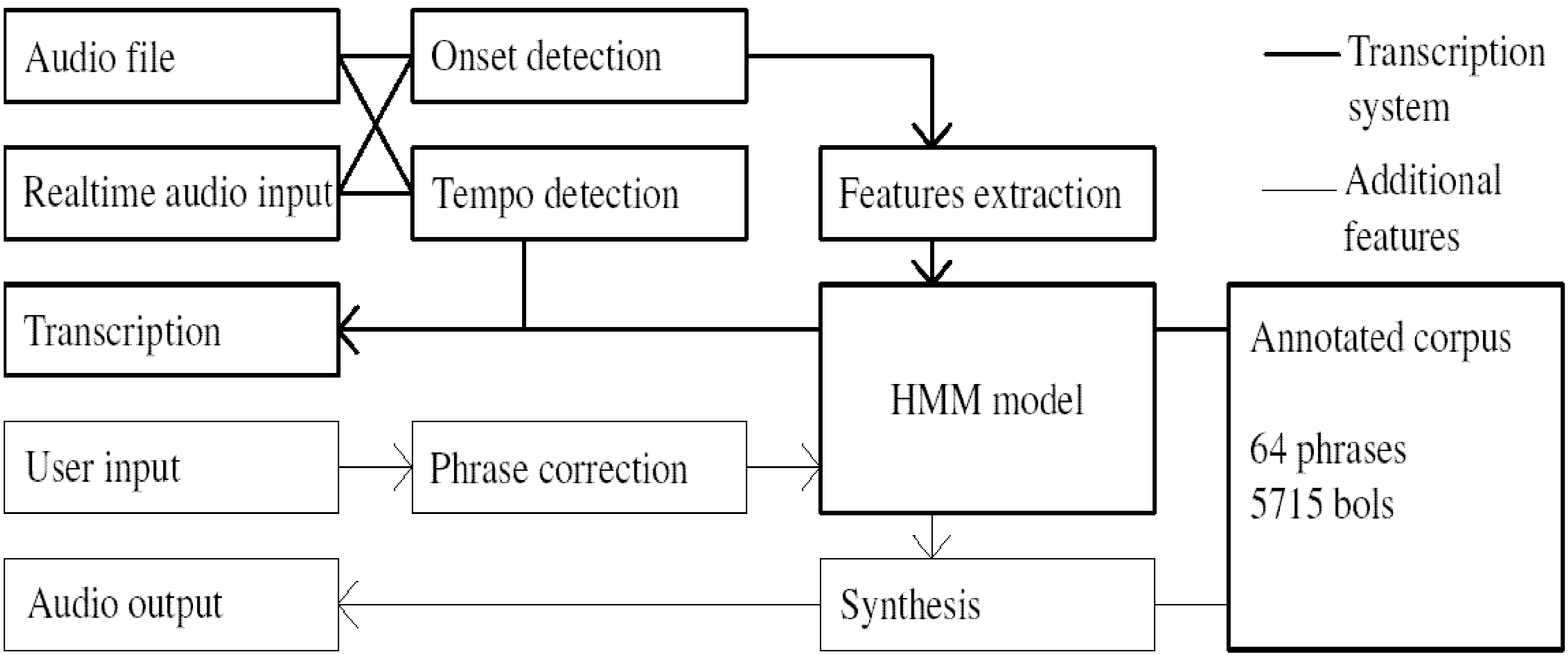
- **Some specificities of this notation system**
  - – Different bols may sound very similar (ex. Ti and Te)
  - – Existence of « words » : « TiReKiTe or « GeReNaGe »
  - – A mnemonic may change depending on the context
  - – Complex rythmic structure based on *Matra* (i.e main beat), *Vibhag* (i.e measure) and *avartan* (i.e phrase)

Presentation of tabla (3)

- **In summary:**
  - A tabla phrase is then composed of successive bols of different duration (note, half note, quarter note) embeded in a rythmic structure
  - Grouping characteristics (words) : similarity with spoken and written languages: Interest of « Language models » or sequence models
- **In this study, the transcription is limited to**
  - the recognition of successives bols
  - The relative duration (note, half note, quarter note) of each bol.

