



Content Based Retrieval of Music Videos

Sumandeep Banerjee
Indian Institute of Technology, Kharagpur
Intern - DFKI

Find Sounds

FindSounds

Search the Web for Sounds

Search for [Help](#)
[Need Examples?](#)

| File Formats | Number of Channels | Minimum Resolution | Minimum Sample Rate | Maximum File Size |
|--|--|--|--|---------------------------------------|
| <input checked="" type="checkbox"/> AIFF | <input checked="" type="checkbox"/> mono | 8-bit <input type="button" value="v"/> | 8000 Hz <input type="button" value="v"/> | 2 MB <input type="button" value="v"/> |
| <input checked="" type="checkbox"/> AU | <input checked="" type="checkbox"/> stereo | | | |
| <input checked="" type="checkbox"/> WAVE | | | | |

What types of sounds can be found on the Web using FindSounds? Below is a partial list. Click on any link below to perform a search, or enter one or more words in the search box above and then click on the Search button.

Animals [alligator](#), [baboon](#), [bat](#), [bear](#), [buffalo](#), [calf](#), [camel](#), [cat](#), [cheetah](#), [chimpanzee](#), [chinchilla](#), [chipmunk](#), [cougar](#), [cow](#), [coyote](#), [deer](#), [dinosaur](#), [dog](#), [dolphin](#), [donkey](#), [elephant](#), [elk](#), [fox](#), [frog](#), [gibbon](#), [goat](#), [gorilla](#), [guinea pig](#), [hippo](#), [horse](#), [hyena](#), [jaguar](#), [kitten](#), [lamb](#), [lemur](#), [leopard](#), [lion](#), [llama](#), [lynx](#), [marmot](#), [monkey](#), [mouse](#), [orca](#), [panda](#), [panther](#), [pig](#), [prairie dog](#), [puppy](#), [raccoon](#), [rat](#), [rattlesnake](#), [rhinoceros](#), [sea lion](#), [seal](#), [sheep](#), [snake](#), [squirrel](#), [tiger](#), [toad](#), [whale](#), [wolf](#), [zebra](#)

Birds [blackbird](#), [blue jay](#), [bluebird](#), [bobwhite](#), [budgie](#), [bunting](#), [canary](#), [cardinal](#), [catbird](#), [chick](#), [chickadee](#), [chicken](#), [coot](#), [cowbird](#), [crane](#), [crow](#), [dove](#), [duck](#), [eagle](#), [falcon](#), [finch](#), [flamingo](#), [flicker](#), [flycatcher](#), [goldfinch](#), [goose](#), [grackle](#), [grebe](#), [grosbeak](#), [grouse](#), [gull](#), [hawk](#), [heron](#), [hummingbird](#), [jay](#), [junco](#), [kestrel](#), [killdeer](#), [kingbird](#), [kingfisher](#), [kinglet](#), [kookaburra](#), [lark](#), [loon](#), [macaw](#), [magpie](#), [mallard](#), [martin](#), [meadowlark](#), [mockingbird](#), [mynah](#), [nightingale](#), [nuthatch](#), [oriole](#), [osprey](#), [ovenbird](#), [owl](#), [parrot](#), [parula](#), [peacock](#), [pheasant](#), [pigeon](#), [quail](#), [raven](#), [redstart](#), [robin](#), [rooster](#), [sandpiper](#), [sapsucker](#), [seagull](#), [siskin](#), [skylark](#), [sparrow](#), [starling](#), [swallow](#), [swan](#), [tanager](#), [tern](#), [thrasher](#), [thrush](#), [towhee](#), [turkey](#), [vireo](#), [vulture](#), [warbler](#), [waxwing](#), [whippoorwill](#), [woodpecker](#), [wren](#)

Holidays Christmas [sleigh bells](#); Halloween [creak](#), [evil laugh](#), [ghost](#), [howl](#), [scream](#), [witch](#); Independence [firecrackers](#), [fireworks](#); New Year's [balloon](#), [party horn](#), [pop cork](#); Thanksgiving [turkey](#)

