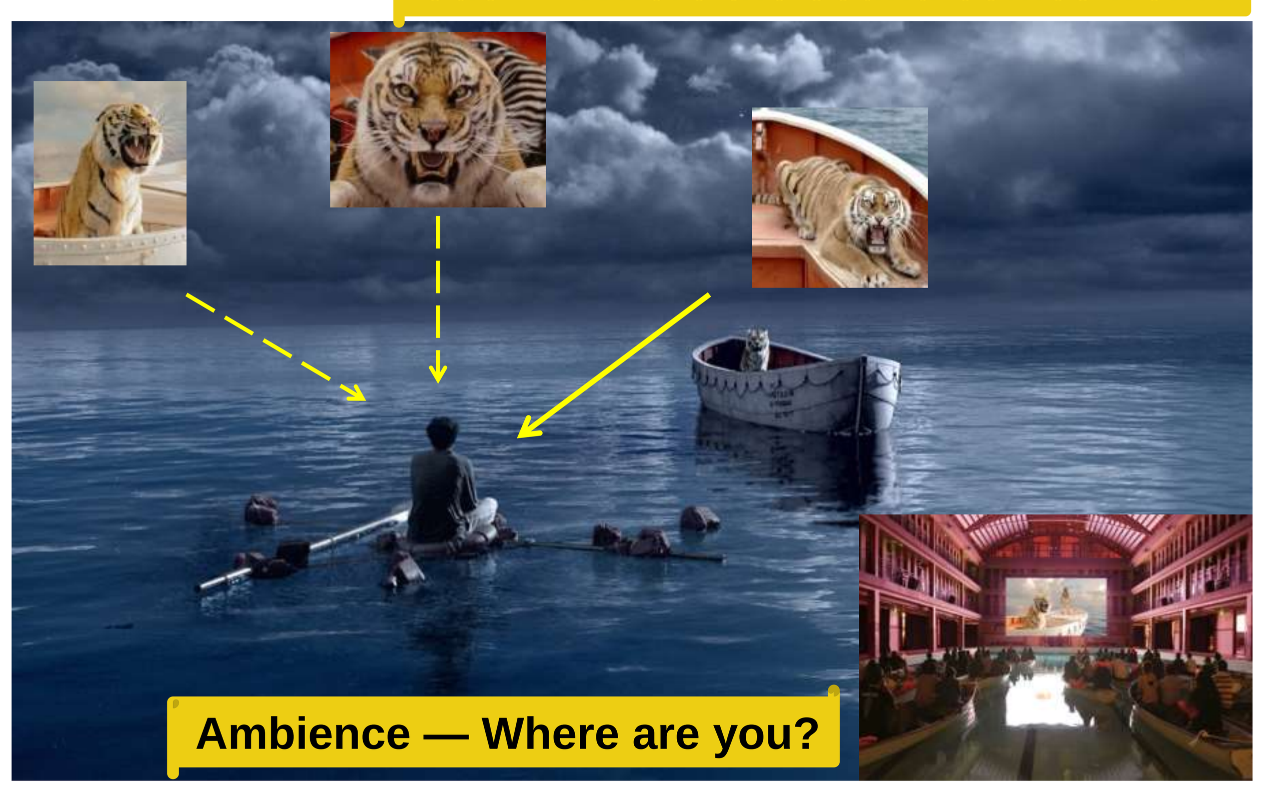
TIME-SHIFTED PRINCIPAL COMPONENT ANALYSIS BASED CUE EXTRACTION FOR STEREO AUDIO SIGNALS

Jianjun HE, Ee-Leng Tan, and Woon-Seng Gan

Digital Signal Processing Lab, Nanyang Technological University, Singapore {jhe007@e.ntu.edu.sg, etanel@ntu.edu.sg, ewsgan@ntu.edu.sg}