Transcription of tabla phrases

- **Architecture of the system**



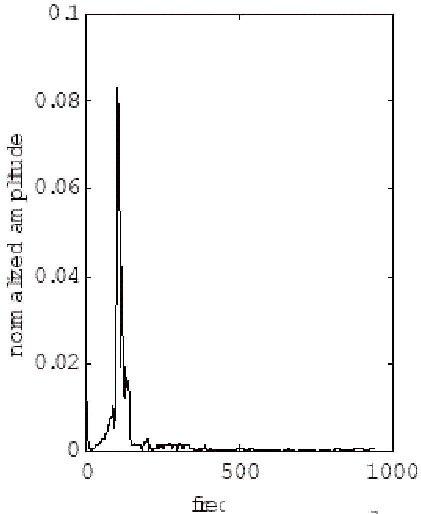
Parametric representation

- **Segmentation in strokes**
  – Extraction of a low frequency envelope (sampled at 220.5 Hz)
  – Simple Onset detection based on the difference between two successives samples of the envelope.

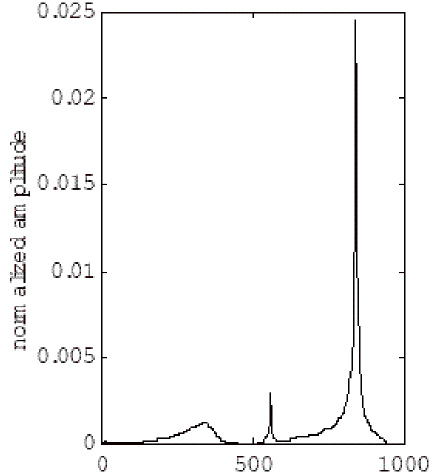
- **Tempo extraction**
  – Estimated as the maximum of the autocorrelation function of the envelope signal in the range {60 – 240 bpm}

Features extraction

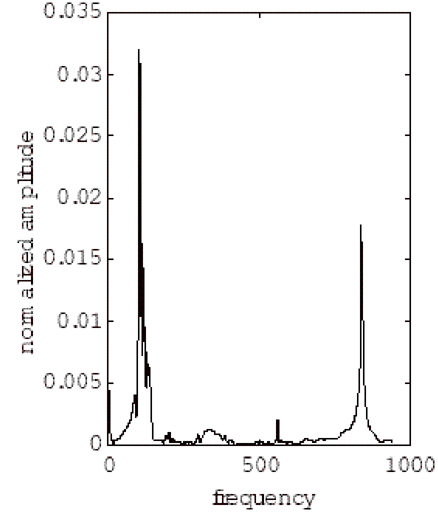
Ge



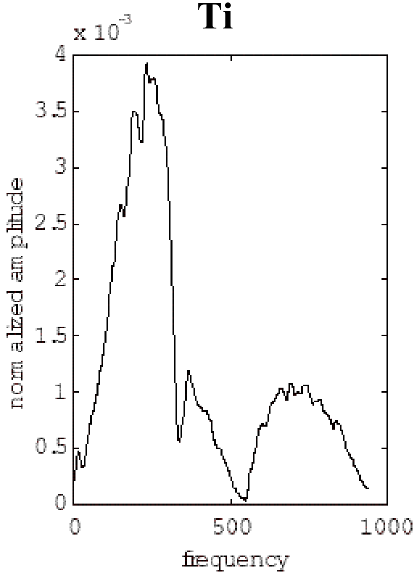
Na



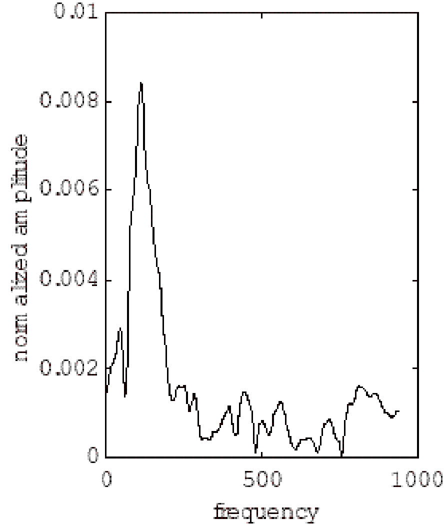
Dha = Ge + Na



Ti



Ke



Features extraction

- **4 frequency bands**
  - B1 = [0 –150] Hz
  - B2 = [150 – 220] Hz
  - B3 = [220 – 380] Hz
  - B4 = [700 – 900] Hz

