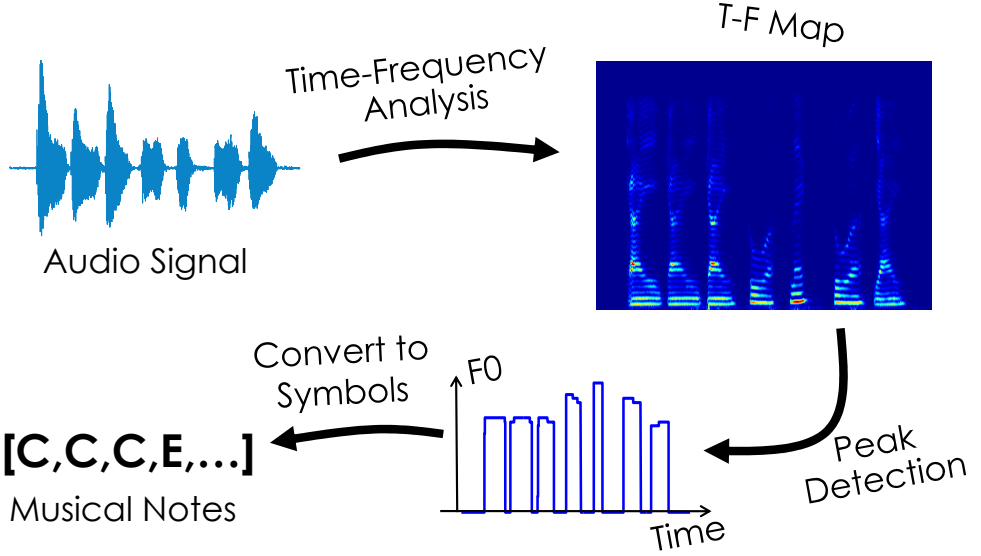
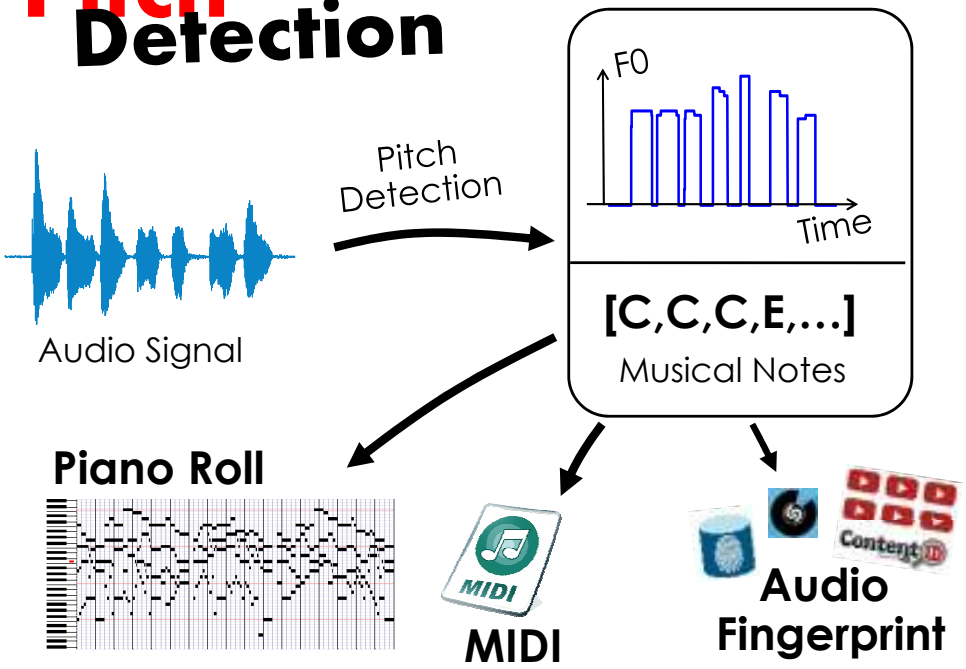
Audio & Speech Technology

