London

Polyphonic Music Retrieval : The *N*-Gram Approach

Shyamala Doraisamy and Stefan Rüger.

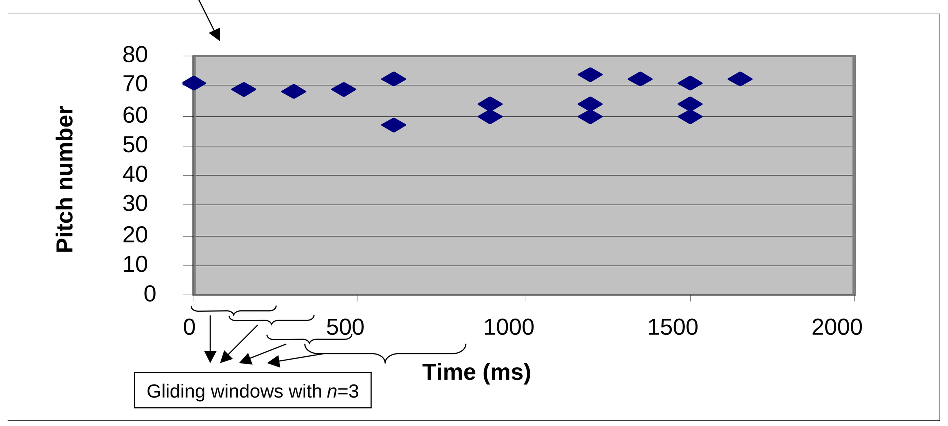
Multimedia Knowledge Management, Department of Computing, Imperial College London, UK.

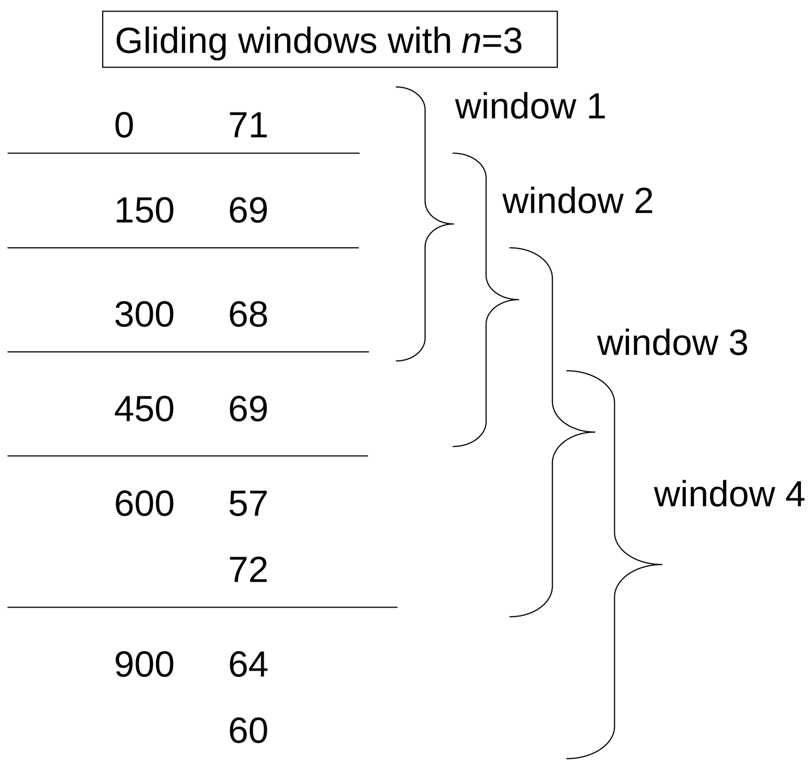
http://km.doc.ic.ac.uk/ {sd3, s.rueger}@imperial.ac.uk

PATTERN EXTRACTION AND ENCODING

Polyphonic excerpt of Mozart's Alla Turca

Concurrent events on a time-line





Sequences from the pitch and rhythm dimensions

Pitch

$$Interval_i = Pitch_{i+1} - Pitch_i$$

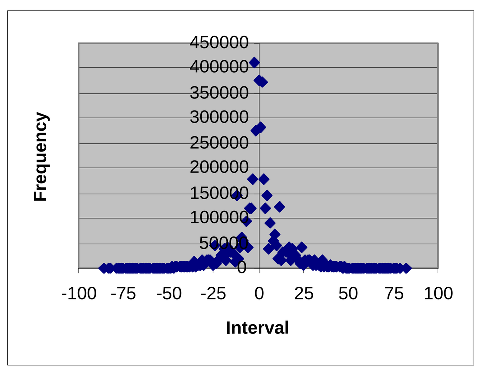
Rhythm

$$Ratio_i = \frac{Onset_{i+2} - Onset_{i+1}}{Onset_{i+1} - Onset_i}$$

N-grams consist of *n*-1intervals and *n*-2 ratios, and encoded to generate a polyphonic musical word document.

