NANYANG TECHNOLOGICAL UNIVERSITY

Spatial Audio Reproduction Using Primary Ambient Extraction

PhD candidate: **Jianjun HE**
Supervisor: Assoc Prof Gan Woon-Seng

Digital Signal Processing Laboratory
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

16th Feb, 2016

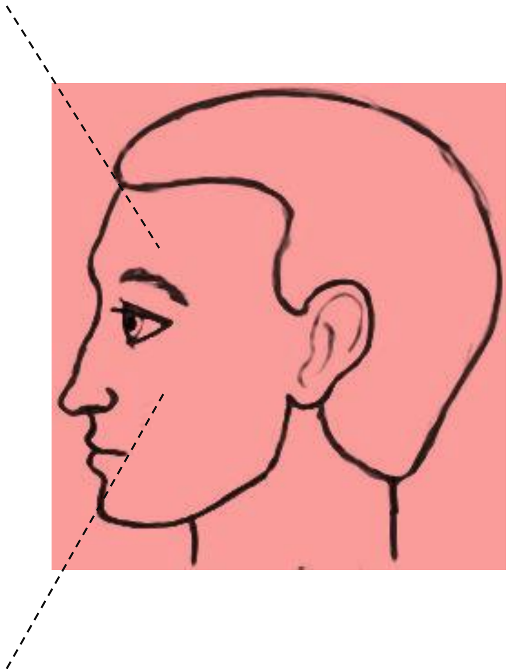
Dsp Digital Signal Processing Lab

Outline

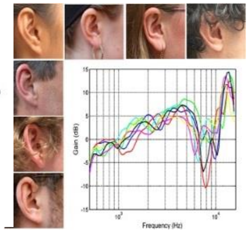
I. Background of Spatial Audio Reproduction
II. Overview of Primary Ambient Extraction (PAE)
III. My Contributions
   1. Linear Estimation based PAE
   2. Ambient Spectrum Estimation based PAE
   3. Time Shifting based PAE
   4. Multi-Source and Multichannel based PAE
   5. Natural Sound Rendering for Headphones
IV. Conclusions and Future work

Hearing is Believing

Immersive visual experience



1. What/where is the sound?
2. How is the sound played?
3. How is the sound heard?



Immersive listening experience ⇐ **Sound content processing** + Personal listening cues

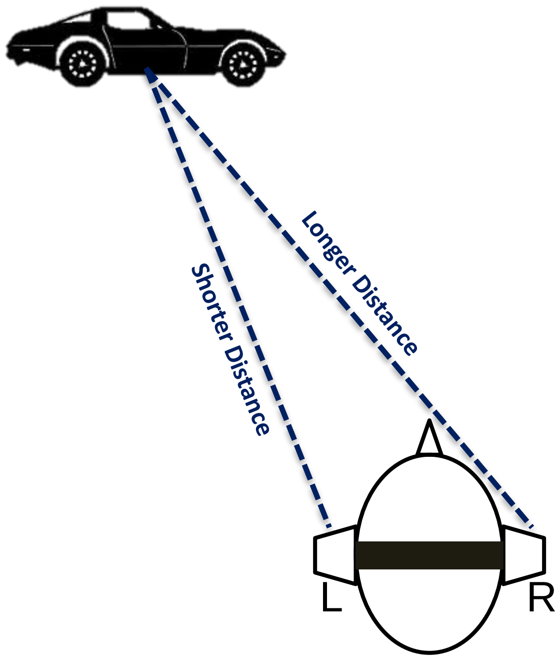
Spatial hearing of sound source

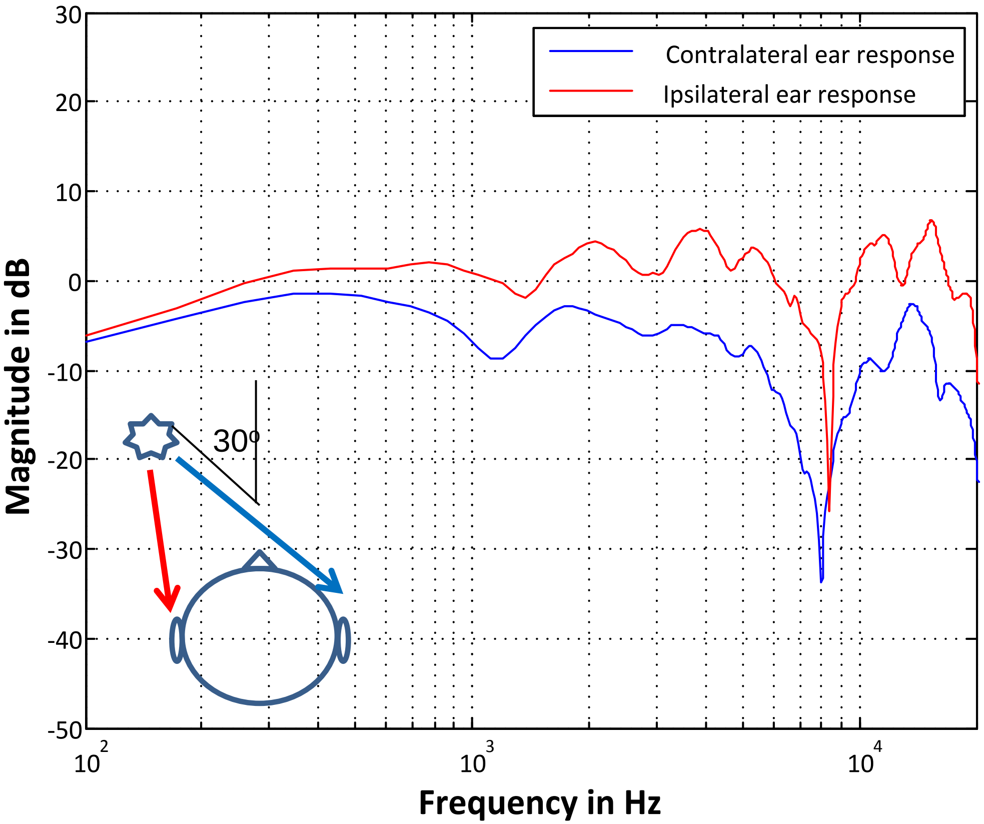
Azimuth, Elevation, Distance

Head Related Transfer Functions (HRTF)

ITD: Interaural Time Difference
ILD: Interaural Level Difference
SC : Spectral Cues

Head Related Transfer Functions (**HRTF**)





HRTF of KEMAR dummy head for an angle of 30 degree azimuth*

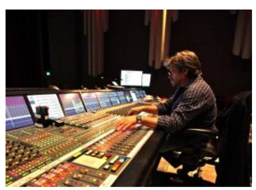
Spatial hearing of sound scene




Immersive Audio
"being there"

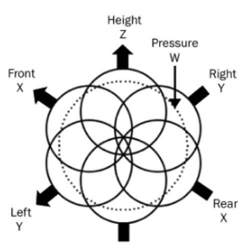
Source
Audio content

- Channel-based



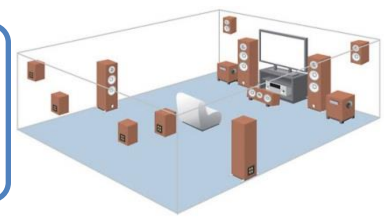
- Object-based

- Transform-domain-based



Medium
Playback system

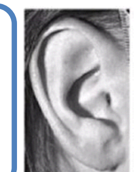
- Loudspeakers



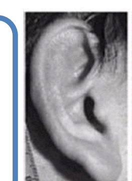
- Headphones

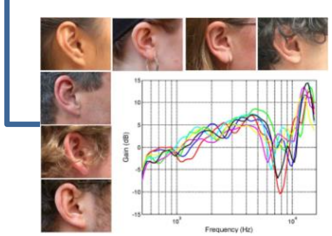
Receiver
Human ear

- Ear A

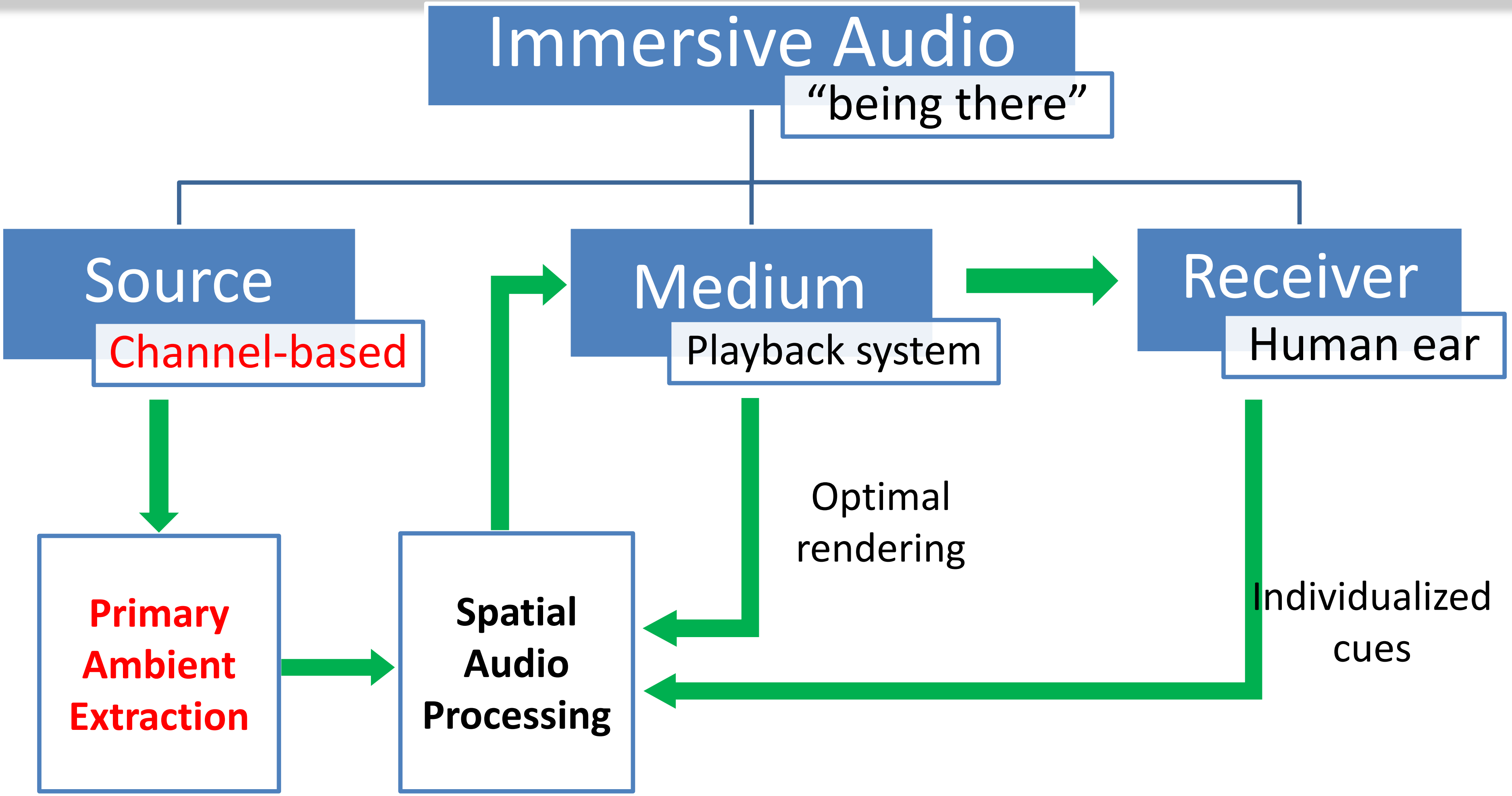


- Ear B





Achieving consistency in source and medium in spatial audio

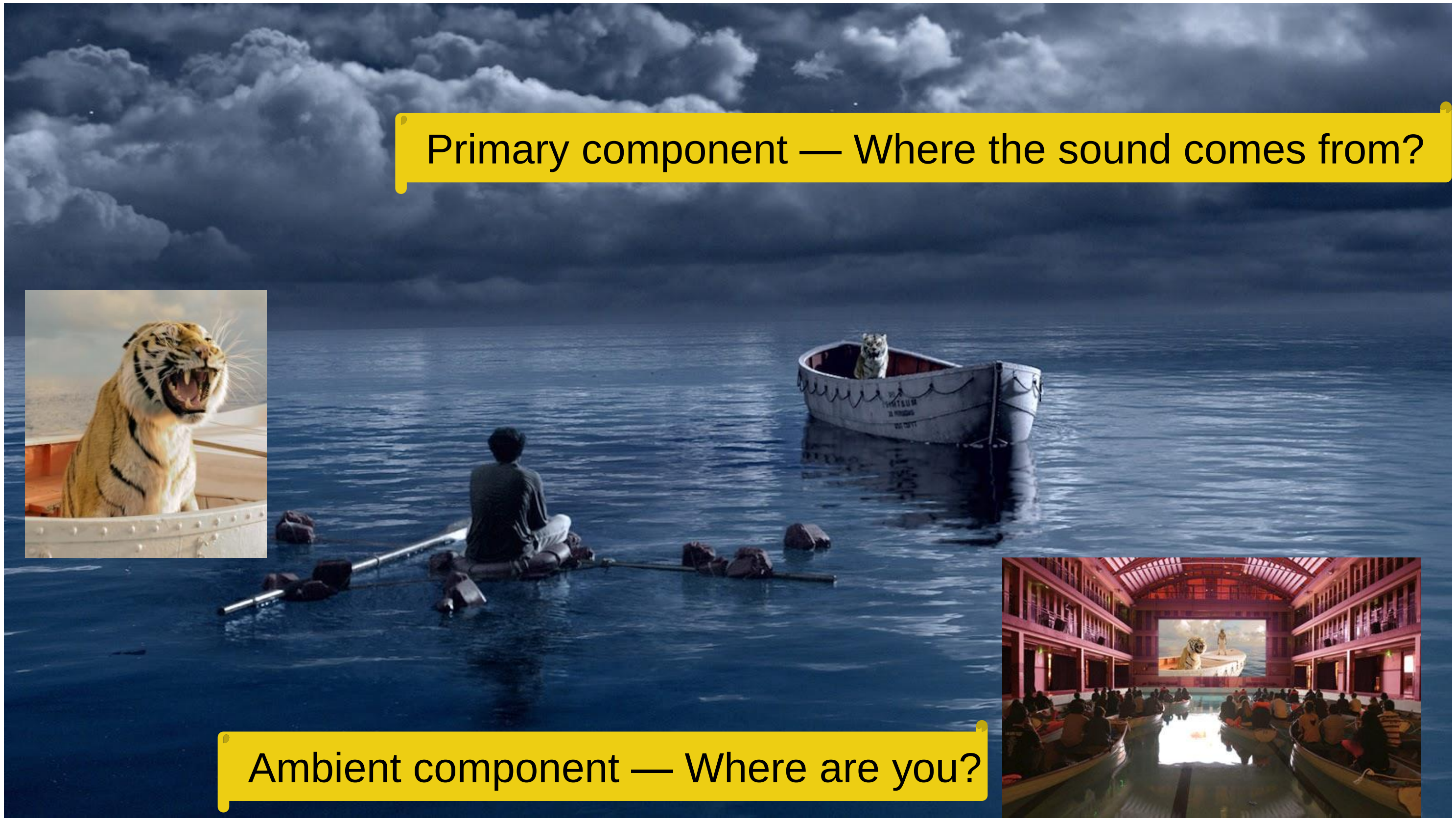


Essentially, PAE serves as a front-end to facilitate **flexible**, **efficient**, and **immersive** spatial audio reproduction.

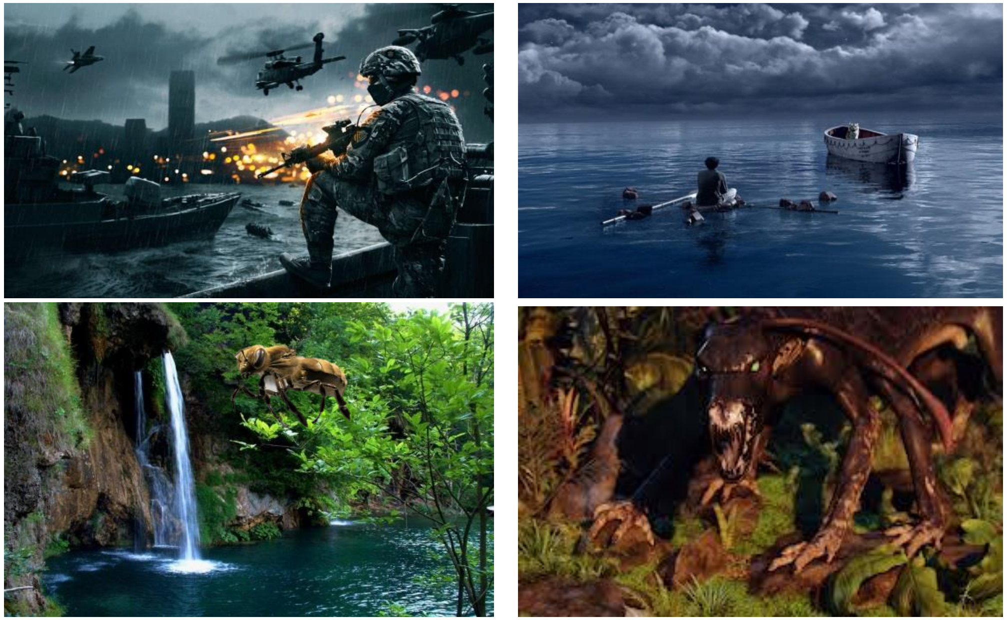
What are primary and ambient components?

Main characteristics

- Primary components are directional
- Ambient components are diffuse



What does PAE do?



Sound scenes in movies and games often consist of:

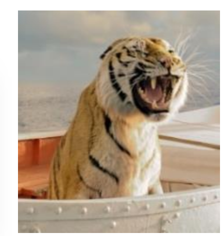
directional sound objects

and **background ambience**, e.g.,

- Circling aircraft in the rain;
- Flying bee at the waterfall;
- Soaring tiger on the sea;
- Attacking wolves in the forest.

