1 Music and nonspeech

- What is ‘nonspeech’?
  - according to research effort: a little music
  - in the world: most everything

<table>
<thead>
<tr>
<th>high</th>
<th>music</th>
</tr>
</thead>
<tbody>
<tr>
<td>speech</td>
<td></td>
</tr>
<tr>
<td>animal</td>
<td></td>
</tr>
<tr>
<td>sounds</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>low</th>
<th>contact/collision</th>
</tr>
</thead>
<tbody>
<tr>
<td>wind &amp; water</td>
<td></td>
</tr>
<tr>
<td>natural</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>man-made</th>
<th>machines &amp; engines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sound attributes

- **Attributes suggest model parameters**
- **What do we notice about ‘general’ sound?**
  - psychophysics: pitch, loudness, ‘timbre’
  - bright/dull; sharp/soft; grating/soothing
  - sound is not ‘abstract’:
    - tendency is to describe by source-events

- **Ecological perspective**
  - what matters about sound is ‘what happened’
  - our percepts express this more-or-less directly

Motivations for modeling

- **Describe/classify**
  - cast sound into model because want to use the resulting parameters
- **Store/transmit**
  - model implicitly exploits limited structure of signal
- **Resynthesize/modify**
  - model separates out interesting parameters
Analysis and synthesis

- **Analysis** is the converse of **synthesis**:

  ![Diagram](image)

  - Model / representation
  - Sound
  - Synthesis
  - Analysis

- **Can exist apart:**
  - analysis for classification
  - synthesis of artificial sounds

- **Often used together:**
  - encoding/decoding of compressed formats
  - resynthesis based on analyses
  - analysis-by-synthesis

Outline

1. Music and nonspeech
2. **Environmental sounds**
   - Collision sounds
   - Sound textures
3. Music synthesis techniques
4. Sinewave synthesis
5. Music analysis
Environmental Sounds

- Where sound comes from: **mechanical interactions**
  - contact / collisions
  - rubbing / scraping
  - ringing / vibrating

- **Interest in environmental sounds**
  - carry information about events around us
    - including indirect hints
  - need to create them in virtual environments
    - including soundtracks

- **Approaches to synthesis**
  - recording / sampling
  - synthesis algorithms

Collision sounds

- **Factors influencing:**
  - colliding **bodies**: size, material, damping
  - local properties at **contact** point (hardness)
  - **energy** of collision

- **Source-filter model**
  - “**source**” = excitation of collision event
    (energy, local properties at contact)
  - “**filter**” = resonance and radiation of energy
    (body properties)

- **Variety of strike/scraping sounds**
  - resonant freqs ~ size/shape
  - damping ~ material
  - HF content in excitation/strike ~ mallet, force

(from Gaver 1993)
Sound textures

- **What do we hear in:**
  - a city street
  - a symphony orchestra

- **How do we distinguish:**
  - waterfall
  - rainfall
  - applause
  - static

- **Levels of ecological description...**

Sound texture modeling (Athineos)

- **Model broad spectral structure with LPC**
  - could just resynthesize with noise

- **Model fine temporal structure in residual with linear prediction in time domain**
  - precise dual of LPC in frequency
  - ‘poles’ model temporal events

- **Allows modification / synthesis?**
Outline

1. Music and nonspeech
2. Environmental sounds
3. Music synthesis techniques
   - Framework
   - Historical development
4. Sinewave synthesis
5. Music analysis

Music synthesis techniques

- **What is music?**
  - could be anything → flexible synthesis needed!

- **Key elements of conventional music**
  - instruments
    → note-events (time, pitch, accent level)
    → melody, harmony, rhythm
  - patterns of repetition & variation

- **Synthesis framework:**
  instruments: common framework for many notes
  score: sequence of (time, pitch, level) note events
The nature of musical instrument notes

- Characterized by instrument (register), note, loudness/emphasis, articulation...

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Time</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piano</td>
<td>0-1</td>
<td>0-4000</td>
</tr>
<tr>
<td>Violin</td>
<td>0-1</td>
<td>0-4000</td>
</tr>
<tr>
<td>Clarinet</td>
<td>0-1</td>
<td>0-4000</td>
</tr>
<tr>
<td>Trumpet</td>
<td>0-1</td>
<td>0-4000</td>
</tr>
</tbody>
</table>

distinguish how?