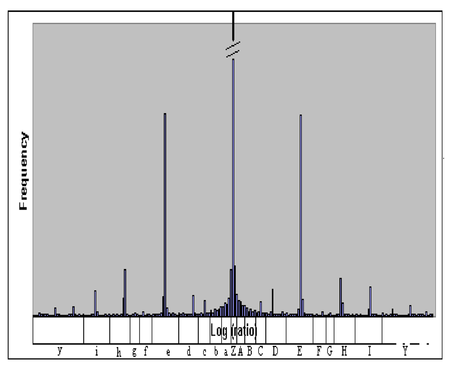
$[I_1 R_1 \ldots I_{n-2} R_{n-2} I_{n-1}]$

Analysis of interval distribution and a non-linear function to map intervals to text letters



$C(I) = \text{int}(X \tanh(I/Y))$

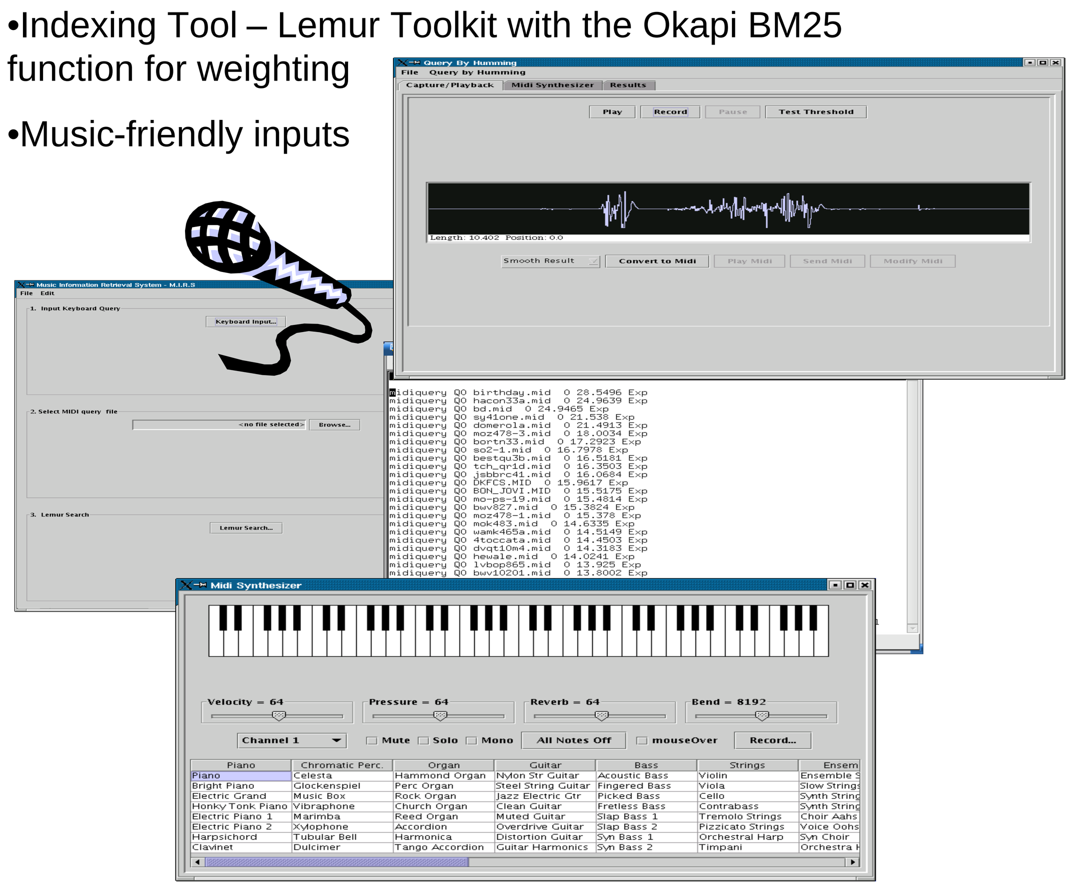
Ratio range quantisation and bins in encoding rhythm information



- •Encode polyphonic music piece as an ordered pair of onset time and pitch
- •Group pitches with similar onset times together as musical events
- •Using the gliding window, the sequence of events are divided into overlapping sub-sequences of *n* different adjacent events, each with its unique onset time
- •All possible monophonic sequences are extracted in constructing the corresponding musical words

INDEXING AND RETRIEVAL SYSTEM

- •Collection – around 10,000 polyphonic musical performances in the MIDI format
- •Indexing Tool – Lemur Toolkit with the Okapi BM25 function for weighting
- •Music-friendly inputs



