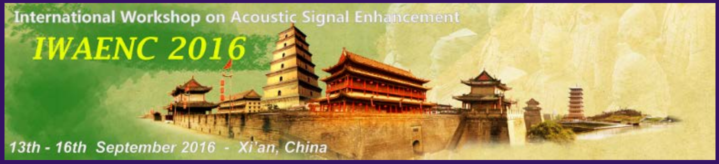
Estimating power spectral density for acoustic signal enhancement

- An effective approach for practical applications -

Department of Mechanical Engineering
University of Auckland
Yusuke Hioka

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

ENGINEERING



Outline

I. Research Background
II. PSD estimation in beamspace
III. Applications using the estimated PSD
   1. Directional sound source separation
   2. Distance distinguishing microphone
   3. "Zooming" microphone array
   4. UAV recording
   5. Blind acoustic characterisation
IV. Summary

Acknowledgement

Many thanks to the colleagues at:

- University of Auckland, New Zealand
  - Dr Michael Kingan
  - Mr Gian Schmid
  - Dr Karl Stol

- NTT Medial Intelligence Laboratories, Japan
  - Dr Kenta Niwa
  - Ms Tomoko Kawase
  - Prof Yoichi Haneda (University of Electro-Communications, Japan)
  - Prof Kenichi Furuya (Oita University, Japan)

I. Research Background

Acoustic signal enhancement

- When do we need acoustic signal enhancement?
  - Recording a dialogue in a noisy public space
  - Sending intelligible speech for hand-free calls
  - Improving speech recognition accuracy
  - Extracting a melody of an instrument for transcription

http://www.nec.co.jp/press/en/0703/0501.html

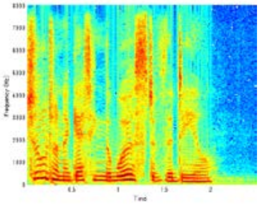
etc. etc.

http://www.btas.com.au/products/main/polycomvsx7000e.shtml

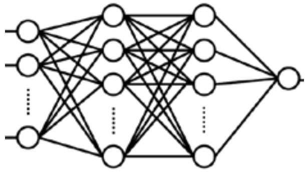
https://en.wikipedia.org/wiki/Orchestra

Available techniques for acoustic signal enhancement

Spectral Manipulation
- Wiener postfilter
- Spectral subtraction
- Nonnegative Matrix Factorisation

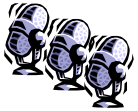


Statistical method / Machine learning
- Independent Component Analysis
- Neural Network



Hardware
- Shotgun
- Parabola
- Cardioid

Microphone Arrays
- Fixed beamforming
- Adaptive beamforming



Software (digital signal processing)

Reality of practical problems

- Limitation on hardware
  - Deviation of devices’ performance
  - Size/space

- Harsh acoustical environments
  - Variety of noise types
  - Reverberation
  - Extremely high noise level

http://techon.nikkeibp.co.jp/article/HONSHI/20060730/119674/

PRACTICALLY EFFECTIVE
Acoustic signal enhancement

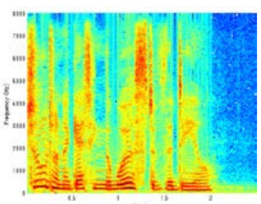
Better performance by combination

Spectral Manipulation

- Wiener postfilter
- Spectral subtraction
- Nonnegative Matrix Factorisation

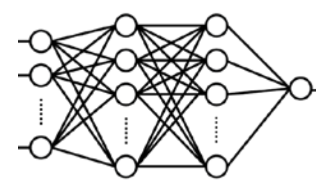
Statistical method / Machine learning

- Independent Component Analysis
- Neural Network



Hardware

- Shotgun
- Parabola
- Cardioid



Microphone Arrays

- Fixed beamforming
- Adaptive beamforming

Software
(digital signal processing)

DSP techniques - Pros & Cons -

Technique	Pros	Cons
Microphone Arrays	• Linear processing • Calculation cost is relatively low	• May be sensitive to errors in practical environment causing significant performance degradation
Spectral manipulation	• Robust to errors in practical environment • Low calculation cost	• May suffer from musical noise • Spectral information of signal/noise is required
Statistical method	• Very high performance if the model fits the actual problem	• High calculation cost • Often sensitive to errors in practical environment • May require training process

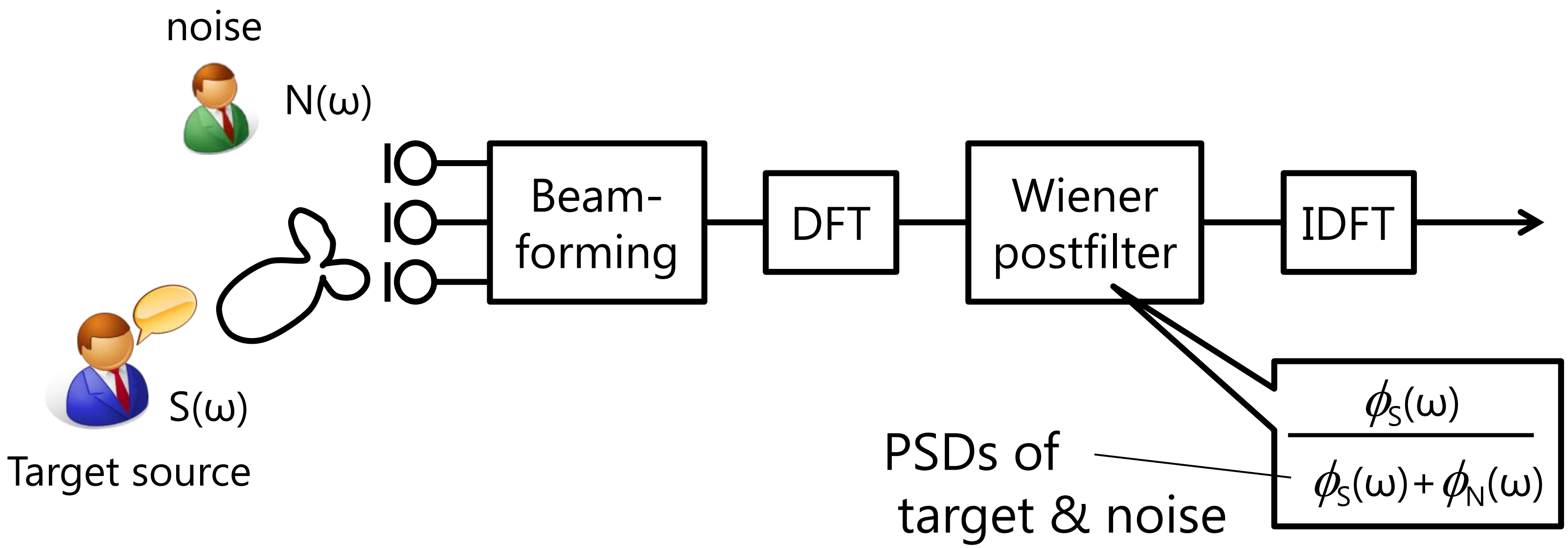
Because of its robustness to errors **Spectral Manipulation** *is often used in combination with* **Microphone Arrays** for practical applications.

Beamforming with Wiener postfilter

- Combination of **Microphone Arrays** & **Spectral Manipulation**
- Pros & Cons

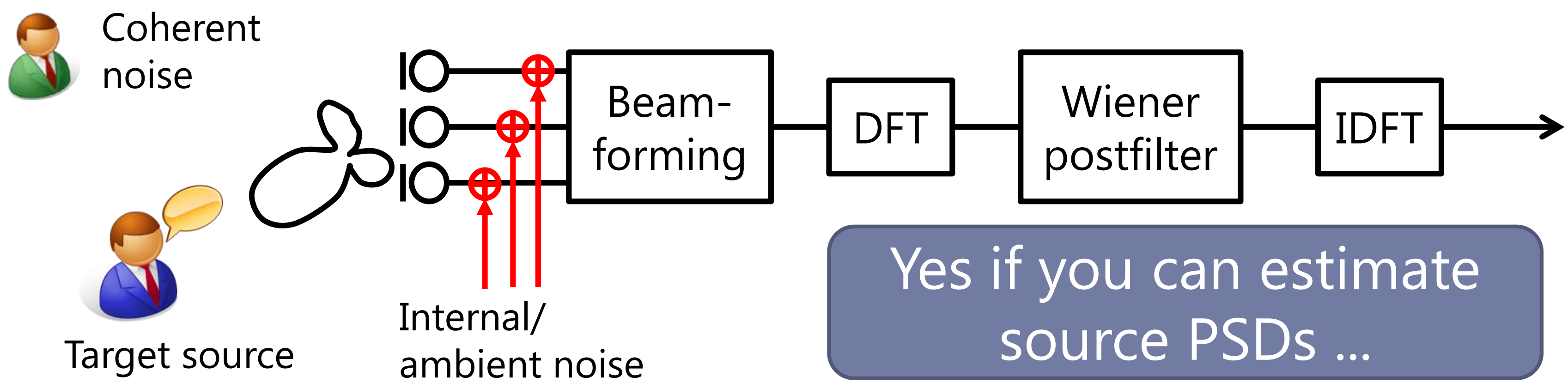
  (+) Robust and high performance in practical environment

  (-) Power spectral density (PSD) of target sound and noise needs to be estimated



BF with Wiener postfilter (cont'd)

- Originally developed for compensating the beamforming being less effective for reducing (*spatially*) *incoherent noise*, including *microphones' internal noise* and *ambient noise* of the environment
- Conventional PSD estimation for microphone arrays
  - Spatially incoherent noise (Zelinski 88)
  - Diffuse noise (McCowan 03)
- Question: Can the technique be extended for reducing other signals e.g. *coherent* noise?



Yes if you can estimate source PSDs …

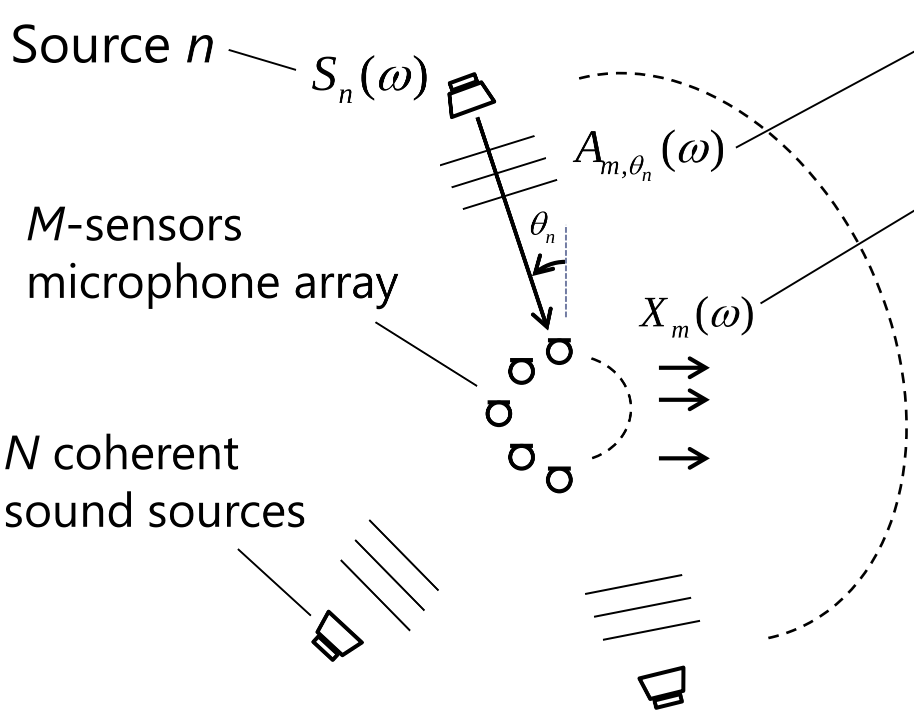
II. PSD estimation in beamspace

Problem (simplest case)

Estimate **PSD** of

- *N* ***coherent*** sound sources located in different angles;
- using an *M*-sensors microphone array;

(in order to calculate the Wiener post-filter).



Transfer function
from source *n* to microphone *m*

Input signal of microphone *m*

$$X_m(\omega) = \sum_{n=1}^{N} A_{m,\theta_n}(\omega) S_n(\omega)$$

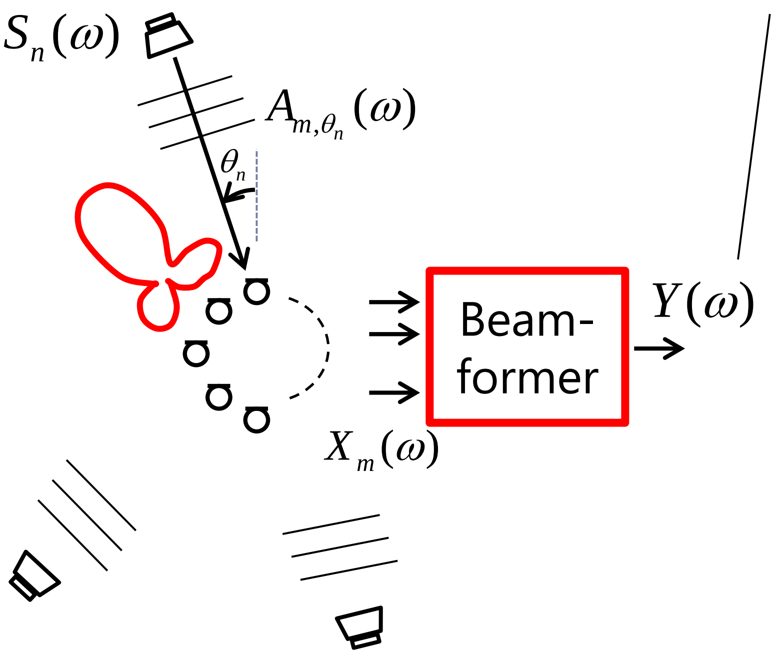
Assumption

All sound sources are
<u>mutually uncorrelated</u>

$$E\left[S_i(\omega) S_j^*(\omega) \right] = 0$$

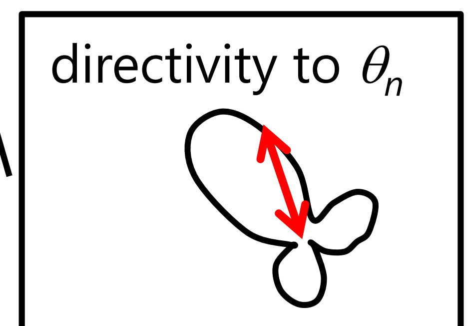
Beamforming

- Assume a beamforming is applied to the microphone array observation.



Output of beamformer

$$Y(\omega) = \sum_{m=1}^{M} W_m(\omega) X_m(\omega)$$
$$= \sum_{m=1}^{M} \sum_{n} W_m(\omega) A_{m,\theta_n}(\omega) S_n(\omega)$$
$$= \sum_{n} \mathbf{w}_l^H(\omega) \mathbf{a}_{\theta_n}(\omega) S_n(\omega)$$
$$= \sum_{n} D_{\theta_n}(\omega) S_n(\omega)$$



$$\mathbf{a}_{\theta_n}(\omega) = \left[A_{1,\theta_n}(\omega) \quad \cdots \quad A_{M,\theta_n}(\omega) \right]^T$$
$$\mathbf{w}_l(\omega) = \left[W_1(\omega) \quad \cdots \quad W_M(\omega) \right]^T$$

PSD of beamformer's output

- PSD of a beamformer's output can be approximated by simple additive model: **Σ(directivity gain × source PSD)**

PSD of beamformer output

Output of a beamformer

$$Y(\omega)=\sum_{n} D_{\theta_n}(\omega)S_n(\omega)$$

$$\phi_Y(\omega)=E[Y(\omega)Y^*(\omega)]$$

$$=\sum_{i}\left|D_{\theta_n}(\omega)\right|^2 E\left[\left|S_n(\omega)\right|^2\right]$$

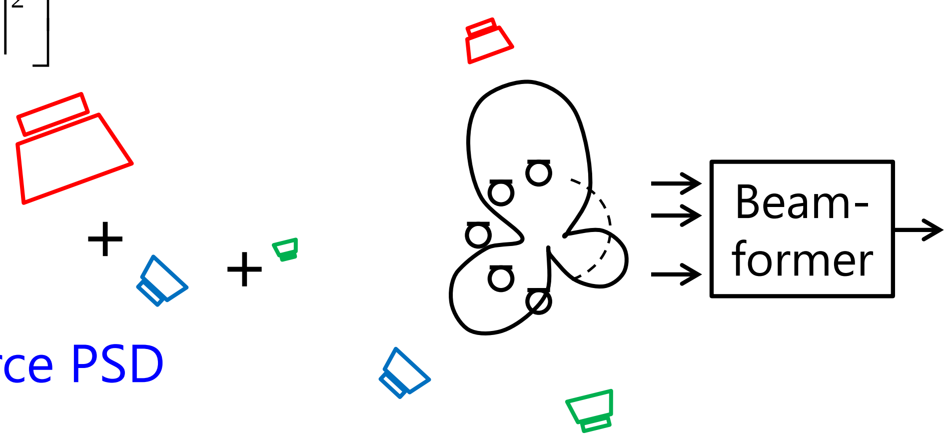
$$+\sum_{n}\sum_{n'\neq n}\left(D_{\theta_n}(\omega)D^*_{\theta_{n'}}(\omega)E\left[S_n(\omega)S^*_{n'}(\omega)\right]+D_{\theta_{n'}}(\omega)D^*_{\theta_n}(\omega)E\left[S_{n'}(\omega)S^*_n(\omega)\right]\right)$$

$$\approx\sum_{n}\left|D_{\theta_n}(\omega)\right|^2 E\left[\left|S_n(\omega)\right|^2\right]$$

$$=\sum_{n}\left|D_{\theta_n}(\omega)\right|^2\phi_n(\omega)$$

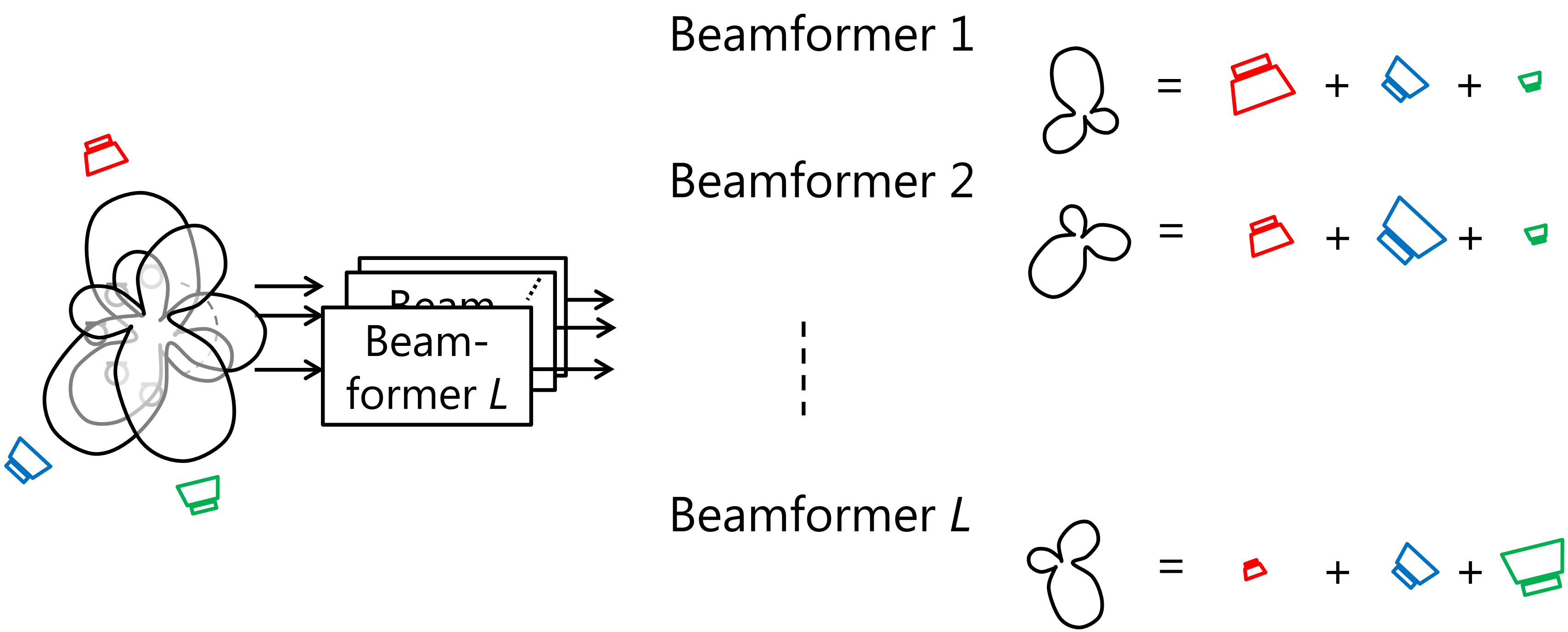
Directivity gain

Source PSD



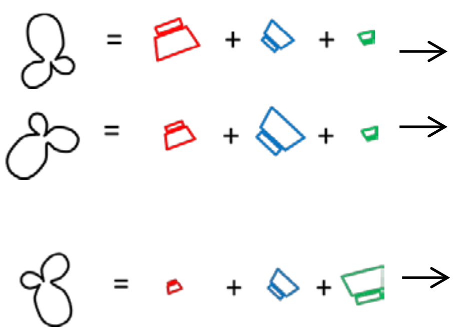
What if we have more beamformers?

- Applying L (≧N) different beamformers will introduce L different combination of directivity gains applied to source PSDs.

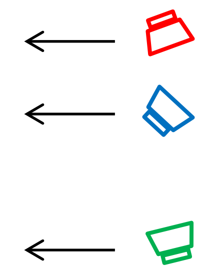


PSD estimation in beamspace

- Relation between beamformers' outputs and source PSDs can be described by a simultaneous equation.



$$\underbrace{\begin{bmatrix} \phi_{Y_1}(\omega) \\ \phi_{Y_2}(\omega) \\ \vdots \\ \phi_{Y_L}(\omega) \end{bmatrix}}_{\Phi_Y(\omega)} \approx \underbrace{\begin{bmatrix} |D_{1,\theta_1}(\omega)|^2 & |D_{1,\theta_2}(\omega)|^2 & \cdots & |D_{1,\theta_N}(\omega)|^2 \\ |D_{2,\theta_1}(\omega)|^2 & & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ |D_{L,\theta_1}(\omega)|^2 & & \cdots & |D_{L,\theta_N}(\omega)|^2 \end{bmatrix}}_{\mathbf{D}(\omega)} \underbrace{\begin{bmatrix} \phi_{S_1}(\omega) \\ \phi_{S_2}(\omega) \\ \vdots \\ \phi_{S_N}(\omega) \end{bmatrix}}_{\Phi_S(\omega)}$$



PSD of beamformers' outputs

Directivity gains

PSD of sound sources

- Source PSDs can be estimated by solving the equation using e.g. least squares method

Estimated source PSDs

$$\hat{\Phi}_S(\omega) = \mathbf{D}^{+}(\omega)\Phi_Y(\omega)$$

Used to calculate Wiener filter $\dfrac{\phi_{S_n}(\omega)}{\sum_n \phi_{S_n}(\omega)}$

Extension for noisy environment

- Diffuse noise can also be included in the model (like many existing studies did).
- PSD of diffuse noise can be separately estimated by adding another column in the gain matrix **D**(*ω*).

$$\underbrace{\begin{bmatrix}\phi_{Y_1}(\omega)\\ \phi_{Y_2}(\omega)\\ \vdots\\ \phi_{Y_L}(\omega)\end{bmatrix}}_{\Phi_Y(\omega)} \approx \underbrace{\begin{bmatrix}\left|D_{1,\theta_1}(\omega)\right|^2 & \left|D_{1,\theta_2}(\omega)\right|^2 & \cdots & \left|D_{1,\theta_N}(\omega)\right|^2 & \int_\theta \left|D_{1,\theta}(\omega)\right|^2 d\theta\\ \left|D_{2,\theta_1}(\omega)\right|^2 & \ddots & \cdots & \vdots & \int_\theta \left|D_{1,\theta}(\omega)\right|^2 d\theta\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \left|D_{L,\theta_1}(\omega)\right|^2 & \cdots & \cdots & \left|D_{L,\theta_N}(\omega)\right|^2 & \int_\theta \left|D_{1,\theta}(\omega)\right|^2 d\theta\end{bmatrix}}_{\mathbf{D}(\omega)} \underbrace{\begin{bmatrix}\phi_{S_1}(\omega)\\ \phi_{S_2}(\omega)\\ \vdots\\ \phi_{S_N}(\omega)\\ \phi_{\bar{V}}(\omega)\end{bmatrix}}_{\Phi_S(\omega)}$$

Squared directivity for diffuse noise

PSD of diffuse noise

- Directivity to diffuse noise is modelled using isotropic power distribution of diffuse noise

[Hioka et al., IWAENC2014]

Analysing "D(ω)" matrix

- Performance of the PSD estimation depends on if the directivity gain matrix **D**(ω) is "well-conditioned" or not.
- Despite the size of **D**(ω) being determined by L and N, the <u>condition</u> of **D**(ω) is determined by:
  - the number of microphones in the array (M) and;
  - the shape of directivity pattern of the beamformers.

# of sources (M)

# of bf's (L)

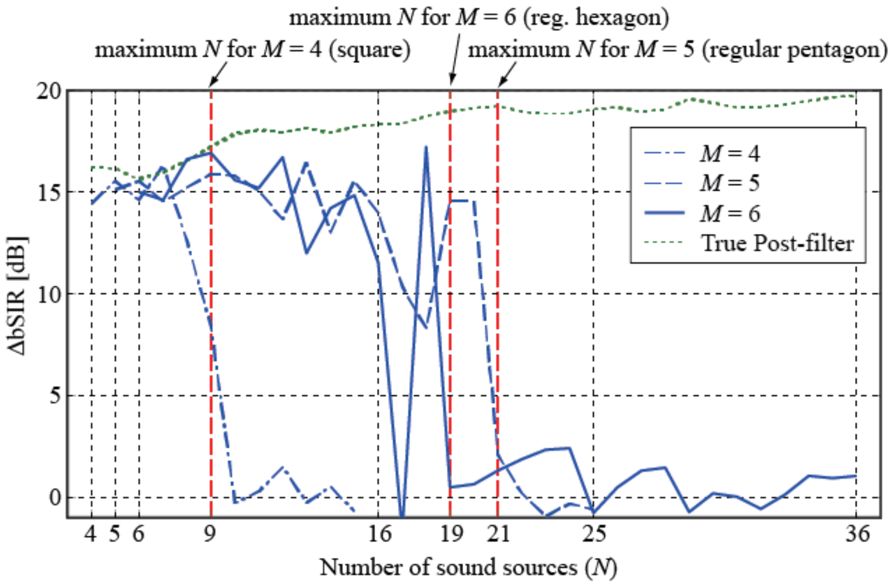
$$\underbrace{\begin{bmatrix} \phi_{Y_1}(\omega) \\ \phi_{Y_2}(\omega) \\ \vdots \\ \phi_{Y_L}(\omega) \end{bmatrix}}_{\Phi_Y(\omega)} \approx \underbrace{\begin{bmatrix} \left|D_{1,\theta_1}(\omega)\right|^2 & \left|D_{1,\theta_2}(\omega)\right|^2 & \cdots & \left|D_{1,\theta_N}(\omega)\right|^2 \\ \left|D_{2,\theta_1}(\omega)\right|^2 & & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \left|D_{L,\theta_1}(\omega)\right|^2 & & \cdots & \left|D_{L,\theta_N}(\omega)\right|^2 \end{bmatrix}}_{\mathbf{D}(\omega)} \underbrace{\begin{bmatrix} \phi_{S_1}(\omega) \\ \phi_{S_2}(\omega) \\ \vdots \\ \phi_{S_N}(\omega) \end{bmatrix}}_{\Phi_S(\omega)}$$

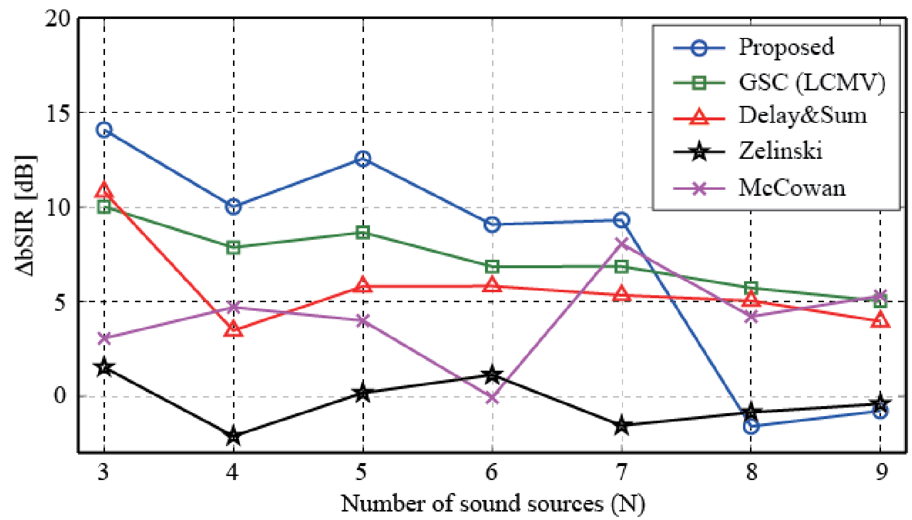
Application to under-determined problems

- Maximum number of separable sources (MNSS) can be analytically derived by looking into the rank of **D**(ω), which is bounded by $M(M\text{-}1)+1$ (M: # of microphones).

May be applied to **under-determined** problems

[Hioka et al. IEEE TASLP2013]





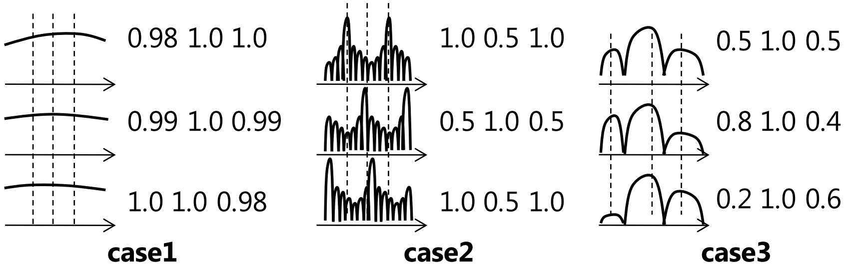
Source separation performance of the Wiener filter designed by using the estimated PSDs for different number of sources. Left: Simulation results for M = 4-6, Right: Experimental results for M=3

Beamformer design

- Beamformers need to be carefully designed in order to avoid causing rank deficiency of **D**(ω).

Cases causing rank deficiency

case1: broad directivity in low frequency
case2: spatial aliasing in high frequency
case3: ill combinations of source angles



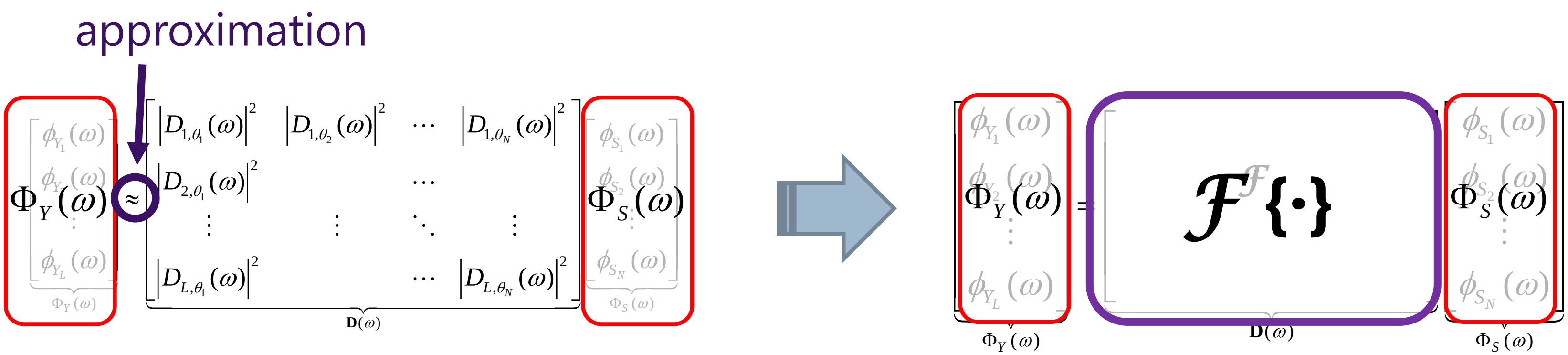
- Attempts to specify a recommended beamformer design:
  - Make $\mathbf{D}^{-1}(\omega)$ (i.e. inverse of **D**(ω)) to be an *M-matrix*

    [Niwa et al., IWAENC2016] **Poster Session I-27**
  - MNSS of 2*M*-1 is guaranteed if a delay-sum beamformer on a *cylindrical array* is utilised.

    [Hioka et al., IEEE WASPAA2013]
- Optimum design of beamformers is still an open problem.

Modelling by nonlinear mapping

- Modelling by linear combination requires approximation.
- More accurate relationship between Φ_Y and Φ_S may be described by a nonlinear mapping.



Mapping by linear combination

Nonlinear mapping

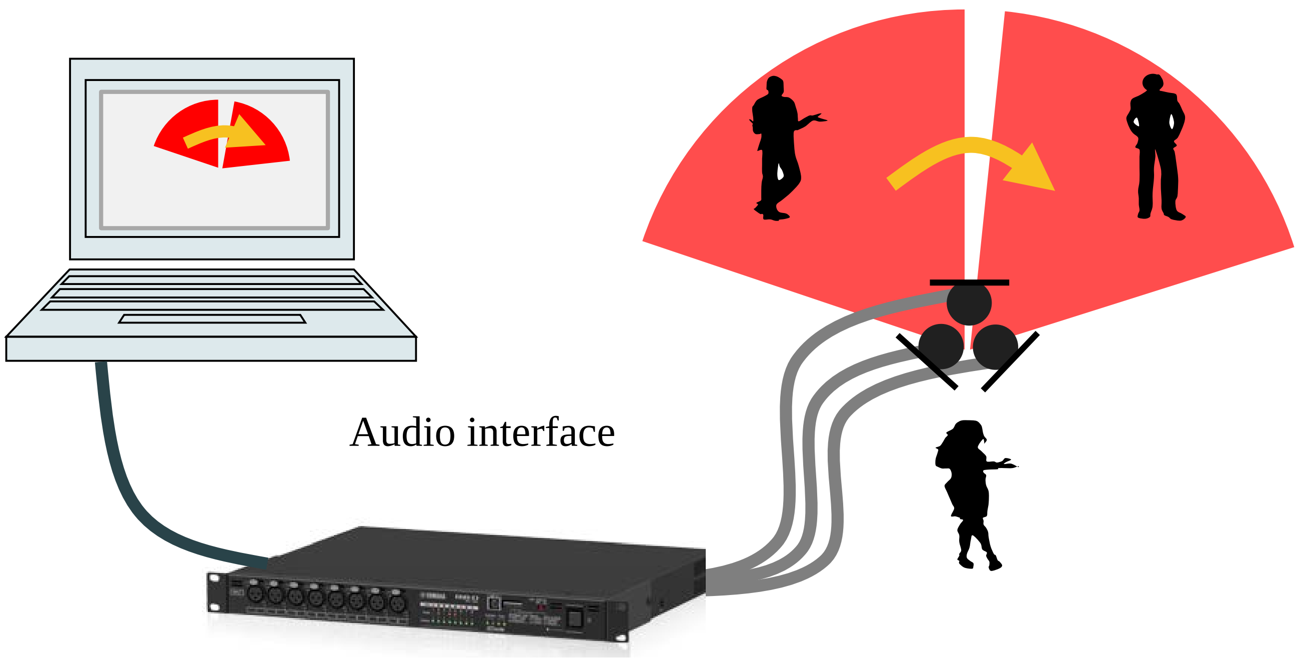
- Attempts to describe the nonlinear mapping by a neural network (NN).
  - Wiener filter estimation using **deep NN** [Niwa et al., ICASSP2016]
  - PSD estimation in beamspace using NN [Kawase et al., IWAENC2016] **Poster Session I-26**

III. Applications using the estimated PSD

1. Directional sound source separation
2. Distance distinguishing microphone
3. “Zooming” microphone array
4. UAV recording
5. Blind acoustic characterisation -DRR estimation-

1. Directional sound separation

- Most straightforward application of the estimated PSD is directional sound separation.
- A source located in particular angle is emphasised by the Wiener filter.



1. Directional sound separation

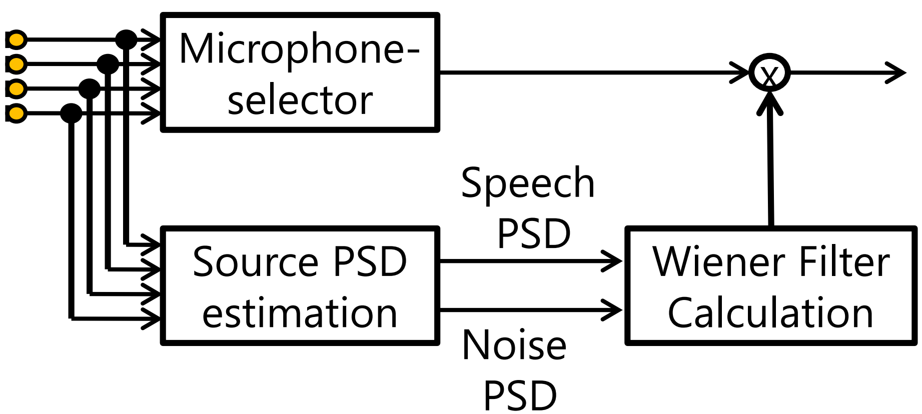
- Demonstration video

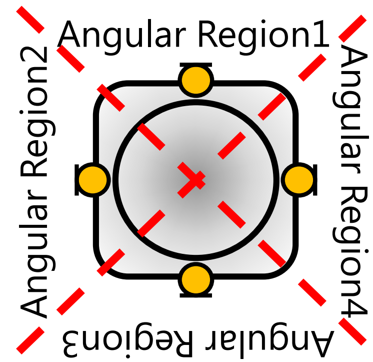
Application to a voice conference system

- Key features
  - 4 mic + 1 loudspeaker
  - RRP: JPY 108,000 (=USD 1,000)
  - Can be connected to various communication devices, e.g. PC, mobile, landline, etc.

(http://www.ntt-at.co.jp/page.jsp?id=1793&content_id=902)

- Key mechanism
  - Each of four angular regions can be muted separately
  - Directional mic + Wiener filter by PSD estimation in beamspace



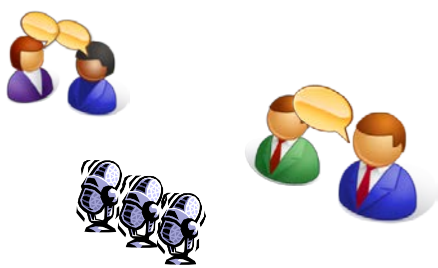


[Hioka et al. IEEE TCE2012]

2. Distance distinguishing microphone

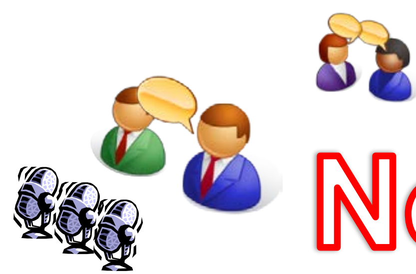
- Most of existing sound source separation techniques are only capable of distinguishing sources located in different **directions** but NOT at different **distances**

Noise in different direction



Yes

Noise in the same direction



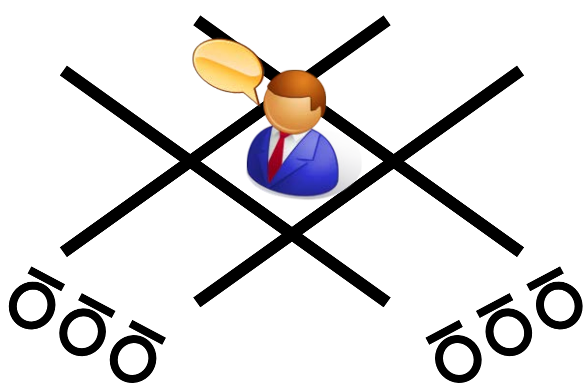
No

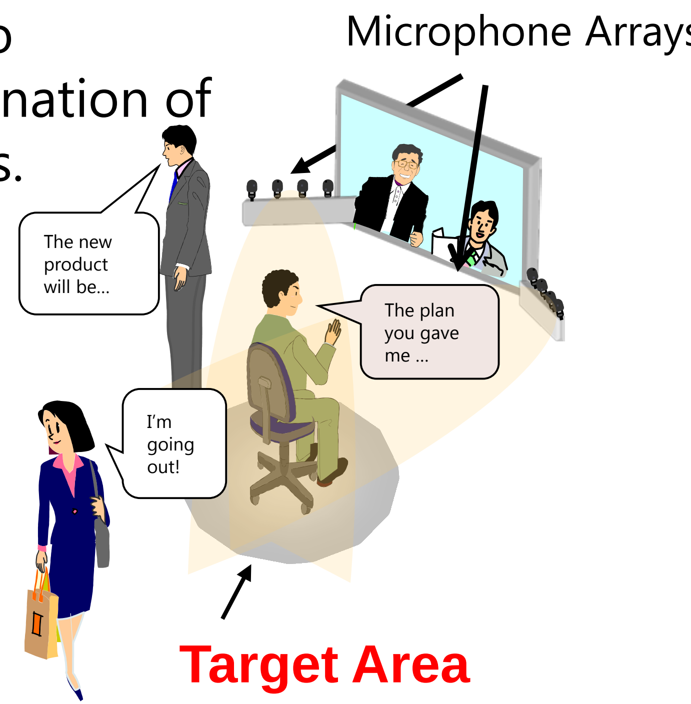
Distance-distinguishing microphone

➡ A sound source located in a particular area (distance & direction) can be extracted

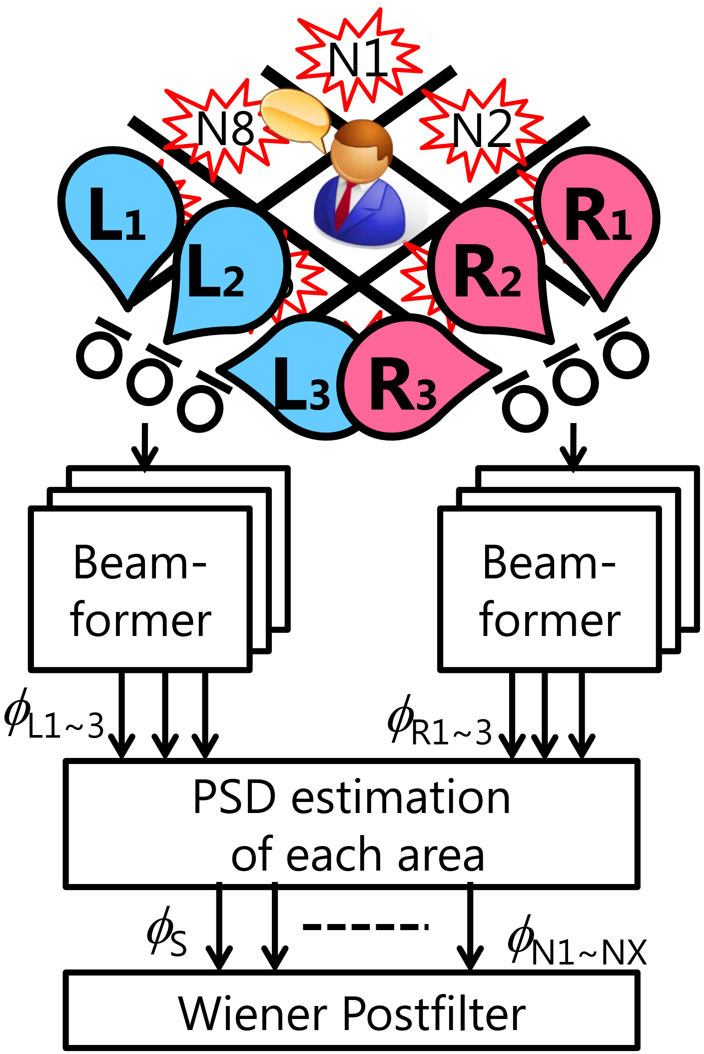
2. Distance distinguishing microphone

- Aim: Emphasise sources in two-dimensional area
- Assumptions:
  - More than one microphone arrays are utilised.
  - Microphone arrays can be located apart.
- Solution:
  - Extend *PSD estimation in beamspace* to 2D scenario by setting ***areas*** by combination of angles from each of microphone arrays.





PSD estimation in beamspace for 2D scenario



1. Define *two-dimensional area* by combination of angles looking from each array
2. Introduce beamformers whose directivity patterns are mutually different (e.g. mainlobe pointing in different direction)
3. Estimate PSD of each area from PSD of multiple beamformers' output

$$\begin{bmatrix} \phi_{L1} \\ \phi_{L2} \\ \phi_{L3} \\ \phi_{R1} \\ \phi_{R2} \\ \phi_{R3} \end{bmatrix} = \begin{bmatrix} D_{L1,S} & \cdots & D_{L1,N8} \\ \vdots & \ddots & \vdots \\ D_{R3,S} & \cdots & D_{R3,N8} \end{bmatrix} \begin{bmatrix} \phi_{S} \\ \phi_{N1} \\ \vdots \\ \phi_{N8} \end{bmatrix}$$

PSD of beamformer's output

directivity gain for each area

area-wise PSD

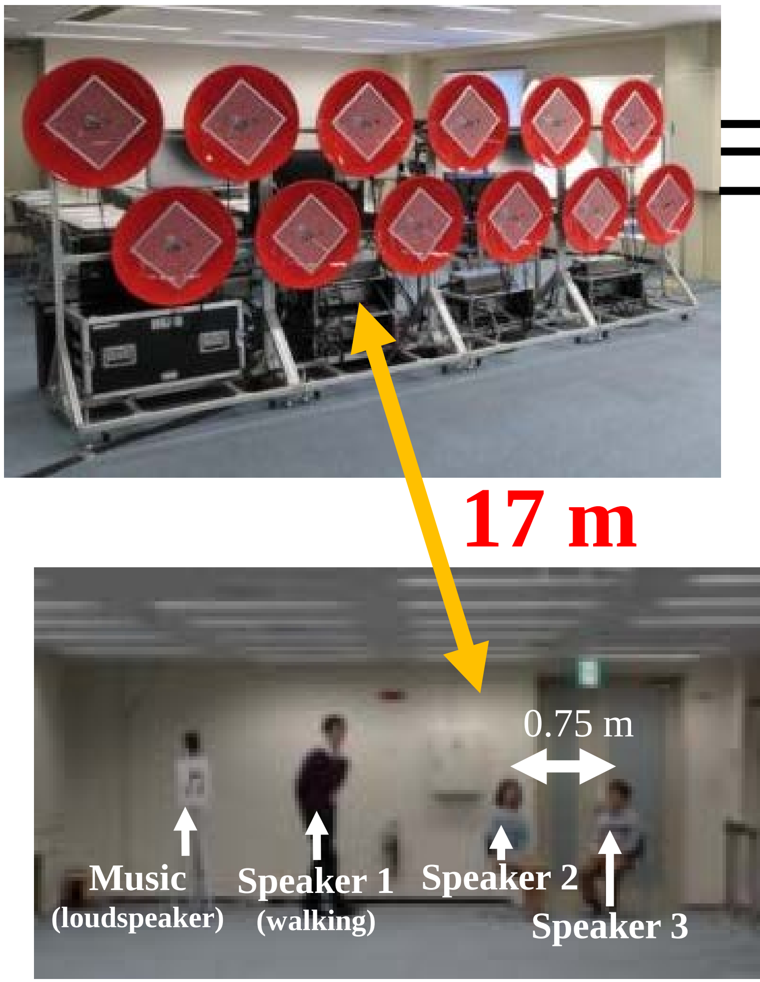
2. Distance distinguishing microphone

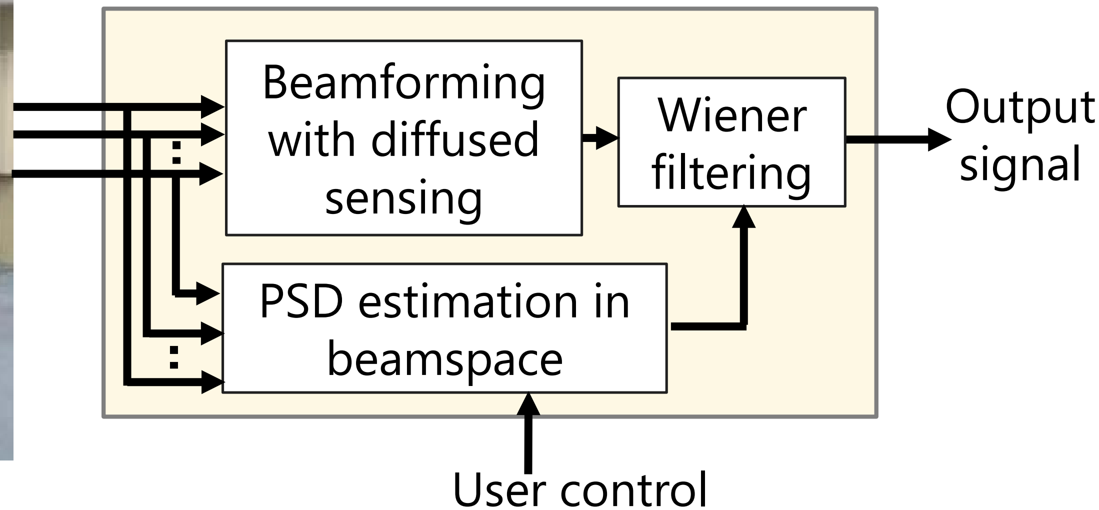
- Demonstration video

On Air in 2009 by TV Tokyo

3. Zooming microphone array

- PSD estimation combined with *optimally designed array*





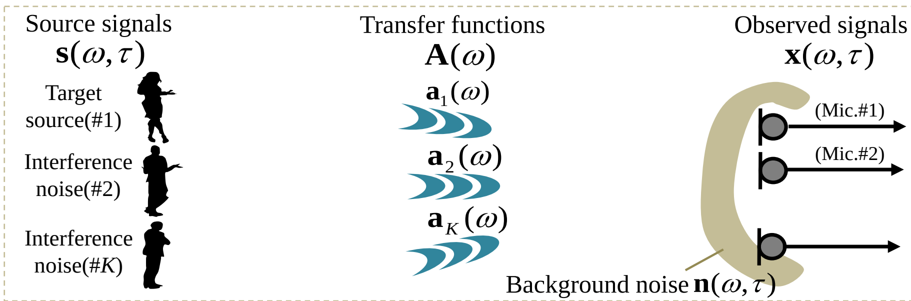
[NIWA et al. IEEE TASLP2016]

Aside: Optimal microphone array design

- A design paradigm of microphone array structure

Problem: Specify microphone array design that maximises the mutual information between **s** and **x**: *I*(**s**;**x**)

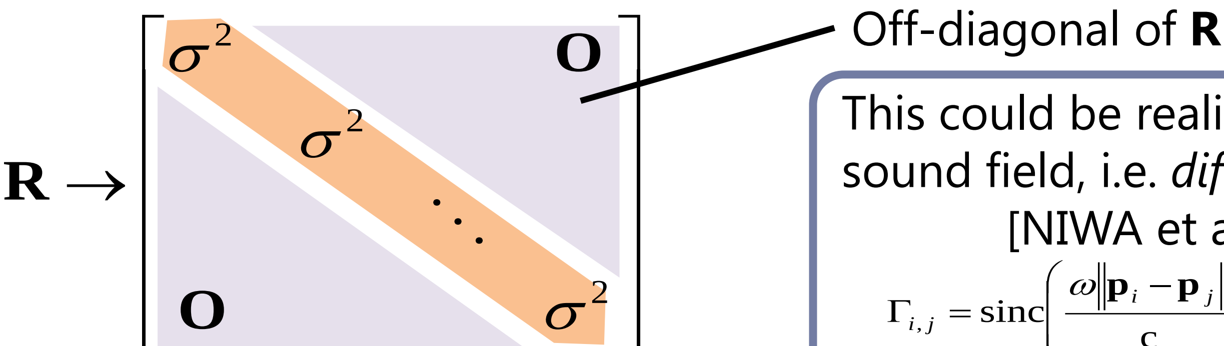
$$\max\{I(\mathbf{s};\mathbf{x})\} = \log_2 \det\left(\sigma_{\mathrm{SN}}^2 \mathbf{R} + \mathbf{I}\right)$$



$$\mathbf{R} = \mathbf{A}\mathbf{A}^H = \begin{bmatrix} \sigma^2 & \Gamma_{1,2} & \cdots & \Gamma_{1,M} \\ \Gamma_{2,1} & \sigma^2 & \cdots & \Gamma_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{M,1} & \Gamma_{M,2} & \cdots & \sigma^2 \end{bmatrix}$$

Spatial correlation matrix

Solution: Design an array that reduces cross-correlation of the transmission paths between microphones (i.e. off-diagonal of **R**).



This could be realised in a diffused sound field, i.e. *diffused sensing*.
[NIWA et al. IEEE TASLP2013]

$$\Gamma_{i,j} = \mathrm{sinc}\left(\frac{\omega\|\mathbf{p}_i - \mathbf{p}_j\|}{c}\right) \rightarrow 0 \text{ if } \|\mathbf{p}_i - \mathbf{p}_j\| \rightarrow \infty$$

3. Zooming microphone array

- Demonstration video

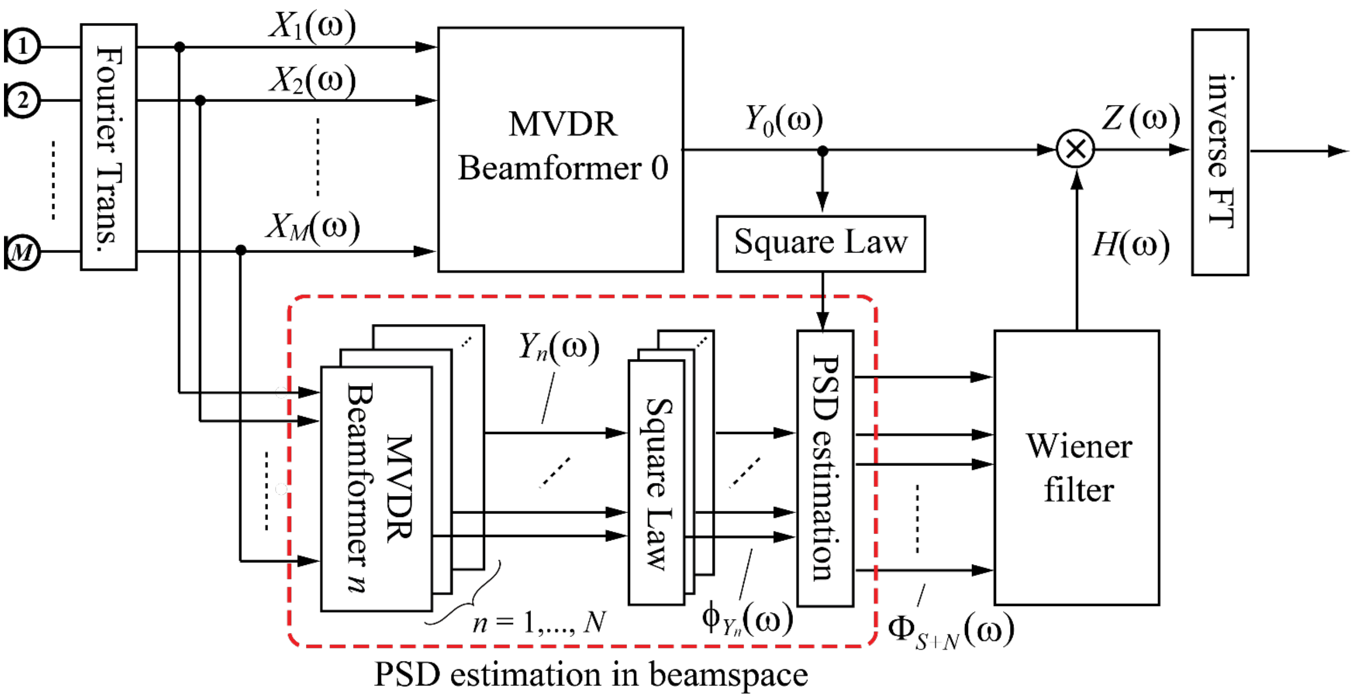
BBC News Oct 2014
http://www.bbc.co.uk/programmes/p029l3hj

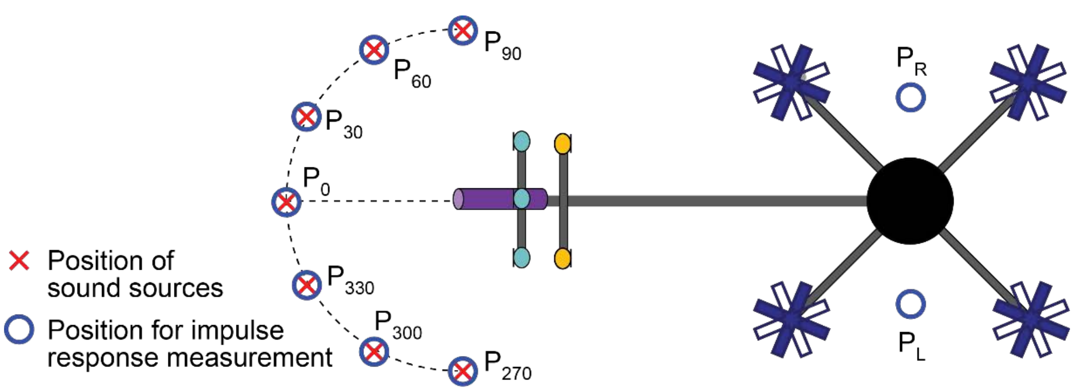
4. UAV recording

- Recently UAVs have been utilised in filming industry but only for video but NOT audio because of huge amount of rotor noise.
- The presenter was a member in one of the six finalist teams of C-Prize, a competition organised by a governmental institute in NZ.
- The developed UAV carried multiple microphones with speech enhancement algorithm based on PSD estimation in beamspace.

http://www.cprize.nz/

4. UAV recording





[Hioka et al. IWAENC2016] **Poster Session I-28**

4. UAV recording

- Demonstration video

Also available on YouTube
https://www.youtube.com/watch?v=X0Rvs7Xoff4

5. Blind acoustic characterisation

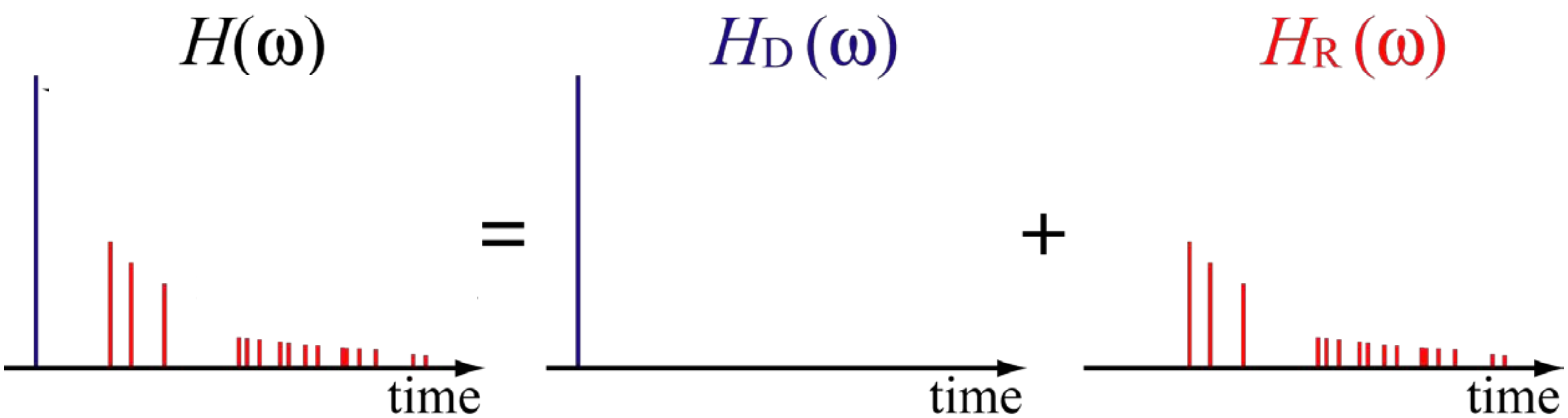
- Estimated PSD can be used for various applications other than acoustic signal enhancement.
- *Blind acoustic characterisation* is an emerging topic that aims to estimate various acoustic parameters, e.g. *reverberation time* (T_{60}) and *direct-to-reverberation ratio* (DRR), without measuring a room impulse response.
- **ACE (acoustic characterisation of environment) Challenge** was held in 2015 where participants competed with others on estimation accuracy of T_{60} and DRR.

ACE Challenge
http://www.ee.ic.ac.uk/naylor/ACEweb/index.html

Direct to reverberation ratio (DRR)

- Energy ratio of **direct sound** and **reverberation** measured at a position in a reverberant room.
- Calculated from a measured room impulse response.

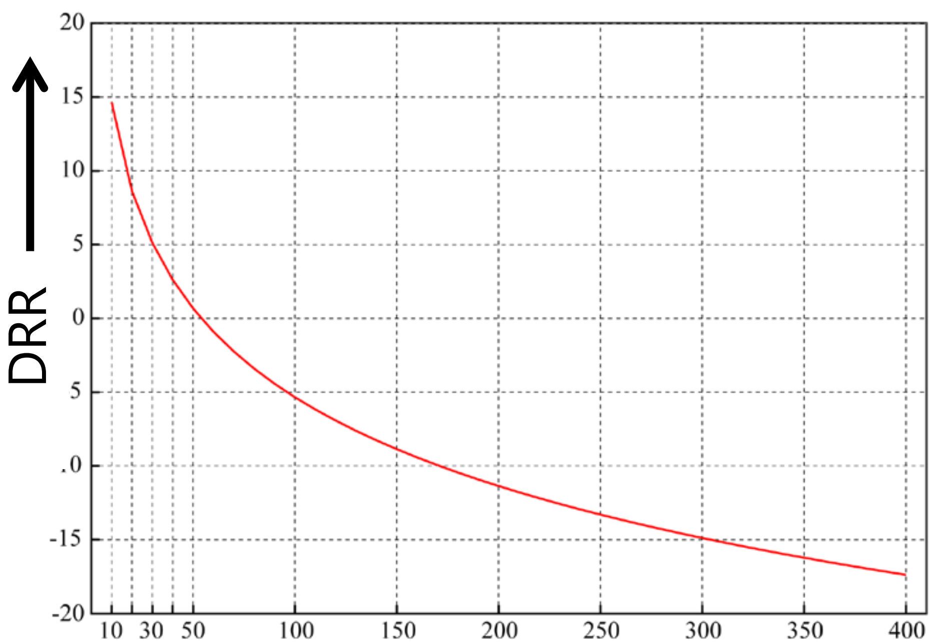
$$\mathrm{DRR} = \frac{\text{Energy of direct sound}}{\text{Energy of reverberant sound}} \left(= 10\log_{10}\left(\frac{\sum_{\omega} |H_{\mathrm{D}}(\omega)|^2}{\sum_{\omega} |H_{\mathrm{R}}(\omega)|^2}\right) \mathrm{dB} \right)$$



Received Signal | Direct Sound | Reverberation

Why estimating DRR?

- DRR can be used for various purposes. For example **source distance** can be calculated from the estimated DRR.



d : distance to source from mic

Theoretical DRR in perfectly diffuse sound field

$$\mathrm{DRR\ [dB]} = 10\log_{10}\left(\frac{S\alpha}{16\pi d^2}\right)$$

d : distance to source
S : surface area of wall
α: absorption coefficient

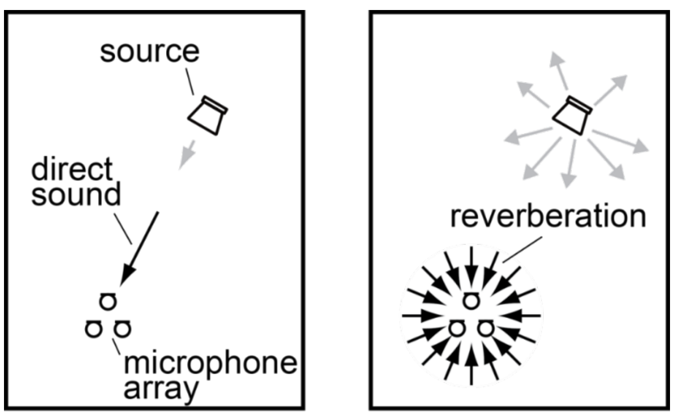
example:
DRR curve in diffuse sound field
(4×6×2.5 [m], α = 0.15)

Calculating DRR from PSD

- DRR can be approximated by the ratio of the PSD of the direct sound and reverberation.

$$\mathrm{DRR}\ \ [\mathrm{dB}] := 10\log_{10}\left(\frac{\sum_{\omega}|H_{\mathrm{D}}(\omega)|^2}{\sum_{\omega}|H_{\mathrm{R},\Omega}(\omega)|^2}\right)$$

$$\approx 10\log_{10}\left(\frac{\sum_{\omega}P_{\mathrm{D}}(\omega)}{\sum_{\omega}\int_{\Omega}P_{\mathrm{R},\Omega}(\omega)d\Omega}\right) = 10\log_{10}\left(\frac{\sum_{\omega}P_{\mathrm{D}}(\omega)}{4\pi\sum_{\omega}\overline{P}_{\mathrm{R}}(\omega)}\right)$$



(a) (b)