Primary Ambient Extraction

Directional sound

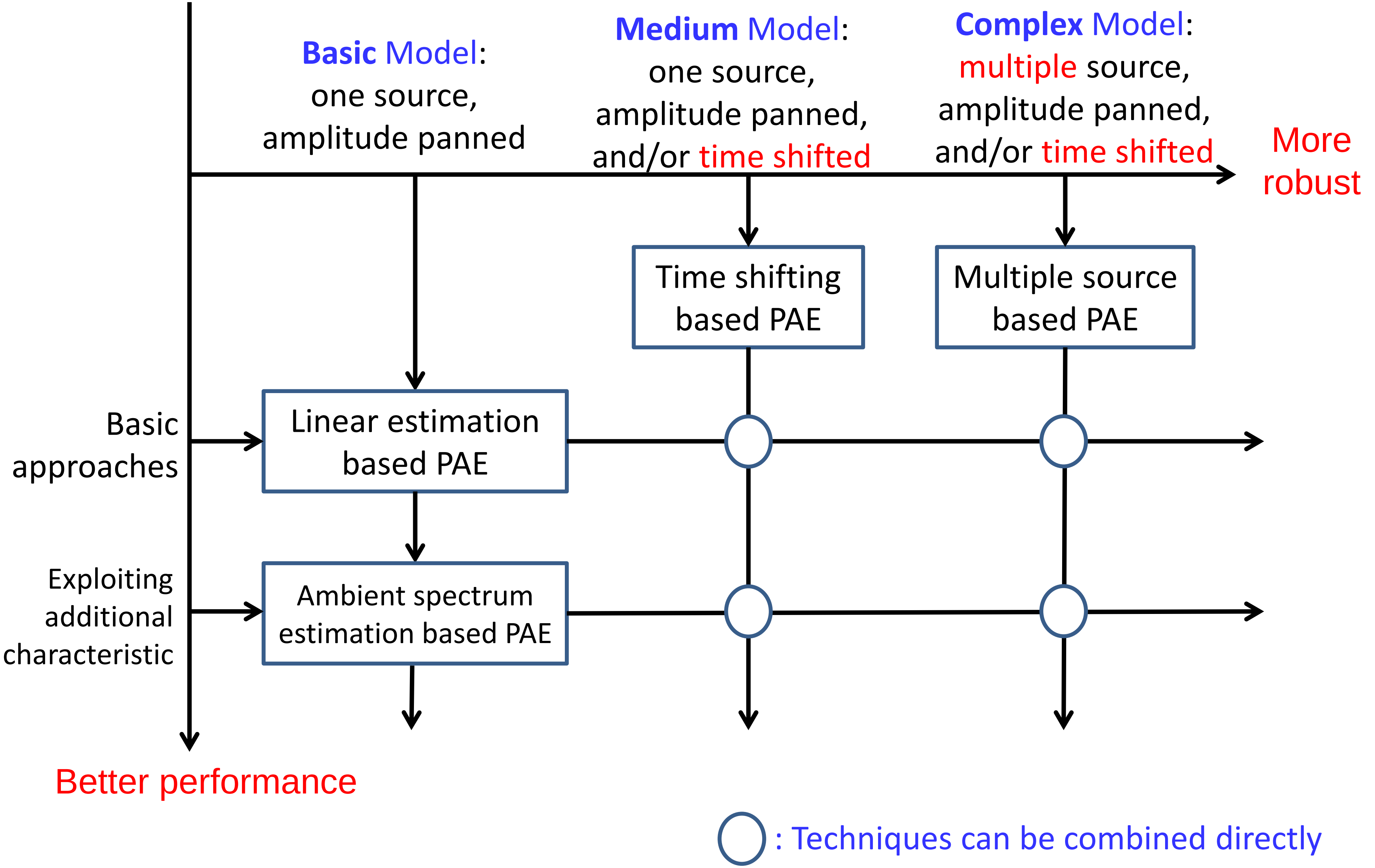


Background ambience

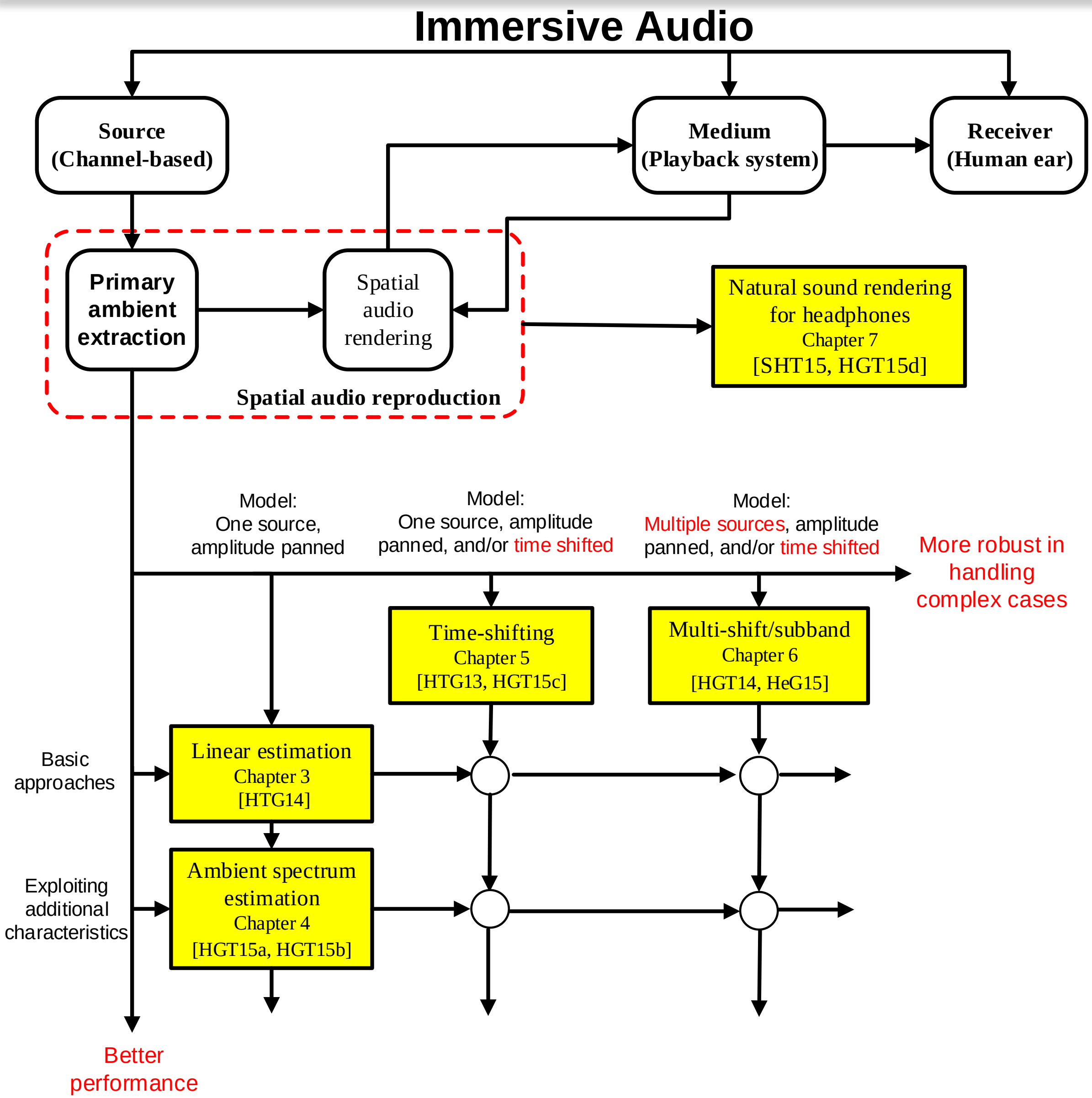
A brief literature of PAE

No. of channels	Complexity of audio scenes		
	Basic (amplitude panned single source)	**Medium (single source)**	**Complex (multiple sources)**
Stereo (2)	**Time-frequency masking:** [AvJ02], [AvJ04], [MGJ07], [Pul07] **PCA**:[IrA02], [BVM06], [MGJ07], [GoJ07b], [BaS07], [God08], [JHS10], [BJP12], [TaG12], [TGC12], [LBP14] **Least-squares**:[Fal06], [Fal07], [JPL10], [FaB11], [UhH15] **Linear estimation**: [HTG14] **Ambient spectrum estimation**: [HGT15a], [HGT15b] **Others**: [BrS08], [MeF10], [Har11]	**LMS**: [UsB07] **Shifted PCA**: [HTG13] **Time-shifting**: [HGT15c]	**PCA**: [DHT12], [HGT14], [HeG15]
Multi-channel	**PCA**: [GoJ07b] **Others**: [GoJ07a], [WaF11], [TGC12], [CCK14]	**ICA & time-frequency masking**: [SAM06] **Pairwise correlations**: [TSW12] **Pairing**: [HeG15b] **Others**: [StM15]	**ICA**: [HKO04]
Single	**NMF**: [UWH07]	**Neural network**: [UhP08]	

My contributions



My contributions



I: Investigation of linear estimation based PAE under the basic signal model.

II: Improving PAE performance for strong ambient power cases using ambient spectrum estimation techniques.

III: Employing time-shifting techniques for PAE with partially correlated primary components

IV: Adaptation of conventional PAE approaches to deal with multiple sources.

V: Applying PAE in natural sound rendering headphone systems.

My contributions in publications

Contribution	Publications
I: Investigation of linear estimation based PAE under the basic signal model.	[J1] **J. He**, E. L. Tan, and W. S. Gan, "Linear estimation based primary-ambient extraction for stereo audio signals," ***IEEE/ACM Trans. Audio, Speech, Lang. Process.***, vol. 22, no. 2, pp. 505-517, Feb. 2014.
II: Improving PAE performance for strong ambient power cases using ambient spectrum estimation techniques.	[J2] **J. He**, W. S. Gan, and E. L. Tan, "Primary-ambient extraction using ambient phase estimation with a sparsity constraint," ***IEEE Signal Process. Letters***, vol. 22, no. 8, pp. 1127-1131, Aug. 2015. [J3] **J. He**, E. L. Tan, and W. S. Gan, "Primary-ambient extraction using ambient spectrum estimation for immersive spatial audio reproduction," ***IEEE/ACM Trans. Audio, Speech, Lang. Process.***, vol. 23, no. 9, pp. 1430-1443, Sept. 2015.
III: Employing time-shifting techniques for PAE with partially correlated primary components	[J4] **J. He**, W. S. Gan, and E. L. Tan, "Time-shifting based primary-ambient extraction for spatial audio reproduction," ***IEEE/ACM Trans. Audio, Speech, Lang. Process.***, vol. 23, no. 10, pp. 1576-1588, Oct. 2015. [C1] **J. He**, E. L. Tan, and W. S. Gan, "Time-shifted principal component analysis based cue extraction for stereo audio signals," in *Proc. ICASSP*, Vancouver, Canada, 2013, pp. 266-270.
IV: Adaptation of conventional PAE approaches to deal with multiple sources.	[C2] **J. He**, W. S. Gan, and E. L. Tan, "A study on the frequency-domain primary-ambient extraction for stereo audio signals," in *Proc. ICASSP*, Florence, Italy, 2014, pp. 2892-2896. [C3] **J. He**, and W. S. Gan, "Multi-shift principal component analysis based primary component extraction for spatial audio reproduction," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015, pp. 350-354.
V: Applying PAE in natural sound rendering headphone systems.	[J5] K. Sunder, **J. He**, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones: Integration of signal processing techniques," ***IEEE Signal Process. Magazine***, vol. 32, no. 2, Mar 2015, pp. 100-113.

Other related publications

- [C4] **J. He**, W. S. Gan, and E. L. Tan, "On the preprocessing and postprocessing of HRTF individualization based on sparse representation of anthropometry features," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015, pp. 639-643.
- [C5] **J. He**, and W. S. Gan, "Applying primary ambient extraction for immersive spatial audio reproduction," *2015 Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference (invited)*, Hong Kong, Dec. 2015.
- [C6] **J. He**, R. Ranjan, and W. S. Gan, "Fast continuous HRTF acquisition with unconstrained movements of human subjects," in *Proc. ICASSP*, Shanghai, China, Mar. 2016, pp.

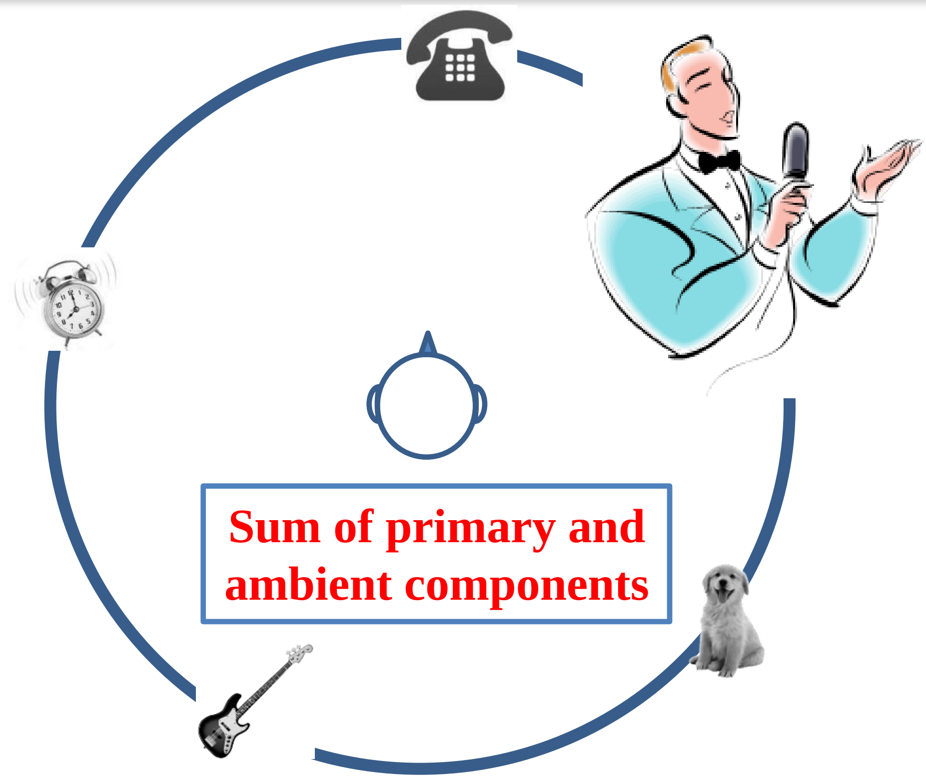
- **[Tutorial]** W. S. Gan, and **J. He**, "Assisted listening for headphones and hearing aids: signal processing techniques," Tutorial at *APSIPA ASC 2015*, Hong Kong, Dec. 2015.

- **[Show & Tell]** D. H. Nguyen, **J. He**, K. K. Phyo, and W. S. Gan, "Real-time audio signal processing platform for natural 3D sound rendering," Show & Tell at *ICASSP 2016*, Shanghai, China, Mar. 2016.

- [J6]] **J. He**, R. Ranjan, W. S. Gan, and K. Sunder, "Scalable data reusing normalized LMS for acoustic system identification with short duration signals," IEEE Sig. Process. Letters, under review.

Objective of PAE

to extract the primary and ambient components from *M* mixtures



Mixtures = primary component + ambient component

$$x_m(n) = p_m(n) + a_m(n)$$

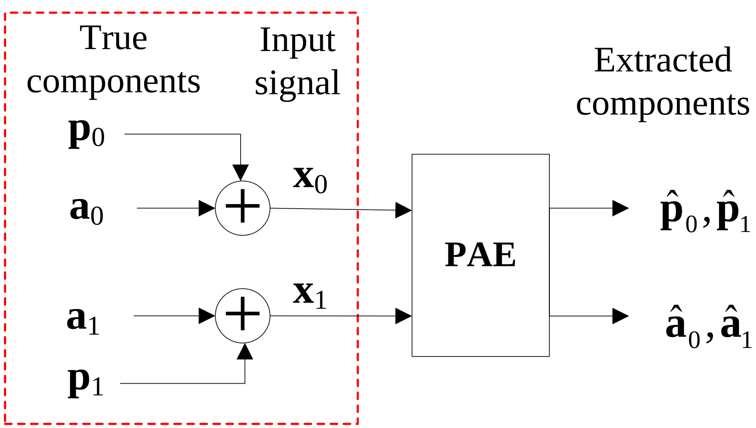
Definitions with Stereo Signal Model

Signal = Primary + Ambient

$$\mathbf{x}_0 = \mathbf{p}_0 + \mathbf{a}_0$$

$$\mathbf{x}_1 = \mathbf{p}_1 + \mathbf{a}_1$$

Stereo signal model



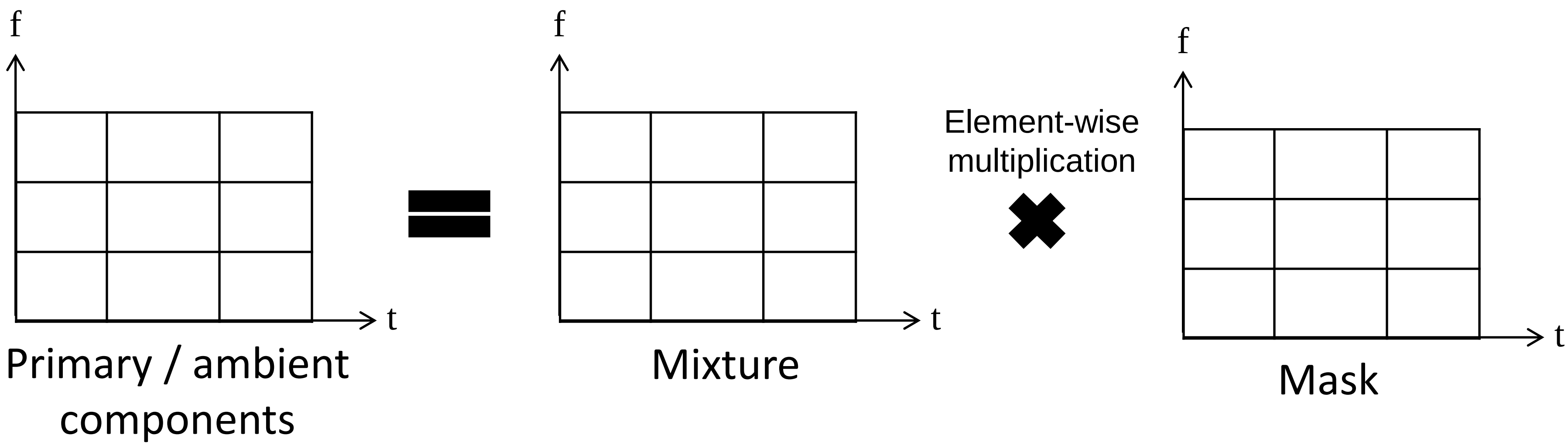
Assumptions

Primary components highly correlated	$\mathbf{p}_1 = k\mathbf{p}_0$
Ambient components uncorrelated	$\mathbf{a}_0 \perp \mathbf{a}_1$
Primary ambient components uncorrelated	$\mathbf{p}_i \perp \mathbf{a}_j$ $\forall i, j \in \{0,1\}$
Ambient power balanced	$P_{\mathbf{a}_0} = P_{\mathbf{a}_1}$

Primary panning factor PPF: $k = \frac{\mathbf{p}_1}{\mathbf{p}_0}$

Primary power ratio PPR: $\gamma = \frac{\text{Total primary power}}{\text{Total signal power}}$

PAE: Time-Frequency Masking



Mask can be constructed using

- **Inter-channel coherence [Avendano and Jot, 2004]**
- **Pairwise correlation [Thompson et al., 2012]**
- **Equal level of ambience [Merimaa et al., 2007]**
- **Diffuseness [Pulkki, 2007]**

Linear Estimation framework in PAE

$$\begin{bmatrix} \hat{\mathbf{p}}_0^T \\ \hat{\mathbf{p}}_1^T \\ \hat{\mathbf{a}}_0^T \\ \hat{\mathbf{a}}_1^T \end{bmatrix} = \begin{bmatrix} w_{\mathrm{P0,0}} & w_{\mathrm{P0,1}} \\ w_{\mathrm{P1,0}} & w_{\mathrm{P1,1}} \\ w_{\mathrm{A0,0}} & w_{\mathrm{A0,1}} \\ w_{\mathrm{A1,0}} & w_{\mathrm{A1,1}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0^T \\ \mathbf{x}_1^T \end{bmatrix} = \mathbf{W} \begin{bmatrix} \mathbf{x}_0^T \\ \mathbf{x}_1^T \end{bmatrix}$$

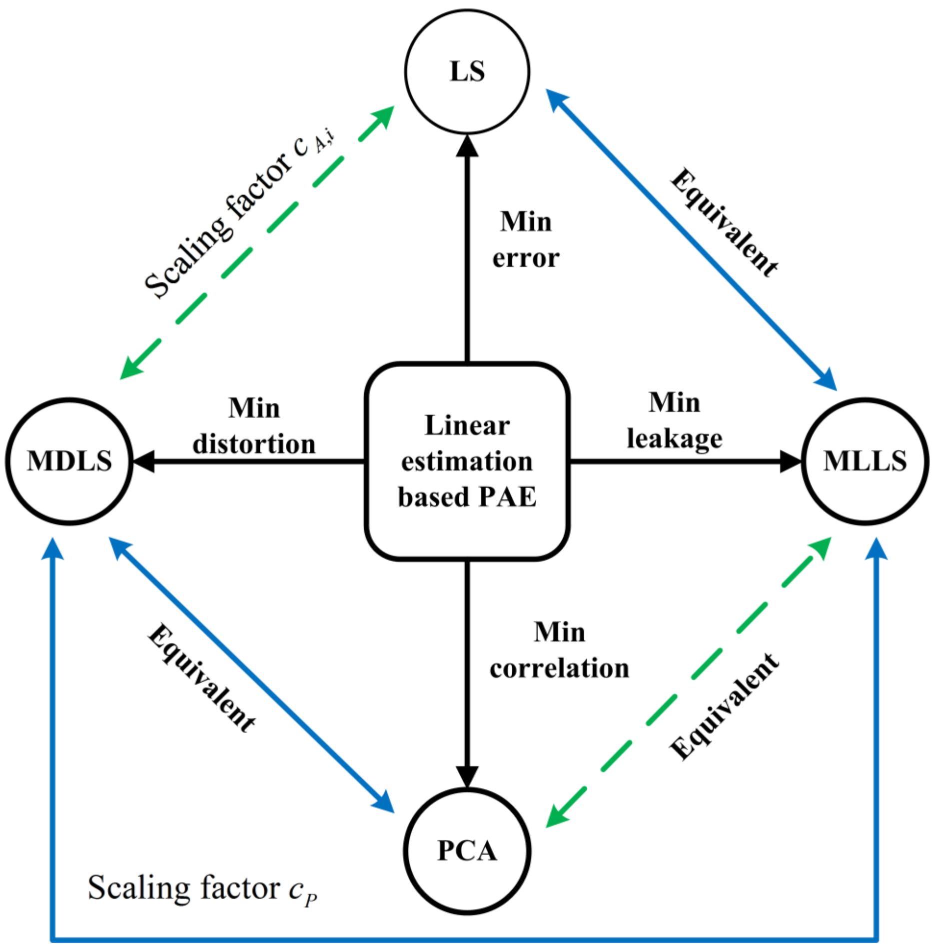
Performance measures

Extraction Accuracy	$\mathbf{e}_{\mathrm{P}} = \left(w_{\mathrm{P}0,0} + k w_{\mathrm{P}0,1} - 1\right)\mathbf{p}_0 + 0 + \left(w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1\right)$ $\mathbf{e}_{\mathrm{A}} = \left(w_{\mathrm{A}0,0} - 1\right)\mathbf{a}_0 + w_{\mathrm{A}0,1}\mathbf{a}_1 + \left(w_{\mathrm{A}0,0}\mathbf{p}_0 + w_{\mathrm{A}0,1}\mathbf{p}_1\right)$ $\frac{P_{\text{Error}}}{P_{\text{True signal}}} \approx \frac{P_{\text{Distortion}}}{P_{\text{True signal}}} + \frac{P_{\text{Interference}}}{P_{\text{True signal}}} + \frac{P_{\text{Leakage}}}{P_{\text{True signal}}}$ $\mathrm{ESR} \approx \mathrm{DSR} + \mathrm{ISR} + \mathrm{LSR}$
Spatial Accuracy	**ICC, ICLD, ICTD**(only for primary)

ESR: Error-to-signal ratio
DSR: Distortion-to-signal ratio
ISR: Interference-to-signal ratio
LSR: Leakage-to-signal ratio

ICC : Inter-channel cross-correlation coefficient
ICLD: Inter-channel level difference
ICTD: Inter-channel time difference

PAE: Linear Estimation



$$\begin{bmatrix} \hat{p}_0(n) \\ \hat{p}_1(n) \\ \hat{a}_0(n) \\ \hat{a}_1(n) \end{bmatrix} = \begin{bmatrix} w_{\mathrm{P}0,0} & w_{\mathrm{P}0,1} \\ w_{\mathrm{P}1,0} & w_{\mathrm{P}1,1} \\ w_{\mathrm{A}0,0} & w_{\mathrm{A}0,1} \\ w_{\mathrm{A}1,0} & w_{\mathrm{A}1,1} \end{bmatrix} \begin{bmatrix} x_0(n) \\ x_1(n) \end{bmatrix}$$

Objectives and relationships of four linear estimation based PAE approaches.

- **Blue** solid lines represent the relationships in the **primary** component;
- **Green** dotted lines represent the relationships in the **ambient** component.
- **MLLS**: minimum leakage LS
- **MDLS**: minimum distortion LS

Performance of the four PAE approaches

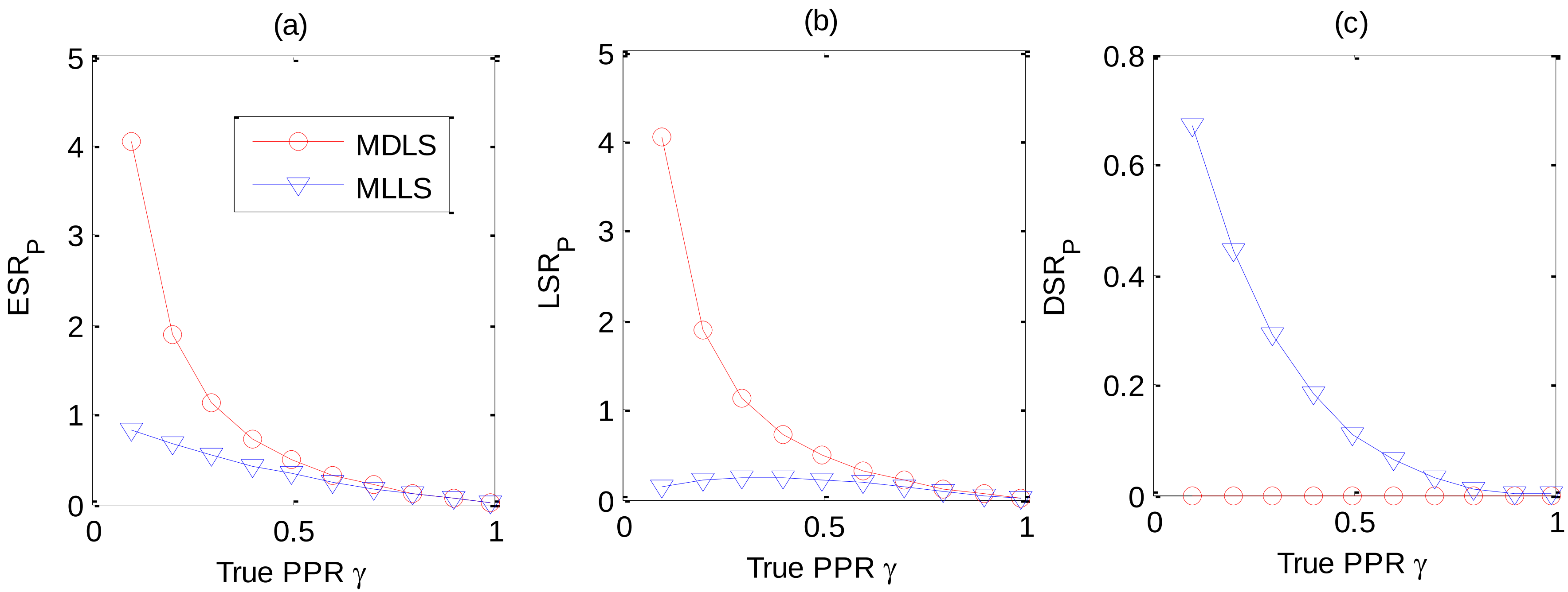
$k : PPF$

$\gamma : PPR$

Performance Measure	Primary component		Ambient component		
	MDLS/PCA	MLLS/LS	MLLS/PCA	LS	MDLS
ESR	$\frac{1-\gamma}{2\gamma}$	$\frac{1-\gamma}{1+\gamma}$	$\frac{1}{1+k^2}$	$\frac{1}{1+k^2}\frac{2\gamma}{1+\gamma}$	$\frac{2\gamma}{1+k^2+(k^2-1)\gamma}$
LSR	$\frac{1-\gamma}{2\gamma}$	$\frac{1-\gamma}{2\gamma}\left(\frac{2\gamma}{1+\gamma}\right)^2$	0	$\frac{1}{1+k^2}\frac{2\gamma(1-\gamma)}{(1+\gamma)^2}$	$\frac{(1+k^2)(1-\gamma)2\gamma}{\left[(1+k^2)(1+\gamma)-2\gamma\right]^2}$
DSR	0	$\left(\frac{1-\gamma}{1+\gamma}\right)^2$	$\left(\frac{1}{1+k^2}\right)^2$	$\left(\frac{1}{1+k^2}\frac{2\gamma}{1+\gamma}\right)^2$	0
ISR	0		$\left(\frac{k}{1+k^2}\right)^2$	$\left(\frac{k}{1+k^2}\frac{2\gamma}{1+\gamma}\right)^2$	$\left[\frac{2k\gamma}{(1+k^2)(1+\gamma)-2\gamma}\right]^2$
ICC(ICTD)	1(0)		1	$\frac{2k\gamma}{\sqrt{(1+k^2)^2-(1-k^2)^2\gamma^2}}$	
ICLD	k^2		$\frac{1}{k^2}$	$\frac{1}{k^2}\frac{1+\gamma+k^2(1-\gamma)}{1+\gamma+\frac{1}{k^2}(1-\gamma)}$	$\frac{1}{k^2}\frac{1-\gamma+k^2(1+\gamma)}{1-\gamma+\frac{1}{k^2}(1+\gamma)}$

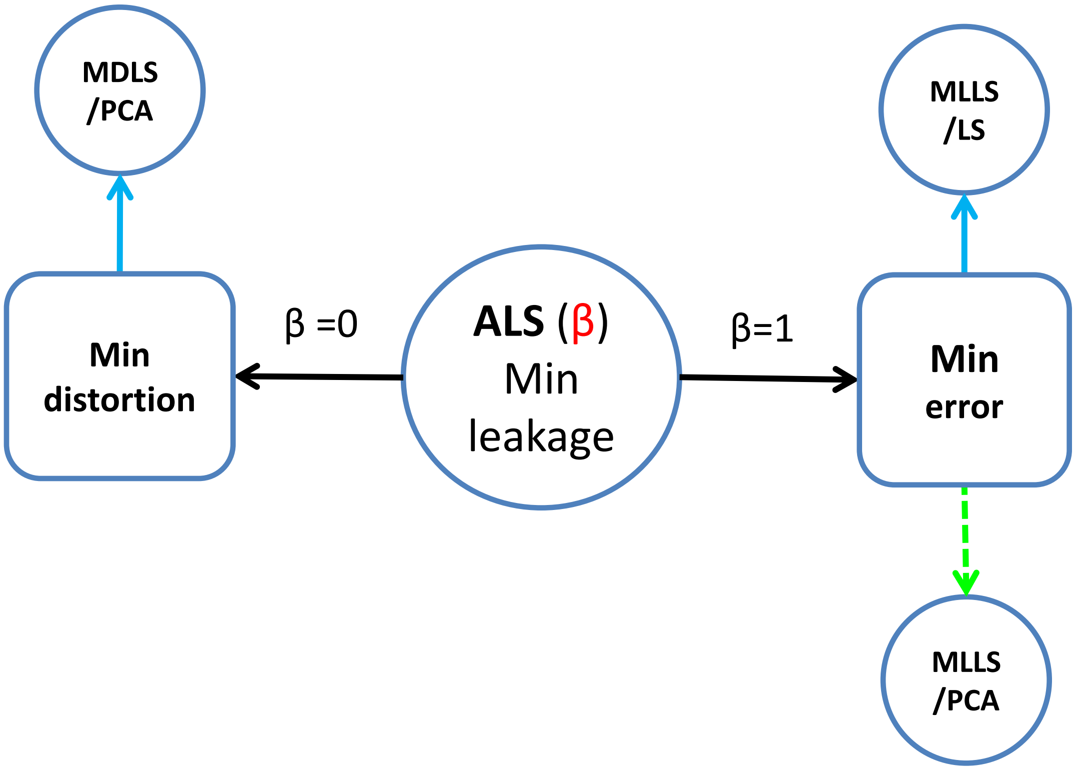
Theoretical results are verified in our experiments!

Experimental results– Primary extraction



Performance of MDLS (or PCA) and MLLS (or LS) in primary extraction
(a) ESR, (b) LSR, (c) DSR.

PAE using Adjustable Least-squares (ALS)



Adjustable factor $\beta \in [0,1]$

$$\mathbf{W}_{\mathrm{ALS}} = \begin{bmatrix} \frac{1}{1+k^2}\left(1-\beta\frac{1-\gamma}{1+\gamma}\right) & \frac{k}{1+k^2}\left(1-\beta\frac{1-\gamma}{1+\gamma}\right) \\ \frac{k}{1+k^2}\left(1-\beta\frac{1-\gamma}{1+\gamma}\right) & \frac{k^2}{1+k^2}\left(1-\beta\frac{1-\gamma}{1+\gamma}\right) \\ 1-\beta\frac{1}{1+k^2} & -\frac{1}{k}\left(1-\beta\frac{1}{1+k^2}\right) \\ -k\left(1-\beta\frac{k^2}{1+k^2}\right) & 1-\beta\frac{k^2}{1+k^2} \end{bmatrix}$$

$$\mathrm{ESR}_{\mathrm{P}} = \frac{1-\gamma}{2\gamma} + \beta(\beta-2)\frac{(1-\gamma)^2}{2\gamma(1+\gamma)}, \quad \mathrm{DSR}_{\mathrm{P}} = \beta^2\left(\frac{1-\gamma}{1+\gamma}\right)^2, \quad \mathrm{LeSR}_{\mathrm{P}} = \frac{1-\gamma}{1+\gamma},$$

$$\mathrm{ESR}_{\mathrm{A}} = \frac{1}{k^2} + \beta(\beta-2)\frac{1}{k^2(k^2+1)}, \quad \mathrm{DSR}_{\mathrm{A}} = \beta^2\left(\frac{1}{1+k^2}\right)^2, \quad \mathrm{LeSR}_{\mathrm{A}} = 0.$$

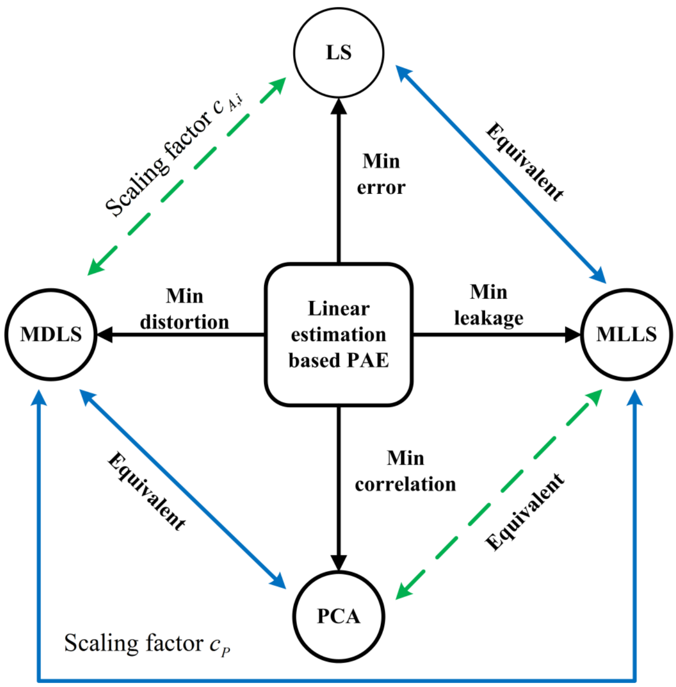
Blue solid lines: primary component; Green dotted lines: ambient component.

Recommendations of Linear Estimation based PAE

Approach	Strengths	Weaknesses	Recommendations
PCA	• No distortion in the extracted primary component; • No primary leakage in the extracted ambient component; • Primary and ambient components are uncorrelated;	• Ambient component severely panned;	Spatial audio coding and interactive audio in gaming, where the primary component is more important than the ambient component.
LS	• Minimum MSE in the extracted primary and ambient components;	• Severe primary leakage in the extracted ambient component;	Applications in which both the primary and ambient components are extracted, processed, and finally mixed together.
MLLS	• Minimum leakage in the extracted primary and ambient components; • Primary and ambient components are uncorrelated;	• Ambient component severely panned;	Spatial audio enhancement systems, and applications in which different rendering or playback techniques are employed on the extracted primary and ambient components.
MDLS	• No distortion in the extracted primary and ambient components;	• Severe interference and primary leakage in the extracted ambient component;	High-fidelity applications in which timbre is of high importance.
ALS	• Performance adjustable;	• Need to adjust the value of the adjustable factor;	For applications without explicit requirements.

Contributions on linear estimation based PAE

1. Formulated the linear estimation framework for PAE.
2. Introduced two groups of performance measures.
   - Extraction accuracy: ESR, DSR, ISR, LSR
   - Spatial accuracy : ICC, ICTD, ICLD
3. Proposed MLLS, MDLS, ALS and compared them with PCA and LS in PAE.
4. Different approaches are recommended in different spatial audio applications.



[J1] J. He, E. L. Tan and W. S. Gan, "Linear estimation based primary-ambient extraction for stereo audio signals," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 2, pp. 505-517, Feb. 2014.

Observation

Ambient components are decorrelated using:

- A small delay
- All-pass filtering with random phase
- Artificial reverberation, and binaural reverberation

Magnitude of the ambient components are usually kept the same.

Element-wise multiplication

Complex with unit magnitude

Ambient components

$$\mathbf{A}_0 = |\mathbf{A}_0| \odot \mathbf{W}_0$$
$$\mathbf{A}_1 = |\mathbf{A}_1| \odot \mathbf{W}_1$$

$$W_0(n,l) = e^{j\theta_0(n,l)}$$
$$W_1(n,l) = e^{j\theta_1(n,l)}$$

Ambient magnitude

$$|\mathbf{A}_0| = |\mathbf{A}_1| = |\mathbf{A}|$$

Ambient phase

From PAE to APE: ambient phase estimation

X: Mixed signal
P: Primary components
k: Primary panning factor

PAE → APE

PAE

$$\mathbf{X}_0 = \mathbf{P}_0 + \mathbf{A}_0$$
$$\mathbf{X}_1 = \mathbf{P}_1 + \mathbf{A}_1$$

with

$$\mathbf{P}_1 = k\mathbf{P}_0,\ \mathbf{A}_0 \perp \mathbf{A}_1$$
$$\mathbf{P}_i \perp \mathbf{A}_j, \forall i, j \in \{0,1\}$$
$$P_{\mathbf{P}_1} = k^2 P_{\mathbf{P}_0},\quad P_{\mathbf{A}_1} = P_{\mathbf{A}_0}$$

APE

$$\boxed{|\mathbf{A}| = (\mathbf{X}_1 - k\mathbf{X}_0)./(\mathbf{W}_1 - k\mathbf{W}_0)}$$

and

$$\mathbf{A}_c = (\mathbf{X}_1 - k\mathbf{X}_0)./(\mathbf{W}_1 - k\mathbf{W}_0) \odot \mathbf{W}_c,$$
$$\mathbf{P}_c = \mathbf{X}_c - (\mathbf{X}_1 - k\mathbf{X}_0)./(\mathbf{W}_1 - k\mathbf{W}_0) \odot \mathbf{W}_c$$
$$\forall \text{channel index } c \in \{0,1\}.$$

→ Find $\mathbf{W}_0, \mathbf{W}_1$ → Find $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ → $\boxed{\text{Find } \boldsymbol{\theta}_1}$

$$\boldsymbol{\theta}_0 = \boldsymbol{\theta} + \arcsin\left[k^{-1}\sin(\boldsymbol{\theta} - \boldsymbol{\theta}_1)\right] + \pi$$
$$\text{where}\quad \boldsymbol{\theta} = \angle(\mathbf{X}_1 - k\mathbf{X}_0)$$

From PAE to APE: ambient phase estimation

PAE → APE

$$\mathbf{X}_0 = \mathbf{P}_0 + \mathbf{A}_0$$
$$\mathbf{X}_1 = \mathbf{P}_1 + \mathbf{A}_1$$

$$\boxed{|\mathbf{A}| = (\mathbf{X}_1 - k\mathbf{X}_0)./(\mathbf{W}_1 - k\mathbf{W}_0)}$$

Is this the only way?

For each time-frequency bin

$$\boxed{\begin{aligned} kX_0 &= P_1 + kA_0 \\ X_1 &= P_1 + A_1 \end{aligned}}$$

→ Complex spectrum shown in complex plane?

From PAE to AME: ambient magnitude estimation

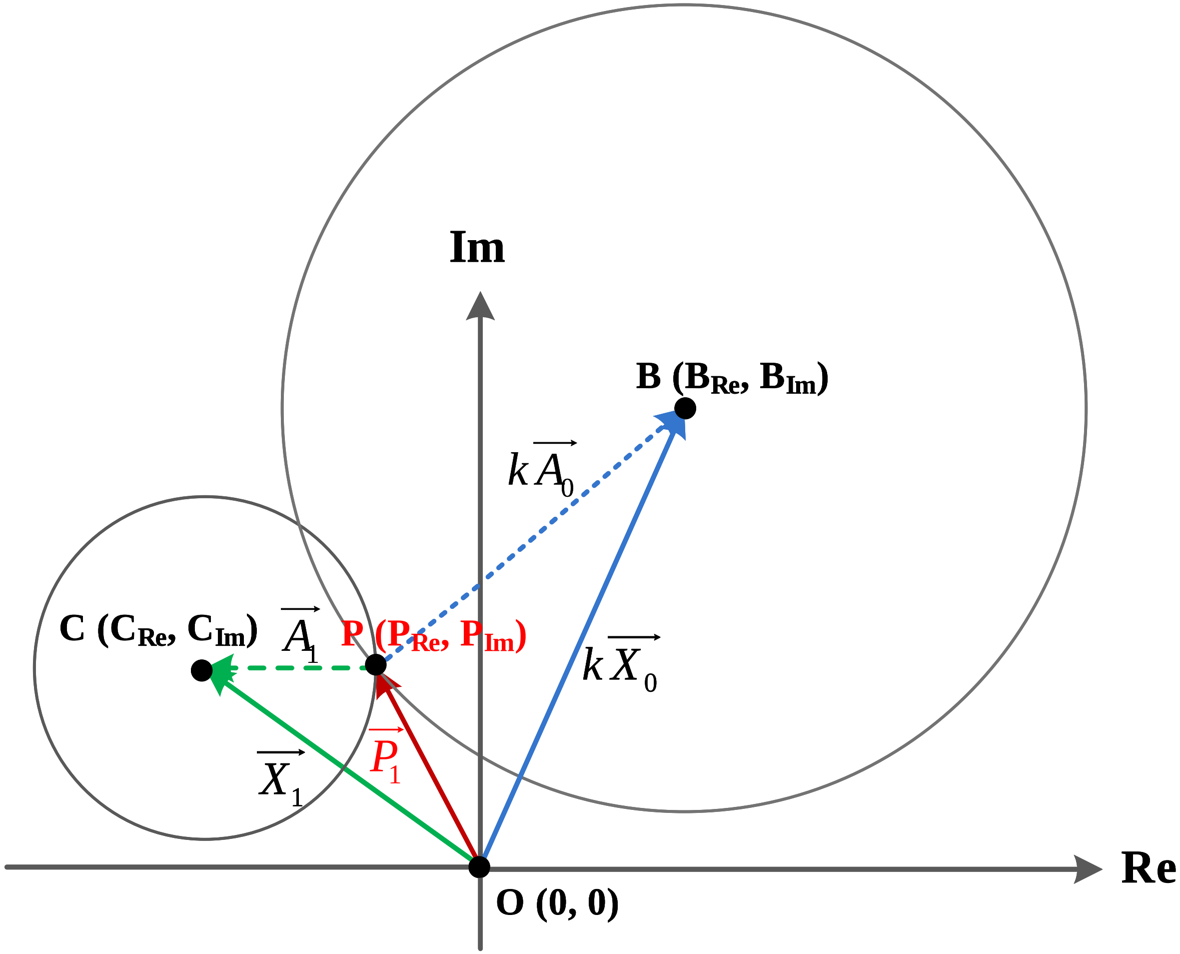
Put the complex spectrum in 2 dimensional vector form

$$k\overrightarrow{X_0} = \overrightarrow{P_1} + k\overrightarrow{A_0}$$
$$\overrightarrow{X_1} = \overrightarrow{P_1} + \overrightarrow{A_1}$$

$$k\overrightarrow{X_0} = \overrightarrow{\mathrm{OB}} = \left(\boxed{\mathrm{B_{Re}}, \mathrm{B_{Im}}}\right)$$
$$\overrightarrow{X_1} = \overrightarrow{\mathrm{OC}} = \left(\boxed{\mathrm{C_{Re}}, \mathrm{C_{Im}}}\right)$$

?

$$\overrightarrow{P_1} = \overrightarrow{\mathrm{OP}} = \left(\boxed{\mathrm{P_{Re}}, \mathrm{P_{Im}}}\right)$$
$$k\overrightarrow{A_0} = \overrightarrow{\mathrm{PB}}$$
$$\overrightarrow{A_1} = \overrightarrow{\mathrm{PC}}$$



From PAE to AME: ambient magnitude estimation

$$r = \left|\overrightarrow{A_0}\right| = \left|\overrightarrow{A_1}\right| \longrightarrow \left|\overrightarrow{PC}\right| = r,\ \left|\overrightarrow{PB}\right| = kr.$$

$$\left(\mathrm{P}_{\mathrm{Re}} - \mathrm{B}_{\mathrm{Re}}\right)^2 + \left(\mathrm{P}_{\mathrm{Im}} - \mathrm{B}_{\mathrm{Im}}\right)^2 = k^2 \hat{r}^2,$$

$$\left(\mathrm{P}_{\mathrm{Re}} - \mathrm{C}_{\mathrm{Re}}\right)^2 + \left(\mathrm{P}_{\mathrm{Im}} - \mathrm{C}_{\mathrm{Im}}\right)^2 = \hat{r}^2.$$

$$\boxed{\begin{aligned}\hat{\mathrm{P}}_{\mathrm{Re}} &= \frac{\mathrm{B}_{\mathrm{Re}} + \mathrm{C}_{\mathrm{Re}}}{2} + \frac{\left(\mathrm{C}_{\mathrm{Re}} - \mathrm{B}_{\mathrm{Re}}\right)\left(k^2 - 1\right)\hat{r}^2 \pm \left(\mathrm{B}_{\mathrm{Im}} - \mathrm{C}_{\mathrm{Im}}\right)\beta}{2\left|\overrightarrow{\mathrm{BC}}\right|^2},\\ \hat{\mathrm{P}}_{\mathrm{Im}} &= \frac{\mathrm{B}_{\mathrm{Im}} + \mathrm{C}_{\mathrm{Im}}}{2} + \frac{\left(\mathrm{C}_{\mathrm{Im}} - \mathrm{B}_{\mathrm{Im}}\right)\left(k^2 - 1\right)\hat{r}^2 \mp \left(\mathrm{B}_{\mathrm{Re}} - \mathrm{C}_{\mathrm{Re}}\right)\beta}{2\left|\overrightarrow{\mathrm{BC}}\right|^2},\end{aligned}}$$

$$\hat{P}_1 = \hat{\mathrm{P}}_{\mathrm{Re}} + j\hat{\mathrm{P}}_{\mathrm{Im}},\ \hat{P}_0 = k^{-1}\left(\hat{\mathrm{P}}_{\mathrm{Re}} + j\hat{\mathrm{P}}_{\mathrm{Im}}\right),$$

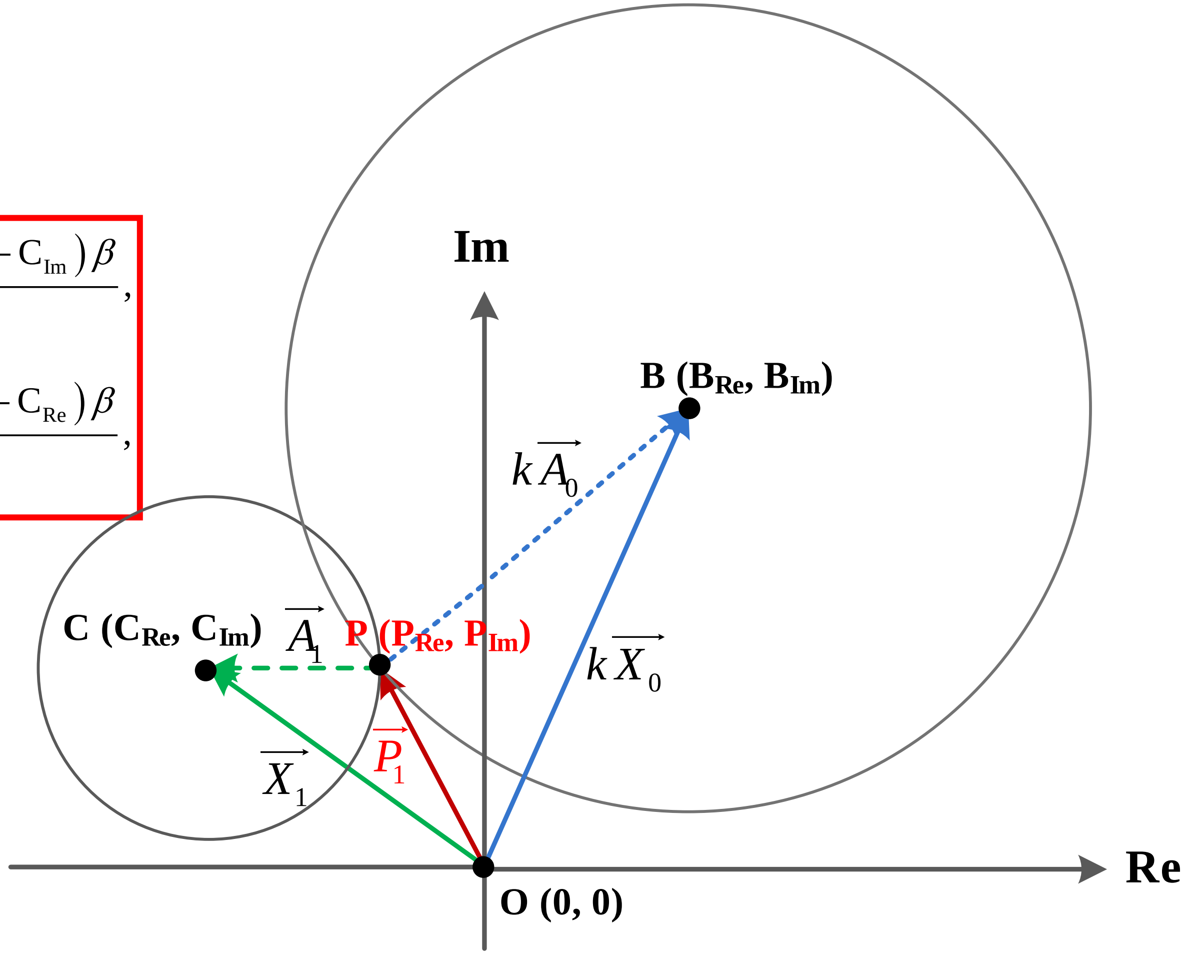
$$\hat{A}_1 = X_1 - \left(\hat{\mathrm{P}}_{\mathrm{Re}} + j\hat{\mathrm{P}}_{\mathrm{Im}}\right),$$

$$\hat{A}_0 = X_0 - k^{-1}\left(\hat{\mathrm{P}}_{\mathrm{Re}} + j\hat{\mathrm{P}}_{\mathrm{Im}}\right).$$

Find r

$$\left|\overrightarrow{\mathrm{BC}}\right| = \sqrt{\left(\mathrm{C}_{\mathrm{Re}} - \mathrm{B}_{\mathrm{Re}}\right)^2 + \left(\mathrm{C}_{\mathrm{Im}} - \mathrm{B}_{\mathrm{Im}}\right)^2},$$

$$\beta = \sqrt{\left[(k+1)^2\hat{r}^2 - \left|\overrightarrow{\mathrm{BC}}\right|^2\right]\left[(k-1)^2\hat{r}^2 - \left|\overrightarrow{\mathrm{BC}}\right|^2\right]}.$$

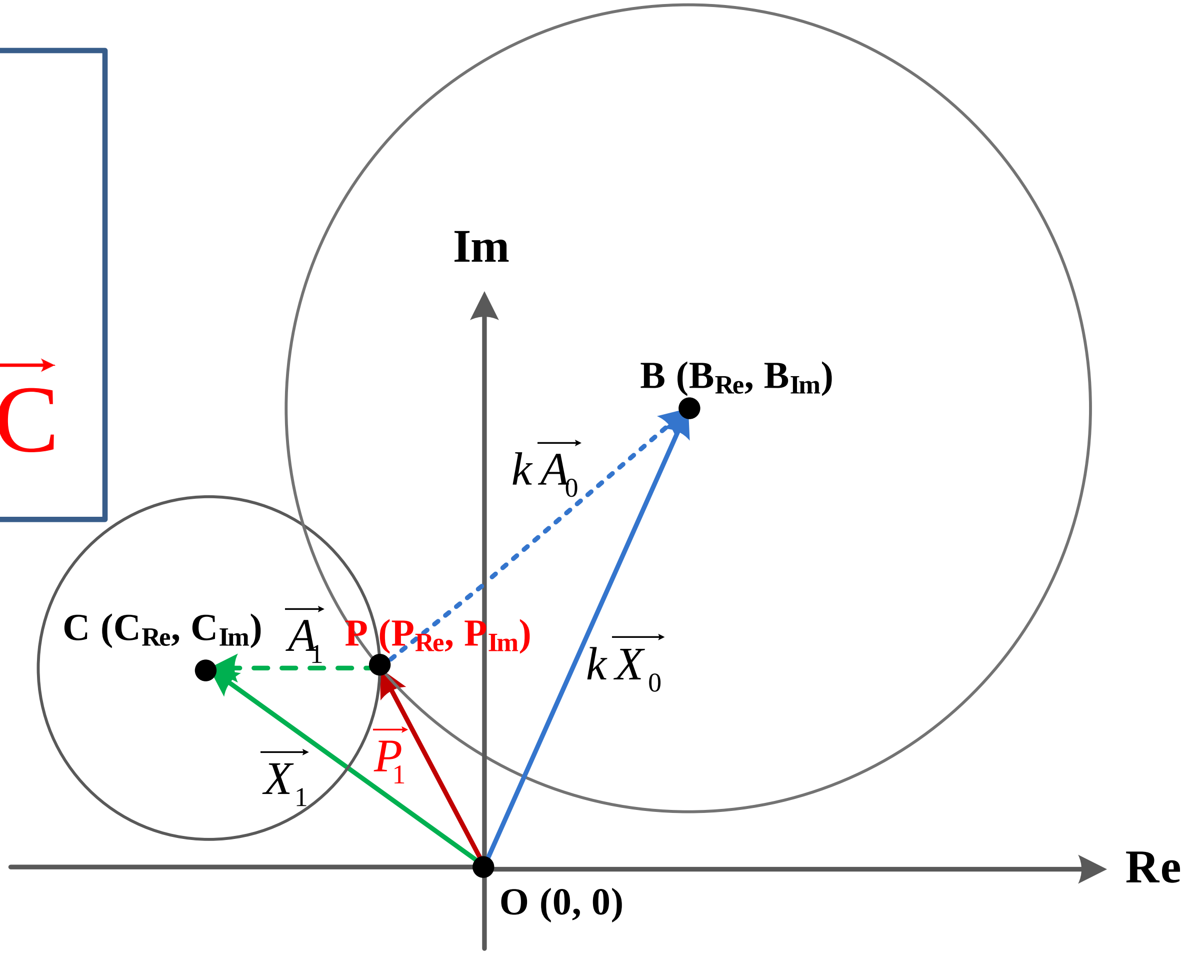


APE ⟺ AME

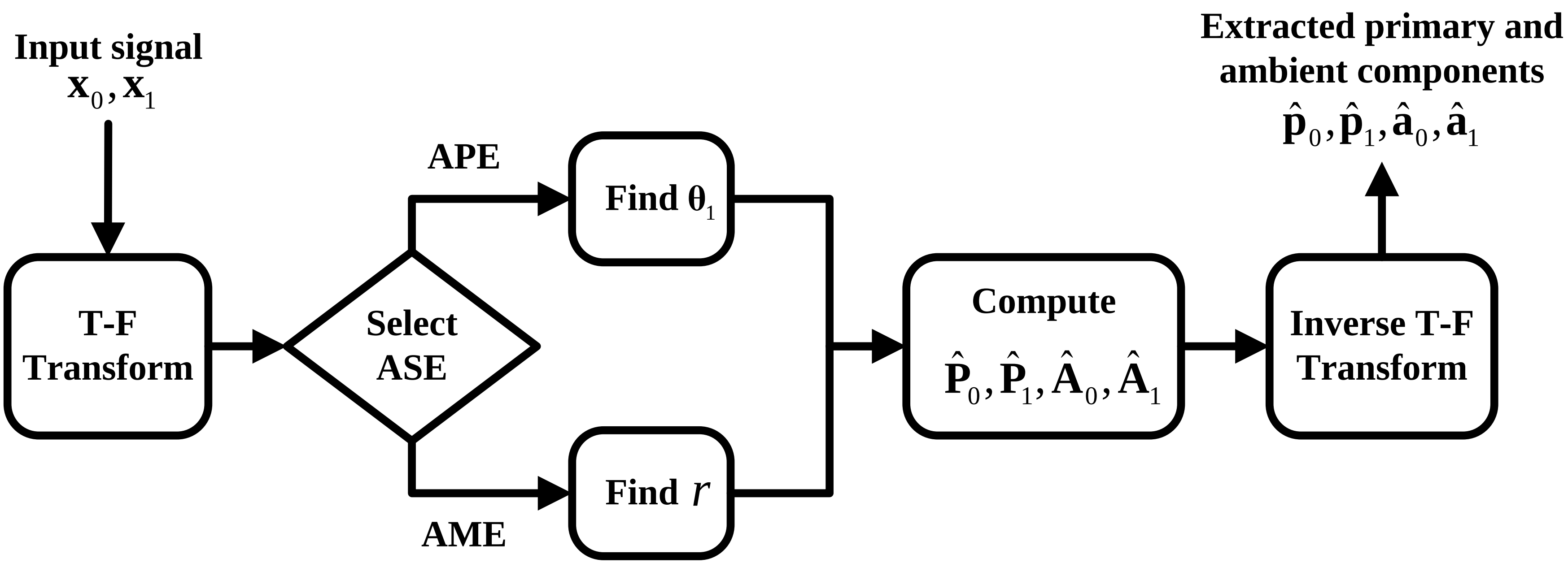
AME: Find $r = \left|\overrightarrow{PC}\right|$

APE: Find $\theta_1 = \angle\overrightarrow{PC}$

Ambient
Spectrum
Estimation

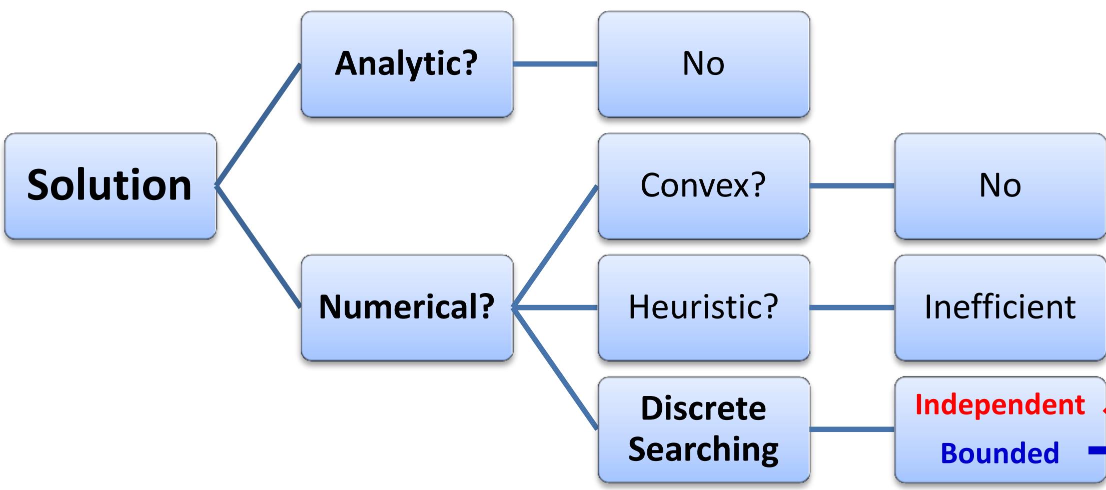


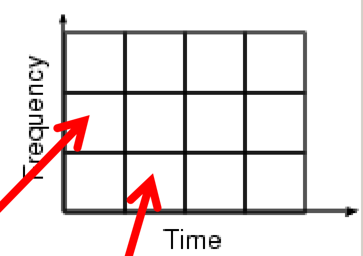
Ambient Spectrum Estimation (ASE)



ASE: how to estimate? Using a Sparsity Constraint

Objective	Find $\boldsymbol{\theta}_1$	Find r
How	$\hat{\boldsymbol{\theta}}_1^* = \arg\min_{\hat{\boldsymbol{\theta}}_1} \left\Vert \hat{\mathbf{P}}_1 \right\Vert_1$	$\hat{\mathbf{r}}^* = \arg\min_{\hat{\mathbf{r}}} \left\Vert \hat{\mathbf{P}}_1 \right\Vert_1$
Method	APES	AMES





$$\left|\hat{P}_1(n_i, l_i)\right| \perp A_1(n_j, l_j)$$

$$\forall n_i \neq n_j \text{ or } l_i \neq l_j$$

$$\hat{\boldsymbol{\theta}}_1 \in [-\pi, \pi]$$

$$\hat{\mathbf{r}} \in [r_{lb}, r_{ub}]$$

Approximate solution APEX and computational cost

Approximate close-form solution APEX :

$$\hat{\boldsymbol{\theta}}_1^* = \begin{cases} \angle \mathbf{X}_1 & , \forall k > 1 \\ \angle \left(\mathbf{X}_1 - \mathbf{X}_0 \right), & \forall k = 1 \end{cases}$$

Computation cost per time-frequency bin

Operations	Square root	Addition	Multiplication	Division	Comparison	Trigonometric operation
APES	D	15D+18	15D+13	4D+6	D-1	**7D+6**
AMES	2D+2	25D+35	24D+24	9D+13	D-1	**0**
APEX	0	13	7	4	1	**7**

D: number of phase or magnitude estimates in discrete searching

Evaluation of PAE- Extraction accuracy

Performance measure: Error-to-Signal Ratio (ESR)

$$\mathrm{ESR_P} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{p}}_c-\mathbf{p}_c\right\|_2^2}{\left\|\mathbf{p}_0\right\|_2^2}\right\},\ \mathrm{ESR_A} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{a}}_c-\mathbf{a}_c\right\|_2^2}{\left\|\mathbf{a}_c\right\|_2^2}\right\}.$$

$$\text{Error} = \textbf{Dist}\text{ortion} + \textbf{Int}\text{erference} + \textbf{Leak}\text{age}$$

$$\text{ESR} \approx \text{DSR} + \text{ISR} + \text{LSR}$$

When there is no analytic solution, how to compute these measures?

We propose an optimization technique to compute these measures.

Optimization method for PAE Extraction accuracy

$$\hat{\mathbf{p}}_c = \mathbf{p}_c + Leak_{\mathbf{p}_c} + Dist_{\mathbf{p}_c} = \mathbf{p}_c + \left(w_{Pc,0}\mathbf{a}_0 + w_{Pc,1}\mathbf{a}_1\right) + Dist_{\mathbf{p}_c},$$

$$\boxed{\left(w^*_{Pc,0}, w^*_{Pc,1}\right) = \arg\min_{\left(w_{Pc,0}, w_{Pc,1}\right)} \left\| \hat{\mathbf{p}}_c - \mathbf{p}_c - \left(w_{Pc,0}\mathbf{a}_0 + w_{Pc,1}\mathbf{a}_1\right)\right\|_2^2,}$$

$$\mathrm{LSR_P} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|w^*_{Pc,0}\mathbf{a}_0 + w^*_{Pc,1}\mathbf{a}_1\right\|_2^2}{\left\|\mathbf{p}_c\right\|_2^2}\right\},$$

$$\mathrm{DSR_P} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{p}}_c - \mathbf{p}_c - \left(w^*_{Pc,0}\mathbf{a}_0 + w^*_{Pc,1}\mathbf{a}_1\right)\right\|_2^2}{\left\|\mathbf{p}_c\right\|_2^2}\right\}.$$

$$\hat{\mathbf{a}}_c = \mathbf{a}_c + Leak_{\mathbf{a}_c} + Intf_{\mathbf{a}_c} + Dist_{\mathbf{a}_c} = \mathbf{a}_c + w_{Ac,c}\mathbf{p}_c + w_{Ac,1-c}\mathbf{a}_{1-c} + Dist_{\mathbf{a}_c},$$

$$\boxed{\left(w^*_{Ac,c}, w^*_{Ac,1-c}\right) = \arg\min_{\left(w_{Ac,c}, w_{Ac,1-c}\right)} \left\| \hat{\mathbf{a}}_c - \mathbf{a}_c - \left(w_{Ac,c}\mathbf{p}_c + w_{Ac,1-c}\mathbf{a}_{1-c}\right)\right\|_2^2,}$$

$$\mathrm{LSR_A} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|w^*_{Ac,c}\mathbf{p}_c\right\|_2^2}{\left\|\mathbf{a}_c\right\|_2^2}\right\},$$

$$\mathrm{ISR_A} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|w^*_{Ac,1-c}\mathbf{a}_{1-c}\right\|_2^2}{\left\|\mathbf{a}_c\right\|_2^2}\right\},$$

$$\mathrm{DSR_A} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{a}}_c - \mathbf{a}_c - \left(w_{Ac,c}\mathbf{p}_c + w_{Ac,1-c}\mathbf{a}_{1-c}\right)\right\|_2^2}{\left\|\mathbf{a}_c\right\|_2^2}\right\}.$$

Objective evaluation

Stimuli

- Primary component:
  - Speech, $k = 2$
- Ambient component:
  - Wave lapping sound
- Primary power ratio (PPR):
  - (0, 1) at an interval of 0.1
- FFT size: 4096

Approaches compared

- Masking
- PCA
- LS
- ASE (APEX)

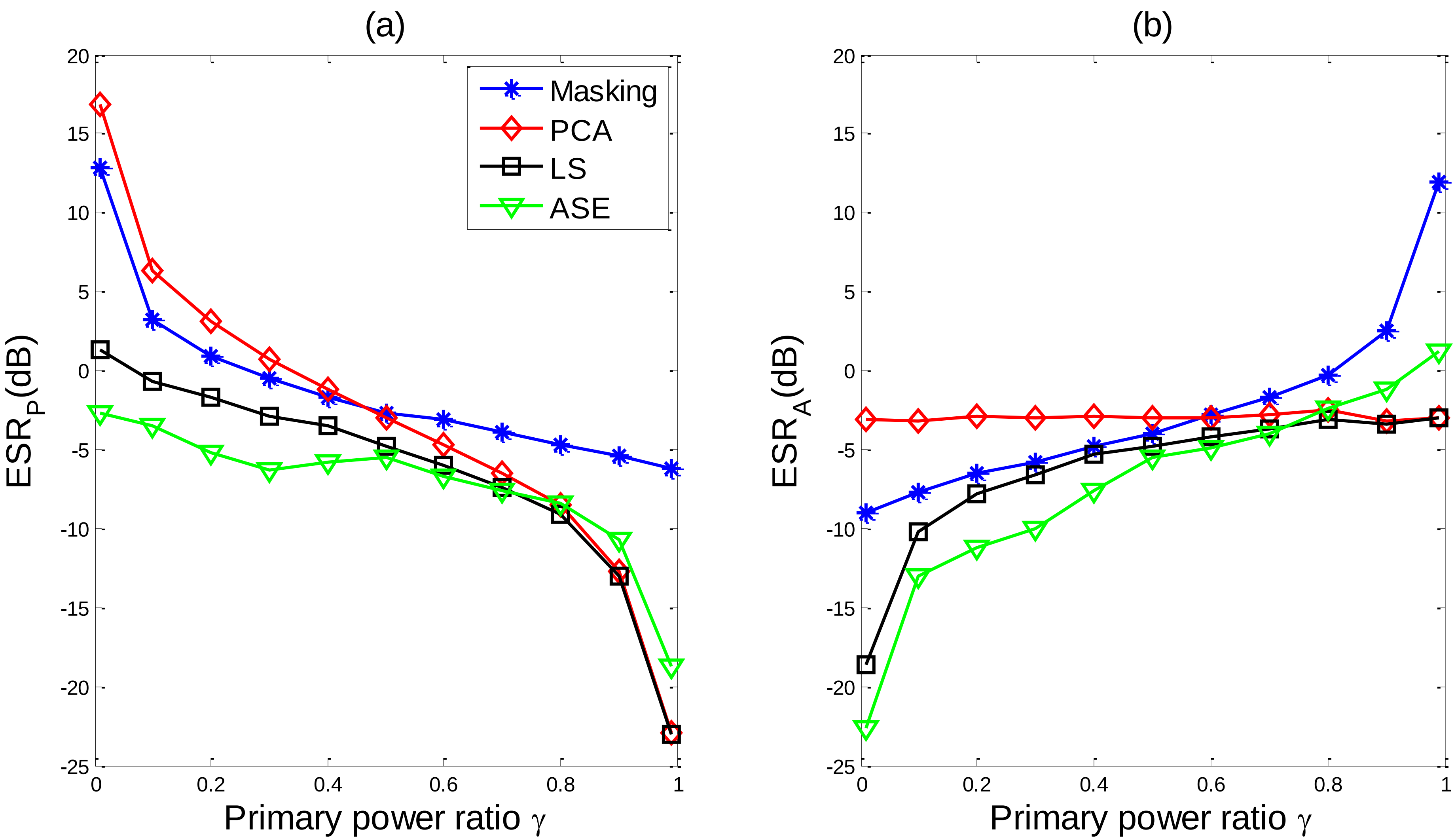
Performance evaluated

1. **Extraction accuracy**: ESR
2. **Spatial accuracy**: ICC, ICLD

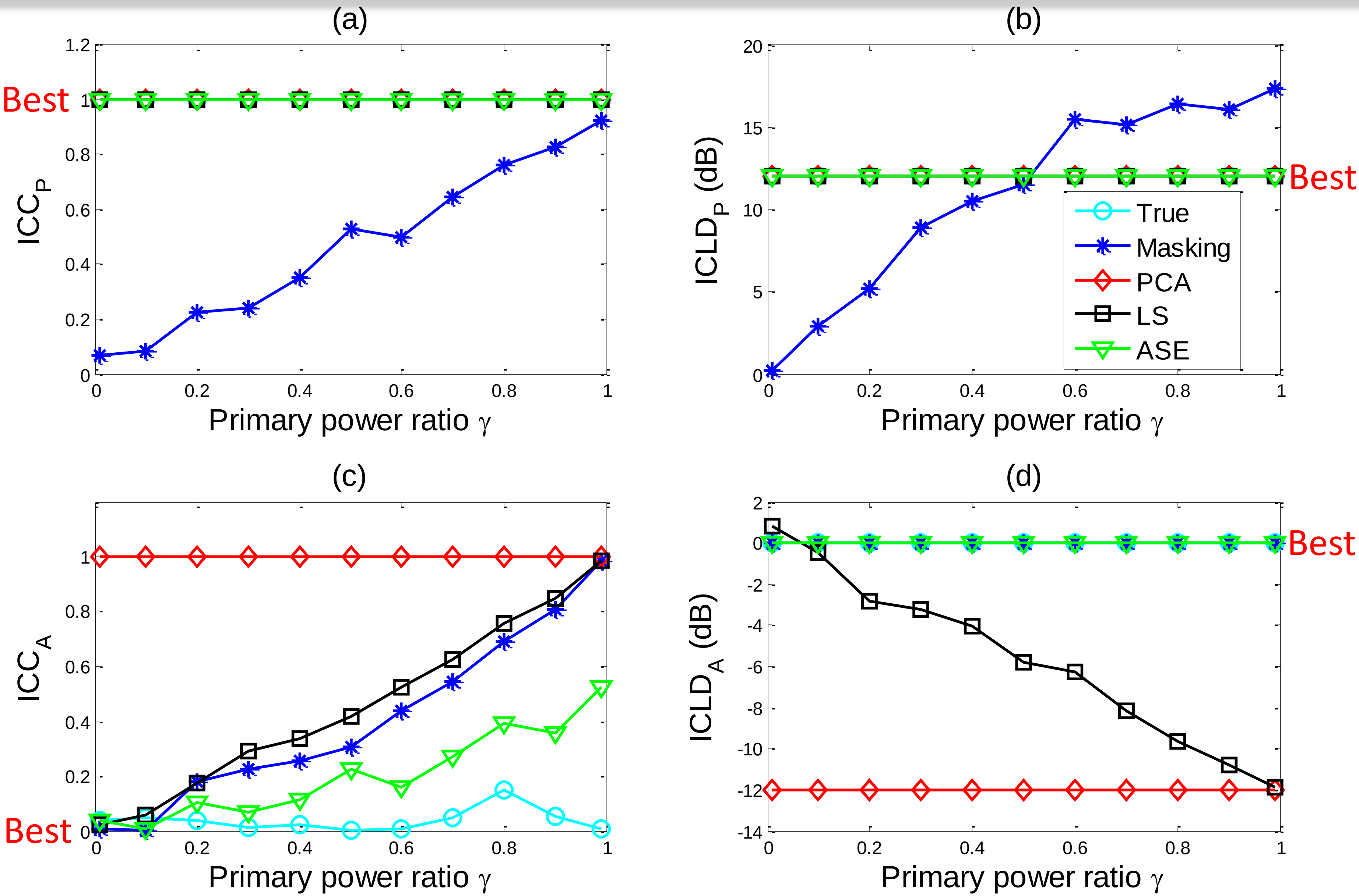
$$\mathrm{ESR}_{\mathrm{P}} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{p}}_c - \mathbf{p}_c\right\|_2^2}{\left\|\mathbf{p}_c\right\|_2^2}\right\},$$

$$\mathrm{ESR}_{\mathrm{A}} = 10\log_{10}\left\{\frac{1}{2}\sum_{c=0}^{1}\frac{\left\|\hat{\mathbf{a}}_c - \mathbf{a}_c\right\|_2^2}{\left\|\mathbf{a}_c\right\|_2^2}\right\}.$$

Extraction accuracy



Spatial accuracy



Subjective evaluation

Stimuli

- Primary component:
  - speech, music, and bee sound, $k = 2$
- Ambient component:
  - forest, canteen, and waterfall sound
- Primary power ratio (PPR):
  - (0.3, 0.7)
- Duration: 2-4 seconds

Approaches compared

- Masking
- PCA
- LS
- ASE (APEX)
- Reference
- Mixture

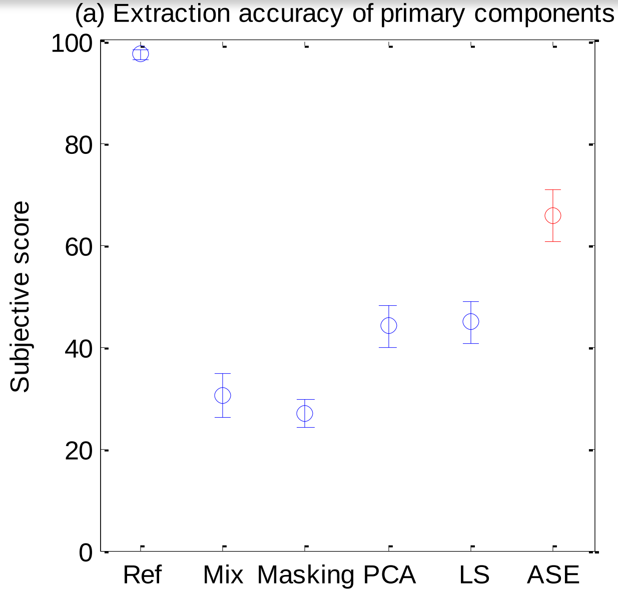
Performance evaluated

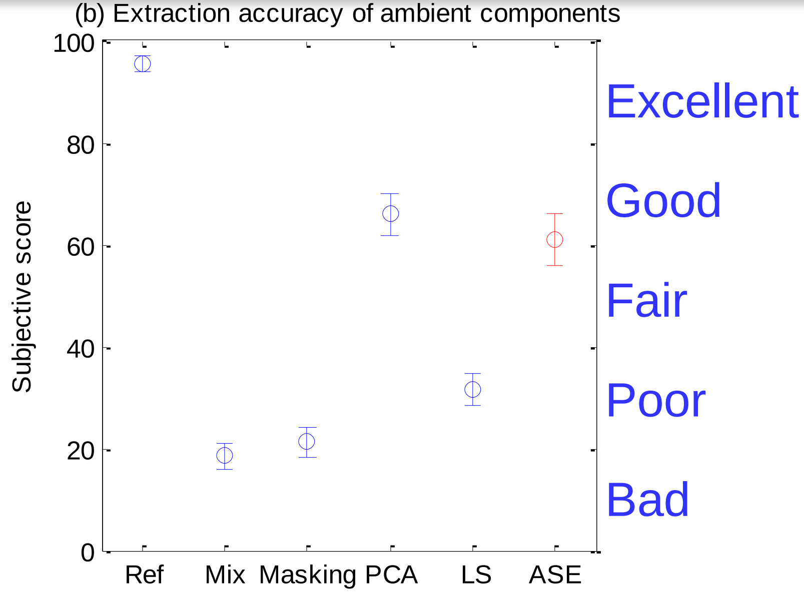
1. **Extraction accuracy**
2. **Ambient diffuseness**

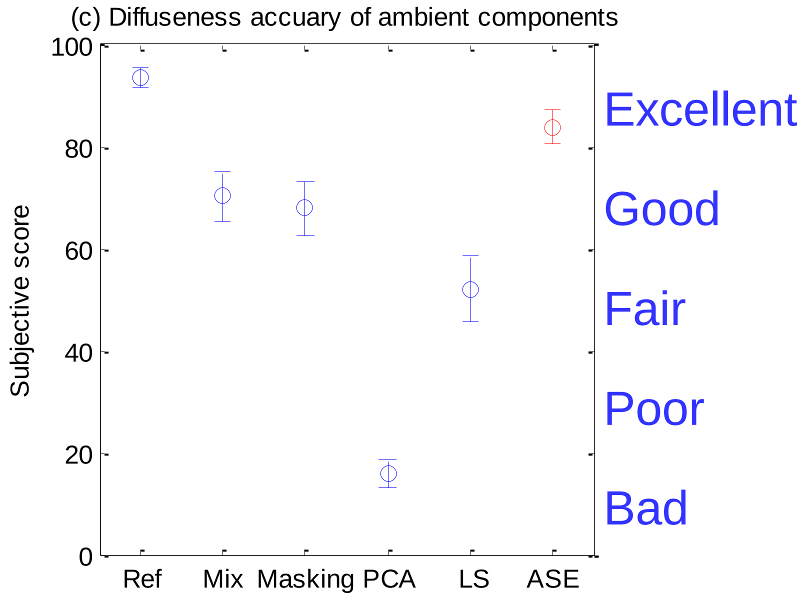
Listening tests

- 17 subjects
- Headphone listening
- Procedure similar to MUSHRA

Subjective scores

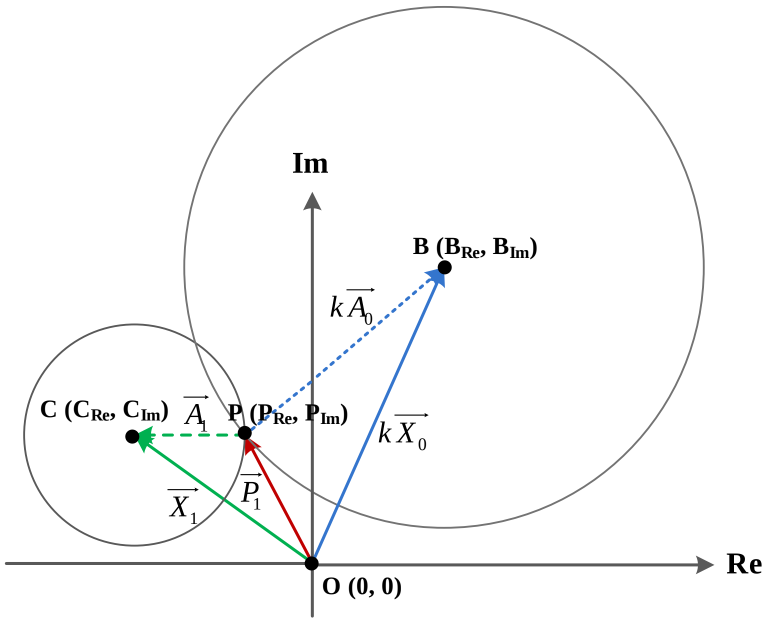






Contributions on ASE based PAE

1. Reformulated PAE as ASE by exploiting the equal ambient magnitude.
2. Solved ASE using the criterion of sparse primary component, resulting in APES, AMES, and APEX.
3. Proposed a technique to compute error measures for PAE approaches without analytic solutions.
4. Validated the improved performance (ESR reduction: 3-6 dB average) in the objective and subjective experiments.



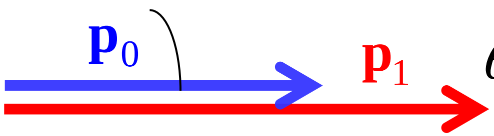
Approximate efficient solution APEX :

$$\hat{\boldsymbol{\theta}}_1^* = \begin{cases} \angle \mathbf{X}_1 & , \forall k > 1 \\ \angle(\mathbf{X}_1 - \mathbf{X}_0), & \forall k = 1 \end{cases}$$

[J2] J. He, W. S. Gan, and E. L. Tan, "Primary-ambient extraction using ambient phase estimation with a sparsity constraint," *IEEE Signal Process. Letters*, vol. 22, no. 8, pp. 1127-1131, Aug. 2015.

[J3] J. He, E. L. Tan, and W. S. Gan, "Primary-ambient extraction using ambient spectrum estimation for immersive spatial audio reproduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1431-1444, Sept. 2015.

PAE in ideal case → complex case

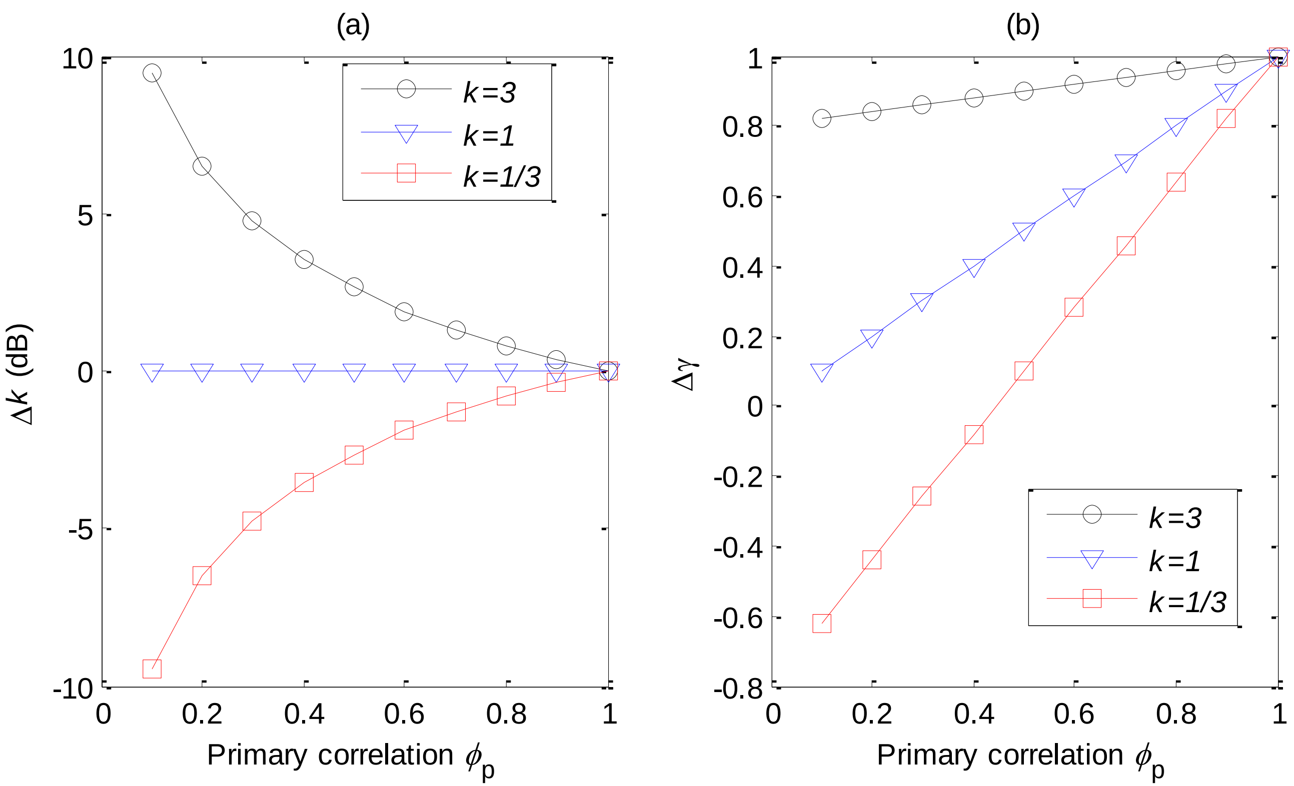


$$\theta_p \neq 0^\circ$$

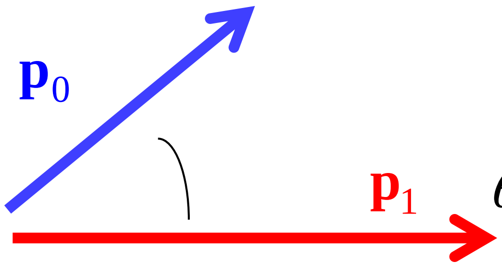
$$\phi_p \leq 1$$

$$\Delta k = \frac{\hat{k}_{ic}}{k} = \frac{k^2 - 1}{2\phi_P k^2} + \sqrt{\left(\frac{k^2 - 1}{2\phi_P k^2}\right)^2 + \frac{1}{k^2}},$$

$$\Delta \gamma = \frac{\hat{\gamma}_{ic}}{\gamma} = \frac{k^2 - 1 + 2\phi_P}{k^2 + 1}.$$



PAE in ideal case → complex case



$$\theta_p \neq 0^\circ$$

$$\phi_p < 1$$

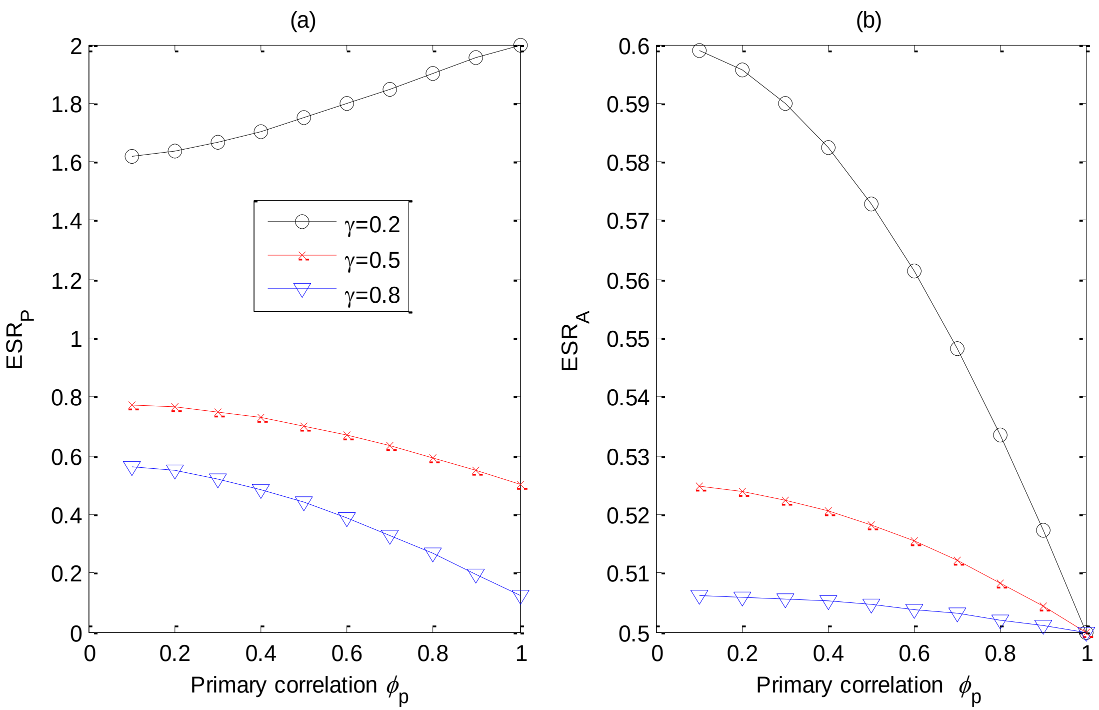
$$\hat{\mathbf{p}}_{\mathrm{PCA},\,0} = \mathbf{p}_0 - \mathbf{v} + \frac{1}{1+\hat{k}_{ic}^{\,2}}\left(\mathbf{a}_0 + \hat{k}_{ic}\mathbf{a}_1\right),$$

$$\hat{\mathbf{p}}_{\mathrm{PCA},\,1} = \mathbf{p}_1 + \frac{1}{\hat{k}_{ic}}\mathbf{v} + \frac{\hat{k}_{ic}}{1+\hat{k}_{ic}^{\,2}}\left(\mathbf{a}_0 + \hat{k}_{ic}\mathbf{a}_1\right),$$

$$\hat{\mathbf{a}}_{\mathrm{PCA},\,0} = \frac{\hat{k}_{ic}^{\,2}}{1+\hat{k}_{ic}^{\,2}}\mathbf{a}_0 + \mathbf{v} + \frac{-\hat{k}_{ic}}{1+\hat{k}_{ic}^{\,2}}\mathbf{a}_1,$$

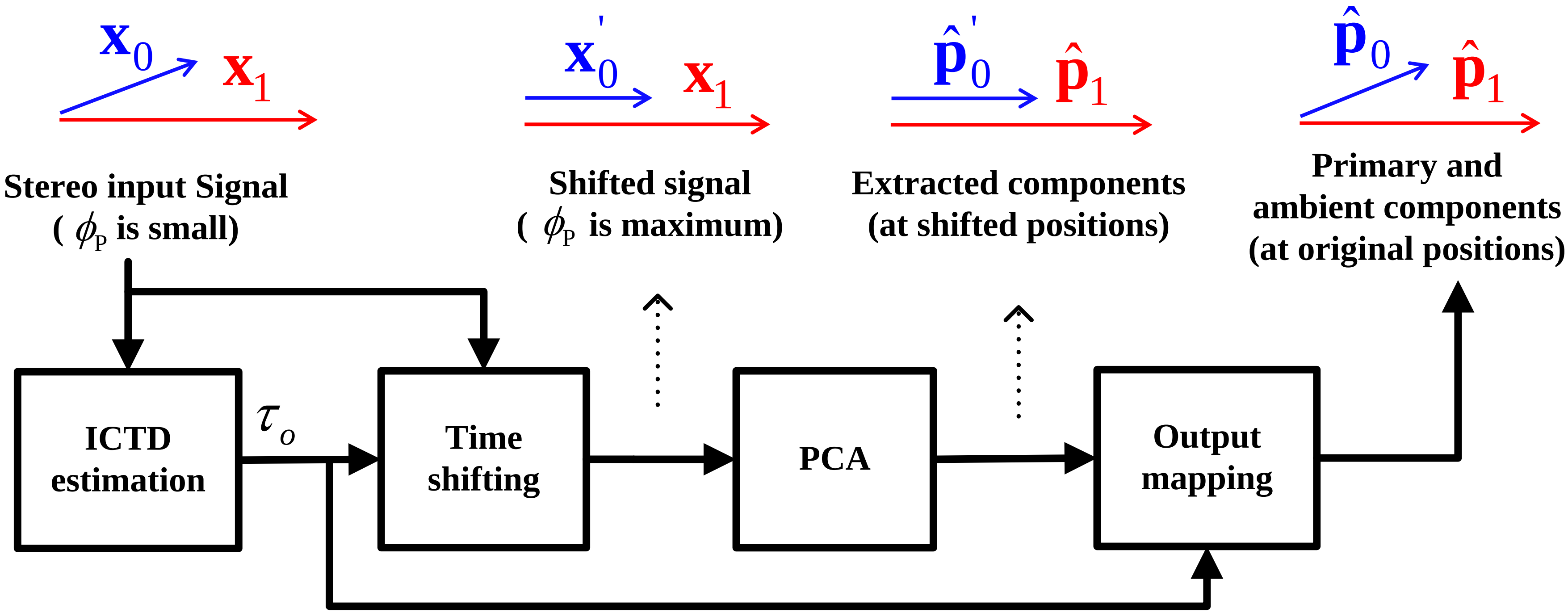
$$\hat{\mathbf{a}}_{\mathrm{PCA},\,1} = \frac{1}{1+\hat{k}_{ic}^{\,2}}\mathbf{a}_1 - \frac{1}{\hat{k}_{ic}}\mathbf{v} + \frac{-\hat{k}_{ic}}{1+\hat{k}_{ic}^{\,2}}\mathbf{a}_0,$$

$$\boxed{\mathbf{v} = \frac{\hat{k}_{ic}}{1+\hat{k}_{ic}^{\,2}}\left(\hat{k}_{ic}\mathbf{p}_0 - \mathbf{p}_1\right)}$$



Time-shifting for PAE

To account for the partial primary correlation (0-lag) caused by the inter-channel time difference (ICTD).

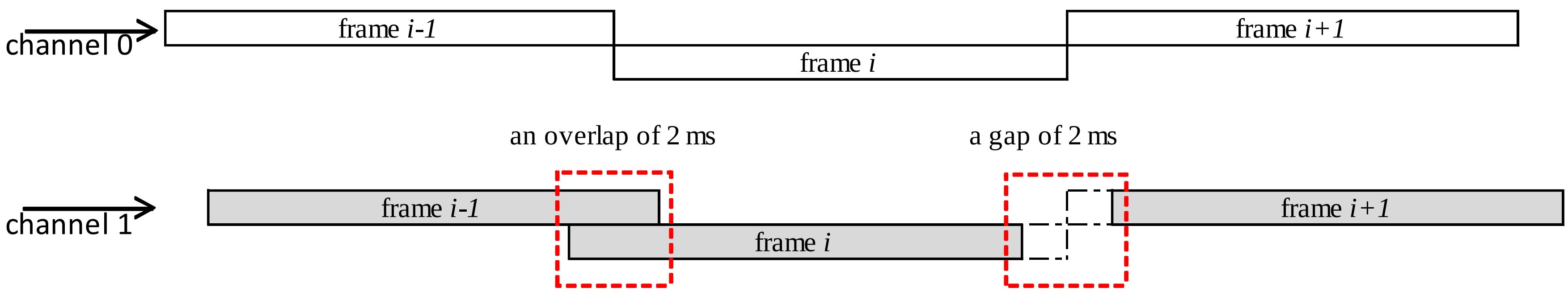


$$\hat{p}_{\text{SPCA, 0}}(n) = \frac{1}{1+\hat{k}_{ic}^2}\left[x_0(n) + \hat{k}_{ic} x_1(n-\tau_o)\right], \quad \hat{p}_{\text{SPCA, 1}}(n) = \frac{\hat{k}_{ic}}{1+\hat{k}_{ic}^2}\left[x_0(n+\tau_o) + \hat{k}_{ic} x_1(n)\right],$$

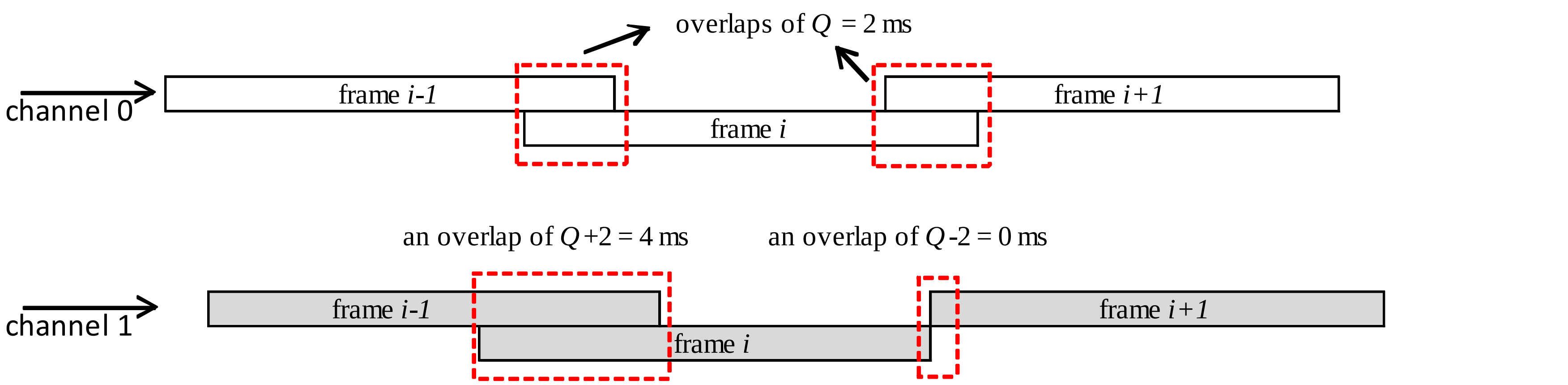
$$\hat{a}_{\text{SPCA, 0}}(n) = \frac{\hat{k}_{ic}}{1+\hat{k}_{ic}^2}\left[\hat{k}_{ic} x_0(n) - x_1(n-\tau_o)\right], \quad \hat{a}_{\text{SPCA, 1}}(n) = -\frac{1}{1+\hat{k}_{ic}^2}\left[\hat{k}_{ic} x_0(n+\tau_o) - x_1(n)\right].$$

Output mapping

Frame	*i-1*	*i*	*i+1*
ICTD (ms)	1	-1	1



(a) Conventional output mapping



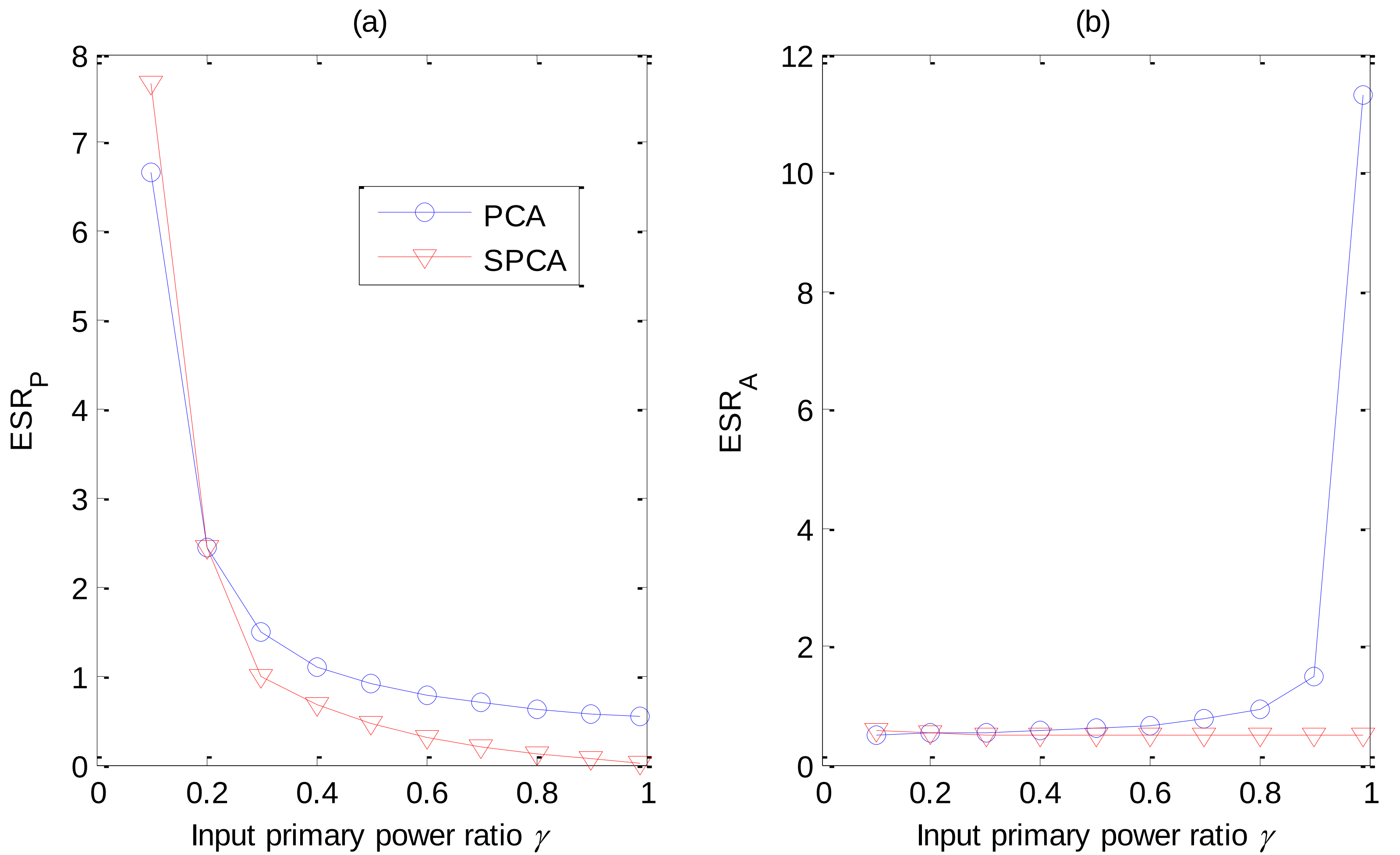
(b) Overlapped output mapping

Overlapping duration: $Q \geq 2$ ms

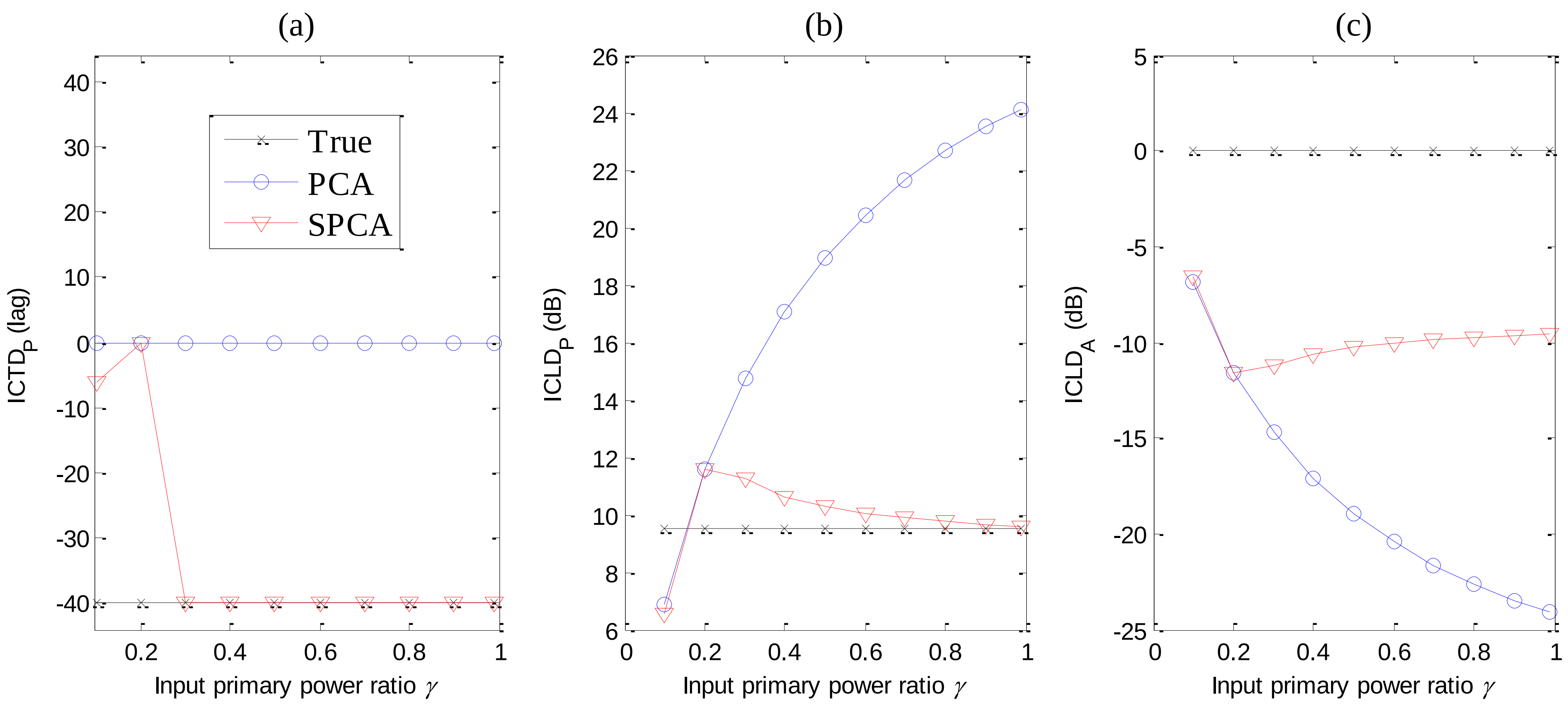
Experiments

Exp	Input signal	Primary component	Ambient component	Settings
1	Synthesized	Speech	Lapping wave	Fixed direction; different values of γ
2	Synthesized	Shaking matchbox	Lapping wave	Panning directions with close γ
3	Synthesized	Direct path of speech	Reverberation of speech	Varying directions with different γ
4	Recorded	Speech	Canteen recording	Three directions with close γ

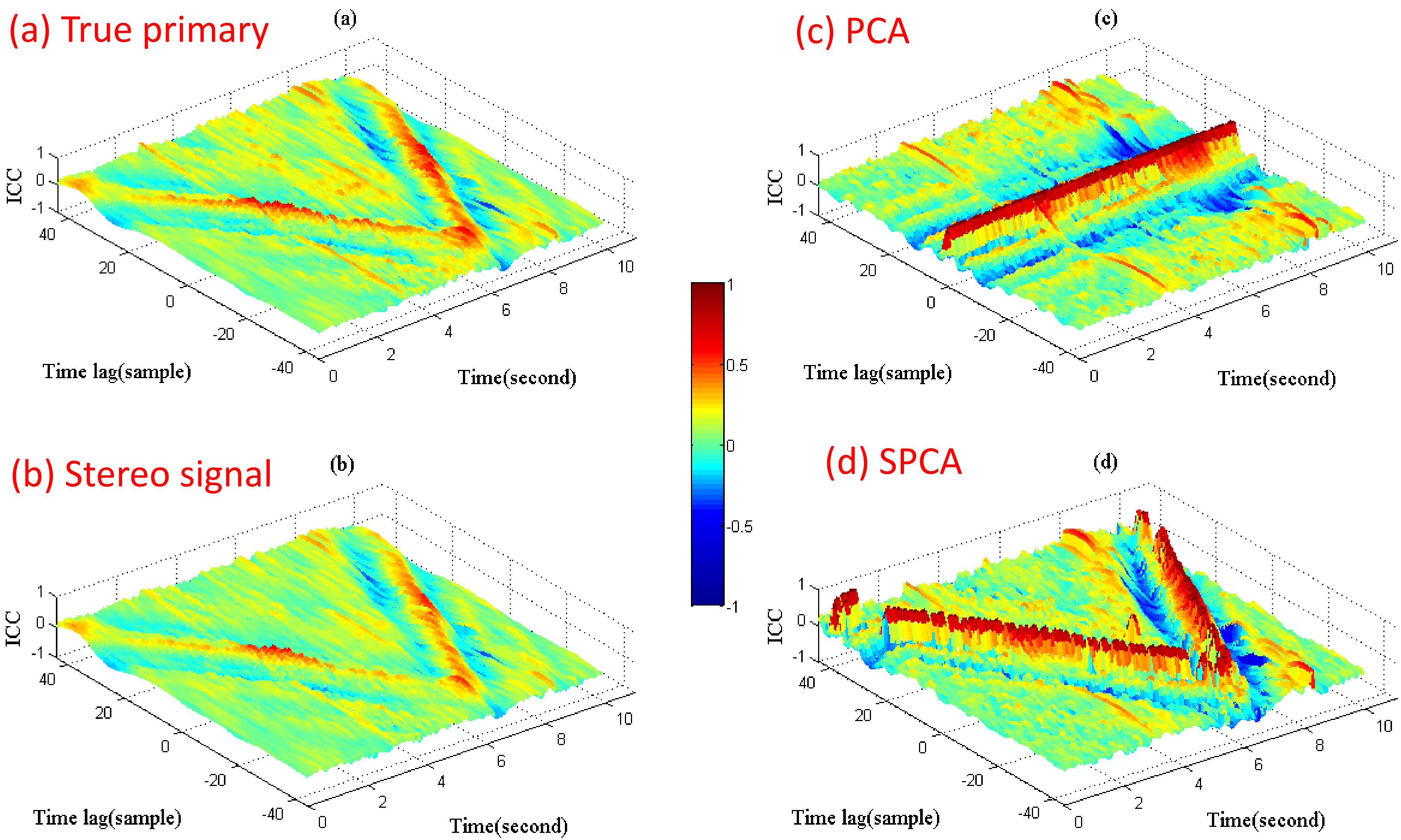
Experiment 1: Extraction accuracy



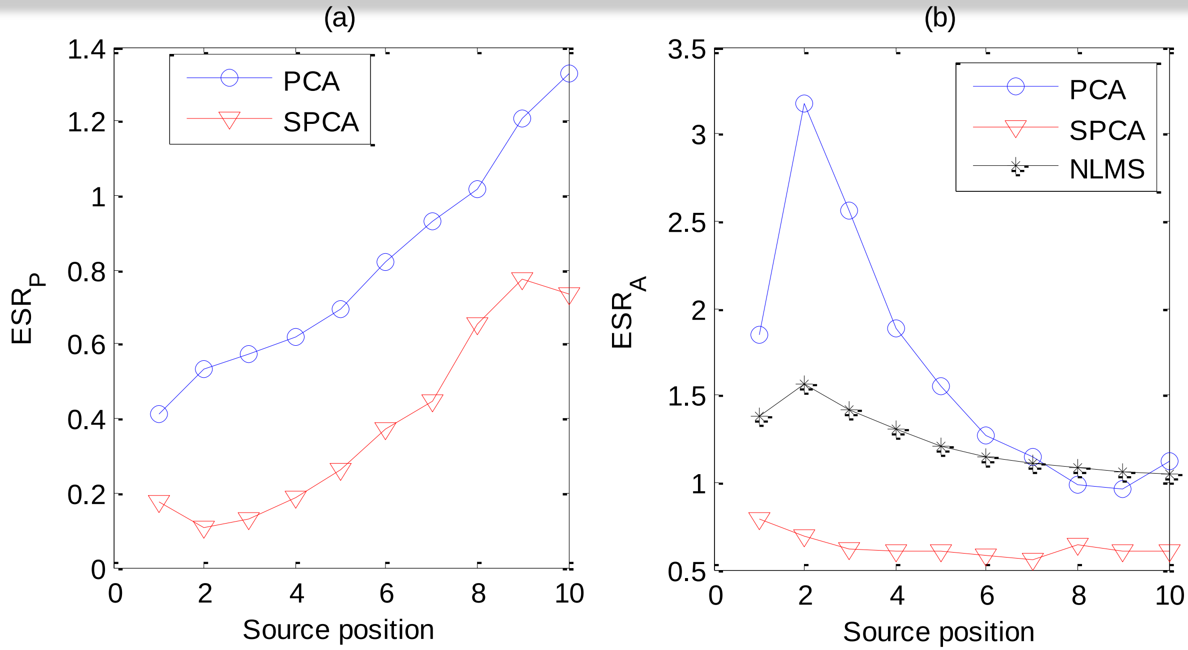
Experiment 1: Spatial accuracy

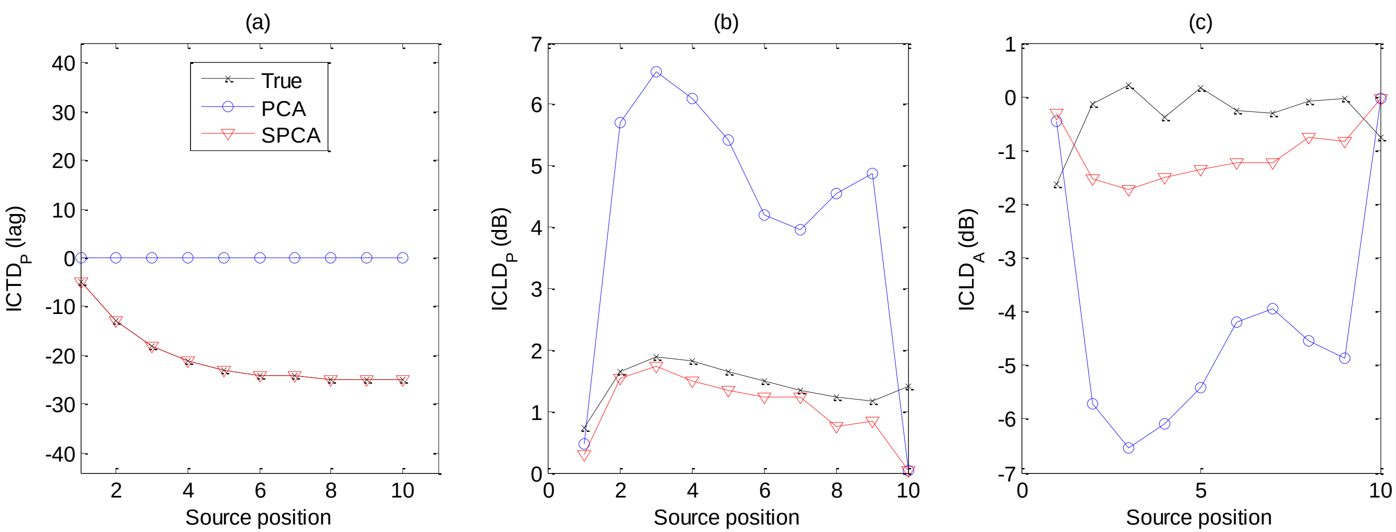


Experiment 2: Tracking of moving source

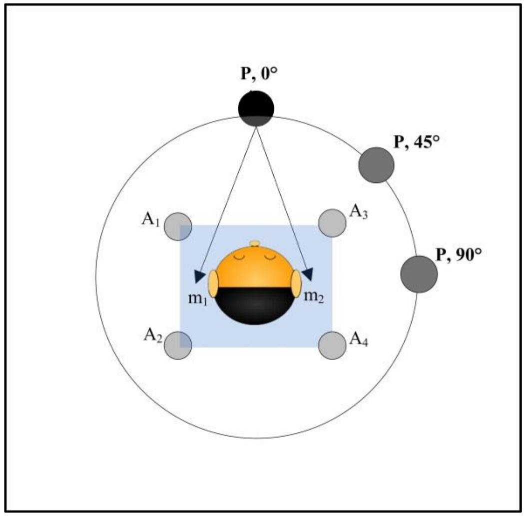


Experiment 3: Reverberation experiment





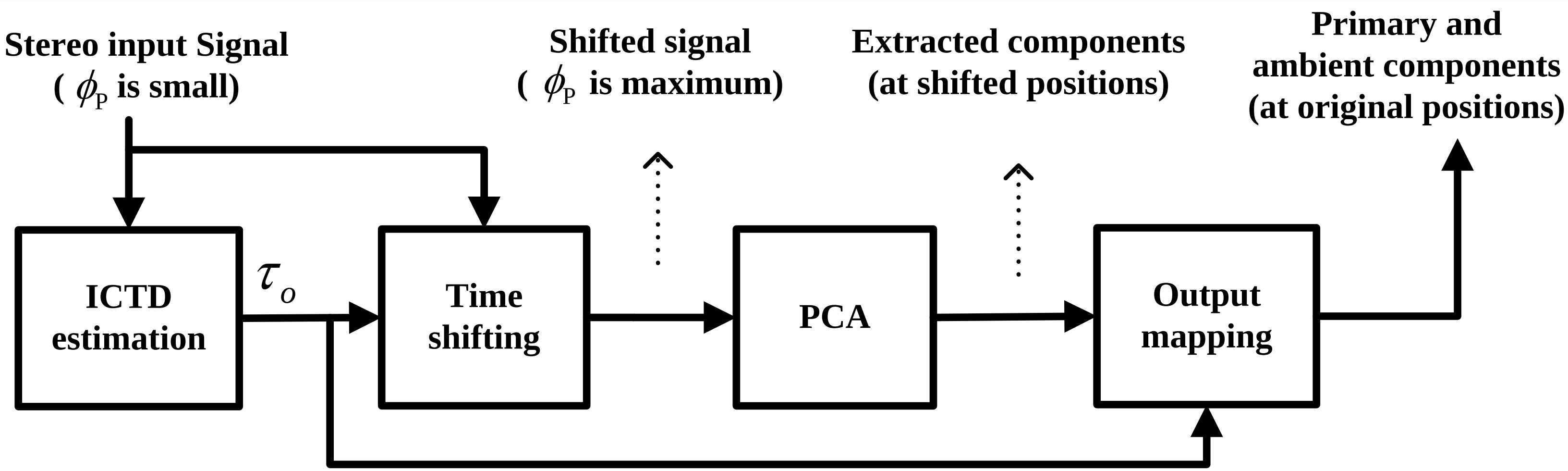
Experiment 4: Recording experiment



ESR	**Primary component**			**Ambient component**		
θ	0°	45°	90°	0°	45°	90°
PCA	0.27	0.64	0.88	1.08	1.89	2.49
SPCA	**0.21**	**0.31**	**0.34**	**0.81**	**1.02**	**1.39**

Spatial	$ICTD_P$			$ICLD_P$ (dB)			$ICLD_A$(dB)		
θ	0°	45°	90°	0°	45°	90°	0°	45°	90°
True	1	-17	-31	-1.02	7.74	11.90	1.03	1.18	1.03
PCA	0	0	0	-1.46	36.03	23.11	1.46	-36.03	-23.11
SPCA	**1**	**-17**	**-31**	**-1.26**	**8.65**	**15.60**	**1.26**	**-8.65**	**-15.60**

Contributions on Time Shifting PAE



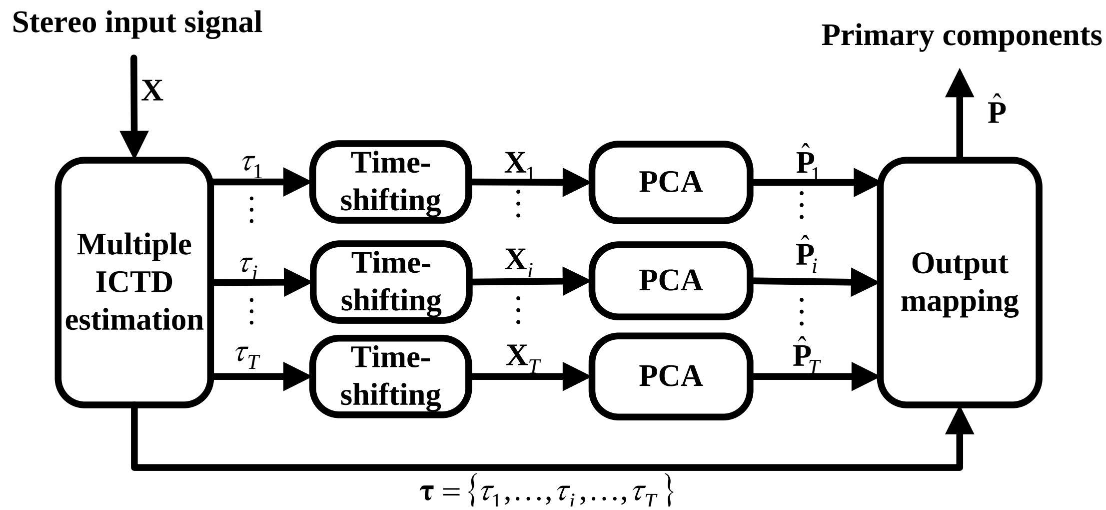
For mixture signals with partially correlated primary components,

- **Analyzed the performance of conventional PAE;**
- **Proposed time shifting technique to improve the PAE performance;**
- **Achieved lower extraction error and more accurate spatial cues.**

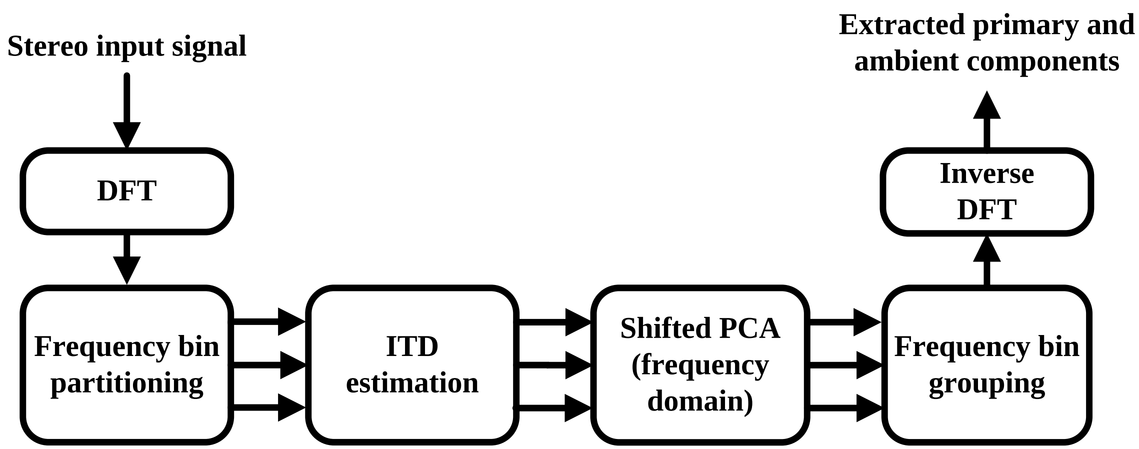
[J4] J. He, W. S. Gan, and E. L. Tan, "Time-shifting based primary-ambient extraction for spatial audio reproduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 10, pp. 1576-1588, Oct. 2015.

Extensions for Multiple Sources

Multi-shift PAE with ICC based output weighting



Subband PAE with frequency bin partitioning

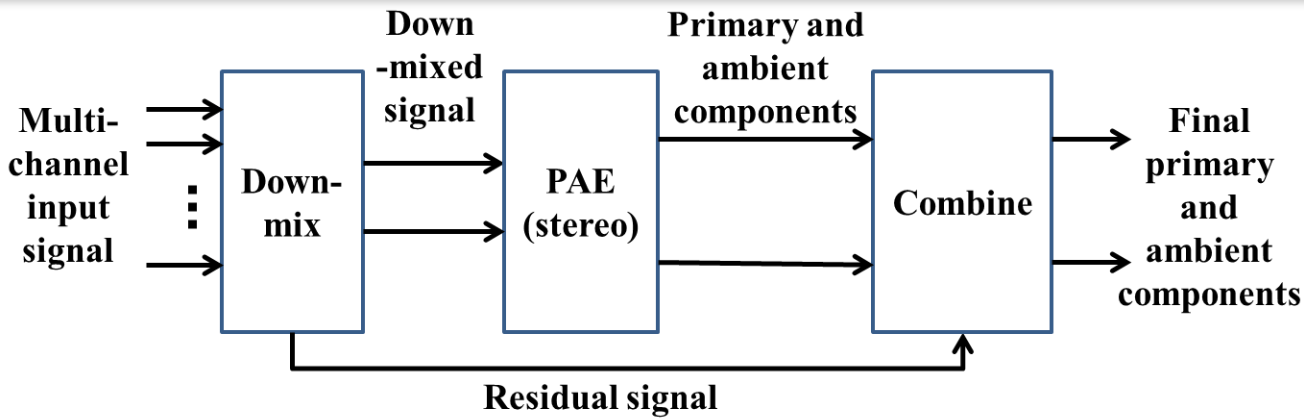


[C3] J. He, and W. S. Gan, "Multi-shift principal component analysis based primary component extraction for spatial audio reproduction," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015, pp. 350-354.

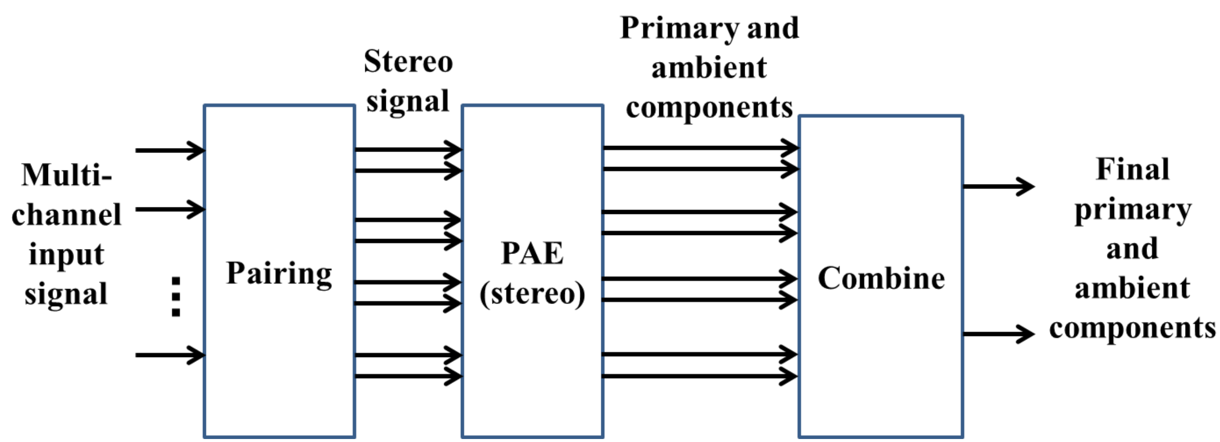
[C2] J. He, E. L. Tan, and W. S. Gan, "A study on the frequency-domain primary-ambient extraction for stereo audio signals," in *Proc. ICASSP*, Florence, Italy, 2014, pp. 2892-2896.

PAE: from stereo to multichannel signals

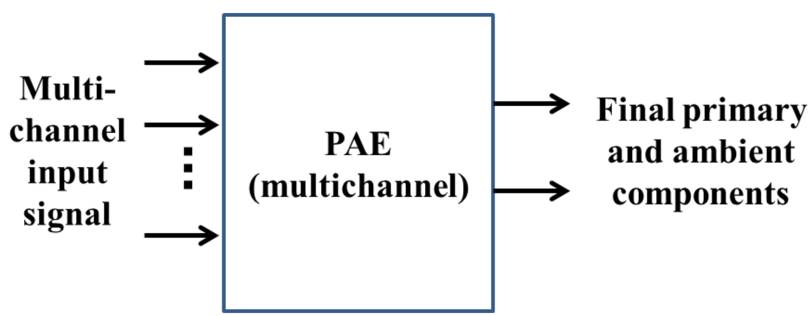
1. Using down-mix



2. Using pairing



3. Direct

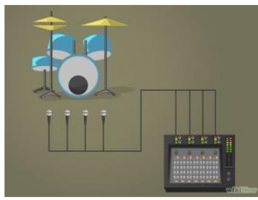
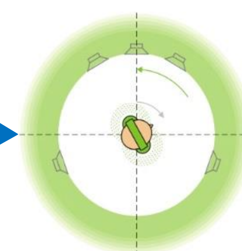


[C5] **J. He**, and W. S. Gan, "Applying primary ambient extraction for immersive spatial audio reproduction," in Proc. *APSIPA Annual Summit and Conference*, Hong Kong, Dec. 2015.

Natural Sound Rendering for Headphones

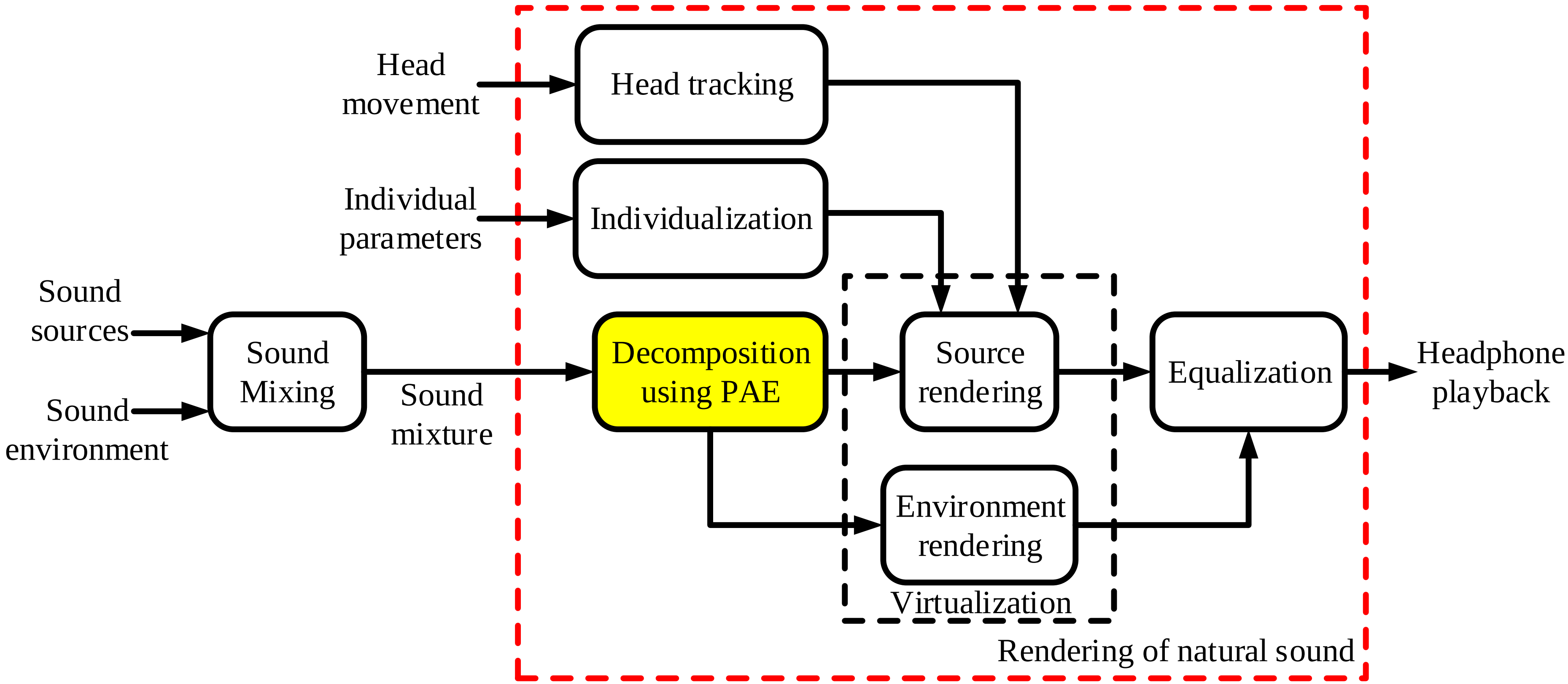
K. Sunder, J. He, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones," ***IEEE Signal Processing Magazine***, vol. 32, no. 2, pp. 100-113, Mar. 2015.

Challenges and solutions

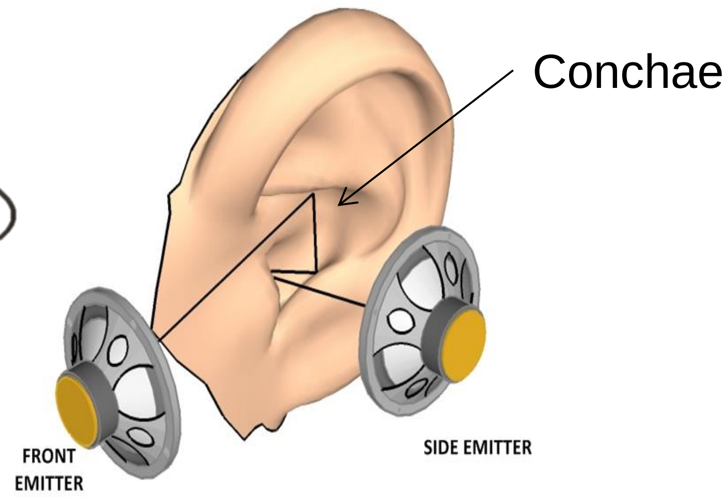
Listening	Headphone listening	Signal processing	Natural listening
Source		Virtualization	
		Sound scene decomposition	
Medium		Equalization	
Receiver		Individualization	
		Head tracking	

D. R. Begault, *3-D sound for virtual reality and multimedia*: AP Professional, 2000.

Integration

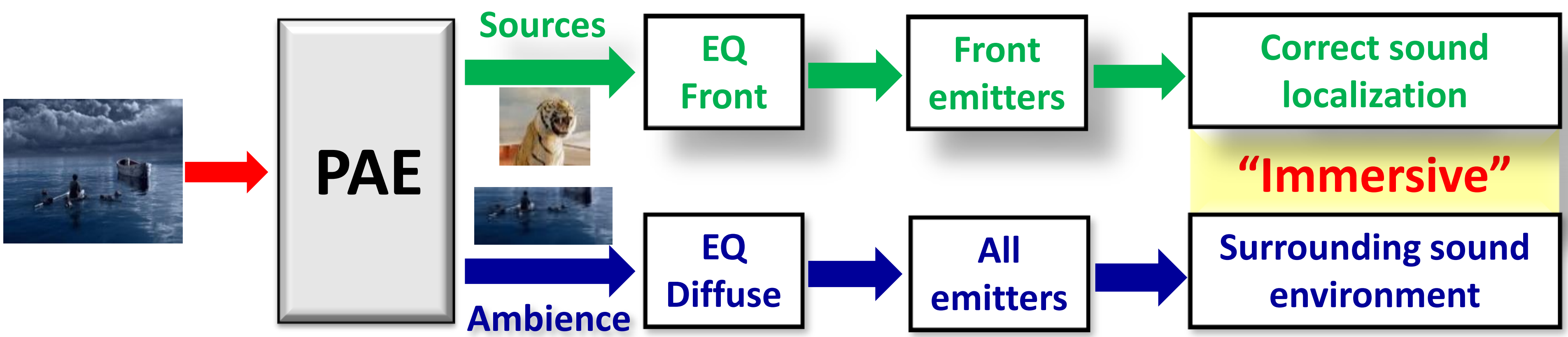


3D Audio Headphone: an example



Key features

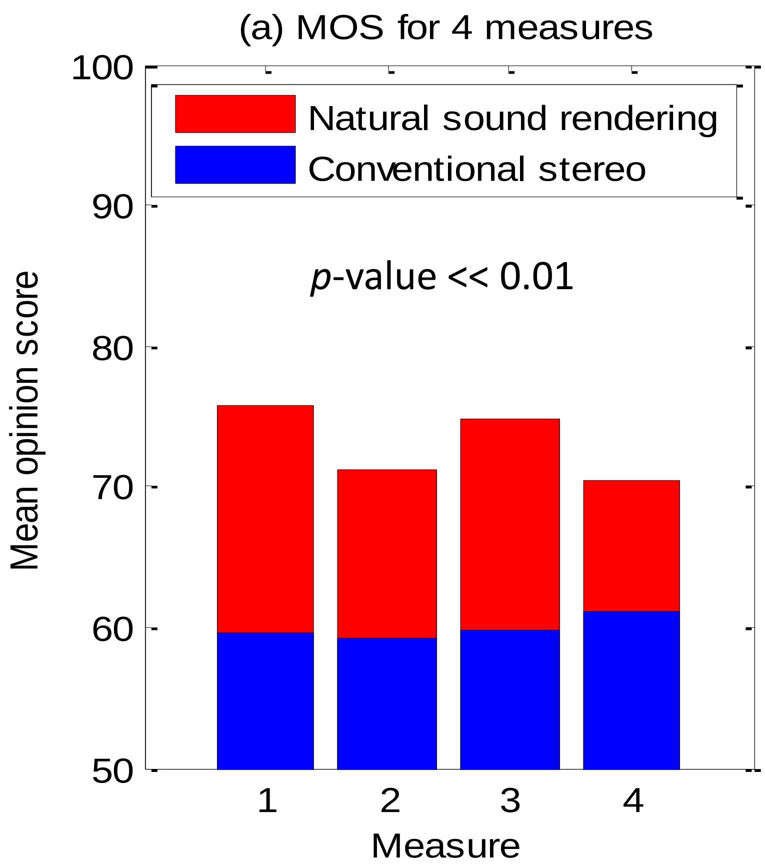
- ✓ Patented structure with strategic-positioned emitters;
- ✓ Individualization via frontal projection; no measurements or training required;
- ✓ Recreate an immersive perception of sound objects with surrounding ambience;
- ✓ Compatible with all existing sound formats.



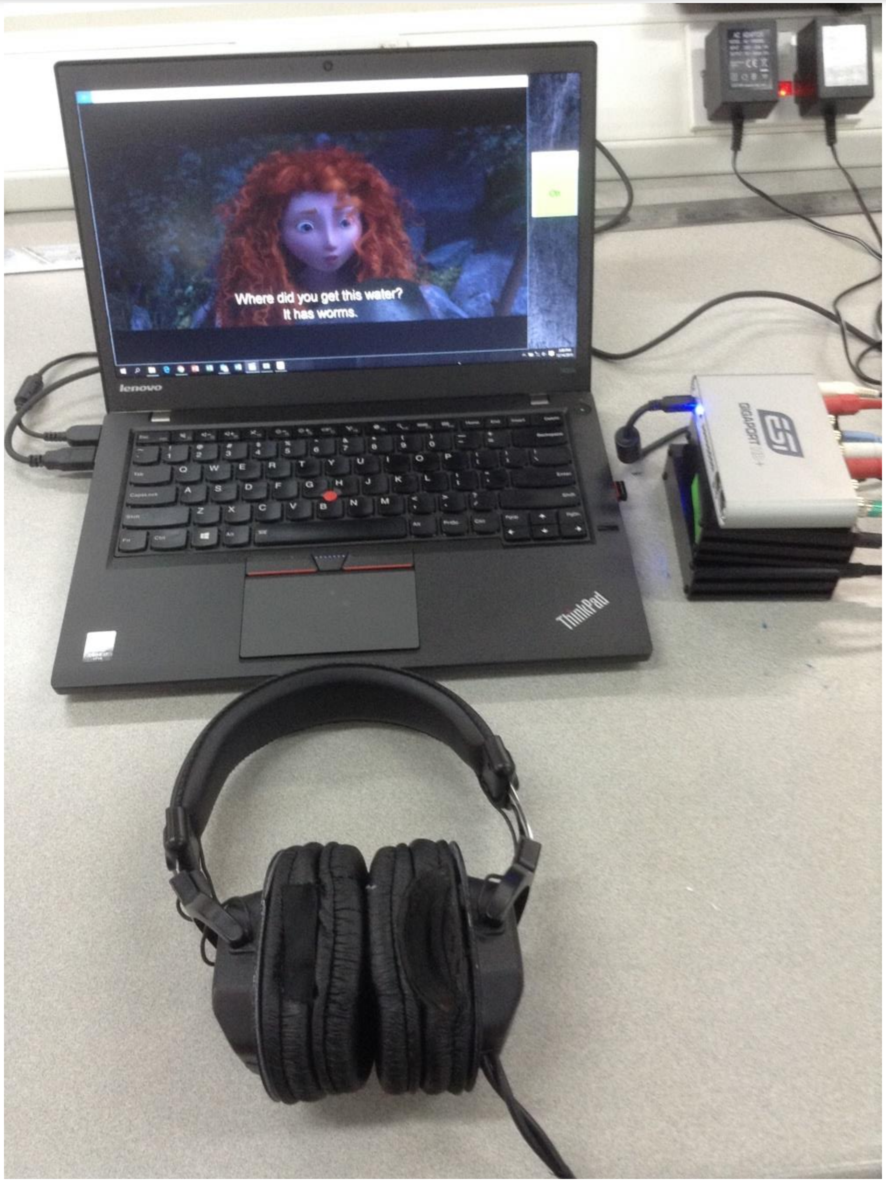
W. S. Gan and E. L. Tan, "Listening device and accompanying signal processing method," US Patent 2014/0153765 A1, 2014.

Subjective evaluation

- 18 subjects, score of 0-100;
- **Stimuli**: binaural, movie and gaming tracks;
- **Four measures**:
  1. Sense of direction,
  2. Externalization,
  3. Ambience,
  4. Timbral quality.



Real-time 3D audio headphones



Show & Tell

at

ICASSP 2016

Conclusions of this Thesis

Primary Ambient Extraction facilitates flexible, efficient, and immersive spatial audio reproduction of channel-based signals for any playback systems

- Comprehensive study on linear estimation based PAE approaches lay the foundation;
- Novel ambient spectrum estimation framework significantly improves PAE performance;
- Time shifting and subband decomposition techniques enhance the robustness of PAE performance in complex scenarios;
- The novel natural sound rendering headphone system validates the advantage of PAE.

Future work

- Extended study of PAE in complex cases for all approaches;
- To achieve tradeoff between timbre and spatial quality in PAE, and objective evaluation system that can mimic subjective performance;
- Evaluate PAE in specific spatial audio reproduction applications (loudspeakers, headphones);
- Incorporate probabilistic approaches, and make use of Big Data and Machine Learning to improve the robustness in complex audio scenes.

Author’s full publication list

Journal papers

[J1] **J. He**, E. L. Tan, and W. S. Gan, “Linear estimation based primary-ambient extraction for stereo audio signals,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 2, pp. 505-517, Feb .2014.

[J2] K. Sunder, **J. He**, E. L. Tan, and W. S. Gan, “Natural sound rendering for headphones: Integration of signal processing techniques,” *IEEE Sig. Process. Mag.*, vol. 32, no. 2, Mar 2015, pp. 100-113.

[J3] **J. He**, W. S. Gan, and E. L. Tan, “Primary-ambient extraction using ambient phase estimation with a sparsity constraint,” *IEEE Signal Process. Letters*, vol. 22, no. 8, pp. 1127-1131, Aug. 2015.

[J4] **J. He**, E. L. Tan, and W. S. Gan, “Primary-ambient extraction using ambient spectrum estimation for immersive spatial audio reproduction,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1430-1443, Sept. 2015.

[J5] **J. He**, W. S. Gan, and E. L. Tan, “Time-shifting based primary-ambient extraction for spatial audio reproduction,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 10, pp. 1576-1588, Oct. 2015.

[J6] **J. He**, R. Ranjan, W. S. Gan, and K. Sunder, “Scalable data reusing normalized LMS for acoustic system identification with short-duration signals,” *IEEE Signal Process. Letters*, under review.

Conference papers

[C1] **J. He**, E. L. Tan, and W. S. Gan, “Time-shifted principal component analysis based cue extraction for stereo audio signals,” in *Proc. ICASSP*, Vancouver, Canada, 2013, pp. 266-270.

[C2] **J. He**, W. S. Gan, and E. L. Tan, “A study on the frequency-domain primary-ambient extraction for stereo audio signals,” in *Proc. ICASSP*, Florence, Italy, 2014, pp. 2892-2896. (*Awarded SPS travel grant*)

[C3] **J. He**, W. S. Gan and Y. K. Chong, “Study on the use of error term in parallel-form narrowband feedback active noise control systems,” in *Proc. 2014 Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference (invited)*, Cambodia, Dec. 2014.

[C4] **J. He**, W. S. Gan, and E. L. Tan, “On the preprocessing and postprocessing of HRTF individualization based on sparse representation of anthropometry features,” in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015, pp. 639-643. (*Awarded SPS travel grant*)

[C5] **J. He**, and W. S. Gan, “Multi-shift principal component analysis based primary component extraction for spatial audio reproduction,” in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015, pp. 350-354. (*Awarded SPS travel grant*)

[C6] S. Fasciani, **J. He**, B. Lam, T. Murao, and W. S. Gan, “Comparative study of cone-shaped versus flat-panel speakers for active noise control of multi-tonal signals in open windows,” in *Proc. Internoise 2015 (invited)*, San Francisco, Aug. 2015.

[C7] **J. He**, and W. S. Gan, “Applying primary ambient extraction for immersive spatial audio reproduction,” *2015 Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference (invited)*, Hong Kong, Dec. 2015.

[C8] **J. He**, R. Ranjan, and W. S. Gan, “Fast continuous HRTF acquisition with unconstrained movements of human subjects,” in *Proc. ICASSP*, Shanghai, China, Mar. 2016, pp.

[Book] **J. He**, “Spatial audio reproduction with primary ambient extraction,” Monograph in preparation to submit to SpringerBriefs in Signal Processing.

[Tutorial] W. S. Gan, and **J. He**, “Assisted listening for headphones and hearing aids: signal processing techniques,” Tutorial at *APSIPA ASC 2015*, Hong Kong, Dec. 2015.

[Show & Tell] D. H. Nguyen, **J. He**, K. K. Phyo, and W. S. Gan, “Real-time audio signal processing platform for natural 3D sound rendering,” Show & Tell at *ICASSP 2016*, Shanghai, China, Mar. 2016.