Musical Instrument Recognition in Combined Electric and Acoustic Cochlear Implant Simulations

Music Induced Hearing Disorders

Audio Engineering Society 47th International Conference

Shaikat Hossain  Peter Assmann
School of Behavioral and Brain Sciences

The University of Texas at Dallas
Hearing Loss

37.1 million adults in the US have some form of hearing loss (CDC, 2010)

Hearing loss often results in:
- an inability to separate multiple talkers
- poor speech perception in noisy conditions
- diminished music perception abilities
Sensorineural Hearing Loss

• **Conduction**
  – Problems with outer or middle ear passing sound into inner ear

• **Sensorineural**
  – Problem with inner ear, auditory nerve, or central processing
Cochlear Implants

-A surgically implanted electronic device that provides a newfound sense of hearing to the profoundly deaf
-Designed for sensorineural hearing loss
- Consists of internal and external hardware
How a CI works

• Sound information is split into separate frequency bands/channels
• The bands are converted into electric pulses which are then sent to the corresponding electrodes along the tonotopy of the cochlea.
How a CI works (continued...
Problems with CI devices

• Reduced spectral resolution (Over 20,000 hair cells to 8-24 electrodes)
• Limited dynamic range (120dB to 10-20 dB)
• Channel interactions between electrodes
Limitations

• Poor speech perception in adverse conditions
  – Background noise
  – Competing talkers
• Music Perception
• Sound localization

Although envelope information is preserved, fine structure is degraded
Music Perception

• Second most important acoustic stimulus in the lives of CI users (Gfeller et al. 2000)
• Rhythm perception comparable to normal hearing listeners
• However, melody and timbre perception remain to be quite poor (Drennan and Rubinstein 2008)
Melody recognition

• Variability in pitch discrimination abilities, ranging from 1 semitone to 2 octaves (McDermott 2004)

• Heavy reliance on temporal cues in discriminating between melodies
Timbre recognition

• Studies investigating CI users’ timbre recognition found that they achieved scores of 44-47% correct compared to 91-97% correct as achieved by normal hearing listeners (Gfeller et al. 2002, McDermott 2004)

• May be due to the huge reduction in dynamic range and spectral resolution
  – Reduced spectral contrast
  – Limited depth of temporal modulations
Electro-acoustic Stimulation (EAS)

• The addition of low frequency residual hearing through use of a hearing aid can help CI user’s ability to:
  – Separate multiple talkers (Cullington & Zeng 2010)
  – Improve sentence recognition in noisy conditions (Dorman et al. 2008)
  – Discriminate between melodies (Kong et al. 2005)
Configurations

- **Hybrid:**
  - Using a short electrode cochlear implant + hearing aid on the same side
  - Often implemented on both sides for those with residual hearing in both ears

- **Bimodal:**
  - Hearing aid on the ear with residual hearing
  - Cochlear Implant on the other
EAS benefit

• Works by introducing low frequency acoustic signal that may contain:
  – Pitch information
  – Amplitude envelope information

• Such cues contribute to perceptual grouping/segregation processes
Purpose of study

• To explore whether the EAS benefit may extend to improving timbre recognition through a closed-set musical instrument recognition task

• Relevance of research
  – Music appreciation
  – Playing music
Hypotheses

• Hypothesis 1: EAS processing condition would have higher scores than Normal CI processing

• Hypothesis 2: Rhythmic instruments would have higher scores than non-rhythmic instruments (McDermott and Looi 2004)
Participants

• 18 Normal Hearing Listeners
• 5 males, 13 females
• Ages 19-44
• Undergraduates recruited for research credits
CI Simulation

- CI processing is based on the channel vocoder
- Used for research purposes to study the psychoacoustic mechanisms underlying CI mediated perception
- Eliminates undesirable confounds: duration of deafness, age of implantation, neurological functioning, etc.
Stimuli

• Loopology sample package bundled with Adobe Audition software
• 7 different musical instrument categories: bass, guitar, horns, percussion, piano, strings, synthesizer
Considerations in choosing Musical Stimuli

• Musical instrument timbre recognition

• Ecological validity
  – Loopology contains real world music clips
  – Faithful to a number of different genres

• Control
  – CAMP corpus (Kang et al. 2009)
    • Isochronous- controlled for rhythm
    • Same note sequence used for timbre judgments
Methods

• MATLAB used for all signal processing and experimental design
• Sounds were controlled for level and duration
• 8 channel simulations of normal CI and EAS processing
  – EAS = normal CI + low pass filtered (cutoff=300 Hz) acoustic signal
• Listeners use headphones in a sound-treated booth
Graphical User Interface (GUI)
Some Audio Examples

• Normal CI

- Guitar
- Piano

• EAS

- Guitar
- Piano
# Average responses (in %)