[2] Melody Extraction and Matching

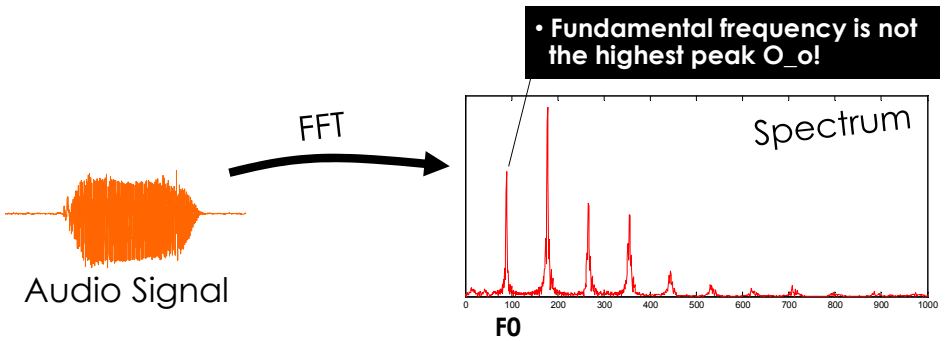
Pitch Detection



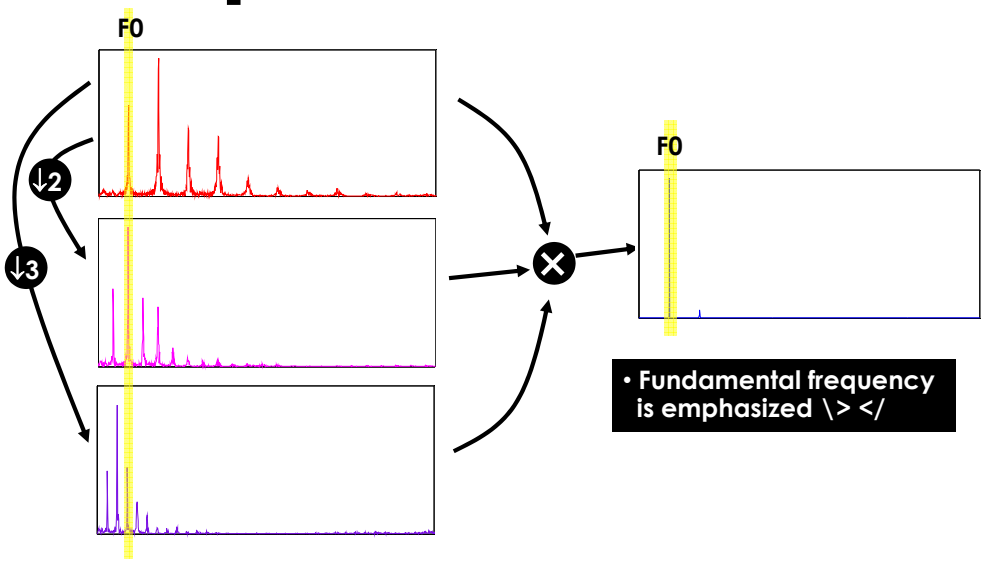
Pitch Detection



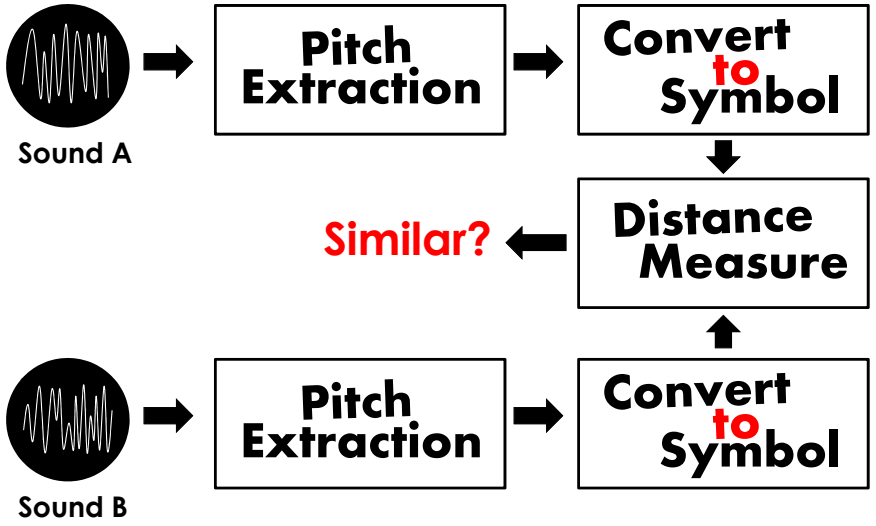
Harmonic Product Spectrum



Harmonic Product Spectrum



Melody Matching

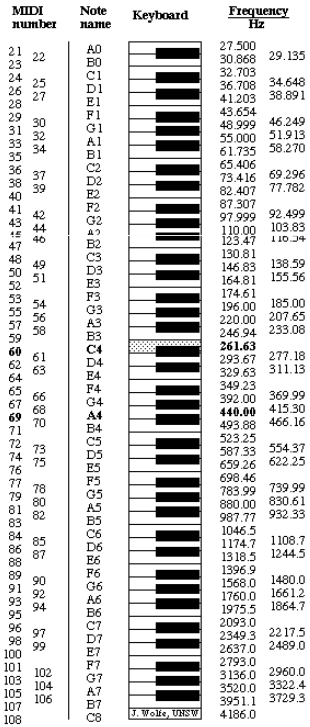


Musical Note

Note	Freq [Hz]	Note	Freq [Hz]	Note	Freq [Hz]	Note	Freq [Hz]
C_0	16.35	C_1	32.7	C_2	65.41	C_3	130.81
$C^{\#}_0/D^{b}_0$	17.32	$C^{\#}_1/D^{b}_1$	34.65	$C^{\#}_2/D^{b}_2$	69.3	$C^{\#}_3/D^{b}_3$	138.59
D_0	18.35	D_1	36.71	D_2	73.42	D_3	146.83
$D^{\#}_0/E^{b}_0$	19.45	$D^{\#}_1/E^{b}_1$	38.89	$D^{\#}_2/E^{b}_2$	77.78	$D^{\#}_3/E^{b}_3$	155.56
E_0	20.6	E_1	41.2	E_2	82.41	E_3	164.81
F_0	21.83	F_1	43.65	F_2	87.31	F_3	174.61
$F^{\#}_0/G^{b}_0$	23.12	$F^{\#}_1/G^{b}_1$	46.25	$F^{\#}_2/G^{b}_2$	92.5	$F^{\#}_3/G^{b}_3$	185
G_0	24.5	G_1	49	G_2	98	G_3	196
$G^{\#}_0/A^{b}_0$	25.96	$G^{\#}_1/A^{b}_1$	51.91	$G^{\#}_2/A^{b}_2$	103.83	$G^{\#}_3/A^{b}_3$	207.65
A_0	27.5	A_1	55	A_2	110	A_3	220
$A^{\#}_0/B^{b}_0$	29.14	$A^{\#}_1/B^{b}_1$	58.27	$A^{\#}_2/B^{b}_2$	116.54	$A^{\#}_3/B^{b}_3$	233.08
B_0	30.87	B_1	61.74	B_2	123.47	B_3	246.94

Musical Note

Note	Freq [Hz]	Note	Freq [Hz]	Note	Freq [Hz]	Note	Freq [Hz]
C_4	261.63	C_5	523.25	C_6	1046.5	C_7	2093
$C^{\#}_4/D^{b}_4$	277.18	$C^{\#}_5/D^{b}_5$	554.37	$C^{\#}_6/D^{b}_6$	1108.73	$C^{\#}_7/D^{b}_7$	2217.46
D_4	293.66	D_5	587.33	D_6	1174.66	D_7	2349.32
$D^{\#}_4/E^{b}_4$	311.13	$D^{\#}_5/E^{b}_5$	622.25	$D^{\#}_6/E^{b}_6$	1244.51	$D^{\#}_7/E^{b}_7$	2489.02
E_4	329.63	E_5	659.26	E_6	1318.51	E_7	2637.02
F_4	349.23	F_5	698.46	F_6	1396.91	F_7	2793.83
$F^{\#}_4/G^{b}_4$	369.99	$F^{\#}_5/G^{b}_5$	739.99	$F^{\#}_6/G^{b}_6$	1479.98	$F^{\#}_7/G^{b}_7$	2959.96
G_4	392	G_5	783.99	G_6	1567.98	G_7	3135.96
$G^{\#}_4/A^{b}_4$	415.3	$G^{\#}_5/A^{b}_5$	830.61	$G^{\#}_6/A^{b}_6$	1661.22	$G^{\#}_7/A^{b}_7$	3322.44
A_4	440	A_5	880	A_6	1760	A_7	3520
$A^{\#}_4/B^{b}_4$	466.16	$A^{\#}_5/B^{b}_5$	932.33	$A^{\#}_6/B^{b}_6$	1864.66	$A^{\#}_7/B^{b}_7$	3729.31
B_4	493.88	B_5	987.77	B_6	1975.53	B_7	3951.07



