Extending *Audacity* for Audio Annotation

Beinan Li, John Ashley Burgoyne, and Ichiro Fujinaga
Music Technology Area, Schulich School of Music, McGill University, Montreal, Quebec
beinan.li@mail.mcgill.ca {ashley, ich}@music.mcgill.ca

Abstract

By implementing a cached region selection scheme and automatic label completion, we extended an open-source audio editor to become a more convenient audio annotation tool for tasks such as ground-truth annotation for audio and music classification. A usability experiment was conducted with encouraging preliminary results.

Keywords: audio annotation, classification, usability.

1. Introduction

Providing training data and supporting objective evaluation, ground truth annotation is both an essential and time-consuming procedure in building general pattern classification systems. In audio segmentation and classification field such as music information retrieval (MIR), manual annotation is increasingly cumbersome for the rapidly growing databases of real-world data.

A few efforts have been contributed to creating specialized annotation tools. Some of them focus on annotating low-level musical events such as onsets [1] and drum patterns [2, 3]. Other publicly available tools work well for everyday audio but lack analytical features such as waveform visualization [4, 5]. In our projects such as automatic track segmentation for digitized phonograph records [6] and chorus detection for popular music, the goal of the ground-truth annotation has been to mark the semantic passages in a complete audio stream with time-stamped textual taxonomic labels and to export them in a human- and machine-readable format. No existing tools meet these requirements exactly. We thus started developing our own annotation software based on the open-source audio editor *Audacity* [7]

2. The Choice of Basic Software Framework

A few pieces of existing software that have potential to become audio annotation tools are introduced in [3]. They are either commercial and not customizable [8] or difficult to adapt to music annotation [9]. In contrast, the open-source audio editor *Audacity* is designed for music-oriented audio applications, is cross-platform, supports MIDI and other major audio formats, and is fully customizable. The current version has a built-in function for creating label tracks that are integrated with an active audio track. Users can create one or more time-stamped labels for a selected region of audio data, and the labels can be exported to a text file. These features makes *Audacity* a suitable basic framework for our tool.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2006 University of Victoria

3. Extending *Audacity*

Our starting point is *Audacity 1.3 Beta* (the latest version), in which the label track function has been improved: the labeled region can be adjusted by mouse dragging. However, a few desired features are still missing and therefore we made several extensions.

3.1 Region Selection

For ground-truth-oriented audio annotation in which semantic boundaries should be placed accurately, human annotators must listen repeatedly to the candidate opening and closing areas of a semantic region before making a selection. Like traditional audio editors, *Audacity* allows users to select an audio region by dragging the mouse, to expand selection with the Shift key plus a mouse click and to input time boundaries manually. However, these features share a common drawback of causing an annotator to lose track of any previously located boundaries—for example, the beginning of a pop-music chorus. As a result, an annotator has to memorize at least one of the boundary positions, which wastes annotation time. Our extension introduces an intermediate cache for the user to store any opening or closing boundary positions so that adjustments are easy to make and little short-term memory is required. To facilitate annotating while listening, a playback position is also cacheable as either boundary (start or end) of the target region.

3.2 Label Organization and Automatic Completion

Audacity considers labels to be independent symbols with arbitrary titles that should be manually typed, which is cumbersome and error-prone for large-scale annotation tasks. Allowing heterogeneous labels to reside in the same label track adds to the difficulty in performing batch operations such as renaming or region adjustment for a specific category of labels. With our extension, multi-class annotation is performed by using parallel multiple label tracks, each corresponding to a single category. In the case of a complementary binary classification, an annotator only needs to annotate one category (such as “Chorus” in our tests) through a single label track, and the annotations

for the other category are automatically completed. An additional converter was also implemented to export the labeling data into the ACE XML format [10].

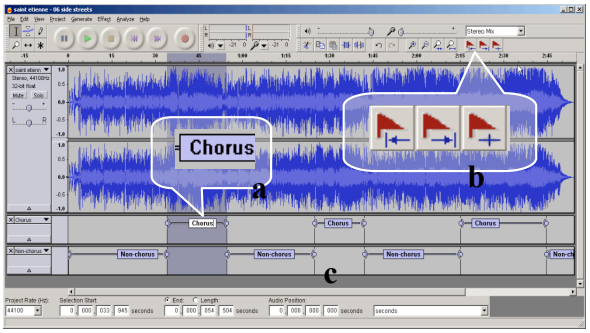

Figure 1. The GUI screen-shot of the extended Audacity. a) A text label in an independent label track. b) Toolbar buttons for the selection. c) Automatic completion.

