

$$\begin{aligned}
& - \frac{\partial \text{KL}(\mathbf{y}_{(\text{ICAL})}(t))}{\partial \mathbf{w}_{(\text{ICAL})}(n)} \cdot \mathbf{W}_{(\text{ICAL})}(z^{-1})^T \mathbf{W}_{(\text{ICAL})}(z) \\
& = \sum_{d=0}^{D-1} \left\{ \left(I\delta(n+d) - \left\langle \varphi\left(x\left(t - \frac{D}{2}\right) - \sum_{l=1}^{L-1} \mathbf{y}_{(\text{ICAL})}(t)\right) \cdot \left(x\left(t - n + d - \frac{D}{2}\right) - \sum_{l=1}^{L-1} \mathbf{y}_{(\text{ICAL})}(t - n + d)^T\right) \right\rangle \right) \cdot \left(I\delta\left(d - \frac{D}{2}\right) + \sum_{l=1}^{L-1} \mathbf{w}_{(\text{ICAL})}(d) \right) \right\}
\end{aligned}$$

学術フロンティア講義

インテリジェントな認識と行動のシステム科学 ～信号処理: 複雑な物理現象からの宝探し～

$$f(x) = x^{k-1} \frac{e^{-x/\theta}}{\Gamma(k)\theta^k}$$

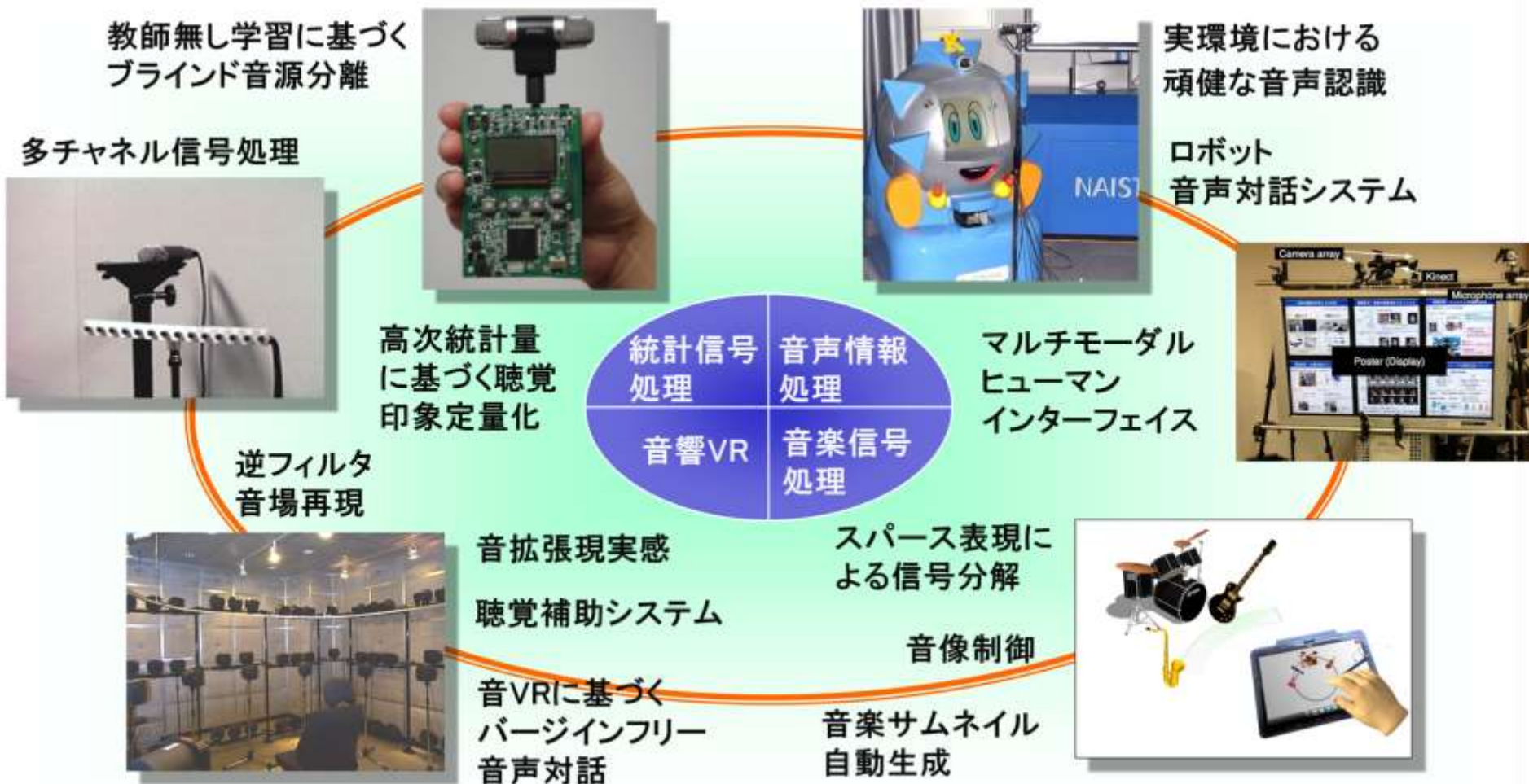
東京大学大学院情報理工学系研究科/計数工学科

システム情報学専攻 第一研究室 教授

猿渡 洋

システム情報第一研の研究俯瞰図

- 音声・音響・音楽メディアに関する信号処理・情報処理
- ヒューマンインターフェイス・コミュニケーションシステムの構築
- 統計的・機械学習論的信号処理、数理最適化問題等を研究



代表的な成果 –いくつかは「世界的に」有名です–

ブラインド音源分離 (BSS)・音楽信号解析

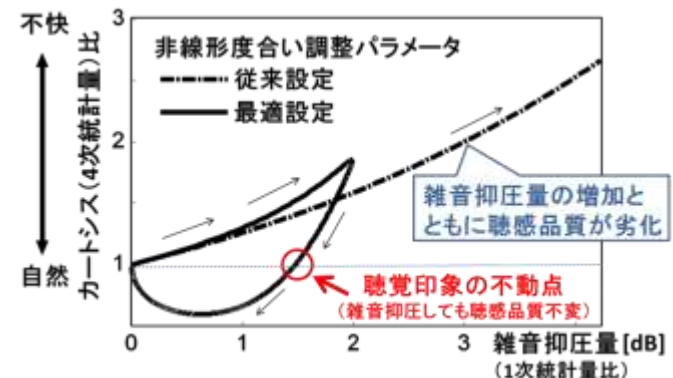
- 独立成分分析(ICA)に基づくBSS(2007年世界コンテスト一等賞、2008年警察備品採用・ICAを実用化した国は世界でも日本だけ)
- BSS補聴器(EUSIPCO2012招待講演)
- 低ランクモデルを導入した音楽信号解析(ASJ粟屋潔賞・育志賞)

高次統計量追跡に基づく非線形システム最適化

- 聴覚定量化理論(市村賞、IEEE JP 学生賞、エリクソン賞、ASJ板倉賞等)

音場分解・再現・音バーチャルリアリティ

- 超解像に基づく音場再現(ASJ板倉賞、TAFテレコム技術賞等)
- スパース信号表現に基づく新しい音場解析(ASJ学生優秀発表賞)



こんな機械に知性を感じる？



君の名前は何？
何ができるの？



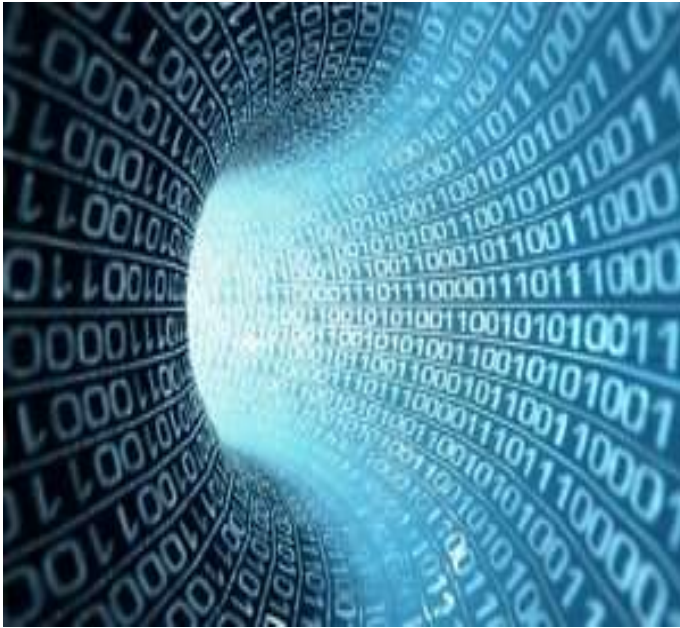
えっ、人間型
ロボットなのに
スイッチで入
力するの？

音メディアの重要性

- **音声は人間が持つ最も原始的なコミュニケーション手段**
 - ・ 特別な道具を必要としない誰でも使える意思伝達手段
 - ・ 人間同士もしくはコンピュータやロボットとのコミュニケーション手段の一つとして昔から注目されてきた
- **人間と人間のインタラクションにおける音声**
 - ・ 音声符号化⇒携帯電話等にて既に実用化
- **機械と人間のインタラクションにおける音声**
 - ・ 知性を感じさせる機械を実現するには音声対話メディアが必須である
 - ・ 必要とされる機能
 - ・ 必要な音を聞き分ける(雑音抑圧・音源分離)
 - ・ ユーザの声を言語として認識理解する(音声認識)
 - ・ ユーザに情報を発信する(音声合成)

音のビッグデータ・IoT？

時代はビッグデータ！
しかし実際の音波動データは
簡単に集められるのか？



むしろ「**スモールデータ(一期一会データ)**」の研究が必要
ではないか？

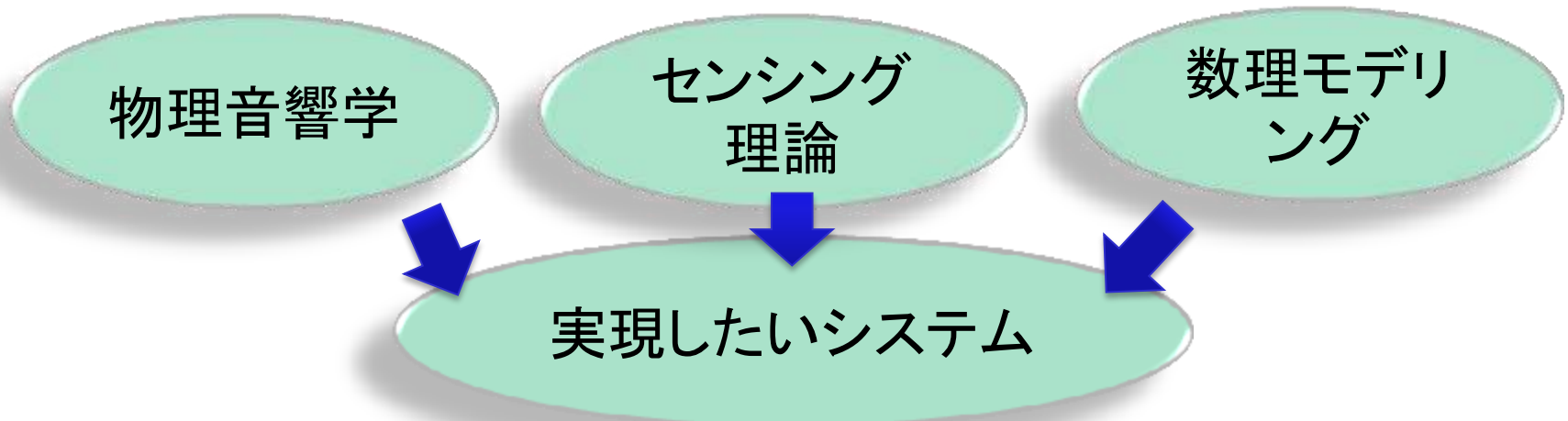
全てのセンサはネットワーク
に繋がり「トリリオン(一兆個)
センシング」も登場。



しかし**ばらまかれた音センサ**
は位相情報を失っているので
活用できない！

音メディアに関する信号処理研究の魅力

- 音メディアに関する信号処理研究の魅力とは？
 - 自然界の音が持つ無限の**多様性** (cf. 無線通信信号)
 - 研究のアプローチに**多面性**あり (決定論的？統計的？)
 - 最後は聴かせてなんぼの評価 ⇒ **芸術性**も併せ持つ
- 「**物理世界(波動)と情報世界(抽象)をまたぐ学問**」であり、かつそれを「**統一的に取り扱うシステム工学**」である。
- 対象の多様性ゆえに「**なんでもあり**」の分野でもある。



音メディアに関する信号処理研究の魅力

波動方程式

室内音響

伝達関数

音生成過程 etc.

物理音響学

離散サンプリング

フーリエ解析

球面調和解析

圧縮センシング etc.

センシング
理論

統計モデリング

最尤・ベイズ推定

機械学習

最適化問題 etc.

数理モデリ
ング

実現したいシステム



本日の講義内容

時系列メディア処理における「音源分離」の紹介

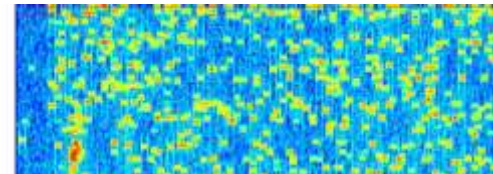
独立成分分析

聞き分けの出来る人工耳



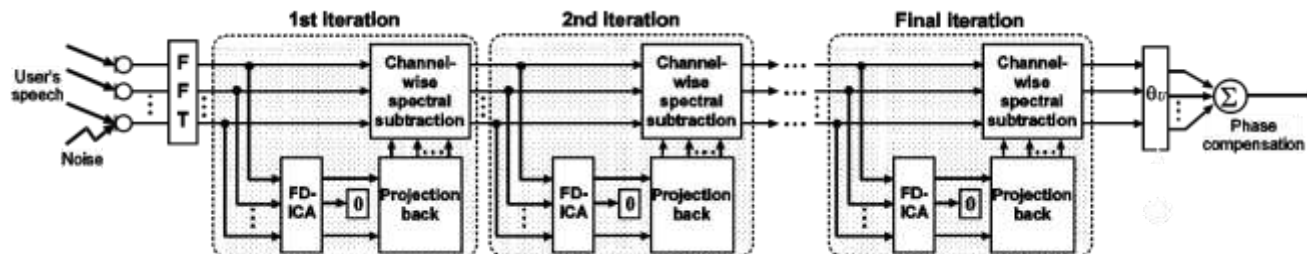
高次統計量音質制御

気持ち悪さを測る



スパース表現信号処理: その威力と活用例

音楽信号処理への応用



本日の講義内容(隠れテーマ)

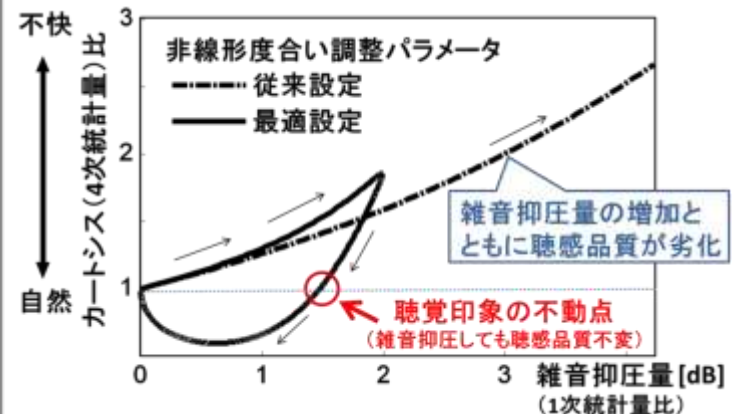
信号処理研究におけるセレンディピティの紹介

- ・セレンディピティとは「偶然に大発見する幸運」の意味
- ・理学と異なり工学では「分かっているもの」を実装することが多いのでセレンディピティを体験することは多くない

