On Musical Performances Identification, Entropy and String Matching

Antonio Camarena-Ibarrola and Edgar Chavez

{camarena,elchavez}@umich.mx

Universidad Michoacana de Sán Nicolás de Hidalgo

Morelia, Michoacán. México

Matching performances

- Are they playing the same Music?
- This problem is also known as *Polyphonic Audio Matching*

Why is it important?

- Radio Broadcast monitoring
- Querying by example
- Querying by humming
- Score-performance following
- Filtering in p2p networks

Related Work

Front End	Aligning	Proposed By	Year
Energy Fundamental Freq ZCR	HMM	Pedro Cano et al	1999
MFCC	DTW	Tzanetakis et al	2003
Spectral envelope Amplitud envelope	Linear DTW	Dixon et al	2005

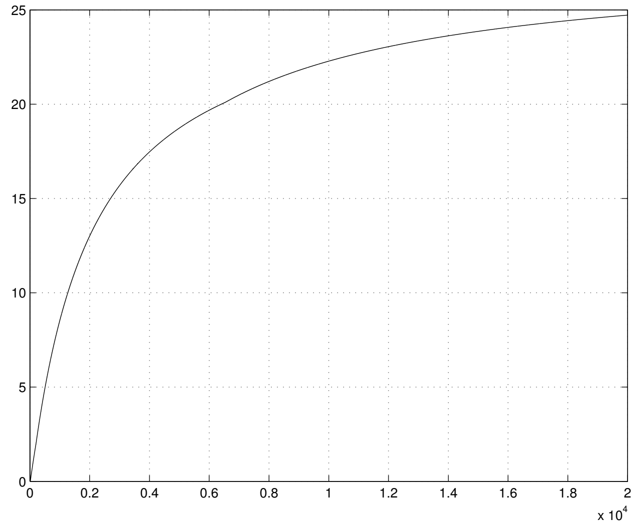
What does the brain perceive?

- Information?
Aoccdrnig to a rsecheearr at an Elingsh uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is that frist and lsat ltteer is at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit a porbelm.
- Entropy H is the expected information content in a sequence.

$$H(x) = E[I(p)] = \sum_{i=1}^{n} p_i I(p) = -\sum_{i=1}^{n} p_i ln(p_i)$$

Information in the perspective of the ear

- Not all frequencies can be heard with the same sensitivity.

