SUPPLEMENTARY MATERIAL

Yuanle Li[1†], *Zhenghan Chen*[2†], *Hongqing Liu*[1], *Yi Zhou*[1*], *Xiaoxuan Liang*[3]

[1]Chongqing University of Posts and Telecommunications, Chongqing, China
[2]Peking University, Beijng, China [3]University of Massachusetts Amherst, Amherst MA, USA

1. EXPERIMENTAL DETAILS AND ADDITIONAL RESULTS

The experimental evaluations are performed using two different background noise types which are often encountered by the listener: Machinery (e.g., factory) and DrivingCar (e.g., street) noise. Both types of recorded background noise represent a wide range of temporal and spectral characteristics, and show nonstationary behavior [1]. The Machinery noise contains some quasiperiodic and periodic components. The DrivingCar noise is mixed with the wind noise. It also includes the Doppler effect as a result of the approaching or receding vehicles. Both noises are recorded by the Statistical Signal Processing Research Laboratory[1] at the University of Texas of Dallas, which can be obtained from either `https://labs.utdallas.edu/ssprl/hearing-aid-project/database/` or `https://utdallas.app.box.com/v/SSPRL-SE`.

Fig. 1 demonstrates the PESQ values achieved by our neural ICA methods with respect to other ICA based approaches, under Machinery noise of SNR 2dB, 5dB, and 10dB. Specifically, Reddy *et al.* combines fastICA with LogMMSE, whereas Hao *et al.* combines fastICA with Wiener filtering.

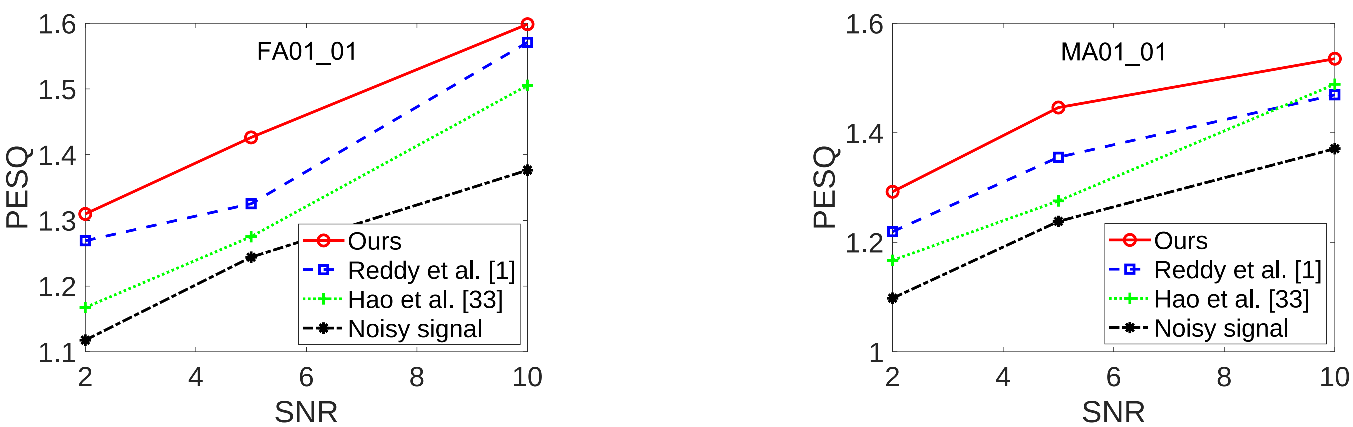

Fig. 1: PESQ under Machinery noise of 2dB, 5dB, 10dB. The black dashed line corresponds to signal without enhancement.

We also report the running time corresponding to Data B ($p = 10$) in Section 3.3. All experiments were conducted on a 2.6-GHz core i7 PC with 32 GB of RAM under the Matlab environment, and averaged over 10 runs. The results corroborate our analysis that the complexity of $\sum_{i=1}^{p}\sum_{j=i+1}^{p} \mathrm{HSIC}\left(s_i, s_j\right)$ is roughly $\mathcal{O}\left(N^2p^2\right)$; whereas the approximated CS-TC and the Rényi's MI (with kernel density estimator) have complexity of about $\mathcal{O}\left(N^2p\right)$. Note that, if we use MINE to estimate the Shannon's MI, the running time is between 3min to 10min.

2. PROOFS

Proposition 1 (Empirical Estimator of CS-TC)**.** *Given N observations $\left\{\left(s_1^i, s_2^i, \ldots, s_p^i\right)\right\}_{i=1}^{N}$, each observation contains p different types of measurements $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2, \ldots, s_p \in \mathcal{S}_p$. Let $Q_k \in \mathbb{R}^{N\times N}$ denote the Gram matrix for the k-th*

*Contact author: zhouy@cqupt.edu.cn; †Co-first authors.
[1]group website:`https://labs.utdallas.edu/ssprl/`

Methods	HSIC	Renyi's MI (KDE)	Ours
Time	0.056±0.009	0.006± 0.002	0.011± 0.002

Table 1: Runing time (s) of different measures to quantify total dependence between $p = 10$ variables.

$(1 \leq k \leq p)$ measurement, i.e., $Q_k(i,j) = G_\sigma\left(s_k^i - s_k^j\right)$, in which G_σ refers to a Gaussian kernel with width σ and takes the form of $G_\sigma\left(s_k^i - s_k^j\right) = \exp\left(-\frac{\|s_k^i - s_k^j\|^2}{2\sigma^2}\right)$. The empirical estimator of CS-TC is given by:

$$
\begin{aligned}
\hat{I}_{CS}(s_1, s_2, \ldots, s_p) &= \log\left(\frac{1}{N^2}\sum_{(i,j)\in\mathbf{i}_2^N}\prod_{k=1}^{p} Q_k(i,j)\right) \\
&+ \log\left(\frac{1}{N^{2p}}\sum_{(i_1,j_1,i_2,j_2,\cdots,i_p,j_p)\in\mathbf{i}_{2p}^N}\prod_{k=1}^{p} Q_k(i_k,j_k)\right) \\
&- 2\log\left(\frac{1}{N^{p+1}}\sum_{(i,j_1,j_2,\cdots,j_p)\in\mathbf{i}_{p+1}^N}\prod_{k=1}^{p} Q_k(i,j_k)\right),
\end{aligned} \tag{1}
$$

*where the index set $\mathbf{i}_r^N$ denotes the set of all r-tuples drawn **with** replacement from $\{1, 2, \cdots, N\}$.*

Proof. By definition, we have:

$$
\begin{aligned}
& I_{CS}(s_1, s_2, \ldots, s_p) := D_{CS}\left(p(s_1, s_2, \ldots, s_p); p(s_1)p(s_2)\ldots p(s_p)\right) \\
&= -\log\left(\frac{\left(\int p(s_1, s_2, \ldots, s_p)p(s_1)p(s_2)\ldots p(s_p)\, ds_1 ds_2 \ldots ds_p\right)^2}{\int p(s_1, s_2, \ldots, s_p)^2\, ds_1 ds_2 \ldots ds_p \int \left(p(s_1)p(s_2)\ldots p(s_p)\right)^2\, ds_1 ds_2 \ldots ds_p}\right) \\
&= \log\left(\int p(s_1, s_2, \ldots, s_p)^2\, ds_1 ds_2 \ldots ds_p\right) \\
&+ \log\left(\int \left(p(s_1)p(s_2)\ldots p(s_p)\right)^2\, ds_1 ds_2 \ldots ds_p\right) \\
&- 2\log\left(\int p(s_1, s_2, \ldots, s_p)p(s_1)p(s_2)\ldots p(s_p)\, ds_1 ds_2 \ldots ds_p\right).
\end{aligned} \tag{2}
$$

Let us discuss the three terms inside the "log":

$$
\begin{aligned}
& \int p(s_1, s_2, \ldots, s_p)^2\, ds_1 ds_2 \ldots ds_p \\
&= \mathbb{E}_{p(s_1, s_2, \ldots, s_p)}\left[p(s_1, s_2, \ldots, s_p)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} p(s_1^i, s_2^i, \ldots, s_p^i) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}\kappa([s_1^i, s_2^i, \ldots, s_p^i]^T - [s_1^j, s_2^j, \ldots, s_p^j]^T)\right) \\
&= \frac{1}{N^2}\sum_{(i,j)\in\mathbf{i}_2^N}\kappa([s_1^i, s_2^i, \ldots, s_p^i]^T - [s_1^j, s_2^j, \ldots, s_p^j]^T) \\
&= \frac{1}{N^2}\sum_{(i,j)\in\mathbf{i}_2^N}\kappa(s_1^i - s_1^j)\kappa(s_2^i - s_2^j)\cdots\kappa(s_p^i - s_p^j) \\
&= \frac{1}{N^2}\sum_{(i,j)\in\mathbf{i}_2^N}\prod_{k=1}^{p} Q_k(i,j),
\end{aligned} \tag{3}
$$

in which the third equation is by the formula of kernel density estimation (KDE) [2], in which κ refers to a Gaussian kernel with width σ and takes the form of $\kappa(x-y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$. The fourth equation is based on the assumption of a diagonal covariance matrix for $[s_1, s_2, \ldots, s_p]^T$, which is common in KDE. In this case, the multivariate kernel reduces to product kernels.

Similarly,

$$\begin{aligned}
&\int p(s_1, s_2, \ldots, s_p) p(s_1) p(s_2) \ldots p(s_p)\, ds_1 ds_2 \ldots ds_p \\
&= \mathbb{E}_{p(s_1, s_2, \ldots, s_p)} \left[p(s_1) p(s_2) \ldots p(s_p) \right] \\
&= \frac{1}{N} \sum_{i=1}^{N} p(s_1^i) p(s_2^i) \ldots p(s_p^i) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left[\left(\frac{1}{N} \sum_{j_1=1}^{N} \kappa(s_1^i - s_1^{j_1}) \right) \left(\frac{1}{N} \sum_{j_2=1}^{N} \kappa(s_2^i - s_2^{j_2}) \right) \cdots \left(\frac{1}{N} \sum_{j_p=1}^{N} \kappa(s_p^i - s_p^{j_p}) \right) \right] \\
&= \frac{1}{N^{p+1}} \sum_{(i, j_1, j_2, \cdots, j_p) \in \mathbf{i}_{p+1}^N} \prod_{k=1}^{p} Q_k(i, j_k),
\end{aligned} \tag{4}$$

and

$$\begin{aligned}
&\int \left(p(s_1) p(s_2) \ldots p(s_p) \right)^2 \, ds_1 ds_2 \ldots ds_p \\
&= \int p^2(s_1) p^2(s_2) \ldots p^2(s_p)\, ds_1 ds_2 \ldots ds_p \\
&= \left[\frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} \kappa(s_1^{i_1} - s_1^{j_1}) \right] \left[\frac{1}{N^2} \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} \kappa(s_2^{i_2} - s_2^{j_2}) \right] \cdots \left[\frac{1}{N^2} \sum_{i_p=1}^{N} \sum_{j_p=1}^{N} \kappa(s_p^{i_p} - s_p^{j_p}) \right] \\
&= \frac{1}{N^{2p}} \sum_{i_1=1}^{N} \sum_{j_1=1}^{N} \sum_{i_2=1}^{N} \sum_{j_2=1}^{N} \cdots \sum_{i_p=1}^{N} \sum_{j_p=1}^{N} \kappa(s_1^{i_1} - s_1^{j_1}) \kappa(s_2^{i_2} - s_2^{j_2}) \cdots \kappa(s_p^{i_p} - s_p^{j_p}) \\
&= \frac{1}{N^{2p}} \sum_{(i_1, j_1, i_2, j_2, \cdots, i_p, j_p) \in \mathbf{i}_{2p}^N} \prod_{k=1}^{p} Q_k(i_k, j_k).
\end{aligned} \tag{5}$$

□

Lemma 1. *Both $I_{CS}(s_1, s_2, \cdots, s_p)$ and $\sum_{k=1}^{p} I_{CS}(s_k; s_{[p]\setminus k})$ reduce to zero if and only if all components $\{s_1, s_2, \cdots, s_p\}$ are independent to each other.*