車の中で全く動かない
独立成分分析から、
非線形音源分離へ



偶然見つけた4次統計量の
不動点と、ミュージカル
ノイズフリー理論



本日の講義内容

時系列メディア処理における「音源分離」の紹介

独立成分分析

聞き分けの出来る人工耳



研究紹介1. ブラインド音源分離

- 音の方向・声質・音量など、事前に何も分かっていなくても、瞬時に音を「聞き分ける」ことの出来るシステムを目指す。
- 独自に開発した高速独立成分分析(ICA)、独立低ランク行列分析(ILRMA)という教師無し数理最適化アルゴリズムに基づいて、音を統計的に独立な成分に分解することにより、別々の音声信号を見つける。

スモールデータ
教師無し最適化

音を聞き分ける耳：マイクロホンアレー

実際、人間も2つの耳で聞くことによって、
音の方向や複数音の聞き分けを行っている

■ 音声処理での一例：マイクロホンアレー

- 複数のマイクによって得られた複数の受信信号のなかから、必要な情報(目的音声)のみを取り出す装置



期待される応用

- 高性能な hands-free 通信
- 雑音にロバストな音声認識

⇒ **ではどういうアルゴリズム
(ソフト)が必要なのか？**

音を聞き分ける耳を作りたい

■ Blind Source Separation (BSS)

- 複数音源信号が混合されている場合、観測信号のみから音源信号を自律的に推定する技術
- 目的音の方位・無音区間情報が不要
- マイク素子位置・特性情報も不要

■ 独立成分分析(ICA)に基づくBSS



1989 J. Cardoso

1990 C. Jutten (高次無相関化)

1994 P. Comon (ICAという言葉进行定義)

1995 A. Bell (infomaxによる定式化)

第一世代

1998 P. Smaragdis, S. Ikeda, **H. Saruwatari** ...
(音響信号へICAを導入)

第二世代

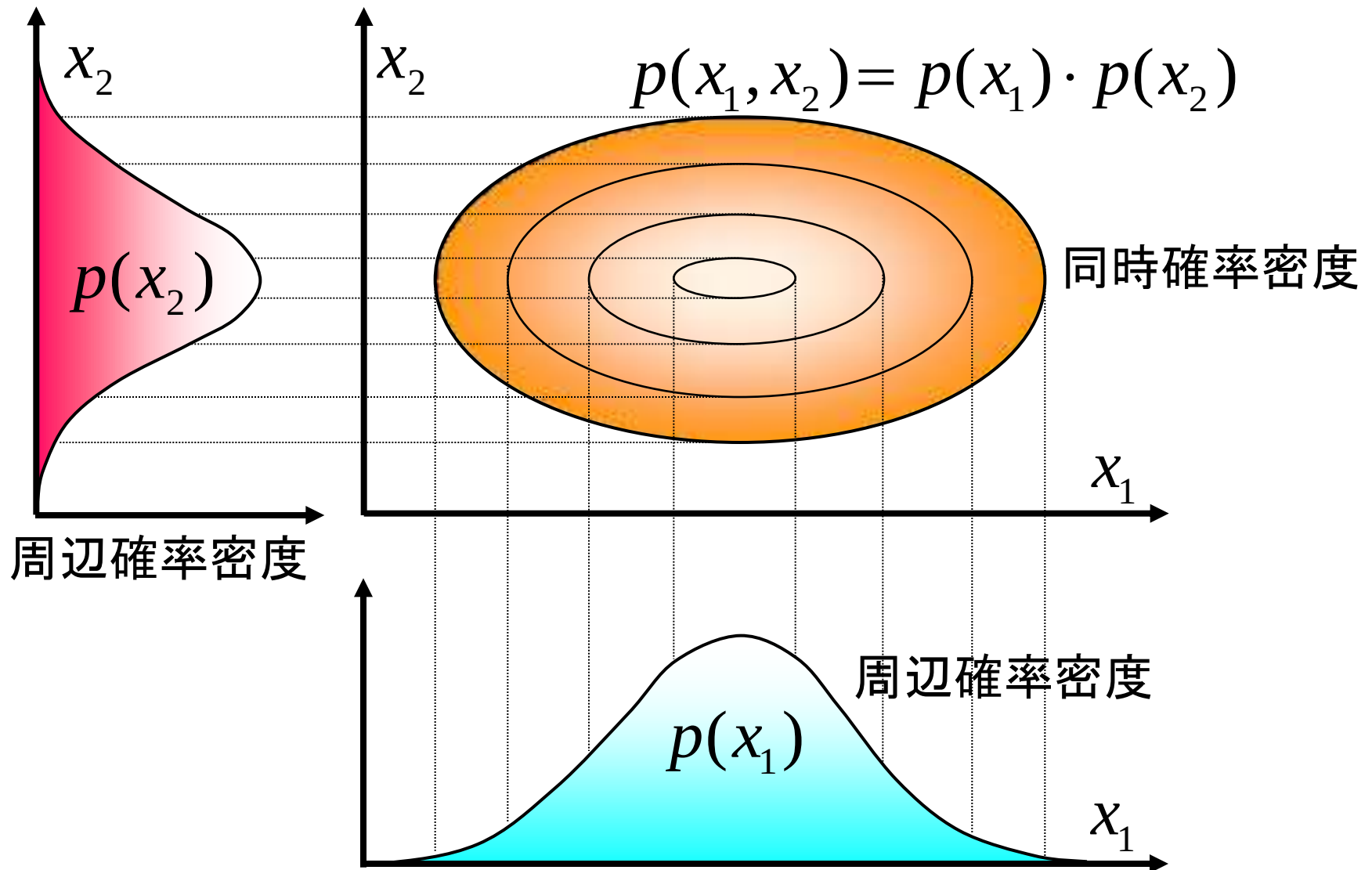
「独立」とは何か？

■ 数学における「独立」の定義：

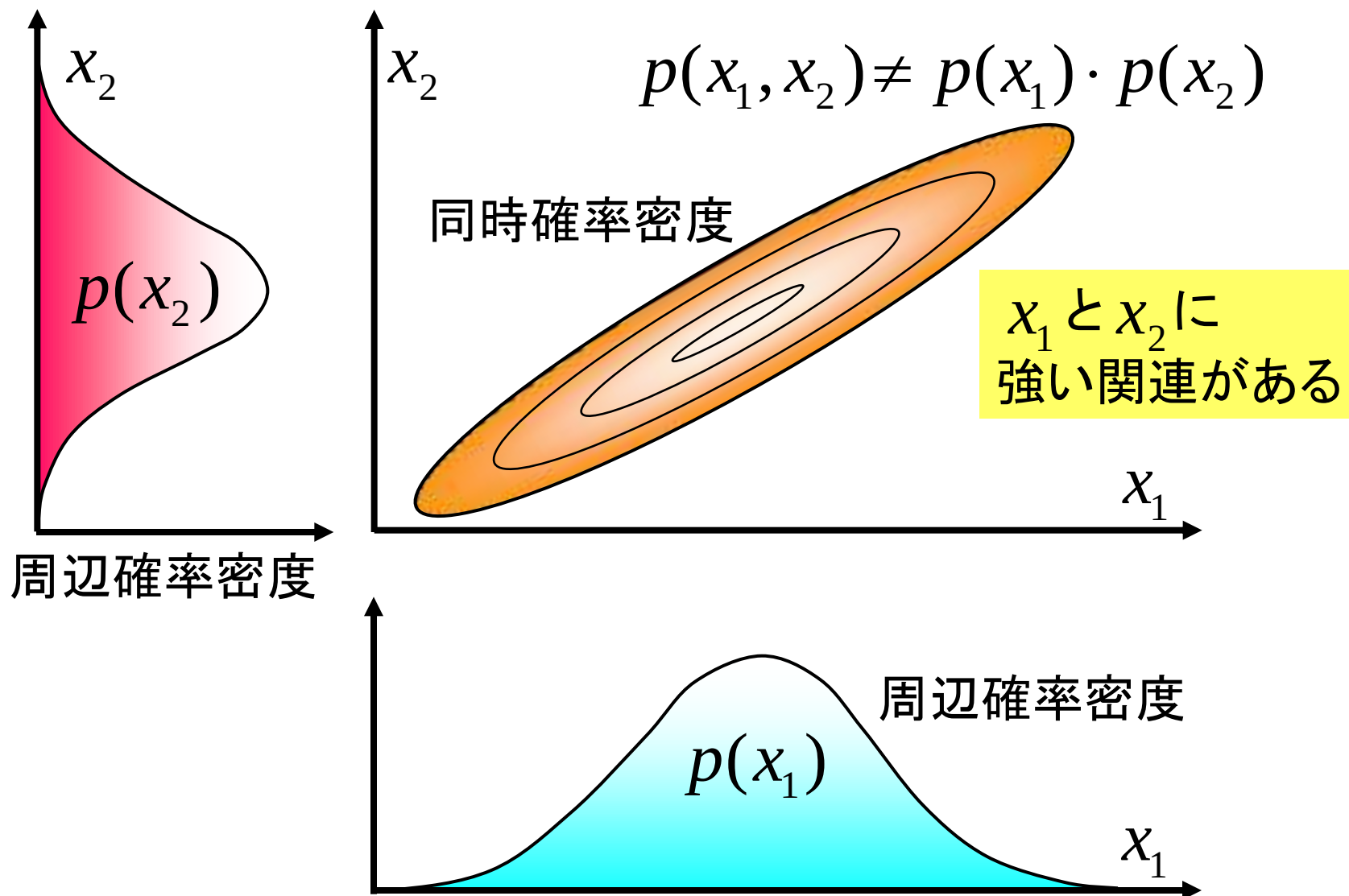
- 2つの確率事象に関する同時確率密度分布 $p(x_1, x_2)$ が、それぞれの事象における周辺密度分布 $p(x_1)$, $p(x_2)$ の積で書ける場合を「**(統計的に)独立**」と呼ぶ.
- つまり

$$\text{独立} \iff p(x_1, x_2) = p(x_1) \cdot p(x_2)$$

独立である場合の例



独立ではない場合の例



独立成分分析(ICA)とは何か？

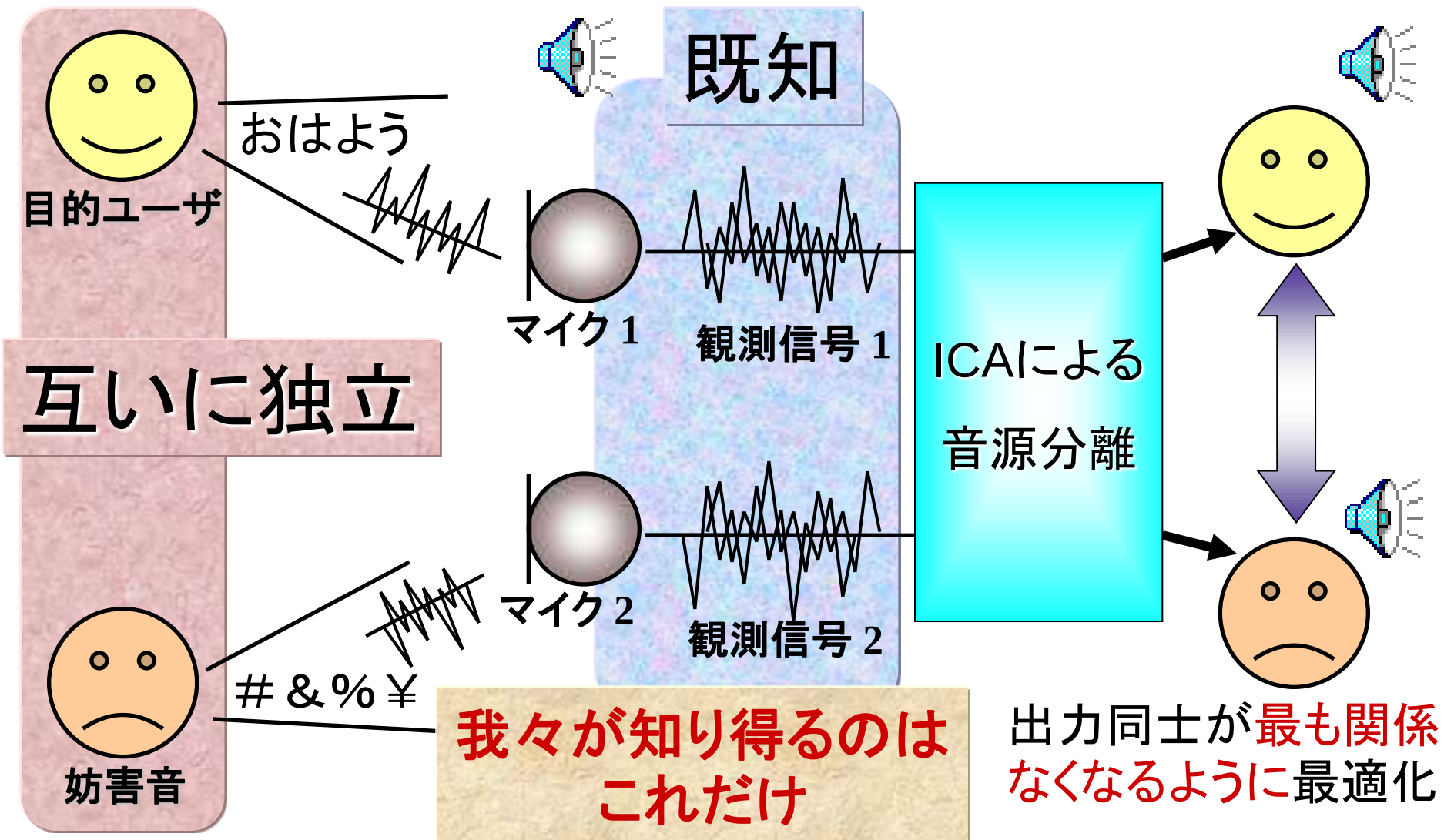
■ 独立な成分の抽出：

- 複数の確率信号が混合された観測系列から、統計的に独立な個々の確率過程を分解する.

■ 特徴：

- 独立性は「無相関性」よりも厳しい尺度であり、確率信号同士の確率密度構造の幾何が問われる. よって「**情報幾何学**」とも呼ばれる.
- 確率密度の構造を測るために、3次以上の統計量が必要とされる. よって統計の分野では、「**高次統計量数理**」の一種でもある.
- 決定論的な目標値を与えずに最適化を行うことより、学習理論の分野では「**教師無し学習**」とも呼ばれる.

ICAに基づくBSSとは？



ICAに基づくBSSの定式化

線形混合過程

$$\begin{bmatrix} A_{11} & \Lambda & A_{1K} \\ M & O & M \\ A_{L1} & \Lambda & A_{LK} \end{bmatrix} \cdot \begin{bmatrix} s_1(t) \\ M \\ s_K(t) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ M \\ x_L(t) \end{bmatrix}$$

混合行列 音源信号 観測信号

分離過程

コスト関数

独立?

$$\begin{bmatrix} y_1(t) \\ M \\ y_K(t) \end{bmatrix} = \begin{bmatrix} W_{11} & \Lambda & W_{1L} \\ M & O & M \\ W_{K1} & \Lambda & W_{KL} \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ M \\ x_L(t) \end{bmatrix}$$

分離信号 分離行列 観測信号

最適化

ICAにおける様々なコスト関数

分離信号ベクトル: $\mathbf{y}(t) = [y_1(t), \dots, y_2(t)]^T$

■ 2次統計量

$$E\{\mathbf{y}(t)\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 信号間相関を最小化 (複数時間区間利用)

■ 高次統計量1

$$E\{\mathbf{y}^3(t)\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 高次相関をも最小化

■ 高次統計量2

$$E\{\Phi(\mathbf{y}(t))\mathbf{y}^T(t)\} \rightarrow \text{diag}$$

- 源信号確率密度関数を仮定 $\Phi(\bullet)$: tanh関数など

非線型関数2の導出

- 独立⇒Kullback Leibler Divergenceの最小化問題
 - 一般にKullback Leibler Divergenceとは2分布間の距離

$$KL(\mathbf{v}, \mathbf{z}) = \int p(\mathbf{z}) \log \frac{p(\mathbf{z})}{p(\mathbf{v})} d\mathbf{z}$$

上式において...

$p(\mathbf{z}) = p(y_1, \Lambda, y_K)$ 分離信号 $y(t)$ の同時分布密度関数

$p(\mathbf{v}) = \prod_{k=1}^K p(y_k)$ 周辺分布密度関数の積

とおき, これらのKLを分離行列 W に関して最小化すれば独立

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{k=1}^K p(y_k)} d\mathbf{y} \quad \Rightarrow \quad \text{最小化}$$

非線型関数2の導出 (cont'd)

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{k=1}^K p(y_k)} d\mathbf{y}$$

$$= -H(\mathbf{Y}; W) + \sum_{k=1}^K H(Y_k; W)$$

1. 結合エントロピー

2. 周辺エントロピー和

$$H(\mathbf{Y}; W) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$$

$$= -\int p(\mathbf{x}) (\log p(\mathbf{x}) - \log |\mathbf{W}|) d\mathbf{x}$$

$$= H(\mathbf{X}) + \log |\mathbf{W}| \quad (\ominus p(\mathbf{y}) = p(\mathbf{x}) / |\mathbf{W}|)$$

$$H(Y_k; W) = -\int p(\mathbf{y}) \log p(y_k) d\mathbf{y}$$

$$= -\int p(\mathbf{x}) \log p(y_k) d\mathbf{x}$$

$$(\ominus p(\mathbf{x}) d\mathbf{x} = p(\mathbf{y}) d\mathbf{y})$$

非線型関数2の導出 (cont'd)

$$KL(W) = \int p(y) \log \frac{p(y)}{\prod_{k=1}^K p(y_k)} dy$$

$$= -H(Y; W) + \sum_{k=1}^K H(Y_k; W)$$

1. 結合エントロピー

2. 周辺エントロピー和

$H(Y; W)$ = この値を最大化

= $\Rightarrow$ 音源間の関連を無くす

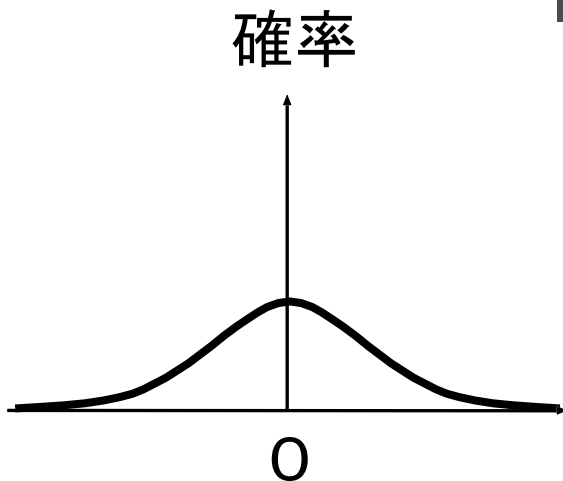
=

$H(Y_k; W)$ = この値を最小化

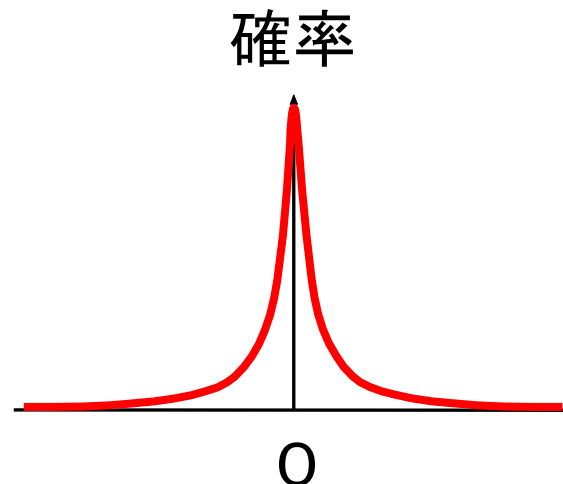
= $\Rightarrow$ 個々の音を非ガウス化

非線型関数2の導出（非ガウス化？）

確率密度関数(PDF)形状



ガウス分布(正規分布)



スーパー(優)ガウス分布
(より尖った形状)

- ・振幅の大きな信号は滅多に出現しない
- ・ほとんど小振幅

非線型関数2の導出 (cont'd)

分離信号の同時確率密度と周辺確率密度積のKL擬距離 $KL(W)$ の W に関する勾配を求め、その逆方向に W を更新学習

$$\Delta \mathbf{W} \propto -\frac{\partial KL(\mathbf{W})}{\partial \mathbf{W}} = \left((\mathbf{W}^T)^{-1} - \int p(x) \phi(y) \mathbf{x}^T dx \right)$$

$$= \left((\mathbf{W}^T)^{-1} - E_x [\phi(y) \mathbf{x}^T] \right)$$

$$= (\mathbf{I} - E_y [\phi(y) \mathbf{y}^T]) (\mathbf{W}^T)^{-1}$$

$$\phi(y) \equiv \left[\frac{\partial \log p(y_1)}{\partial y_1}, \dots, \frac{\partial \log p(y_K)}{\partial y_K} \right]^T$$

音声の場合はSigmoid関数で近似可能

様々なバリエーション

- EMアルゴリズムによる $p(y)$ の同時推定
- 二次統計量による $p(y)$ の推定 + 高次統計量ICA [Saruwatari, ICASSP2009]
- ICAと音響信号処理の類似点を明らかにし、相補性を生かした高速収束アルゴリズム [Saruwatari, IEEE Trans. SAP 2003 & 2006]

教師無し最適化としてのICA

■ 従来の教師有り最適化：目標値が与えられる

- 子育てで言えば「医者にするにはどうするか？」
- 最小化関数 = $\lceil (\text{医者} - \text{子供の現在})^2$
- コスト関数が可計算、その微分勾配も可計算



■ ICA等の教師無し最適化：目標値が無い！

- 子育てで言えば「良い大人になりなさい！（でも具体的には？）」
- 最小化関数 = $\text{div}(??? || \text{子供の現在}) \Rightarrow$ 陽に計算不可
- コスト関数は可計算ではないが、その勾配は可計算
- 独立性自体をデータから測ることは困難であるが、どの方向に動けば独立性が高まるかは計算可能
- 「勉強しよう」、「運動しよう」、「約束は守る」、「友達は大切に」...

「聖徳太子」マイクの実用化

- 実際に音源分離をやってみましょう



ワシは10人
聞き分けれ
るぞよ。



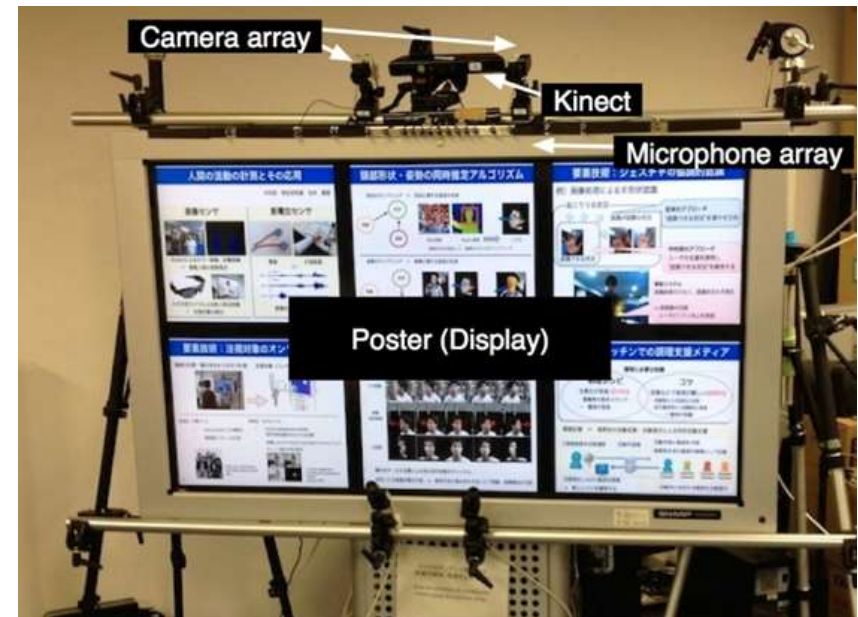
BSSを利用したアプリケーション

- 世界で初めてリアルタイムBSSモジュールが商用化され、2008年には警察備品として採用された。

- ドコモモバイルサイエンス賞



- 京大NAIST-CRESTプロジェクトにて「場の雰囲気を読むポスタセッションアーカイブシステム」に導入された。



ポスタ会場



発表者



質問者

BSSを利用したアプリケーション

質問者の質問の数や

「へえ～」「ふーん」

の数を数えることにより、

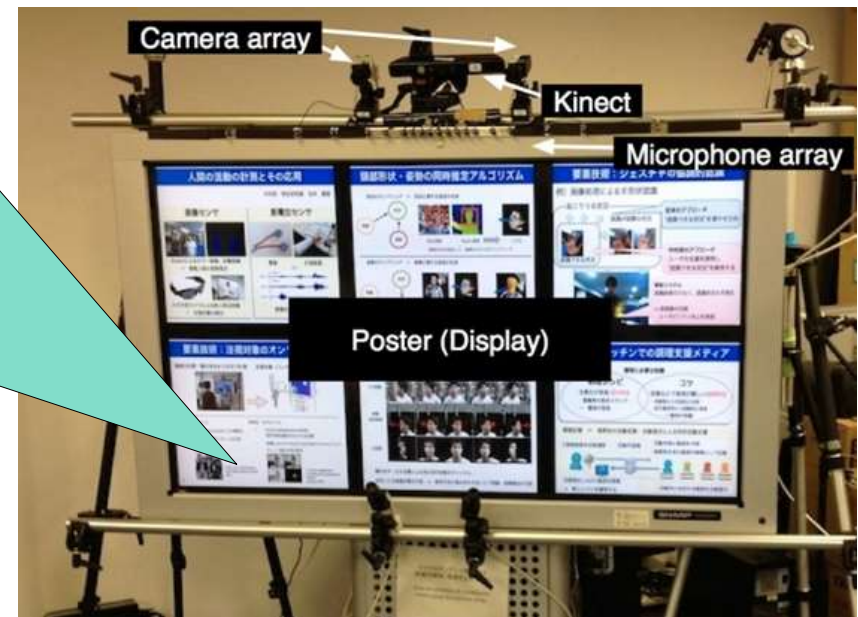
場の盛り上がりや

コミュニケーションの上達

度を測ることができます。



- 京大NAIST-CRESTプロジェクトにて「場の雰囲気を読むポスタセッションアーカイブシステム」に導入された。



ポスタ会場



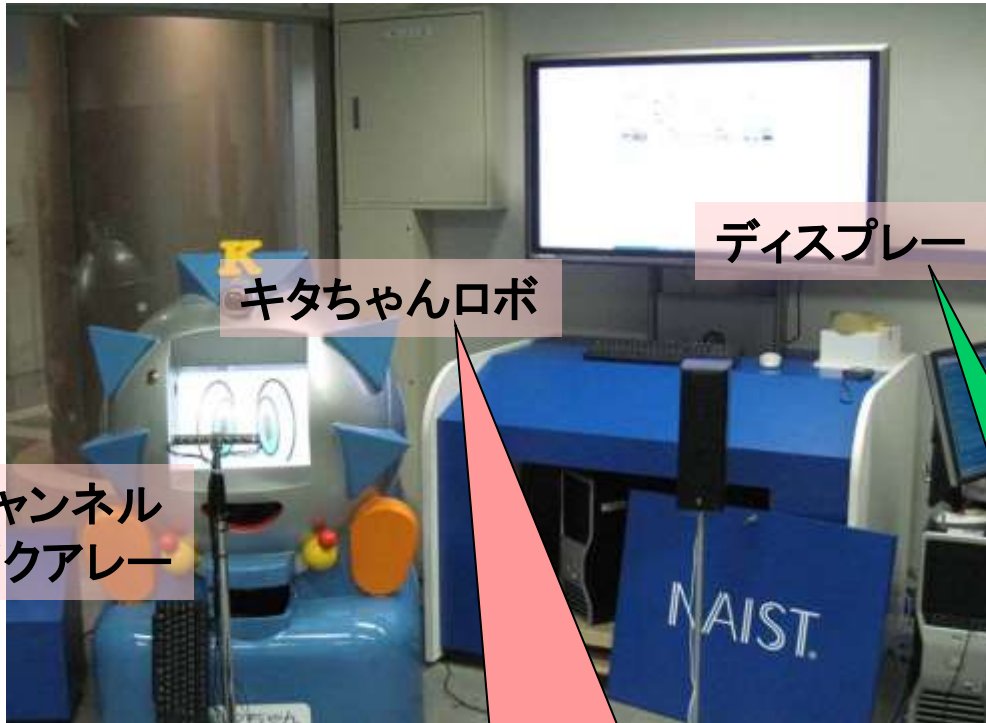
発表者



質問者

ハンズフリー音声対話ロボシステムの構築

- 各種モジュールを統合した実環境動作可能な音声対話システム



典型的な駅騒音の中で
遠隔発話した場合でも

90%以上
の単語認識率を達成

リアルタイム
ブラインド空間
サブトラクションアレー

音声発話検出 &
音声認識デコーダ

対話管理処理

応答音声生成

各種情報提示処理

(今後)
ロボット動作
との連携

ハンズフリー音声対話ロボシステムの構築

ロボット対話 デモンストレーション

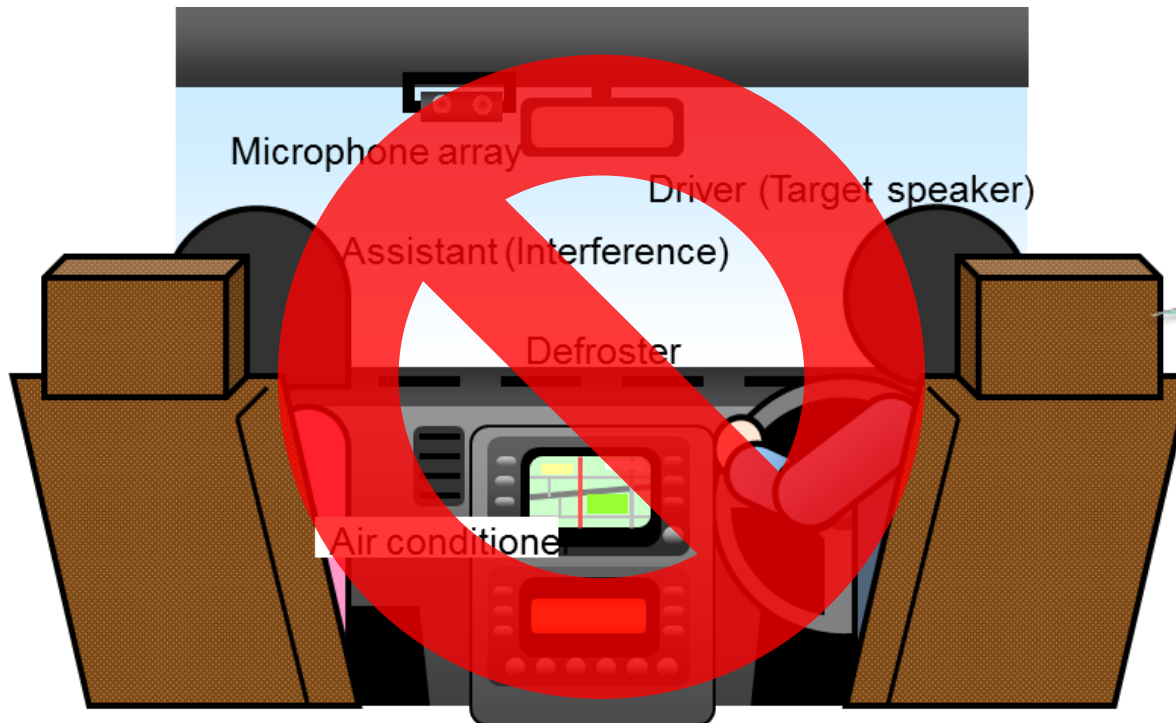
Demonstration for
spoken-oriented hands-free
interactive robot.