Motivation

- In spatial audio, one of the key issues is to decompose a signal into ambience and cue based on their spatial features that ambience is more diffuse while cue is more directional. The applications of ambience cue extraction include (but not limited to):
  - Spatial audio coding
  - Audio mixing
  - 3D audio systems
- Principal component analysis (PCA) has been widely employed in cue extraction from stereo signals .
- However, the performance of PCA based cue extraction is highly dependent on the assumptions of the input signal model, which are often unmet. For example, the stereo signal model does not take the time difference of the cue into consideration. But practically, cues are often time shifted.
- To overcome this problem, time shifted PCA (SPCA) is proposed in this paper by time-shifting the input signal according to the estimated inter-channel time difference (ITD) of the input signal before cue extraction using PCA.



Stereo Signal Model

Signal = Cue + Ambience

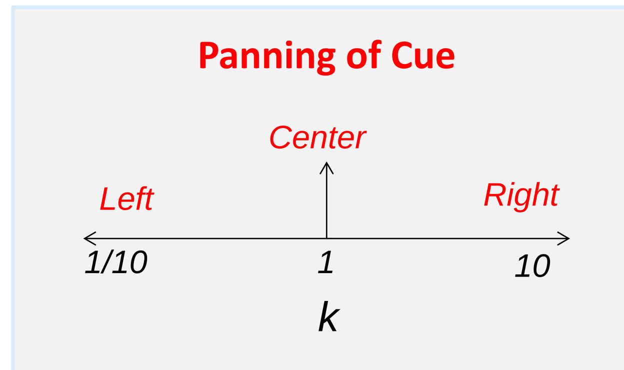
$$\vec{x}_L = \vec{c}_L + \vec{a}_L$$
$$\vec{x}_R = \vec{c}_R + \vec{a}_R$$

Cues highly correlated	$\vec{c}_R = k\vec{c}_L$
Ambience uncorrelated	$\vec{a}_L \perp \vec{a}_R$
Cue ambience uncorrelated	$\vec{c}_{L(R)} \perp \vec{a}_{L(R)}$
Ambience power balanced	$P_{\vec{a}_L} \approx P_{\vec{a}_R}$

Characterization

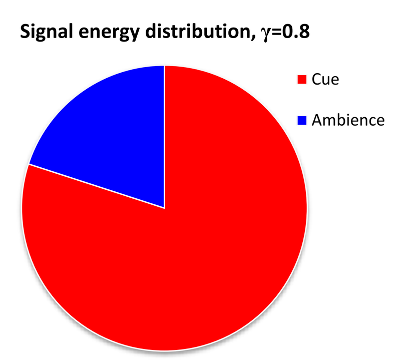
Cue panning factor CPF: $k = \frac{\vec{c}_R}{\vec{c}_L}$

$$k = \frac{r_{RR} - r_{LL}}{2r_{LR}} + \sqrt{\left(\frac{r_{RR} - r_{LL}}{2r_{LR}}\right)^2 + 1}$$



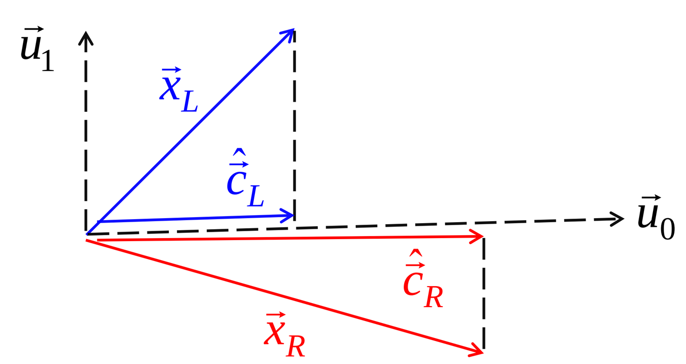
Cue energy ratio CER: $\gamma = \frac{\text{Cue energy}}{\text{Signal energy}}$

$$\gamma = \frac{2r_{LR} + (r_{RR} - r_{LL})k}{(r_{RR} + r_{LL})k}, \quad \gamma \in [0,1]$$



PCA based Cue Extraction

Objective: $\vec{u}_0 = \arg\max_{\vec{u}_0} \left(\left\| \vec{u}_0^H \vec{x}_L \right\|^2 + \left\| \vec{u}_0^H \vec{x}_R \right\|^2 \right)$



$$\lambda_0 = 0.5(r_{LL} + r_{RR} + \sqrt{(r_{LL} - r_{RR})^2 + 4r_{LR}^2}),$$
$$\vec{u}_0 = r_{LR}\vec{x}_L + (\lambda_0 - r_{LL})\vec{x}_R.$$
$$\hat{\vec{c}}_L = \frac{\vec{u}_0^H \vec{x}_L}{\vec{u}_0^H \vec{u}_0}\vec{u}_0, \quad \hat{\vec{c}}_R = \frac{\vec{u}_0^H \vec{x}_R}{\vec{u}_0^H \vec{u}_0}\vec{u}_0.$$

Results: $\hat{\vec{c}}_L = (1+k^2)^{-1}(\vec{x}_L + k\vec{x}_R), \ \hat{\vec{c}}_R = k\hat{\vec{c}}_L.$

PCA based Cue Extraction in Practical Case

Ideal case: $\vec{c}_L$, $\vec{c}_R$; $\theta_c = 0^\circ$, $\phi_c = 1$