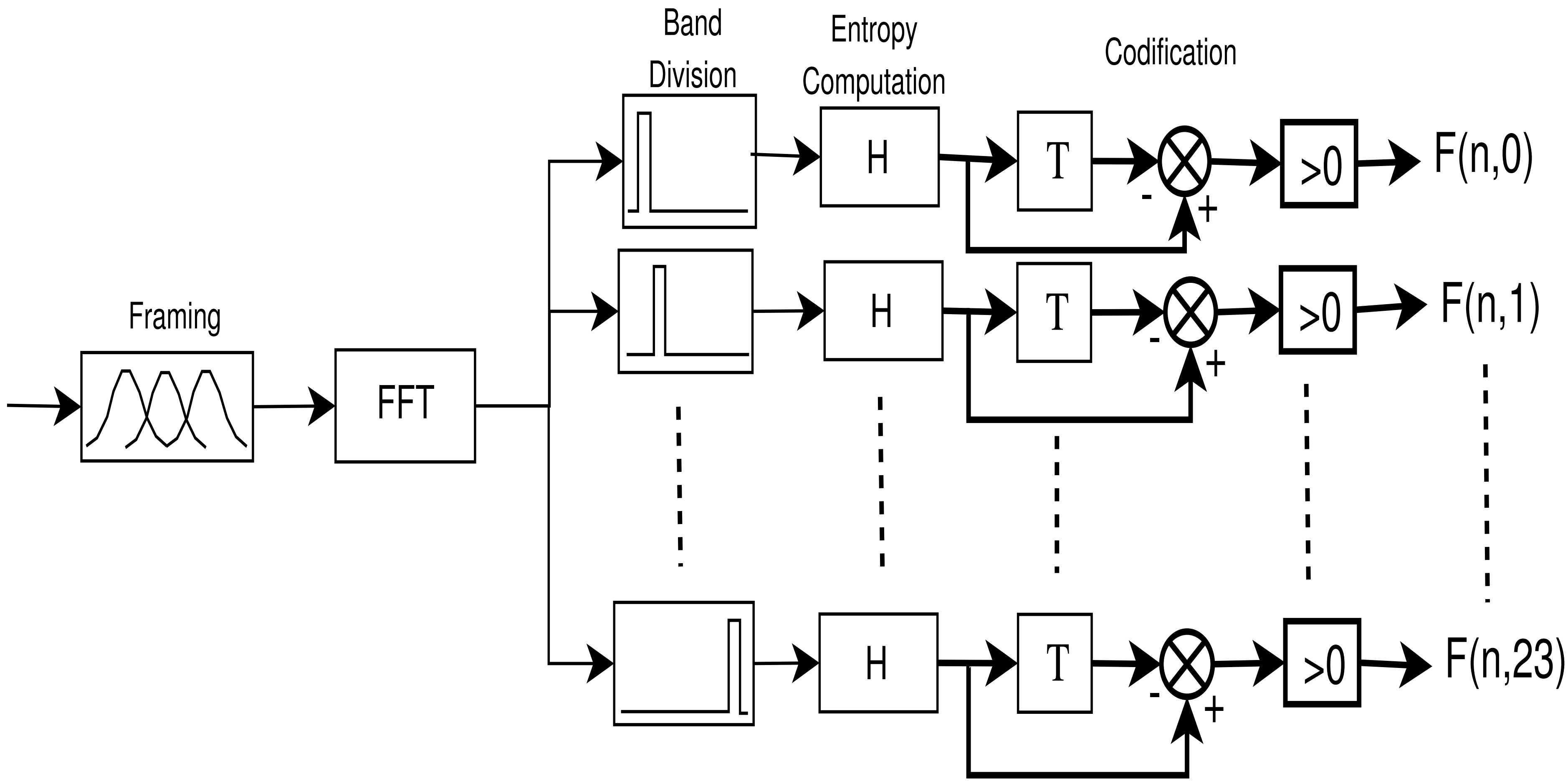


$$z = \frac{16.81f}{1960+f} - 0.53$$

$$z' = z + 0.22(z - 20.1) \quad \forall \quad z > 20.1$$

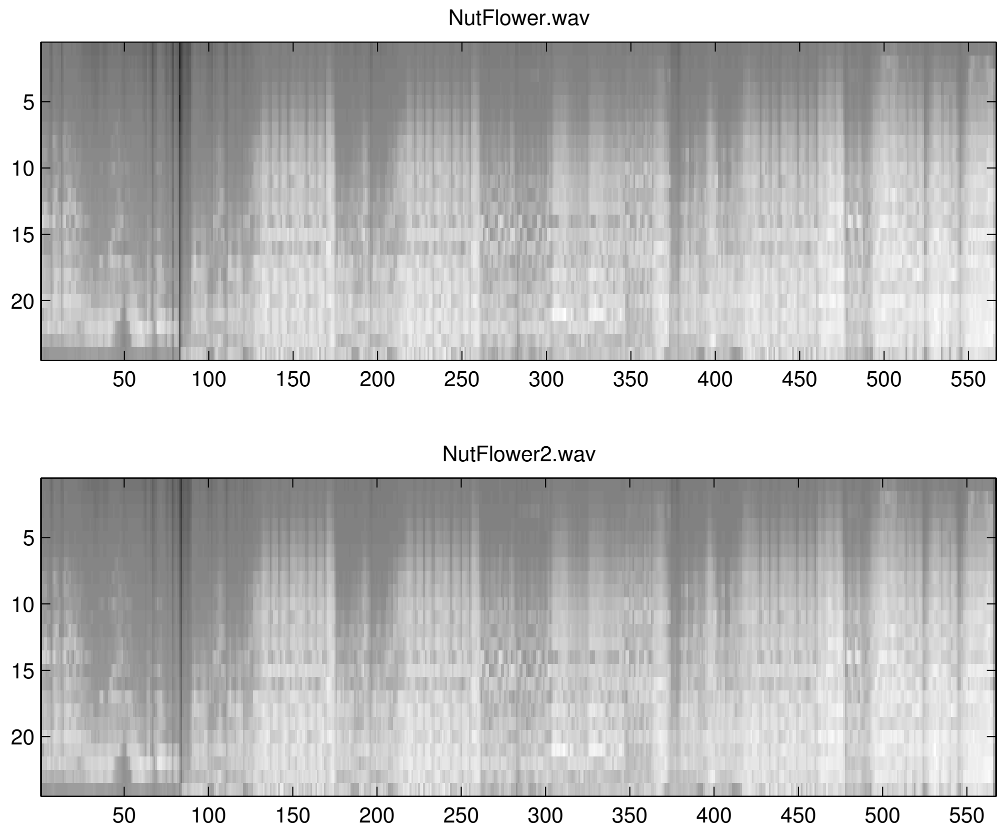
$$z' = z + 0.15(2 - z) \quad \forall \quad z < 2$$