<table>
<thead>
<tr>
<th>Normal CI</th>
<th>Bass</th>
<th>Guitar</th>
<th>Horns</th>
<th>Percussion</th>
<th>Piano</th>
<th>Strings</th>
<th>Synthesizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bass</td>
<td>13.9</td>
<td>13.89</td>
<td>2.22</td>
<td>6.11</td>
<td>1.67</td>
<td>5</td>
<td>57.22</td>
</tr>
<tr>
<td>Guitar</td>
<td>5</td>
<td><strong>54.44</strong></td>
<td>2.78</td>
<td>9.44</td>
<td>4.44</td>
<td>13.33</td>
<td>10.56</td>
</tr>
<tr>
<td>Horns</td>
<td>4.44</td>
<td>10</td>
<td><strong>43.33</strong></td>
<td>3.33</td>
<td>8.33</td>
<td>21.67</td>
<td>8.89</td>
</tr>
<tr>
<td>Percussion</td>
<td>0</td>
<td>1.67</td>
<td>0</td>
<td><strong>90</strong></td>
<td>0</td>
<td>1.11</td>
<td>7.22</td>
</tr>
<tr>
<td>Piano</td>
<td>2.22</td>
<td>8.89</td>
<td>4.44</td>
<td>15.56</td>
<td><strong>58.33</strong></td>
<td>5</td>
<td>5.56</td>
</tr>
<tr>
<td>Strings</td>
<td>3.89</td>
<td>5</td>
<td>21.11</td>
<td>3.33</td>
<td>2.78</td>
<td><strong>55</strong></td>
<td>8.89</td>
</tr>
<tr>
<td>Synthesizer</td>
<td>6.11</td>
<td>16.11</td>
<td>10</td>
<td>8.33</td>
<td>5</td>
<td>25</td>
<td><strong>29.44</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EAS</th>
<th>Bass</th>
<th>Guitar</th>
<th>Horns</th>
<th>Percussion</th>
<th>Piano</th>
<th>Strings</th>
<th>Synthesizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bass</td>
<td><strong>37.22</strong></td>
<td>12.78</td>
<td>1.67</td>
<td>2.78</td>
<td>0.56</td>
<td>2.78</td>
<td>42.22</td>
</tr>
<tr>
<td>Guitar</td>
<td>6.67</td>
<td><strong>82.78</strong></td>
<td>1.11</td>
<td>0.56</td>
<td>0.56</td>
<td>2.22</td>
<td>6.11</td>
</tr>
<tr>
<td>Horns</td>
<td>5.56</td>
<td>7.22</td>
<td><strong>52.78</strong></td>
<td>1.11</td>
<td>7.22</td>
<td>16.67</td>
<td>9.44</td>
</tr>
<tr>
<td>Percussion</td>
<td>4.44</td>
<td>0.56</td>
<td>0</td>
<td><strong>81.11</strong></td>
<td>0</td>
<td>0.56</td>
<td>13.33</td>
</tr>
<tr>
<td>Piano</td>
<td>8.89</td>
<td>10.56</td>
<td>1.67</td>
<td>6.11</td>
<td><strong>67.78</strong></td>
<td>4.44</td>
<td>0.56</td>
</tr>
<tr>
<td>Strings</td>
<td>21.67</td>
<td>3.89</td>
<td>24.44</td>
<td>0</td>
<td>0.56</td>
<td><strong>43.89</strong></td>
<td>5.56</td>
</tr>
<tr>
<td>Synthesizer</td>
<td>25</td>
<td>18.89</td>
<td>6.67</td>
<td>2.78</td>
<td>1.67</td>
<td>13.33</td>
<td><strong>31.67</strong></td>
</tr>
</tbody>
</table>
Results

• Two-factor ANOVA on percent correct responses revealed:
  – A main effect of processing condition, $F(1,17)=13.80; p<0.0017$
  – A main effect of instrument type, $F(6,17)=28.46; p<0.0001$
  – An interaction between processing condition and instrument type, $F(6,102)=8.19; p<0.0001$
Two-way interaction

* Indicate significant differences
Discussion

• EAS condition lead to higher recognition scores for particular instruments
• High variability in responses
• Rhythmic instruments (percussion, piano) among the highest recognized (McDermott and Looi 2004)
Discussion continued

• Possible fusion of residual pitch/amplitude envelope information with temporal structure of sound provided through CI simulation

• Recognition vs appraisal based approaches

• Confusions in instrument categories:
  – Bass vs synthesizer
  – Categorical boundaries- largely culturally based
Auditory Modeling Initiative

• Currently in progress
• Multidimensional Scaling techniques
• Perceptual: Analysis of confusion errors
  – Using ExPosition software package written in R
• Proposed physical measurements:
  – Spectral energy distribution
  – Analysis of transients
  – Spectral fluctuation over time
Correspondence Analysis

Component 1 variance: 37.924%

Component 2 variance: 25.287%
Hierarchical Clustering

Cluster Dendrogram

```
dist(data.out$ExPosition.Data$fi)
hclust (*, "complete")
```
Future Directions

• Replication of present study with CI users
• Auditory Modeling initiative
  – Additional experiments using CAMP corpus
  – Effects of CI Processing on the Timbre Space
• Melody recognition
• More extensive experiments including:
  – More instrument categories
  – With/without rhythmic cues

Acknowledgements: Special thanks to Derek Beaton and Dr. Philip Loizou
References