人工知能研究会優秀賞・平成23年度 RIEC Award を受賞

BSS開発におけるセレンディピティ

■ 実はICAは**実環境**で全く動かなかった！

- ある自動車株式会社との共同研究
- 自動車内の雑音・妨害話者音とドライバーの声を分離・認識
- プロトタイプシステムを製作して偉い方々へお披露目したが...



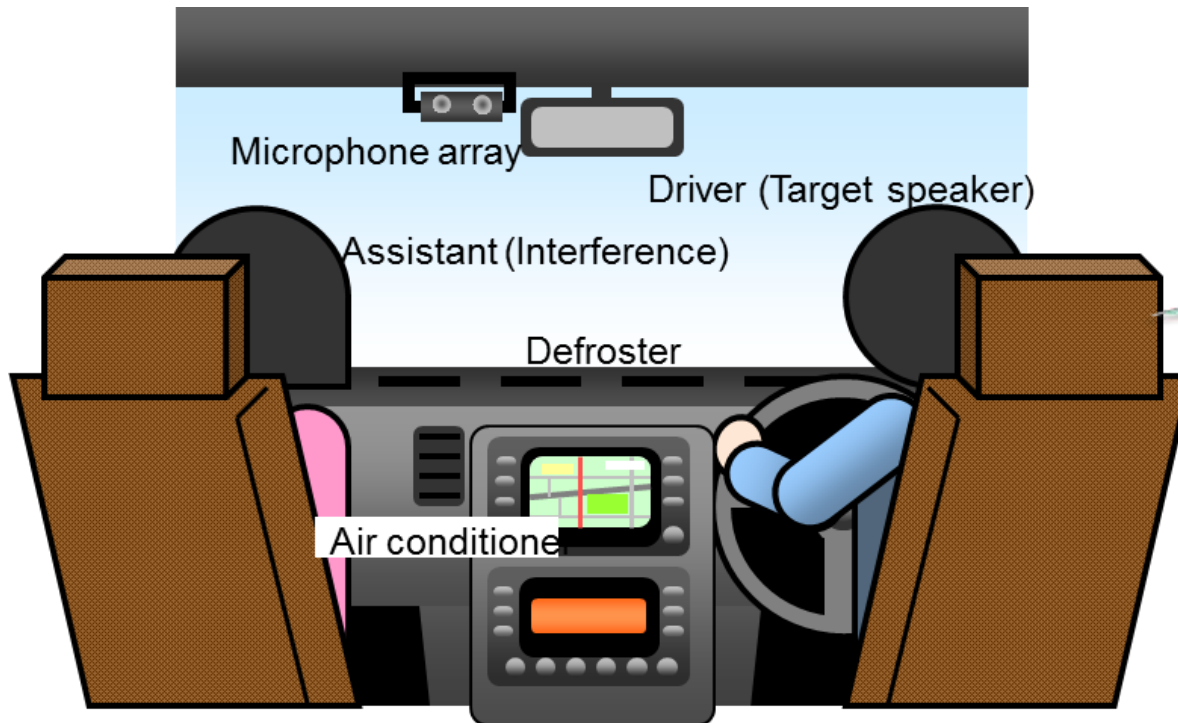
なんだ、全然雑音が減らないじゃないか！



BSS開発におけるセレンディピティ

■ 偶然、ICAの「音声出力」ではない方を流してしまった

- ドライバーが消えた！（と思えるほど音声が消えた）
- 車室内の雑音しか聞こえない！
- 「なんだ、ICAは噂ほどの実力はないんだな...」

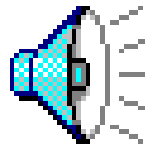


ドライバーの声が
消えた！

従来のICAは実環境雑音を消せるか？

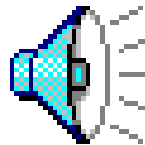
ICAによる音声 & 実環境雑音(掃除機)分離結果

観測信号



Speech & cleaner noise

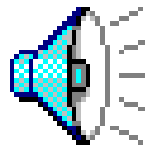
音声推定音



× Bad!

(雑音が残留している)

雑音推定音



○ Good!

(音声が消えている)

ICA は音声推定よりも雑音推定が得意！

BSS開発におけるセレンディピティ

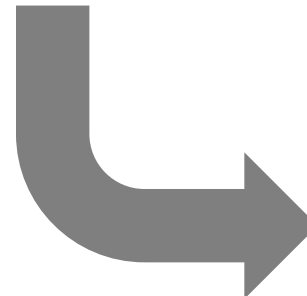
■ 実はICAは「雑音推定の方が得意」だった

- 当時は誰もがICAの音声出力の品質を上げる努力をしていた
 - 「実環境では雑音数が多いからダメなんだ」
 - 「マイク数を増やしてICAの次元を増やせば
音声分離はうまくいくんだ」



- しかし理論解析を重ねるうち、雑音側出力の精度が非常に高いことが数学的に証明された(これなら例え2chでも稼働する)
- 「皆はICAの使い方を間違っている。本当は**音声消去**装置だ」

非線形処理に基づく
音声抽出の考案



本日の講義内容

時系列メディア処理における「音源分離」の紹介

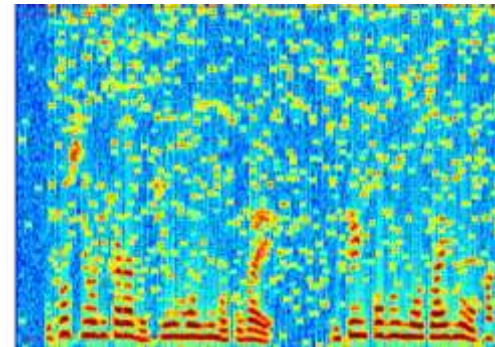
独立成分分析

聞き分けの出来る人工耳



高次統計量音質制御

気持ち悪さを測る



研究紹介2. 気持ち悪さを測る

- 各種の統計的音声推定を行う場合、非常に不愉快な人工雑音が残留し、出力音の「聴感的な印象」を下げてしまう。
- 統計推定手法ごとに「聴感的印象」は異なる。つまり、統計的推定には「音の個性」がある。我々は芸術的観点から統計的推定問題を眺める。
- 聴感的印象を数値化し、その値が不動となるパラメータの存在を世界で初めて発見した。

非線形システム
高次統計量解析

非線形処理の問題:アーチファクトの発生

- ◆ 一般に、非線形雑音抑圧信号処理において、不快なアーチファクト(これは**ミュージカルノイズ**と呼ばれる)が発生する。これは「人が聴く」用途へ適用する際に、大きな問題となってしまう。
- ◆ 信号処理を改良することで多少のミュージカルノイズ軽減は可能である。また、統計推定手法ごとに「聴感的印象」は異なる。つまり、統計的推定には「音の個性」がある。我々は**芸術的観点**から統計的推定問題を眺める。
- ◆ 本来、ミュージカルノイズに関しては、数理解析がほとんどなされていないという現状がある。よって、
 - (1) まずミュージカルノイズの定量指標を定める必要がある。
 - (2) 次に、ミュージカルノイズを低減する信号処理を開発する。

統計推定における音色の差を聞いてみよう！

白色ノイズの場合

観測音



最尤推定



ベイズ推定



ミュージカル
ノイズフリー



人ごみノイズの場合

観測音



最尤推定



ベイズ推定



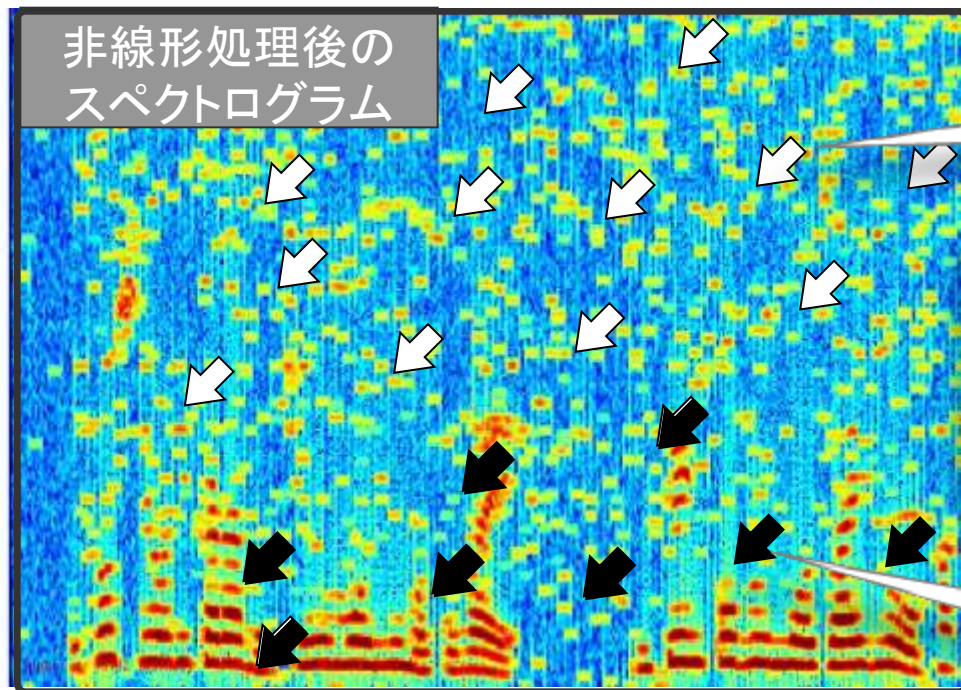
ミュージカル
ノイズフリー



どの推定方式が「自然」だと感じましたか？

ミュージカルノイズの原因成分

- ミュージカルノイズ＝「トーンを感じる歪み」
- トーンは、音声・音楽信号に豊富に含まれ、スペクトログラム上にて卓越成分があると、人は「トーンとして感じる」
- 音声・音楽信号以外で卓越した成分＝ミュージカルノイズ



雑音部分

ミュージカルノイズの評価
＝卓越成分の評価

音声部分

高次統計量の導入

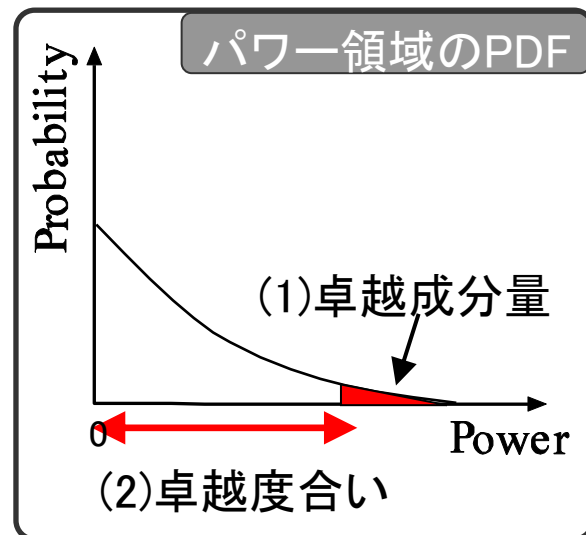
■ カートシス (尖度)

- 4次統計量に基づく「PDF裾野量尺度」
- 卓越成分の数値評価が可能となる

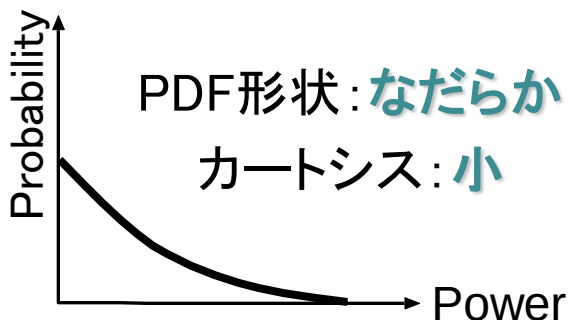
$$\text{kurtosis} = \frac{\mu_4}{\mu_2^2} \quad \mu_n \text{ はPDF } P(x) \text{ の } n \text{ 次モーメント}$$

$$\mu_n = \int_0^{\infty} x^n P(x) dx$$

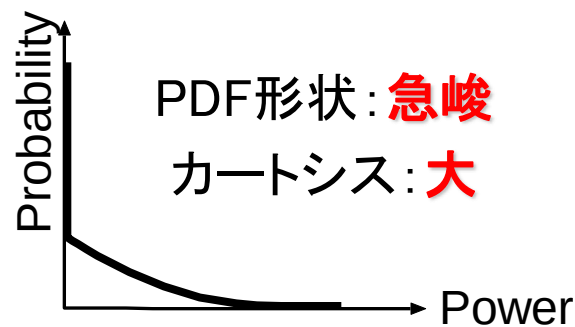
※ 中心化モーメントではないため厳密な尖度とは異なる



PDF形状とカートシス



卓越成分が少ない

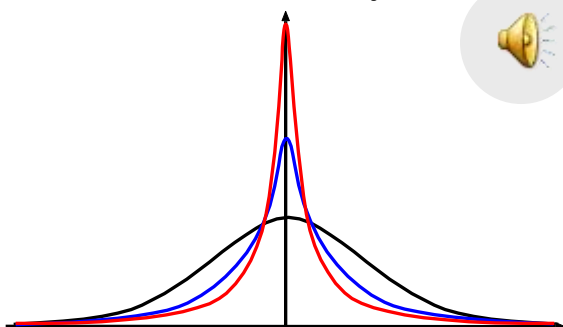
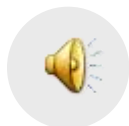


卓越成分が多い

音色と PDF の関係

PDF 形状

Probability



ガウス分布



スーパーガウス分布（やや尖った形状）



スーパーガウス分布（より尖った形状）

と信号の従う PDF 形状を変化させて
作成した音源.

3秒ごとに、より急峻な分布に変わる.

※ 非線形処理は一切行っていない

ミュージカルノイズはカートシスと強く関連していることが分かる

カートシス比 [IWAENC2009]

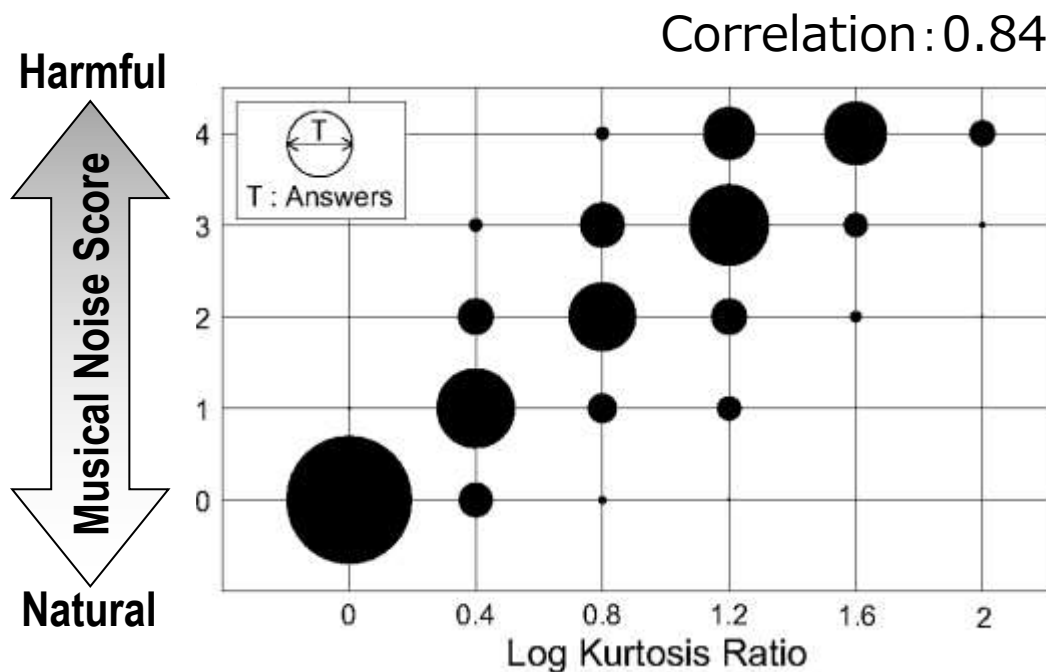
■ 処理前後のカートシス変化比（源信号尖度で正規化）

$$\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{org}}}$$

kurt_{org} 入力信号の尖度
 $\text{kurt}_{\text{proc}}$ 出力信号の尖度

カートシス比が1 $\Rightarrow$ ミュージカルノイズ無し

カートシス比が1より大 $\Rightarrow$ ミュージカルノイズ多い



雑音抑圧量 (Noise Reduction Ratio: NRR)

NRR

出力SN比－入力SN比

$$\text{NRR} = 10\log_{10} \frac{E[s_{\text{out}}^2]/E[n_{\text{out}}^2]}{E[s_{\text{in}}^2]/E[n_{\text{in}}^2]}$$

処理後の音声信号: s_{out}

処理後の音声信号: n_{out}

処理前の音声信号: s_{in}

処理前の雑音信号: n_{in}

ここで信号処理における雑音成分の抑圧量は音声成分に影響を与えないと仮定すると

$$\text{NRR} = 10\log_{10} \frac{E[n_{\text{in}}^2]}{E[n_{\text{out}}^2]} = 10\log_{10} \frac{\mu'_1}{\mu_1}$$

処理前の1次モーメント: μ_1

処理後の1次モーメント: μ'_1

1次モーメントで**量**を、2次・4次モーメントで**質**を表す

信号のモデリング

- 入力雑音のパワースペクトルの分布が、以下のガンマ分布に従うと仮定する

$$P(x) = \frac{x^{\alpha-1} \exp(-x/\theta)}{\Gamma(\alpha)\theta^{\alpha}}$$

α : 形状母数

θ : 尺度母数

$\Gamma(\cdot)$: ガンマ関数

- 形状母数は、分布の形状を決定するパラメータ
 - $\alpha = 1$ のときは、ガウス信号を2乗したものが従う分布
 - $\alpha < 1$ のときは、優ガウス信号(音声, 音楽)を2乗したものが従う分布

各種信号処理における高次モーメント関数

一般化スペクトル減算 [EURASIP JASP 2010]

$$\mu_m^{\text{SS}} = \theta^m \mathcal{M}(\alpha, \beta, \eta, m)$$

$$\mathcal{M}(\alpha, \beta, \eta, m) = \frac{1}{\Gamma(\alpha)} \sum_{l=0}^m (-\beta\alpha)^l \frac{\Gamma(m+1)}{\Gamma(l+1)\Gamma(m-l+1)} \Gamma(\alpha+m-l, \beta\alpha) + \eta^{2m} \frac{\gamma(\alpha+m, \beta\alpha)}{\Gamma(\alpha)}$$

$\gamma(\alpha, z) = \int_0^z t^{\alpha-1} \exp(-t) dt$: 第1種不完全ガンマ関数

$\Gamma(\alpha, z) = \int_z^\infty t^{\alpha-1} \exp(-t) dt$: 第2種不完全ガンマ関数

一般化ウィーナフィルタ [IEEE Trans. ASLP 2011]

$$\mathcal{M}_{\text{QPWF}}(\alpha, \beta, m, \xi, \eta) = \frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{t^{(\xi\eta+1)m+\alpha-1}}{\left\{ t^{\frac{\xi}{2}} + \beta \frac{\Gamma(\alpha+\frac{\xi}{2})}{\Gamma(\alpha)} \right\}^{2m\eta}} \exp(-t) dt$$

各種信号処理における高次モーメント関数

ベイズアンMMSE音声振幅推定 [IWAENC2012]

$$\begin{cases} \mu_1 = K_1(|Y(f, \tau)|^2) \\ \mu_2 = K_2(|Y(f, \tau)|^2) + K_1^2(|Y(f, \tau)|^2) \\ \mu_4 = K_4(|Y(f, \tau)|^2) + 4K_3(|Y(f, \tau)|^2)K_1(|Y(f, \tau)|^2) \\ \quad + 3K_2^2(|Y(f, \tau)|^2) + 6K_2(|Y(f, \tau)|^2)K_1^2(|Y(f, \tau)|^2) \\ \quad + K_1^4(|Y(f, \tau)|^2) \end{cases}$$

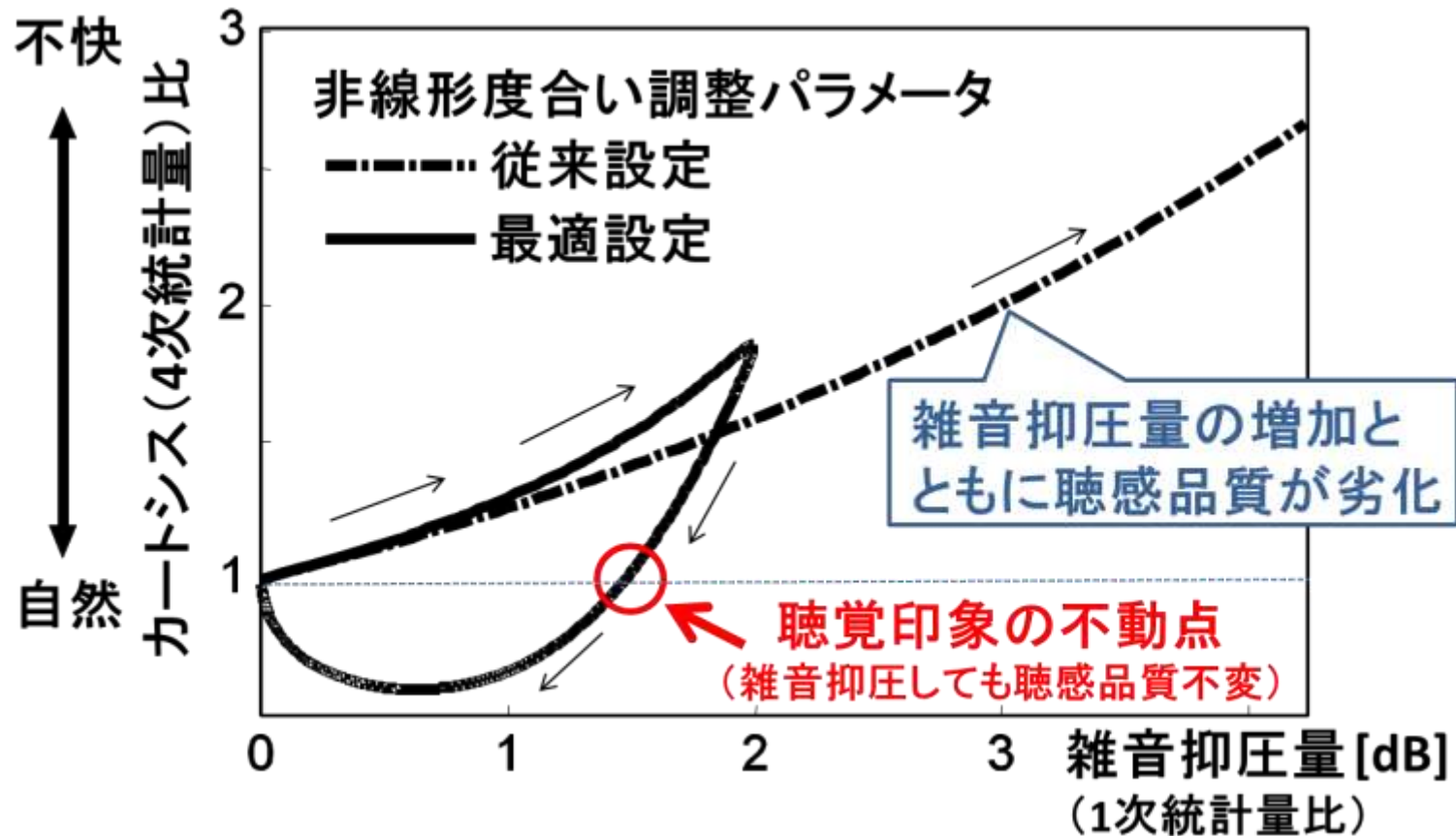
where $K_m(|Y(f, \tau)|^2) =$

$$\sum_{i=1}^2 (-1)^{i-1} (1-\alpha)^{im} (R(\beta)\eta\theta)^m Z_i(m) K_m((\xi^{\text{ml}}(f, \tau))^i),$$

$$\begin{cases} K_1((\xi^{\text{ml}}(f, \tau))^i) = \mu_{i[SS]} \\ K_2((\xi^{\text{ml}}(f, \tau))^i) = \mu_{2i[SS]} - \mu_{i[SS]}^2 \\ K_3((\xi^{\text{ml}}(f, \tau))^i) = \mu_{3i[SS]} - 3\mu_{2i[SS]}\mu_{i[SS]} + 2\mu_{i[SS]}^3 \\ K_4((\xi^{\text{ml}}(f, \tau))^i) = \mu_{4i[SS]} - 4\mu_{3i[SS]}\mu_{i[SS]} - 3\mu_{2i[SS]}^2 \\ \quad + 12\mu_{2i[SS]}\mu_{i[SS]}^2 - 6\mu_{i[SS]}^4 \end{cases}$$

$$\mu_m[SS] = \sum_{l=0}^m (-\eta)^l \frac{\Gamma(m+1)\Gamma(\eta+m-l, \eta)}{\Gamma(\eta)\Gamma(l+1)\Gamma(m-l+1)}.$$

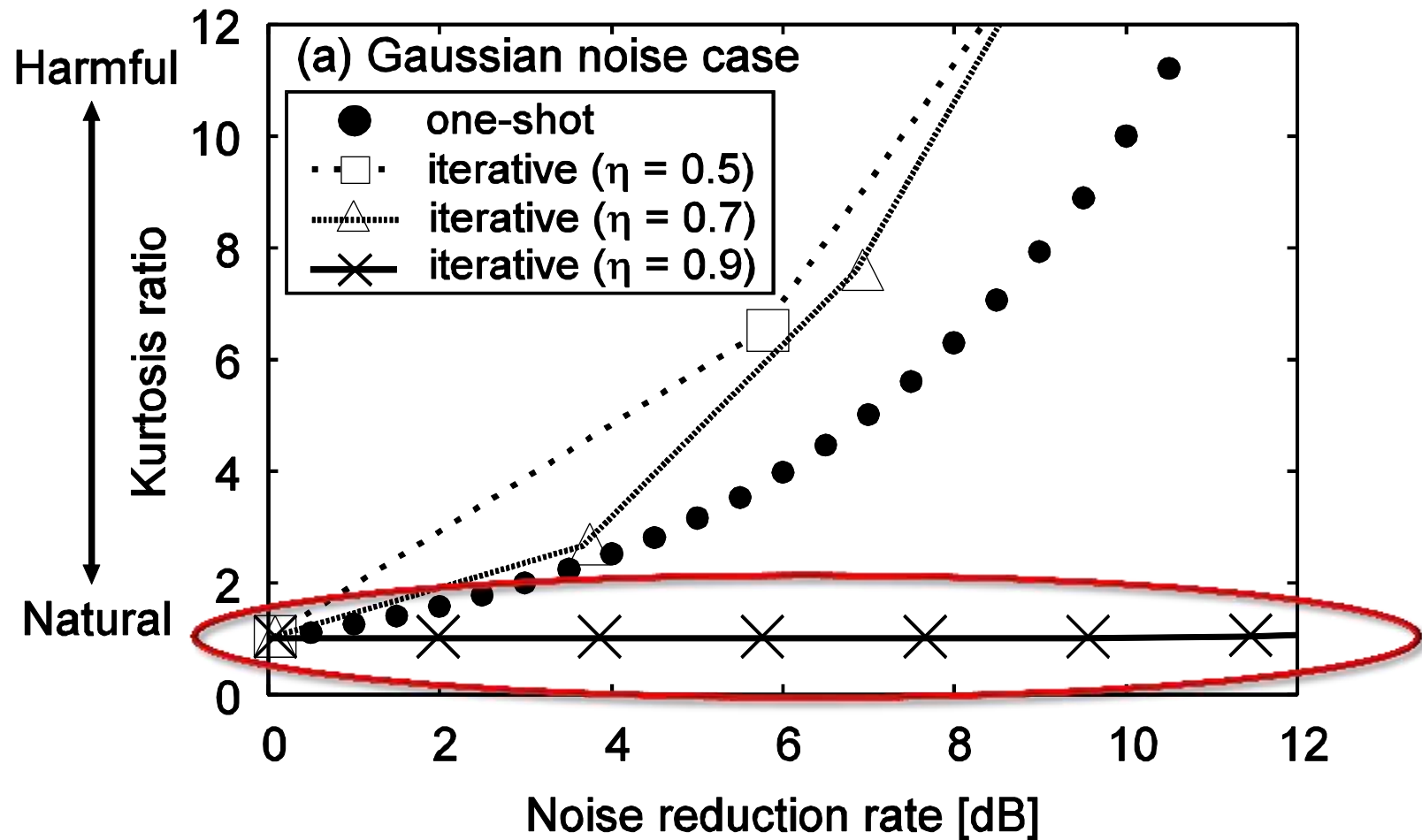
高次統計量空間での不動点の発見



ミュージカルノイズフリー信号処理 [IEEE Trans. ASLP 2012]

- カートシス比不動点の存在は聴覚印象の不動点を表す
- 一次統計量の増分が少ない場合は本処理を繰り返せば良い

スペクトル減算を繰り返すと？



$\alpha = 1.0, \beta = 2.4, \eta = 0.9$ でミュージカルノイズフリー状態を達成→以降, 数学的に解を求める

カートシス比不動点の解析例(スペクトル減算法)

$$\text{kurt}(\alpha_0, \beta, \eta) = \frac{\mathcal{S}(\alpha_0, \beta, 4) + \eta^8 \mathcal{F}(\alpha_0, \beta, 4)}{(\mathcal{S}(\alpha_0, \beta, 2) + \eta^4 \mathcal{F}(\alpha_0, \beta, 2))^2} = \frac{(\alpha_0 + 3)(\alpha_0 + 2)}{(\alpha_0 + 1)\alpha_0}$$

処理後のカートシス

観測信号のカートシス

$$\mathcal{S}(\alpha, \beta, m) = \frac{1}{\Gamma(\alpha)} \sum_{l=0}^m (-\beta\alpha)^l \frac{\Gamma(m+1)}{\Gamma(l+1)\Gamma(m-l+1)} \cdot \Gamma(\alpha+m-l, \beta\alpha)$$

$$\mathcal{F}(\alpha, \beta, m) = \frac{\gamma(\alpha+m, \beta\alpha)}{\Gamma(\alpha)}$$

$\mathcal{H} = \eta^4$ と置くことで, 2次方程式に帰着

$$\mathcal{H} = \{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^{-1}$$

$$\begin{aligned} & \left[\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2) \right. \\ & \pm \left[\{\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^2 \right. \\ & \left. - \{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\} \right. \\ & \left. \left. \{\mathcal{S}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{S}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\} \right]^{\frac{1}{2}} \right] \end{aligned}$$

