Authoring soundscapes with user-generated content and automatic audio classification

Jordi Janer
Senior Researcher at Universitat Pompeu Fabra
Barcelona, Europe

GAME DEVELOPERS CONFERENCE
SAN FRANCISCO, CA
MARCH 5-9, 2012
EXPO DATES: MARCH 7-9
2012

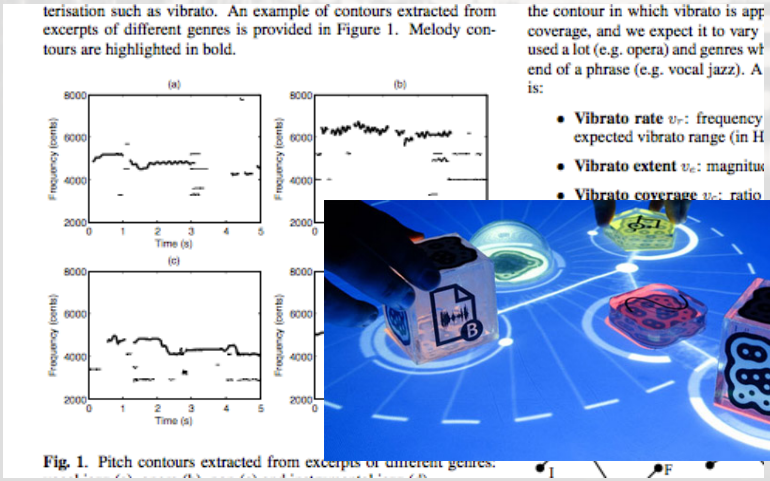
about us

Music Technology Group

Basic and applied research on **sound and music computing**

Key figures:

- 40+ researchers
- 13 patents
- 50+ publications / year
- ~1,4 M€ annual income from projects:
  - Public
  - Industrial


terisation such as vibrato. An example of contours extracted from excerpts of different genres is provided in Figure 1. Melody contours are highlighted in bold.

the contour in which vibrato is app
coverage, and we expect it to vary
used a lot (e.g. opera) and genres wh
end of a phrase (e.g. vocal jazz). A
is:

- **Vibrato rate** v_r: frequency
expected vibrato range (in H
- **Vibrato extent** v_e: magnitu
- **Vibrato coverage** v_c: ratio

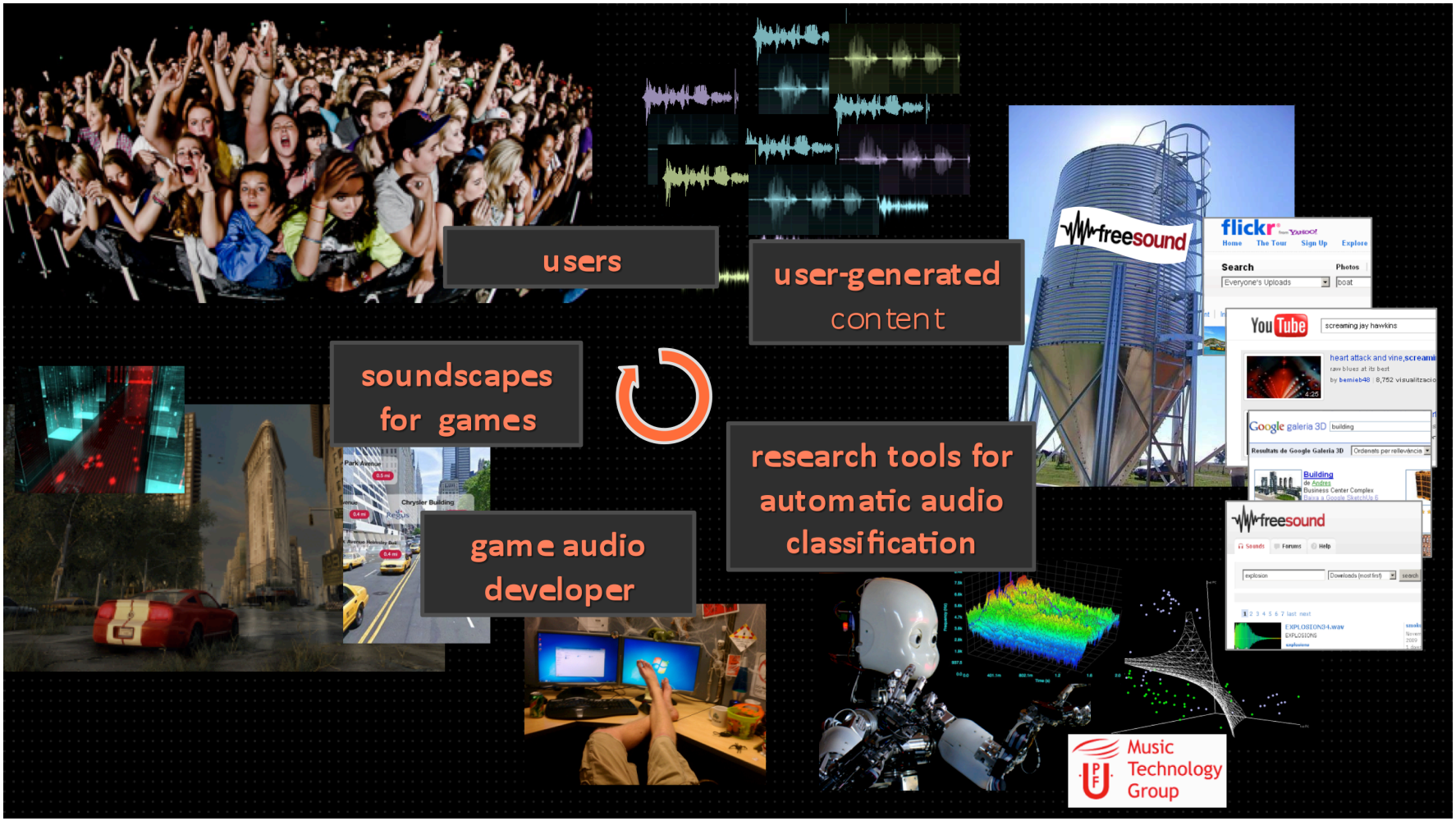
Fig. 1. Pitch contours extracted from excerpts of different genres.