N-Gramming Strategies – addressing various problems and issues

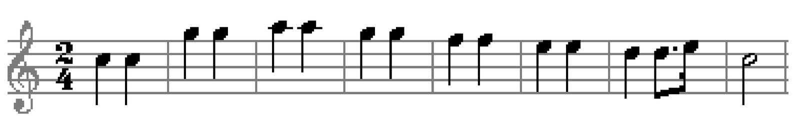
1. Alternate onsets – intercepting accompaniment onsets

Monophonic query

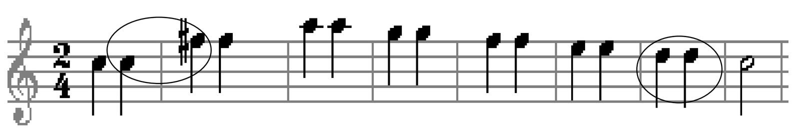
Polyphonic collection

2. Encoding precision - robustness with erroneous queries

Perfect query



Erroneous query



3. Envelopes – large number of possible terms

Orchestral score

4. Proximity with 'overlaying' words – improving precision

Position Indexing for Polyphonic Musical Words

1 bYaYA_2 aYAYI 2 aYAYC 3 AYIFG 3 AYCFh 3 AYIFC 3 AYCFI 4 IFGYJ 4 CFhYJ 4 IFCYN 4 CFIYN 4 IFGY8 4 CFhY8 4 IFCYD 4 CFIYD 4 IFGYd 4 CFhYd 4 IFCY8 4 CFIY8 5 GYJfb 5 hYJfb 5 CYNfb 5 IYNfb 5 GY8fH 5 hY8fH 5 CYDfH 5 IYDfH 5 GYdfL 5 hYdfL 5 CY8fL 5 IY8fL 6 JfbYa 6 NfbYa 6 8fHYa 6 DfHYa 6 dfLYa 6 8fLYa 6 JfbYh 6 NfbYh 6 8fHYh 6 DfHYh ..

EXPERIMENTAL DESIGN

Index Files and Experimental Factors

Index	Pitch	Rhythm	*n*	*Y*	#R.Bins	#Terms
P3	yes	-	3	24		$2,809 = 53^2$
PR3	yes	yes	3	24	21	$58,989 = 53^2 \cdot 21$
PR3CP1	yes	yes	3	48	21	$58,989 = 53^2 \cdot 21$
PR3CP2	yes	yes	3	72	21	$58,989 = 53^2 \cdot 21$
PR3CP1CR	yes	yes	3	48	11	$30,899 = 53^2 \cdot 11$
PR3CP2CR	yes	yes	3	72	11	$30,899 = 53^2 \cdot 11$
P4	yes	-	4	24		$148,877 = 53^2$
PR4	yes	yes	4	24	21	$65,654,757 = 53^2 \cdot 21^2$
PR4CP1	yes	yes	4	48	11	$65,654,757 = 53^2 \cdot 21^2$
PR4CP2	yes	yes	4	72	11	$65,654,757 = 53^2 \cdot 21^2$
PR4CP1CR	yes	yes	4	48	11	$18,014,117 = 53^2 \cdot 11^2$
PR4CP2CR	yes	yes	4	72	11	$18,014,117 = 53^2 \cdot 11^2$
PR4ENV	yes	yes	4	24	21	$65,654,757 = 53^2 \cdot 21^2$
PR5ENV	yes	yes	5	24	21	$3,479,702,121 = 53^4 \cdot 21^2$

Error Simulation

- • Monophonic queries - Query by Humming (QBH)

Error Models of QBH systems

- • Polyphonic queries - Query by Example (QBE)

$NewInterval_k = Interval_k + D_i \cdot \varepsilon$

$NewRatio_k = Ratio_k \cdot \exp(D_r \cdot \varepsilon)$

Test Collection

•Documents

Collection – Around 10,000 music documents obtained from the Internet and divided into training and test sets

•Queries – simulated using themes based on the Dictionary of Musical Themes, Barlow and Morgenstern (1948)

•Relevance judgements – assumptions based on Uitdenbogerd and Zobel (1999)

Proximity Analysis

Structured and Proximity-Based Operators

Sum operator : #sum($T_1 \ldots T_n$)

The terms or nodes contained in the sum operator are treated as having equal influence on the final result. The belief values p rovided by the arguments of the sum are averaged to produce the belief v alue of the #sum node.

Ordered Distance Operator : #odN ($T_1 \ldots T_n$)

The terms within an ODN operator must be found in any order with in a window of N words of each other in the text in order to contribu te to the document's belief value.

And operator : #and ($T_1 \ldots T_n$)

The more terms contained in the AND operator which are found in the document , the higher the belief value of the document.

Or operator: #or ($T_1 \ldots T_n$)

One of the terms within the OR operator must be found in a docum ent for that document to get credit for this operator.