Development of music synthesis

- **Goals of music synthesis:**
  - generate realistic / pleasant new notes
  - control / explore timbre (quality)
- **Earliest computer systems in 1960s**
  (voice synthesis, algorithmic)
- **Pure synthesis approaches:**
  - 1970s: Analog synths
  - 1980s: FM (Stanford/Yamaha)
  - 1990s: Physical modeling, hybrids
- **Analysis-synthesis methods:**
  - sampling / wavetables
  - sinusoid modeling
  - harmonics + noise (+ transients)
  others?
**Analog synthesis**

- The minimum to make an 'interesting' sound

  ![Analog synthesis diagram]

  **Elements:**
  - harmonics-rich oscillators
  - time-varying filters
  - time-varying envelope
  - modulation: low frequency + envelope-based

  **Result:**
  - time-varying spectrum, independent pitch

**FM synthesis**

- Fast frequency modulation → sidebands:

  \[
  \cos(\omega_c t + \beta \sin(\omega_m t)) = \sum_{n=-\infty}^{\infty} J_n(\beta) \cos((\omega_c + n\omega_m)t)
  \]

  - a harmonic series if \( \omega_c = r\omega_m \)

- \( J_n(\beta) \) is a Bessel function:

  ![Bessel function graph]

  \( J_n(\beta) \approx 0 \)
  
  for \( \beta < n - 2 \)

  \( \omega_c = 2000Hz \)
  
  \( \omega_m = 200Hz \)

  ![Complex harmonic spectra by varying \( \beta \)]
Sampling synthesis

- **Resynthesis from real notes**
  → vary pitch, duration, level

- **Pitch**: stretch (resample) waveform

- **Duration**: loop a ‘sustain’ section

- **Level**: cross-fade different examples

- need to ‘line up’ source samples

Outline

1. Music and nonspeech
2. Environmental sounds
3. Music synthesis techniques
4. **Sinewave synthesis** (detail)
   - Sinewave modeling
   - Sines + residual ...
5. Music analysis
Sinewave synthesis

- If patterns of harmonics are what matter, why not generate them all explicitly:
  \[ s[n] = \sum_k A_k[n] \cos(k \cdot \omega_0[n] \cdot n) \]
  - particularly powerful model for pitched signals

- Analysis (as with speech):
  - find peaks in STFT \(|S[\omega,n]|\) & track
  - or track fundamental \(\omega_0\) (harmonics / autoco)
    & sample STFT at \(k \cdot \omega_0\)
  \(\rightarrow\) set of \(A_k[n]\) to duplicate tone:

- Synthesis via bank of oscillators

Steps to sinewave modeling - 1

- The underlying STFT:
  \[ X[k, n_0] = \sum_{n=0}^{N-1} x[n + n_0] \cdot w[n] \cdot \exp\left(-j \frac{2\pi k n}{N}\right) \]

  What value for \(N\) (FFT length & window size)?
  What value for \(H\) (hop size: \(n_0 = rH, r = 0, 1, 2...\))?

- STFT window length determines freq. resol’n:
  \[ X_w(e^{j\omega}) = X(e^{j\omega}) \ast W(e^{j\omega}) \]

- Choose \(N\) long enough to resolve harmonics
  \(\rightarrow\) 2-3x longest (lowest) fundamental period
  - e.g. 30-60 ms = 480-960 samples @ 16 kHz
  - choose \(H \leq N/2\)

- \(N\) too long \(\rightarrow\) lost time resolution
  - limits sinusoid amplitude rate of change
Steps to sinewave modeling - 2

- Choose candidate sinusoids at each time by picking peaks in each STFT frame:

- Quadratic fit for peak:

  \[ y = ax(x-b) \]

  \[ + \text{linear interpolation of unwrapped phase} \]

Steps to sinewave modeling - 3

- Which peaks to pick?
  Want ‘true’ sinusoids, not noise fluctuations
  - ‘prominence’ threshold above smoothed spec.

