INTEGRATION OF AUDITORY MASKS INTO A LOCALLY COMPETITIVE ALGORITHM FOR SPARSE REPRESENTATIONS OF AUDIO SIGNALS

Christopher Srinivasa*, Ramin Pichevar^, Hossein Najaf-Zadeh^, and Martin Bouchard*

University of Ottawa* Communications Research Centre Canada^

OVERVIEW

Goal:
• Investigate the impact of auditory masking when obtaining sparse representations for audio signals

Strategy:
• Integrate a masking model into a Locally Competitive Algorithm (LCA) designed to obtain such representations and compare the LCA with Masking (LCAM) to the LCA

Outcome:
• The LCAM allows for larger residual error between an input signal and its reconstructed and it does so by shaping the residual error such that this larger error is inaudible

LOCALLY COMPETITIVE ALGORITHM

• LCA Terminology:
  - \( s \): Input signal; vector of length \( D \)
  - \( M \): Number of neurons
  - \( \Phi \): Receptive field matrix of size \( D \times M \); each column is a vector corresponding to the receptive field for one neuron
  - \( \mathbf{b} \): Input to all neurons; vector of length \( M \)
  - \( \mathbf{a} \): Internal states of all neurons; vector of length \( M \)
  - \( \mathbf{s} \): Output coefficients of all neurons; vector of length \( M \)
  - \( \mathbf{r} \): Reconstructed input to all neurons; vector of length \( M \)

• Internal states and output coefficients are related via a bio-inspired integrate-and-fire mechanism (where \( \lambda \) is user defined):
  \[
  a_m = \begin{cases} 
  u_m, & |u_m| \geq \lambda \\
  0, & \text{otherwise}
  \end{cases}
  \]

• The LCA solves the sparse signal decomposition problem by finding the set of coefficients which minimize
  \[
  E = \frac{1}{2} \| s - \mathbf{r} \|^2 + \lambda C(\mathbf{a})
  \]
  where \( C \) is the cost function related to the sparsity of the coefficients and \( \mathbf{r} \) is computed as \( \Phi \mathbf{a} \)

• It is an iterative process, where at each iteration \( k \), the update is
  \[
  u_m^{k+1} = u_m^k + \frac{1}{\tau} \delta_m^k
  \]
  where \( \tau \) is a step size and \( \delta_m^k = \beta_m^k - u_m^k \)

• \( \alpha \) is the representational responsibility of a neuron:
  \[
  \alpha_m^k = \left( \beta_m^k - u_m^k \right) + a_m^k = r_m^k + a_m^k
  \]

• In audio, the LCA is used to obtain sparse time-frequency representations, dubbed spikegrams, by using Gammatone kernels waveforms as receptive fields for the neurons

AUDITORY MASKING MODEL

• Considers both temporal and frequency domain masking effects
• Specifically designed for Gammatone kernels

Fig. 2. Overall masking effects from single masker.

LCAM: INTEGRATION OF THE MASKS IN THE LOCALLY COMPETITIVE ALGORITHM

• Change LCA update process:
  \[
  \delta_m^k = \beta_m^k - u_m^k
  \]

• \( \beta \) is the unmasked representational responsibility of a neuron:
  \[
  \beta_m^k = \begin{cases} 
  \alpha_m^k, & |\alpha_m^k| > |\delta_m^k| \\
  0, & \text{otherwise}
  \end{cases}
  \]

• \( v \) is the masking effect felt by a neuron from the respective masking model instances induced by all other neurons

RESULTS

• For the same broadcast-level perceptual quality, the LCAM extracts a lot less energy from the original signal

Table 1. Signal-to-Noise Ratios of LCA and LCAM

<table>
<thead>
<tr>
<th>Audio</th>
<th>LCA SNR</th>
<th>LCAM SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castanet</td>
<td>27.4767</td>
<td>20.5149</td>
</tr>
<tr>
<td>Percussion</td>
<td>28.2317</td>
<td>24.7087</td>
</tr>
<tr>
<td>Speech</td>
<td>38.7884</td>
<td>38.7707</td>
</tr>
</tbody>
</table>

• The LCAM is capable of tolerating larger error by shaping the error to follow the original signal such that the masking effects stemming from it cover up the error, making it inaudible

Fig. 3. Original signal time-frequency surface.

Fig. 4. LCA error time-frequency surface.

Fig. 5. LCAM error time-frequency surface.