Sound Texture Classification Using Statistics from an Auditory Model

Gabriele Carotti-Sha, Evan Penn, Daniel Villamizar
Fall 2015 CS229 Machine Learning Class Project, Stanford University

Motivation
Sound textures may be defined as a category of sounds produced by a superposition of many similar acoustic events. Extensive work has been done to (1) analyze the features of sounds that are used by the auditory system as features of sound textures, and (2) develop synthesis techniques to generate realistic-sounding textures based on the relevant features. Machine recognition of sound textures becomes a natural extension of this work and can be useful for location-awareness, security, and robotics applications among others.

Methodology
- **Supervised Learning**
  - Audio recording (5-10 min) → Training samples (~7s) → Extract statistics → Cross-validation

- **Textures**
  - Applause
  - Birds
  - Cricket sounds
  - Crowded room
  - Busy road
  - Train station
  - Subway station
  - Thunderstorm
  - Rolling train
  - Walking sounds

- **Measured Statistics**
  - Sub-band correlations
  - Pre-modulation moments
  - Mean
  - Variance
  - Skew
  - Kurtosis
  - Modulated channel
  - Power

- **Future Work**
  - Implement full synthesis features:
    - C1-C2 Correlations
  - Parameter regularization for hardware optimization:
    - Sample rate
  - Full-scale amplitude (SNR)
  - Extended texture analysis:
    - Timbre
    - Time-varying statistics

Feature Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Error</th>
<th>Test Error</th>
<th>Logistic Regression</th>
<th>SVM, rbf kernel</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimental Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Training &amp; Test Error Rate</th>
<th>Confusion Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References: