On Communicating Computational Research

Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio

Dept. Electrical Eng., Columbia Univ., NY USA

dpwe@ee.columbia.edu http://labrosa.ee.columbia.edu/

1. Computational Research
2. Five Ways to Present Results
3. Sharing Code
4. Conclusions

Lab ROSA

Laboratory for the Recognition and Organization of Speech and Audio

Columbia University in the City of New York

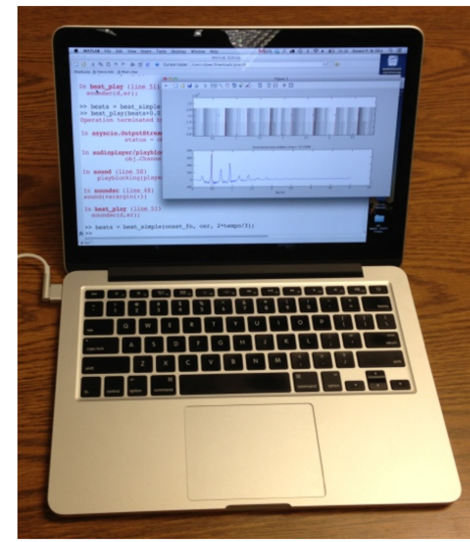
1. Computational Research

- 50 years of computational research (in SP):

1963

1988

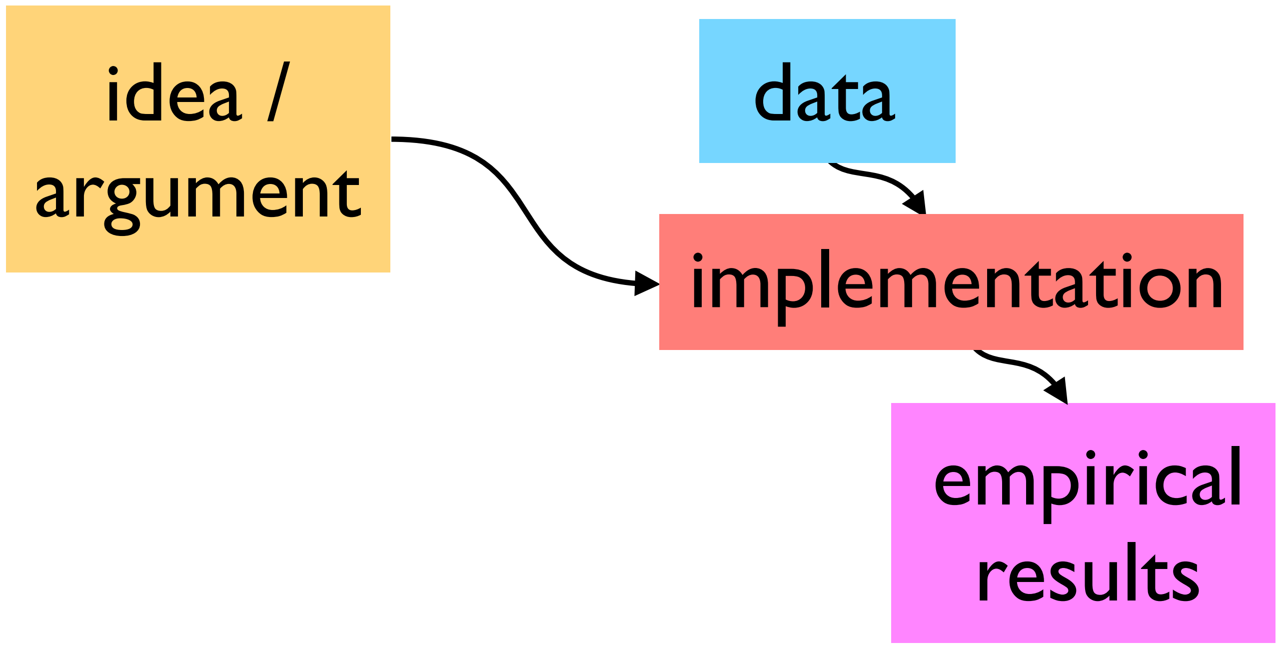
2013



Accessible..	Within lab	Researchers	Anyone
Coding time (FFT)	1 month	1 day	<1 minute