Spin-off company specialized in ***voice processing*** *for music, films and… games!*

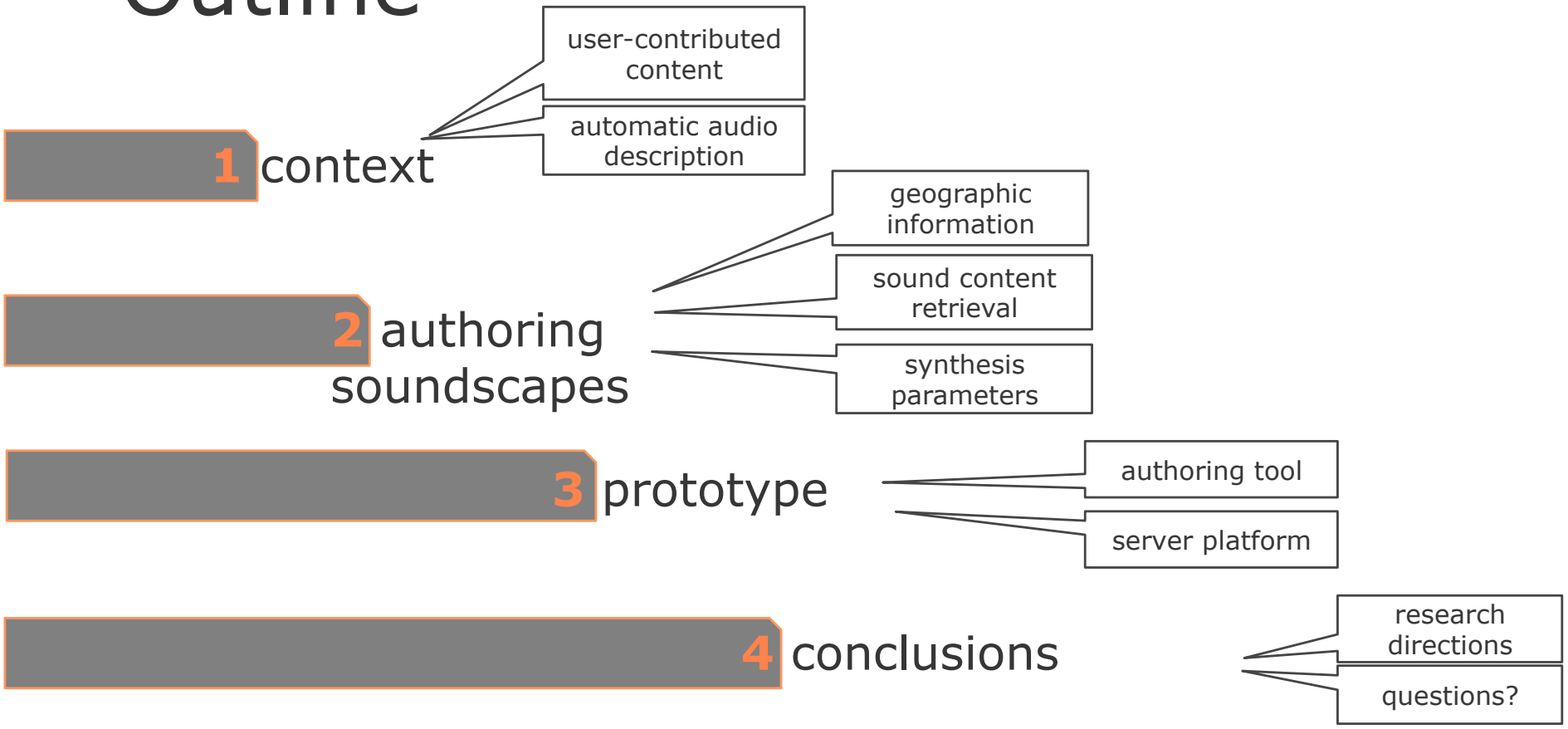
VOCTRO LABS

2005 2009 2011

...but let's go back to our topic.


users
user-generated content
soundscapes for games
research tools for automatic audio classification
game audio developer
freesound
flickr
YouTube
Google galeria 3D
Music Technology Group

