The VU sound corpus
Adding More Fine-grained Annotations to the Freesound Database

Emiel van Miltenburg   Benjamin Timmermans   Lora Aroyo

`{emiel.van.miltenburg, b.timmermans, lora.aroyo}@vu.nl`

Vrije Universiteit Amsterdam

What are sound events?

Sound events: everyday sounds that we hear around us

- Dogs barking, cats purring
- Doors opening and closing
- Toilet flushing
- Cars in a traffic jam
- Water dripping from the tap
- Footsteps

Questions

- What kind of language do we use to describe sound events?
- Can we reliably crowdsource keywords describing those events?
- What is the variation and range of those keywords?
- Is there a difference between 'expert' authors and the crowd?
- Could the keywords be useful for sound retrieval?

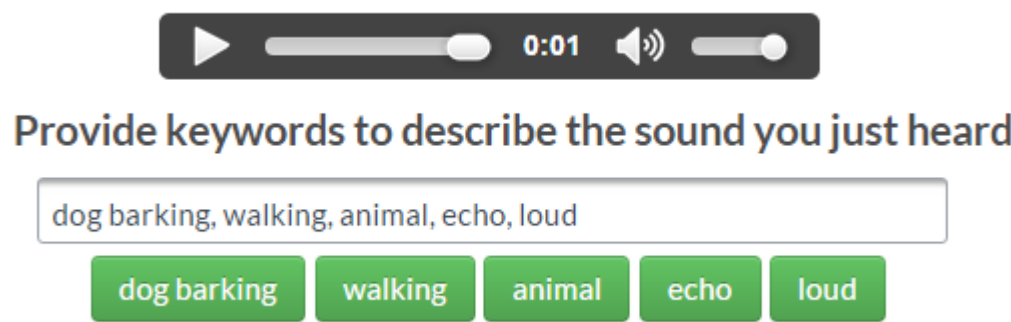## Data

Freesound.org (Font et al. 2013):
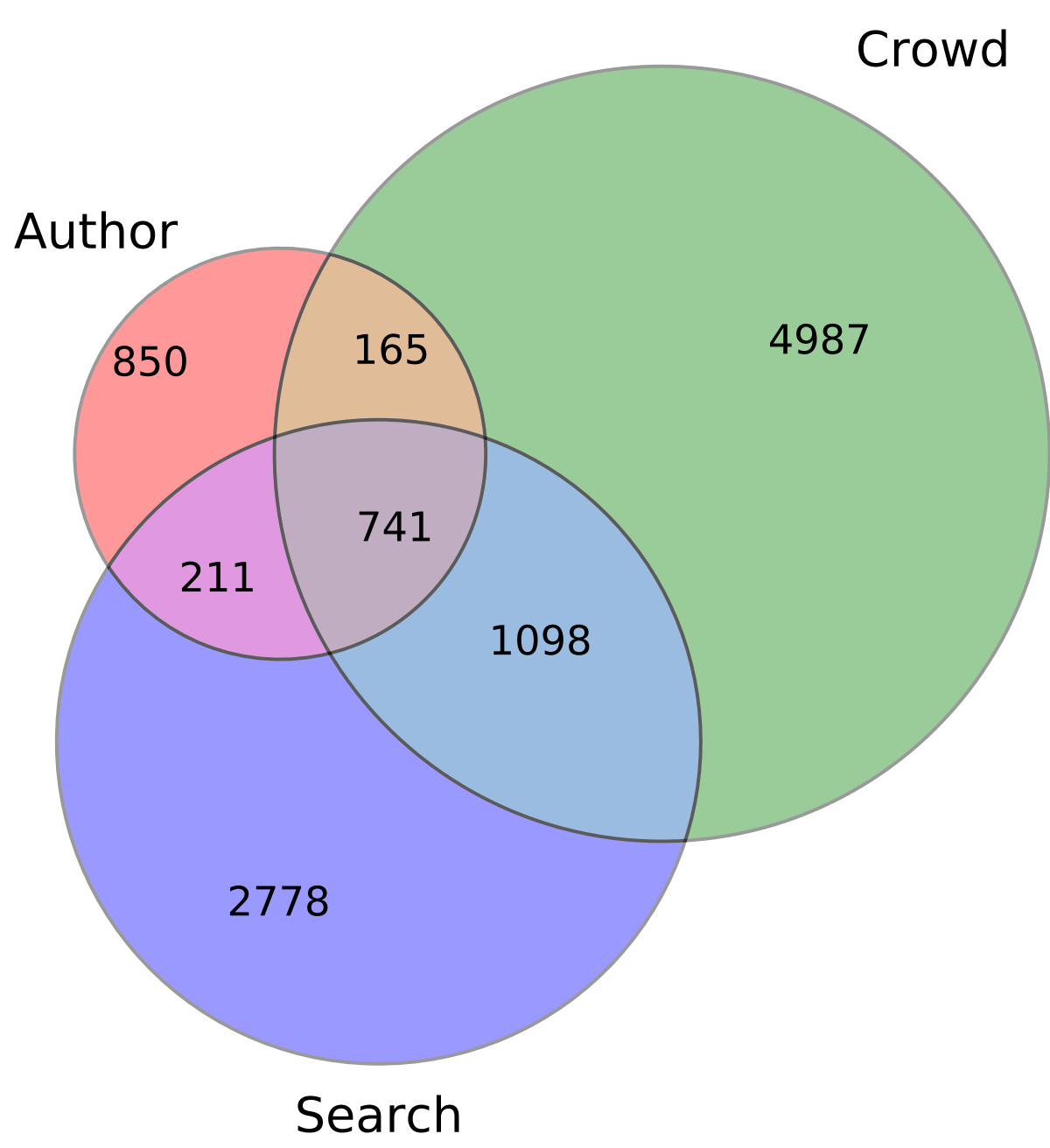
- Collaborative database of Creative Commons-licensed sounds
- Almost 300 000 sounds, with tags and descriptions
- 4 000 manually categorized as sound-effects (Font et al. 2014)
- Our final dataset: 2 133 sounds