The Paradigm

- Computational approaches can be very complex
- Proposed techniques are often elude theoretical analysis
- Empirical results are “the proof of the pudding”

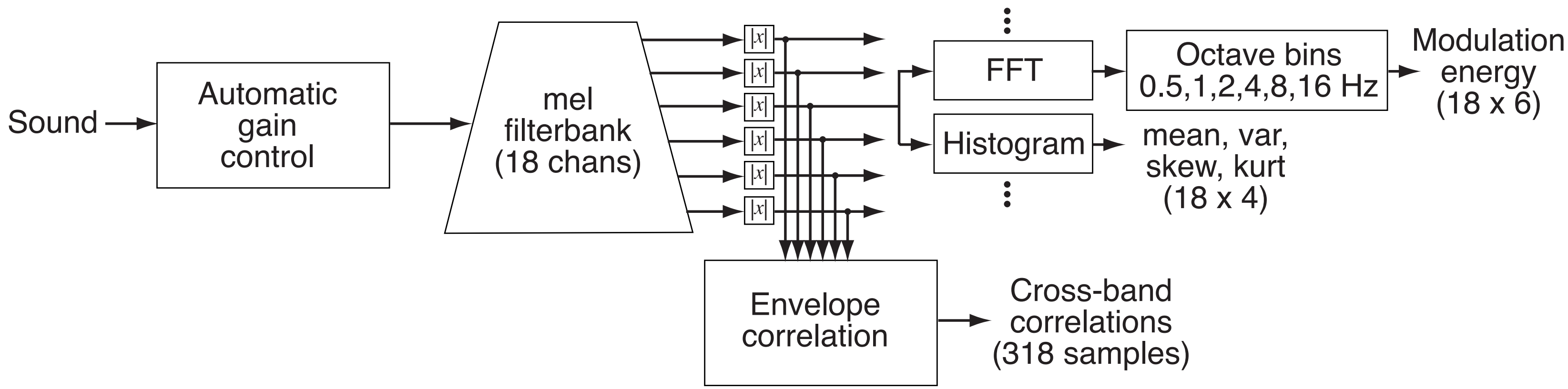


- statistical
- comparative

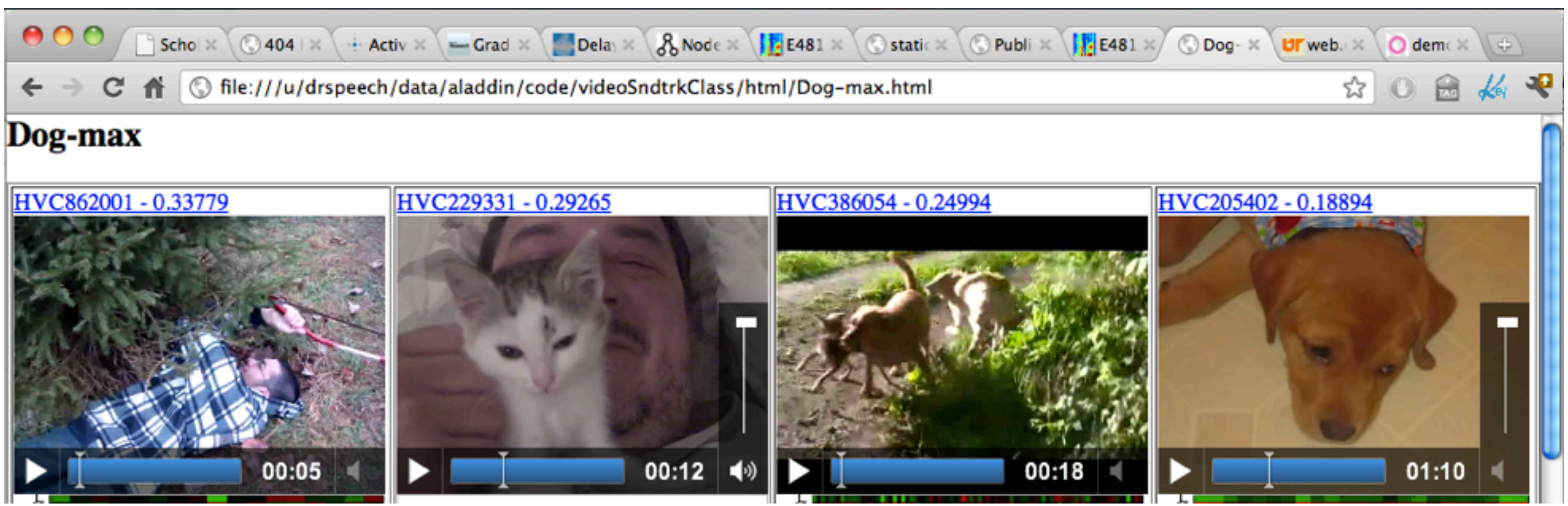
Example: Soundtrack Classification