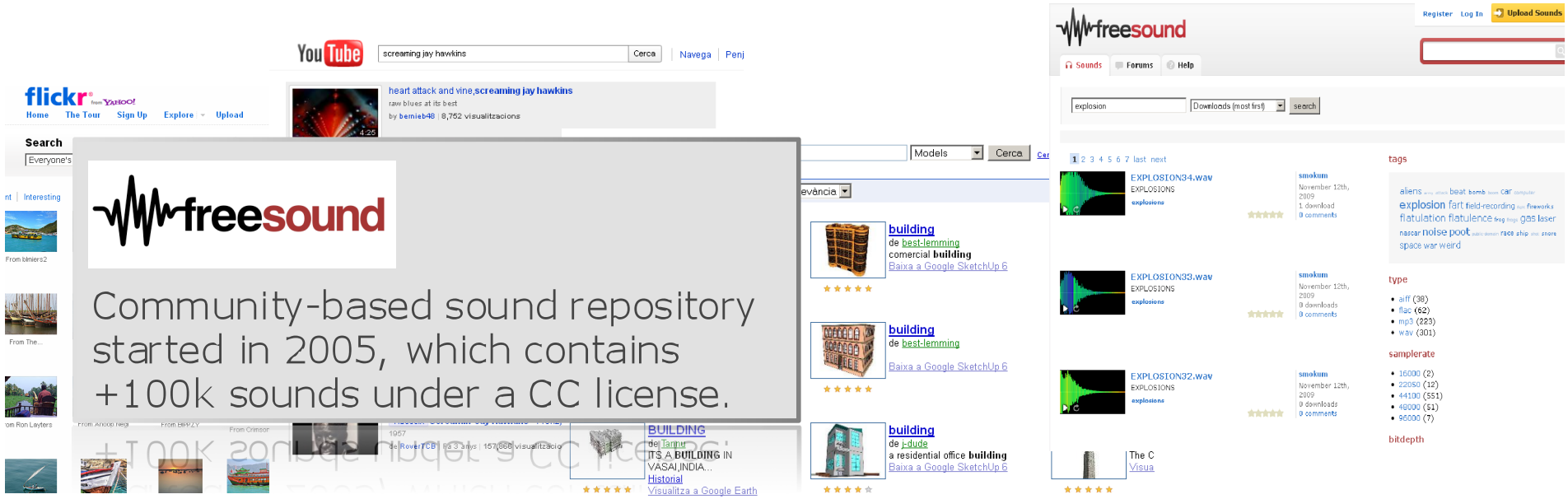
Outline


1 context
user-contributed content
automatic audio description
2 authoring soundscapes
geographic information
sound content retrieval
synthesis parameters
3 prototype
authoring tool
server platform
4 conclusions
research directions
questions?

context

User-contributed media available:

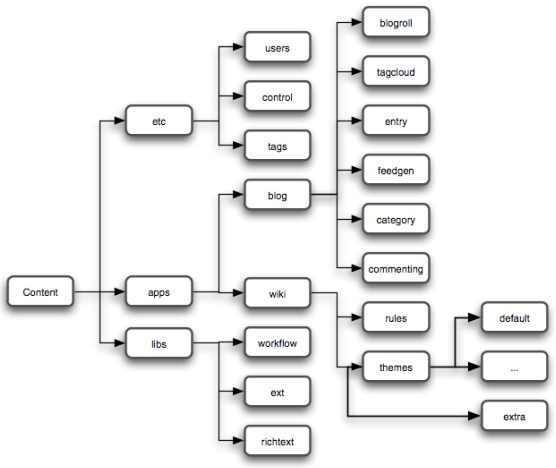
- Photos, videos, 3D models, sounds
- Community-based, different licensing schema

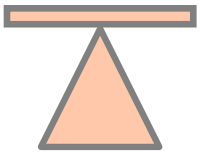


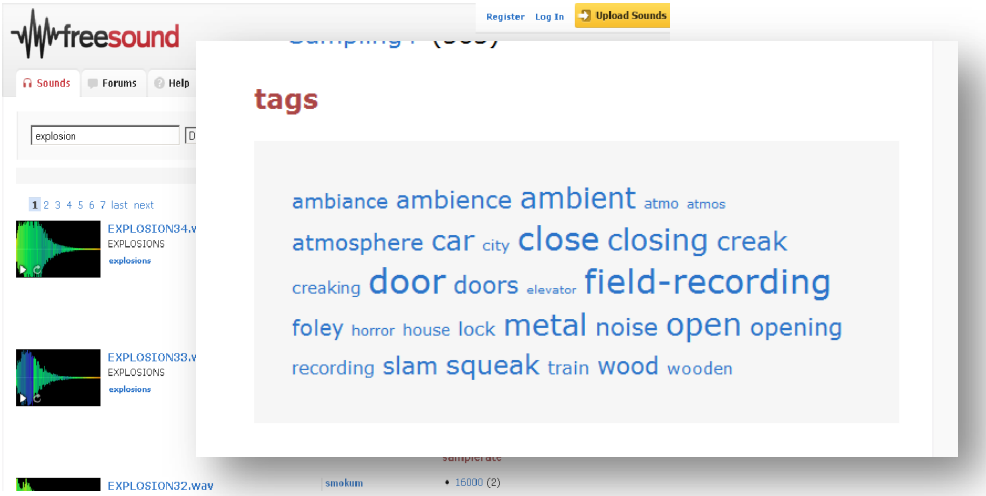
drawbacks of user-contributed media assets

context

1) Inconsistent (audio) quality
2) Unstructured repositories







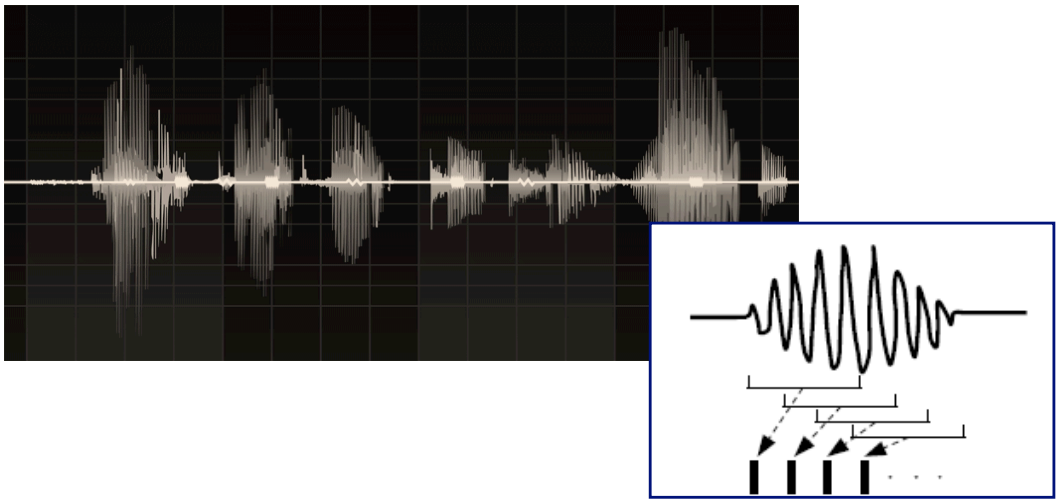
STRUCTURED REPOSITORIES
Publisher libraries
(e.g. soundsnap.com, soundideas.com)

