COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Tunable Music

Haiyu Huang

E6820: Speech & Audio Processing & Recognition

Advisor: Prof. Dan Ellis

Why music need to be tunable?

Once a music track has been made, it is forever fixed and we are unable to edit it, especially to change its melody, even if the song really need to be improved or the listeners really want some creative input to that expression, or, even if music works really need constant changes through the time to prevent people from being tired of it.

To do list

1. Make the real digital records (wav,mp3) tunable, with higher performance and computation cost comparing to existing methods and software.
2. Setup a tunable music database system to support an online interactive music service and produce.
3. Explore the possibility to enable the computer automatically edit music tracks. The changes generated by computers would be pseudorandom or based on learning algorithm but should be enjoyable to human listeners.

Pitch Shifting Research

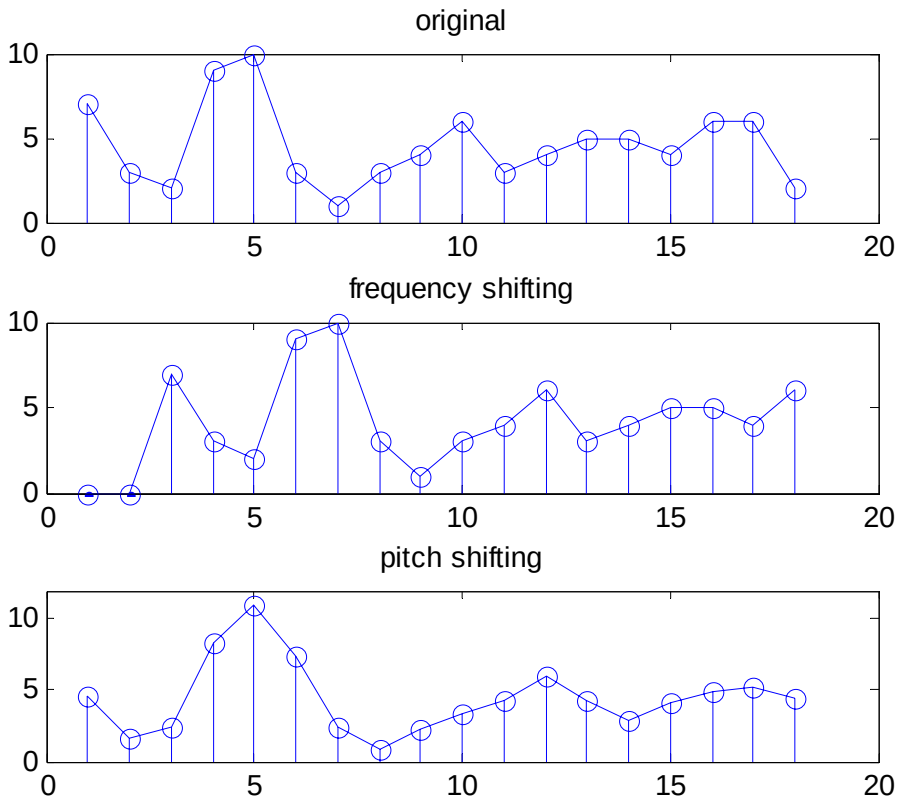
- As opposed to the process of pitch transposition achieved using a simple sample rate conversion, Pitch Shifting is a way to change the pitch of a signal without changing its length.
- In most practical applications, pitch shift is achieved by changing the length of a sound and then performing a sample rate conversion to change the pitch. This is based on Time Domain Harmonic Scaling.
- Time Compression/Expansion, as what we have learned in lecture 1, is the reciprocal process to Pitch Shifting. So we could also realize pitch shifting via STFT, in the frequency domain. (This is what I choose)

Pitch Shifting vs Frequency Shifting

- A true Frequency Shift (as obtainable by modulating an analytic signal by a complex exponential) will shift the spectrum of a sound, while Pitch Shifting will dilate it, retaining the harmonic relationship of the sound. Frequency Shifting yields a metallic, inharmonic sound which may well be an interesting special effect but which is a totally inadequate process for changing the pitch of any harmonic sound except a single sine wave.
- Pitch Shifting is more difficult to achieve than frequency shifting if conducting in the frequency domain.