Ellis, Zheng, McDermott '11

- Trained models using "texture" features



- Results on 9,317 videos (210 hours)

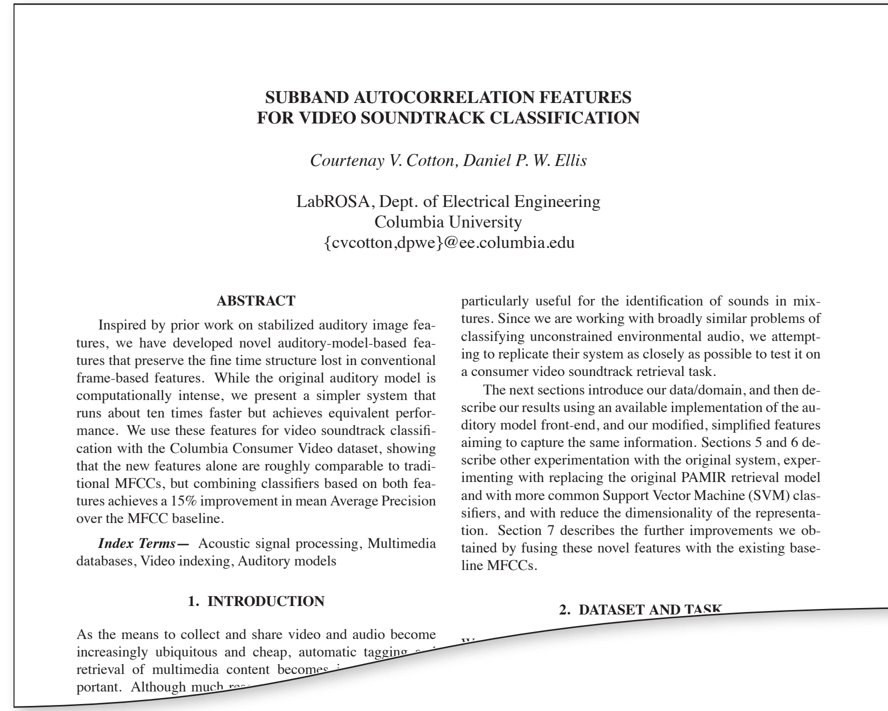


Presenting Results

- Traditional paper:
  - Compute times:

Table 1. Comparison of feature properties. Calculation times are over the 210 h CCV data set on a single CPU.