Practical case: $\vec{c}_L$, $\vec{c}_R$; $\theta_c \neq 0^\circ$, $\phi_c < 1$

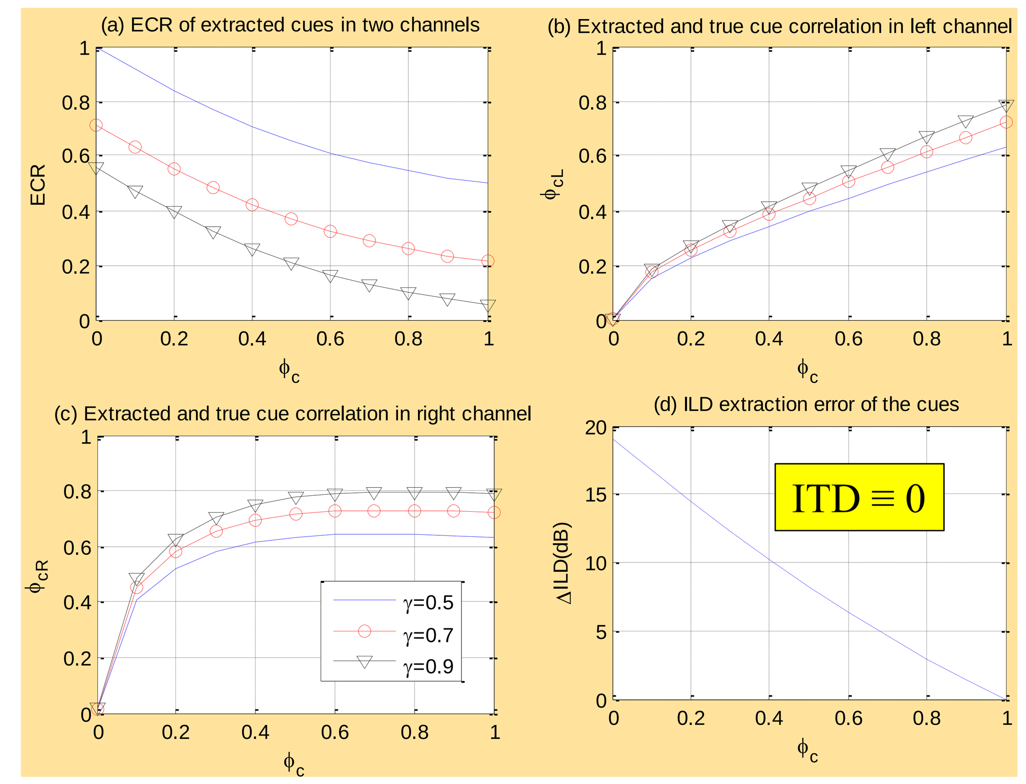
ϕ_c :Cue correlation

$$\hat{\vec{c}}_L = \vec{c}_L + \frac{1}{1+k^2}(\vec{a}_L + k\vec{a}_R) + \frac{k}{1+k^2}(\vec{n}_R - k\vec{n}_L)$$
$$\hat{\vec{c}}_R = \vec{c}_R + \frac{k}{1+k^2}(\vec{a}_L + k\vec{a}_R) + \frac{1}{1+k^2}(k\vec{n}_L - \vec{n}_R)$$

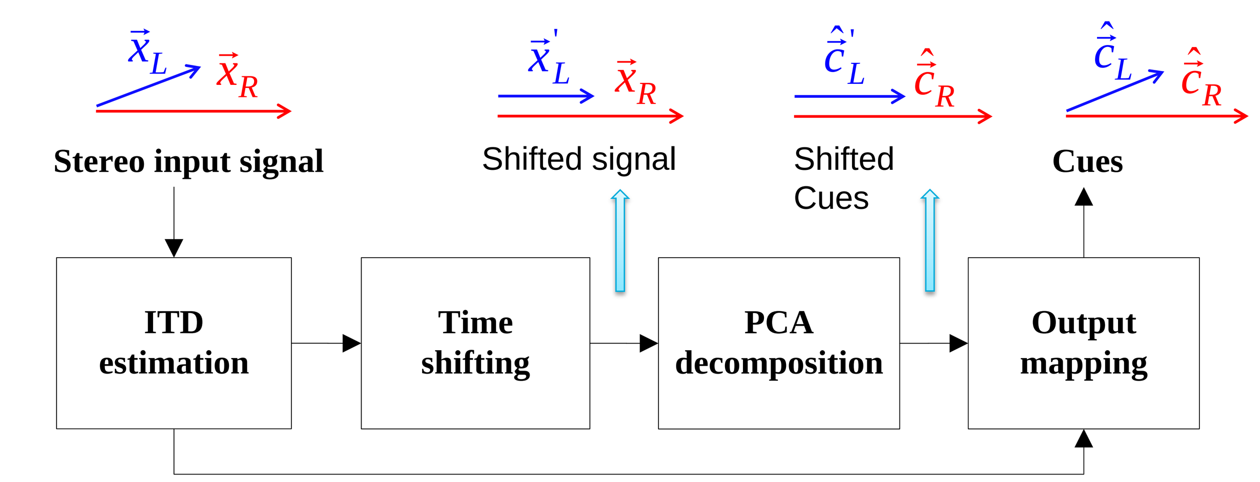
Error

Performance of PCA based cue extraction with varying cue correlation, k=3

- (a) ECR = Error power / True cue power
- (b-c) Extraction Similarity$_{L(R)}$ = Correlation $(\vec{c}_{L(R)}, \hat{\vec{c}}_{L(R)})$
- (d) Cue localization: ITD & ILD



