

工学研究科 電気工学専攻 生体機能工学分野

new!

教授：吉井 和佳 講師：伊藤 陽介 助教：上田 博之



自己紹介

• 学歴

- 1999－2003：京都大学 工学部 情報学科
- 2003－2005：同 情報学研究科 知能情報学専攻 修士課程
- 2005－2008：同 博士課程

• 職歴

- 2008－2014：産総研 情報技術研究部門 研究員
- 2014－現在：情報学研究科 知能情報学専攻
 - 2014－2018：講師
 - 2018－現在：准教授
- 2017－現在：理研 AIPセンター
 - 音響情景理解チームリーダー (兼任)

音響信号処理
音楽情報処理

視聴覚信号処理



Kazuyoshi Yoshii
2021年2月13日の4:49



琵琶湖一周

208.24 km 1,167 m 6時間 49分
大中ストレート北行き 自己最高記録 (3:26)
萩の浜口～ローソン 自己最高記録 (7:17)



FTP 273w (4.4w/kg)

研究環境

- 居室・実験室：桂キャンパス A1棟 2階
 - 保有設備：MRI, 脳磁界計測装置, GPUサーバなど

研究室見学会
A1-218号室



研究備品を大幅に増強
Dell U4021QW & MacBook Air (M2)
小型ワークステーション (+GPU)

研究・教育哲学

確率モデルに基づく統計的アプローチ

- 順問題：観測変数 \mathbf{X} に対する生成モデルの定式化

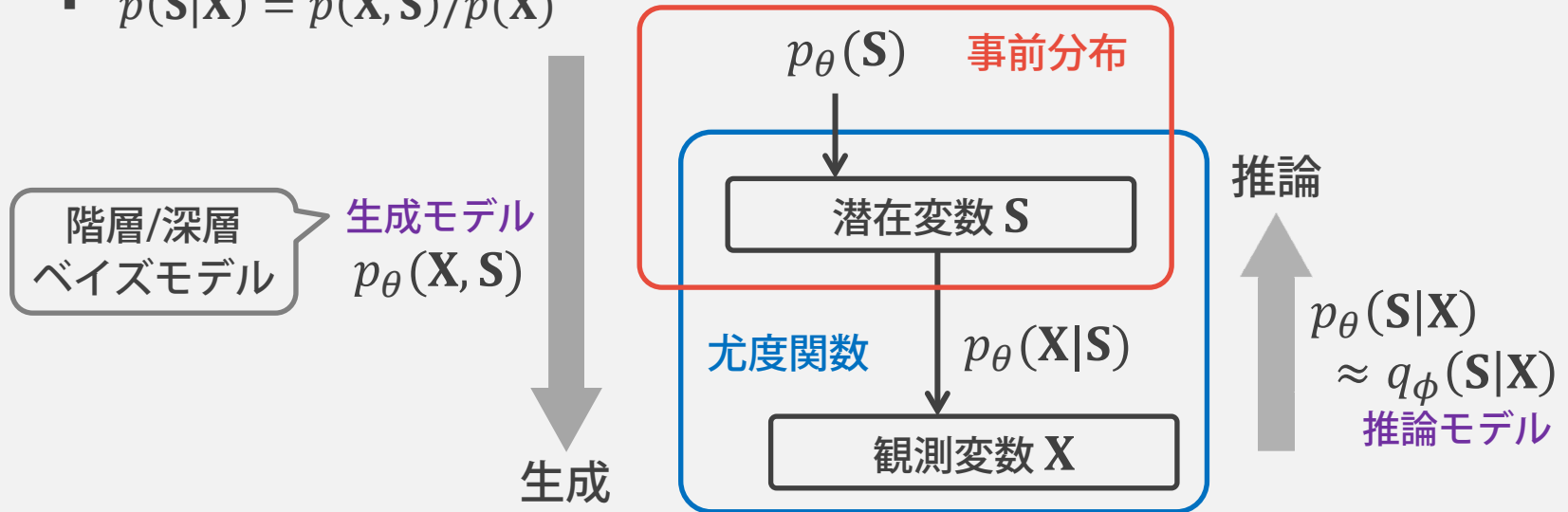
物理拘束・専門知識
線形/非線形モデル

- $p(\mathbf{X}, \mathbf{S}) = p(\mathbf{X}|\mathbf{S})p(\mathbf{S})$

- 逆問題：潜在変数 \mathbf{S} に対する推論モデルの導出

解析的導出
反復最適化

- $p(\mathbf{S}|\mathbf{X}) = p(\mathbf{X}, \mathbf{S})/p(\mathbf{X})$



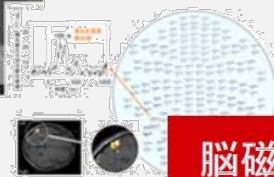
基礎理論に根ざした研究教育

- 一貫した統計的アプローチは優れた教育効果
 - 電気工学の専門知識に立脚した統計的機械学習
 - データサイエンスリテラシー・洞察力の涵養
 - まずは工学・情報学の基礎理論・技術を習得
 - 生体信号 (視聴覚情報・脳情報) は格好の題材
 - 背後には何らかの物理法則や数理モデルが存在
 - 将来どのような道に進むにも“潰しが効く” = 普遍的な技術習得が可能
 - 物理モデル + 統計モデル + 深層学習は要着目分野
 - まだブルーオーシャンでいち早く取り掛かるべき