	MFCC	SAI (reduced)	SBPCA
Feature extraction	5.6 h	1087 h	310 h
Feature/patch dims	60	48	60
# patches/codebooks	1	24 (8)	4
Codebook size	3000	1000	1000
Histogram size	3000	24000 (8000)	4000



SUBBAND AUTOCORRELATION FEATURES
FOR VIDEO SOUNDTRACK CLASSIFICATION

Courtenay V. Cotton, Daniel P. W. Ellis

LabROSA, Dept. of Electrical Engineering
Columbia University
{cvcotton,dpwe}@ee.columbia.edu

ABSTRACT

Inspired by prior work on stabilized auditory image features, we have developed novel auditory-model-based features that preserve the fine time structure lost in conventional frame-based features. While the original auditory model is computationally intense, we present a simpler system that runs about ten times faster but achieves equivalent performance. We use these features for video soundtrack classification with the Columbia Consumer Video dataset, showing that the new features alone are roughly comparable to traditional MFCCs, but combining classifiers based on both features achieves a 15% improvement in mean Average Precision over the MFCC baseline.

Index Terms— Acoustic signal processing, Multimedia databases, Video indexing, Auditory models

1. INTRODUCTION

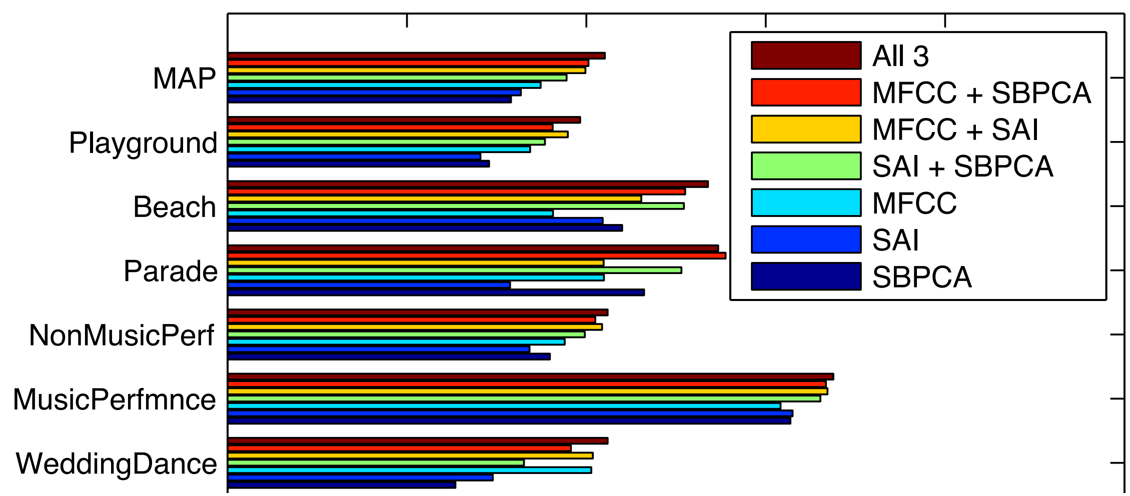
As the means to collect and share video and audio become increasingly ubiquitous and cheap, automatic tagging and retrieval of multimedia content becomes i... portant. Although much re...

particularly useful for the identification of sounds in mixtures. Since we are working with broadly similar problems of classifying unconstrained environmental audio, we attempting to replicate their system as closely as possible to test it on a consumer video soundtrack retrieval task.

The next sections introduce our data/domain, and then describe our results using an available implementation of the auditory model front-end, and our modified, simplified features aiming to capture the same information. Sections 5 and 6 describe other experimentation with the original system, experimenting with replacing the original PAMIR retrieval model and with more common Support Vector Machine (SVM) classifiers, and with reduce the dimensionality of the representation. Section 7 describes the further improvements we obtained by fusing these novel features with the existing baseline MFCCs.

