Database Integration of MPEG-7 Low-Level Audio Metadata

Dr. Michael Casey

City University, London /

Goldsmiths College, University of London

AES25 Workshop on Practical Issues on MPEG-7

Overview

- MPEG-7 Low level audio descriptors
- Audio-based information retrieval
- LLAMAS low-level audio management
- Video retrieval by audio
- Database schemas
- Scalability of features
- Performance Evaluation

ISO/IEC JTC1/SC29 WG11

MPEG

MOVING PICTURE EXPERTS GROUP

MPEG-7
International Standard

- Features (Descriptors)
  - Visual D
  - Audio LLDs (Low-Level Descriptors)
- Description Schemes (Feature Sets)
  - Multimedia Description Schemes
  - Tailored to applications
- Segment Decomposition
- Statistical Models
- Similarity Metrics

Audio Descriptions

Header

```
<!-- DAFx 2003 MPEG-7 Audio Processing Examples -->
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
        <Description xsi:type="ContentEntityType">
                <MultimediaContent xsi:type="AudioType">
                        <Audio xsi:type="AudioSegmentType">
                                <MediaTime>
                                        <MediaTimePoint>T00:00:00</MediaTimePoint>
                                        <MediaDuration>PT15S450N1000F</MediaDuration>
                                </MediaTime>
                                <AudioDescriptor xsi:type="AudioWaveformType">
                                        <SeriesOfScalar hopSize="PT10N1000F" totalNumOfSamples="1545">
                                                <Min>
-6.10352e-05 -0.000274658 -0.0010376 -0.00161743 -0.00210571 -0.00216675 -0.00259399 -0.00473022 -0.00
1 -0.00299072 -0.0043335 -0.00762939 -0.00497437 -0.00683594 -0.00408936 -0.0071106 -0.00296021 -0.010
-0.0118713 -0.00765991 -0.00363159 -0.0100403 -0.0131226 -0.0138245 -0.0131226 -0.00772095 -0.00854492
-0.0020752 -0.00265503 -0.00158691 -0.00140381 -0.0010376 -0.000518799 -0.000610352 -0.000183105
                                                </Min>
                                                <Max>
0.000335693 0.000671387 0.00152588 0.000396729 0.00177002 0.00125122 0.00286865 0.00119019 0.00280762
74658 0.0032959 0.0027771 0.00552368 0.00445557 0.00588989 0.00494385 0.00674438 0.00964355 0.0007324
0.00210571 0.00167847 0.000671387 0.00109863 0.00112915 0.000488281 0.000274658
                                                </Max>
                                        </SeriesOfScalar>
                                </AudioDescriptor>
                        </Audio>
                </MultimediaContent>
        </Description>
</Mpeg7>
```

Audio Descriptions

```
<!-- DAFx 2003 MPEG-7 Audio Processing Examples -->
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
        <Description xsi:type="ContentEntityType">
                <MultimediaContent xsi:type="AudioType">
                        <Audio xsi:type="AudioSegmentType">
                                <MediaTime>
                                        <MediaTimePoint>T00:00:00</MediaTimePoint>
                                        <MediaDuration>PT15S450N1000F</MediaDuration>
                                </MediaTime>
                                <AudioDescriptor xsi:type="AudioWaveformType">
                                        <SeriesOfScalar hopSize="PT10N1000F" totalNumOfSamples="1545">
                                                <Min>
-6.10352e-05 -0.000274658 -0.0010376 -0.00161743 -0.00210571 -0.00216675 -0.00259399 -0.00473022 -0.00
1 -0.00299072 -0.0043335 -0.00762939 -0.00497437 -0.00683594 -0.00408936 -0.0071106 -0.00296021 -0.010
-0.0118713 -0.00765991 -0.00363159 -0.0100403 -0.0131226 -0.0138245 -0.0131226 -0.00772095 -0.00854492
-0.0020752 -0.00265503 -0.00158691 -0.00140381 -0.0010376 -0.000518799 -0.000610352 -0.000183105
                                                </Min>
                                                <Max>
0.000335693 0.000671387 0.00152588 0.000396729 0.00177002 0.00125122 0.00286865 0.00119019 0.00280762
74658 0.0032959 0.0027771 0.00552368 0.00445557 0.00588989 0.00494385 0.00674438 0.00964355 0.0007324
0.00210571 0.00167847 0.000671387 0.00109863 0.00112915 0.000488281 0.000274658
                                                </Max>
                                        </SeriesOfScalar>
                                </AudioDescriptor>
                        </Audio>
                </MultimediaContent>
        </Description>
</Mpeg7>
```

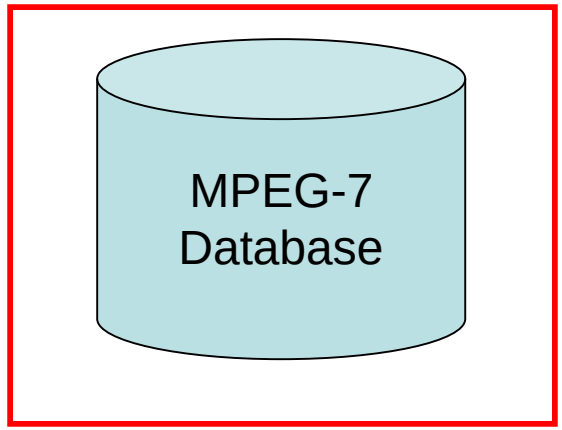
Segments

Audio Descriptions

```
<!-- DAFx 2003 MPEG-7 Audio Processing Examples -->
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
        <Description xsi:type="ContentEntityType">
                <MultimediaContent xsi:type="AudioType">
                        <Audio xsi:type="AudioSegmentType">
                                <MediaTime>
                                        <MediaTimePoint>T00:00:00</MediaTimePoint>
                                        <MediaDuration>PT15S450N1000F</MediaDuration>
                                </MediaTime>
                                <AudioDescriptor xsi:type="AudioWaveformType">
                                        <SeriesOfScalar hopSize="PT10N1000F" totalNumOfSamples="1545">
                                                <Min>
-6.10352e-05 -0.000274658 -0.0010376 -0.00161743 -0.00210571 -0.00216675 -0.00259399 -0.00473022 -0.00
1 -0.00299072 -0.0043335 -0.00762939 -0.00497437 -0.00683594 -0.00408936 -0.0071106 -0.00296021 -0.010
-0.0118713 -0.00765991 -0.00363159 -0.0100403 -0.0131226 -0.0138245 -0.0131226 -0.00772095 -0.00854492
-0.0020752 -0.00265503 -0.00158691 -0.00140381 -0.0010376 -0.000518799 -0.000610352 -0.000183105
</Min>
                                                <Max>
0.000335693 0.000671387 0.00152588 0.000396729 0.00177002 0.00125122 0.00286865 0.00119019 0.00280762
74658 0.0032959 0.0027771 0.00552368 0.00445557 0.00588989 0.00494385 0.00674438 0.00964355 0.0007324
0.00210571 0.00167847 0.000671387 0.00109863 0.00112915 0.000488281 0.000274658
</Max>
                                        </SeriesOfScalar>
                                </AudioDescriptor>
                        </Audio>
                </MultimediaContent>
        </Description>
</Mpeg7>
```

