RESIC: A Tool for Music Stretching Resistance Estimation

Jun CHEN, Chaokun WANG

School of Software, Tsinghua University, Beijing 100084, China

Tsinghua National Laboratory for Information Science and Technology

Introduction

In scenarios like audio-video synchronization and animation products, music pieces can be stretched in the time domain to a preferred length, say 0.80 times as long as the original one, with the time stretching and the music resizing techniques. However, these techniques suffer from the perceptual artifacts while changing the time length, and the acceptance of the stretched music piece for general audience is not always guaranteed when it is over-stretched (either over-compressed or over-elongated as illustrated in Fig. 1). Thus, the knowledge about the music stretchability of is highly demanded to avoid over-stretching. Specifically, a tool or a computing methodology to estimate the music stretchability is favorable.

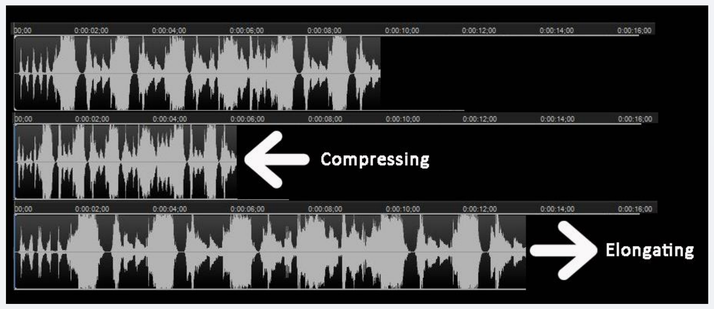

Fig. 1 Music stretching – Compression & Elongation

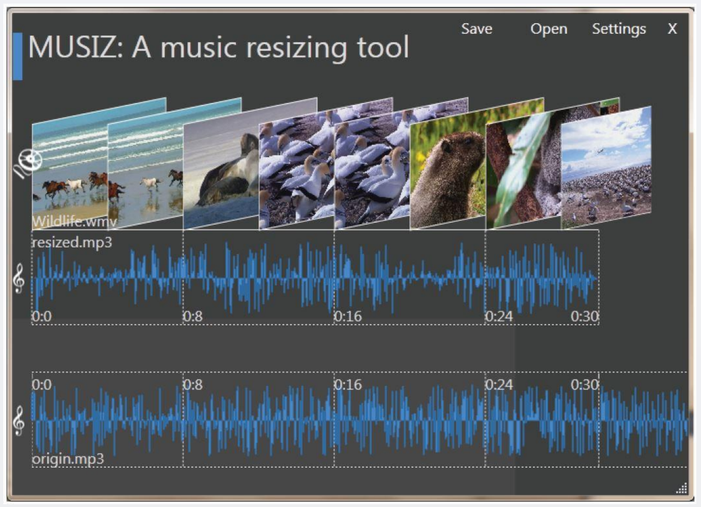

Fig. 2 MUSIZ – A music resizing tool (from Z. Liu, MM'11)

Fig. 2 shows the music resizing tool, *MUSIZ*, which can resize the time length of a music piece to adapt to a target video clip while preserving music features like pitch, timbre and voice quality. However, if the targetted video clip is several times longer or shorter than the original music piece, the resized version of music piece will sound uncomfortable and not acceptable by general audience.

Music stretching may be overdone if the acceptable stretching range is not known. The compression and elongation will change the event density of the music piece as illustrated in Fig. 3, which leads to the stretching boundaries in the time domain considering the human listening. Thus, it is important to estimate the stretchability of music pieces to avoid over-stretching.

