Why converted speech sounds muffled?

Modulation Spectrum (MS) degradation
[Takamichi et al., 2014.]

MS recovery will improve quality!

Modulation Spectrum-Based Post-Filter
for GMM-based Voice Conversion

Shinnosuke Takamichi,
Tomoki Toda,
Alan W. Black,
Satoshi Nakamura

Carnegie Mellon University (USA)
Nara Institute of Science and Technology (Japan)

Evaluation Setup

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Female (F) and Male (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM</td>
<td>F-to-M/M-to-F, 64-mixture</td>
</tr>
<tr>
<td>Training data</td>
<td>50 Japanese sentences</td>
</tr>
<tr>
<td>Test data</td>
<td>50 Japanese sentences</td>
</tr>
<tr>
<td>Spectral parameter</td>
<td>25 mel-cepstral coefficients</td>
</tr>
<tr>
<td>Excitation parameters</td>
<td>Log F0, 5 band-aperiodicity</td>
</tr>
</tbody>
</table>

Preference AB tests comparing:
Conv: non-filtered parameters (consider GV.)
MSPF: filtered parameters

Results

Speech quality
Speaker individuality

Preference score

: Conv  : MSPF  : 95% conf. interval