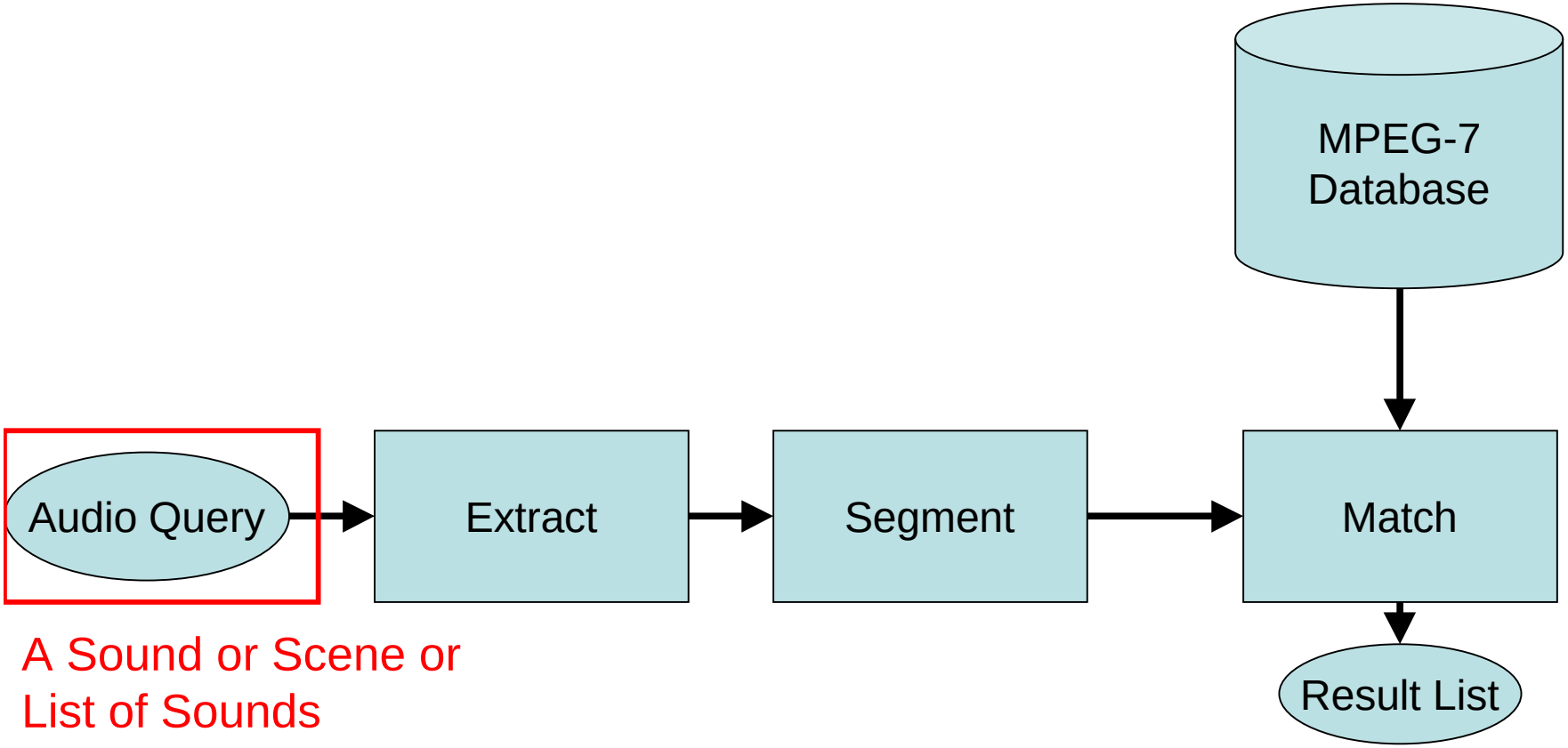
Descriptor

Audio Information Retrieval

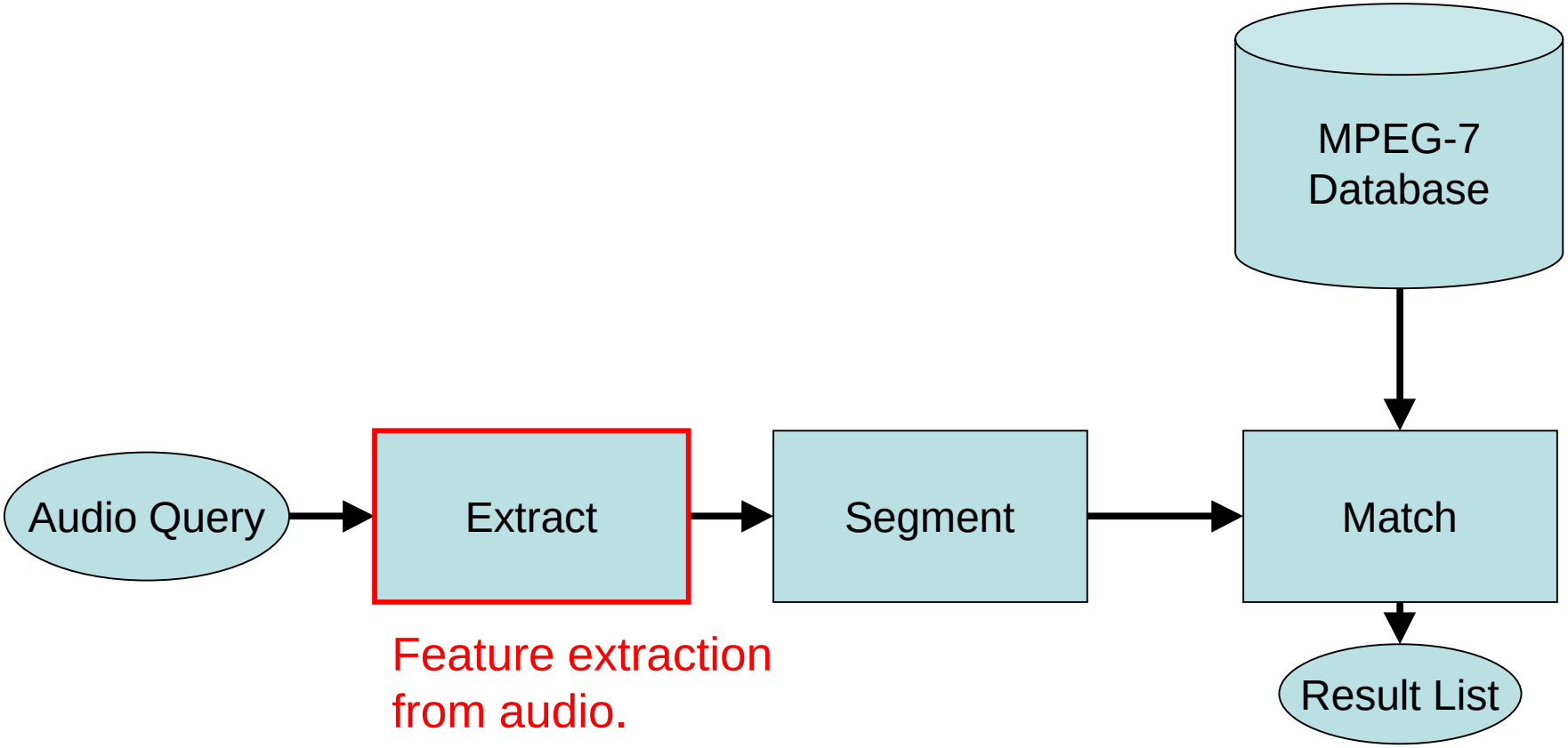


A pre-indexed Collection of Sounds

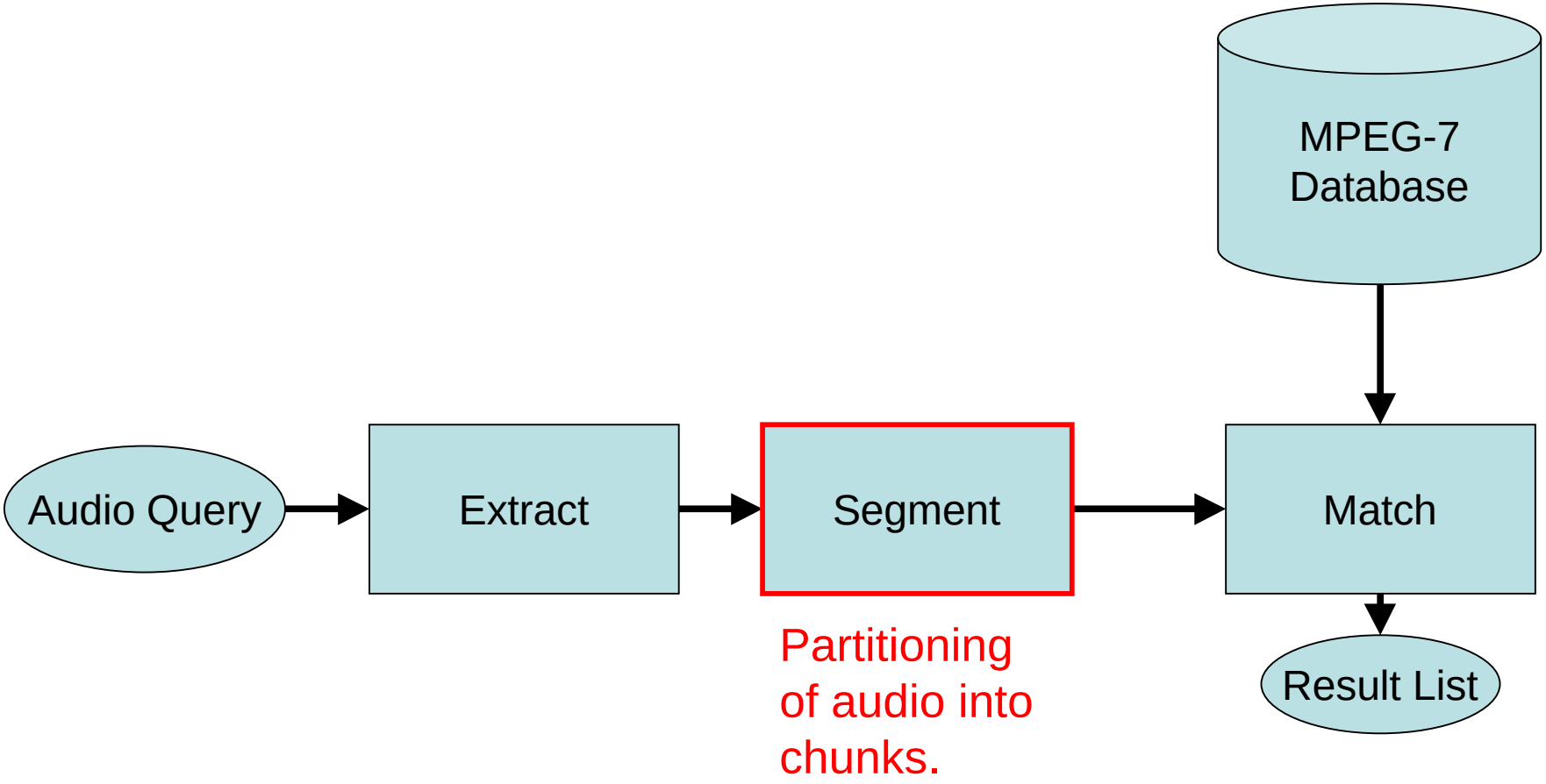
Audio Information Retrieval



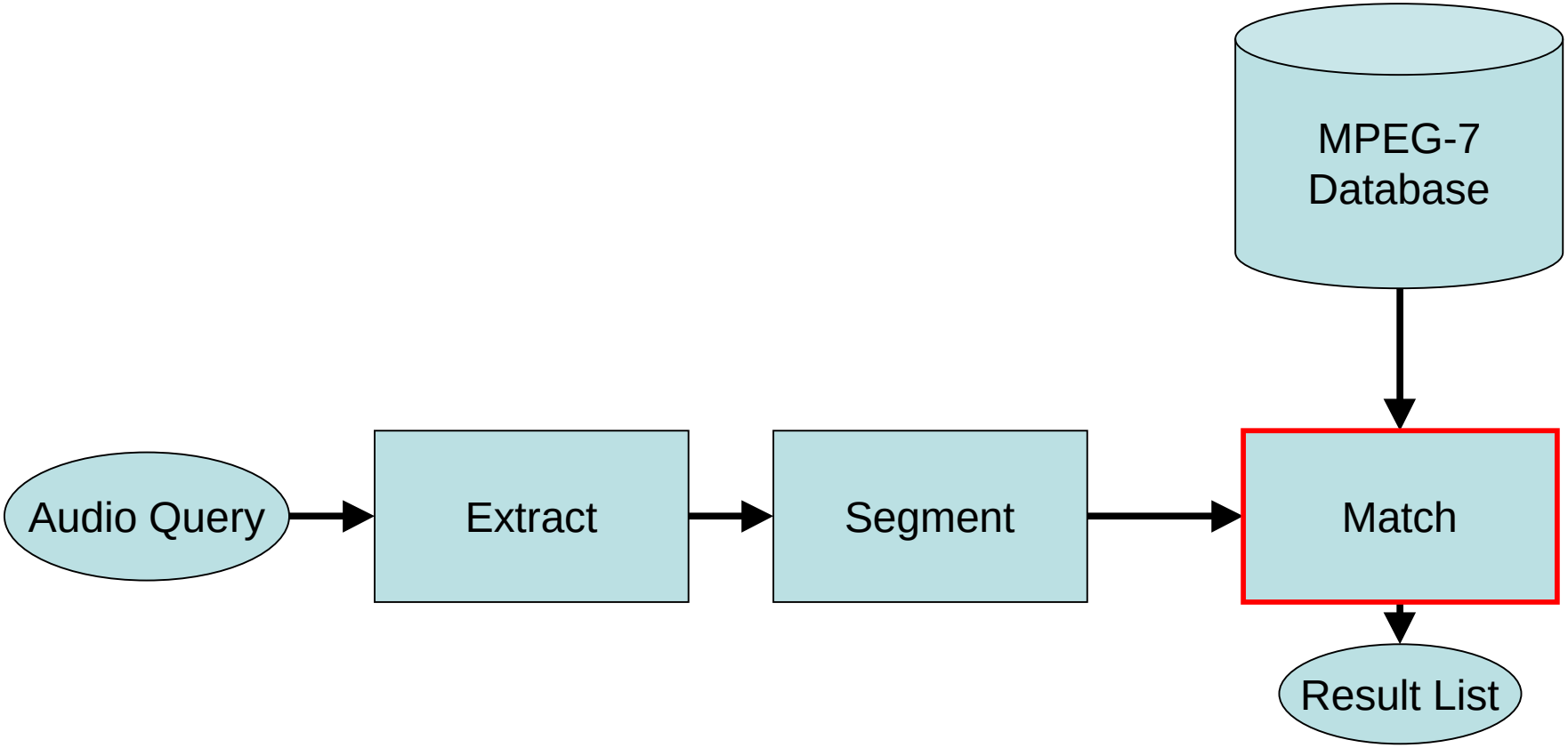
Audio Information Retrieval