Household [blender](#), [blinds](#), [boiling](#), [bottle](#), [bubbles](#), [can](#), [clock](#), [coffee](#), [cork](#), [dishes](#), [door](#), [door bell](#), [drain](#), [drawer](#), [drip](#), [dryer](#), [fire extinguisher](#), [gate](#), [ice cubes](#), [kettle](#), [keys](#), [latch](#), [lighter](#), [lock](#), [match](#), [oven](#), [phone](#), [pot](#), [pour liquid](#), [scissors](#), [soda](#), [spray](#), [straw](#), [switch](#), [tape](#), [timer](#), [toaster](#), [toothbrush](#), [toy](#), [trash](#), [vacuum cleaner](#), [washing machine](#), [window](#), [zipper](#)

Insects [bee](#), [cicada](#), [cricket](#), [fly](#), [katydid](#), [mosquito](#), [wasp](#)

Find Sounds – Search results

FindSounds

Search the Web for Sounds

Search for [Help](#)

[Need Examples?](#)

| File Formats | Number of Channels | Minimum Resolution | Minimum Sample Rate | Maximum File Size |
|--|--|------------------------------------|--------------------------------------|-----------------------------------|
| <input checked="" type="checkbox"/> AIFF | <input checked="" type="checkbox"/> mono | <input type="text" value="8-bit"/> | <input type="text" value="8000 Hz"/> | <input type="text" value="2 MB"/> |
| <input checked="" type="checkbox"/> AU | <input checked="" type="checkbox"/> stereo | | | |
| <input checked="" type="checkbox"/> WAVE | | | | |

Sounds 1-7 of 7 labelled "robot"



 <http://jupiter2.freeyellow.com/root/laugh.wav>
robot's laugh
179k, mono, 8-bit, 22050 Hz, 8.3 seconds ([show page](#) | [e-mail this sound](#))



 <http://www.buffaloshideout.com/wavs/robots.wav>
robot
60k, mono, 8-bit, 11025 Hz, 5.6 seconds ([show page](#) | [e-mail this sound](#))



 <http://www.paulies-world.com/OLD-ZONE-ACCOUNT/MIROM/ROBOTS/robot.wav>
robot
254k, mono, 16-bit, 44100 Hz, 3.0 seconds ([show page](#) | [e-mail this sound](#))



 http://saveandromeda.com/tarantulas/questusers/user03/robot_b9/robot_laugh.wav
Lost in Space: robot's laugh
50k, mono, 8-bit, 22050 Hz, 2.3 seconds ([show page](#) | [e-mail this sound](#))



 <http://www.lostinspace.tv.com/ART/download/Robot/RobotLaugh.aif>
robot laugh
66k, mono, 8-bit, 22050 Hz, 3.1 seconds ([e-mail this sound](#))

Google videos

Google
Video BETA

[New! Upload and share your own videos](#)

[Top 100](#) [Comedy](#) [Music videos](#) [Movies](#) [Sports](#) [Animation](#) [TV shows](#) [Google Picks](#)

 **New! See your own videos on Google Video.**
 Share it with others instantly.

FREE TODAY

[More »](#)

You can now watch selected videos for free - thanks to our sponsors. These videos usually sell from \$0.99 to \$14.99.

[See hundreds more »](#)

 The
Documentary
Channel

[Documentary Channel](#)



[Music News](#)



[Charlie Rose](#)

Popular

[RSS](#) - [Top 100](#) - [New! Movers & Shakers »](#)



[hi Jack](#)
13 sec



[President Bush
Impersonati...](#)
★★★★★
11 min 24 sec



[Girl Caught
Cheating Remix](#)
★★★★★
20 sec



[The Evolution of
Dance](#)
★★★★★
5 min 59 sec



[Amazing Juggling
Finale](#)
★★★★★
4 min 27 sec



[aRCHeRYs
PaRTY WoLuM...](#)
★★★★★
4 min 53 sec



[nike](#)
★★★★★
4 min 28 sec



[R. VS. I](#)
★★★★★
1 min 26 sec



[Warcraft III and
Google Eart...](#)
★★★★★
3 min 59 sec



[The reason](#)
★★★★★
4 min 13 sec



[Shakira - Hips
Don't Lie](#)
★★★★★
3 min 42 sec



[Alizee - Shake](#)
★★★★★
3 min 8 sec












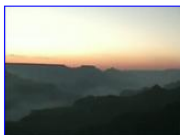
Google videos – Search results

sumandeep.banerjee@gmail.com | [Purchased Videos](#) | [Uploaded Videos](#) | [My Account](#) | [Sign Out](#)