Pitch Shifting vs Frequency Shifting

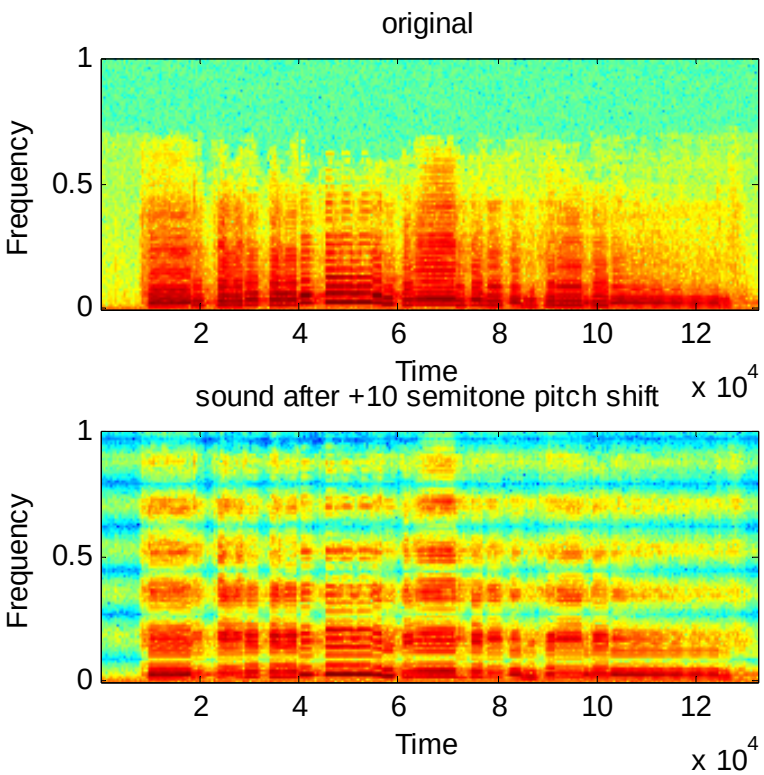
- Frequency shift is linear in spectrum while pitch shift is non- linear transformation.
- Because of finite resolution, we need numerical method to achieve pitch shift.
- Here I use linear interpolation resampling

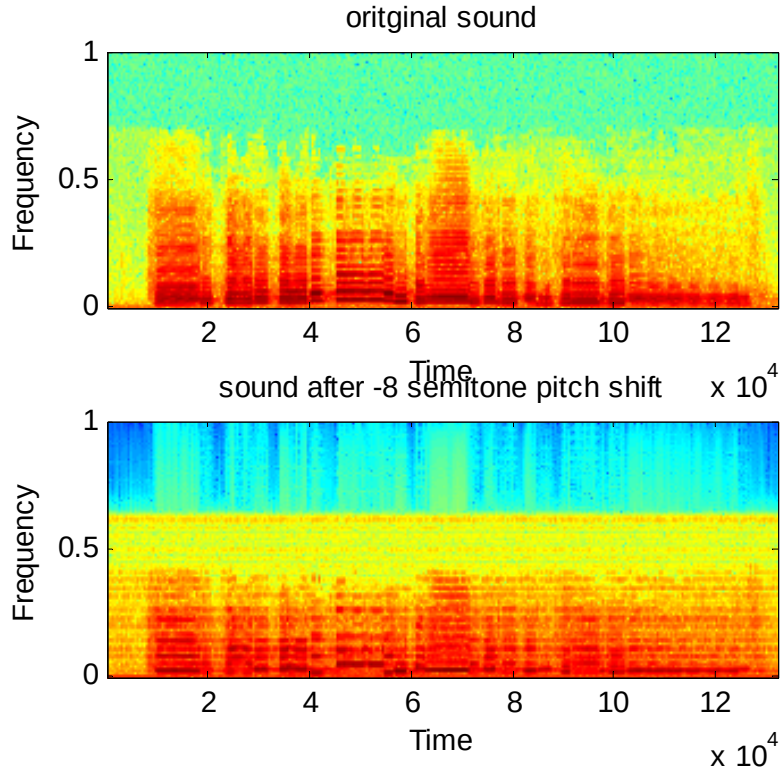


Pitch Shifting Research

- As opposed to the process of pitch transposition achieved using a simple sample rate conversion, Pitch Shifting is a way to change the pitch of a signal without changing its length.
- In most practical applications, pitch shift is usually achieved by changing the length of a sound and then performing a sample rate conversion to change the pitch.