Audio Information Retrieval

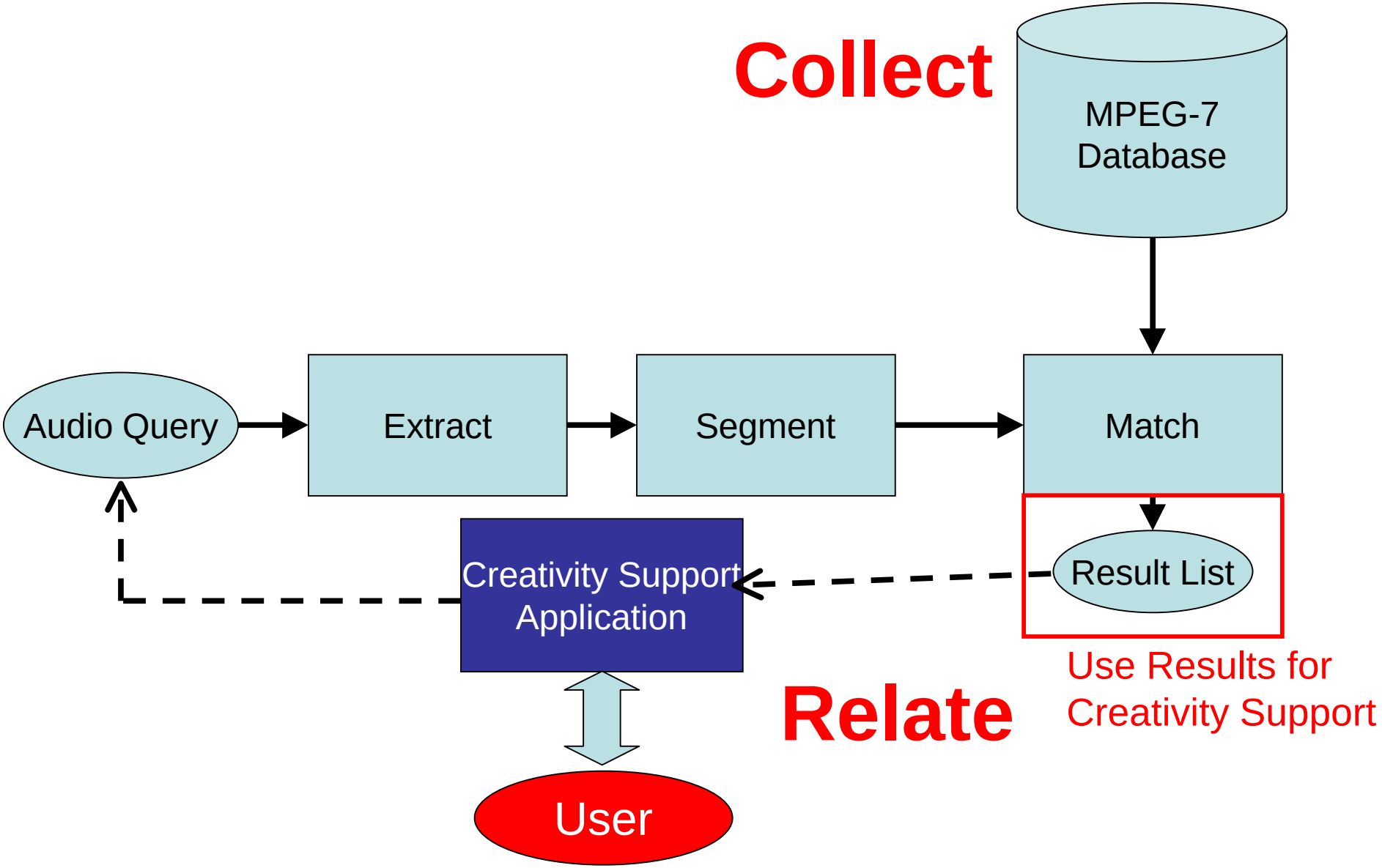


Audio Information Retrieval

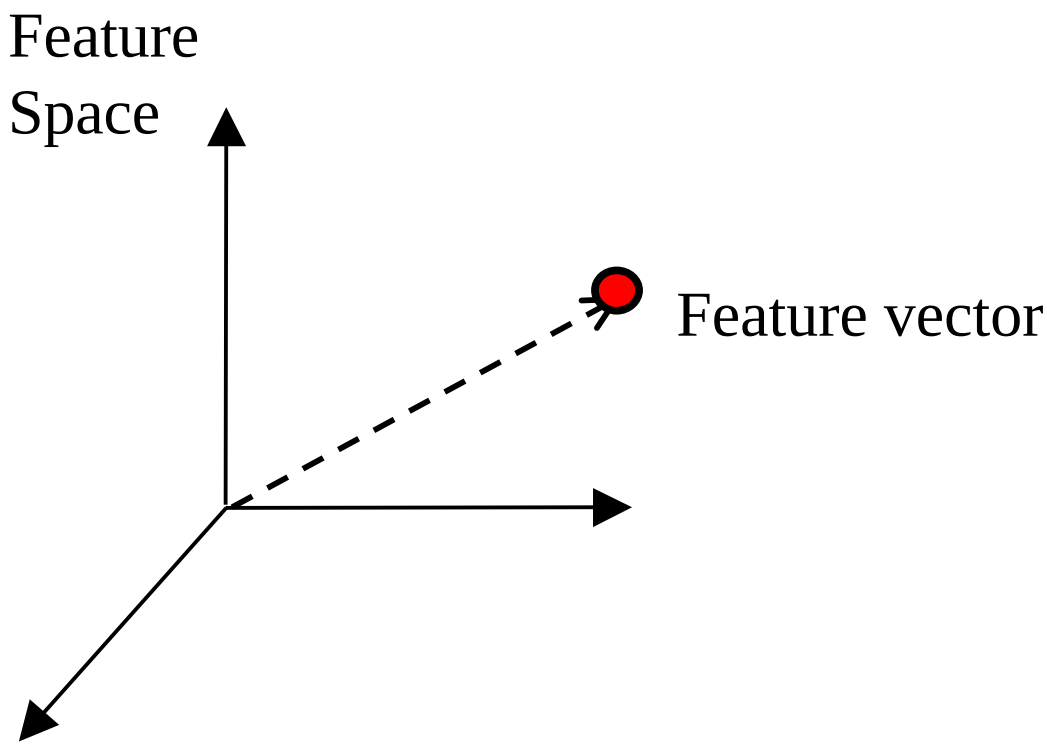


Find similar chunks of Audio

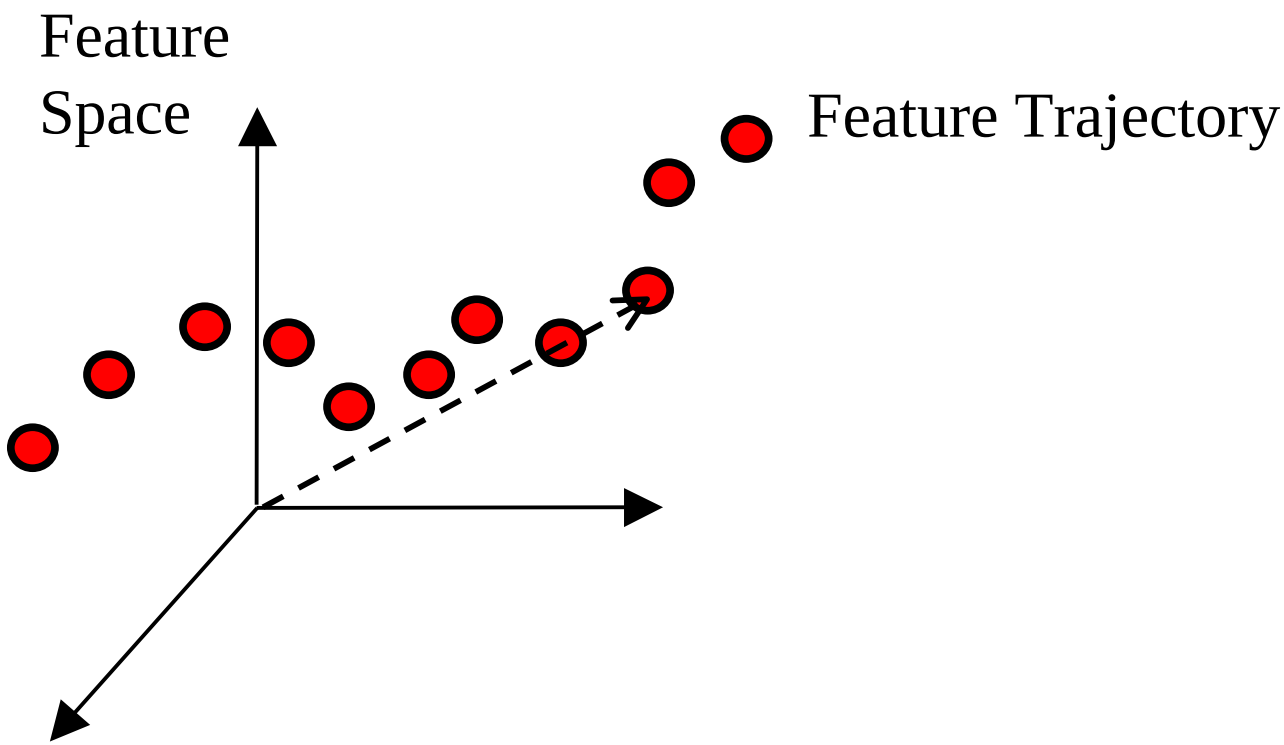
Audio Information Retrieval



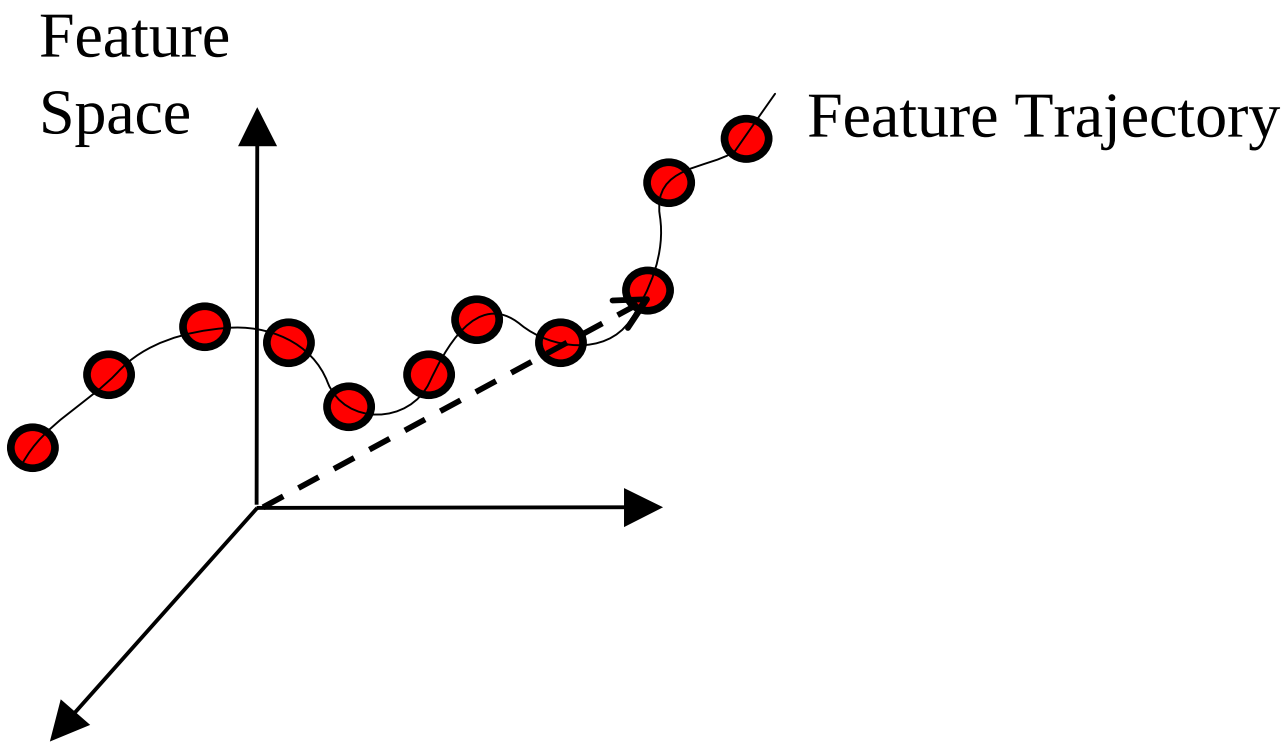
Feature Extraction