UNSTRUCTURED REPOSITORIES
User-contributed content
(e.g. freesound.org)

automatic audio description

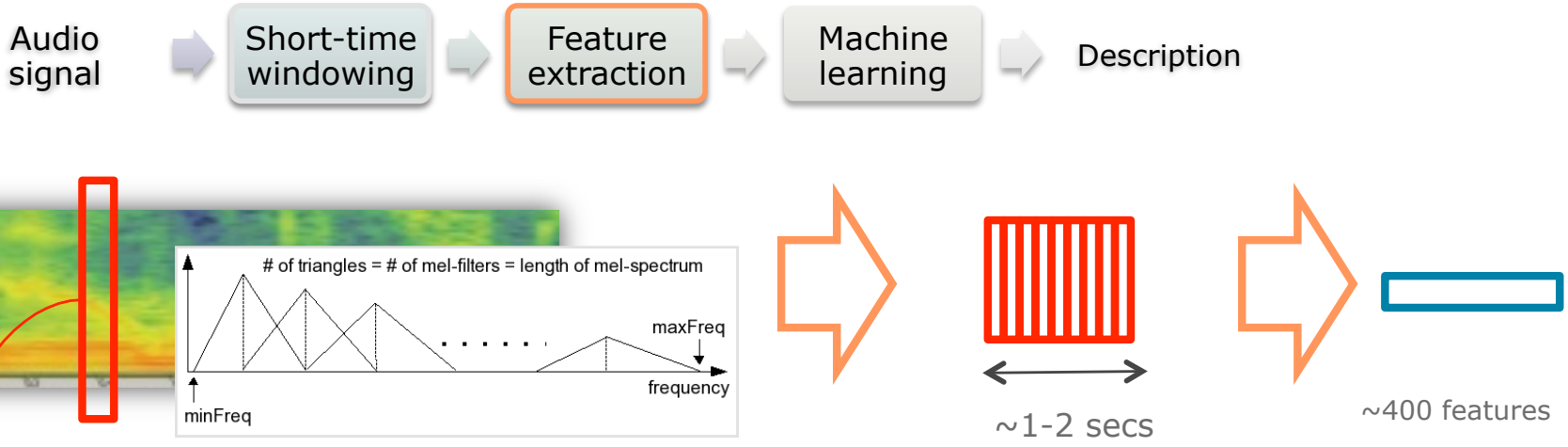
context

Audio signal → **Short-time windowing** → Feature extraction → Machine learning → Description



automatic audio description

context



Frame features (~100) are typically derived from the spectral analysis:

Timbre (e.g. Mel-Frequency Cepsturm Coefficients), Harmonicity, Spectral moments (centroid, kurtosis), other…

To capture time evolution we compute statistics of features over several frames (in red)

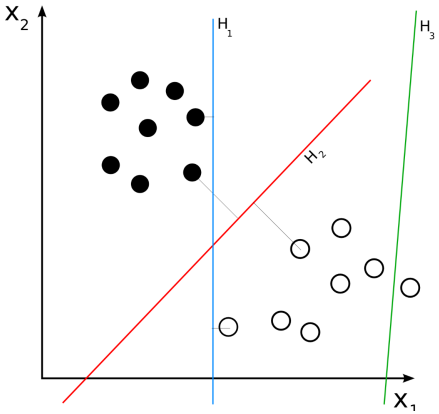
We can consider it as a single features vector (in blue)

automatic audio description

context

Several methods/applications:

- *Pattern recognition* (*item matching as used in audio fingerprinting*)
- *Clustering* (*unsupervised grouping of instances*)
- ***Classification*** (*assign a predetermined label to a new instance*)

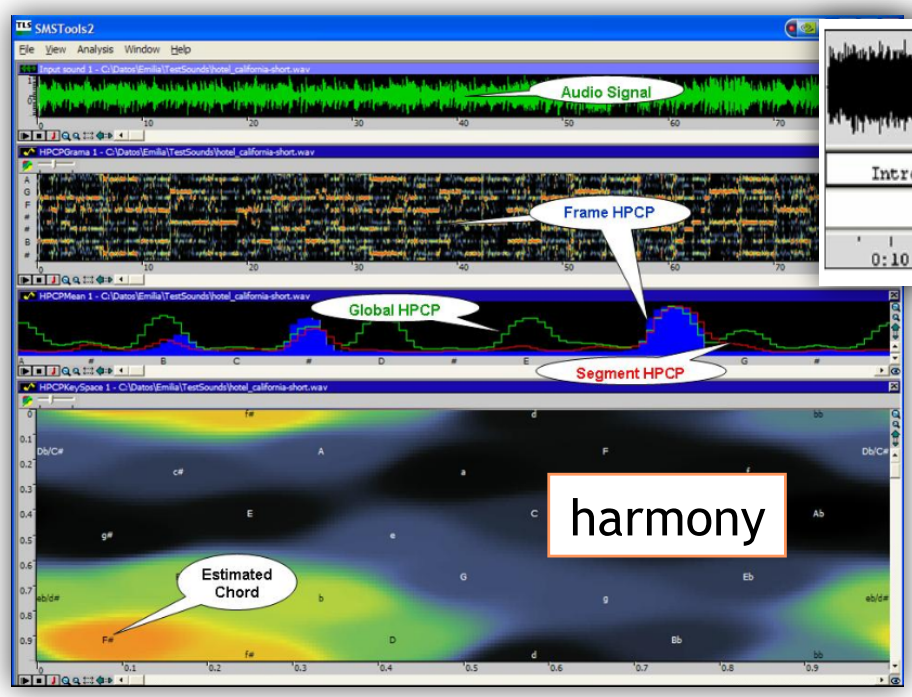


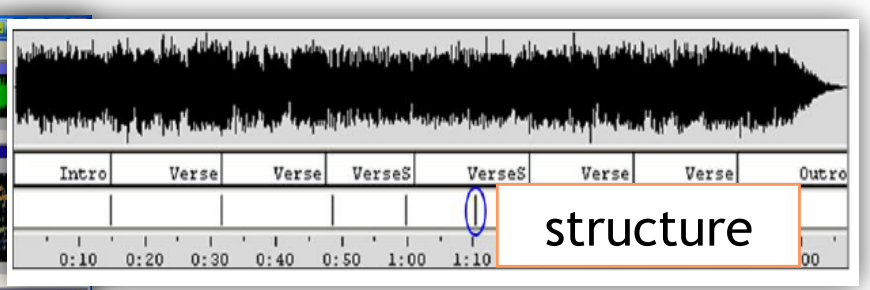
Automatic classification:

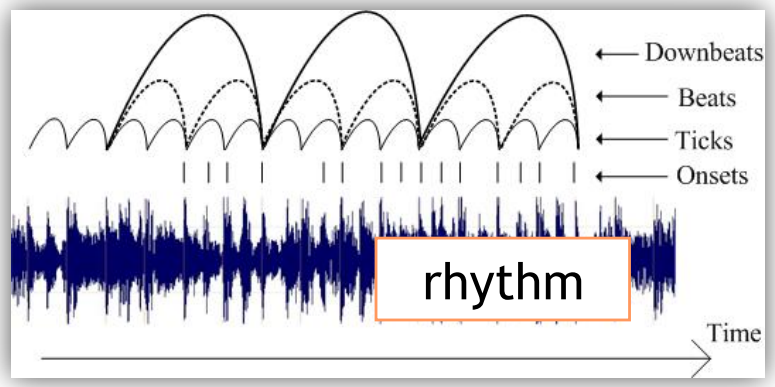
- *Training*: requires annotated datasets to train a model
- *Prediction*: given a model, a new instance is labeled.
- A variety of statistical algorithms are available:
  - e.g. SVM, Decision-trees, Gaussian models.

analysis and description of music

context

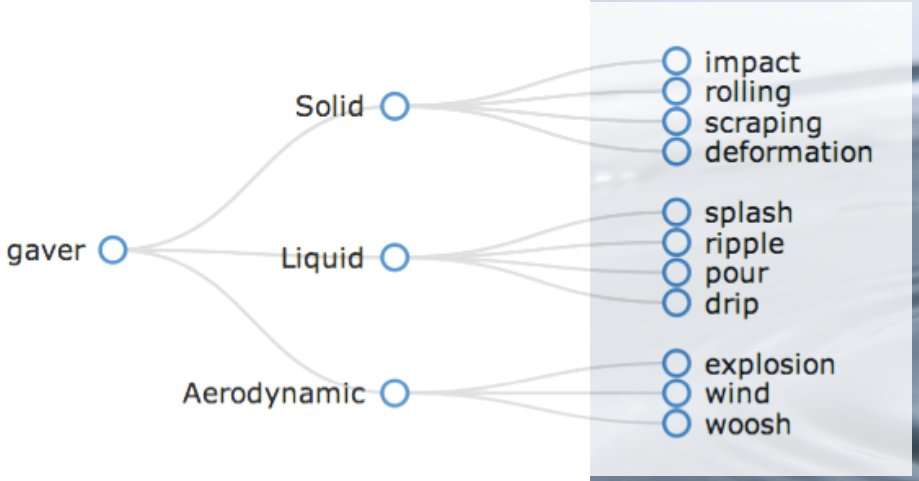






analysis and description of environmental sounds

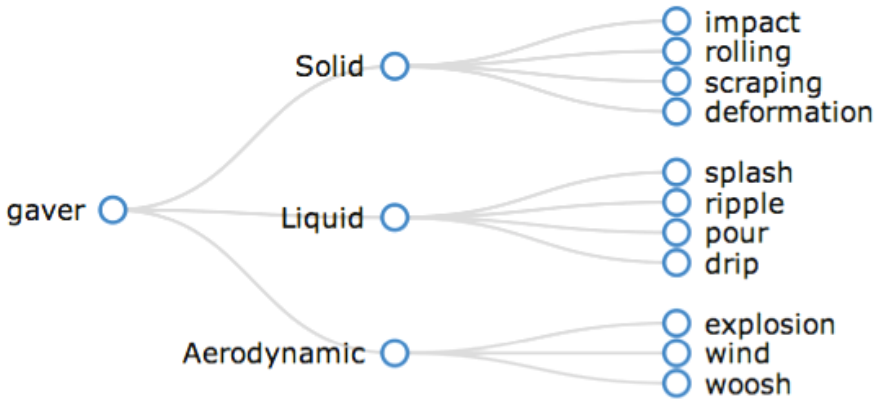
context



Taxonomy based on ***ecological acoustics*** *as proposed by W. Gaver (1994)*

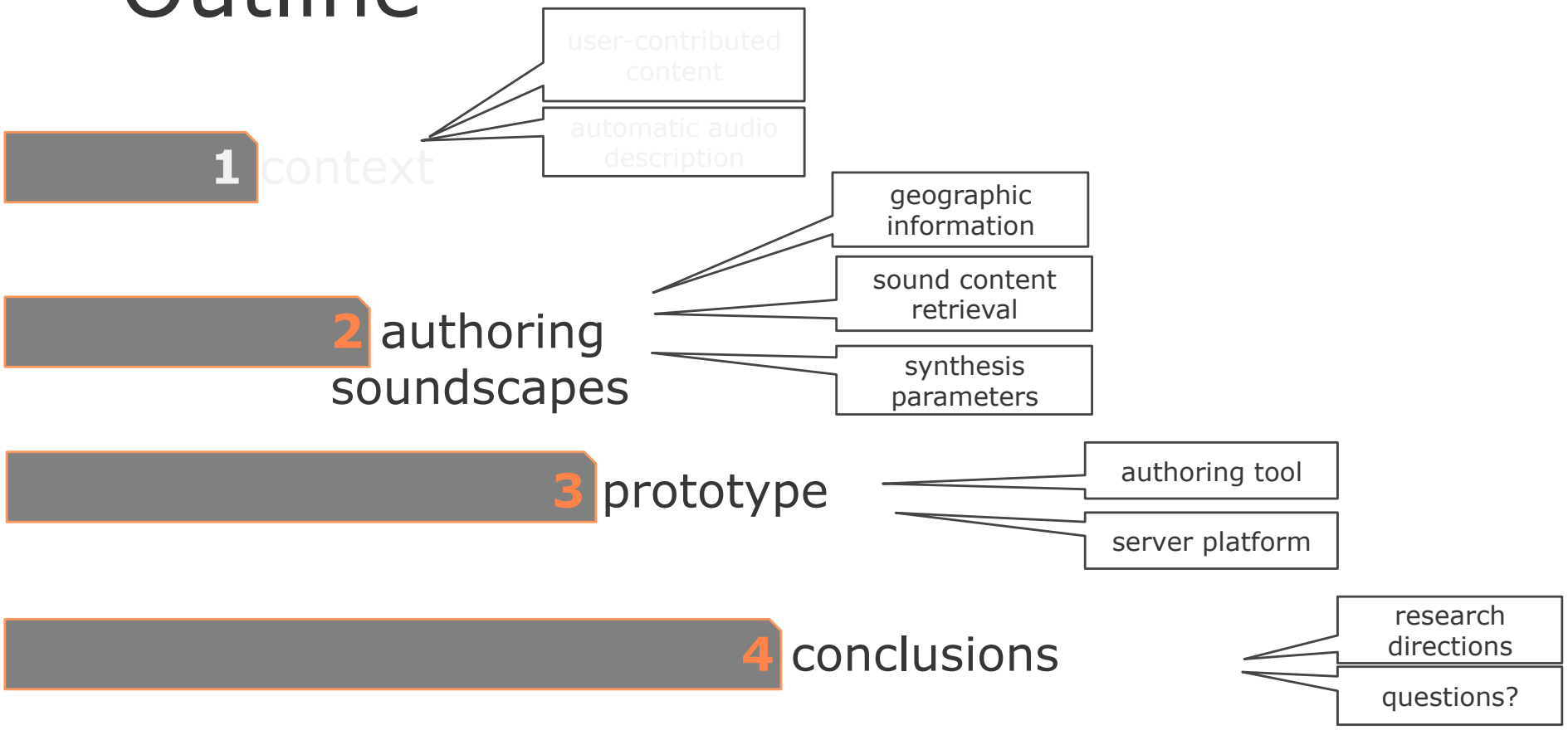
context

analysis and description of environmental sounds



Taxonomy based on
ecological acoustics
as proposed by W. Gaver (1994)

Outline


user-contributed content
automatic audio description
1 context
geographic information
sound content retrieval
synthesis parameters
2 authoring soundscapes
authoring tool
server platform
3 prototype
research directions
questions?
4 conclusions

authoring soundscapes

But what's a soundscape?

an acoustic environment or an environment created by sound

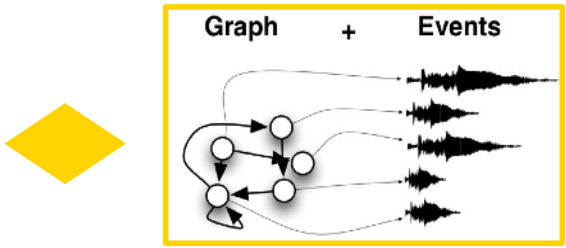
- Background sonic ambiance that reconstructs the sound of a given real or virtual space.
- Only a part of all game audio content
  - e.g. not dialogs, no synched events,...
- Limited spatialization
  - e.g. 2D, no room acoustics simulation

other definitions

"The sonic environment. Technically, any portion of the sonic environment regarded as a field for study. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an environment." *(R.M. Schafer, 1977: 275)*

authoring soundscapes

CONCEPT: a graph model sequencer and a set of sound events (samples) perceived as a single semantic unit.