Google Video BETA [New!](#) [Upload and share your own videos](#)

[Top 100](#) | [Comedy](#) | [Music videos](#) | [Movies](#) | [Sports](#) | [Animation](#) | [TV shows](#) | [Google Picks](#)

Results for sunrise All prices ▾ All durations ▾ Sort by relevance ▾ 1 - 18 of about 693 (0.00 sec)

| | | |
|--|--|--|
|  <p>Sunrise / Back to the Future Festival 1989 [jloveparade] 4 min 31 sec Sunrise / Back to the Future Festival 1989 Longwick UK</p> |  <p>Pan across sunrise behind Monument Valley, Arizona, Utah, southwest www.dvarchive.com 15 sec To download and license this royalty free clip, please go to..</p> |  <p>Haleakala Sunrise Artifact Studios 20 sec high speed sunrise from the 10,000 ft. summit of Haleakala o..</p> |
|  <p>Sunrise ★★★★★ (1 rating) nme 11 sec cool sunrise vid</p> |  <p>Brits Sunrise 2003 Pautz Productions Pretoria 43 sec Sunrise over the Kranskop Gliding Club at Brits Airfield (S2)..</p> |  <p>Bali Hai Sunrise (Timelapse) Graham Curran 30 sec Sunrise over Bali Hai (on island of Kauai) during my vacatio..</p> |
|  <p>Duluth Sunrise time-lapse The Ravenwood Gang 33 sec This short video is mostly many shots from the same angle as..</p> |  <p>Timelapse Sunrise Alan Chan 18 sec sunrise over Culver City, CA Canon Digital Rebel EOS, 210mm..</p> |  <p>Jason Engelstein Sunrise in the Philippines Jason Engelstein 1 min 14 sec After taking a night bus from Baguio to Manila, getting some..</p> |
|  <p>sailing into the sunrise on frances fleet fishing, rhode island 01/28/06 ★★★★★ (1 rating) william pease 42 sec we left a 5am to go fishing on the frances fleet. seas were ..</p> |  <p>'Sunrise Tai Chi' DVD preview: perfect for beginners, YMAA Taijiquan YMAA Publication Center 2 min 35 sec Free preview. 4-hour DVD produced in high-definition. Tai..</p> |  <p>Grand Canyon Sunrise Santoor Shiv kumar Sharma Milind Makwana 4 min 55 sec This video of Sun rise was taken during our July 2005 trip t..</p> |



Content based retrieval of music videos

- Music videos
 - Relevant information in the audio track
- Only audio tracks are used for analysis
- Query results
 - Matching segments of songs (not just the entire song)



Content based retrieval of music videos

- Offline processing
 - Audio track extraction from videos
 - Feature extraction from the audio tracks
 - Indexing
- Query processing
 - Search algorithm
 - Ranked list of hits



Audio track extraction

- Done using Media Coder
 - <http://www.rarewares.org/mediacoder/>
- Open source software for transcoding
- Videos are transcoded to the audio track only



Feature extraction

- Mel Frequency Cepstral Coefficients (MFCC)
 - Signal sampling rate : 44.1KHz
 - Coefficients : 13
 - Window function : Hamming
 - Frame size : 512 (11ms)
 - Overlap : 256 (6ms)
 - Coefficients downsampled to 250 ms



Indexing mechanism

- Very large number of frames in the database
 - $\sim 10^5$ for 65 videos from our database
- Matching a query to retrieve segments of songs requires us to slide a window
 - Computationally expensive