Results using Interpolation



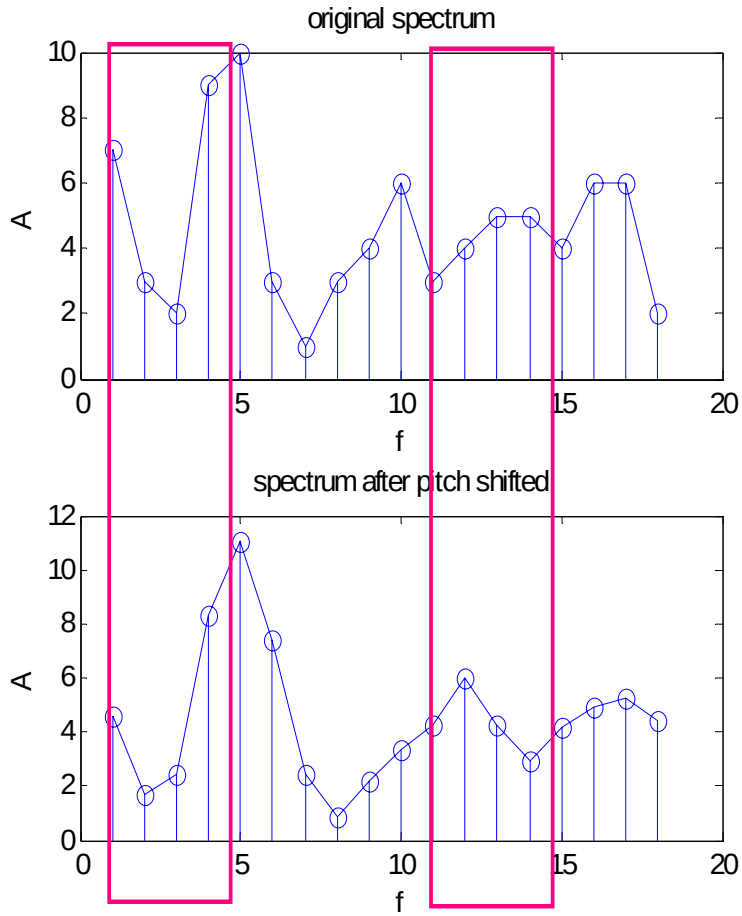


Original Shifted by +10 Original Shifted by -8

Doesn't change the pitch, but the instrument timbre instead !!

Why?

- Exact reason still unknown.
- One possible reason is because of the lower resolution of pitch shift in the small f range and higher resolution in the large f range.
- Resolution of human hearing perception is not constant along the frequency domain !!



Improve DFT for Cepstral domain

We define the "discrete pitch transform" (DPT) as

DPT $$X(p) \triangleq \sum_{n=1}^{N-1} x(n) \bullet \exp(-j2\pi f(p) n / N)$$

IDPT $$x(n) \triangleq \frac{1}{N} \sum_{p=p_1}^{p_N} X(p) \bullet \exp(j2\pi f(p) n / N)$$