ZONE: part of the soundscape that presents a specific characteristic. Composed by a set of *concepts*.

SOUNDSCAPE: complex temporal-spatial structure of sound objects, organized as a set of layers or *zones*.

authoring soundscapes

geographic information

Examples of a soundscape of a real location.

```
- <Placemark>
    <name>children</name>
  - <LookAt>
      <longitude>2.173519684377638</longitude>
      <latitude>41.40316147827902</latitude>
      <altitude>0</altitude>
      <heading>1.447342568145065</heading>
      <tilt>3.158027485318621</tilt>
      <range>448.8122098079919</range>
      <altitudeMode>relativeToGround</altitudeMode>
      <gx:altitudeMode>relativeToSeaFloor</gx:altitudeMode>
    </LookAt>
    <styleUrl>#msn_ylw-pushpin6</styleUrl>
  - <Point>
      <altitudeMode>clampToGround</altitudeMode>
      <gx:altitudeMode>clampToSeaFloor</gx:altitudeMode>

        <coordinates>2.173021098769079,41.4030310155222
    </Point>
```

Exported as a standard KML file

Authoring applications

authoring soundscapes

sound content retrieval

Next videos compare the results obtained by querying:

textual search
results ranked by popularity (downloads)

"CONCEPT CATEGORY"

***faceted* search**
results ranked by automatic classification

"CONCEPT" CATEGORY

* Results longer than 20 secs were discarded

sound content retrieval

#1 water pour

sound content retrieval

#2 metal impact

sound content retrieval

#3 metal scraping

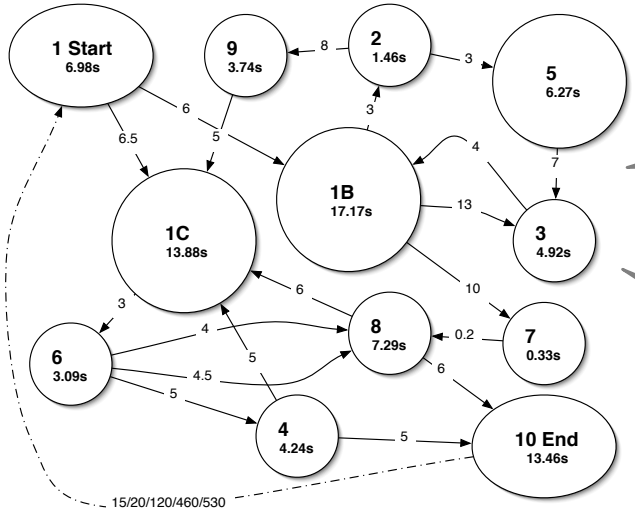
sound content retrieval

#4 gun explosion

authoring soundscapes

real-time synthesis engine

- Based on Concatenative Sound Synthesis (CSS):
- Real-time autonomous generation
- A *sound concept* is a graph model with multiple samples