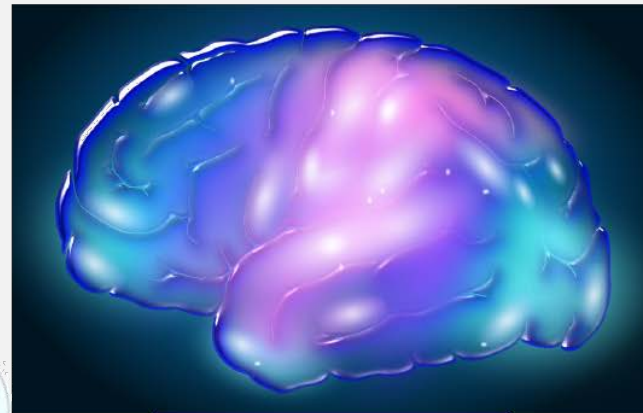
伝統的な電気工学を専攻する学生に統計的アプローチを教える効果は甚大
(薬学研究科との共同研究経験から確信)

研究目標

- 「見えないもの」を「見えるもの」に！
 - 簡便・手軽・低コストな計測・解析・イメージング技術
 - OPMを用いた生体磁気計測およびMRI
 - 脳神経磁場を計測するfMRI

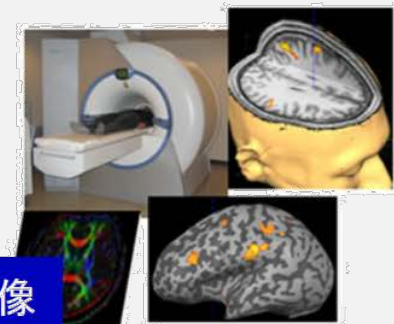


脳磁界
MEG



脳波
EEG

(機能的)磁気共鳴画像
(functional) MRI

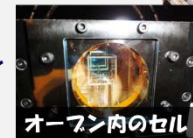
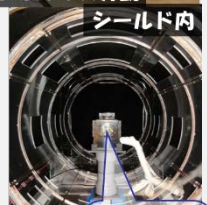
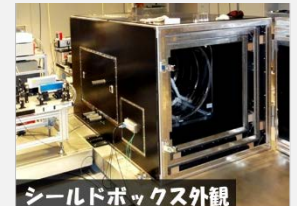
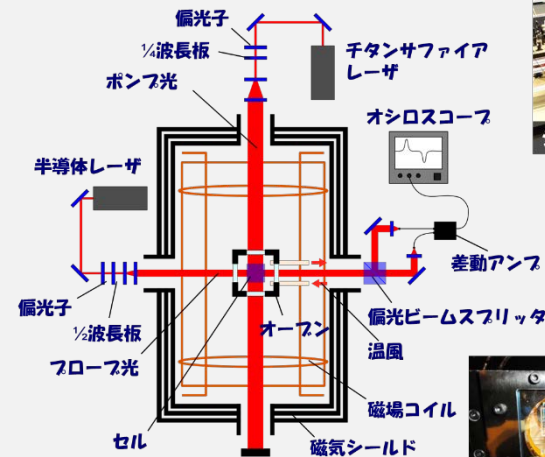
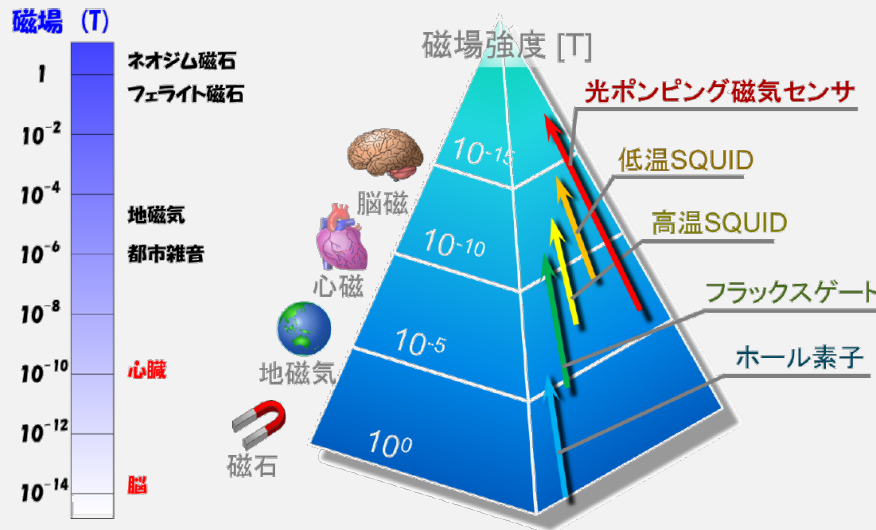


磁気計測

光ポンピング磁気センサの開発

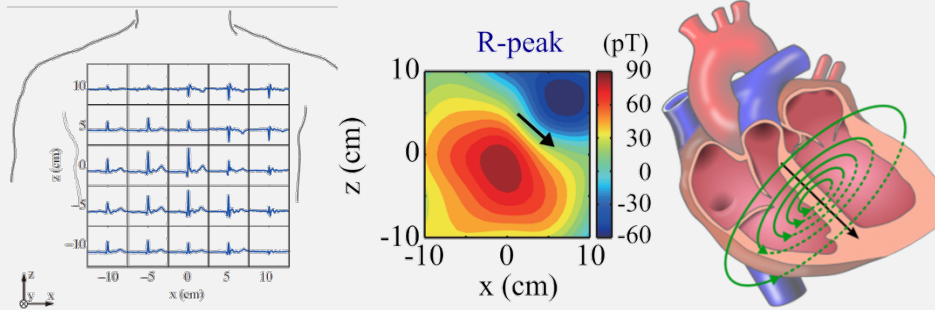
- 超高感度
- 液体ヘリウム等の冷媒が不要 → 低コスト
- SQUIDよりも信号源に近いところで計測が可能