MIDI Note Number

$$p = 69 + 12\log_2 \frac{f}{440}$$

http://www.phys.unsw.edu.au/jw/notes.html

MIDI File Format

- MIDI = Musical Instrument Digital Interface

HEADER CHUNK

TRACK CHUNK

⋮

TRACK CHUNK

MIDI File Format

HEADER CHUNK

4D 54 68 64	00 00 00 06	ff ff	nn nn	dd dd

- 4D 54 68 64 = Start of MIDI File
- 00 00 00 06 = Header Length (in bytes)
- ff ff = Format (single-track, multiple tracks)
- nn nn = Number of track chunks
- dd dd = Division (delta timing)

MIDI File Format

4 Track chunks after header chunks

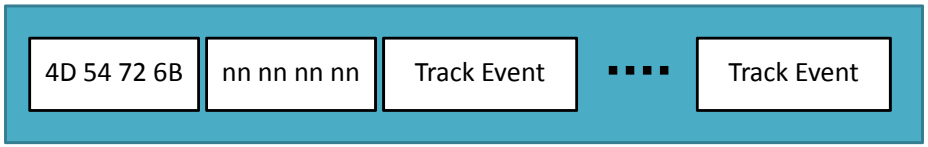
Format = 1 (multiple track)

208 Pulses per Quarter Note

```
4D 54 68 64 00 00 00 06 00 01 00 04 00 C0 4D 54
72 6B 00 00 00 19 00 FF 51 03 05 B8 D8 00 FF 58
04 06 03 12 08 00 FF 59 02 FD 00 00 FF 2F 00 4D
54 72 6B 00 00 24 A5 00 FF 03 05 50 69 61 6E 6F
00 FF 20 01 00 00 B0 07 64 00 C0 01 00 90 54 72
00 90 4F 72 00 90 4B 72 00 90 48 72 60 80 48 00
00 80 4B 00 00 80 4F 00 00 80 54 00 60 90 57 72
60 80 57 00 00 90 56 72 00 90 4F 72 00 90 4B 72
00 90 47 72 60 80 47 00 00 80 4B 00 00 80 4F 00
00 80 56 00 00 90 57 72 60 80 57 00 00 90 56 72
60 80 56 00 00 90 54 72 00 90 4F 72 00 90 4B 72
00 90 46 72 60 80 46 00 00 80 4B 00 00 80 4F 00
00 80 54 00 60 90 57 72 60 80 57 00 00 90 56 72
00 90 4B 72 00 90 4E 72 00 90 45 72 82 20 80 45
```

MIDI File Format

TRACK CHUNK



- 4D 54 72 6B = Start of track chunk
- nn nn nn nn = Length of this track chunk

- Delta Time = Elapsed time (delta time) from the previous event to this event

MIDI File Format

EXAMPLE OF MIDI EVENT

MIDI EVENT	DESCRIPTION
8x nn vv	Note off
9x nn vv	Note on

nn = Note number
vv = Note velocity

EXAMPLE OF META EVENT

META EVENT	DESCRIPTION
FF 01 nn tt…	Text event
FF 04 nn tt…	Track instrument name
FF 05 nn tt…	Lyric

nn = Length of Text
tt… = Text characters

MIDI File Format

1st Track has 25-byte track event

2nd Track has 9381-byte track event

Header Chunks

```
4D 54 68 64 00 00 00 06 00 01 00 04 00 C0 4D 54
72 6B 00 00 00 19 00 FF 51 03 05 B8 D8 00 FF 58
04 06 03 12 08 00 FF 59 02 FD 00 00 FF 2F 00 4D
54 72 6B 00 00 24 A5 00 FF 03 05 50 69 61 6E 6F
00 FF 20 01 00 00 B0 07 64 00 C0 01 00 90 54 72
00 90 4F 72 00 90 4B 72 00 90 48 72 60 80 48 00
00 80 4B 00 00 80 4F 00 00 80 54 00 60 90 57 72
60 80 57 00 00 90 56 72 00 90 4F 72 00 90 4B 72
00 90 47 72 60 80 47 00 00 80 4B 00 00 80 4F 00
00 80 56 00 00 90 57 72 60 80 57 00 00 90 56 72
60 80 56 00 00 90 54 72 00 90 4F 72 00 90 4B 72
00 90 46 72 60 80 46 00 00 80 4B 00 00 80 4F 00
00 80 54 00 60 90 57 72 60 80 57 00 00 90 56 72
00 90 4B 72 00 90 4E 72 00 90 45 72 82 20 80 45
```