- **In the case of single mixture, each band is modelled by a Gaussian**

➡ **Feature vector $F = f_1..f_{12}$ (mean, variance and relative weight of each of the 4 Gaussians)**

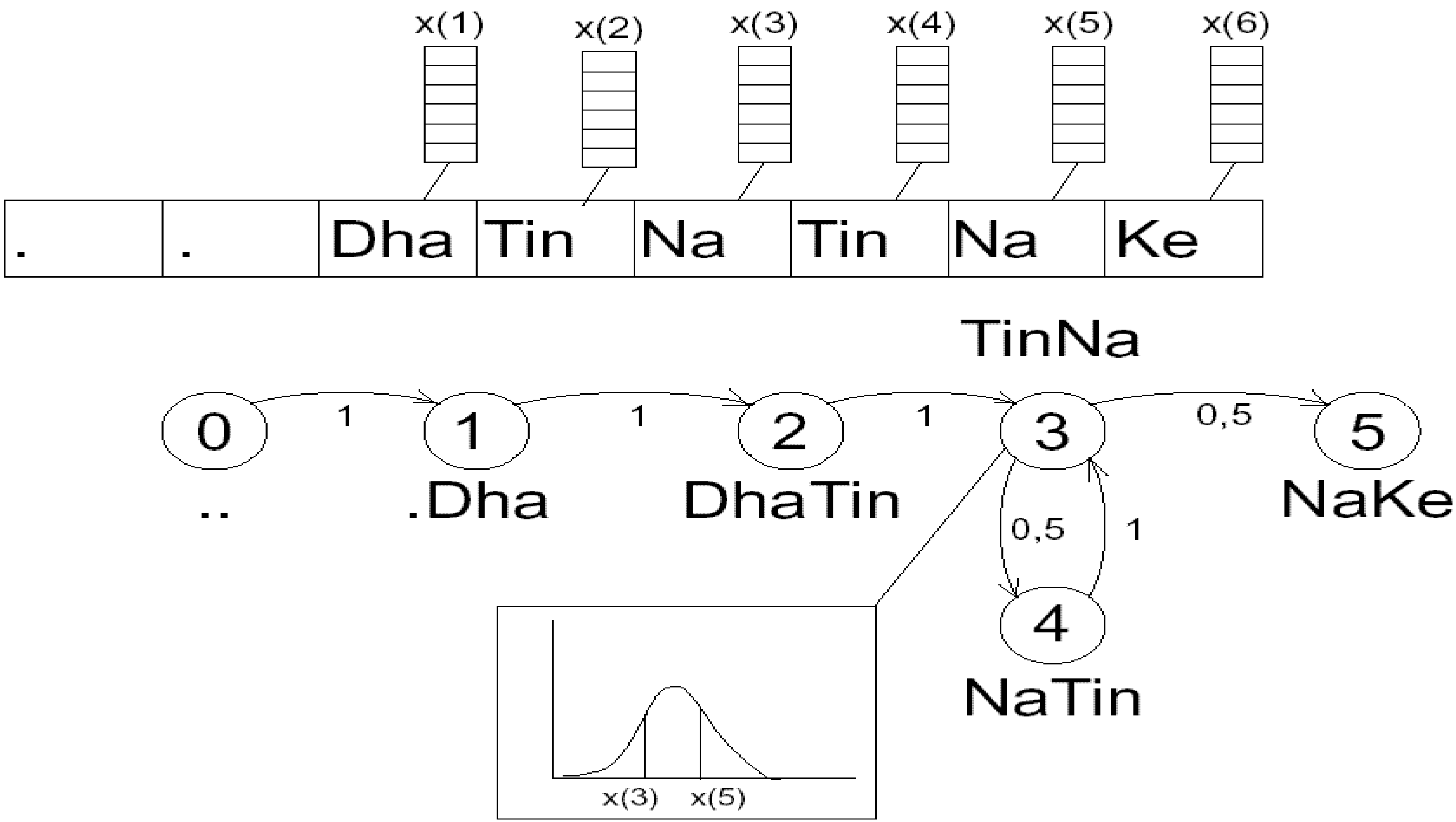
Learning and Classification of bols

- **4 classification techniques were used.**
  - K-nearest Neighbors (k-NN)
  - Naive Bayes
  - Kernel density estimator

