Locality-Sensitive Hashing

- phoneme proximity
  
  - feature sampling
    * create a single (concatenated) feature vector for each phoneme
    * to avoid performing feature vector upsampling, reduce the feature extraction frame step from 10ms to 2ms
    * for each phoneme, keep only 5 equally spaced feature frames
    * for comparison, consider situation where the 5 feature vectors are not concatenated
  
  - LSH parameters
    * feature vector concatenation: \( R \) range: 40 → 70, step: 2
    * no feature vector concatenation: \( R \) range: 10 → 20, step: 1
    * \( P_1 \) range: 0.80 → 0.98, step: 0.05
    * TIMIT corpus speakers: \( abc0 \), \( ask0 \), \( cth0 \)
    * Note: different values for \( R \) are necessary because the space dimension in the two situations is different, which leads to different distance values. The choice of \( R \) was made so that a similar amount of nearest-neighbours if found in both situations
  
  - analysis of the results
    * only results for speaker \( abc0 \) are presented, as they are representative for the 3 speakers
    * figures 1-2 present the results using feature concatenation
    * figures 3-4 present the results without feature concatenation
    * table 1 presents a summary of the obtained (best) results
    * current approach produces worse results, both in phoneme classification and positive examples detection
Figure 1: Using concatenation, phoneme classification, speaker abc0.

Figure 2: Using concatenation, positive examples detection, speaker abc0.
(a) With h# phoneme.
(b) Without h# phoneme.

Figure 3: Without concatenation, phoneme classification, speaker *abc0*.

(a) With h# phoneme.
(b) Without h# phoneme.

Figure 4: Without concatenation, positive examples detection, speaker *abc0*.
<table>
<thead>
<tr>
<th></th>
<th>phoneme classification</th>
<th>positive examples detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with h#</td>
<td>without h#</td>
</tr>
<tr>
<td>initial approach</td>
<td>0.643  0.663</td>
<td>0.532  0.549</td>
</tr>
<tr>
<td>using feature concatenation</td>
<td>0.157  0.167</td>
<td>0.196  0.204</td>
</tr>
<tr>
<td>without feature concatenation</td>
<td>0.329  0.340</td>
<td>0.367  0.374</td>
</tr>
</tbody>
</table>

Table 1: Results comparison summary