- Sinusoids exhibit stability...
  - of amplitude in time
  - of phase derivative in time
  \[ \rightarrow \text{compare with adjacent time frames to test?} \]
Steps to sinewave modeling - 4

- 'Grow' tracks by appending newly-found peaks to existing tracks:
  - ambiguous assignments possible
  - Unclaimed new peak
    - 'birth' of new track
    - backtrack to find earliest trace?
  - No continuation peak for existing track
    - 'death' of track
    - or: reduce peak threshold for hysteresis

Resynthesis of sinewave models

- After analysis, each track defines contours in frequency, amplitude $f_k[n]$, $A_k[n]$ (+ phase?)
  - use to drive a sinewave oscillators & sum up

- 'Regularize' to exactly harmonic $f_k[n] = k \cdot f_0[n]$
Modification in sinewave resynthesis

- Change duration by warping timebase
  - may want to keep onset unwarped

- Change pitch by scaling frequencies
  - either stretching or resampling envelope

- Change timbre by interpolating params

Sinusoids + residual

- Only ‘prominent peaks’ became tracks
  - remainder of spectral energy was noisy?
    → model residual energy with noise

- How to obtain ‘non-harmonic’ spectrum?
  - zero-out spectrum near extracted peaks?
  - or: resynthesize (exactly) & subtract waveforms
    \[ e_s[n] = s[n] - \sum_k A_k[n] \cos(2\pi n \cdot f_k[n]) \]
    .. must preserve phase!

- Can model residual signal with LPC
  → flexible representation of noisy residual
Sinusoids + noise + transients

- Sound represented as sinusoids and noise:

\[ s[n] = \sum_k A_k[n] \cos(2\pi n \cdot f_k[n]) + h_n[n] \ast b[n] \]

**Sinusoids**

**Residual** \( e_s[n] \)

Parameters are \( \{A_k[n], f_k[n]\}, h_n[n] \)

- Separate out abrupt transients in residual?

\[ e_s[n] = \sum_k t_k[n] + h_n[n] \ast b'[n] \]

  - more specific \( \rightarrow \) more flexible

Outline

1. Music and nonspeech
2. Environmental sounds
3. Music synthesis techniques
4. Sinewave synthesis
5. Music analysis
   - Instrument identification
   - Pitch tracking
Music analysis

- What might we want to get out of music?

  - Instrument identification
    - different levels of specificity
    - ‘registers’ within instruments

  - Score recovery
    - transcribe the note sequence
    - extract the ‘performance’

  - Ensemble performance
    - ‘gestalts’: chords, tone colors

  - Broader timescales
    - phrasing & musical structure
    - artist / genre clustering and classification

Instrument identification

- Research looks for perceptual ‘timbre space’

- Cues to instrument identification
  - onset (rise time), sustain (brightness)

- Hierarchy of instrument families
  - strings / reeds / brass
  - optimize features at each level
Pitch tracking

- Fundamental frequency (→ pitch) is a key attribute of musical sounds
  → pitch tracking as a key technology

- Pitch tracking for speech
  - voice pitch & spectrum highly dynamic
  - speech is voiced and unvoiced

Applications
- voice coders (excitation description)
- harmonic modeling

Pitch tracking for music

- Pitch in music
  - pitch is more stable (although vibrato)
  - but: multiple pitches

Applications
- harmonic modeling
- music transcription (→ storage, resynthesis)
- source separation

Approaches: “place” & “time”
**Meddis & Hewitt pitch model**

- Autocorrelation (time) based pitch extraction
  - fundamental period $\rightarrow$ peak(s) in autocorrelation
  $$x(t) \approx x(t + T) \quad \rightarrow \quad r_{xx}(T) = \int x(t)x(t + T) \approx \text{max}$$

- Compute separately in each frequency band & ‘summarize’ across (perceptual) channels

**Tolonen & Karjalainen simplification**

- Multiple frequency channels can have different dominant pitches ...

- But equalizing (flattening) the spectrum works:

  - Summary AC as a function of time:

  - ‘Enhancement’ = cancel subharmonics
Post-processing of pitch tracks

- Remove outliers with median filtering

![Median Filtering Diagram]

- Octave errors are common:
  - if $x(t) \approx x(t + T)$ then $x(t) \approx x(t + 2T)$ etc.

  $\rightarrow$ dynamic programming/HMM

- Validity
  - “is there a pitch at this time?”
  - voiced/unvoiced decision for speech

- Event detection
  - when does a pitch slide indicate a new note?

Summary

- ‘Nonspeech audio’
  - i.e. sound in general
  - characteristics: ecological

- Music synthesis
  - control of pitch, duration, loudness, articulation
  - evolution of techniques
  - sinusoids + noise + transients

- Music analysis
  - different aspects: instruments, pitches, performance

  and beyond?