小型モジュールの開発に成功
(企業との共同研究)



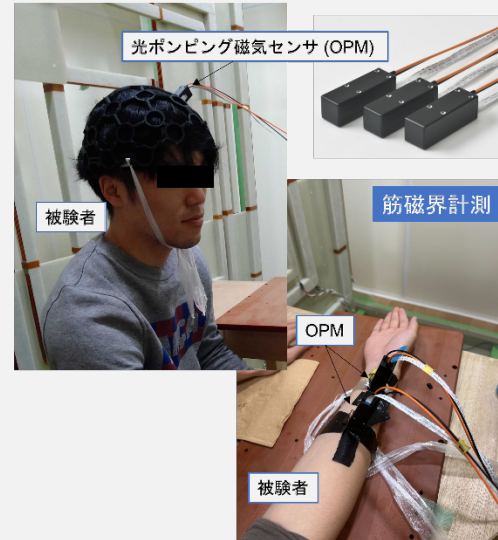
生体磁気計測への応用

心磁図

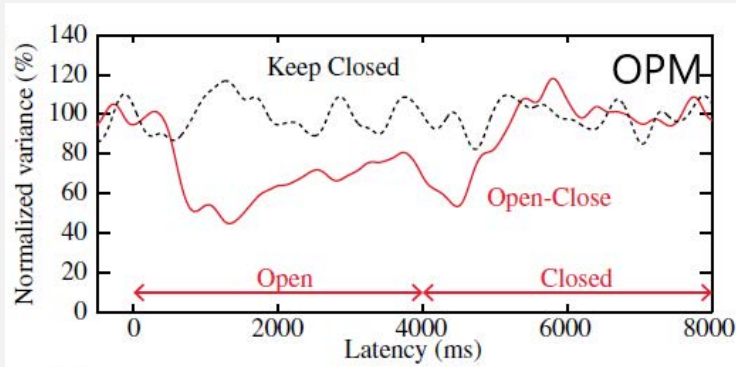
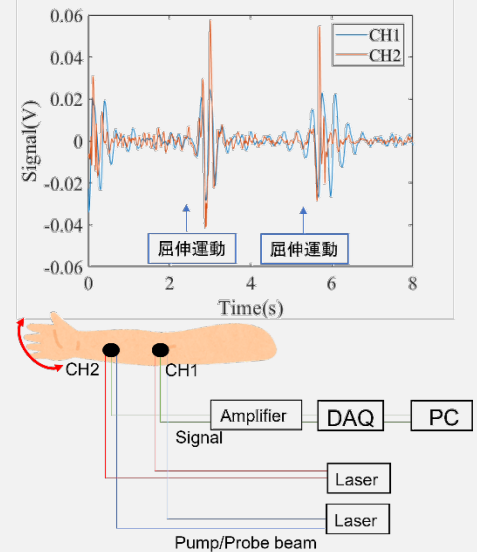


脳磁界・筋磁界の計測

手首屈伸運動時の脳磁界計測



手首屈伸運動時の筋磁界



開眼時の α 波の事象関連脱同期

研究成果

— 視聽覺信号处理 —

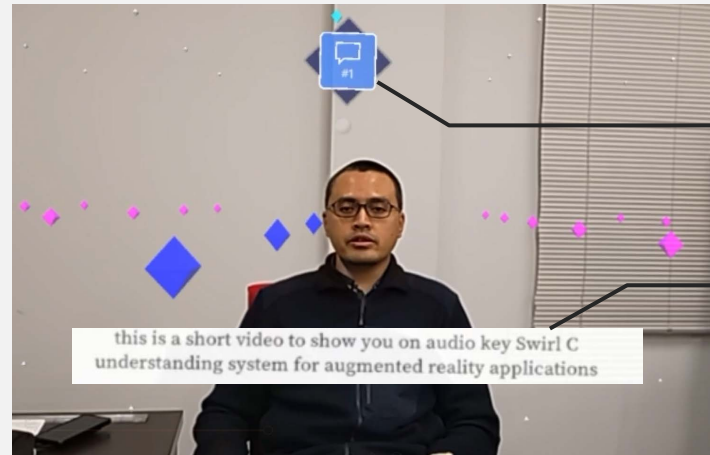
研究の概要

- ARスマートグラスを用いた音環境理解支援システム
 - リアルな雑音環境下で動作
 - ユーザが指定した目的話者のみの音声を認識・翻訳
 - **リアルタイム音声強調**と残響除去
 - 認識結果をAR表示

RGBカメラ・マイク5個を搭載



Microsoft HoloLens 2



目的話者を
示すマーカ

認識結果

ARディスプレイの表示

国際的な位置づけ

- ARグラス向け視聴覚理解技術を世界に先駆けて学術発表
 - 発表業績
 - Interspeech 2022 (音声情報処理のトップ会議)
 - IROS 2022 (ロボティクスのトップ会議)
 - IWAENC 2022 (多チャネル信号処理の会議)
 - 光学シースルー型
 - HoloLens2 (Microsoft)
 - Ray-Ban Metaスマートグラス (Meta)
 - プロトタイプ (Google)
 - ビデオシースルー型
 - Vision Pro (Apple), Quest 3 (Meta)



キラーアプリケーション創出に向けて

- 計測技術 (工学) と解析技術 (情報学) の融合

ハードウェア

ソフトウェア

マイクアレイ

カメラアレイ

深度センサ

IMUセンサ

Robot Operating
System (ROS)

ARグラス
(HoloLens 2)

