**Paper Title**

Nearfield acoustic holography based on the equivalent source method and reproducing kernels

**Authors**

You-Siang Chen, National Tsing Hua University  
Da-Rong Tseng, National Tsing Hua University  
Mingsian Bai, National Tsing Hua University

**Reviewer 1**31A1

* Importance/Relevance to ICASSP 2024→ Of limited interest
* Justification of Importance/Relevance Score

The topic of the paper, nearfield acoustical holography (NAH), is of limited interest to the audience of ICASSP and is better suited to an acoustics conference.

* Novelty/Originality → Minor originality
* Justification of Novelty/Originality Score

Several regularization have been proposed for the NAH problem. This regularization used here is quadratic, which is well-known (Tikhonov regularization).

* Technical Correctness → Has major problems
* Justification of Technical Correctness Score

The proposed regularization is based on the assumption that the distribution of sources to be estimated is a Gaussian process with the zero-th order spherical Bessel function as autocorrelation function.

This is not clear from reading the paper, and no justification is proposed for this model. In [21] and [22], it is the soundfield itself, and not a distribution of sources, which is assumed to be a realization of such a random process. The acoustical wave equation combined with the assumption that the soundfield is diffuse, support this model.

There is no reason that the distribution of sources should follow this model. In this particular case, the source of the acoustical field is the vibration of a plate, which is not governed by the acoustical wave equation.

* Experimental Validation → Lacking in some respect
* Justification of Experimental Validation Score

It turns out that on the simulation shown in the paper, the proposed method works relatively well. This not sufficient to recommend acceptation. The method should be tested on real data.

* Clarity of Presentation → Clear enough
* Justification of Clarity of Presentation Score

The paper is mostly clear, except the introduction of the kernel which is not justified.

* Reference to Prior Work → Does not cite relevant work
* Justification of Reference to Prior Work Score

Compressive sampling for NAH was introduced in G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, R. Gribonval, Near-field acoustic holography using sparse regularization and compressive sampling principles , J. Acoust. Soc. Am., 2012, This method can be applied on the examples used in the numerical validation.

Switched name and surname for Ingrid Daubechies in [27]

**Reviewer 23B05**

* Importance/Relevance to ICASSP 2024 → Of sufficient interest
* Novelty/Originality → Moderately original
* Technical Correctness → Probably correct
* Experimental Validation → Lacking in some respect
* Justification of Experimental Validation Score

This reviewers has concerns regarding the evaluation in that it is based on simulations only whereby the simulated signals are entirely free of noise, and there not position errors of the microphones etc. assumed. This is far away from reality.

* Clarity of Presentation → Very clear
* Reference to Prior Work → References adequate
* Additional comments to author(s)

The paper is well written, and the mathematical development of the presented method appears to be good, too. However, the evaluation is performed based on simulations under the most ideal possible conditions. There is no noise whatsoever in the signals, which is a fundamental shortcoming.

**Reviewer 342B3**

* Importance/Relevance to ICASSP 2024 → Of sufficient interest
* Novelty/Originality → Moderately original
* Technical Correctness → Probably correct
* Experimental Validation → Limited but convincing
* Clarity of Presentation → Clear enough
* Reference to Prior Work → References adequate
* Additional comments to author(s)

The submitted article discusses the use of kernel ridge regression as a means of inferring the distribution of virtual sources in NAH. The authors address the least squares regression problem under Tikhonov regularization, by applying the classical "kernel trick". This allows them to avoid direct discretization of the continuous distribution of virtual sources, and efficiently solve the infinite-dimensional optimization problem. The distribution has a parametric form, which, in turn, enables them to increase the spatial resolution in the pressure domain.

The results on a small scale simulation experiment are convincing. It would be welcome to provide an intuition on the observed phenomenon that all methods tend to improve with the increase in frequency.

The known problem with kernel ridge regression in sound field interpolation is the fitness of isotropic kernels, such as the chosen zero-order spherical Bessel function, for anisotropic sound fields. Can the authors comment on that?

### Paper Title

Blind estimation of acoustic transfer functions with application to dereverberation using convolutive transfer functions

### Authors

An-Chi Yuan, National Tsing Hua University  
You-Siang Chen, National Tsing Hua University  
Mingsian Bai, National Tsing Hua University

**Reviewer 1309**

* Importance/Relevance to ICASSP 2024 → Of limited interest
* Justification of Importance/Relevance Score

See general comments.

* Novelty/Originality → Has been done before
* Justification of Novelty/Originality Score

See general comments.

* Technical Correctness → Contains minor errors
* Justification of Technical Correctness Score

See general comments.

* Experimental Validation → Limited but convincing
* Clarity of Presentation → Clear enough
* Reference to Prior Work → References missing
* Justification of Reference to Prior Work Score

See general comments.

* Additional comments to author(s)

The level of this paper is low, no noticeable original contribution is presented. The details of getting the reference signal via delay and sum beamformer are omitted. The references are from 2010 or older!

**Reviewer 2 2B**BB

* Importance/Relevance to ICASSP 2024 → Of sufficient interest
* Novelty/Originality → Moderately original
* Technical Correctness → Probably correct
* Experimental Validation → Limited but convincing
* Clarity of Presentation → Very clear
* Reference to Prior Work → References adequate
* Additional comments to author(s)

This paper proposes a blind estimation of acoustic transfer functions (room impulse responses) using microphone arrays. It lives in the STFT domain and uses long term convolution for allowing the estimation of long impulse responses. The blind estimation of the source signal is achieved using a delay and sum beamformer for estimating signal statistics and Wiener filtering or recursive least squares.

The paper is well written and provides a nice introduction of all the methods used. It is a good read also for readers who not familiar with all related art documents. The evaluation of the methods shows a good performance in a dereverberation scenario using simulated data.

Without knowing all the related work in this field, the approach presented by the authors as novel seems straight forward. I am surprised that this has not been tried before as it is a basic combination of the convolution in STFT domain with beamforming for source estimation and Wiener filtering.

**Reviewer 343**0A

* Importance/Relevance to ICASSP 2024 → Of sufficient interest
* Novelty/Originality → Minor originality
* Justification of Novelty/Originality Score

see comments below

* Technical Correctness → Probably correct
* Experimental Validation → Insufficient validation
* Justification of Experimental Validation Score

see comments below

* Clarity of Presentation → Clear enough
* Reference to Prior Work → References missing
* Justification of Reference to Prior Work Score

see comments below

* Additional comments to author(s)

This paper presents a method for blind system identification (BSI) of room impulse responses (RIRs) / acoustic transfer functions (ATFs). First, a crude estimate of the source signal is estimated using the weighted prediction error (WPE) method followed by a delay-and-sum beamformer. Next, the convolutive transfer function (CTF) coefficients are estimated either using a Wiener filter or using the recursive least squares algorithm. Finally, the ATFs are recovered by convolving the STFT of a unit pulse with the estimated CTFs. The performance of the proposed method is evaluated using a 30-element uniform microphone array for different reverberation times (without noise) and compared with the WPE method. Since several important details are missing and the proposed method is not compared to other state-of-the-art BSI methods, in my opinion the paper can not be accepted in its present form.

Major comments:

1) In the flow chart (page 3) it is mentioned that {\bf s}\_{DAS}(n) is obtained by WPE followed by delay-and-sum beamforming. Several details are missing:

- I assume that a MIMO version of WPE is used (since followed by delay-and-sum beamforming) - this is never mentioned in the paper.

- the filter length and the prediction delay for WPE are not mentioned in the paper.

- it should be clearly mentioned in the paper that the delay-and-sum beamformer requires both the microphone array configuration as well as the direction-of-arrival of the source to be known. This implies that the proposed method is not fully blind !

2) The number of CTF coefficients L is not mentioned in the paper. From Fig. 2 I assume that the length of the simulated RIRs was equal to 6000 - it is never mentioned in the paper how long the length of the estimated RIRs is (this depends on STFT length, overlap and L).

3) It is unclear from the paper if the proposed method was compared to WPE or to WPE followed by a delay-and-sum beamformer (to be fair, it needs to be the latter !).

4) Several BSI methods for RIRs have been proposed in the literature, none of which has been cited or compared with:

- Y. Huang and J. Benesty, “Adaptive multi-channel least mean square and Newton algorithms for blind channel identification,” Signal Processing, vol. 82, no. 8, pp. 1127–1138, Aug. 2002.

- Y. Huang and J. Benesty, “A class of frequency-domain adaptive approaches to blind multichannel identification,” IEEE Trans. Signal Processing, vol. 51, no. 1, pp. 11–24, Jan. 2003.

- M. K. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, “Improving robustness of blind adaptive multichannel identification algorithms using constraints,” in Proc. European Signal Processing Conference (EUSIPCO), 2005.

- M. K. Hasan and P. A. Naylor, “Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: robustness issue,” in Proc. European Signal Processing Conference (EUSIPCO), 2006.

- M. A. Haque and M. K. Hasan, “Noise robust multichannel frequency-domain LMS algorithms for blind channel identification,” IEEE Signal Processing Letters, vol. 15, pp. 305–308, 2008.

5) Most state-of-the-art BSI methods (see comment above) were shown to be rather sensitive to noise and the (assumed to be known) length of the RIRs.

- instead of only using simulated RIRs (with known length), it is essential to include simulations with recorded microphone signals

- simulations with (a limited amount of) noise should be included.

Minor comments:

1) p.3: the recursive averaging -> recursive averaging

2) p.4: please mention the reference signal used for PESQ and SDR

3) the simulations have been performed with a large number of microphones (30). It would be interesting to include simulations for a limited number of microphones (or show the performance of WPE and the proposed method as a function of the number of microphones).

4) the authors should explain why matching error (ME) instead of normalized projection misalignment (NPM) was used as performance metric.

5) it would be nice to include a website with audio samples

**Reviewer 40**70E

* Importance/Relevance to ICASSP 2024 → Of sufficient interest
* Novelty/Originality → Moderately original
* Technical Correctness → Probably correct
* Experimental Validation → Lacking in some respect
* Justification of Experimental Validation Score

"State-of-the-art" method is from 2010, probably not state-of-the-art anymore?

* Clarity of Presentation → Clear enough
* Reference to Prior Work → References missing
* Justification of Reference to Prior Work Score

The latest paper cited is from 2019. I'm not an expert of dereverberation but I believe there're newer paper out there with meaningful results.