Reducing Interference with Phase Recovery in DNN-based Monaural Singing Voice Separation
Paul Magron\textsuperscript{1}, Konstantinos Drossos\textsuperscript{1}, Stylianos Ioannis Mimilakis\textsuperscript{2}, Tuomas Virtanen\textsuperscript{1}

\textsuperscript{1}Laboratory of Signal Processing, Tampere University of Technology, Finland
\textsuperscript{2}Fraunhofer IDMT, Ilmenau, Germany

Introduction

Problem setting
- Monaural singing voice separation.
- State-of-the-art approach: estimate STFT magnitudes with DNNs, and retrieve the sources using the mixture’s phase.
- This re-introduces interferences in the separated signals.

Contributions
- Phase recovery algorithms instead of the mixture’s phase.
- Reduction of interferences between the estimated sources.

Magnitude estimation: MaD TwinNet
- Most up-to-date DNN system for singing voice separation \cite{1}.
- Used here as a baseline for magnitude estimation.
- The Masker predicts and applies a time-frequency mask to the mixture’s magnitude.
- The Denoiser reduces remaining interferences.
- The Twin Network is a regularization that accounts for long-term temporal patterns (melody, rhythm).

Phase recovery algorithms

Consistent anisotropic Wiener filter \cite{2}
- A phase-aware extension of the Wiener filter.
- Based on an anisotropic Gaussian model (i.e., with non-uniform phase) → accounts for the sinusoidal phase model.
- Includes a consistency regularization.

Iterative procedure \cite{3}
- Minimizes the mixing error between the sources.
- Uses the sinusoidal phase as an initial phase estimate.
- Does not modify the magnitude estimates.

Experiments

- DSD100: 100 songs with singing voice + musical background.
- A learning set (50 songs) is used for tuning the consistent anisotropic Wiener filter parameters.
- Separation on the test set measured with the signal to distortion, interference and artifact ratios (SDR, SIR and SAR):

<table>
<thead>
<tr>
<th></th>
<th>SDR</th>
<th>SIR</th>
<th>SAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture’s phase</td>
<td>4.57</td>
<td>8.17</td>
<td>5.97</td>
</tr>
<tr>
<td>Iterative procedure</td>
<td>4.52</td>
<td>8.87</td>
<td>5.52</td>
</tr>
<tr>
<td>Consistent anisotropic Wiener</td>
<td>4.46</td>
<td>10.32</td>
<td>4.97</td>
</tr>
</tbody>
</table>

Conclusion

Advanced phase recovery algorithms reduce interference at the cost of limited additional distortion and artifacts.

Future work:
- Phase recovery with DNNs.
- Joint magnitude and phase estimation.

References


http://arg.cs.tut.fi/demo/phase-madtwinnet/
https://github.com/magronp/phase-madtwinnet