Feature Trajectories <SeriesOfVectors>

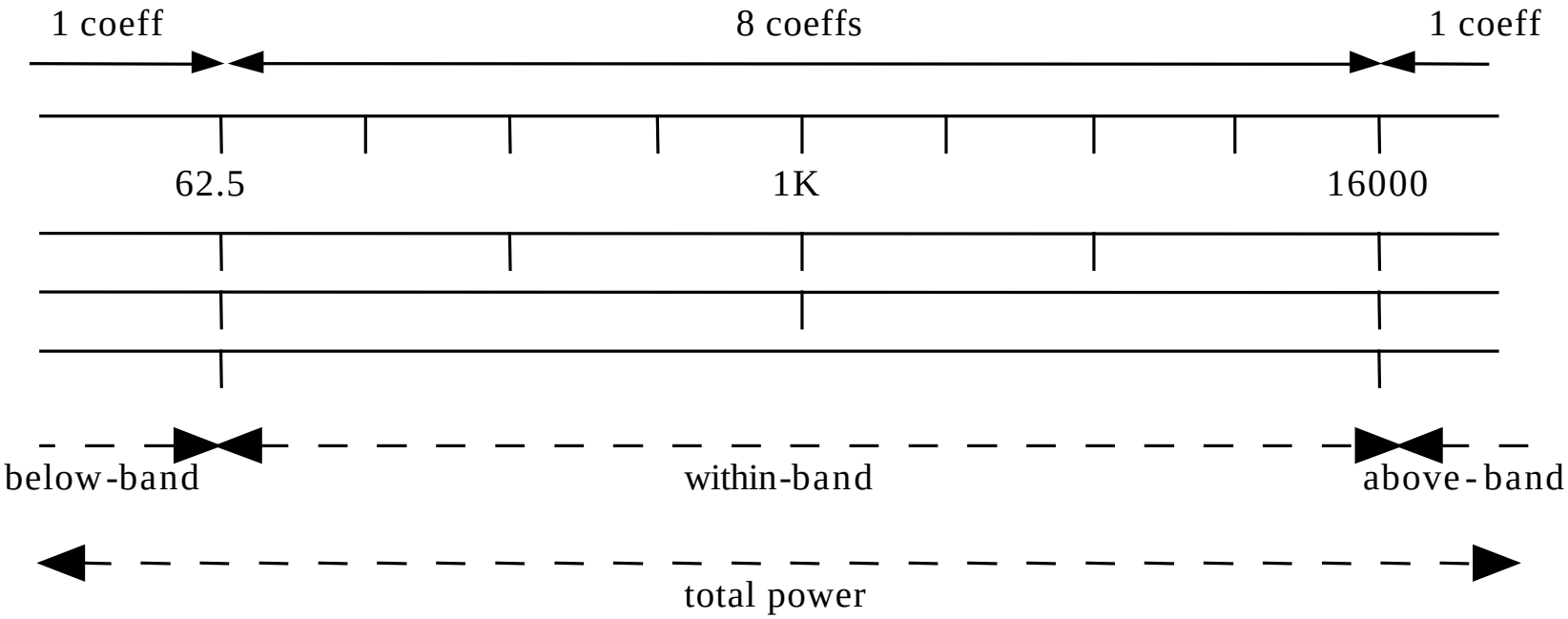


Feature Trajectories <SeriesOfVectors>



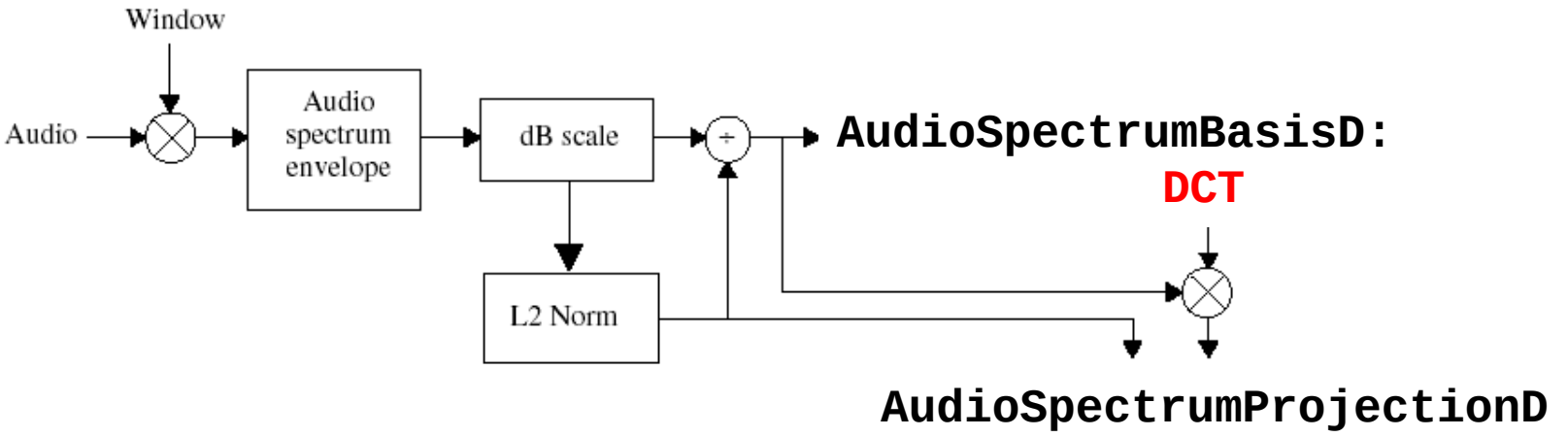
AudioSpectrumEnvelope

- Log frequency scale spectral power coefficients
- Total power preserved across logarithmic bands