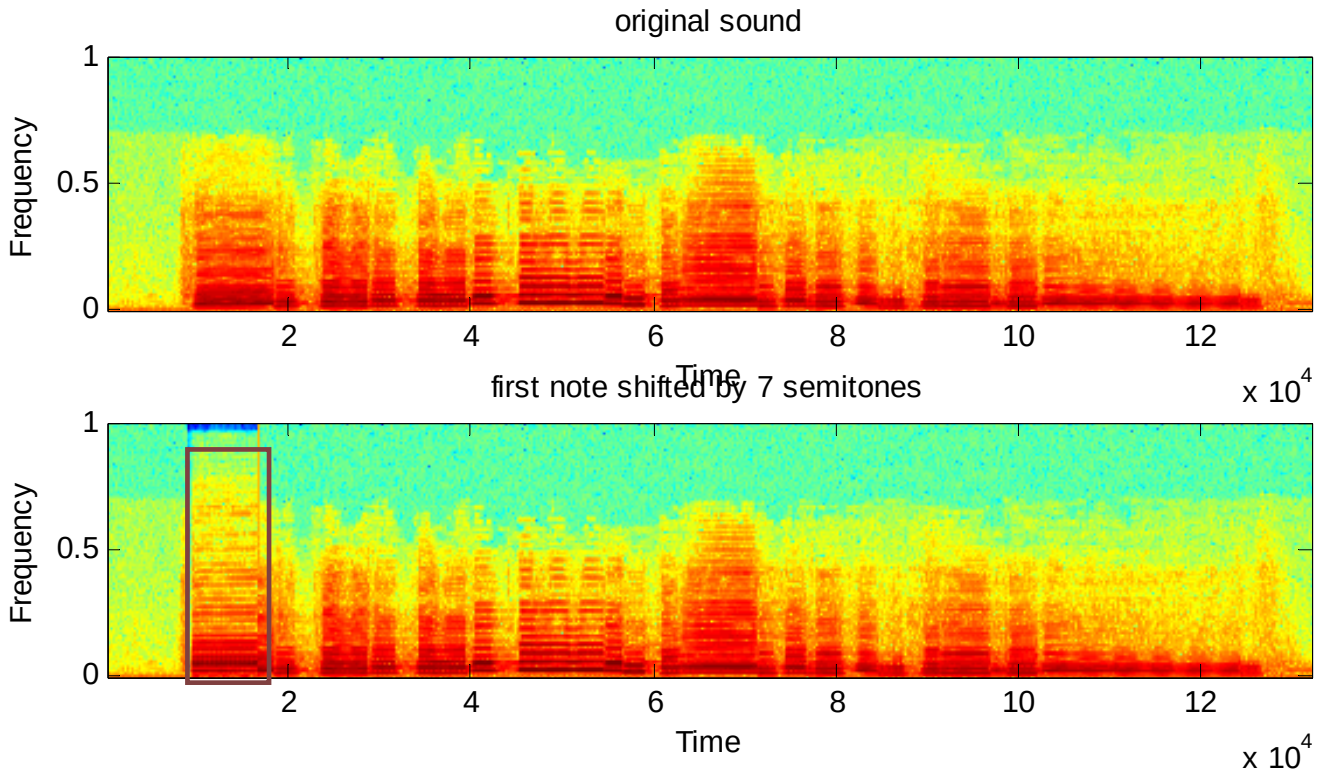
Where $$f(p) = 440 \bullet 2^{\frac{p-69}{12}} \qquad p_i = p_N - \Delta p \bullet (N - i)$$

$$p_N = 69 + 12 \bullet \log_2(\frac{\mathrm{Fs}/2}{440}) \quad , \ \Delta p \text{ is predefined}$$

Improve DFT for Cepstral domain

- The short- time pitch transform (STPT) can be easily derived from STFT, in which we replace the step of DFT and IDFT by DPT and IDPT respectively.
- The result of DPT has constant resolution in pitch domain, or in another word, has unequal distribution in the frequency domain, which matches the human hearing system better.
- Advantage for pitch shift: In cepstral domain (or pitch domain), pitch shift would be linear transform!
- Results: Original single pitch shifted by +7 shifted by +10

Application to Music Recomposition



Original

Shifted by +10

Why the shifted pitch sounds a little weird?-IDPT definition was wrong!

IDPT $x(n) = \frac{1}{N} \sum_{p=p_1}^{p_N} X(p) \bullet \exp(j2\pi f(p)n/N)$

The vectors $\exp(-j2\pi f(p)n/N)$ do not form an orthogonal basis over the set of N-dimensional complex vectors:

$$\sum_{n=1}^{N-1} \exp(-j2\pi f(p)n/N) \bullet \exp(j2\pi f(q)n/N) \neq N\delta_{pq}$$

which means $x(n)$ cannot be recovered from this expression. Must try to form an orthogonal basis!