Proof. Lemma 1 is obvious. □

Lemma 2. *Total correlation is closely related to the sum of mutual information between individual component s_i and all rest components $s_{[p]\setminus i}$, in particular:*

$$\frac{p}{p-1} I(s_1, s_2, \cdots, s_p) \leq \sum_{k=1}^{p} I(s_k; s_{[p]\setminus k}) \leq p I(s_1, s_2, \cdots, s_p). \tag{6}$$

Proof. According to Lemma 4.3 in [3], we have:

$$I(s_1, s_2, \cdots, s_p) + \mathrm{DTC}(s_1, s_2, \cdots, s_p) = \sum_{k=1}^{p} I(s_k; s_{[p]\setminus k}), \tag{7}$$

in which $\mathrm{DTC}(s_1, s_2, \cdots, s_p)$ is also called the dual total correlation (DTC) [4], which is an alternative way to measure the total amount of dependence among p random variables [5] and is expressed as:

$$\mathrm{DTC}(s_1, s_2, \cdots, s_p) := H(s_1, s_2, \cdots, s_p) - \sum_{k=1}^{p} H(s_k | s_{[p]\setminus k}), \tag{8}$$

and $H(s_k|s_{[p]\setminus k})$ is the conditional entropy of s_k given all remaining variables $s_{[p]\setminus k}$.

Further, according to Lemma 4.13 in [3], we have:

$$\frac{I(s_1, s_2, \cdots, s_p)}{p-1} \leq \mathrm{DTC}(s_1, s_2, \cdots, s_p) \leq (p-1) I(s_1, s_2, \cdots, s_p). \tag{9}$$

Combining Eqs. (7) and (9), we obtain Eq. (6). □

Proposition 2 (Empirical Estimator of $I_{CS}(s_k; s_{[p]\setminus k})$ [6, 7]). *Given N observations $\left\{\left(s_1^i, s_2^i, \ldots, s_p^i\right)\right\}_{i=1}^N$, each observation contains p different types of measurements $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2, \ldots, s_p \in \mathcal{S}_p$. Let Q and L denote, respectively, the Gram matrices for variable s_k and all rest variables $s_{[p]\setminus k} = [s_1, \cdots, s_{k-1}, s_{k+1}, \cdots, s_p]$, e.g., $L(i,j) = \exp\left(-\frac{\left\|s^i_{[p]\setminus k} - s^j_{[p]\setminus k}\right\|^2}{2\sigma^2}\right)$. The empirical estimator of $I_{CS}(s_k; s_{[p]\setminus k})$ is given by:*

$$\widehat{I}_{CS}(s_k; s_{[p]\setminus k}) = \log\left(\frac{1}{N^2}\sum_{i,j}^{N} Q_{ij}L_{ij}\right) + \log\left(\frac{1}{N^4}\sum_{i,j,q,r}^{N} Q_{ij}L_{qr}\right) - 2\log\left(\frac{1}{N^3}\sum_{i,j,q}^{N} Q_{ij}L_{iq}\right). \tag{10}$$

Proof. By definition, we have:

$$\begin{aligned}
I_{\mathrm{CS}}(s_k, s_{[p]\setminus k}) &= D_{\mathrm{CS}}(p(s_k, s_{[p]\setminus k}); p(s_k)p(s_{[p]\setminus k})) \\
&= -\log\left(\frac{\left|\int p(s_k, s_{[p]\setminus k})p(s_k)p(s_{[p]\setminus k}) ds_k ds_{[p]\setminus k}\right|^2}{\int p^2(s_k, s_{[p]\setminus k}) ds_k ds_{[p]\setminus k} \int p^2(s_k)p^2(s_{[p]\setminus k}) ds_k ds_{[p]\setminus k}}\right) \\
&= \log\left(\int p^2(s_k, s_{[p]\setminus k}) ds_k ds_{[p]\setminus k}\right) + \log\left(\int p^2(s_k)p^2(s_{[p]\setminus k}) ds_k ds_{[p]\setminus k}\right) \\
&\quad - 2\log\left(\int p(s_k, s_{[p]\setminus k})p(s_k)p(s_{[p]\setminus k}) ds_k ds_{[p]\setminus k}\right).
\end{aligned} \tag{11}$$

