Q: How to search for a sound that matches the concept in your head?
A: Current ways: through its name or other semantic labels.
Q: What if you don’t remember its name, or what you are looking for simply doesn’t have a semantic meaning?
A: Imitate the concept with your voice!

**Dog barking sound:** infantile bark threat bark

**Synthesized sound:**

**Acoustic instr.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline IMISOUND</td>
<td>0.450 ± 0.308</td>
<td>0.126 ± 0.340</td>
<td>0.450 ± 0.308</td>
<td></td>
</tr>
<tr>
<td>Proposed Untied</td>
<td>0.377 ± 0.200</td>
<td>0.014 ± 0.020</td>
<td>0.377 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Partial</td>
<td>0.384 ± 0.200</td>
<td>0.145 ± 0.020</td>
<td>0.384 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Tied</td>
<td>0.401 ± 0.200</td>
<td>0.340 ± 0.020</td>
<td>0.401 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Untied + Partial</td>
<td>0.438 ± 0.200</td>
<td>0.012 ± 0.020</td>
<td>0.438 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Untied + IMISOUND</td>
<td>0.470 ± 0.200</td>
<td>0.168 ± 0.020</td>
<td>0.470 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Partial + IMISOUND</td>
<td>0.476 ± 0.200</td>
<td>0.035 ± 0.020</td>
<td>0.476 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Tied + IMISOUND</td>
<td>0.504 ± 0.200</td>
<td>0.171 ± 0.020</td>
<td>0.504 ± 0.200</td>
<td></td>
</tr>
<tr>
<td>Proposed Untied + Partial + Tied</td>
<td>0.520 ± 0.200</td>
<td>0.188 ± 0.020</td>
<td>0.520 ± 0.200</td>
<td></td>
</tr>
</tbody>
</table>

**Evaluation Measure:**

\[
MRR = \frac{1}{q} \sum_{i=1}^{q} \frac{1}{\text{rank}_i}
\]

**Before:**

**Step 1: Pre-processing**

Constant-Q Transform (CQT) with 2 seconds raw input

**Step 2: Feature Extraction**

Use Convolutional Neural Networks (CNN) to learn features in two towers with tied, partially tied (layers in shade are weight-shared), and untied weights

**Step 3: Metric Learning**

Use Fully Connected Networks (FCN) to learn the pair-wise similarity and generate a single value output in [0, 1]

**Step 4: Sound Retrieval**

Pair the imitation query with each recording in the library to calculate its likelihood of being a positive pair. Likelihood scores are ranked in descending order

---

**References:**