Shifted PCA based Cue Extraction

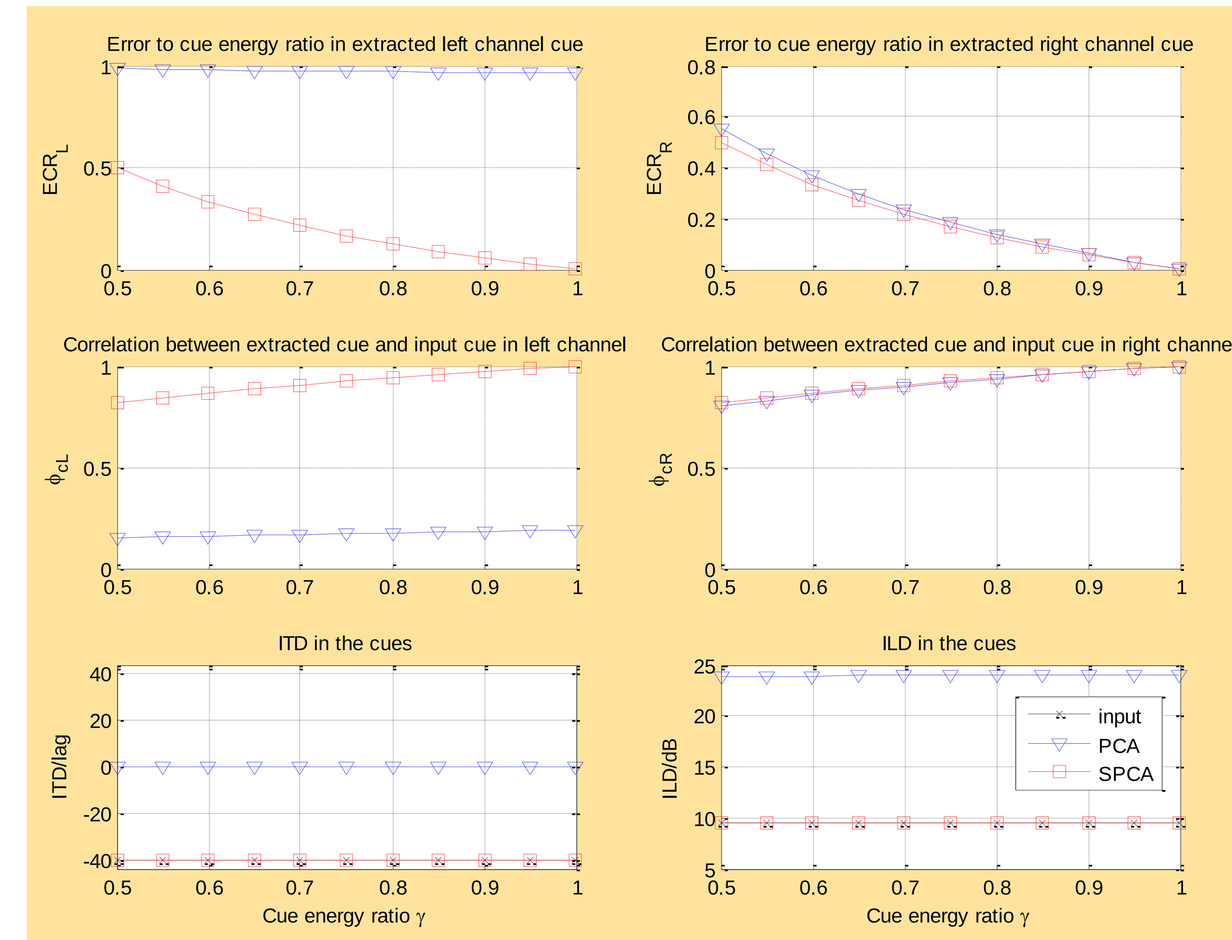


Block diagram and signal flow of shifted PCA based cue extraction.

Experimental results

Performance comparison of cue extraction using PCA and SPCA.

Synthesized signals:
- Cue: speech amplitude panned by 3 and shifted by 40 time units
- Ambience: uncorrelated white Gaussian noise
- Cue energy ratio CER: [0.5, 1]



SPCA	Extraction performance	PCA
Retained	ITD	Lost
Corrected	ILD	Exaggerated
Reduced	Error	High
Increased	Similarity	Low

CONCLUSIONS

In practice, cues in stereo signals can be time shifted as well as amplitude panned. As the input cues become partially correlated at zero lag, the performance of PCA based cue extraction degrades drastically. The proposed shifted PCA method overcomes the problem by strategically time-shifting input signals prior to PCA decomposition. This approach extracts cue having the correct ITD and ILD, and increases the similarity of the extracted cue to the original cue.