AudioSpectrumProjection

General Linear transform of Log Spectrum

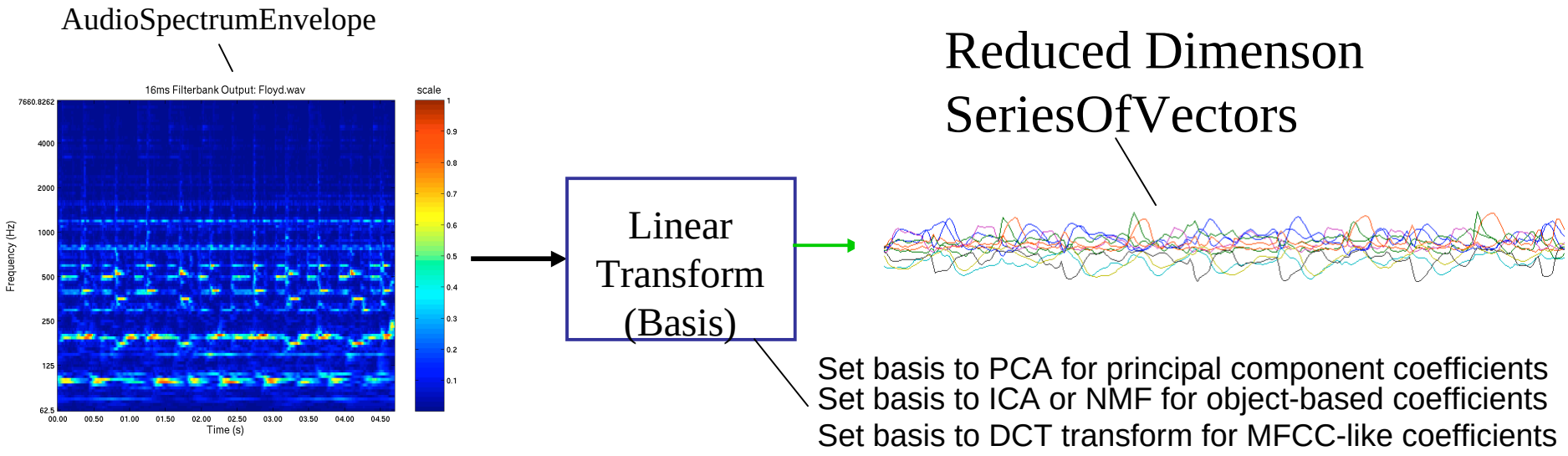


Set basis to PCA for principal component coefficients

Set basis to ICA or NMF for object-based coefficients

Set basis to DCT transform for MFCC-like coefficients

Spectrum Basis Projection Features



Log Frequency

Log amplitude

- Like MFCC using log frequency scale

Free windows MP7 Encoder

Download windows executable from *musicstructure.com*

```
AudioSpectrumProjection.exe -d 100 -b 21 -x foo.xml foo.wav foo.bin
XML output to text file:foo.xml
sr=44100.000000, channels = 1, sampleWidth=2, bufSize=8192
ASE loEdge:439.063
ASE hiEdge:8000
ASE octave:0.0625
ASE FFTsiz:2048
ASE NumEdg:70
ASE NumBan:69
ASE HopSiz:441
ASE Window:1323
ASE     fs:44100
ASE     wP:524.369

HH:MM:SS.mmm
01:01:00.000
>
```

Example Output of AudioSpectrumProjection.exe

SeriesOfVectors (ASCII, XML or Big-Endien Binary File [network byte order])

616.195	-0.898	0.325	-0.175	0.136	-0.104	0.082	-0.075	0.040	-0.058	0.041	-0.039	0.038	-0.014	0.037	-0.031	0.028	-0.030	0.020	-0.013	0.026
574.751	-0.888	0.323	-0.193	0.150	-0.110	0.070	-0.093	0.066	-0.062	0.033	-0.055	0.019	-0.029	0.041	-0.034	0.056	-0.017	0.027	-0.027	0.007
652.088	-0.874	0.358	-0.187	0.172	-0.105	0.086	-0.081	0.056	-0.052	0.044	-0.055	0.031	-0.029	0.029	-0.032	0.029	-0.031	0.017	-0.028	0.021
723.976	-0.892	0.336	-0.152	0.171	-0.089	0.083	-0.049	0.071	-0.058	0.060	-0.047	0.034	-0.031	0.024	-0.031	0.023	-0.036	0.015	-0.032	0.020
623.821	-0.881	0.345	-0.163	0.178	-0.094	0.095	-0.062	0.080	-0.054	0.054	-0.051	0.027	-0.039	0.012	-0.062	0.016	-0.042	0.016	-0.021	0.023
717.812	-0.883	0.346	-0.191	0.144	-0.105	0.075	-0.078	0.073	-0.055	0.053	-0.048	0.029	-0.022	0.033	-0.033	0.041	-0.023	0.019	-0.027	0.010
606.720	-0.892	0.331	-0.188	0.148	-0.107	0.054	-0.084	0.030	-0.058	0.046	-0.044	0.034	-0.021	0.034	-0.039	0.022	-0.047	0.013	-0.010	0.025
543.379	-0.896	0.303	-0.205	0.122	-0.125	0.050	-0.100	0.021	-0.032	0.067	-0.006	0.061	0.004	0.053	-0.048	0.010	-0.065	0.008	0.001	0.031
562.814	-0.886	0.328	-0.159	0.166	-0.129	0.066	-0.102	0.053	-0.066	0.026	-0.040	0.039	-0.011	0.062	-0.016	0.068	-0.040	0.013	-0.043	-0.009
502.962	-0.886	0.335	-0.196	0.140	-0.110	0.085	-0.086	0.055	-0.030	0.055	-0.033	0.042	-0.037	0.030	-0.029	0.051	-0.009	0.041	-0.033	-0.006
485.651	-0.864	0.364	-0.214	0.163	-0.106	0.089	-0.089	0.066	-0.050	0.049	-0.039	0.037	-0.031	0.049	-0.034	0.034	-0.037	0.023	-0.026	0.015
461.432	-0.863	0.367	-0.207	0.161	-0.117	0.089	-0.095	0.061	-0.043	0.068	-0.031	0.051	-0.037	0.033	-0.041	0.031	-0.025	0.027	-0.027	0.022

- Floating point representation;
    - IEEE 64-bit doubles as Base64 – simple to read/write
    - ASCII via libc.a printf ; needs parsing (costly)
- Fixed time offset between feature vectors
    - Fast access using MediaTime locators
- Efficient storage/retrieval/search for MM database applications

Similarity of Features

- Compute distance between feature pairs

- Similarity Metric
  - `dist(a,b) >= 0`
  - `dist(a,b)== 0 iff a==b`
  - `dist(a,b) + dist(b,c) >= dist(a,c)`
- Vector Dot Product

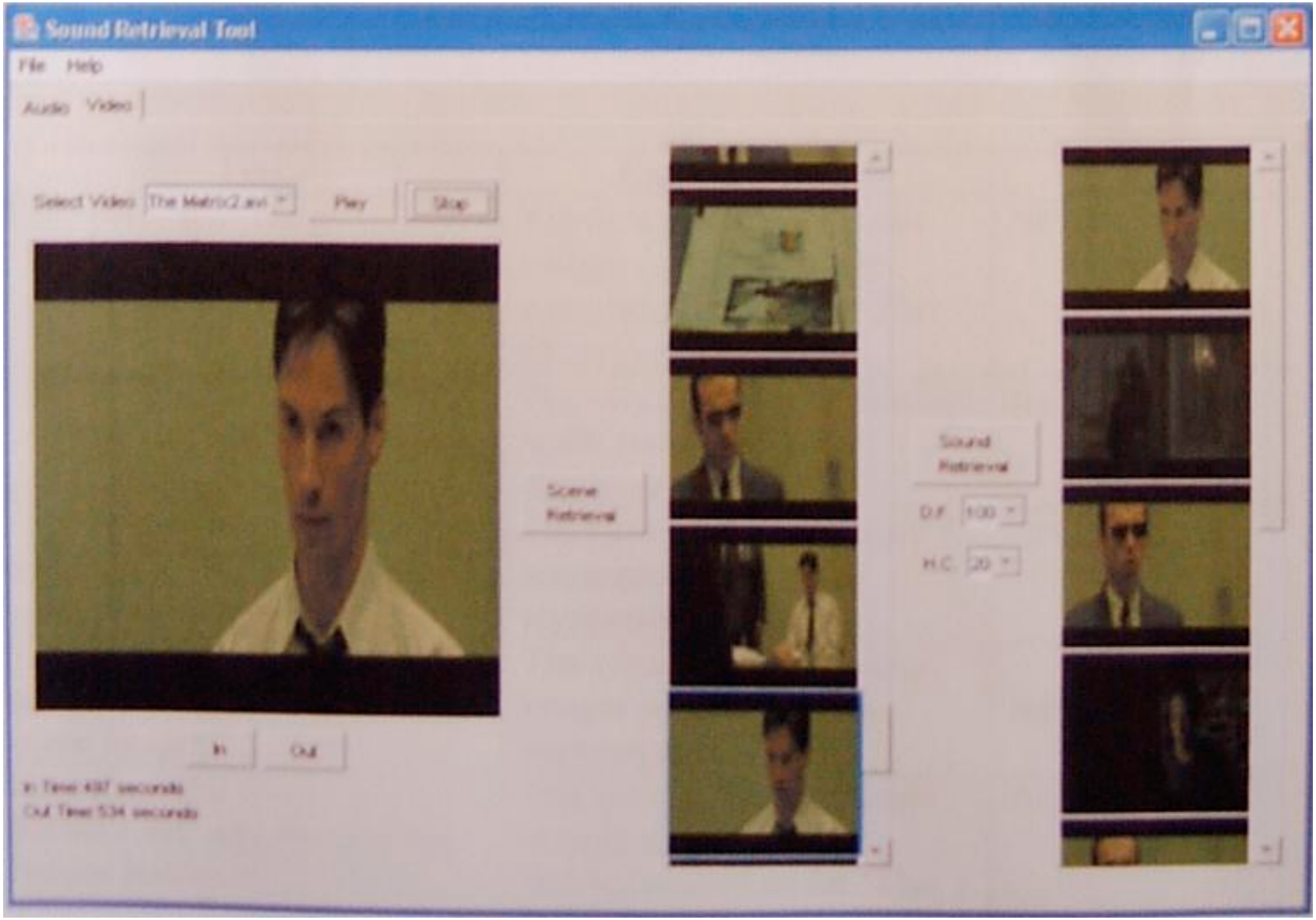
$$\mathrm{d}(a,b) = \left[\frac{\mathbf{a}^T \mathbf{b}}{\| \mathbf{a} \| \| \mathbf{b} \|} \right]$$

L L A M A S

LLAMA Service

- Portable light-weight GNU C/C++
- musicstructure.com uses LLAMAS
- video retrieval application with LLAMAS
- industry collaboration
- Multi platform:
  - Windows under Cygwin
  - .NET integration
  - JAVA Callable via JNI
  - MAC OSX version to run with XSERV
  - Callable as a Web Service
- http://musicstructure.com