音源分離・音声強調

雑音抑制・残響除去

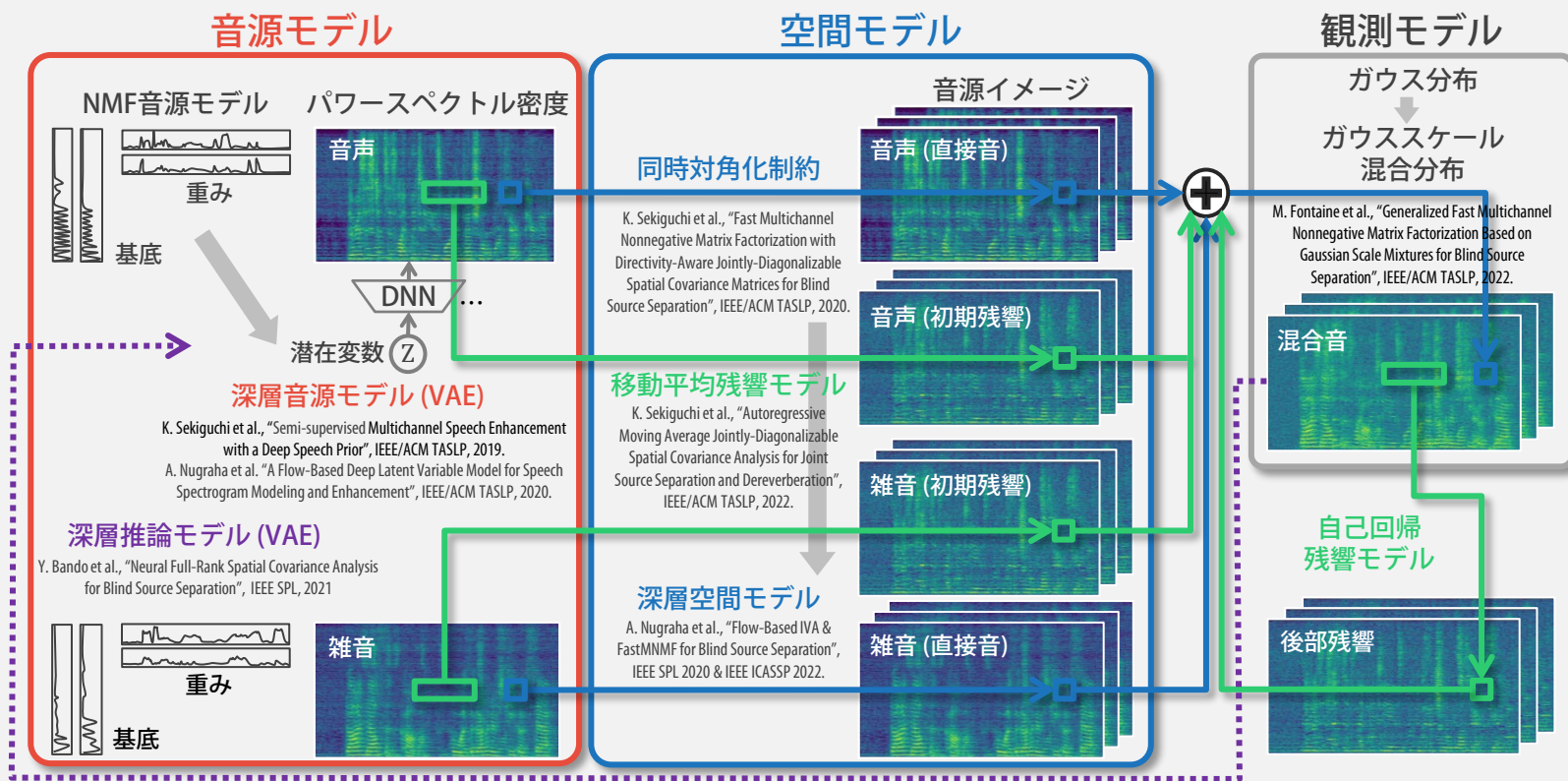
音声認識・翻訳

環境適応

プラットフォーム

統計的音響信号処理の垂直展開

- 世界最先端のブラインド音源分離・残響除去技術を保有
 - 統一的な確率的定式化・深層学習との融合 (多数のIEEE TASLP発表)



研究成果

— 音楽情報処理 —

計算論的聴覚情景分析

- 人間の「聴き分け」能力の数理モデルを作りたい
 - 人間は混合音を一瞬にして「聴き分ける」ことができる
 - ブラインド音源分離・定位
 - 音響イベント検出

音楽を聴き分けられる

→ 単チャンネル音響信号処理

人間は2つの耳をもつ

→ 多チャンネル音響信号処理



今こそ真の自動採譜を

- 人間が読める**楽譜形式**で出力したい！
 - 既存研究でいう「自動採譜」とは「ピアノロール推定」
 - 「楽譜推定」までには大きなギャップ (皆が避けてきた)
 - ピアノ以外の楽器も扱いたい

ピアノロール形式



時間軸は秒 [s]



楽譜形式

時間軸は量子化

人間が再演奏可能!