The process of obtaining the AFP



Entropygrams

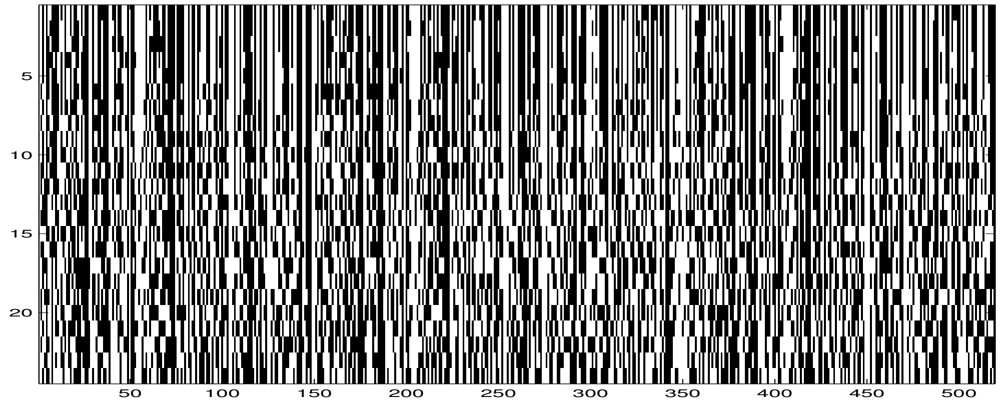
The information level for every band and frame

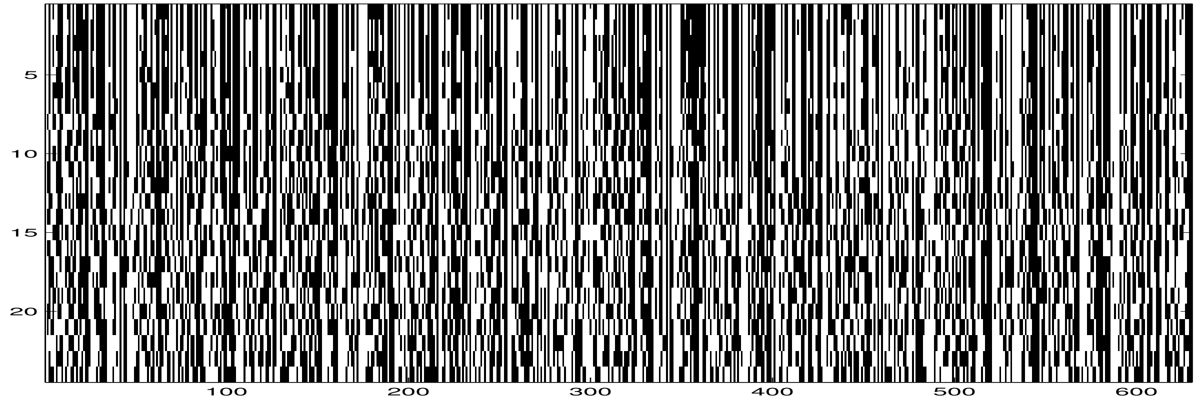


Tchaikovsky's Nutcracker waltz of the flower

Low Res. spectral entropy string AFP

Aligning technique required!