Audio Based Video Retrieval
Michael Casey and Vikrant Ardhawa



LLAMAS for VIDEO Retrieval

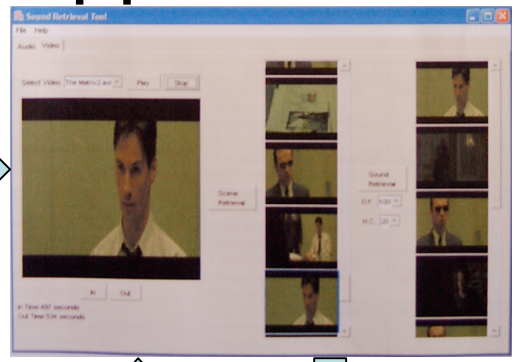
Acquire / Ingest

Segmenter

Encoded Media Store

Textual Metadata Store

Application



XML Result

XML Query

MPEG7 Encoder

Low-Level Metadata Store

Binary Vector Store

LLAMAS

Query / Retrieval

```
Media Unique ID (INT)
Media Locator (URI)
Media Time Locator (ASCII)
DescriptorAttributes (ASCII)
SeriesOfVectorsRef(URI)
```

Fast/Compact Database Access

- Covert XML to simple database schema
- BLOB fields
  - fast binary access
  - Structure determined by client applications
- SeriesOfVectors memory-mapped files
  - fast access via core image (seekable)
  - Use large segments for efficiency
    - » entire VOB file (DVD)
    - » Use Scene Segmentation Tool (IBM VideoAnnEx)

Example Database Schema

access unit

MediaSegmentTable

SegmentID	StartTime	MediaID
73	01:07:15.27	1
74	01:07:39.02	1
75	01:08:12.17	1
76	00:00:00.01	2

MediaTable

MediaID	MediaFile	AudioFeaturesID
1	Videos/TheMatrix1.vob	1171
2	Videos/TheMatrix2.vob	1172

AudioFeaturesTable

SeriesOfVectorsLocatorURI	Attributes	AudioFeaturesID
AudioFeatures/KC07040411.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1171
AudioFeatures/KC07040412.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1172

SeriesOfVectors (ASCII, XML or Big-Endien Binary File)

```
616.195 -0.898  0.325 -0.175  0.136 -0.104  0.082 -0.075  0.040 -0.058  0.041 -0.039  0.038 -0.014  0.037 -0.031  0.028 -0.030  0.020 -0.013  0.026
574.751 -0.888  0.323 -0.193  0.150 -0.110  0.070 -0.093  0.066 -0.062  0.033 -0.055  0.019 -0.029  0.041 -0.034  0.056 -0.017  0.027 -0.027  0.007
652.088 -0.874  0.358 -0.187  0.172 -0.105  0.086 -0.081  0.056 -0.052  0.044 -0.055  0.031 -0.029  0.029 -0.032  0.029 -0.031  0.017 -0.028  0.021
723.976 -0.892  0.336 -0.152  0.171 -0.089  0.083 -0.049  0.071 -0.058  0.060 -0.047  0.034 -0.031  0.024 -0.031  0.023 -0.036  0.015 -0.032  0.020
623.821 -0.881  0.345 -0.163  0.178 -0.094  0.095 -0.062  0.080 -0.054  0.054 -0.051  0.027 -0.039  0.012 -0.062  0.016 -0.042  0.016 -0.021  0.023
717.812 -0.883  0.346 -0.191  0.144 -0.105  0.075 -0.078  0.073 -0.055  0.053 -0.048  0.029 -0.022  0.033 -0.033  0.041 -0.023  0.019 -0.027  0.010
606.720 -0.892  0.331 -0.188  0.148 -0.107  0.054 -0.084  0.030 -0.058  0.046 -0.044  0.034 -0.021  0.034 -0.039  0.022 -0.047  0.013 -0.010  0.025
543.379 -0.896  0.303 -0.205  0.122 -0.125  0.050 -0.100  0.021 -0.032  0.067 -0.006  0.061  0.004  0.053 -0.048  0.010 -0.065  0.008  0.001  0.031
562.814 -0.886  0.328 -0.159  0.166 -0.129  0.066 -0.102  0.053 -0.066  0.026 -0.040  0.039 -0.011  0.062 -0.016  0.068 -0.040  0.013 -0.043 -0.009
502.962 -0.886  0.335 -0.196  0.140 -0.110  0.085 -0.086  0.055 -0.030  0.055 -0.033  0.042 -0.037  0.030 -0.029  0.051 -0.009  0.041 -0.033 -0.006
485.651 -0.864  0.364 -0.214  0.163 -0.106  0.089 -0.089  0.066 -0.050  0.049 -0.039  0.037 -0.031  0.049 -0.034  0.034 -0.037  0.023 -0.026  0.015
461.432 -0.863  0.367 -0.207  0.161 -0.117  0.089 -0.095  0.061 -0.043  0.068 -0.031  0.051 -0.037  0.033 -0.041  0.031 -0.025  0.027 -0.027  0.022
```

Locating feature segments in Scalable SeriesOfVectors

- Segments delimited by
  - unique media ID
  - unique start time with media file
- Feature access
  - D=decimation factor
  - Vectors sampled every D/100 seconds
  - feature time = ROW*(D/100)+StartTime
  - Memory mapped files are fastest
  - ASCII base 16 or base 64 encoding for constant field-width floating-point access

Database Schema

access unit

MediaSegmentTable

SegmentID	StartTime	MediaID
73	01:07:15.27	1
74	01:07:39.02	1
75	01:08:12.17	1
76	00:00:00.01	2

MediaTable

MediaID	MediaFile	AudioFeaturesID
1	Videos/TheMatrix1.vob	1171
2	Videos/TheMatrix2.vob	1172

AudioFeaturesTable

SeriesOfVectorsLocatorURI	Attributes	AudioFeaturesID
AudioFeatures/KC07040411.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1171
AudioFeatures/KC07040412.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1172

SeriesOfVectors (ASCII, XML or Big-Endien Binary File)

```
616.195 -0.898  0.325 -0.175  0.136 -0.104  0.082 -0.075  0.040 -0.058  0.041 -0.039  0.038 -0.014  0.037 -0.031  0.028 -0.030  0.020 -0.013  0.026
574.751 -0.888  0.323 -0.193  0.150 -0.110  0.070 -0.093  0.066 -0.062  0.033 -0.055  0.019 -0.029  0.041 -0.034  0.056 -0.017  0.027 -0.027  0.007
652.088 -0.874  0.358 -0.187  0.172 -0.105  0.086 -0.081  0.056 -0.052  0.044 -0.055  0.031 -0.029  0.029 -0.032  0.029 -0.031  0.017 -0.028  0.021
723.976 -0.892  0.336 -0.152  0.171 -0.089  0.083 -0.049  0.071 -0.058  0.060 -0.047  0.034 -0.031  0.024 -0.031  0.023 -0.036  0.015 -0.032  0.020
623.821 -0.881  0.345 -0.163  0.178 -0.094  0.095 -0.062  0.080 -0.054  0.054 -0.051  0.027 -0.039  0.012 -0.062  0.016 -0.042  0.016 -0.021  0.023
717.812 -0.883  0.346 -0.191  0.144 -0.105  0.075 -0.078  0.073 -0.055  0.053 -0.048  0.029 -0.022  0.033 -0.033  0.041 -0.023  0.019 -0.027  0.010
606.720 -0.892  0.331 -0.188  0.148 -0.107  0.054 -0.084  0.030 -0.058  0.046 -0.044  0.034 -0.021  0.034 -0.039  0.022 -0.047  0.013 -0.010  0.025
543.379 -0.896  0.303 -0.205  0.122 -0.125  0.050 -0.100  0.021 -0.032  0.067 -0.006  0.061  0.004  0.053 -0.048  0.010 -0.065  0.008  0.001  0.031
562.814 -0.886  0.328 -0.159  0.166 -0.129  0.066 -0.102  0.053 -0.066  0.026 -0.040  0.039 -0.011  0.062 -0.016  0.068 -0.040  0.013 -0.043 -0.009
502.962 -0.886  0.335 -0.196  0.140 -0.110  0.085 -0.086  0.055 -0.030  0.055 -0.033  0.042 -0.037  0.030 -0.029  0.051 -0.009  0.041 -0.033 -0.006
485.651 -0.864  0.364 -0.214  0.163 -0.106  0.089 -0.089  0.066 -0.050  0.049 -0.039  0.037 -0.031  0.049 -0.034  0.034 -0.037  0.023 -0.026  0.015
461.432 -0.863  0.367 -0.207  0.161 -0.117  0.089 -0.095  0.061 -0.043  0.068 -0.031  0.051 -0.037  0.033 -0.041  0.031 -0.025  0.027 -0.027  0.022
```

Database Schema

access unit

MediaSegmentTable

SegmentID	StartTime	MediaID
73	01:07:15.27	1
74	01:07:39.02	1
75	01:08:12.17	1
76	00:00:00.01	2

MediaTable

MediaID	MediaFile	AudioFeaturesID
1	Videos/TheMatrix1.vob	1171
2	Videos/TheMatrix2.vob	1172

AudioFeaturesTable

SeriesOfVectorsLocatorURI	Attributes	AudioFeaturesID
AudioFeatures/KC07040411.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1171
AudioFeatures/KC07040412.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1172

SeriesOfVectors (ASCII, XML or Big-Endien Binary File)

```
616.195 -0.898  0.325 -0.175  0.136 -0.104  0.082 -0.075  0.040 -0.058  0.041 -0.039  0.038 -0.014  0.037 -0.031  0.028 -0.030  0.020 -0.013  0.026
574.751 -0.888  0.323 -0.193  0.150 -0.110  0.070 -0.093  0.066 -0.062  0.033 -0.055  0.019 -0.029  0.041 -0.034  0.056 -0.017  0.027 -0.027  0.007
652.088 -0.874  0.358 -0.187  0.172 -0.105  0.086 -0.081  0.056 -0.052  0.044 -0.055  0.031 -0.029  0.029 -0.032  0.029 -0.031  0.017 -0.028  0.021
723.976 -0.892  0.336 -0.152  0.171 -0.089  0.083 -0.049  0.071 -0.058  0.060 -0.047  0.034 -0.031  0.024 -0.031  0.023 -0.036  0.015 -0.032  0.020
623.821 -0.881  0.345 -0.163  0.178 -0.094  0.095 -0.062  0.080 -0.054  0.054 -0.051  0.027 -0.039  0.012 -0.062  0.016 -0.042  0.016 -0.021  0.023
717.812 -0.883  0.346 -0.191  0.144 -0.105  0.075 -0.078  0.073 -0.055  0.053 -0.048  0.029 -0.022  0.033 -0.033  0.041 -0.023  0.019 -0.027  0.010
606.720 -0.892  0.331 -0.188  0.148 -0.107  0.054 -0.084  0.030 -0.058  0.046 -0.044  0.034 -0.021  0.034 -0.039  0.022 -0.047  0.013 -0.010  0.025
543.379 -0.896  0.303 -0.205  0.122 -0.125  0.050 -0.100  0.021 -0.032  0.067 -0.006  0.061  0.004  0.053 -0.048  0.010 -0.065  0.008  0.001  0.031
562.814 -0.886  0.328 -0.159  0.166 -0.129  0.066 -0.102  0.053 -0.066  0.026 -0.040  0.039 -0.011  0.062 -0.016  0.068 -0.040  0.013 -0.043 -0.009
502.962 -0.886  0.335 -0.196  0.140 -0.110  0.085 -0.086  0.055 -0.030  0.055 -0.033  0.042 -0.037  0.030 -0.029  0.051 -0.009  0.041 -0.033 -0.006
485.651 -0.864  0.364 -0.214  0.163 -0.106  0.089 -0.089  0.066 -0.050  0.049 -0.039  0.037 -0.031  0.049 -0.034  0.034 -0.037  0.023 -0.026  0.015
461.432 -0.863  0.367 -0.207  0.161 -0.117  0.089 -0.095  0.061 -0.043  0.068 -0.031  0.051 -0.037  0.033 -0.041  0.031 -0.025  0.027 -0.027  0.022
```