ポピュラー音楽の自動採譜



10 = 120 = 122 = 125 = 122 = 120 = 122 = 120 = 125 = 120 = 122 = 120 = 122 = 125 = 120 = 122

Vocal

Drums

Chords

5 = 120 = 122 = 120 = 122 = 120 = 120 = 125 = 120 = 122 = 120 = 122 = 120

Vocal

Drums

Chords

9 = 122 = 120 = 122 = 125 = 120 = 122

Vocal

Drums

Chords

12 = 120 = 122 = 125 = 120 = 125 = 118 = 120 = 122 = 125 = 122 = 120 = 122

Vocal

Drums

Chords

AIST-MDB-2020-No.79

16 = 120 = 125 = 120 = 122 = 120 = 125 = 122

Vocal

Drums

Chords

20 = 118 = 122 = 120 = 122 = 120 = 125 = 122 = 120

Vocal

Drums

Chords

24 = 125 = 122 = 120 = 125 = 118 = 125 = 120 = 122 = 125 = 122 = 120 = 122 = 120 = 122 = 120 = 122 = 120 = 125 = 122

Vocal

Drums

Chords

29 = 120 = 122 = 120 = 125 = 120 = 122 = 120 = 122 = 120 = 122

Vocal

Drums

Chords

ピアノ採譜例

- 超絶技巧のピアノ演奏でも高精度に楽譜化可能に



<https://www.youtube.com/watch?v=bol0tThhTZE>

4/4拍子の一般的な
ポピュラー楽曲なら
実用レベルに近い精度
(応用上十分に広い
範囲をカバー)