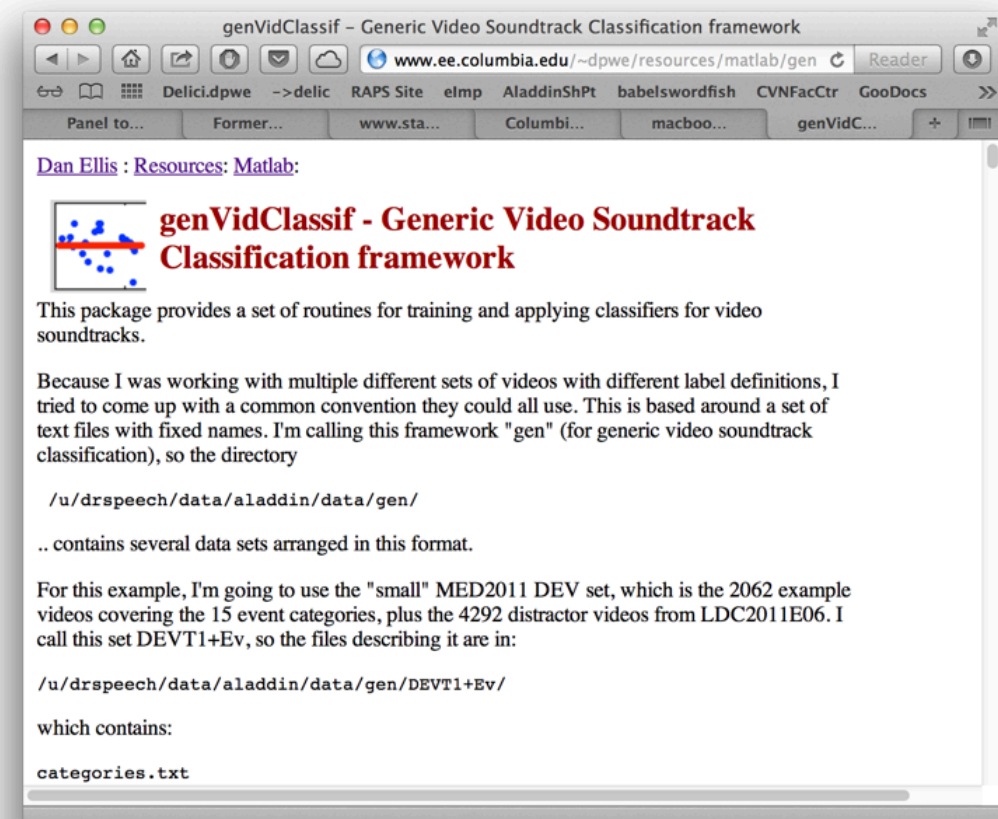
2. DATASET AND TASK

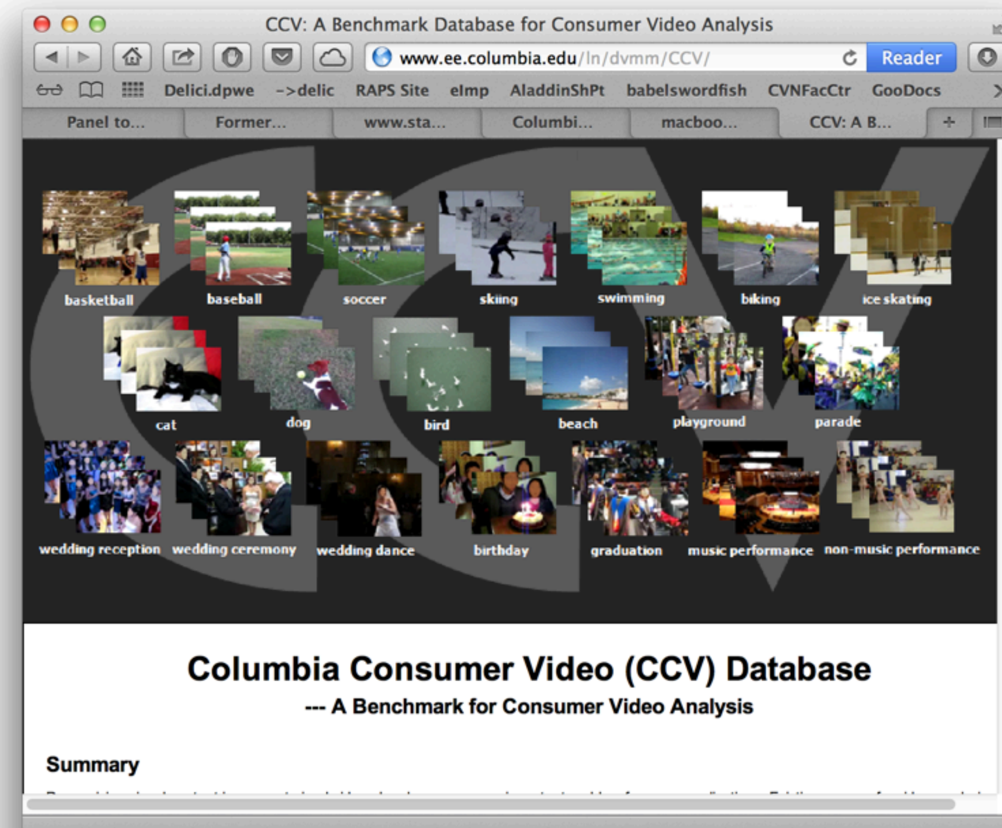
  - Results:



Presenting Results

- Code & Data release
  - ~5000 lines of Matlab
  - ~5 GB of data





2. Five Ways to Present Results

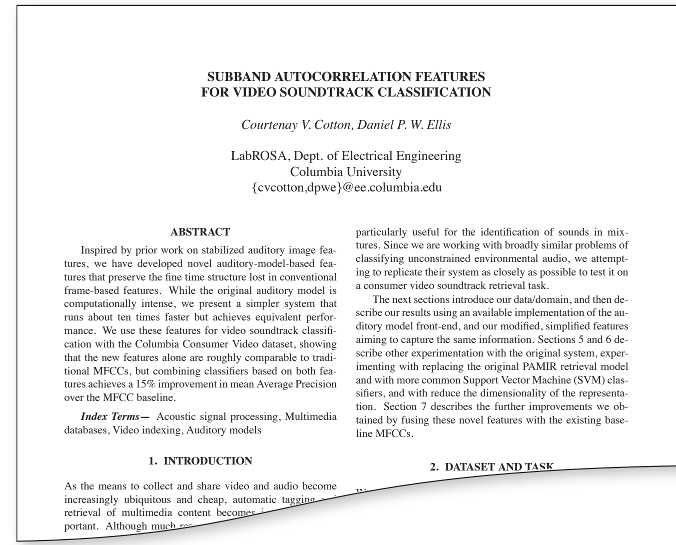
- Researchers want to present their results
  - "publish or perish"
  - paper & citation counts
  - impact & fame

- Math/Humanities model
  - the paper is the product

- Science/Engineering model
  - the paper describes the product

Five Ways...

1. **Traditional Publications**
   - pro: Present your "spin"
   - con: Not the whole story



SUBBAND AUTOCORRELATION FEATURES
FOR VIDEO SOUNDTRACK CLASSIFICATION

Courtenay V. Cotton, Daniel P. W. Ellis

LabROSA, Dept. of Electrical Engineering
Columbia University
{cvcotton,dpwe}@ee.columbia.edu

ABSTRACT

Inspired by prior work on stabilized auditory image features, we have developed novel auditory-model-based features that preserve the fine time structure lost in conventional frame-based features. While the original auditory model is computationally intense, we present a simpler system that runs about ten times faster but achieves equivalent performance. We use these features for video soundtrack classification with the Columbia Consumer Video dataset, showing that the new features alone are roughly comparable to traditional MFCCs, but combining classifiers based on both features achieves a 15% improvement in mean Average Precision over the MFCC baseline.

Index Terms— Acoustic signal processing, Multimedia databases, Video indexing, Auditory models

1. INTRODUCTION

As the means to collect and share video and audio become increasingly ubiquitous and cheap, automatic tagging ... retrieval of multimedia content becomes ... portant. Although much ...

particularly useful for the identification of sounds in mixtures. Since we are working with broadly similar problems of classifying unconstrained environmental audio, we attempting to replicate their system as closely as possible to test it on a consumer video soundtrack retrieval task.