Graph model:
each node is a sample and edges contain transition probabilities that control the sequencing behaviour

Multiple agents can navigate the graph simultaneously

prototype

Authoring tool

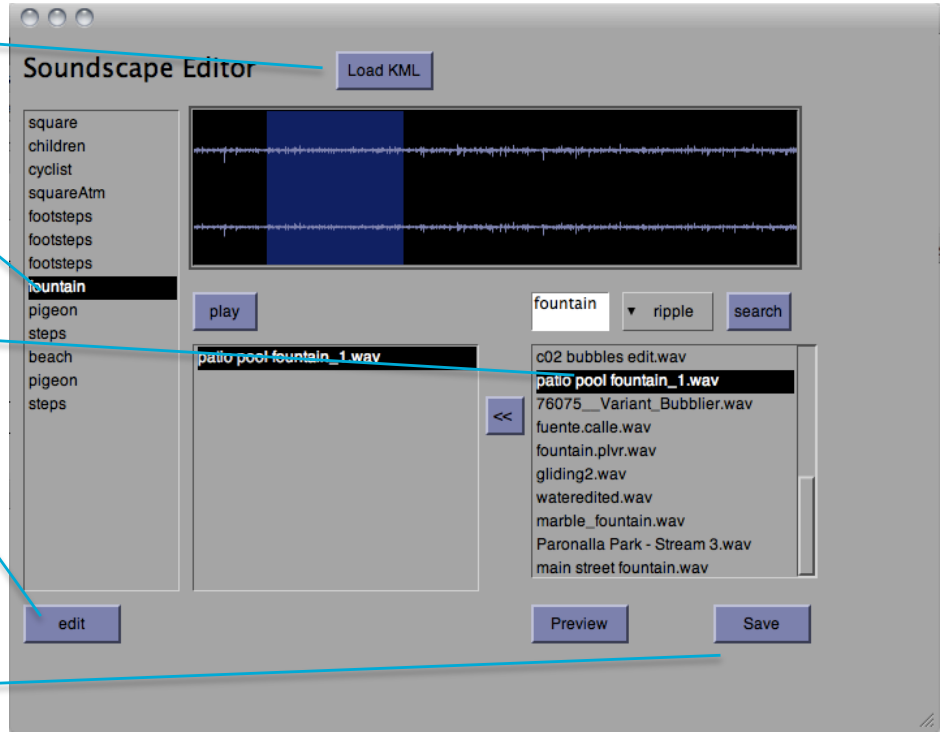
1- Import KML file

2- select a sound concept

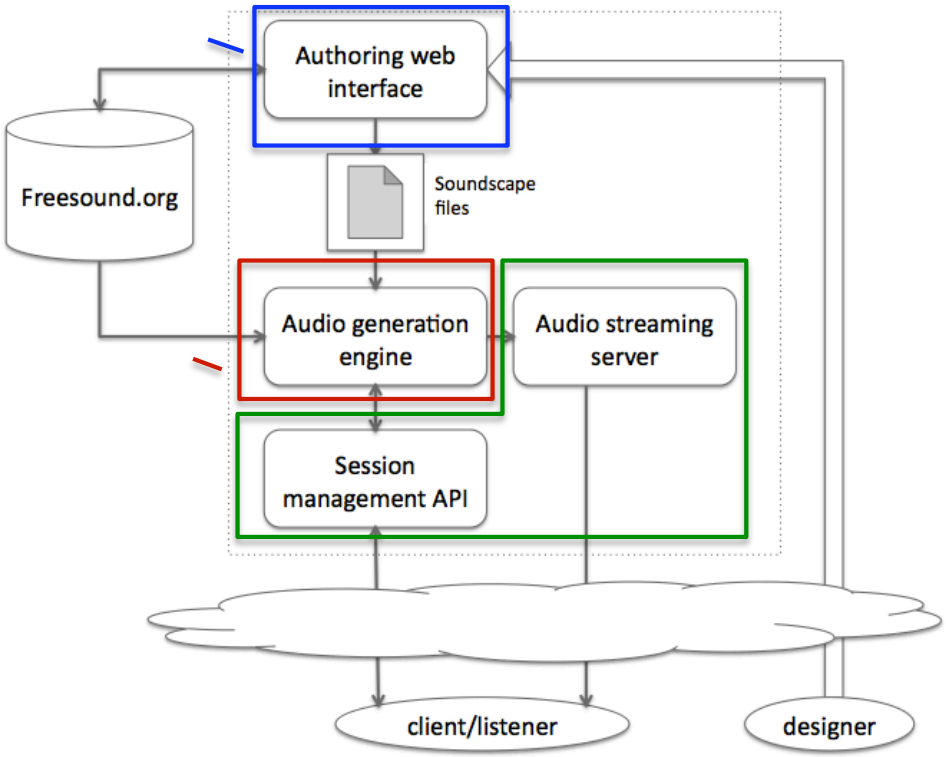
3- search and assign samples to a concept