Derive IDPT from continuous FT

Inverse FT: $x(t) = \int_{-\infty}^{\infty} X(f) \bullet \exp(j2\pi ft) df$

Define $f(p) = 440 \bullet 2^{\frac{p-69}{12}}$. For real $x(t)$: $X(-f) = X^*(f)$

$$x(t) = \int_0^{\infty} (X(f) \bullet \exp(j2\pi ft) + X^*(f) \bullet \exp(-j2\pi ft)) df$$

$$= \int_0^{\infty} (X(p) \bullet \exp(j2\pi f(p)t) + X^*(p) \bullet \exp(-j2\pi f(p)t)) \bullet f(p) \bullet \frac{\ln 2}{12} dp \Rightarrow$$

$$x[n] = \sum_{p=p_1}^{p_N} (X(p) e^{j2\pi f(p)\frac{n}{N}} + X^*(p) e^{-j2\pi f(p)\frac{n}{N}}) f(p) \frac{\ln 2}{12} \Delta p$$

What still need to do with DPT

- Proof that the vectors in IDPT expression do form an orthogonal basis.
- Experimentally testify the validation of the new IDPT expression.
- Derive a fast computing algorithm for DPT like FFT does.
- Apply the correct DPT and IDPT and its fast computation method to pitch shift tasks. Hopefully build up DPT based pitch shift software for mp3 file with friendly GUI.

A Novel Music Database

- Different from any other music database. It forms a interactive music creating and evaluating community in which the customer is no longer a passive listener. Anyone can rate the music tracks, and edit the original version as a composer or lyrics writer.
- The most distinguishing idea is, all the evaluation and edition feedbacks will be considered by the music companies to improve the music tracks and hence produce newer editions which contains contribution from customers. Of course, customers will also gain corresponding profits from that.
- If this platform becomes popular, traditional music industry will be changed. Music quality will become much better than any time before. The underlying enthusiasm of everyone to create music will eventually escalate!!!

Proposal for the database

- There are five major entities involved in the design of the database – Music Companies, Musicians, Professional Editors, Music Tracks, and Customers.
- A single music track is created by only one musician and published by only one music company, while it can be edited by one or more customers as generate different customer's editions.
- The music company will improve the original music based on the customer's editions and all the evaluation result. Hence a final version of a music track will be released containg the portion (percentage) of contribution of each customers.
- Any customer with contribution will gain corresponding profits when the final music track is sold.

Entity Relationship diagram Tables and SQL codes

Next step with the database

- Put the database online by the end of spring break.
- Collect large amount of editable music tracks for testing service. Because mp3 files are still untunable so far, maybe I will first put MIDI format files instead.
- Seek for help from real musicians, request them to put their own original works in the database system if possible.
- Attract more people to participate.

Computer musicians

- Automatic Composition -- Idea cool but extremely difficult to realize.
- I have tried artificial neuron networks to train the model but totally failed.
- Other researchers, such as David Cope and David Birchfield have tried different methods to “teach” computer to create music here is a sample track generated by computer using genetic algorithm.
- There are rarely computational aesthetic models for music composition, so generating totally new song by computer is impossible yet.

Computer Improvisation

- Computer improvisation is different from computer composition, the rhythm and baseline is fixed, only some certain parts of melody can be adjusted.
- Computer improvisation is comparatively realizable. The simplest way is to treat pitch note sequence as digital signal, and the computer generate pseudorandom pitch shifts to the note points which themselves are also pseudorandomly selected by computer.
- There should be some constraints applying to computer improvisation. First, no pitch shift can be more than 12 semitones. Second, no more than 20% of notes can be shifted during one playing. Last and most important, we should keep the “envelope” of the original note sequence while only change the “fine structure”.

Computer Improvisation

- According to aesthetic and psychological theory, people would like stuffs that are new but not that new. People will be bored with music if the melody is forever fix. Computer improvisational edit might be an alternative way for people enjoying music in the future.
- Find out the optimum “novelty” coefficient will significantly improve the performance of computer improvisation.
- If we can collect the information on how people usually change the melody from the tunable music database, we can train the computer do the similar thing, and hence ensure what the computers do will match the favors of large population.

Summary & Discussion

- . Changed the timebre of a music track (although not intended to do it) using linear interpolating
- . Improved DFT, derive a new transform named “discrete pitch transform” in order to realize non- uniform resolution distribution in the frequency domain while uniform in pitch domain, which enable fast pitch shifting with good quality
- . Established a tunable music database for interactive music service.
- . Made computer change the pitch improvisationally, generating a kind of “random music”, which provide a fresh concept for music appreciation.

Any Questions or Suggestions??

Thanks for your attention !