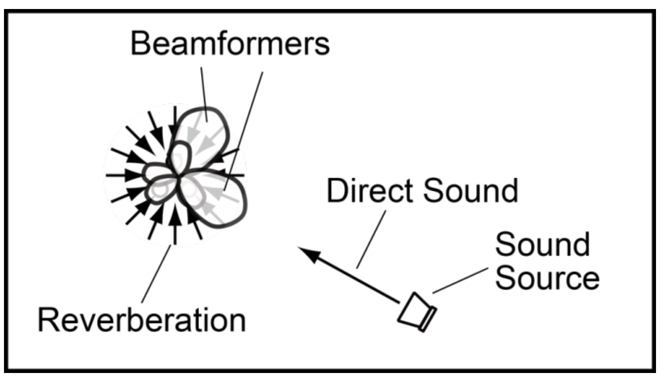
Assumed sound propagation model

- **Direct sound** arrives to a microphone without being reflected or diffracted by rigid materials (a)
- **Reverberation** arrives from every angle with uniform power distribution, i.e. Isotropic (b)

PSD estimation

- Use TWO beamformers which have different directivity patterns for PSD estimation in beamspace.
- Assuming the reverberation being spatially diffuse, same model as that for diffuse noise can be set.

$$\underbrace{\begin{bmatrix} P_{\mathrm{BF},1}(\omega) \\ P_{\mathrm{BF},2}(\omega) \end{bmatrix}}_{\mathbf{P}_{\mathrm{BF}}(\omega)} = \underbrace{\begin{bmatrix} G_{1,\Omega_\mathrm{D}}(\omega) & \int_\Omega G_{1,\Omega}(\omega) d\Omega \\ G_{2,\Omega_\mathrm{D}}(\omega) & \int_\Omega G_{2,\Omega}(\omega) d\Omega \end{bmatrix}}_{\mathbf{G}(\omega)} \underbrace{\begin{bmatrix} P_\mathrm{D}(\omega) \\ \overline{P}_\mathrm{R}(\omega) \end{bmatrix}}_{\mathbf{P}_\mathrm{S}(\omega)}$$



PSD of direct sound and reverberation are estimated.

$$\mathbf{P}_\mathrm{S}(\omega) = \mathbf{G}^{-1}(\omega)\mathbf{P}_{\mathrm{BF}}(\omega)$$

Evaluation using ACE Challenge corpus

- Performance evaluated using the ACE Challenge corpus recorded by a triangular microphone array (Mobile).
- Direction of arrival (DOA) of the direct sound was estimated by the steered beamformer based method with a delay-and-sum beamformer.
- Voice activity detection (VAD) was applied for extracting frames with reasonable amount of speech components used for the PSD estimation.

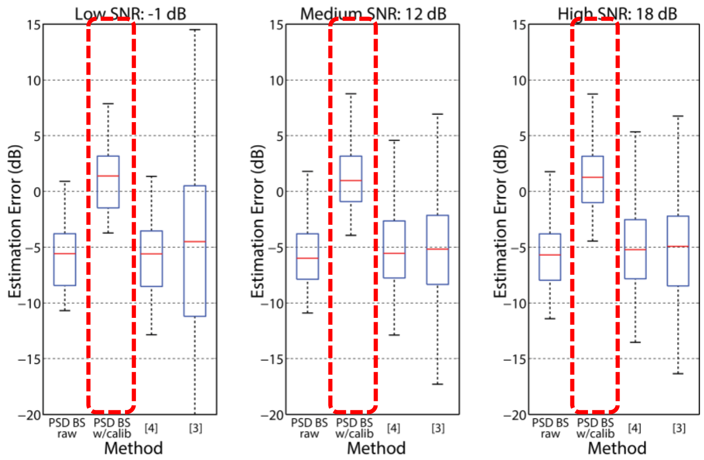
Parameter	Value	Parameter		Value
Sampling rate	16,000 Hz	Beamformer for PSD estimation	Type	Delay & Sum
Frame size	512 samples		Mainlobe angle	$\Omega_1=\{\theta_D,\varphi_D\}$ $\Omega_2=\{\theta_D+\pi/3,\varphi_D\}$
Frame shift	256 samples	Resolution of DOA estimation		Azimuth: $\pi/72$ Zenith: $\pi/60$

[Hioka et al. ACE Challenge Workshop 2015]

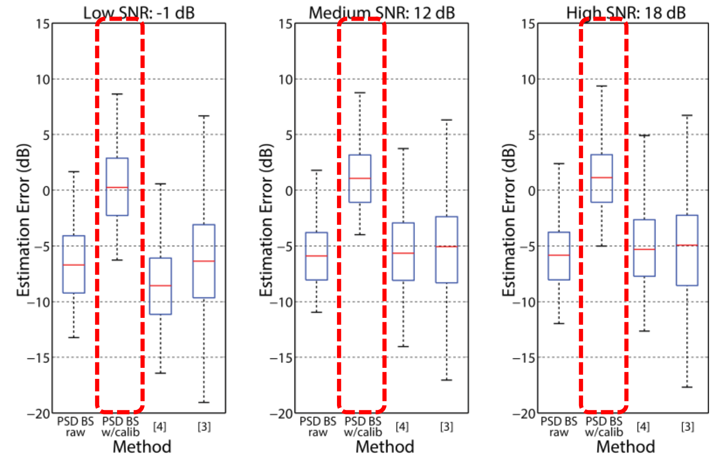
Evaluation using ACE Challenge corpus

- Estimation accuracy and distribution across rooms

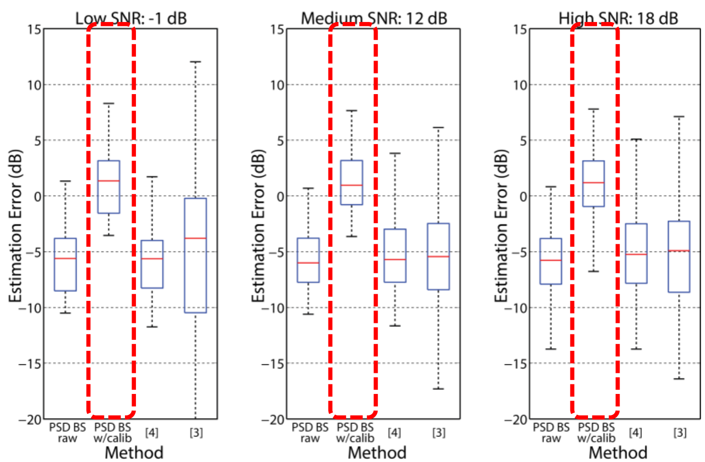
(a) Ambient



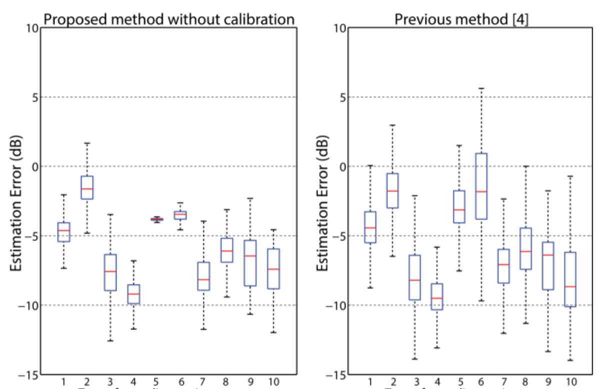
(b) Fan



(c) Babble



DRR estimation performance in different rooms



IV. Summary

Summary

- For realising **practically effective** acoustic signal enhancement, spectral manipulation using Wiener filter has been focused.
- **PSD estimation in beamspace** was developed for calculating the PSD of each sound source located in different angles/locations.
- Various applications that utilise PSD estimation in beamspace are introduced:
  - Directional sound source separation
  - Distance distinguishing microphone
  - "Zooming" microphone array
  - UAV recording
  - Blind acoustic characterisation – DRR estimation –

References

- Y. Hioka, K. Furuya, K. Kobayashi, S. Sakauchi, and Y. Haneda, "Angular region-wise speech enhancement for hands-free speakerphone," IEEE Trans. on Consumer Electronics, Vol. 58, No. 4, pp. 1403-1410, 2012.
- Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, and Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," IEEE Trans. on Audio, Speech, and Language Proc., Vol. 21, No. 6, pp. 1240-1250, 2013.
- K. Niwa, Y. Hioka, K. Furuya, and Y. Haneda, "Diffused sensing for sharp directive beamforming," IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, Issue 11, pp. 2346–2355, 2013.
- Y. Hioka and T. Betlehem, "Under-determined source separation based on power spectral density estimated using cylindrical mode beamforming" in Proc. WASPAA 2013, 2013.
- Y. Hioka and K. Niwa, "PSD estimation in beamspace for estimating direct-to-reverberant ratio from a reverberant speech signal," ACE Challenge Workshop, a satellite event of IEEE-WASPAA 2015.
- K. Niwa, Y. Koizumi, T. Kawase, K. Kobayashi, and Y. Hioka, "Pinpoint extraction of distant sound source based on DNN mapping from multiple beamforming outputs to prior SNR" in Proc. of ICASSP 2016, pp. 435-439, 2016.
- Y. Hioka, M. Kingan, G. Schmid, and K.A. Stol, "Speech enhancement using a microphone array mounted on an unmanned aerial vehicle," in Proc. of IWAENC 2016, 2016.
- K. Niwa and Y. Hioka, "PSD estimation in beamspace using property of M-matrix," in Proc. of IWAENC 2016, 2016.
- T. Kawase, K. Niwa, K. Kobayashi, and Y. Hioka, "Application of neural network to source PSD estimation for Wiener filter based array sound source enhancement," in Proc. of IWAENC 2016, 2016.
- K. Niwa, Y. Hioka, and K. Kobayashi, "Optimal Microphone Array Observation for Clear Recording of Distant Sound Sources," IEEE Transactions on Audio, Speech and Language Processing, Vol. 24, Issue 10, pp. 1785 - 1795, 2016.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

ENGINEERING

Thank you for your attention