Task

- Using the Crowdtruth framework (Dumitrache et al. 2015)
- Through www.crowdflower.com
- Free labeling task
- Three sounds per task (short, med, long)
- Live segmentation of input into labels

Provide keywords to describe the sound you just heard
dog barking, walking, animal, echo, loud
[dog barking] [walking] [animal] [echo] [loud]

Results

Set	% Overlap
Author tags	57
Search terms	72
Both	55
Non-overlapping	26

Type overlap   Token overlap

Main result: Authors differ from the crowd in the labels they use.

*Sound #158802.*

Author (a commercial party)
Film; Radio; Future; Alien; Futuristic; effects; Broadcasting; Recording; fx; Music-Production; Video; Screen; media; TV; space; Remixing; alien-sound-effects; Sound-Effects; pod-Cast; DVD; Home-Videos

Crowd-tags
ultro sound; video pings; mysterious; computerized; chaotic; robotic; robot; scales; buttons; computer; bleeping; synthetic; chimes; random; descending; tones; discordant; electronic; chaos; science; beeps; technology; playing music backward; high; pitchy; mix

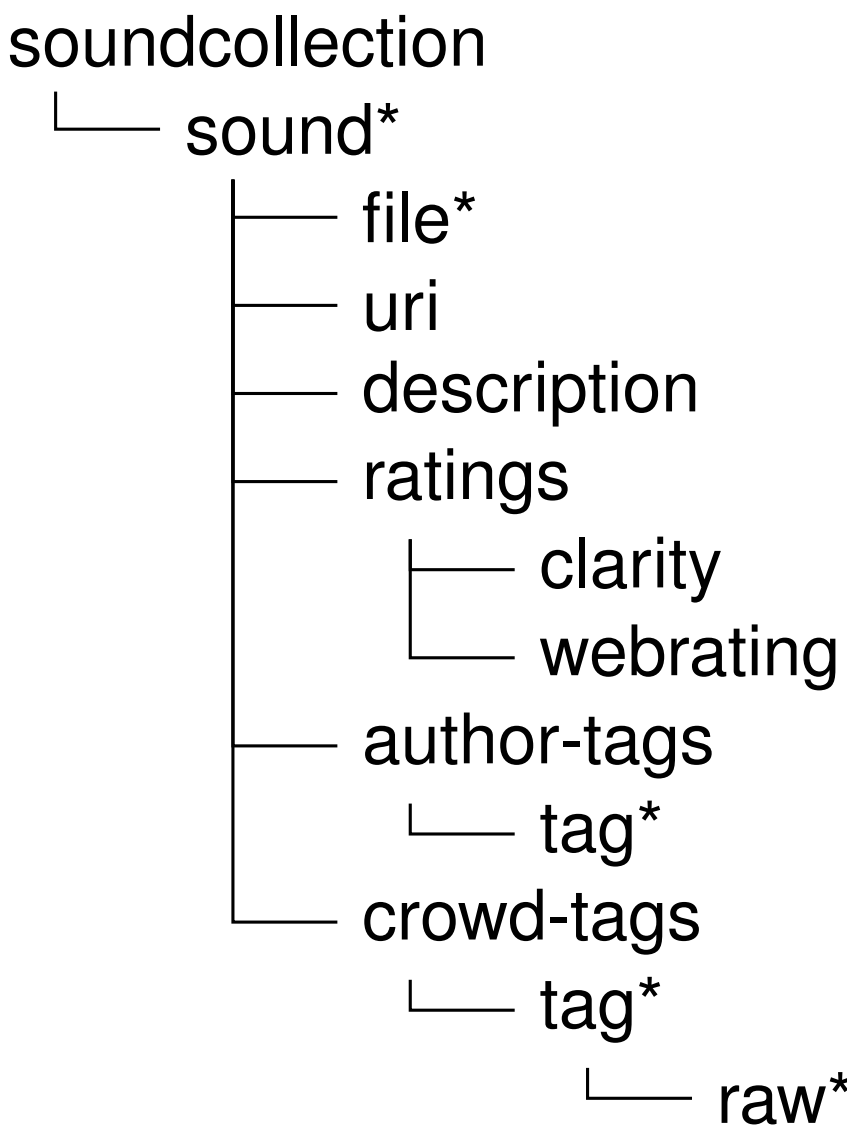## Results (continued)
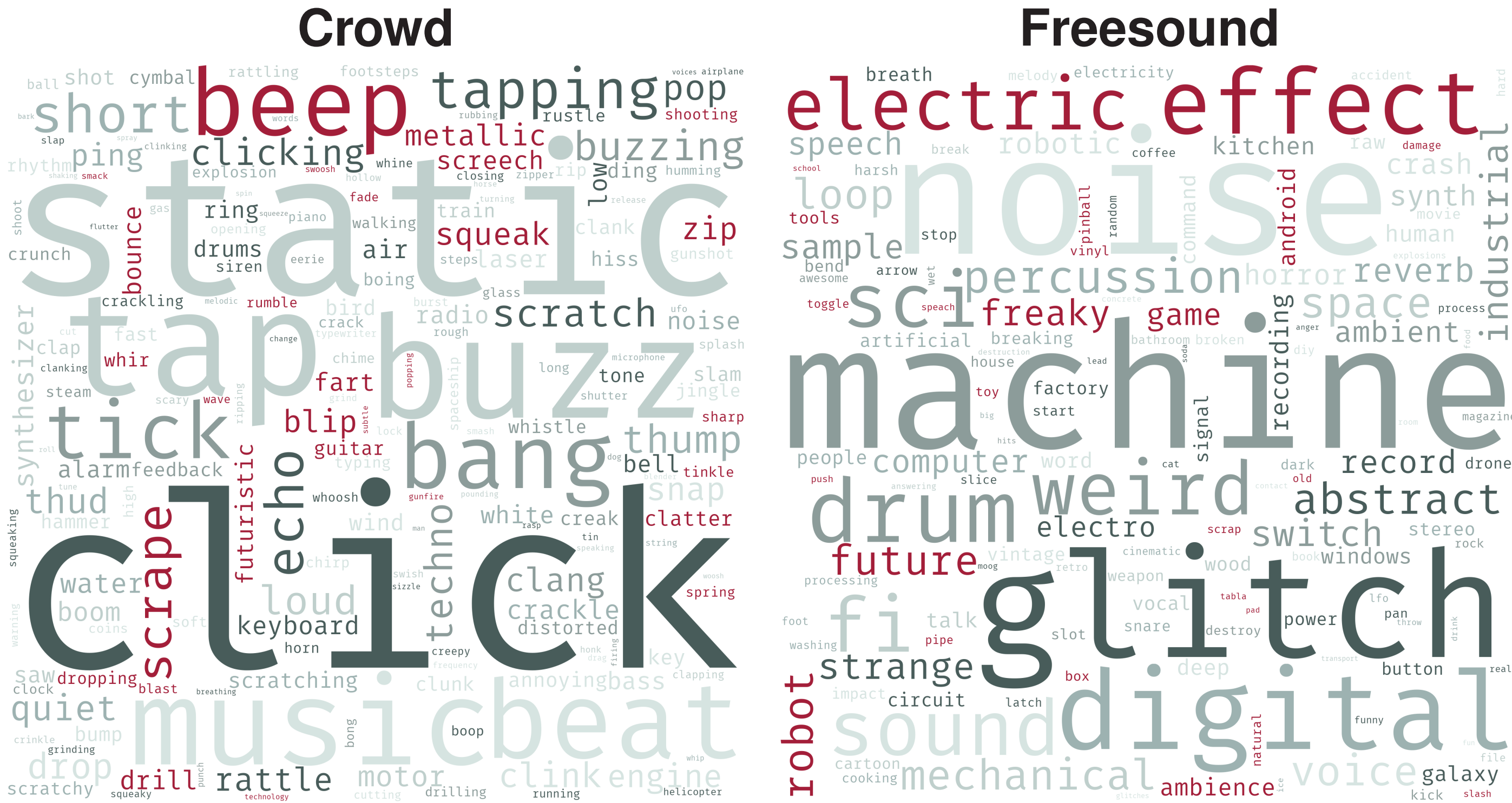
Crowd

