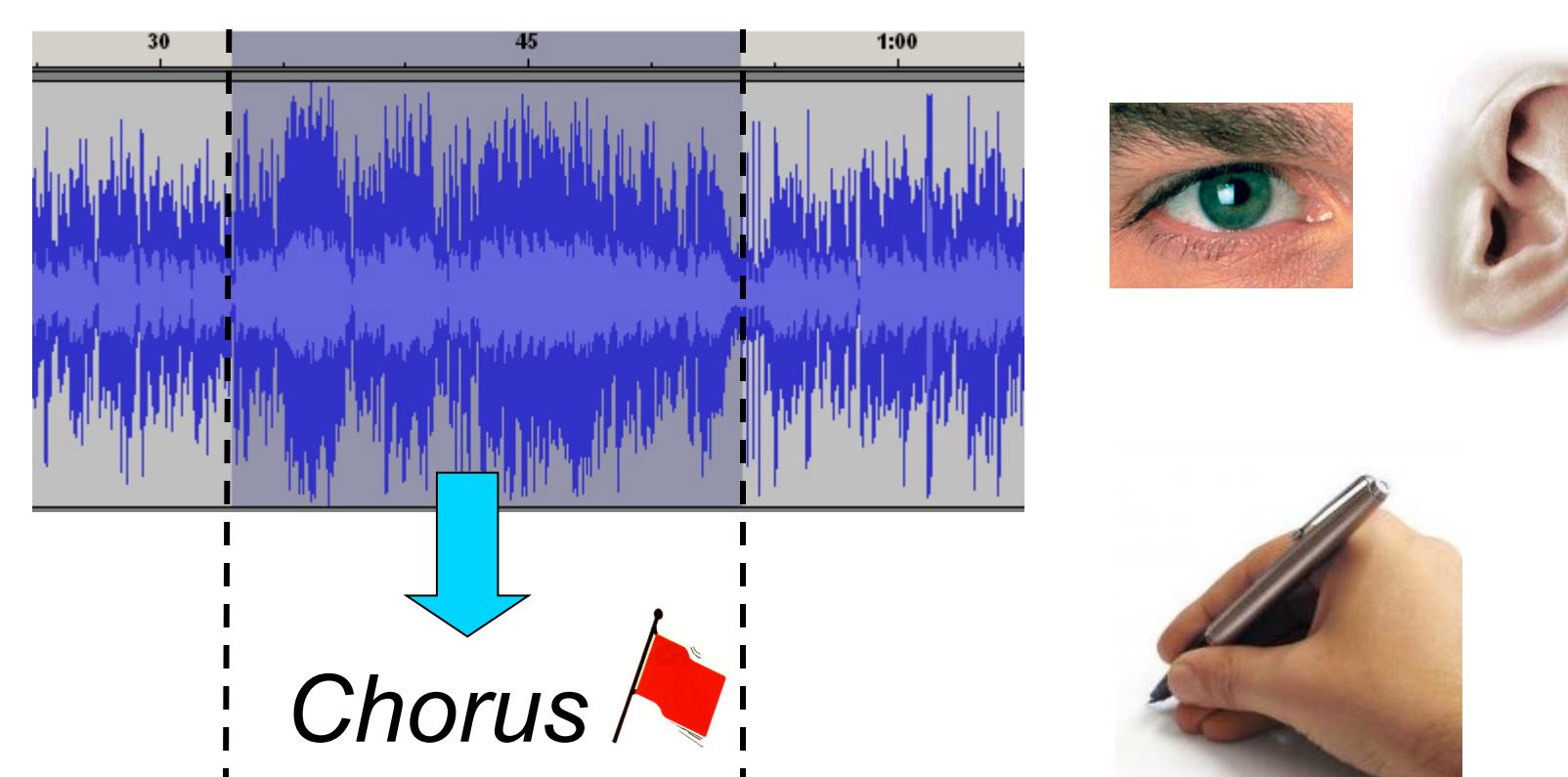


Extending Audacity for Audio Annotation

Beinan Li, John Ashley Burgoyne, Ichiro Fujinaga

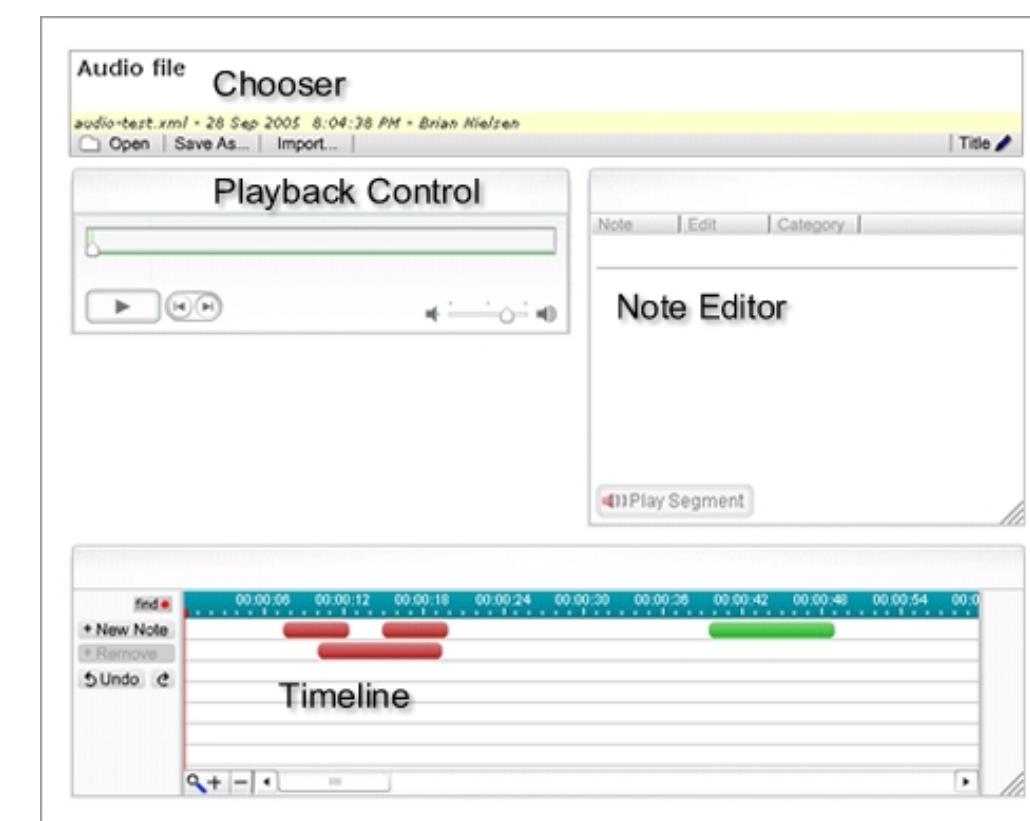
Music Technology Area, Schulich School of Music, McGill University, and CIRMMT, Montreal, Canada

Manual Audio Annotation



Audio classification systems calls for custom manual annotation software.

Existing tools lack features for audio classification purposes and are not customizable, e.g. :



Project Pad:
No waveform viewer

Choice of Software Framework

MIR needs open-source software with:

- Full audio playback control
- Support for creating text notes
- Audio signal visualization
- Audio format compatibility
- Cross-platform compatibility

Pratt
(version 4.4.31):



- Written in C
- Mainly for speech analysis
- No support for MP3, and many other popular compressed audio formats
- Self-implemented GUI, hard to extend

Audacity
(version 1.3beta):