生成モデルに基づくアプローチ

- 生成モデル＝音楽信号の確率的な生成過程を記述
 - 逆問題を解くことが目標

言語モデル



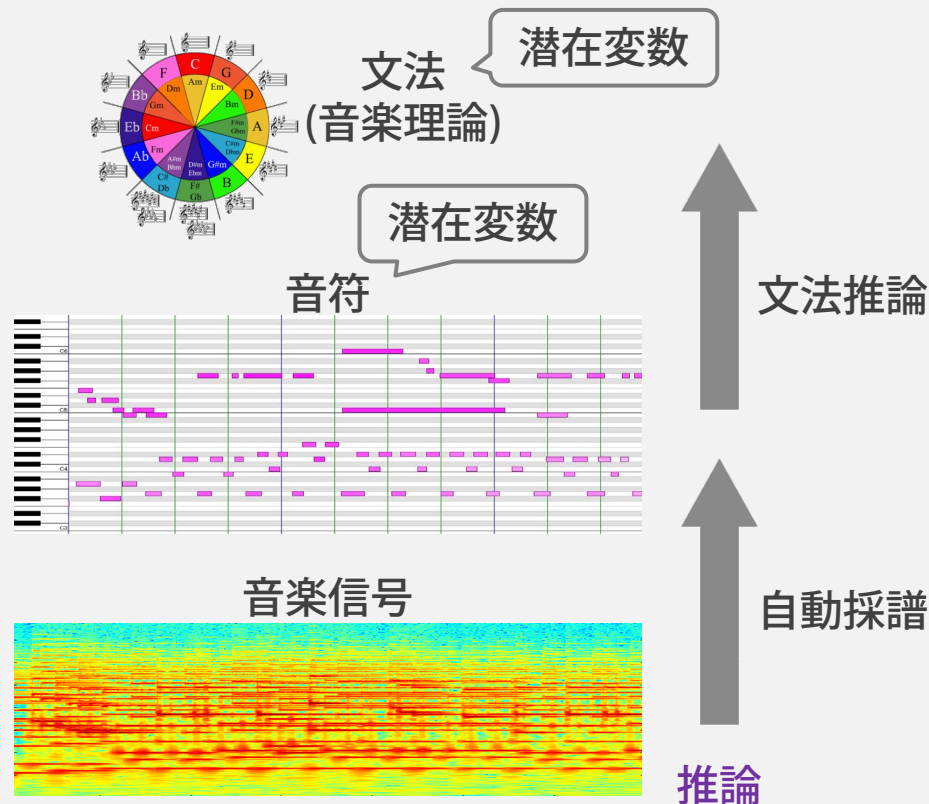
作曲

音響モデル



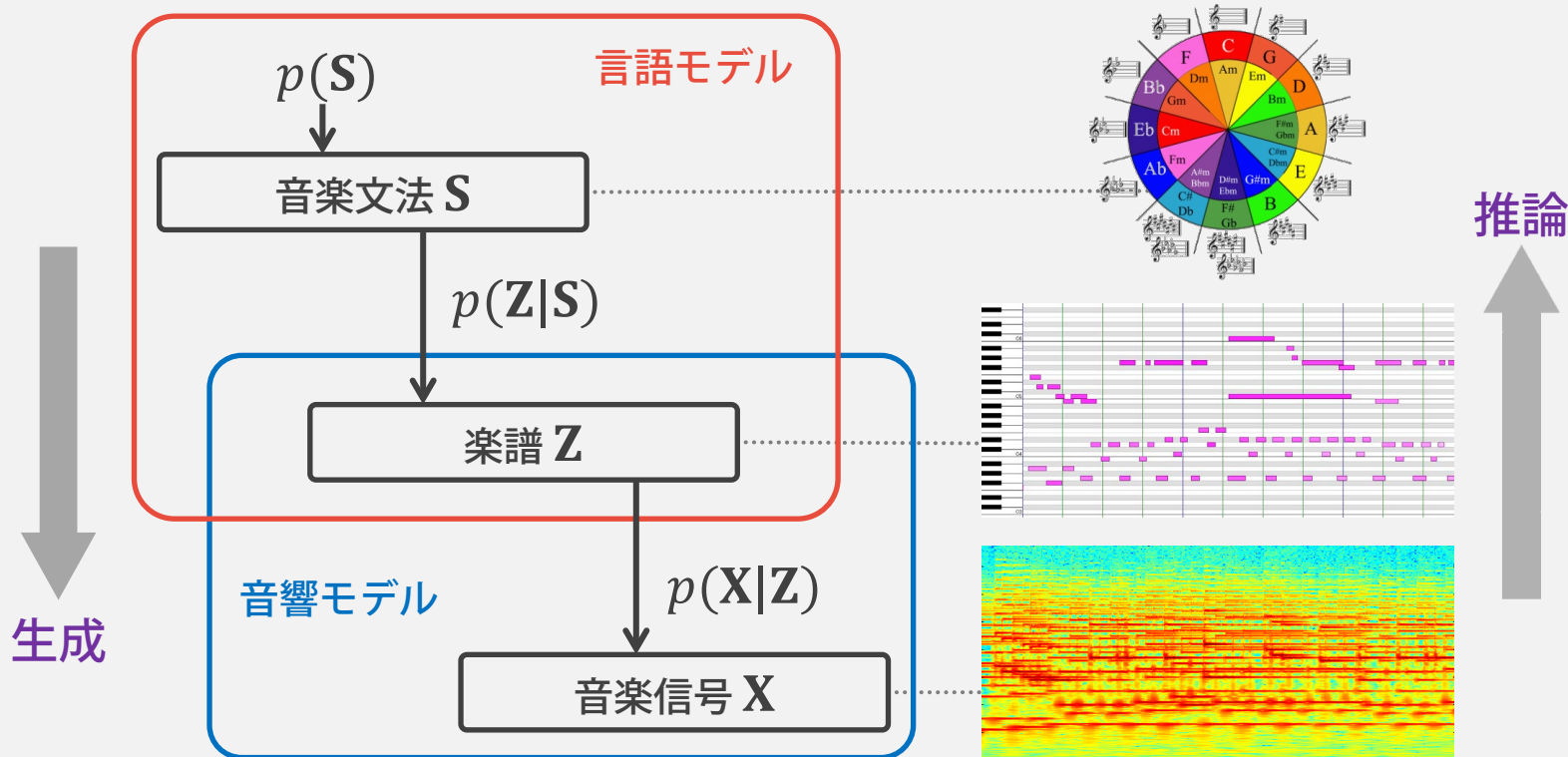
演奏

生成



“深層”生成モデルに基づくアプローチ

- 音響モデルと言語モデルを統合した確率モデル
 - 深層ニューラルネットを「適材適所に用いて」表現力を向上





研究構想

— 物理ベース深層学習 —

キラアアプリケーション創出

- 視聴覚メディアの究極形を提案・実現

- 過去は変えられないという常識が動揺するような体験を創出

従来のメディア

過去の鑑賞
(レッドオーシャン)



究極のメディア

過去の追/再体験
(ブルーオーシャン)

- 再生デバイスを通じた間接的体験
 - リアル (高精細・高忠実) の追求
- 固定視点での鑑賞
 - ユーザの位置・向きは固定

- 自らの視聴覚による直接的体験
 - リアリティ (真実味) の追求
- 自由視点での鑑賞
 - ユーザは自由な移動が可能
- 過去への干渉
 - 「たら」「れば」を具現化

2024開始



ニューラルタイムマシン：
時空間視聴覚場の分析合成系

ニューラルタイムマシン

• 拡張現実 (AR) スマートグラスを用いたタイムマシン

- 過去にその場に存在した人・物体に対する

- 自由視点鑑賞 (追体験)
- リアルタイムインタラクション (再体験)

▪ 使用例

- リビングルームに多数のカメラやマイクを設置
- 日々の家族の団欒を記録・時空間視聴覚場を推定
- あとから時空間をジャンプして任意時間・座標での観測信号を再現
 - 子供が声や音を出しながら走り回る姿を別の角度で眺める
 - 体を押したら、驚いた表情で振り返って会話してくれる
 - 究極的には触覚・味覚・嗅覚も再現



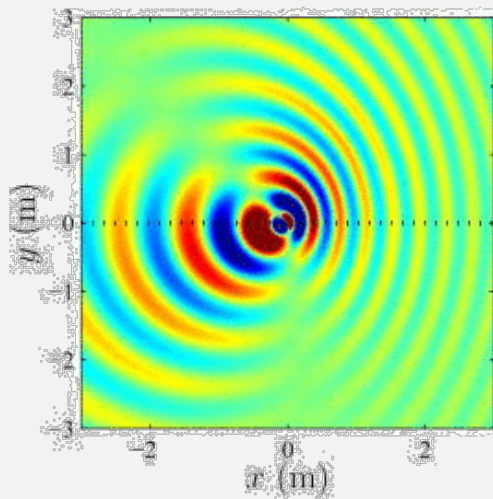
映画“Kingsman”の一場面

時空間上に定義された
輝度・音圧の連続関数

場の計測・推論

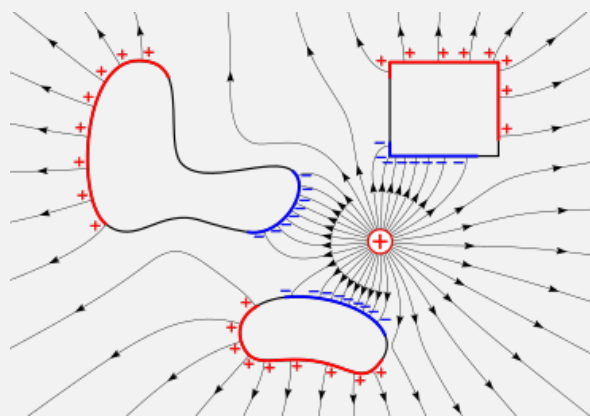
- 時空間上に定義された連続関数 $f(x, y, z, t)$ を推定したい
 - 工学アプローチ：機器による測定 → 時間・空間分解能に限界
 - 情報学アプローチ：統計的推論 → 計測限界を突破

融合が有望



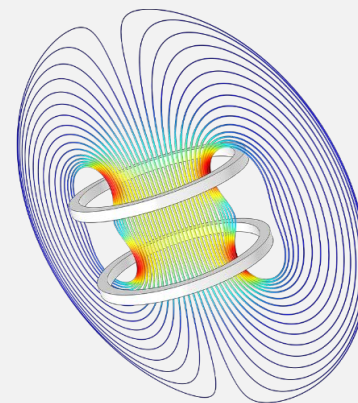
音場

ヘルムホルツ方程式により記述



電場

マクスウェル方程式により記述



磁場



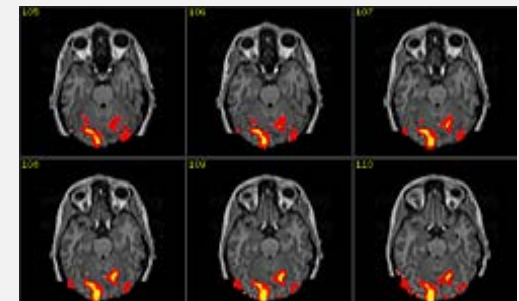
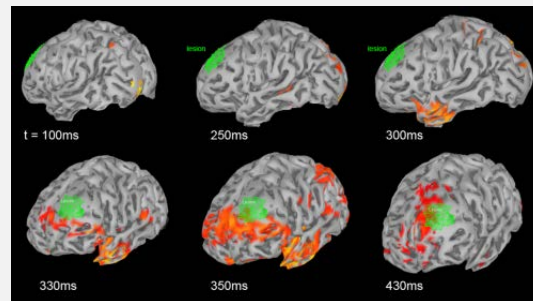
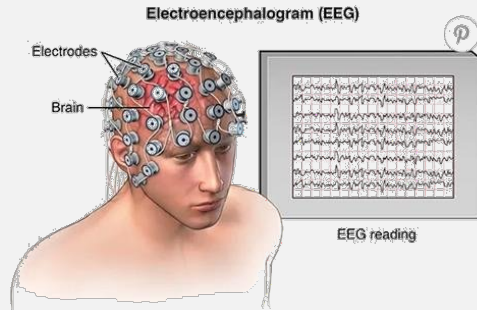
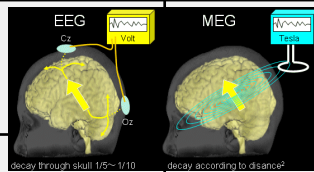
研究構想 — 生体信号処理 —

脳信号計測・解析

工学＋情報学の融合による計測限界の突破

- マルチモーダル計測 → **統一的生成モデル**による情報統合

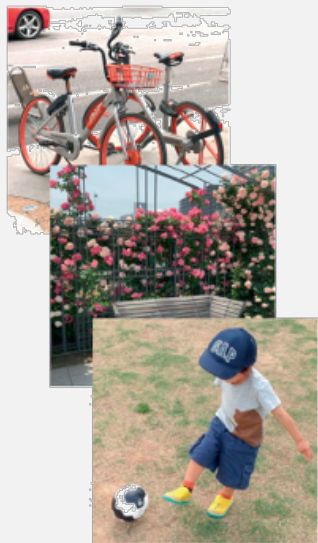
<p>脳波計測 (EEG)</p>	<p>脳磁図 (MEG)</p>	<p>磁気共鳴機能画像法 (fMRI)</p>
<p>神経細胞の 電気活動の総和を観察</p>	<p>神経細胞の電気活動で 生じる磁場を測定</p>	<p>脳の血流変化で 生じる磁場を測定</p>
<p>脳深部の活動も検出</p>	<p>主に脳表層の信号を検出</p>	<p>主に脳表層の信号を検出</p>



脳情報処理の新展開

- 変分自己符号化器 (VAE) の定式化と推論
 - 脳エンコーディング+脳デコーディングの同時学習

自然で多様な
知覚・認知体験
(例：動画視聴)



順問題：脳活動計測 (従来)

