EXPLOITING EXPLICIT MEMORY INCLUSION FOR ARTIFICIAL BANDWIDTH EXTENSION

Pramod Bachhav, Massimiliano Todisco and Nicholas Evans
EURECOM, France
Emails: {bachhav, todisco, evans}@eurecom.fr

Introduction
- Traditional telephony infrastructure is typically limited to a bandwidth of 0.3-3.4 kHz, referred as narrowband (NB).
- Voiced phonemes exhibit significant information beyond NB.
- Wider bandwidths generally correspond to higher quality speech.
- Artificial bandwidth extension (AVE) methods estimate missing highband (HB) components at 3.4-4kHz.
- Use of dynamic information or memory to improve AVE performance is common and can be captured using back-end regression models or via front-end features.
- Memory inclusion via delta features for AVE has been investigated thoroughly [1,2] via information theoretic analysis.
- A quantitative analysis of the benefit of explicit memory from neighbouring frames, without significant increases to complexity and latency is missing.

Contributions
- Assessment of explicit memory through information theoretic analysis.
- Explicit memory inclusion for AVE without affecting complexity of a standard regression model.
- Application of principal component analysis as a dimensionality reduction transform.

Mutual Information
- Correlation between NB and HB features is usually measured using mutual information (MI).
- The mutual information between two continuous random variables X and Y with joint probability density function (PDF) \( f_{XY}(x,y) \) is defined according to:
  \[
  I(X;Y) = \int \int f_{XY}(x,y) \log \frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} \, dx \, dy
  \]
- The integral can be written as an expectation approximated by the sample mean over X samples as follows:
  \[
  I(X;Y) \approx \frac{1}{K} \sum_{k=1}^{K} \log \frac{f_{XY}(x_k,y_k)}{f_X(x_k)f_Y(y_k)}
  \]
- The joint PDF \( f_{XY}(x,y) \) is usually modelled using a Gaussian mixture model (GMM).
- A GMM of 128 components was used to estimate MI between NB and HB representations, using TIMIT database.

Mutual information assessment results

Benefit of memory to ABE
- An illustration of MI estimation with contextual information from neighbouring frames. Vertical bars represent NB (bottom) and HB (top) feature vectors. Red boxes represent the pair of NB \((X_{t-1}, \delta = 1.0,1)\) and HB \((Y = Y_t)\) components used for MI calculations.

Experimental setup and results
- Database: TIMIT database divided into training (3696 utterances) and test (1344 utterances) sets.
- NB features: 10 log Mel filter (logMFE) coefficients; HB features: 10 linear prediction (LP) coefficients including LP gain.
- Proposed ABE system with memory \(M_2\) uses 4 neighboring frames (NB features - \(\delta = 1.2\)).
- Baseline B1 uses static NB features \(E^B\).
- Baseline B2 uses NB and HB features formed by appending 5 static features with corresponding 5 second order delta coefficients. (A variant of the approach presented in [2])

Conclusions and future work
- Explicit memory inclusion for AVE is presented without significant impact on computational complexity.
- Use of PCA is as a dimensionality reduction transform.
- Potential of the memory is demonstrated through information theoretic analysis.
- Memory produces bandwidth-extended speech signals with better speech quality.

Future Work
- Investigation of dimensionality reduction techniques designed to preserve speech quality rather than feature variance.

Subjective assessment results in terms of CMOS. Files used for the subjective evaluation are available at http://audio.eurecom.fr/content/media

Comparison B → A

<table>
<thead>
<tr>
<th></th>
<th>CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_2\rightarrow NB)</td>
<td>0.69</td>
</tr>
<tr>
<td>(M_2\rightarrow B1)</td>
<td>0.51</td>
</tr>
<tr>
<td>(M_2\rightarrow WB)</td>
<td>-0.78</td>
</tr>
</tbody>
</table>

Comparison

\[
\begin{align*}
I(X;Y_t) & = 1.24 \\
I(X;Y_t) & = 1.34 \\
\end{align*}
\]

Mutual information assessment results

Selected References

ABE method | dMFE-LQO | dCOSH | MOS-LQO |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>9.2 (1.2)</td>
<td>2.4 (0.7)</td>
<td>2.4</td>
</tr>
<tr>
<td>B2</td>
<td>10.1 (1.7)</td>
<td>3.6 (1.2)</td>
<td>2.2</td>
</tr>
<tr>
<td>M2</td>
<td>8.2 (0.9)</td>
<td>2.2 (0.6)</td>
<td>2.8</td>
</tr>
<tr>
<td>M2</td>
<td>8.1 (0.9)</td>
<td>2.1 (0.6)</td>
<td>2.9</td>
</tr>
<tr>
<td>M2</td>
<td>8.2 (0.9)</td>
<td>2.2 (0.7)</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Objective assessment results. Lower values of \(d_{MFE-LQO}\) and \(d_{COSH}\) indicate better performance whereas as higher MOS-LQO indicates better quality.

A block diagram of the ABE system with memory inclusion.

A comparison of true NB LP gain \(g_{NB}^M\) to estimated WB LP gain \(g_{WB}^M\) for systems \(M_2\) and B1.