MIDI File Format

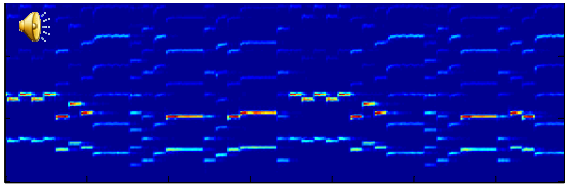
META EVENT
NOTE ON
CONTROL/PROGRAM CHANGE
NOTE OFF

```
4D 54 68 64 00 00 00 06 00 01 00 04 00 C0 4D 54
72 6B 00 00 00 19 00 FF 51 03 05 B8 D8 00 FF 58
04 06 03 12 08 00 FF 59 02 FD 00 00 FF 2F 00 4D
54 72 6B 00 00 24 A5 00 FF 03 05 50 69 61 6E 6F
00 FF 20 01 00 00 B0 07 64 00 C0 01 00 90 54 72
00 90 4F 72 00 90 4B 72 00 90 48 72 60 80 48 00
00 80 4B 00 00 80 4F 00 00 80 54 00 60 90 57 72
60 80 57 00 00 90 56 72 00 90 4F 72 00 90 4B 72
00 90 47 72 60 80 47 00 00 80 4B 00 00 80 4F 00
00 80 56 00 00 90 57 72 60 80 57 00 00 90 56 72
60 80 56 00 00 90 54 72 00 90 4F 72 00 90 4B 72
00 90 46 72 60 80 46 00 00 80 4B 00 00 80 4F 00
00 80 54 00 60 90 57 72 60 80 57 00 00 90 56 72
00 90 4B 72 00 90 4E 72 00 90 45 72 82 20 80 45
```

Parsons Code

- Symbol Sequence for representing Melodic Contours

R = "Repeat" : Current note is the same previous note
U = "Up" : Current note is higher than previous note
D = "Down" : Current note is lower than previous note



Parsons Code

*UDUDUDDDUUU
DUUUDUDUDUD
UDDDUUUDUDD

Levenshtein Distance

- Edit distance between two string/symbol sequence
- Levenshtein Distance $LD(X,Y)$ = Minimum no. of operations need for editing sequence X to Y

Operations

INSERTION : "**KARN**" ⇒ "**KARNY**", "**พ่อ**" ⇒ "**พ่อง**"
DELETION : "**SPORT**" ⇒ "**SORT**", "**ควาย**" ⇒ "**คาย**"
SUBSTITUTION : "**3.99**" ⇒ "**1.99**", "**DUCK**" ⇒ "**LUCK**"

Levenshtein Distance

LD("**KARN**" ,"**KAK**") =

LD("**LOSO**" ,"**SOLO**") =

LD("**หอม**" ,"**หมอน**") =

X = [69 69 72 74 72 69 71 71 71 75]
Y = [72 72 69 71 74 71]

LD(X,Y) =

Levenshtein Distance

- Can be compute by using dynamic programming

$$X = [x_1 \quad x_2 \quad \cdots \quad x_n] \text{ and } Y = [y_1 \quad y_2 \quad \cdots \quad y_m]$$

$$LD(i,j) = LD(x_1 x_2 \cdots x_i, y_1 y_2 \cdots y_j)$$

$$LD(0,0) = 0$$

$$LD(i,0) = i \quad \text{(i deletion)}$$

$$LD(0,j) = j \quad \text{(j insertion)}$$

$$LD(i,j) = \min \begin{cases} LD(i-1,j)+1 \\ LD(i,j-1)+1 \\ LD(i-1,j-1)+d \end{cases} \begin{matrix} \text{(deletion)} \\ \text{(insertion)} \end{matrix}$$

$$d = \begin{cases} 0 & ; x_i = y_j \quad \text{(match)} \\ 1 & ; x_i \neq y_j \quad \text{(substitution)} \end{cases}$$