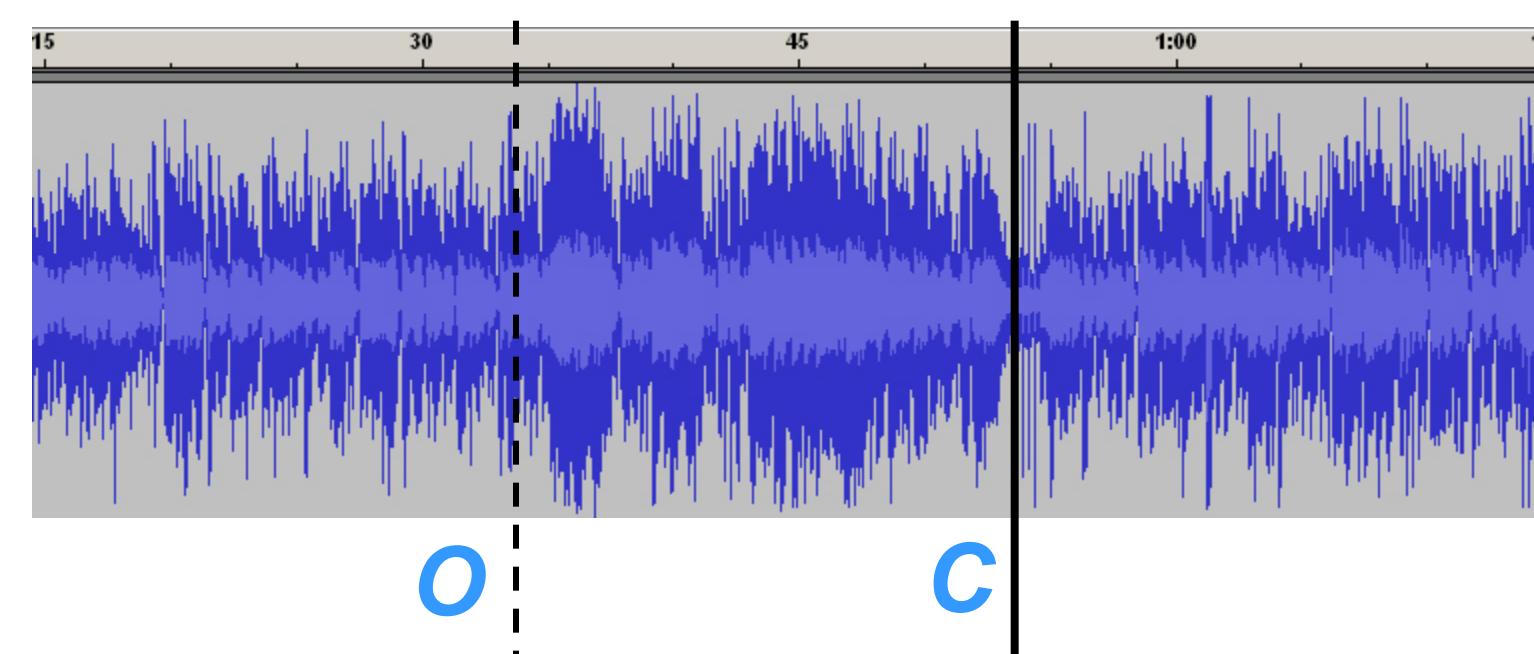
- Written in C++
- General audio editing
- Support for labeling tracks
- Support for popular uncompressed and compressed audio formats
- GUI based on the open-source framework *wxWidgets*, easy to extend

Limitations of Audacity

- **Region selection:** Cannot store temporary boundaries.
- **Labeling tracks:** No automatic label creation or naming.

