Audio Transport

Trevor Henderson

April 26, 2017

Abstract

yoyoyo.

1   Introduction

2   Related Work

Vocoders existed in the 70's. vocoders take a monophonic source and spread it out over a range of pitches. requires harmonic similarity between the sources. Voices have lots of noise so it is easy - Daft punk.

realtime pitch shifting has become very popular in software and hardware. since the early 2000's. Traditional pitch shifters were granular. Updates to algorithms use phase vocoders.

Real time pitch shifting is a key element to autotune, which charachterized the music of the mid 2000's and is still popular today.

Harmonizers sound similar to vocoders. Take a pitch and pitch shift it into new chords.

In 2008 Melodyne pitch shifts individual elements without distoring the sound. They must be processed beforehand and then can be manipulated in realtime

The 2012 product can manipulate polyphonic audio in realtime. http://www.zynaptiq.com/pitch

Most of this work has been proprietary and out of research.

So far as I can

processing but no synthesis. `http://www.maths.lu.se/fileadmin/maths/personal_staff/Andreas_Jakobsson/ElvanderAKJ16_icassp_final.pdf`

http://remi.flamary.com/biblio/flamary2016optimal.pdf

1

3   Optimal Transport

Discussion of optimal transport.

Distance.

It also allows for you to interpolate between the two shapes along the trajectory.

In 1 dimension it is a solved problem.

For any dimensionality higher than 1, fast computation of the wasserstein2 distance is

Optimal transport is however only defined for, nonnegative masses. Audio has phase. Initial attempts to tackle the problem of phase involved formulating the transport problem differently. But you can't hear the distance between two sinusoids that are played at the same pitch but at different phases But you can hear the "distance" between two pitches.

4   Phase Vocoders

As noted by, audio quality is .

Vertical and horizontal incoherence.

5   Algorithm Overview

5.1   STFT

Weighted overlap add root hann window Let M be the window size, R be the hopsize.

We are guaranteed perfect reconstruction in the absense of spectral modifications if

$$\sum_m w^2(n - mR) = 1, \ \forall n \in \mathbb{Z}$$

I choose the to use the root Hann window.

$$w(n) = \cos(\pi n/M), n \in \left[-\frac{M-1}{2}, \frac{M-1}{2} \right]$$

5.2 Spectrum Segmentation

In order to fix the spec Techniques mention peak finding. However peak finding does not give a great sense of where to put boundaries. The reassigned frequency can be computed by:

$$\hat{\omega}(\omega) = \omega + \Im\left\{\frac{X_{\mathcal{D}}(\omega) \cdot X^*(\omega)}{|X(\omega)|^2}\right\} \tag{1}$$

where $X_{\mathcal{D}}$ is the spectrum of the x with the window $w_{\mathcal{D}}(n) = \frac{d}{dt}w(n)$:

$$X_{\mathcal{D}}(\omega) = \text{FFT}\left(w_{\mathcal{D}}(n)x(n)\right) \tag{2}$$

d(root hann window)

$$w_{\mathcal{D}}(n) = -\frac{\pi f_s}{M}\sin\left(\frac{\pi n}{M}\right)$$

Nice plot.

5.3 Phase Accumulation

Horizontal coherence.

$$\theta' = \theta + \frac{M}{f_s}\hat{\omega}$$

5.4 Transport

Combinations of frequencies: rotating in the same direction

$$X_{01}(\omega) = X_0(\omega)^{1-k}X_1(\omega)^k \tag{3}$$

This maintains that the energy of the signal is constant and that the phase rotates from one state to the other. However, this rotation can happen in multiple directions:

6 Implementation

Real time. `fftw`
   `portaudio portmidi`
   video
   github link

7 Conclusion

8 References

Fundemental theory
    that phase vocoder one
    1d transport paper?
    STFT
    weighted overlap add
    fftw
    portaudio
    portmidi
    cite justin solomon