Monophonic query

bYaYA aYAYC AYCIB CIBib BibYa bYaYA aYAYD AYDIA DIAia AiaYa aYaYA aYAYG AYGYb GYbYa bYaYA aYAYB AYBYb BYbYa bYaYA aYAYC

Query Formulation using Nested Phrase Operators

#SUM(#ODN3(bYaYA aYAYC)
#ODN3(AYCIB CIBib)
…
#ODN3(bYaYA aYAYC))

RESULTS AND CONCLUSIONS

Experimental Series: 1. Preliminary Investigation, 2. Query by Melody, Fault-Tolerance and Comparative Study, 3. Robustness and Envelopes and 4. Proximity Analysis

% of relevant documents retrieved within the top 15

	Perfect Queries				Humming Errors (*p*=20%)			
Song ID	PR4	AL1	AL2	AM	PR4	AL1	AL2	AM
1	100	0	0	100	0	0	0	2
2	50	50	25	50	28	23	5	50
3	0	0	0	0	0	0	0	0
4	0	0	0	100	0	0	0	90
5	100	100	100	100	82	84	9	90
6	13	0	0	100	9	0	0	78
7	100	100	50	100	70	75	35	70
8	33	33	0	67	30	7	0	40
9	50	50	50	50	50	50	50	50
10	86	71	43	86	86	71	26	82
W.A.	58	40	34	80	38	32	10	58

Exp. 2 A promising performance despite the large number of index terms with full-music indexing of polyphonic music – enabling retrieval from a polyphonic collection without the need for melody extraction algorithms

MRR Performance of monophonic queries

Index	Perfect Query	*p* = 10%	*p* = 20%	*p* = 30%	*p* = 50%
P3	0.05±0.15	0.03±0.11	0.03±0.12	0.02±0.10	0.03±0.11
PR3	0.31±0.45	0.19±0.35	0.16±0.31	0.15±0.30	0.10±0.25
PR3CP1	0.17±0.31	0.19±0.36	0.17±0.34	0.16±0.34	0.10±0.26
PR3CP2	0.13±0.30	0.10±0.26	0.09±0.24	0.10±0.26	0.07±0.21
PR3ENV	0.39±0.43	0.27±0.38	0.23±0.36	0.17±0.32	0.11±0.26
P4	0.47±0.45	0.29±0.42	0.25±0.40	0.18±0.33	0.14±0.31
PR4	0.61±0.41	0.48±0.44	0.42±0.43	0.37±0.43	0.26±0.40
PR4CP1	0.70±0.39	0.54±0.43	0.49±0.44	0.44±0.45	0.29±0.39
PR4CP2	0.48±0.44	0.41±0.44	0.38±0.44	0.34±0.44	0.26±0.39
PR4ENV	0.75±0.32	0.65±0.39	0.61±0.42	0.47±0.44	0.34±0.41
PR5ENV	0.80±0.40	0.78±0.39	0.71±0.43	0.56±0.47	0.39±0.46

Exp. 3 Heuristics of restrictions to the variations of the upper and lower envelopes of polyphonic music improved retrieval precision

MRR Performance of polyphonic queries

Index	Perfect Query	$D_i = 1$, $D_r = 0.02$	$D_i = 2$, $D_r = 0.02$
P3	0.00±0.00	0.00±0.00	0.00±0.00
PR3	0.35±0.45	0.05±0.11	0.04±0.10
PR3CP1CR	0.11±0.30	0.01±0.04	0.01±0.03
PR3CP2CR	0.10±0.30	0.00±0.01	0.00±0.01
PR3ENV	0.62±0.47	0.13±0.31	0.05±0.16
P4	0.76±0.38	0.27±0.38	0.10±0.27
PR4	0.85±0.31	0.36±0.39	0.25±0.38
PR4CP1CR	0.81±0.38	0.23±0.35	0.14±0.30
PR4CP2CR	0.60±0.49	0.16±0.32	0.12±0.30
PR4ENV	1.00±0.00	0.57±0.45	0.33±0.43
PR5ENV	1.00±0.00	0.56±0.47	0.27±0.42

Retrieval Measures with Adjacent and Concurrent N-G

		W.A.	MRR
PR4		52	0.63
PR4ENV		62	0.66
PPR4	Sum	48	0.57
	Nested	58	0.70
	AND	49	0.56
	OR	62	0.61
PPR4ENV	Sum	62	0.62
	Nested	62	0.66
	AND	56	0.65
	OR	56	0.62

Exp. 4 Nested phrase operators improve retrieval performance - a proximity-based operator for music retrieval is being investigated

- •Proven the usefulness of using n-grams in polyphonic music retrieval – retrieving from a polyphonic music collection with both monophonic and polyphonic queries
- •Introduced a non-linear function to map intervals to text letters, and analysed how to best quantise ratio ranges in order to incorporate rhythm information
- •Development of a state-of-the-art test collection and the use of error models for fault-tolerance investigation