The next sections introduce our data/domain, and then describe our results using an available implementation of the auditory model front-end, and our modified, simplified features aiming to capture the same information. Sections 5 and 6 describe other experimentation with the original system, experimenting with replacing the original PAMIR retrieval model and with more common Support Vector Machine (SVM) classifiers, and with reduce the dimensionality of the representation. Section 7 describes the further improvements we obtained by fusing these novel features with the existing baseline MFCCs.

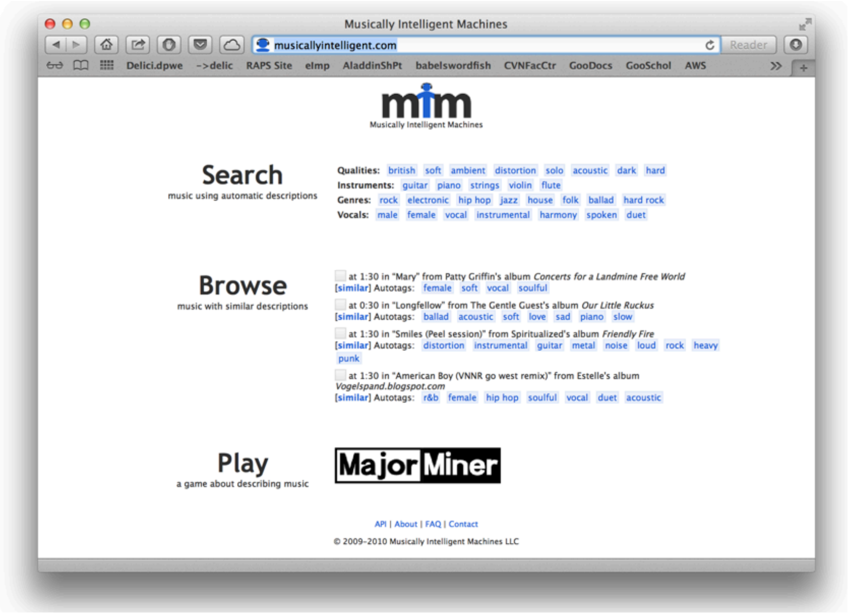
2. DATASET AND TASK

2. **Talks / Demos / videos**
   - pro: Quick hook
   - con: Distorting

Five Ways...

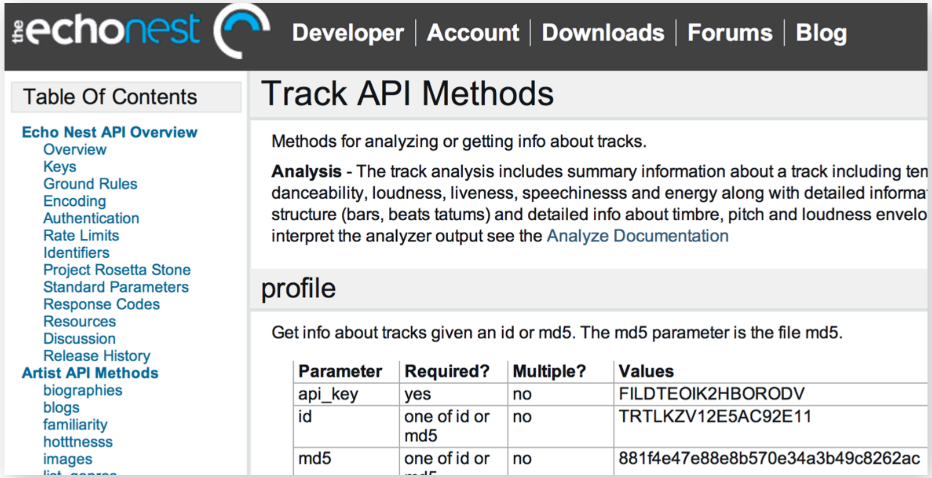
3. Interactive demos

- pro: Let people ask their own questions
- con: Substantial additional development