符号化 (エンコード) モデル

視覚特徴
聴覚特徴
言語特徴
情動特徴

逆問題：原因推定

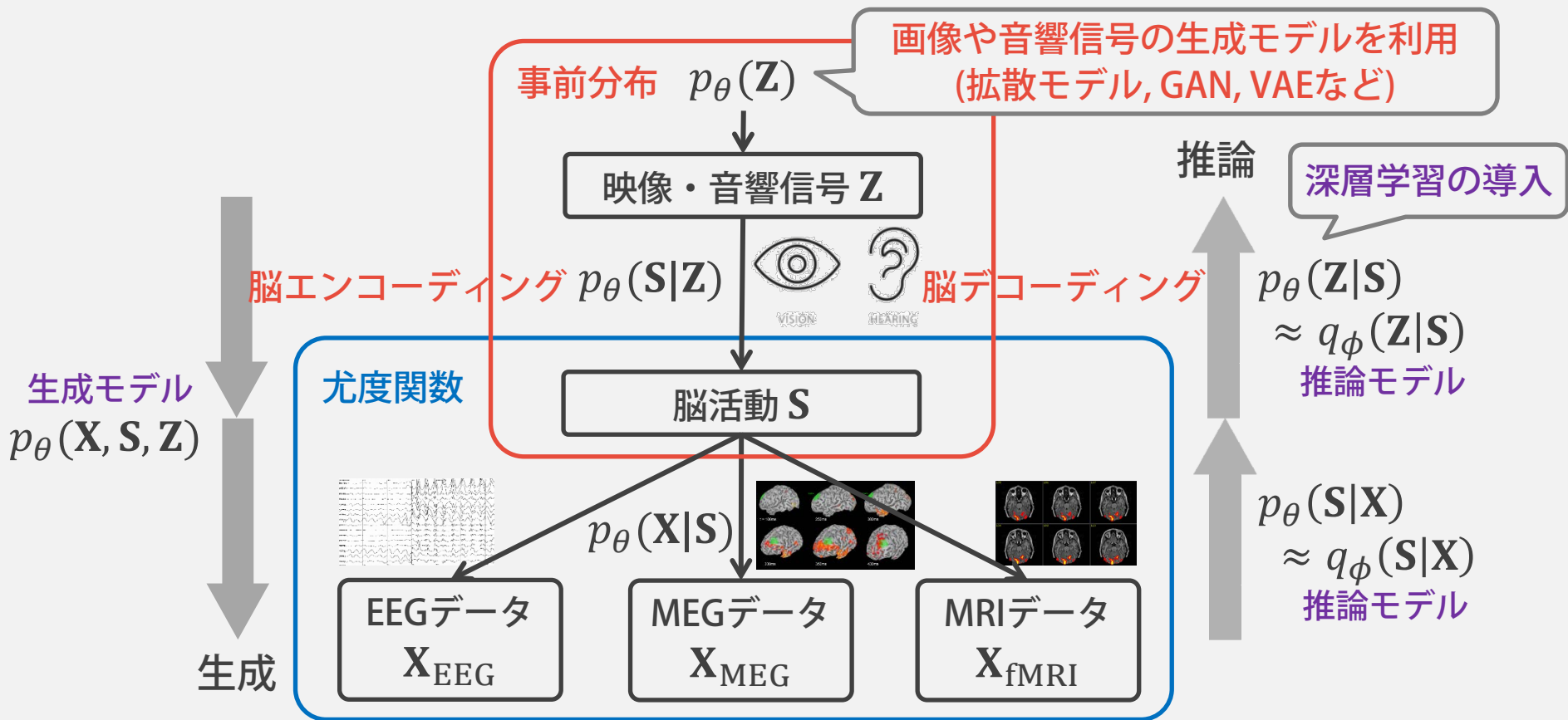
逆符号化 (デコード) モデル

脳神経活動
(例：fMRI計測)



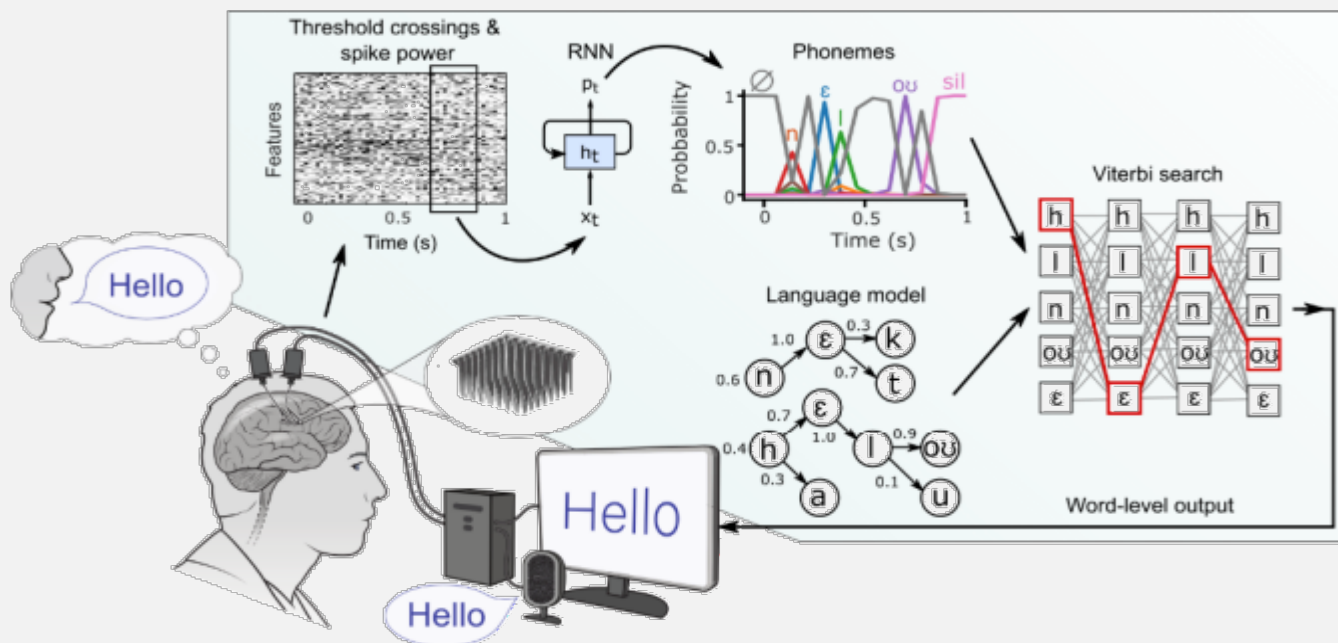
統計的視聴覚脳情報処理

- 知覚と計測を内包する階層生成モデルの定式化・推論



ブレインマシンインタフェース (BMI)

- ARグラス向け環境理解支援システムへの搭載を目指して
 - 機器をフリーハンドで操作・想像した文字を入力
 - 計測技術と情報技術の両輪での研究開発が必須



<https://www.biorxiv.org/content/10.1101/2023.01.21.524489v1.article-info>