Database Schema

access unit

MediaSegmentTable

SegmentID	StartTime	MediaID
73	01:07:15.27	1
74	01:07:39.02	1
75	01:08:12.17	1
76	00:00:00.01	2

MediaTable

MediaID	MediaFile	AudioFeaturesID
1	Videos/TheMatrix1.vob	1171
2	Videos/TheMatrix2.vob	1172

AudioFeaturesTable

SeriesOfVectorsLocatorURI	Attributes	AudioFeaturesID
AudioFeatures/KC07040411.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1171
AudioFeatures/KC07040412.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1172

SeriesOfVectors (ASCII, XML or Big-Endien Binary File)

```
616.195 -0.898  0.325 -0.175  0.136 -0.104  0.082 -0.075  0.040 -0.058  0.041 -0.039  0.038 -0.014  0.037 -0.031  0.028 -0.030  0.020 -0.013  0.026
574.751 -0.888  0.323 -0.193  0.150 -0.110  0.070 -0.093  0.066 -0.062  0.033 -0.055  0.019 -0.029  0.041 -0.034  0.056 -0.017  0.027 -0.027  0.007
652.088 -0.874  0.358 -0.187  0.172 -0.105  0.086 -0.081  0.056 -0.052  0.044 -0.055  0.031 -0.029  0.029 -0.032  0.029 -0.031  0.017 -0.028  0.021
723.976 -0.892  0.336 -0.152  0.171 -0.089  0.083 -0.049  0.071 -0.058  0.060 -0.047  0.034 -0.031  0.024 -0.031  0.023 -0.036  0.015 -0.032  0.020
623.821 -0.881  0.345 -0.163  0.178 -0.094  0.095 -0.062  0.080 -0.054  0.054 -0.051  0.027 -0.039  0.012 -0.062  0.016 -0.042  0.016 -0.021  0.023
717.812 -0.883  0.346 -0.191  0.144 -0.105  0.075 -0.078  0.073 -0.055  0.053 -0.048  0.029 -0.022  0.033 -0.033  0.041 -0.023  0.019 -0.027  0.010
606.720 -0.892  0.331 -0.188  0.148 -0.107  0.054 -0.084  0.030 -0.058  0.046 -0.044  0.034 -0.021  0.034 -0.039  0.022 -0.047  0.013 -0.010  0.025
543.379 -0.896  0.303 -0.205  0.122 -0.125  0.050 -0.100  0.021 -0.032  0.067 -0.006  0.061  0.004  0.053 -0.048  0.010 -0.065  0.008  0.001  0.031
562.814 -0.886  0.328 -0.159  0.166 -0.129  0.066 -0.102  0.053 -0.066  0.026 -0.040  0.039 -0.011  0.062 -0.016  0.068 -0.040  0.013 -0.043 -0.009
502.962 -0.886  0.335 -0.196  0.140 -0.110  0.085 -0.086  0.055 -0.030  0.055 -0.033  0.042 -0.037  0.030 -0.029  0.051 -0.009  0.041 -0.033 -0.006
485.651 -0.864  0.364 -0.214  0.163 -0.106  0.089 -0.089  0.066 -0.050  0.049 -0.039  0.037 -0.031  0.049 -0.034  0.034 -0.037  0.023 -0.026  0.015
461.432 -0.863  0.367 -0.207  0.161 -0.117  0.089 -0.095  0.061 -0.043  0.068 -0.031  0.051 -0.037  0.033 -0.041  0.031 -0.025  0.027 -0.027  0.022
```

Database Schema: Scene 74

access unit

MediaSegmentTable

SegmentID	StartTime (HH:MM:SS)	MediaID
73	01:07:15	1
74	01:07:39	1
75	01:08:12	1
76	00:00:00	2

MediaTable

MediaID	MediaFile	AudioFeaturesID
1	Videos/TheMatrix1.vob	1171
2	Videos/TheMatrix2.vob	1172

AudioFeaturesTable

SeriesOfVectorsLocatorURI	Attributes	AudioFeaturesID
AudioFeatures/KC07040411.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1171
AudioFeatures/KC07040412.bin	LoEdge,439,ScaleRatio,25,ScalingMethod,Mean	1172

SeriesOfVectors (ASCII, XML or Big-Endien Binary File)

```
616.195 -0.898  0.325 -0.175  0.136 -0.104  0.082 -0.075  0.040 -0.058  0.041 -0.039  0.038 -0.014  0.037 -0.031  0.028 -0.030  0.020 -0.013  0.026
574.751 -0.888  0.323 -0.193  0.150 -0.110  0.070 -0.093  0.066 -0.062  0.033 -0.055  0.019 -0.029  0.041 -0.034  0.056 -0.017  0.027 -0.027  0.007
652.088 -0.874  0.358 -0.187  0.172 -0.105  0.086 -0.081  0.056 -0.052  0.044 -0.055  0.031 -0.029  0.029 -0.032  0.029 -0.031  0.017 -0.028  0.021
723.976 -0.892  0.336 -0.152  0.171 -0.089  0.083 -0.049  0.071 -0.058  0.060 -0.047  0.034 -0.031  0.024 -0.031  0.023 -0.036  0.015 -0.032  0.020
633.831 -0.881  0.345 -0.163  0.178 -0.094  0.095 -0.063  0.080 -0.054  0.054 -0.051  0.037 -0.039  0.013 -0.063  0.016 -0.043  0.016 -0.031  0.023
717.812 -0.883  0.346 -0.191  0.144 -0.105  0.075 -0.078  0.073 -0.055  0.053 -0.048  0.029 -0.022  0.033 -0.033  0.041 -0.023  0.019 -0.027  0.010
606.720 -0.892  0.331 -0.188  0.148 -0.107  0.054 -0.084  0.030 -0.058  0.046 -0.044  0.034 -0.021  0.034 -0.039  0.022 -0.047  0.013 -0.010  0.025
543.379 -0.896  0.303 -0.205  0.122 -0.125  0.050 -0.100  0.021 -0.032  0.067 -0.006  0.061  0.004  0.053 -0.048  0.010 -0.065  0.008  0.001  0.031
562.814 -0.886  0.328 -0.159  0.166 -0.129  0.066 -0.102  0.053 -0.066  0.026 -0.040  0.039 -0.011  0.062 -0.016  0.068 -0.040  0.013 -0.043 -0.009
502.962 -0.886  0.335 -0.196  0.140 -0.110  0.085 -0.086  0.055 -0.030  0.055 -0.033  0.042 -0.037  0.030 -0.029  0.051 -0.009  0.041 -0.033 -0.006
485.651 -0.864  0.364 -0.214  0.163 -0.106  0.089 -0.089  0.066 -0.050  0.049 -0.039  0.037 -0.031  0.049 -0.034  0.034 -0.037  0.023 -0.026  0.015
461.432 -0.863  0.367 -0.207  0.161 -0.117  0.089 -0.095  0.061 -0.043  0.068 -0.031  0.051 -0.037  0.033 -0.041  0.031 -0.025  0.027 -0.027  0.022
```

01:07:39

01:08:12

Scalability

- Scale-Ratio [10,25,50,100]
- <SeriesOfVecors> <mean>
  - moving average of SeriesOfVectors
- Performance
  - retrieval results for video
  - retrieval results for music
  - retrieval results for speech

Application: Video Retrieval using Audio

- Input: XML audio metadata query
- LLAMAS accepts XML query
- Uses LLAMAS to query database for similar features
- LLAMAS returns a FileLocator and MediaTime start/stop points in a result list
- Test tasks:
  - actor-based scene retrieval
    - Theo, Trinity, Morpheus
  - *action scene retrieval*