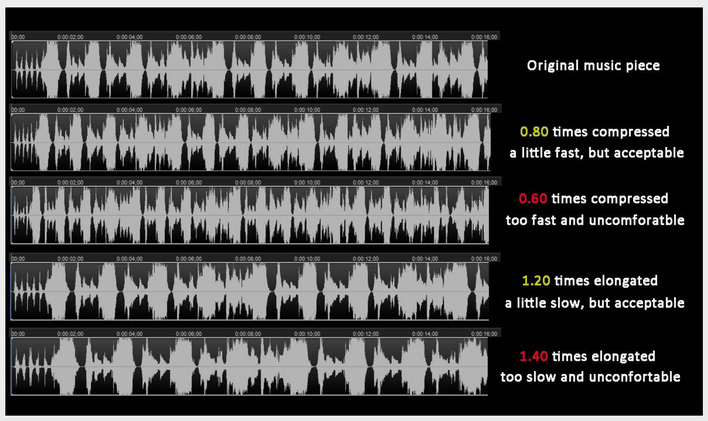

Fig. 3 Music stretching inside/outside the acceptable range

Definition 1 – *Music Stretching Resistance*

Music stretching resistance (MSR) is a new property of music piece, which characterizes the ability of music to be stretched (either compressed or elongated) in the time domain while guaranteeing the acceptance for general audiences. MSR consists of two parameters – the minimal compressing rate α_{min} and the maximal elongating rate α_{max}.

System Overview

RESIC is composed of two parts, the back-end and the front-end. The back-end deals with the audio signal processing and analysis, automatic musical genre classification and MSR estimation, while the front-end interacts with users.

As shown in Fig. 4, the back-end of RESIC contains the following four components:

(1) **Audio Feature Extraction**: We extract overall mean of MFCC, Chroma, Spectral features (Centroid, Rolloff, Flux) and ZeroCrossing rate for the input music piece.

(2) **Normalization**: Normalize the audio features into [0.0, 1.0] range since the value ranges of these features vary a lot, and later metric learning can learn the weights for different dimensions.

(3) **Musical Genre Setting**: To automatically set the genre of the input music piece using musical genre classification methods or manually set it. Musical genre has the ability to describe a family of similar music pieces and distinguish them from the others. Considering that the automatic musical genre classification may induce extra error in the estimation of the musical genre, the manual setting is more suggested if the user has a knowledge about musical genre.

(4) **Estimation**: The estimation is performed by using the metric learning technique using the audio features extracted and normalized in previous steps. We treat the MSR estimation problem as a classification problem, and discretize the numerical stretching rate axis into intervals. Each interval corresponds to a class of MSR-similar music pieces and represents a class label. The KNN classifier is used to estimate the most probable MSR class for a given music piece with the weight parameters obtained through the metric learning.

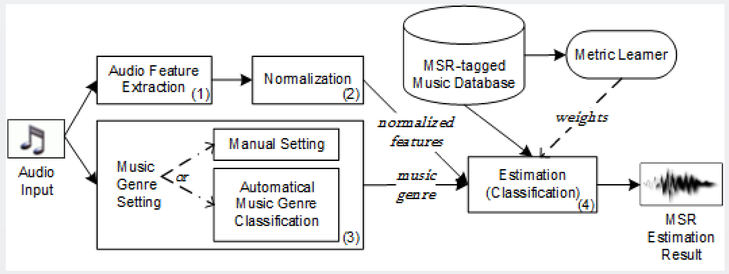

Fig. 4 The back-end architecture and workflow of RESIC

Demonstrations

Two front-ends are implemented as shown in Fig. 5 (stand-alone version) and Fig. 6 (web version). Both front-ends follow the regular routine for a given music piece, i.e., decoding (if needed), audio feature extraction, normalization, musical genre setting and MSR estimation.

The stand-alone version was designed to provide MSR estimation off-line where processing time is guaranteed. The estimation results will be presented intuitively through the color change of the slider bar and also the display of MSR values. The red bars represent the "dangerous" and non-suggested ranges while the green bars mean the "safe" ranges. The red-to-green and green-to-red gradual bars are the transition ranges from "dangerous" to "safe" and from "safe" to "dangerous", respectively. Users can experience what over-compressed or over-elongated music tracks sound like when resizing at rates within red bars, and also what acceptable stretched music tracks sound like when resizing within green bars. Thus, once they have a music track to resize, RESIC will alert to them what is the minimum (or maximum) rate it can be compressed (or elongated). Besides the stand-alone version, the web version takes full advantage of the remote interaction through a web browser with a friendly user interface, related audio, video samples and readings.