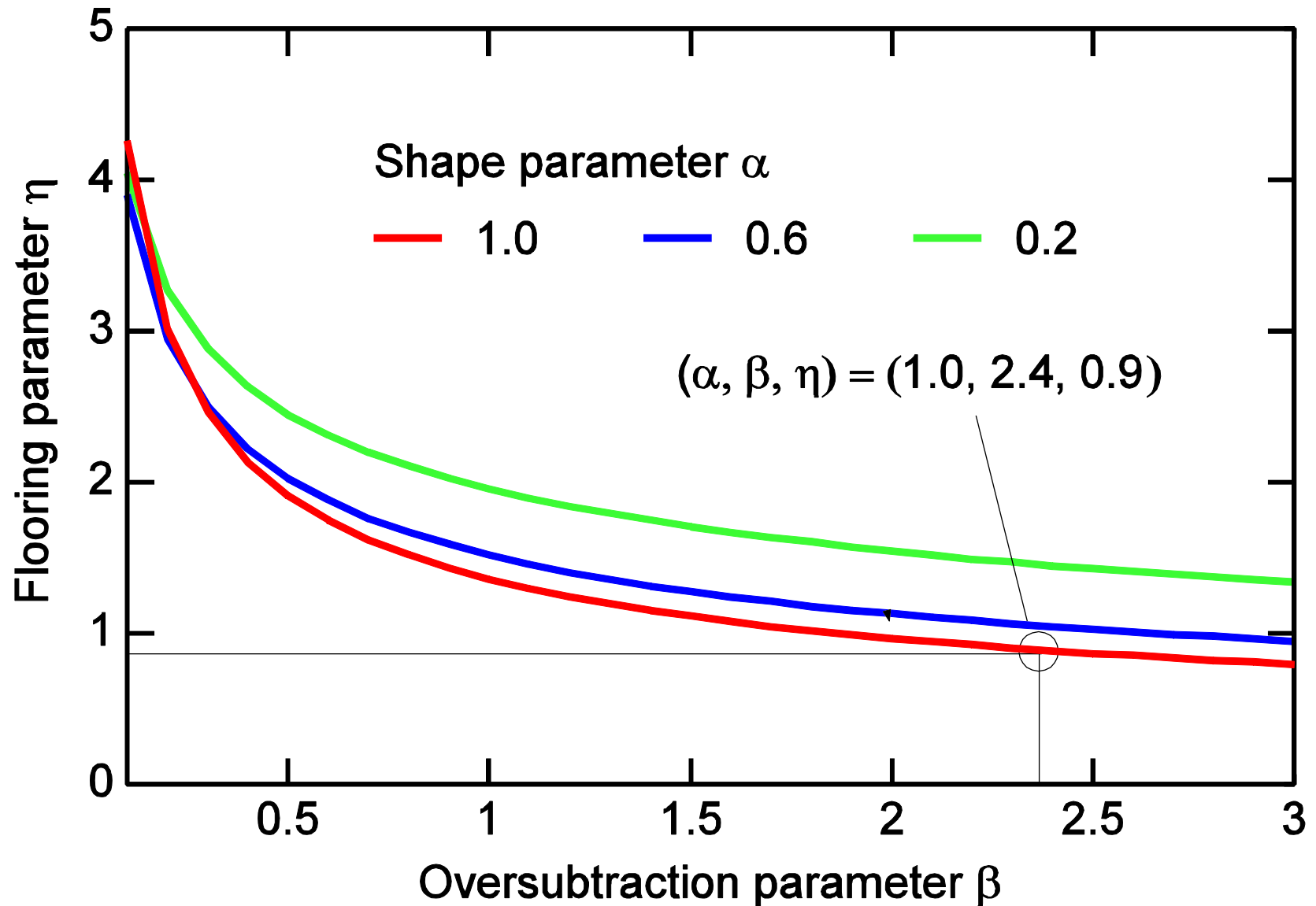
一部の信号処理

は代数的に最適化

を行うことができる

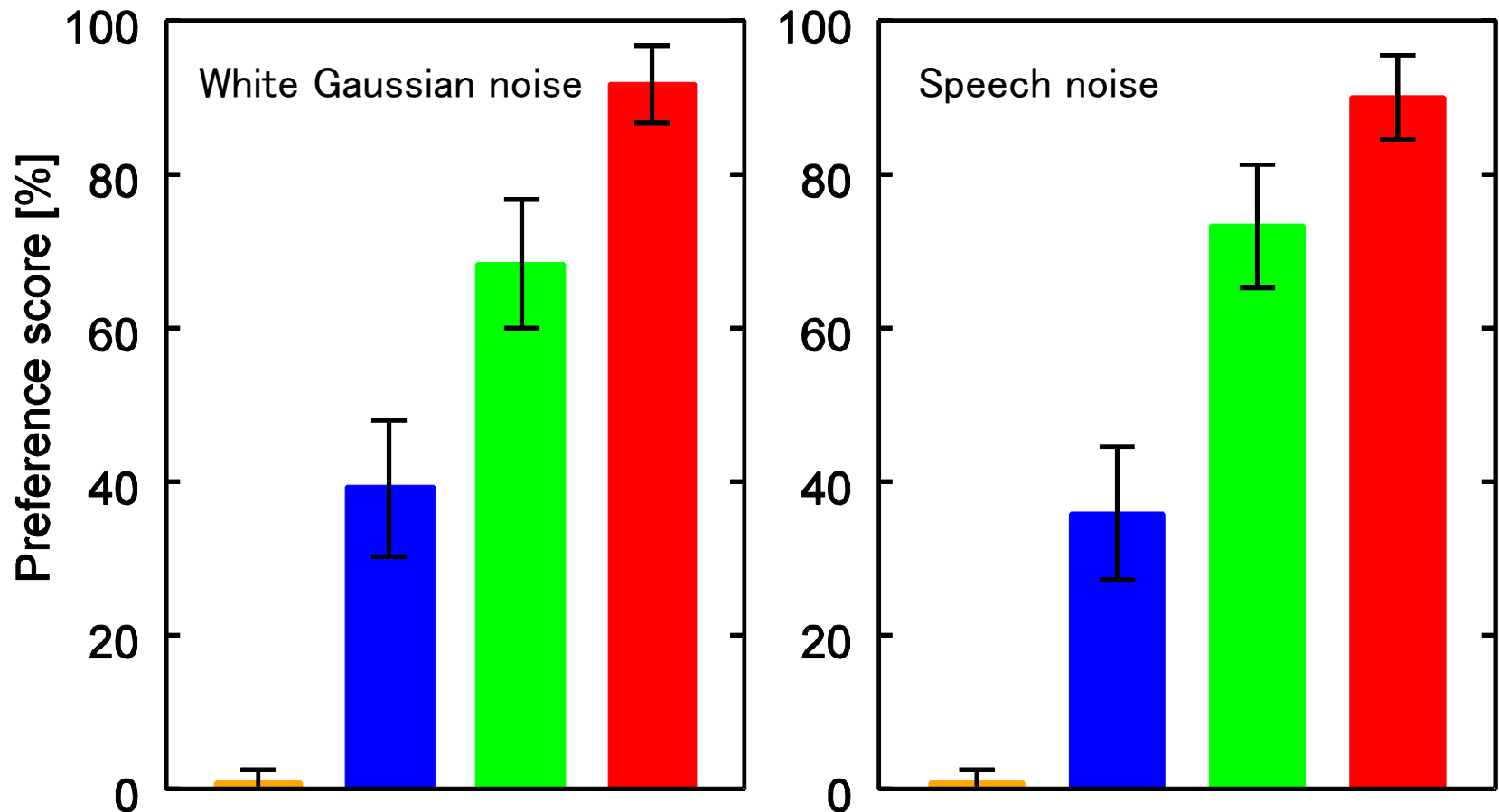
[2013年市村学術賞]

ミュージカルノイズフリーを満たすパラメータの例



主観評価実験結果

One-shot spectral subtraction: ■ Wiener filtering: ■
MMSE STSA estimator: ■ Iterative spectral subtraction: ■
95% confidence intervals: 



不動点発見におけるセレンディピティ

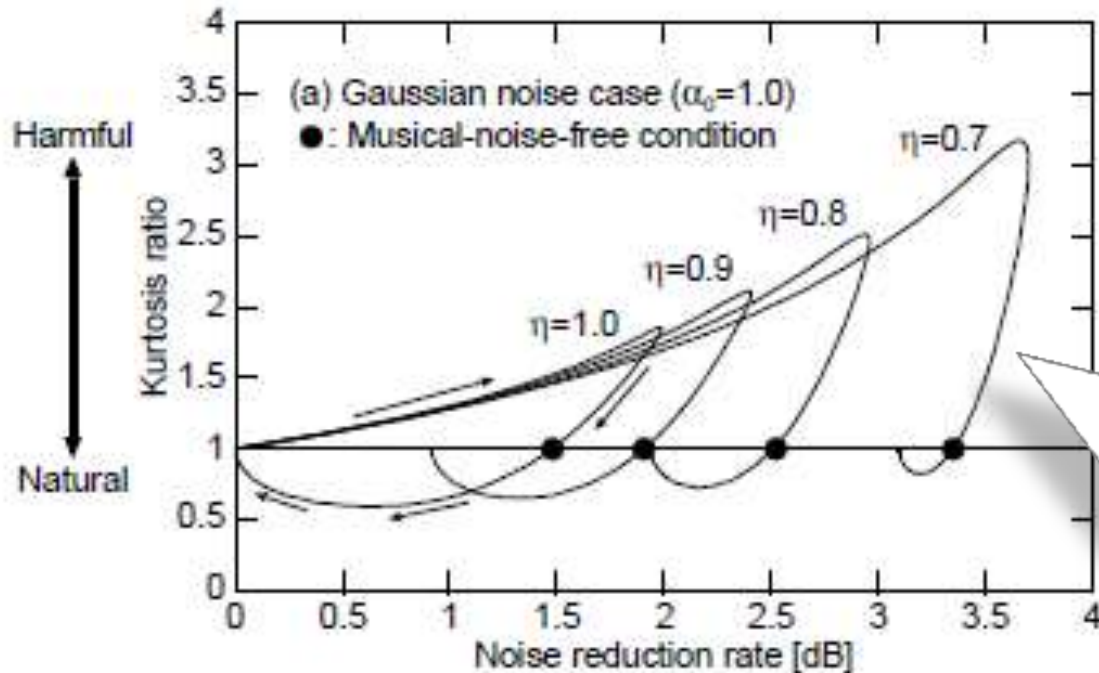
- 高次統計量(カートシス)の不動点はなぜ見つかったのか
 - 「不動点を見つけよう」と狙っていたわけではない
 - ある学生が「**雑音を過大に引き、かつ原音を過大に足すと、音色が元に戻る時がある！**」と報告してきた
 - 誰も信じなかった。学生のプログラムのバグだと思っていた
 - そもそも、雑音を減らすための信号処理において「雑音を足しなおす」というのは**常識的におかしい！**



不動点発見におけるセレンディピティ

■ 高次統計量空間における「ヒステリシス」の確認

- 本当にあっているかどうか数理的に確認してみた
- すると、なんと、高次統計量空間には非線形性に由来する**ヒステリシス現象が存在する**ことが分かった
- 存在が証明されれば後はそれが「何時生じるか」を予測するだけ



海外論文誌に投稿した際に、査読で「**意味が分からんので削除せよ**」と言われたヒステリシスの存在証明図。

後に国際会議招待講演にて喝采をあびた。

本日の講義内容

時系列メディア処理における「音源分離」の紹介

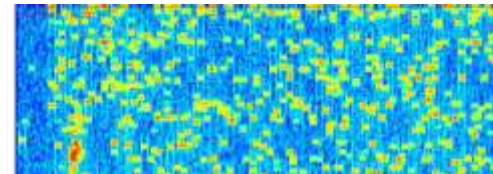
独立成分分析

聞き分けの出来る人工耳



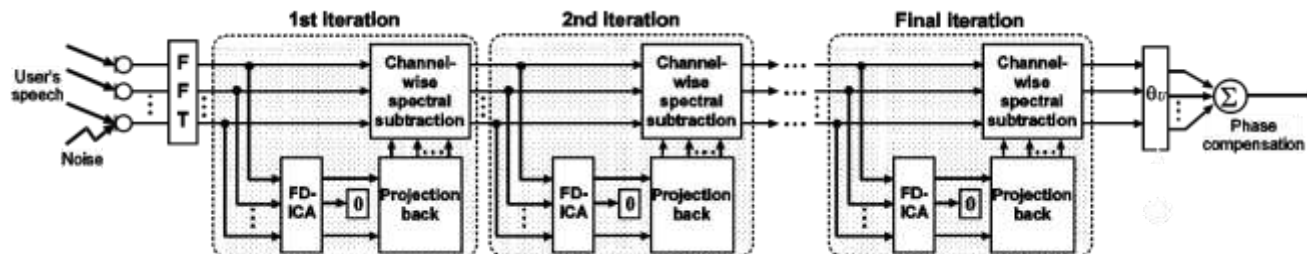
高次統計量音質制御

気持ち悪さを測る



スパース表現信号処理: その威力と活用例

音楽信号処理への応用



研究紹介3. 音楽信号解析

- 様々な楽器がまじりあった音楽信号の中から、自分の好きな楽器を見つけ出し、自分の好みのリミックス版を製作する。
- 非負値行列因子分解(NMF)という数理アルゴリズムに基づいて、音を事前に学習した「より簡単(低ランクかつ疎:スパース)な頻出パターン」に分解することにより、信号を解析する。

スパース分解
低ランクモデル

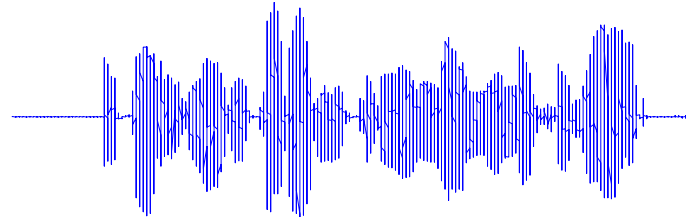
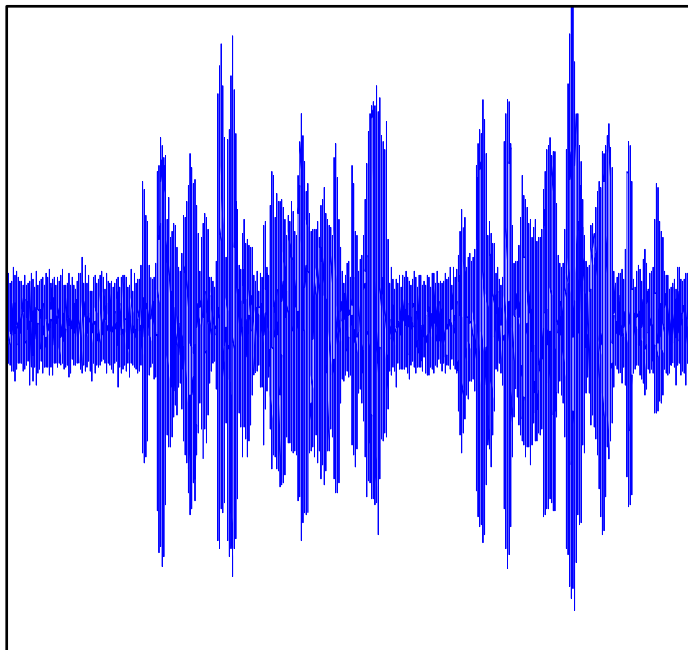
インタラクティブ音オブジェクトコントローラ

- 🖱 実際の曲の楽器の配置を変えてリミックスを作成
- 🖱 指定した楽器を左右に振ったり、遠くに飛ばしたり...