4. Experiment and Results

A usability experiment was conducted. During the experiment, six annotators with solid musical background (4 years of professional music training) were asked to annotate the chorus and non-chorus passages of six pop songs ranging in length from 3'54" to 4'18" by using both *Audacity 1.3 Beta* (AB) and our extension (AE). The annotation time for each song by each subject was recorded and then analyzed by employing a statistical model. Half the songs are annotated with AB and the other half AE, with the compositions shuffled for each subject. The mean labeling times for a test song for AE and AB were 605 sec. (σ = 189 sec.) and 698 sec. (σ = 221 sec.).

In our model, the usability performance is represented by a normal linear model for log labeling time:

$$E\left[\log(T)\right] = A + S + V \qquad (1)$$

The factors on time consumption include the subject annotator (A), the selected song (S), and the annotation tool in use (V). Errors are assumed to be normal and additive, and a Kolmogorov-Smirnov test on the model residuals shows no significant departures from normality. We also investigated the inclusion of other factors, such as the presumed level of annotation difficulty, mean and standard deviation of the length of the choruses, and the total length of the song; none of these additional factors proved to be significant. Under the chosen model, the use of our version of the software was significant (p = 0.0411) and showed an average reduction in labeling time of 17.1 percent when AE was in use (with the 95-percent confidence interval ranging from 7.9 to 25.6 percent).

Note that during the experiment, AB was reported to have crashed during annotations four times in total and two subjects had to restart their annotation of the affected songs. In these cases, we used the time spent on the second attempt as the resulting time to avoid the performance of AB being artificially degraded.

5. Conclusion and Future Work

An audio annotation tool that is capable of ground-truth annotation for audio and music segmentation and classification is introduced by extending the open-source audio editor *Audacity*. A usability experiment shows that users can perform annotation faster with our extension than with the current version of *Audacity*.

Traditional audio editors offer only limited means of audio visualization, typically waveform and spectrogram. More helpful visual cues derived from audio features that have proven to be reliable semantic indicators might facilitate annotation further. By displaying these data, pre-processed with a tool like ACE [10], the annotator could be aided by a larger number of visual cues.

6. Acknowledgments

We would like to thank CIRMMT, the Canada Foundation for Innovation, and Schulich Scholarship program at McGill University for their generous financial support.

References

[1] P. Leveau P, L. Daudet, and G. Richard. "Methodology and Tools for the Evaluation of Automatic Onset Detection Algorithms in Music," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc.*, 2004, pp. 72–75.

[2] F. Gouyon, N. Wack, and S. Dixon. "An Open Source Tool for Semi-Automatic Rhythmic Annotation," in *DAFx 2004 Seventh Int. Conf. on Digital Audio Effects Proc.*, 2004, pp. 193–196.

[3] K. Tanghe, M. Lesaffre, S. Degroeve, M. Leman, B. De Baets, and J.-P. Martens. "Collecting Ground Truth Annotations for Drum Detection in Polyphonic Music," in *ISMIR 2005 Sixth Int. Conf. on Music Inf. Retr. Proc.*, 2005, pp. 50–57.

[4] Academic Technologies. "Project Pad – Documentation." [Web site] 2006, [2006 July 8], Available: http://dewey.at.northwestern.edu/ppad2/documents/help/audio.html

[5] The Center for Humane Arts, Letters, and Social Sciences Online. "[MediaMatrix]." [Web site] 2006, [2006 July 8], Available: http://www.matrix.msu.edu/~mmatrix/

[6] Marvin Duchow Music Library. "David Edelberg Handel LPs." [Web site] 2006, [2006 April 22], Available: http://coltrane.music.mcgill.ca/handel/lp/search.php

[7] Audacity Development Team. "Audacity: Free Audio Editor and Recorder." [Web site] 2006, [2006 April 22], Available: http://audacity.sourceforge.net/

[8] Twelve Tone Systems, Inc. "SONAR 5" [Web site] 2006, [2006 April 22], Available: http://www.cakewalk.com/Products/SONAR/default.asp

[9] P. Boersma and D. Weenink. "Praat: Doing Phonetics by Computer" [Web site] 2006, [2006 April 22], Available: http://www.fon.hum.uva.nl/praat/

[10] C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga. "ACE: A Framework for Optimizing Music Classification," in *ISMIR 2005 Sixth Int. Conf. on Music Inf. Retr. Proc.*, 2005, pp. 42–49.