Levenshtein Distance

$$LD(i,j) = \min\begin{cases} LD(i-1,j)+1 \\ LD(i,j-1)+1 \\ LD(i-1,j-1)+d \end{cases}$$

$$d = \begin{cases} 0 & ;x_i = y_j \\ 1 & ;x_i \neq y_j \end{cases}$$

$LD(i-1,j)$

$x_1x_2\cdots x_{i-1}$ $y_1y_2\cdots y_j$

$x_1x_2\cdots x_{i-1}$ x_i $y_1y_2\cdots y_j$

$LD(i,j) = LD(i-1,j)+1$

$LD(i,j-1)$

$x_1x_2\cdots x_i$ $y_1y_2\cdots y_{j-1}$

$x_1x_2\cdots x_i$ $y_1y_2\cdots y_{j-1}$ y_j

$LD(i,j) = LD(i,j-1)+1$

$LD(i-1,j-1)$

$x_1x_2\cdots x_{i-1}$ $y_1y_2\cdots y_{j-1}$

$x_1x_2\cdots x_{i-1}$ x_i $y_1y_2\cdots y_{j-1}$ y_j

$LD(i,j) = LD(i-1,j-1)+\{0,1\}$

Levenshtein Distance

X = [69 69 72 74 72 69 71 71 71 75]
Y = [72 72 69 71 74 71]

LD(0,1) = LD([],[72]) =

LD(0,3) = LD([],[72 72 69]) =

LD(2,0) = LD([69 69],[]) =

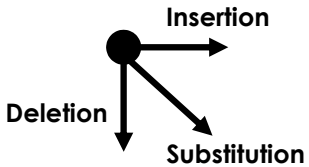
LD(2,3) = LD([69 69],[72 72 69]) =

LD(2,4) = LD([69 69],[72 72 69 71]) =

LD(3,3) = LD([69 69 72],[72 72 69]) =

LD(3,4) = LD([69 69 72],[72 72 69 71]) =

Levenshtein Distance



X \ Y	∅	72	72	69	71	74	71
∅	0	1	2	3	4	5	6
69	1						
69	2	2	2	2	3	4	5
72	3	2	2	3	3	4	5
74	4	3	3	3	4	3	4
72	5						
69	6	5	4	3	4	5	5
71	7	6	5	4	3	4	5
71	8	7	6	5	4	4	4
71	9	8	7	6	5	5	4
75	10	9	8	7	6	6	5

Levenshtein Distance

X = [DUDRR]
Y = [UURURDUUDURDRRDDU]

X \ Y	∅	U	U	R	U	R	D	U	U	D	U	R	D	R	R	D	D	U
∅	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
D	1	1	2	3	4	5	5	6	7	8	9	10	11	12	13	14	15	16
U	2	1	1	2	3	4	5	5	6	7	8	9	10	11	12	13	14	15
D	3	2	2	2	3	4	4	5	6	6	7	8	9	10	11	12	13	14
R	4	3	3	2	3	3	4	5	6	7	7	7	8	9	10	11	12	13
R	5	4	4	3	3	3	4	5	6	7	8	7	8	8	9	10	11	12

Levenshtein Distance

Fuzzy Substring Matching

X = [DUDRR]
Y = [UURURDUUDURDRRDDU]

X \ Y	∅	U	U	R	U	R	D	U	U	D	U	R	D	R	R	D	D	U
∅	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	0	0	1
U	2	1	1	2	1	2	1	0	1	1	0	1	1	1	2	1	1	0
D	3	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	1	1
R	4	3	3	2	3	2	3	2	2	2	2	1	2	1	2	3	2	2
R	5	4	4	3	3	3	3	3	3	3	3	2	2	2	1	2	3	3

Probabilistic String Matching

X = [42 45]

⬇

Y = [42 42 47 43 45 44]

LD(X,Y) = 4

(4 Insertions)

X = [41 41 47 42 44 43]

⬇

Y = [42 42 47 43 45 44]

LD(X,Y) = 5

(5 Substitutions)

Probabilistic String Matching

Prob. of Matching = 0.3
Prob. of Insertion = 0.1
Prob. of Deletion = 0.05
Prob. of Substitution (Error = 1 Midi no.) = 0.2
Prob. of Substitution (Error = 2 Midi no.) = 0.05

X = [42 45] ➡ Y = [42 42 47 43 45 44]

$P(X,Y) = P(\text{Match})^2\, P(\text{Insert})^4 = 0.3^2 \times 0.1^4 = 9 \times 10^{-6}$

X = [41 41 47 42 44 43] ➡ Y = [42 42 47 43 45 44]

$P(X,Y) = P(\text{Match})\, P(\text{Error} = 1)^5 = 0.3 \times 0.2^5 = 9.6 \times 10^{-5}$