All three terms inside the "log" can be estimated by KDE as follows [8, 9]:

$$\begin{aligned}
\int p^2(s_k, s_{[p]\setminus k}) ds_k ds_{[p]\setminus k} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \kappa(s_k^i - s_k^j)\kappa(s^i_{[p]\setminus k} - s^j_{[p]\setminus k}) \\
&= \frac{1}{N^2}\sum_{i,j}^{N} Q_{ij}L_{ij},
\end{aligned} \tag{12}$$

in which κ refers to a Gaussian kernel with width σ and takes the form of $\kappa(x-y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$.

$$\begin{aligned}
\int p(s_k, s_{[p]\setminus k})p(s_k)p(s_{[p]\setminus k}) ds_k ds_{[p]\setminus k} &= \mathbb{E}\left[p(s_k)p(s_{[p]\setminus k})\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{N}\sum_{j=1}^{N}\kappa(s_k^i - s_k^j)\right)\left(\frac{1}{N}\sum_{q=1}^{N}\kappa(s^i_{[p]\setminus k} - s^q_{[p]\setminus k})\right)\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{q=1}^{N}\kappa(s_k^i - s_k^j)\kappa(s^i_{[p]\setminus k} - s^q_{[p]\setminus k}) \\
&= \frac{1}{N^3}\sum_{i,j,q}^{N} Q_{ij}L_{iq},
\end{aligned} \tag{13}$$

$$
\begin{aligned}
\int p^2(s_k)p^2(s_{[p]\setminus k})ds_k ds_{[p]\setminus k} &= \left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(s_k^i - s_k^j)\right]\left[\frac{1}{N^2}\sum_{q=1}^{N}\sum_{r=1}^{N}\kappa(s_{[p]\setminus k}^q - s_{[p]\setminus k}^r)\right] \\
&= \frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{q=1}^{N}\sum_{r=1}^{N}\kappa(s_k^i - s_k^j)\kappa(s_{[p]\setminus k}^q - s_{[p]\setminus k}^r) \\
&= \frac{1}{N^4}\sum_{i,j,q,r}^{N} Q_{ij}L_{qr}.
\end{aligned} \tag{14}
$$

Combine Eqs. (12)-(14) with Eq. (11), we obtain Eq. (10).

□

3. REFERENCES

[1] Nikhil Shankar, Gautam Shreedhar Bhat, Issa Panahi, Stephanie Tittle, and Linda M Thibodeau, "Smartphone-based single-channel speech enhancement application for hearing aids," *The Journal of the Acoustical Society of America*, vol. 150, no. 3, pp. 1663–1673, 2021.

[2] Emanuel Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[3] Tim Austin, "Multi-variate correlation and mixtures of product measures," *Kybernetika*, vol. 56, no. 3, pp. 459–499, 2020.

[4] TH Sun, "Linear dependence structure of the entropy space," *Inf Control*, vol. 29, no. 4, pp. 337–68, 1975.

[5] Shujian Yu, Francesco Alesiani, Xi Yu, Robert Jenssen, and Jose Principe, "Measuring dependence with matrix-based entropy functional," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 10781–10789.

[6] "Information theory meets deep learning," `https://rb.gy/44t0p`, 2023, IEEE ICASSP Tutorial.

[7] Jose C Principe, *Information theoretic learning: Renyi's entropy and kernel perspectives*, Springer Science & Business Media, 2010.

[8] Sohan Seth and José C Príncipe, "On speeding up computation in information theoretic learning," in *2009 International Joint Conference on Neural Networks*. IEEE, 2009, pp. 2883–2887.

[9] Leonardo Barroso Gonçalves and José Leonardo Ribeiro Macrini, "Rényi entropy and cauchy-schwartz mutual information applied to mifs-u variable selection algorithm: a comparative study," *Pesquisa Operacional*, vol. 31, pp. 499–519, 2011.