4- edit segmentation and synthesis parameters

5- export extended KML and dataset XML files



prototype



Online platform

- HTTP API
  - Session management (add/remove listeners)
  - Client (listener) sends position and orientation update messages to the server
- Streaming server
  - Each client receives a personalized MP3 stream
  - Latency < 1-2 sec
- Client
  - Applications supporting MP3 streams
  - Virtual worlds (SL), Games (Unity 3D) or Mobile web browsers (HTML5)

conclusions

conclusions

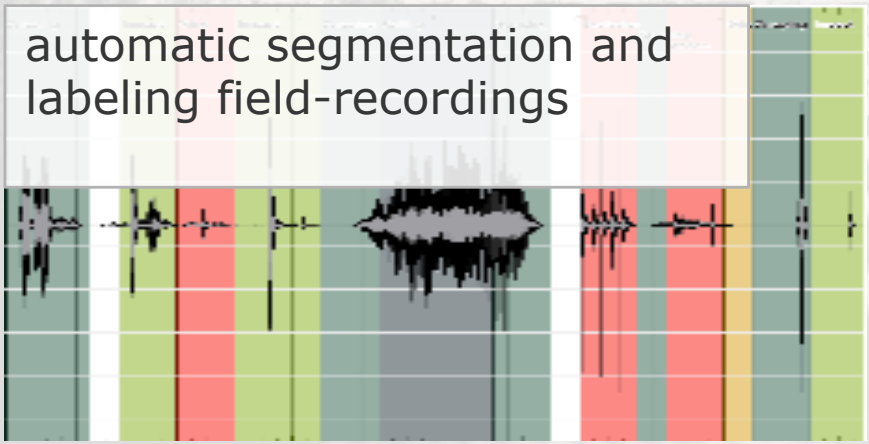
demo

Beach ambiance http://goo.gl/B92At

conclusions

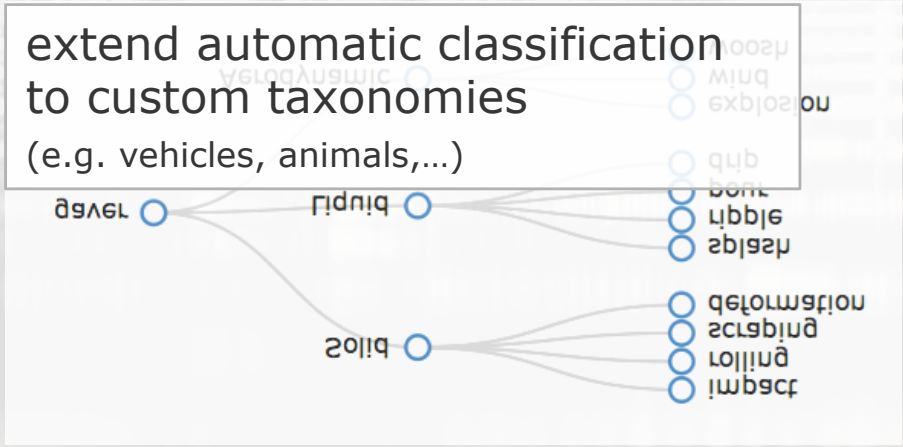
Current limitations
and future research directions

automatic segmentation and
labeling field-recordings



extend automatic classification
to custom taxonomies

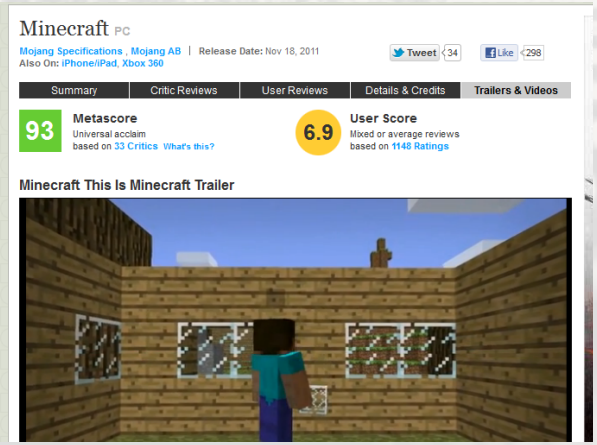
(e.g. vehicles, animals,…)



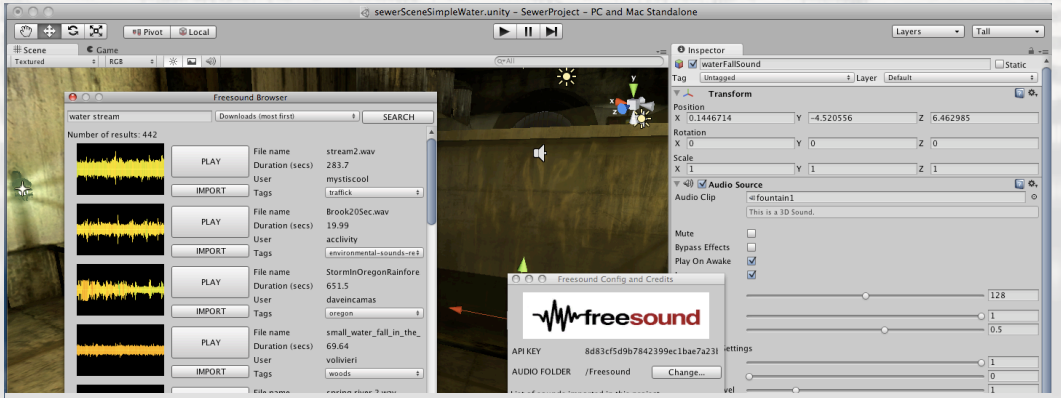
conclusions

We encourage you to use Freesound.org…

by using sound content in your games: ex. *Minecraft*



or by integrating Freesound API in your development tools: e.g. *Unity 3D package*



https://github.com/jorgegarcia/UnityFreesound

thanks!

Jordi Janer
jordi.janer@upf.edu

More information and additional video demos:
http://mtg.upf.edu/technologies/soundscapes

Partially funded by
Generalitat de Catalunya

TECNIO
Be tech. Be competitive

GDC 12

GAME DEVELOPERS CONFERENCE
SAN FRANCISCO, CA
MARCH 5-9, 2012
EXPO DATES: MARCH 7-9
2012