Probabilistic String Matching

X = [42 45 42] ➡ Y = [42 44 47 43]

Wrong because of humming or detection

⬇

Match 42	$P(x = 42, y = 42)$	Prob. of Note 42 is detected correctly
Substitute 45 with 44	$P(x = 45, y = 44)$	Prob. of Note 44 is detected as 45
Substitute 42 with 47	$P(x = 42, y = 47)$	Prob. of Note 47 is detected as 42
Insert 43	$P(x = \phi, y = 43)$	Prob. of Note 43 is missing

Match 42	$P(x = 42, y = 42)$	Prob. of Note 42 is detected correctly
Insert 44	$P(x = \phi, y = 44)$	Prob. of Note 44 is missing
Substitute 45 with 47	$P(x = 45, y = 47)$	Prob. of Note 47 is detected as 45
Substitute 42 with 43	$P(x = 42, y = 43)$	Prob. of Note 43 is detected as 42

Probabilistic String Matching

Prob. of Matching = 0.3
Prob. of Insertion = 0.1
Prob. of Deletion = 0.05
Prob. of Substitution (Error = 1 Midi no.) = 0.2
Prob. of Substitution (Error = 2 Midi no.) = 0.05
Prob. of Substitution (Error = 3 Midi no.) = 0.01
Prob. of Substitution (Error = 4 Midi no.) = 0.005
Prob. of Substitution (Error = 5 Midi no.) = 0.002

X = [42 45 42] ➡ Y = [42 44 47 43]

Match 42 $P(x=42, y=42) = 0.3$
Substitute 45 with 44 $P(x=45, y=44) = 0.2$
Substitute 42 with 47 $P(x=42, y=47) = 0.002$
Insert 43 $P(x=\phi, y=43) = 0.1$

$P = 0.3 \times 0.2 \times 0.002 \times 0.1$
$= 1.2 \times 10^{-5}$

Match 42 $P(x=42, y=42) = 0.3$
Insert 44 $P(x=\phi, y=44) = 0.1$
Substitute 45 with 47 $P(x=45, y=47) = 0.05$
Substitute 42 with 43 $P(x=42, y=43) = 0.2$

$P = 0.3 \times 0.1 \times 0.05 \times 0.2$
$= 3 \times 10^{-4}$

Probabilistic String Matching

Prob. of Matching = 0.3
Prob. of Insertion = 0.1
Prob. of Deletion = 0.05
Prob. of Substitution (Error = 1 Midi no.) = 0.2
Prob. of Substitution (Error = 2 Midi no.) = 0.05
Prob. of Substitution (Error = 3 Midi no.) = 0.01
Prob. of Substitution (Error = 4 Midi no.) = 0.005
Prob. of Substitution (Error = 5 Midi no.) = 0.002

X = [42 45 42] ➡ Y = [42 44 47 43]

Insertion
Deletion
Substitution

	∅	42	44	47	43
∅	1	I 0.1	I 0.01	I 0.001	I 1×10-4
42	D 0.05	S 0.3	I 0.03		
45	D 0.0025				
42	D 1.25×10-4				

0.05, 0.05, 0.1

0.3 = max(1×0.3,0.1×0.05,0.05×0.1)
0.03 = max(0.1×0.05,0.01×0.05,0.3×0.1)

Probabilistic String Matching

Prob. of Matching = 0.3
Prob. of Insertion = 0.1
Prob. of Deletion = 0.05
Prob. of Substitution (Error = 1 Midi no.) = 0.2
Prob. of Substitution (Error = 2 Midi no.) = 0.05
Prob. of Substitution (Error = 3 Midi no.) = 0.01
Prob. of Substitution (Error = 4 Midi no.) = 0.005
Prob. of Substitution (Error = 5 Midi no.) = 0.002

X = [42 45 42] ➡ Y = [42 44 47 43]

Insertion
Deletion
Substitution

	∅	42	44	47	43
∅	1	I 0.1	I 0.01	I 0.001	I 1×10-4
42	D 0.05	S 0.3	I 0.03	I 0.003	I 3×10-4
45	D 0.0025	D 0.015	S 0.06	I 0.006	I 6×10-4
42	D 1.25×10-4	S 7.5×10-4	D 0.003	I 3×10-4	S 0.0012

0.3 = max(1×0.3,0.1×0.05,0.05×0.1)
0.03 = max(0.1×0.05,0.01×0.05,0.3×0.1)