Improving Design of Input Condition Invariant Speech Enhancement

Wanyou Zhang\textsuperscript{1,2,*}, Jee-woon Jung\textsuperscript{2,\dagger}, Yanmin Qian\textsuperscript{1}

\textsuperscript{1}Shanghai Jiao Tong University, China \textsuperscript{2}Carnegie Mellon University, USA

C: \#channels  \ D: embedding dim  \ N: bottleneck dim  \ H: hidden dim  \ F: frequency dim  \ T: time dim  \ K: \#blocks

5.
4.
3.
2.
1.

Experiments

1. Investigations on CHIME-4 simulated and real data
   - All models are only trained on 8 kHz data, and tested on 16 kHz data. (See the table for detailed results)

2. Experiments on combined SE datasets covering diverse input conditions
   - Combination of VCTK+DEMAND, DNS-2020, CHIME-4, REVERB, and WHAMR!
   - Significantly improved performance in realistic conditions
   - Check our paper for more information.

Highlights

1. Extension of our prior work \cite{1}: input condition invariant speech enhancement (SE)
   - A single SE model unifying various input conditions including different sampling frequencies (SF), microphone variations, and input lengths
   - Decoupled 1ch and multi-channel modeling
2. Two-stage training strategy: 1ch → c-ch
3. Exploration of T-F window-wise modeling \cite{2,3} in frequency-domain SE
4. A novel channel modeling design combining TAC \cite{4} and channel-wise attention: TA\textsuperscript{ch}
5. Open-source implementation in ESPNet

Experiments

1. Investigations on CHIME-4 simulated and real data
2. Experiments on combined SE datasets covering diverse input conditions
3. Combination of VCTK+DEMAND, DNS-2020, CHIME-4, REVERB, and WHAMR!
4. Significantly improved performance in realistic conditions
5. Check our paper for more information.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
No. & Model & \#Param & \#MACs (G/s) & Test (Simu) & Test (Real) \\
\hline
1 & No processing & - & - & 7.5 & 1.46 & 6.7 \\
2 & \textit{USES} (baseline) \cite{1} & 3.05 M & 65.3 & 98.0 & 20.6 & 4.2 \\
3 & w/ decoupled proc. & 3.05 M & 60.8 & 98.0 & 18.2 & 5.2 \\
4 & + 2-stage training & 3.05 M & 60.8 & 98.0 & 15.6 & 6.8 \\
5 & No.3 + \textit{Att}+1 + TAC+2 & 3.02 M & 60.8 & 97.5 & 22.2 & 4.0 \\
6 & + 2-stage training & 3.02 M & 60.8 & 97.5 & 19.0 & 4.4 \\
7 & No.3 + TA\textsuperscript{ch}×3 & 3.47 M & 60.8 & 116.9 & 18.7 & 5.2 \\
8 & + 2-stage training & 3.47 M & 60.8 & 116.9 & 19.8 & 4.2 \\
9 & \textit{USES2-Swin} & 2.92 M & 37.7 & 75.5 & 20.6 & 4.2 \\
10 & + 2-stage training & 2.92 M & 37.7 & 75.5 & 21.1 & 4.1 \\
11 & \textit{USES2-Comp} & 2.53 M & 52.4 & 83.0 & 20.4 & 4.2 \\
12 & + 2-stage training & 2.53 M & 52.4 & 83.0 & 18.8 & 4.6 \\
\hline
\end{tabular}
\end{table}

References