The main contribution of this demonstration is to provide a tool which is capable of estimating the MSR values of music pieces. Also, it can be used to estimate MSR values for different segments of the same music track and resize it non-homogeneously.

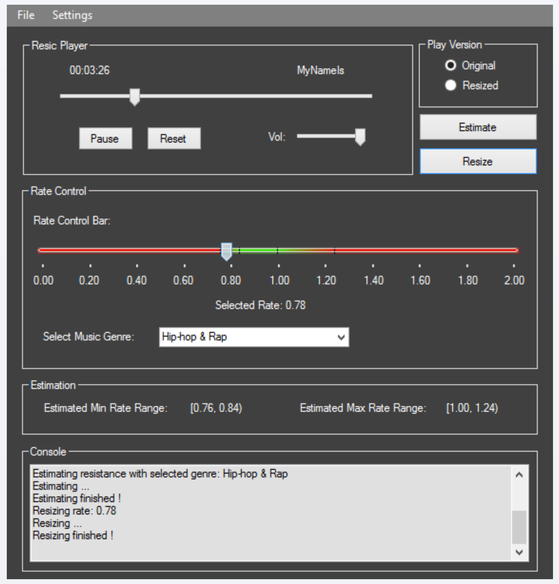

Fig. 5 Stand-alone version of RESIC

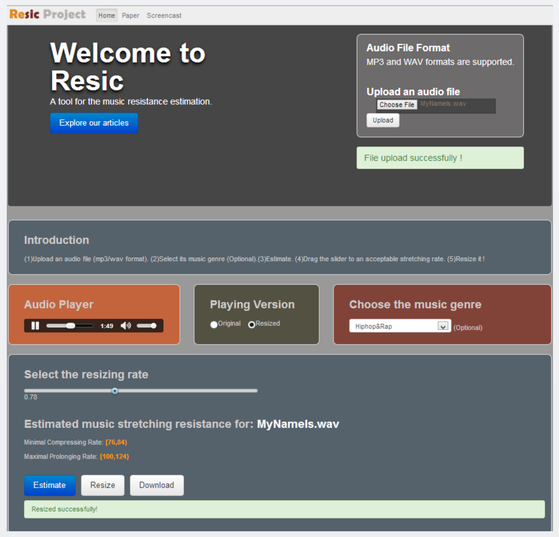

Fig. 6 Web version of RESIC

REFERENCES

1. Liu, Z., Wang, C., Wang J., Wang, H. and Bai, Y.: Adaptive music resizing with stretching, cropping and insertion. Multimedia Systems. 19(4), pp. 359-380. (2013)
2. Liu, Z., Wang, C., Bai, Y., Wang, H. and Wang, J.: MUSIZ: a generic framework for music resizing with stretching and cropping. ACM Multimedia. pp. 523-532. (2011)
3. Liu, Z., Wang, C., Guo, L., Bai, Y. and Wang J.: Lydar: a lyrics density based approach to non-homogeneous music resizing. IEEE ICME. pp. 310-315. (2010)
4. Xing, E., Ng, A., Jordan, M. and Russell, S.: Distance metric learning, with appli-cation to clustering with side-information. NIPS. pp. 505-512. (2002)

ACKNOWLEDGEMENT

The work is supported by the National Natural Science Foundation of China (No. 61373023, No. 61170064, No. 60803016), National High Technology Research and Development Program of China (No. 2013AA013204), and the Opening Foundation of Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, the Chinese Academy of Sciences under Grant (No. IIP 2012-5).

CONTACT

Email: chenjun12@mails.thu.edu.cn, chaokun@tsinghua.edu.cn.

Address: Room 11-421, East Main Building, Tsinghua University, Beiing 100084, China.