Indexing mechanism

- Idea: Cluster the MFCC frames
 - K-means
- Search for potential matches in the cluster centres
- Advantage: Reduces the search complexity to sub-linear depending on the number of clusters



Algorithm - QueryDB

Input : Query i.e., audio signal in time domain

Step 1: Compute MFCCs for the query

Step 2: Do for each MFCC frame

- a. Find nearest cluster center
- b. Cluster members are potential hits
- c. Collect all cluster members

Step 3: Group the hits (from step 2) by their song ID

Step 4: Do for each song

- a. Identify contiguous / semi-contiguous segments of the song
- b. Apply a smoothing function to the segments

Step 5: Rank the identified segments using the Hamming distance between the original and the smoothed version of the segments

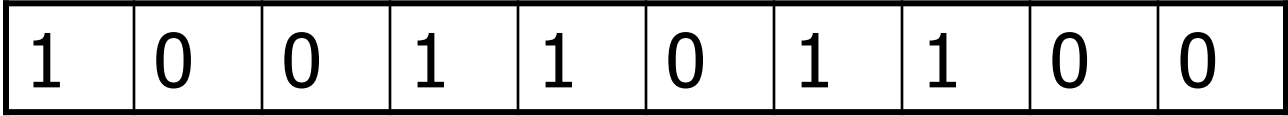
Output: Ranked list of song segments matching the query



Song Representation

Start frame

End frame



| | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|---|
| 1 | | | | | | | | | 0 |
|---|--|--|--|--|--|--|--|--|---|

Hits



Smoothing the frame sequence

Frame
sequence

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|

AR-LPF

Smoothed
sequence

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|

Start time

End time



Ranked list of hits

- Remove outliers in the frame sequence
 - Apply auto regressive low pass filter
 - AR- LPF : Triangular, order = 11
- Measure smoothness of frame sequence
 - Normalized Hamming distance between raw sequence and smoothened sequence
- Lower the normalized Hamming distance, better the match



Typical query result

| Rank | Normalized Hamming Distance | Matching segment length (Sec) | Song ID | Start position (Sec) | End position (Sec) |
|------|-----------------------------|-------------------------------|-----------|----------------------|--------------------|
| 1 | 6 | 42 | 37 Heartb | 14 | 56 |
| 2 | 16 | 71 | 07 Smile. | 37 | 108 |
| 3 | 16 | 34 | 07 Smile. | 199 | 234 |
| 4 | 19 | 24 | 18 Leaf H | 6 | 30 |
| 5 | 19 | 17 | 21 Us.wav | 82 | 99 |
| 6 | 19 | 27 | 36 Breath | 193 | 220 |
| 7 | 19 | 17 | 63 New Sl | 406 | 423 |
| 8 | 21 | 42 | 25 Your E | 41 | 83 |
| 9 | 21 | 27 | 36 Breath | 162 | 190 |
| 10 | 21 | 28 | 49 Going | 253 | 282 |
| 11 | 23 | 19 | 10b Hitch | 270 | 289 |
| 12 | 23 | 17 | 53 Mushab | 257 | 274 |
| 13 | 24 | 36 | 07 Smile. | 162 | 198 |
| 14 | 26 | 18 | 32 Walk A | 233 | 252 |
| 15 | 31 | 21 | 37 Heartb | 130 | 151 |

Table 1. Query result for a typical query of 15 second snippet from song 37. Query time was 0.968 seconds



Internship - Contributions

- Duration : 8 weeks
- Developed independent modules
 - Extraction of MFCCs (MARSYAS)
 - K-means clustering
 - Proposed QueryDB algorithm
- All modules written in C++
- Detailed documentation is available
- Internship report



Internship - Contributions

- Sub modules developed
 - IIR Filtering
 - Windowing
 - FFT
 - WAVE read / write
 - Quick Sort
- All main and sub modules can be used independently for other projects



Conclusion and Further work

- Sub-linear complexity of query matching algorithm
- Satisfactory result retrieval
- Further work
 - Optimal selection of number of clusters
 - Varying of AR-LPF shape and order
 - GUI Interface for the modules



Thank you

Questions ?