Audio Processing

Chaiwoot Boonyasiriwat
October 8, 2020
An example of an audio processing system is given as

- In ADC converters, analog sound signal is first filtered by an anti-aliasing filter to prevent aliasing before it is sampled.

---

Christensen (2019, p.35)
Audio Sampling Frequencies

Common audio sampling frequencies and examples of their applications are given below.

<table>
<thead>
<tr>
<th>Sampling frequency</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 kHz</td>
<td>Narrowband speech, telephony</td>
</tr>
<tr>
<td>16 kHz</td>
<td>Wideband speech, telephony, VoIP</td>
</tr>
<tr>
<td>44.1 kHz</td>
<td>CD, audio equipment, sound cards</td>
</tr>
<tr>
<td>48 kHz</td>
<td>DAT, video recorders</td>
</tr>
<tr>
<td>96 kHz</td>
<td>DVD-audio, HD DVD, Blu-ray</td>
</tr>
<tr>
<td>192 kHz</td>
<td>DVD-audio, Blu-ray</td>
</tr>
</tbody>
</table>

Christensen (2019, p.36)
A note is a symbol denoting a musical sound.

A note can represent the pitch of a sound in musical notation or a pitch class (e.g., A, B, C, D, E, F, G).

A pitch is a perceptual property of sounds and is closely related to frequency. High pitch → high frequency.

Notes are the building blocks of music.

A0 = 27.5 Hz
A1 = 55.0 Hz
A2 = 110 Hz
A3 = 220 Hz
A4 = 440 Hz
Music Theory

- Most Western music is based on the twelve-tone equal temperament (TET) which is a tuning system that has 12 notes or semitones (C, C#/D♭, D, D#/E♭, E, F, F#/G♭, G, G#/A♭, A, A#/B♭, B) within an octave.

- An octave is the interval between one musical pitch and another with double its frequency. For example, A₃ and A₄ are one octave apart.

- The ratio of the frequency of two consecutive notes, e.g., C₄ and C♯₄, is always equal to $\sqrt[12]{2}$.

- A₄ (440 Hz) is the reference note.

- Any note can be expressed as $f(k) = 2^{k/12} \times 440$ Hz where $k$ is an integer.

Christensen (2019, p.69)
Music Theory

- The ratio between two frequencies is called an interval in music and can also be thought of as a difference on a logarithmic scale.

- Since human perceives sounds in a logarithmic scale, equal intervals are perceived as a difference in pitch.

- An interval can be measured in terms of semitones, octaves, or cents.

- Cent is a sub-semitone unit. There are 100 cents per semitone, i.e., 1200 cents per octave.

- The interval between two frequencies $f_1$ and $f_2$ can be computed in cents as

$$\Delta = 1200 \log_2 \left( \frac{f_1}{f_2} \right)$$

Christensen (2019, p.70)
In MIDI Tuning Standard, a pitch denoted as $F_0$ is computed by

$$F_0 = 69 + 12 \log_2 \left( \frac{f_0}{440} \right)$$

where $f_0$ is the pitch in Hz.

- When $f_0$ is equal to a semitone, $F_0$ is an integer.
- $A_4$ (440 Hz) corresponds to MIDI note 69.
Music Theory

- “A scale is a set of notes defined by the intervals of the notes in relative to the root note or tonic.”
- “For example, the A minor scale, where A is the root note, comprises A, B, C, D, E, F#, and G where the intervals between consecutive notes, expressed in semitones, are 2, 1, 2, 2, 1, 2.”
- “A **chord** is a set of two or more notes played simultaneously.”
- For example, A minor chord consists of the notes A, C, and D (root, 3rd, 5th notes of the A minor scale).
- A major chord consists of the notes A, C#, and D (root, 3rd, 5th notes of the A major scale).

Christensen (2019, p.71)
Audio Effect: Echo

- A single echo can be generated by an inverse comb filter represented by the difference equation
  \[ y_k = x_k + c x_{k-d} \]
  where \( c < 1 \) determines how loud the echo is relative to the original sound, and \( d \) is the delay time in samples.

- Multiple echoes can be generated by
  \[ y_k = x_k + c_1 x_{k-d_1} + c_2 x_{k-d_2} + c_3 x_{k-d_3} + \cdots \]
  representing multiple inverse comb filters connected in parallel.

- Multiple echoes can also be generated by a comb filter
  \[ y_k = x_k + cy_{k-d} \]

Christensen (2019, p.120)
Audio Effect: Vibrato

- Vibrato is a sound effect generated by time-varying delay which is a frequency modulation (FM)
  \[ y_k = x_{k-d(k)} \]

- “The delay \( d(k) \) is typically in the range of 0 – 10 ms while it varies at a frequency of 0.1 – 5 Hz.

- An example of a delay function is
  \[ d(k) = \frac{D}{2} \left(1 - \cos\left(\frac{2\pi kf}{f_s}\right)\right) \]
  where \( D \) is called the depth measured in samples and \( f \) is the frequency (speed) of the time-varying delay. The value of \( d(k) \) will vary between 0 and \( D \).

Christensen (2019, p.120)
Audio Effect: Vibrato

Christensen (2019, p.122)
Audio Effect: Tremolo

- Tremolo is a amplitude-modulation (AM) sound effect which can be generated by the filter

\[ y_k = x_k m_k \]

where \( m_k \) is called the modulating signal and \( x_k \) is called the carrier.

- For the tremolo effect, the modulating signal has the form

\[ m_k = 1 - A \cos \left( \frac{2\pi nf}{f_s} \right) \]

where \( f < 20 \) Hz and \( 0 < A \leq 1 \).

- “If \( f \) is too high, the effect will not be perceived as a time-varying loudness but as adding roughness to the input signal.”

Christensen (2019, p.123)
Audio Effect: Tremolo

---

Christensen (2019, p.124)
Audio Effect: Chorus

- “The chorus effect imitates the effect of several musical instruments playing the same part while not being completely in sync and playing at the same volume.”
- To emulate two instruments playing together, we can use an inverse comb filter
  \[ y_k = x_k + g x_{k-d(k)} \]
- Here, the time-varying delay \( d(k) \) must be so small that they are not perceived as distinct echoes.
- To emulate multiple instruments playing together, we can use multiple comb filters connected in parallel
  \[ y_k = x_k + g_1 x_{k-d_1(k)} + g_2 x_{k-d_2(k)} + g_3 x_{k-d_3(k)} + \cdots \]

Christensen (2019, p.125)
Audio Effect: Chorus

- The delay function can be

\[ d_i(k) = F_i + \frac{D_i}{2} \left( 1 - \cos\left(\frac{2\pi nf_i}{f_s}\right) \right) \]

where \( F_i \) is the delay offset in samples, \( D_i \) is the depth in samples, and \( f_i \) the frequency.

- Typical values for these parameters are \( F \) corresponding to 10 ms, frequency \( f \) of 0.2 Hz, and depth \( D \) corresponding to 20 ms.

Christensen (2019, p.126)
References

- Christensen, M. G., 2019, Introduction to Audio Processing, Springer.
- http://www.ee.columbia.edu/~ronw/dsp/
- https://pages.mtu.edu/~suits/notefreqs.html
- https://en.wikipedia.org/wiki/Guitar_tunings