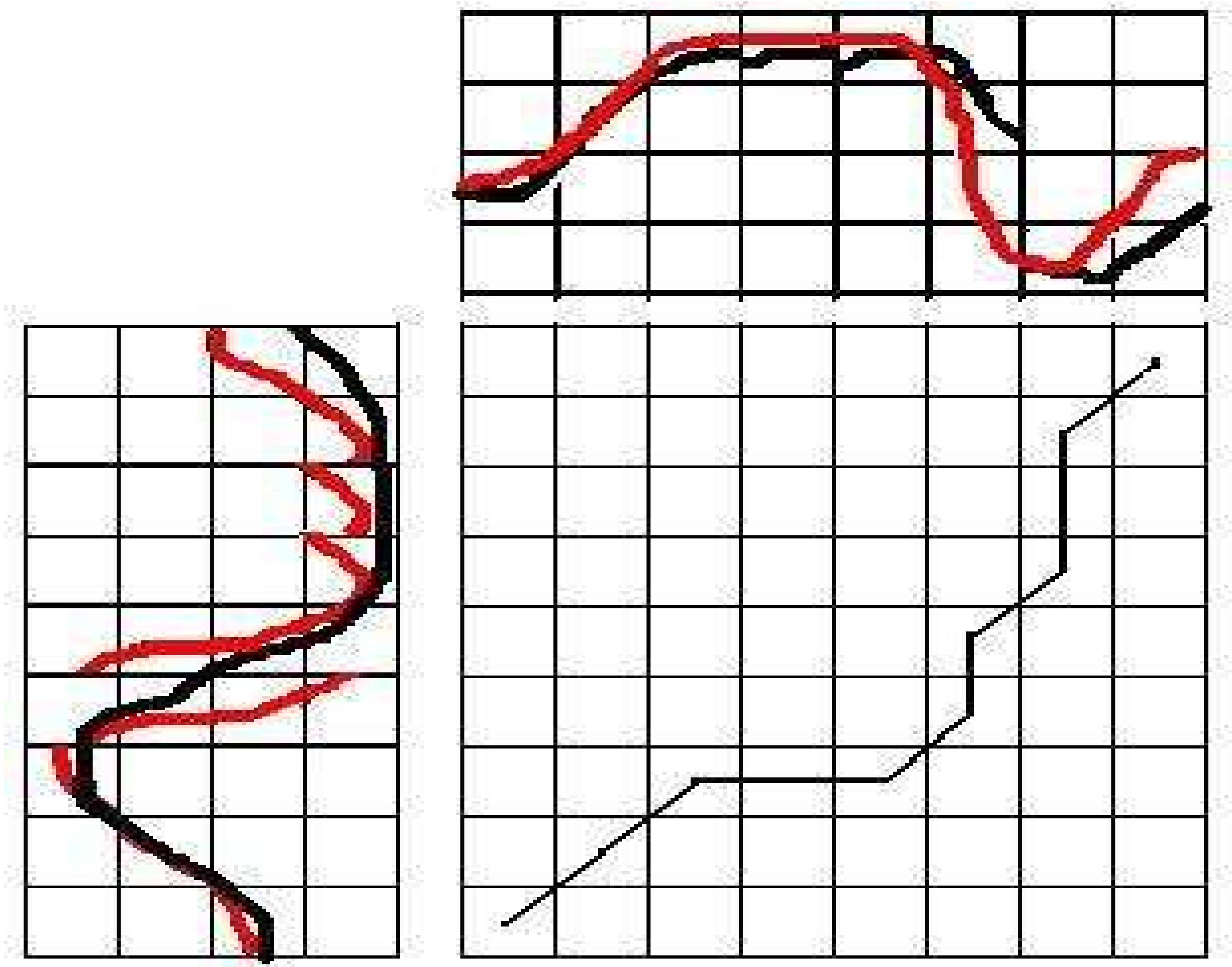


Mozart’s Serenade Number 13 Allegro

Aligning techniques used

- Classical Approach. Dynamic Time Warping (DTW)
- Flexible string matching techniques. Succesfull in text retrieval and computational biology (Matching DNA sequences). Levenshtein Distance and Longest Common Subsequence Distance (LCS)

Dynamic Time Warping



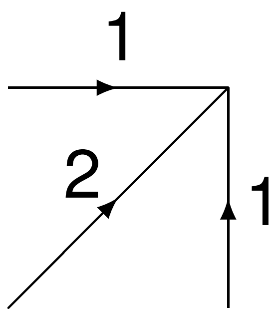
Dynamic Time Warping

$$D_{i,0} = \sum_{k=0}^{i} d_{i,0}$$

$$D_{0,j} = \sum_{k=0}^{j} d_{0,j}$$

$$D_{i,j} = min \begin{cases} D_{i-1,j-1} + 2d_{i,j} \\ D_{i-1,j} + d_{i,j} \\ D_{i,j-1} + d_{i,j} \end{cases}$$

$d_{i,j}$ is the Hamming distance between row i of one performance's AFP and row j of the other performance's AFP.