  - **HMM sequence modelling**

Learning and Classification of bols

- **Context-dependant models (HMM)**



Learning and Classification of bols

- **Hidden Markov Models**

  - **States:** a couple of Bols B_1B_2 is associated to each state q_t
  - **Transitions:** if state *i* is labelled by B_1B_2 *and j by* B_2B_3 then the transition from state $q_t = j$ to state $q_{t-1} = i$ is given by:

$$\begin{aligned} a_{ij} &= p(q_t = j | q_{t-1} = i) \\ &= p(b_t = B_3 | b_{t-1} = B_2, b_{t-2} = B_1) \end{aligned}$$

  - **Emissions probabilities:** Each state i labelled by B_1B_2 emits a feature vector according to a distribution $b_i(x)$ characteristics of the bol B_2 preceded by B_1

$$\begin{aligned} b_i(x) &= p(O_t = x | q_t = i) \\ &= p(O_t = x | b_t = B_2, b_{t-1} = B_1) \end{aligned}$$

Learning and Classification of bols

- **Training**
  - Transition probabilities are estimated by counting occurrences in the training database
  - Emission probabilities are estimated with
    - mean and variance estimators on the set of feature vectors in the case of simple Gaussian model
    - 8 iterations of the Expectation-Maximisation (EM) algorithm in the case of a mixture model
- **Recognition**
  - Performed using the traditionnal Viterbi algorithm

Experimental results

- **Database**
  - 64 phrases with a total of 5715 bols
  - A mix of long compositions with themes / variations (*kaïda*), shorter pieces (*kudra*) and basic *taals*.
  - **3 specific sets corresponding to three different tablas:**

	Tabla quality	Dayan tuning	Recording quality
Tabla #1	Low (cheap)	in C#3	Studio equipment
Tabla #2	High	In D3	Studio equiment
Tabla #3	High	In D3	Noisier environment

Evaluation protocols

- **Protocol #1:**
  - Cross-validation procedure
    - Database split in10 subsets (randomly selected)
    - 9 subsets for training, 1 subset for testing
    - Iteration by rotating the 10 subsets
    - Results are average of the 10 runs
- **Protocol #2:**
  - Training database consists in 100% of 2 sets
  - Test is 100% of the remining sets

➡ Different instruments and/or conditions are used for training and testing

Experimental results (protocol #1)

Database # of *bols*	All 5715	Tabla #1 1678	Tabla #2 2216	Tabla #3 1821
Classification using only features of stroke n				
Kernel density estimator	81.7%	81.8%	82.4%	85.2%
5-NN	83.0%	81.7%	83.3%	85.6%
Naive Bayes	76.6%	79.4%	78.6%	78.5%
Classification using features of stroke $n, n-1, n-2$				
Kernel density estimator	86.8%	86.0%	88.7%	92.0%
5-NN	88.9%	87.2%	88.4%	90.6%
Naive Bayes	81.8%	86.5%	83.8%	85.8%
Classification using language modelling				
HMM, 3-grams, 1 mixture	88.0%	90.6%	89.9%	92.6%
HMM, 4-grams, 2 mixtures	**93.6%**	**92.0%**	**91.9%**	**93.4%**

Experimental results (protocol #2)

Training set Test set	Tabla #1 & Tabla #2 Tabla #3 (noisy rec.)	Tabla #2 & Tabla #3 Tabla #1 (cheap quality)
5-NN	79.8 %	78.2 %
HMM, 3-grams, 1 mixture	**90.2 %**	**88.4 %**
HMM, 4-grams, 2 mixtures	84.5 %	85.0 %