Freesound

Typical items for the crowd-annotations versus the author-tags

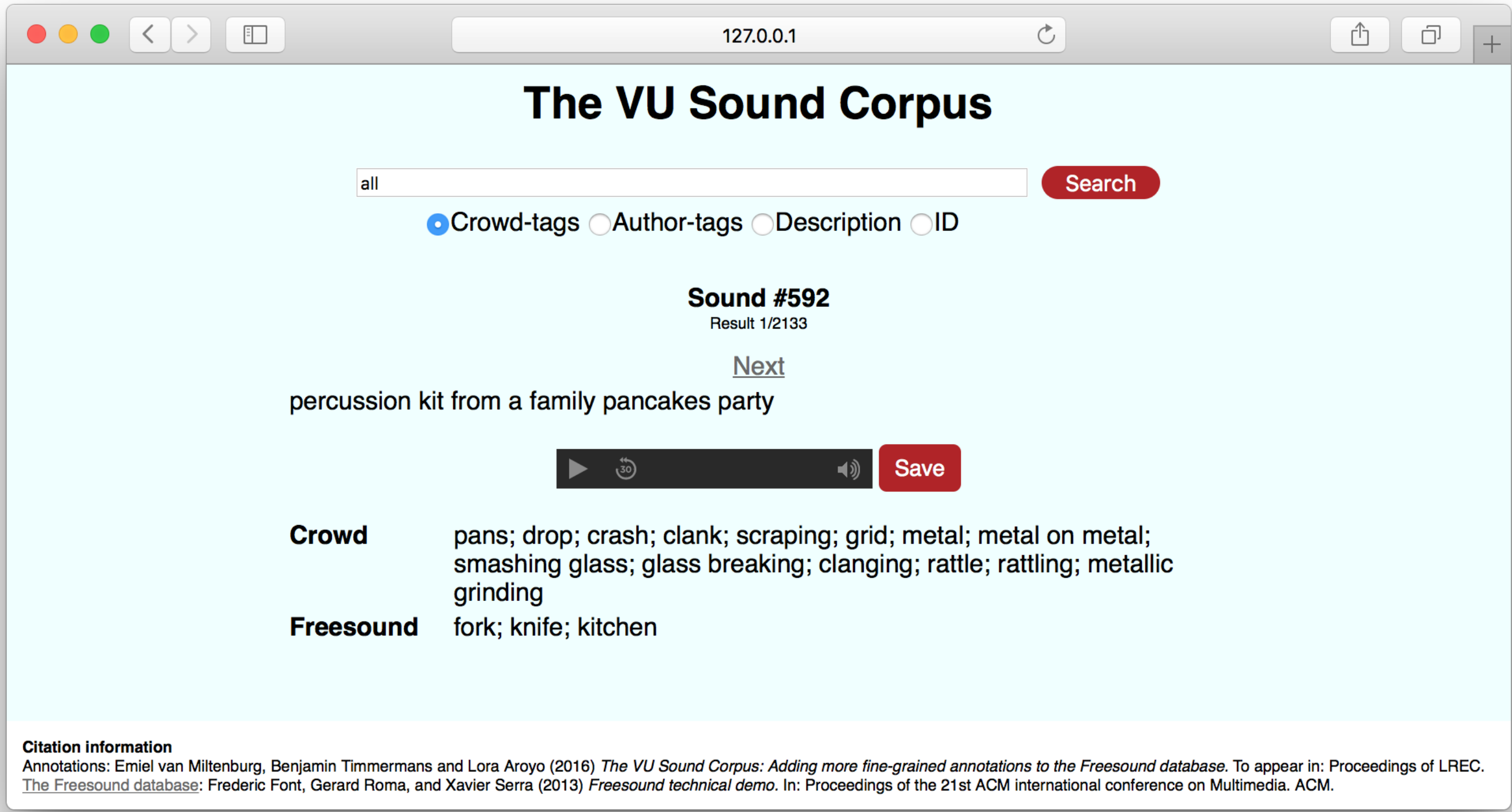(Generated using `https://github.com/amueller/word_cloud`, with `relative_scaling` = 0.5)

XML format

- Sounds part of sound collection
- Linked to `Freesound.org` URIs
- Provided with *id, batch, name, type, samplerate, duration, channels, bitrate* and *bitdepth*
- Batch number provides transparency about data collection


```
soundcollection
└── sound*
        ├── file*
        ├── uri
        ├── description
        ├── ratings
        │     ├── clarity
        │     └── webrating
        ├── author-tags
        │     └── tag*
        └── crowd-tags
              └── tag*
                    └── raw*
```

`https://github.com/CrowdTruth/vu-sound-corpus`

Sound browser interface

`https://github.com/evanmiltenburg/SoundBrowser`

References

Dumitrache, Anca, Oana Inel, Benjamin Timmermans, Lora Aroyo & Robert-Jan Sips. 2015. Crowdtruth: Machine-human computation framework for harnessing disagreement in gathering annotated data. In Grotov, Van Gysel, Kanoulas, Azarbonyad, Voskarides, Best & Li (eds.), *Proceedings of the 14th dutch-belgian information retrieval workshop (DIR)*, 15.

Font, Frederic, Gerard Roma & Xavier Serra. 2013. Freesound technical demo. In *Proceedings of the 21st acm international conference on multimedia*, 411–412. ACM.

Font, Frederic, Joan Serrà & Xavier Serra. 2014. Audio clip classification using social tags and the effect of tag expansion. In *Audio engineering society conference: 53rd international conference: Semantic audio*, Audio Engineering Society.