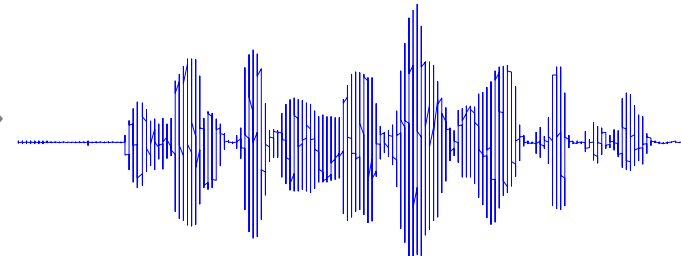
スパース信号解析としてのICA

■ スパース(疎)信号解析とICAの密接な関係

- スパースな生起を有する信号の確率密度は**非ガウス**
- より非ガウス分布に従う確率信号へ分解するのがスパース解析
- ICAによって分解される信号は、実際、より「疎」なものになる



優ガウス



優ガウス



ガウス

研究背景

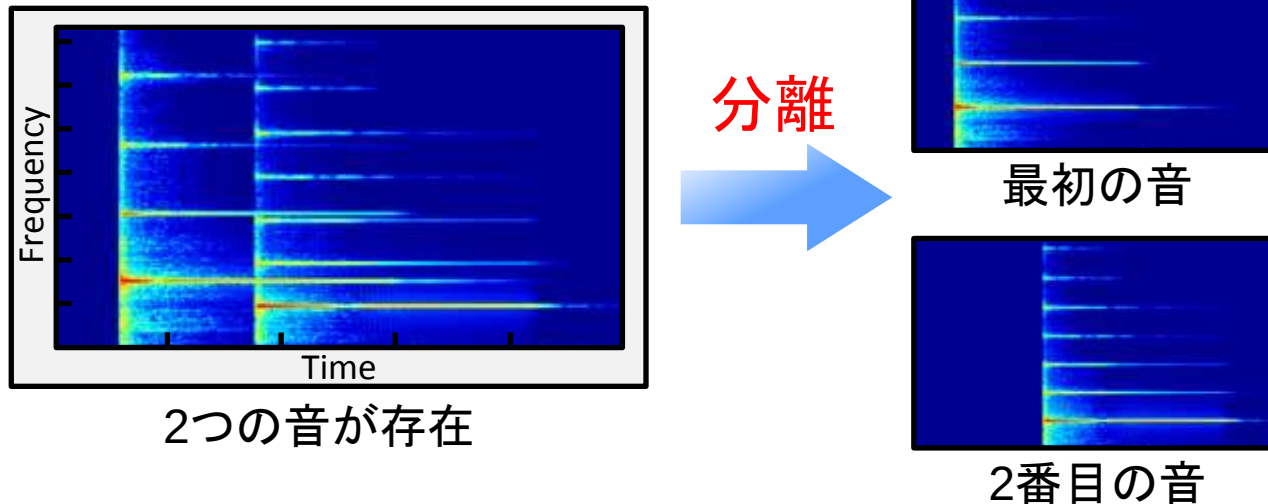
- 音源分離技術

- 複数の音源が混合された信号を個々の音源に分離する信号処理
- 音声と雑音の分離, 個々の音源の編集, 音拡張現実感への基盤技術



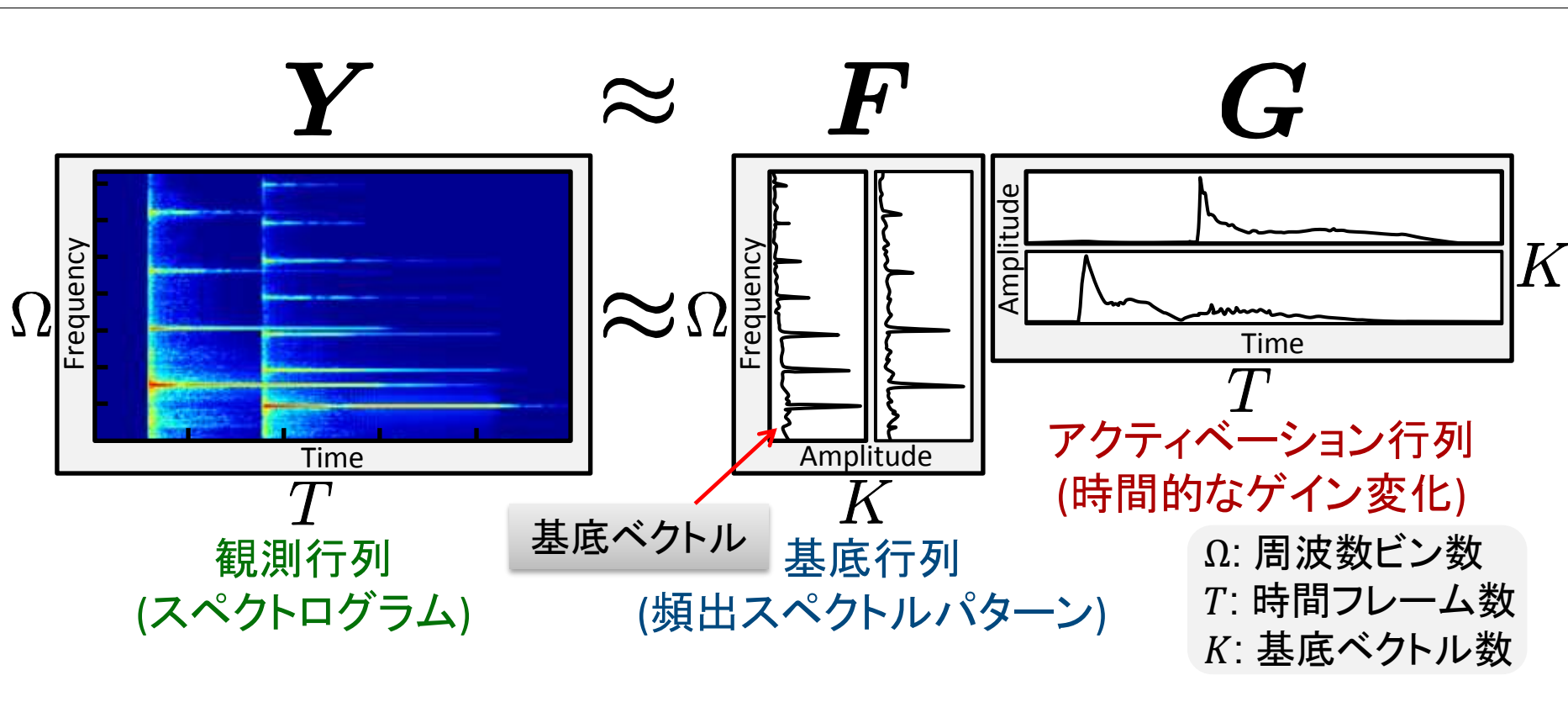
- 代表的な音源分離方法

- 時間-周波数表現されたスペクトログラム上で音源ごとに分解



非負値行列因子分解 [Lee, et al., 2012]

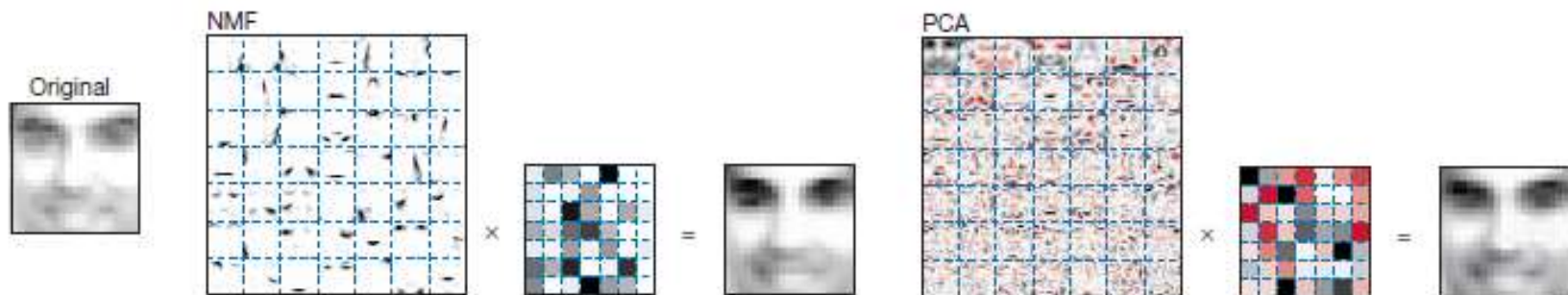
- 非負値行列因子分解 (nonnegative matrix factorization: **NMF**)
 - 非負値に限定したスパース分解表現による特徴量抽出法



- 全て非負値の要素から構成される

なぜ非負値なのか？

- 観測データの非負性
 - 世の中の多くの物事は非負値で表現するのが自然
 - パワースペクトログラム, 文書中の各単語の出現回数, 画像データ etc.
- 基底行列の非負性
 - 非負データの構成要素は非負であるべき
 - 負のスペクトル, 負の出現回数, 負の画素値は解釈できない
- アクティベーション行列の非負性
 - 構成要素の混ざり方は足し算のみ
- 「非負制約を与えた低ランク近似」は暗に「スパースな分解」をする
 - スパースな解は有意な情報を表現する (可能性が高い)

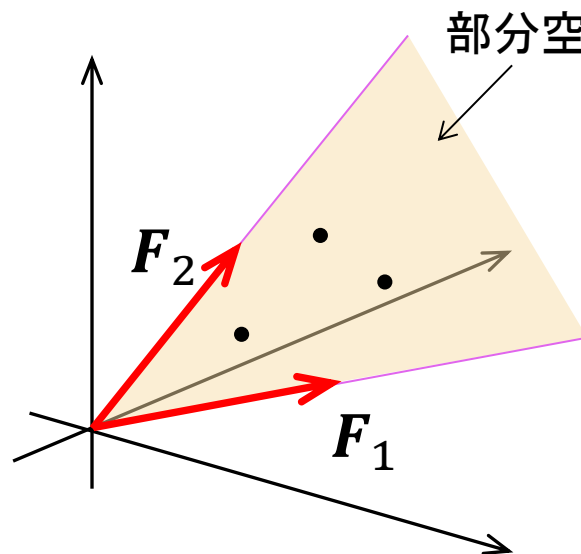


基底行列の意味

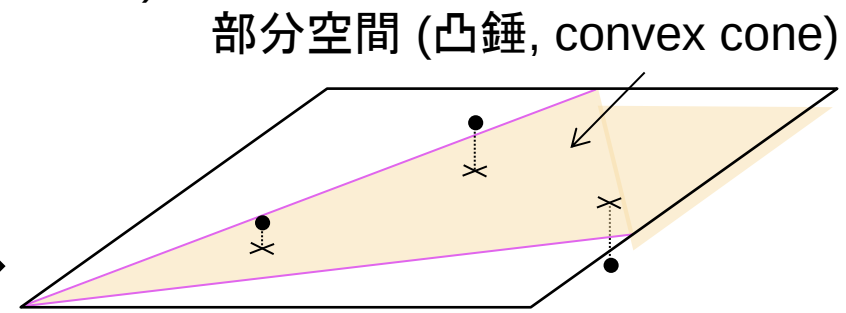
- NMFで分解された基底ベクトルとは？

$$\begin{array}{c} \text{次元数} \\ \left\{ \begin{pmatrix} 1 & 3 & 2 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \right. = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \end{pmatrix} \\ \left. \begin{array}{c} \text{データ数} \end{array} \right\} \quad \left. \begin{array}{c} \text{基底数} \end{array} \right\} \end{array}$$

- 3次元空間の3点を2本の基底ベクトルで張られる凸錐の部分空間で表現



部分空間 (凸錐, convex cone)



部分空間 (凸錐, convex cone)

様々なデータを入力していくと、
だんだんと基底間の角度が広がる
→ スパースな基底が構成される

NMFの最適化

- 変数 F と G をどのように見つけるのか
- 目的関数 $\mathcal{J}$ を定義して最小化する F と G を求める

$$\begin{aligned} \arg \min_{F, G} \mathcal{J} = & \|Y - FG\|_{\text{Fr}}^2 \\ & \text{subject to } f_{\omega, k}, g_{k, t} \geq 0 \end{aligned}$$

➡ 不等式制約条件付き最適化問題に帰着

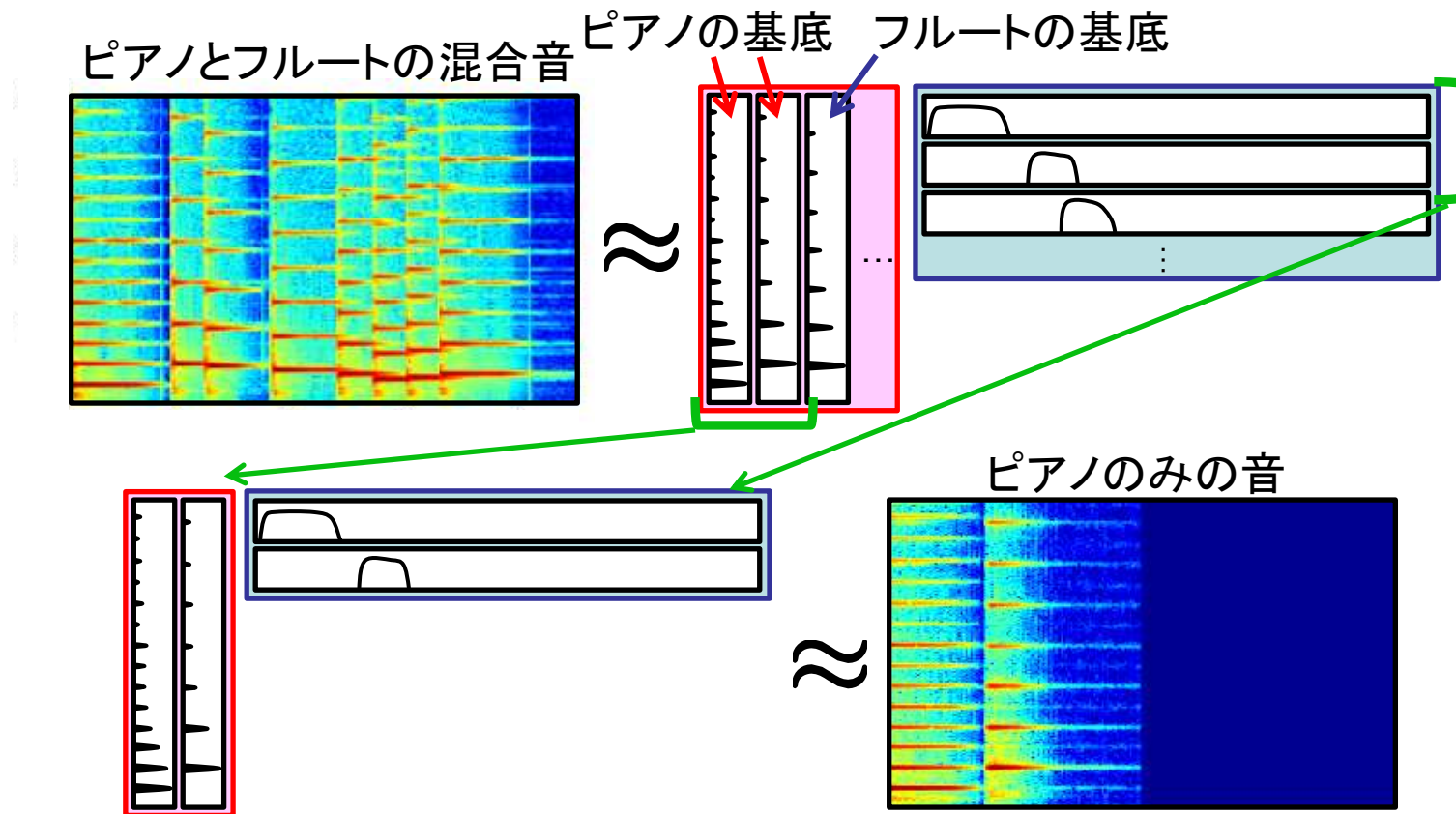
- 解法はいろいろあるが、最終的には効率的な反復型最適化式が導出される [Lee, et al., 2001]

$$f_{\omega, k} \leftarrow f_{\omega, k} \frac{[Y G^T]_{\omega, k}}{[F G G^T]_{\omega, k}}, \quad g_{k, t} \leftarrow g_{k, t} \frac{[F^T Y]_{k, t}}{[F^T F G]_{k, t}}$$