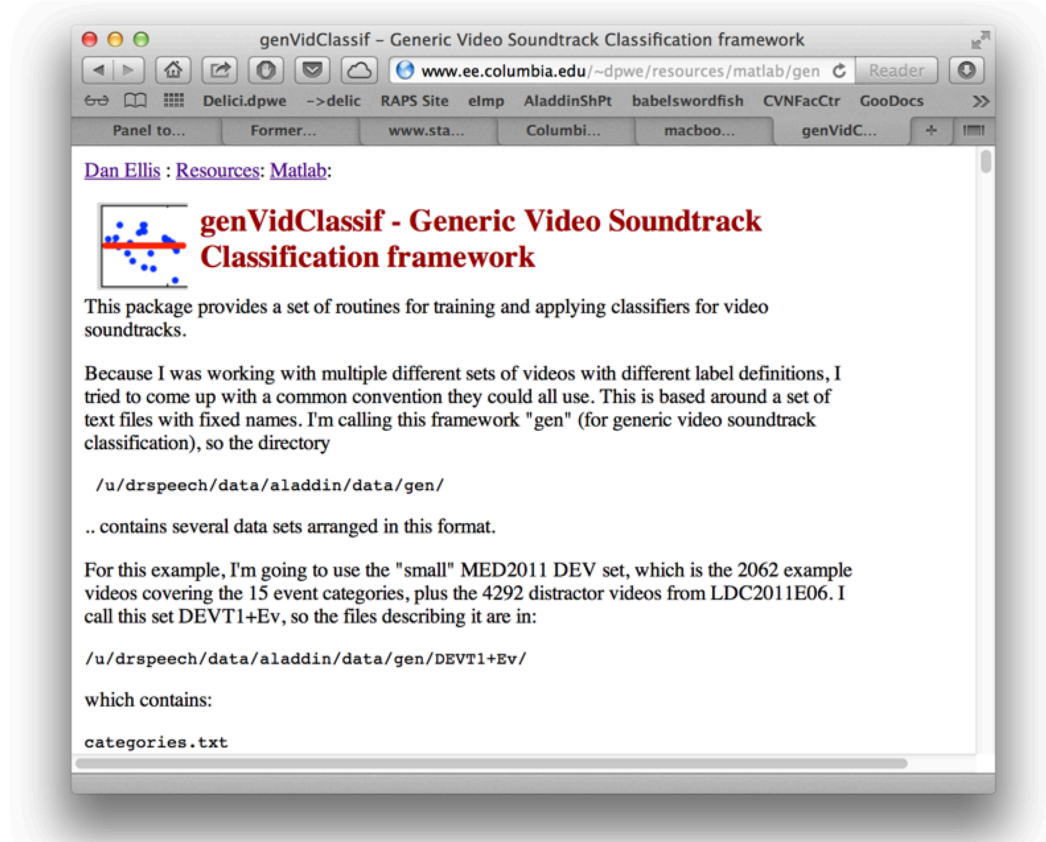
4. Libraries / APIs

- pro: Promotes uptake
- con: Development and support liability



The Fifth Way: Code Sharing

- Complete description of what you did
  - "share the research equipment"
- Pros
  - every detail, regardless of your spin
  - allows replication & reuse
  - the best way to uncover bugs
- Cons
  - time to prepare
  - dirty laundry
  - competitive edge



3. Sharing Code

- **Code Sharing & the Scientific Mission**
  - Scientific fields traditionally struggle to develop reproducible protocols
  - Commodity computers & software support unprecedented reproducibility
- **Barriers** [Stodden 2010]

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	-
34%	Legal Barriers (ie. copyright)	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

The Future of Sharing Code

- **The opportunity to share code is novel**
  - Better, more consistent, high-level platforms
  - Open Source movement
- **There are drawbacks**
  - Time to prepare
  - Fear of humiliation
- **There are huge advantages**
  - Scientific mission: reproduction, verification, debugging
  - Impact

4. Conclusions

- Computational research is qualitatively different
  - and great
  - but hard to comprehend
- Traditional publications describe superficially
  - not a good match
  - editorial choices about "what matters"
- Cheap & powerful computers support code sharing
  - "if I can't fix it, I don't own it"
  - but: airing dirty laundry
- Waiting for a generational change...