- **HMM approaches are more robust to variability**
- **Simpler classifiers fail to generalise and to adapt to different recording conditions or instruments**

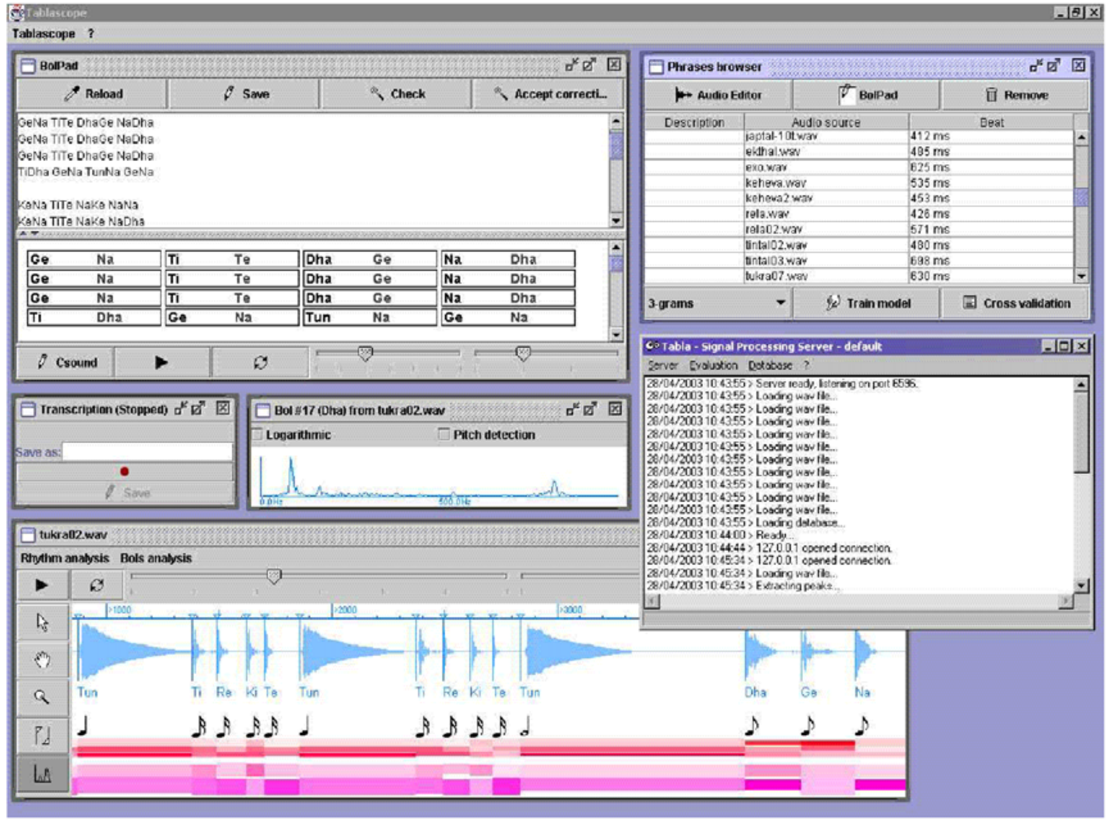
Experimental results

- **Confusion matrix by bol category**
  (*HMM 4-grams, 2 mixture classifier*)

a	b	c	d	e	¡- classified as
1241	22	2	7	8	a: resonant *dayan* strokes (Tin, Na, Tun...)
20	1076	2	1	5	b: *bayan+dayan* strokes (Dhin, Dha...)
1	3	766	5	20	c: resonant *bayan* strokes (Ge, Gi...)
8	2	2	448	61	d: non-resonant *bayan* strokes (Ke, Ki...)
11	7	6	50	1938	e: non-resonant *dayan* strokes (Te, Ti, Tek...)

Tablascope: a fully integrated environment

- **Applications:**
  - Tabla transcription
  - Tabla sequence synthesis
  - Tabla-controlled synthesizer



Conclusion

- A system for automatic labelling of tabla signals was presented
- Low error rate for transcription (6.5%)
- Several applications were integrated in a friendly environment called Tablascope.
- This work can be generalised to other types of percussive instruments

- …still need a larger database to confirm the results…..