解析的な解は求まらない (不良設定の逆問題の為)

NMFによる音源分離の可能性

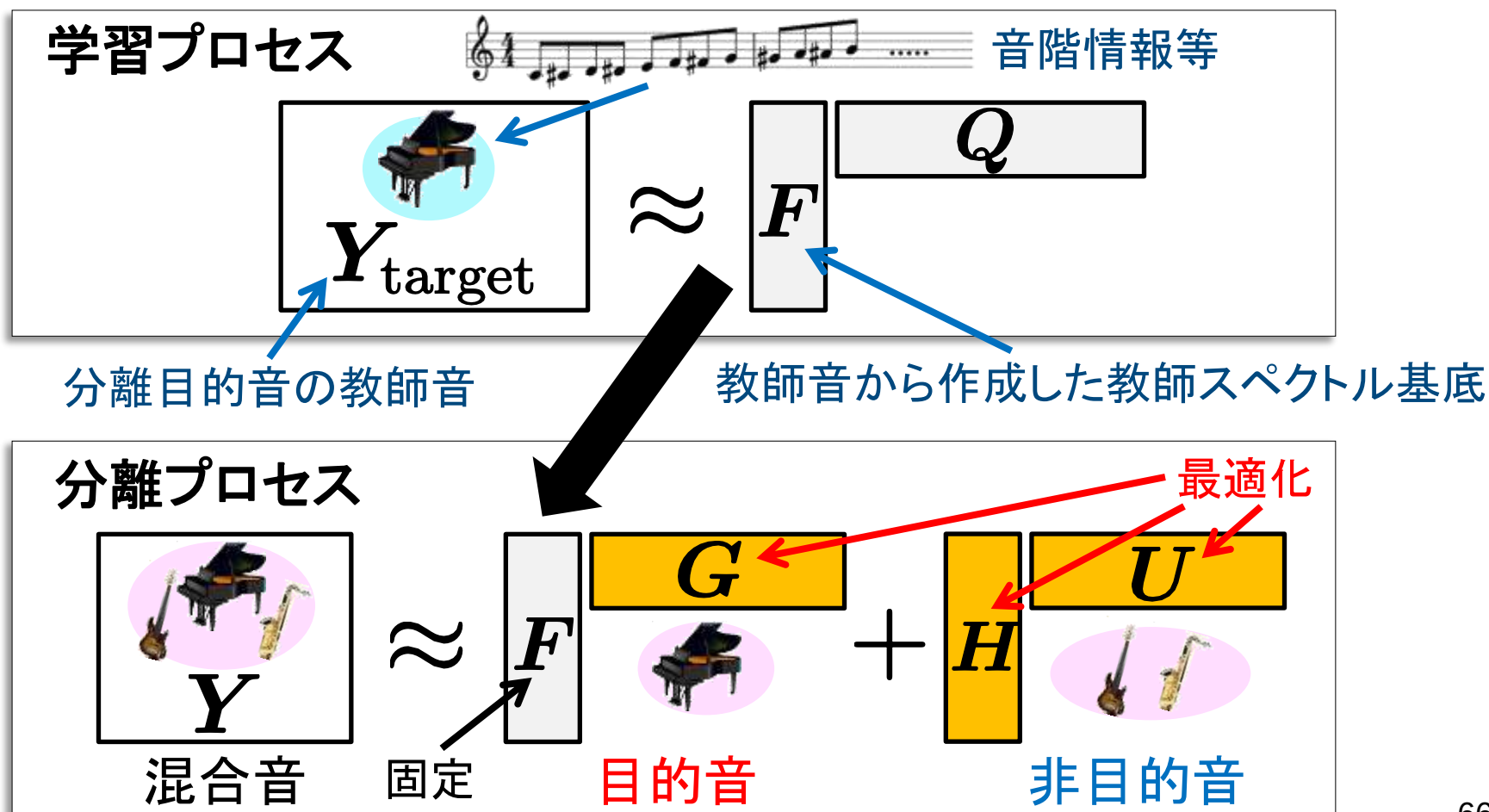
- NMFによって抽出された基底が「何の音なのか」が分かれば音源分離ができる



- 音源毎に基底を選別する必要がある
 - 出来るのか？

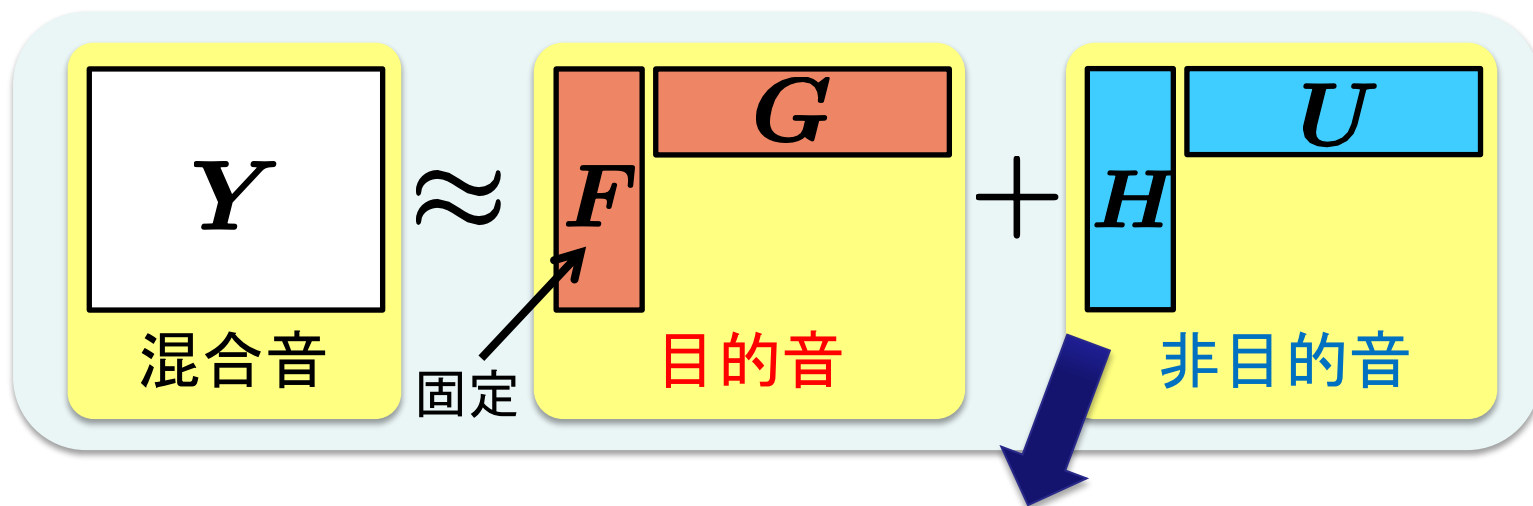
教師ありNMF (SNMF) [Smaragdakis, et al., 2007]

- 分離したい目的音の教師 (サンプル) 音を事前に学習
 - 学習プロセスで教師スペクトル基底 (dictionary) F を作成
 - 分離プロセスで目的音 FG と, 非目的音 HU に分離



罰則条件付きSNMF [Kitamura, et al., 2014]

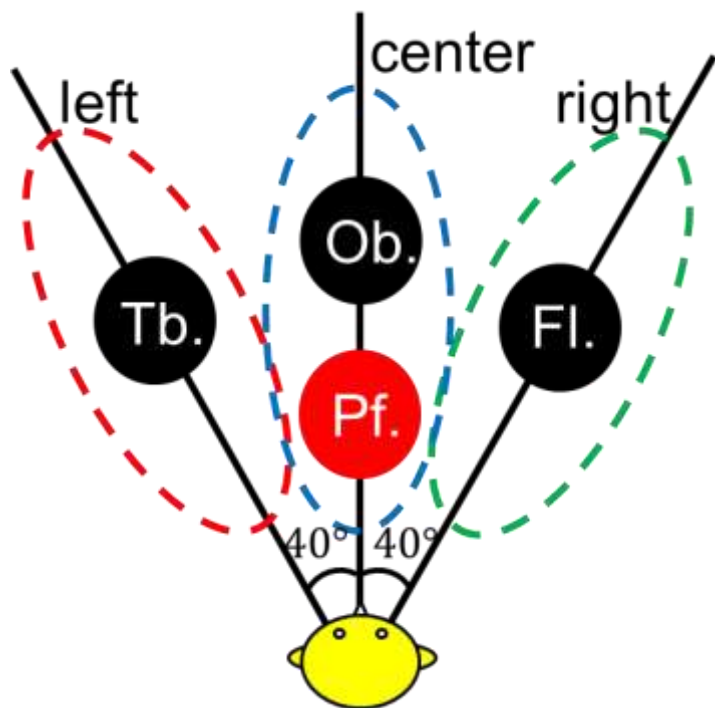
- 教師と同じスペクトル基底がその他の基底に現れる
➡ 基底共有問題が発生, 分離した目的音が欠落
- その他の基底 H を教師基底 F と (できるだけ) **無相関にする罰則条件** をコスト関数に付与する
- 罰則条件付きSNMF (Penalized SNMF: PSNMF)



なるべく F と無相関となるように最適化

多チャンネル音楽信号分離デモ1

4楽器から成るステレオ曲を実際に分解してみた。



全てのメロディが聞き取れたら、君もプロミュージシャン！

原曲



分離音 1



分離音 2



分離音 3



分離音 4



簡単



フルート
聞き取れた？

難しい



多チャンネル音楽信号分離デモ2

📺 実際の演奏曲を教師有りNMFで分解してみた。

原曲



教師1



分離音1



教師2



分離音2



多チャンネル音楽信号分離デモ3

🖥️ プロレコーディングに対応できる品質を目指して。

原曲（プロ演奏）



Saxのみを抜いた
伴奏部分



Copyright © 2014 Yamaha Corp.
All rights reserved.

サックス奏者が
消えた!?



研究紹介おまけ. 統計的音声合成

- テキストもしくは自分の声を入れるだけで、誰の声でも、どんな訛りでも、何語でも、喋ることが出来るような統計的音声合成システムを実現する。
- 深層学習 (Deep Neural Net) の枠組みを活かし、「AIオレオレ詐欺師」と「AI防犯課刑事」を対決させて、お互いに精度を高めるAnti-Spoofing 敵対学習理論を独自に提唱

ビッグデータ
深層敵対学習



音声合成の統計モデリング



入力 x と出力 y の関係をどう記述するか？ → **統計的逆問題**

声の**ゆらぎ**をどう扱うか？

人間らしい声とは何か？ 「人間らし

東京大学は世界一！

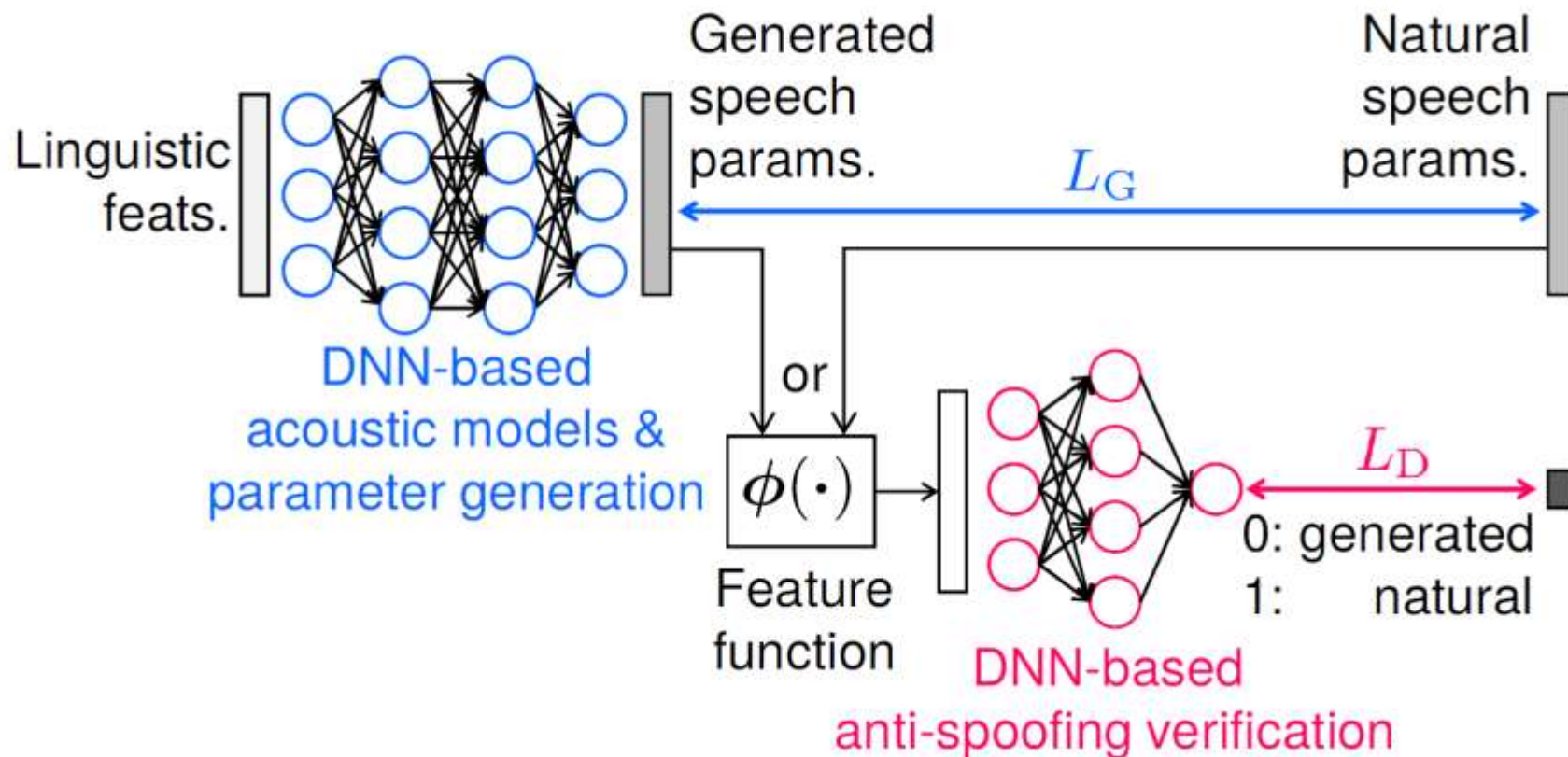
昼飯はとんこつラーメンに限る！

2015年
テキスト音声合成国際コンペ一位
2016年
音声変換国際コンペ一位





Anti-Spoofingと敵対する音響モデル学習理論



まとめ

- 音メディアを対象とする様々な信号処理に関して、ブラインド音源分離・音質制御・スパース信号処理の観点より概説した。
- 独立成分分析に基づくブラインド音源分離処理を解説し、ハードウェア実装例などを紹介した。
- カートシス比を用いてミュージカルノイズの発生量を定量化できることを示し、それに基づくミュージカルノイズフリー理論を提案した。
- 信号のスパース表現に関して、その意味と効果を解説した。実装例として、非負値行列因子分解を紹介し、音楽信号分離へ応用した例を示した。

最後に: セレンディピティと理論の架け橋

- セレンディピティは誰にでも訪れる
 - それを見逃さない努力は必要
- セレンディピティの先を目指す
 - 未開の現象を数学によって徹底的に説明してみる
 - 説明が出来れば、もはや偶然が必然になる
 - セレンディピティと数理解析の橋渡しが重要



連絡先

- 講義の質問、感想、等は以下まで。

hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

もしくは「猿渡洋」でググっても見つかります。