Levenshtein distance

$$C_{i,0} = i \qquad \forall \quad 0 \le i \le N$$
$$C_{0,j} = j \qquad \forall \quad 0 \le j \le M$$

$$C_{i,j} = \begin{cases} C_{i-1,j-1} & t_i = p_j \\ min[C_{i-1,j-1}+1, C_{i,j-1}+1, C_{i-1,j}+1] & t_i \neq p_j \end{cases}$$

		y	e	l	l	o	w
	0	1	2	3	4	5	6
h	1	**1**	2	3	4	5	6
e	2	2	**1**	2	3	4	5
l	3	3	2	**1**	2	3	4
l	4	4	3	2	**1**	2	3
o	5	5	4	3	2	**1**	**2**

(i.e replace the “h” by a “y” and insert a “w” at the end)

Levenshtein distance in one column

$$C_i^{'} = \begin{cases} C_{i-1} & t_i = p_j \\ min[C_{i-1} + c_s, C_{i-1}^{'} + 1, C_i + 1] & t_i \neq p_j \end{cases}$$

To monitor occurrences of a pattern in a text set $C_0^{'} = 0$

		A	T	C	A	T	T
	0	0	0	0	0	0	0
A	1	0	1	1	0	1	1
T	2	1	0	2	1	0	1
T	3	2	**1**	**1**	2	**1**	**0**

As substitution cost (c_s) we use $Hamming(t_i, p_j)/24$ since t_i and p_j are made from 24 bits and we want it to be a value between 0 and 1.

Longest Common Subsequence distance

		Levenshtein	LCS
computer	courtain	6	10
computer	cooommpuuteeer	6	6

$$C_{i,0} = i \qquad \forall \quad 0 \leq i \leq N$$
$$C_{0,j} = j \qquad \forall \quad 0 \leq j \leq M$$
$$C_{i,j} = \begin{cases} C_{i-1,j-1} & t_i = p_j \\ min[C_{i,j-1}, C_{i-1,j}] + 1 & t_i \neq p_j \end{cases}$$

LCS computation in one column

$$C_i^{'} = \begin{cases} C_{i-1} & t_i = p_j \\ min[C_{i-1}^{'}, C_i] + 1 & t_i \neq p_j \end{cases}$$

We consider two symbols as different only if the Hamming distance between two frames is greater than 7.

The Test Set

Name	Dur1	Dur2
All my loving	2:09	2:08
All you need is love	3:49	3:46
Come together	4:18	4:16
Eleanor Rigby	2:04	2:08
Here comes the sun	3:07	3:04
Lucy in the sky	3:27	3:27
Nowhere man	2:40	2:44
The Nutcracker flowers	7:09	7:06
The Nutcracker Reeds	2:39	2:41
Octopus's garden	2:52	2:48

Name	Dur1	Dur2
Mozart's Ser 13 Menuetto	2:19	2:10
Mozart's Ser 13 Allegro	6:30	7:52
Mozart's Ser 13 Romance	6:45	5:47
Mozart's Ser 13 Rondo	3:24	2:55
Sgt Pepper's Lonely hearts	2:01	2:01
Something	3:02	2:59
Swan Lake theme	3:14	3:13
Symphony 41	8:50	8:55
With a little help	2:44	2:43
Yellow submarine	2:37	2:35

