Cited papers in related work

[20] D. Wang and J. Chen, ‘‘Supervised speech separation based on

deep learning: An overview,’’ IEEE/ACM Trans. Audio, Speech, Lan-

guage Process., vol. 26, no. 10, pp. 1702–1726, Oct. 2018, doi:

10.1109/TASLP.2018.2842159.

[21] L. Pfeifenberger and F. Pernkopf, ‘‘Nonlinear residual echo suppression

using a recurrent neural network,’’ in Proc. Interspeech, Oct. 2020,

pp. 3950–3954, doi: 10.21437/interspeech.2020-1473.

[22] H. Zhang and D. Wang, ‘‘Deep learning for acoustic echo cancellation

in noisy and double-talk scenarios,’’ in Proc. Interspeech, Sep. 2018,

pp. 1–26, doi: 10.21437/interspeech.2018-1484.

[23] A. Fazel, M. El-Khamy, and J. Lee, ‘‘Deep multitask acoustic

echo cancellation,’’ in Proc. Interspeech, 2019, pp. 4250–4254, doi:

10.21437/Interspeech.2019-2908.

[24] H. Zhang, K. Tan, and D. Wang, ‘‘Deep learning for joint acoustic echo

and noise cancellation with nonlinear distortions,’’ in Proc. Interspeech,

Sep. 2019, pp. 4255–4259, doi: 10.21437/interspeech.2019-2651.

[25] Y. Zhang, C. Deng, S. Ma, Y. Sha, and H. Song, ‘‘Deep multi-task network

for delay estimation and echo cancellation,’’ 2020, arXiv:2011.02109.

[26] N. L. Westhausen and B. T. Meyer, ‘‘Dual-signal transformation LSTM

network for real-time noise suppression,’’ in Proc. Interspeech, Oct. 2020,

pp. 2477–2481, doi: 10.21437/interspeech.2020-2631.

[27] N. Sun, H. Liu, L. Gan, Y. Zhao, Z. Luo, and Y. Zhou, ‘‘Time-frequency

complex mask network for echo cancellation and noise suppression,’’ in

Proc. IEEE 24th Int. Conf. High Perform. Comput. Commun., 8th Int. Conf.

Data Sci. Systems, 20th Int. Conf. Smart City, 8th Int. Conf. Dependability

Sensor, Cloud Big Data Syst. Appl. (HPCC/DSS/SmartCity/DependSys),

Dec. 2022, pp. 2275–2279.

[28] J. C. Brown, ‘‘Calculation of a constant Q spectral transform,’’ J. Acoust.

Soc. Amer., vol. 89, no. 1, pp. 425–434, Jan. 1991.

[29] A. T. Patil, K. Khoria, and H. A. Patil, ‘‘Voice liveness detection using

constant-Q transform-based features,’’ in Proc. 30th Eur. Signal Process.

Conf. (EUSIPCO), Aug. 2022, pp. 110–114.

[30] Z. Shi, H. Lin, L. Liu, R. Liu, and J. Han, ‘‘Is CQT more suitable

for monaural speech separation than STFT? An empirical study,’’ 2019,

arXiv:1902.00631.

[31] R. Hemavathi and R. Kumaraswamy, ‘‘A study on unsupervised monaural

reverberant speech separation,’’ Int. J. Speech Technol., vol. 23, no. 2,

pp. 451–457, Jun. 2020, doi: 10.1007/s10772-020-09706-x.

[32] N. Holighaus, M. Dörfler, G. A. Velasco, and T. Grill, ‘‘A framework

for invertible, real-time constant-Q transforms,’’ IEEE Trans. Audio,

Speech, Language Process., vol. 21, no. 4, pp. 775–785, Apr. 2013, doi:

10.1109/TASL.2012.2234114.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,

L. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ 2017,

arXiv:1706.03762.

[34] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, and J. Zhong,

‘‘Attention is all you need in speech separation,’’ in Proc. IEEE Int. Conf.

Acoust., Speech, Signal Process. (ICASSP), Jun. 2021, pp. 21–25, doi:

10.1109/ICASSP39728.2021.9413901.

[36] L. Xu, Z. Wei, S. F. A. Zaidi, B. Ren, and J. Yang, ‘‘Speech enhancement

based on nonnegative matrix factorization in constant-Q frequency

domain,’’ Appl. Acoust., vol. 174, Mar. 2021, Art. no. 107732, doi:

10.1016/j.apacoust.2020.107732.

[39] X. Sun, C. Cao, Q. Li, L. Wang, and F. Xiang, ‘‘Explore relative and

context information with transformer for joint acoustic echo cancellation

and speech enhancement,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal

Process. (ICASSP), May 2022, pp. 9117–9121.

[40] Y. Zhang, X. Xu, and W. Tu, ‘‘Improving acoustic echo cancellation by

exploring speech and echo affinity with multi-head attention,’’ in Proc.

IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2024,

pp. 401–405.