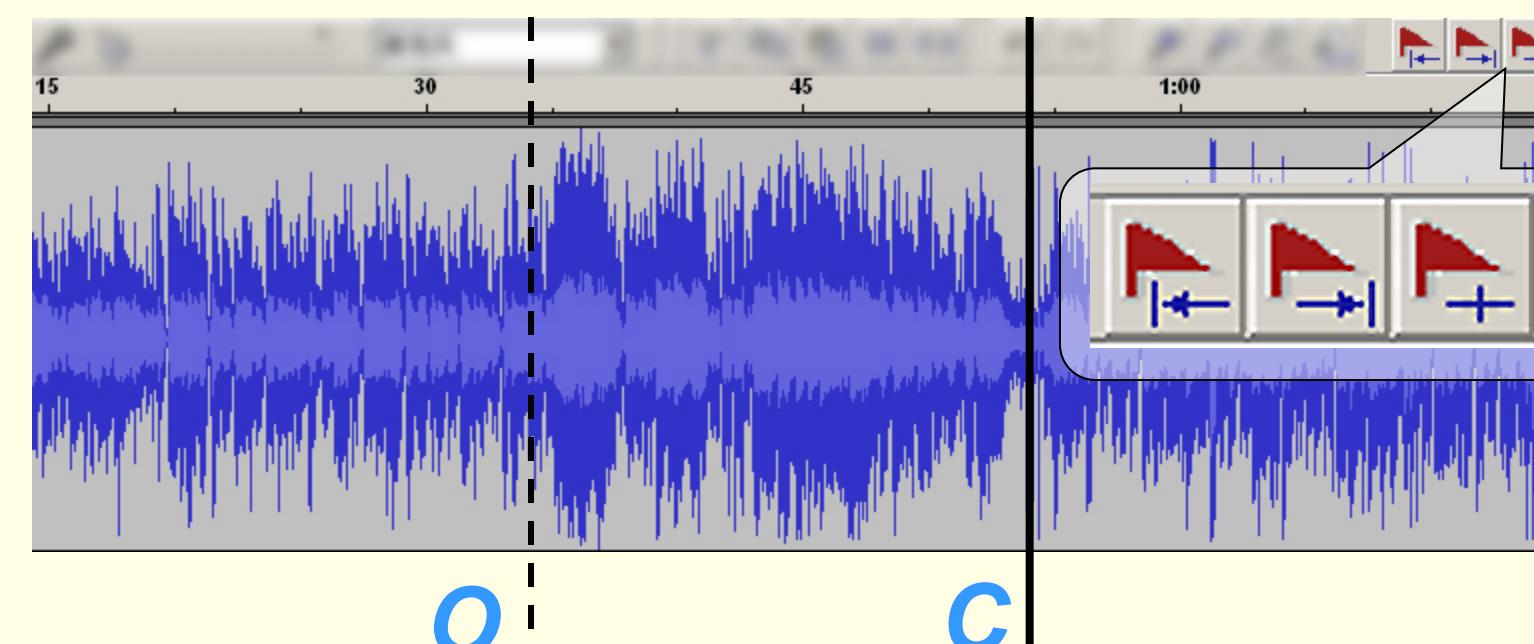
Region Selection

In classic Audacity:



The opening boundary **O** of an audio segment cannot be saved while the user listens for the closing boundary near another playback location **C**.

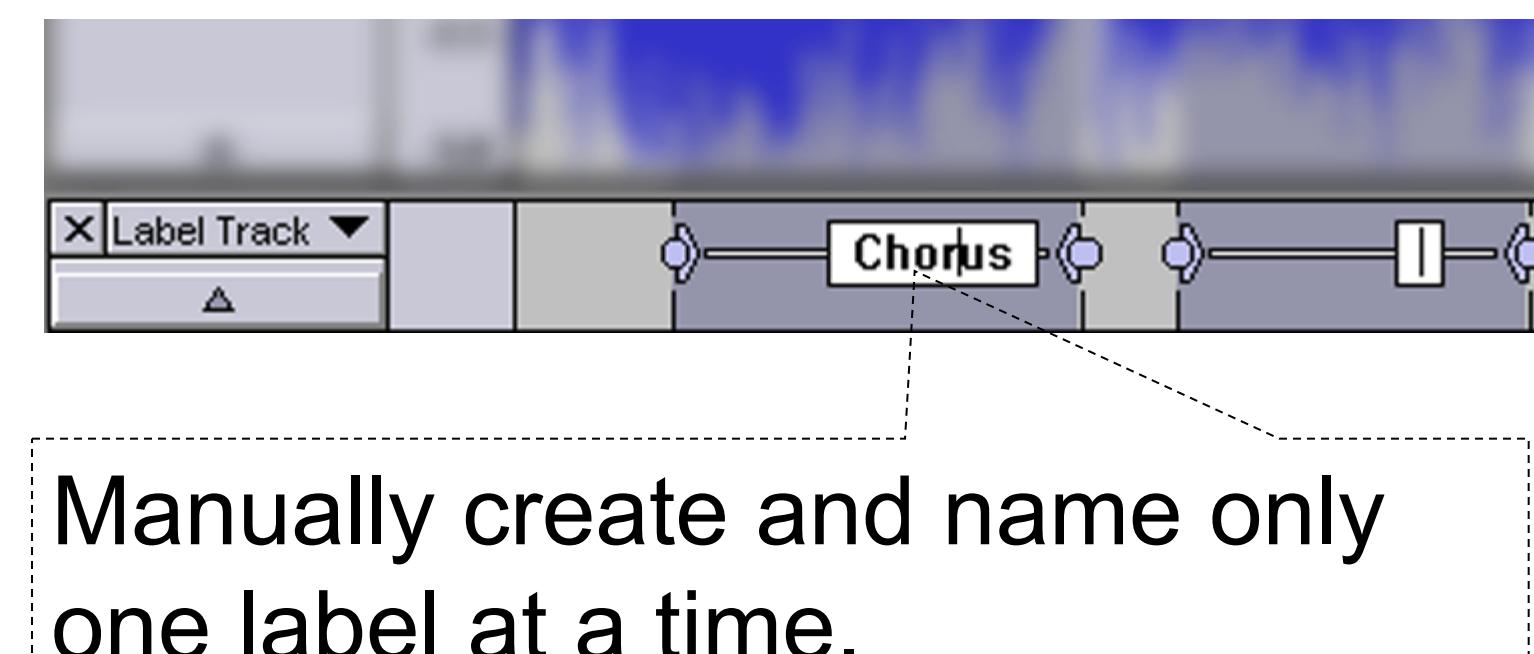
In our extended Audacity:



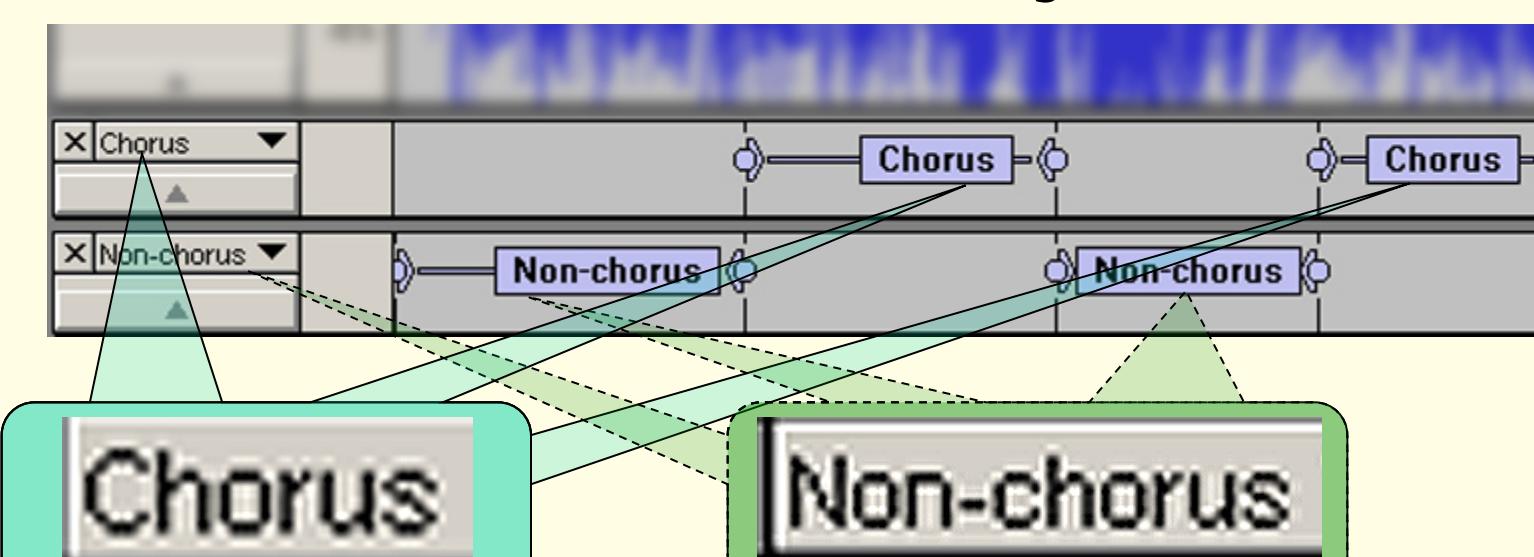
- Cache a location (e.g., **O**) as a candidate opening boundary.
- Cache a location (e.g., **C**) as a candidate closing boundary.
- Finalize both candidate boundaries and create a label.

Label Tracks and Auto-Completion

In classic Audacity:



In our extended Audacity:



Export Results in ACE XML Format

```
<classifications_file>
  <data_set>
    <data_set_id>So Young.mp3</data_set_id>
    <misc_info info_type="Artist">Suede</misc_info>
    <role>training</role>
    <classification>
      <section>
        <start>0</start>
        <stop>43.66146</stop>
        <class>Non-chorus</class>
      </section>
```

Usability Study

Six human subjects:

- All trained musicians
- No specific training for annotation

Six popular songs:

- Three levels of difficulty
- Length range of 3'54" to 4'18"

Block design:

- Original and extended versions of *Audacity*
- No subject annotates the same song more than once.

Subject	Song					
	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Faster Annotation

Three significant factors:

- Annotator (A)
- Selected song (S)
- Version of *Audacity* (V)
- Additive effect on log annotation time (T) with normal errors.

$$E[\log T] = A + S + V$$

Reduced annotation time:

- Average reduction in labelling time of 17.1% with extended *Audacity*
- 95%-confidence range of 7.9% to 25.6% improvement.

Future Work

- Provide more visual cues by visualizing various audio features to human annotators.
- Provide realtime audio effects and enhancement that can help the listening of the human annotator, e.g., variable playback speed in real time.