Using DTW

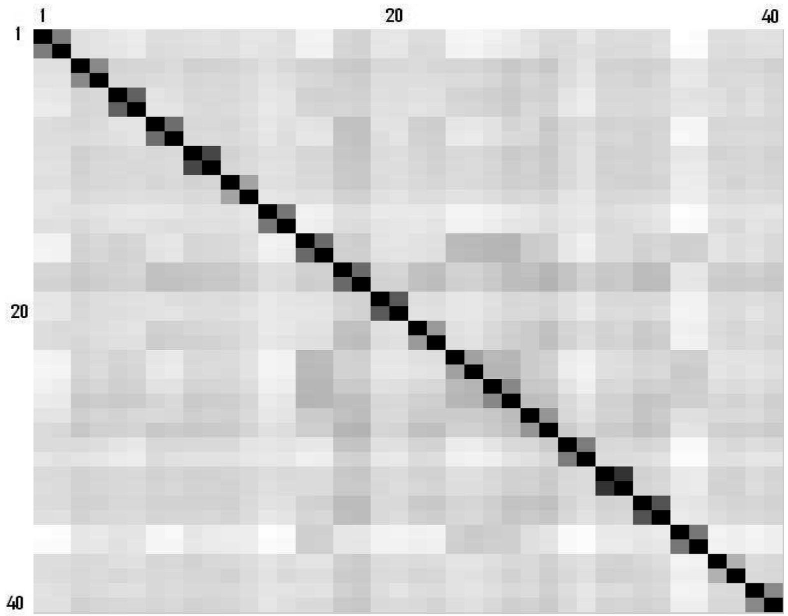


Figure 1: Confusion Matrix result from using DTW

Using LCS

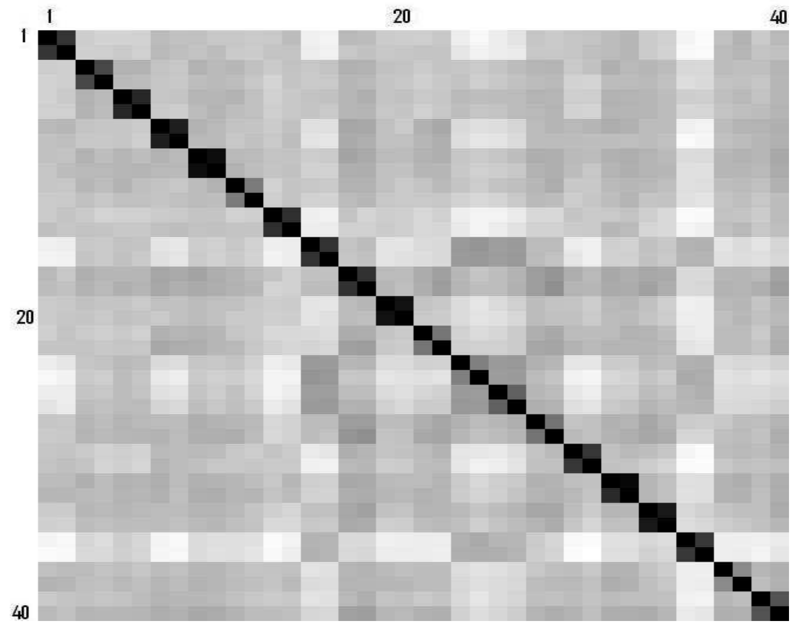


Figure 2: Confusion Matrix result from using LCS distance

Using Levenshtein

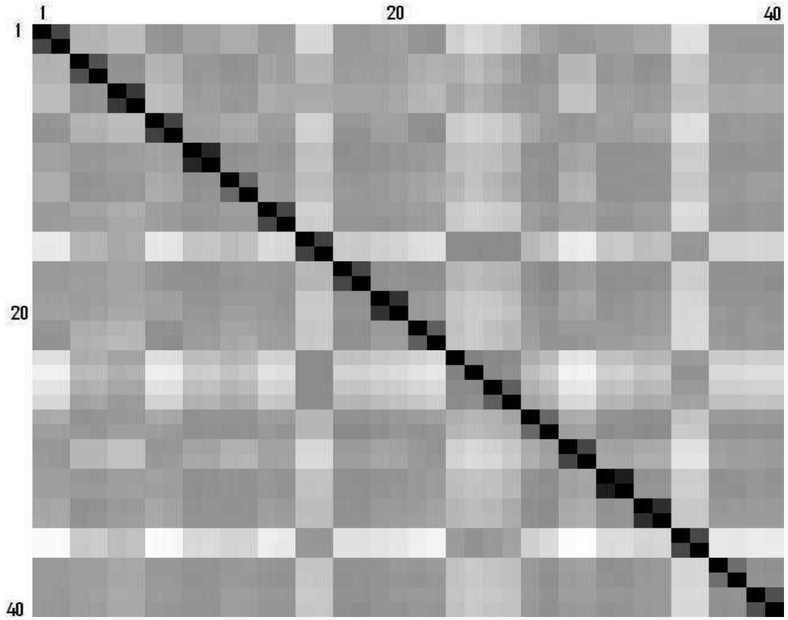


Figure 3: Confusion Matrix result from using Levenshtein distance

Conclusions

- The spectral entropy AFP is a feature that is very usefull in identifying musical renditions
- LCS and Levenshtein are as effective as DTW as distance measures for matching performances
- LCS and Levenshtein are more flexible than DTW (can be used in radio broadcast monitoring)
- LCS and Levenshtein do not require training as HMM do

Thanks!!!

Questions?

{camarena,elchavez}@umich.mx