LLAMAS for VIDEO Retrieval

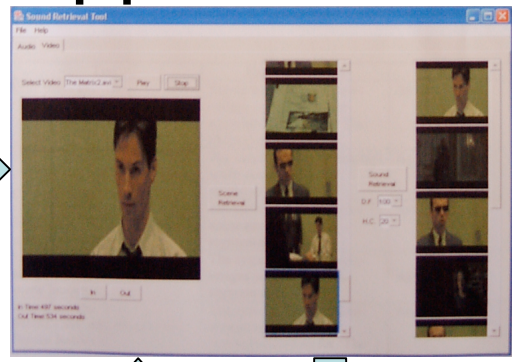
Acquire

Segmenter

Encoded Media Store

Textual Metadata Store

Application



XML Result

XML Query

MPEG7 Encoder

Low-Level Metadata Store

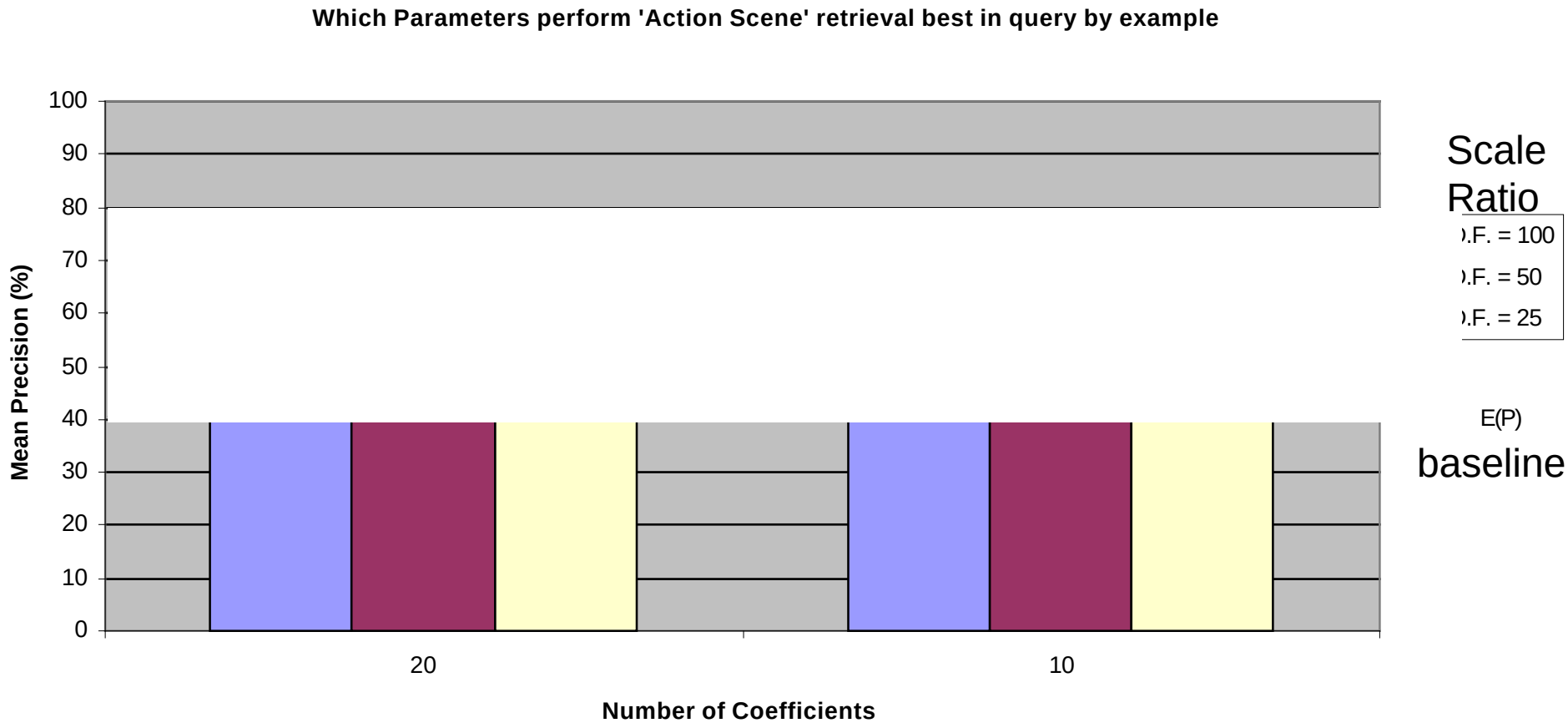
Binary Vector Store

LLAMAS

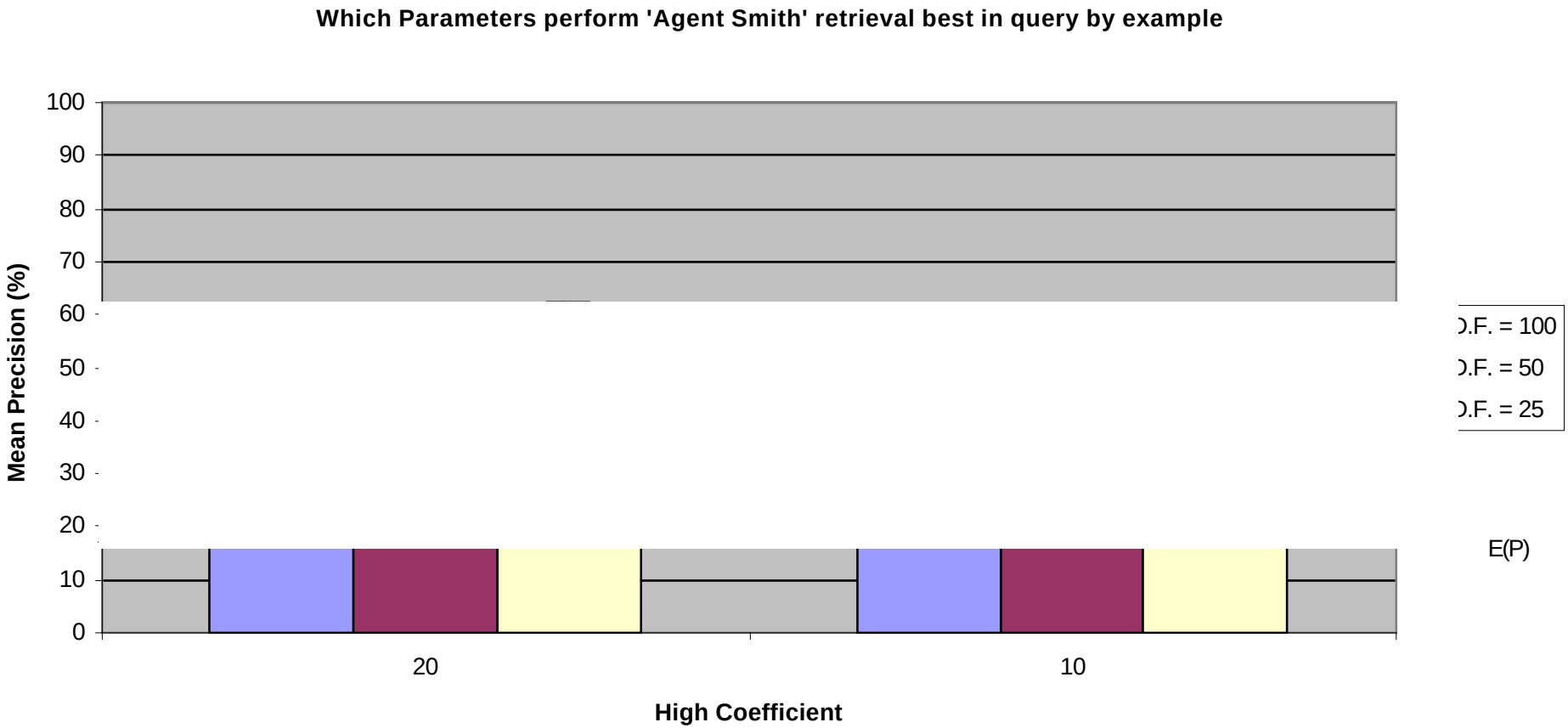
Query / Retrieval

```
Media Unique ID (INT)
Media Locator (URI)
Media Time Locator (ASCII)
DescriptorAttributes (ASCII)
SeriesOfVectorsRef(URI)
```

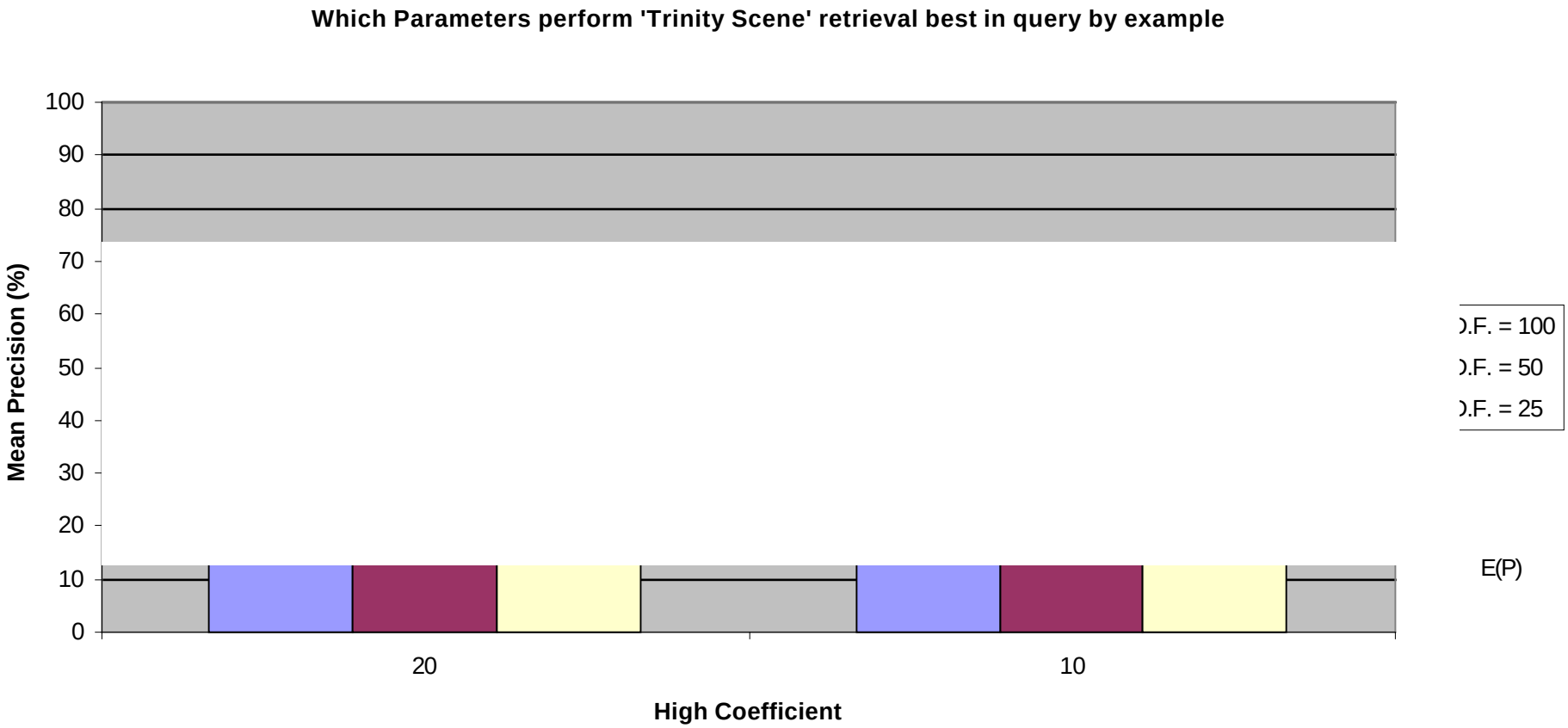
Action Scene Retrieval Scalable Descriptor



Character Retrieval: “Agent Smith”



Character Retrieval: “Trinity”



Summary

- Audio-based information retrieval
- LLAMAS low-level audio management
- Video retrieval by audio
- Database schemas
- Scalability of features
- Performance Evaluation