



メディア工学領域 マルチモーダル研究 室

菅木 禎史
ちさき よしふみ

教授

津田沼キャンパス 1号館9階



自己紹介

菅木 禎史（ちさき よしふみ）

- マルチモーダル研究室ですが、原点は音響屋です。
- 音を中心とした興味ある研究を実施しています。
- 卒業論文 → マイクロフォンアレー
- 修士論文 → マイクロフォンアレー
- 博士論文 → バイノーラル関連と実時間信号処理
Real-time speech processing on pitch estimation
and speech enhancement using binaural information
- 研究領域の変遷
電気音響，デジタル信号処理
 - + マルチモーダル
 - + 音声特徴抽出，音声秘話
 - + バイノーラル信号処理，騒音制御(ANC)
 - + 屋外拡声



自己紹介

これまでの外部の業務

- 日本音響学会 理事，役員，編集委員長ほか複数の委員会委員長，委員
- 電子情報通信学会 基礎・境界ソサエティ 副会長

- 国際協力機構(JICA) 高等教育移転関連業務 プロジェクト副総括
- 国際会議 運営業務各種委員

- 東北大学 電気通信研究所 客員教授

•所属学会

日本音響学会(ASJ)

電子情報通信学会(IEICE) 2009年～シニア会員

人工知能学会(JSAI)

米国電気電子学会(IEEE)

米国音響学会(ASA)

情報処理学会(IPSJ)

日本教育工学会(JSET)

超臨場感コミュニケーション産学官フォーラム特別会員

The Asia-Pacific Signal and Information Processing Association (APSIPA)

日本騒音制御工学会(INCE-J)

Audio Engineering Society(AES)



マルチモーダル情報処理とは

マルチモーダルとは、視覚や聴覚など複数の感覚（入力）のことです。

人は複数の感覚の情報に基づき認知して、生活をしています。

昨今では、マルチモーダルAIという言葉もあり、
AIでも複数の入力を活用しています。

知覚、理解、思考をコンピュータで模倣もしくは人以上の性能で実現できると、豊かな世界の構築が期待できます。



例) 聴覚の特性を研究した成果と応用

例えば、聴覚特性であるマスキング効果といわれる、

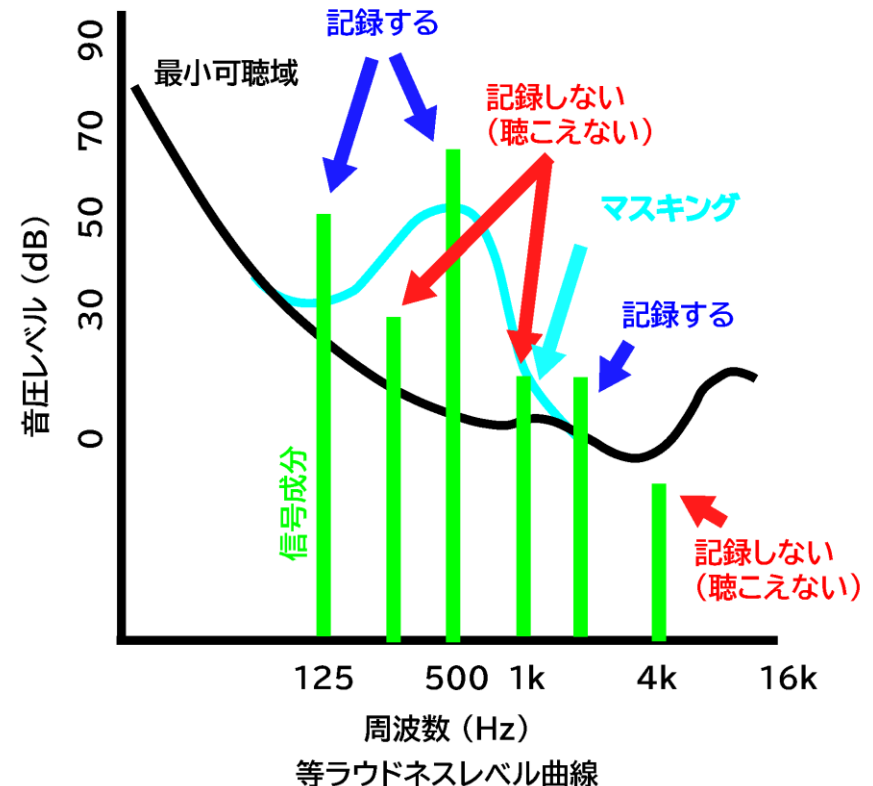
「ある音が別の音に妨害されて聞きづらくなる」

ということを人に実験音を聞かせて明らかにしました。

“The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear”, R. L. Wegel & C. E. Lane, Physical Review, 23(2), 266-285, 1924

(100年前！)

これは、みなさんの身近なことへの応用としては、日頃利用しているスマホなどで聞いている音楽の情報圧縮において聞こえない音は記録・伝送しないことに活用されています。



マルチモーダルの特性を研究

この例のように、人の特性を明らかにし、工学分野で応用されていることは、マルチモーダル処理においても存在します。

VR/ARでの没入感を向上させるために、視覚と聴覚の相互作用（クロスモーダル効果）を活用

例)

- 仮想空間での音像生成とアバターの連携
- オブジェクトのリアリティ性の向上
- 聴覚が視覚を補完し、触れていなくても触れているように感じさせる触覚的リアリティなど

巨木研での研究例 その1

空間音響技術を活用した 仮想空間の臨場感向上に関する研究 (以前に企業と研究したテーマを継続)



直近では、仮想空間内に見えている
オブジェクトの放射特性を考慮して、
音の再現をよりリアルにしようとしています。
(2025.3 オープンキャンパスでデモ)



- ・プログラミングスキルが必要 Unity, C#
- ・選択科目：離散数学(B2前期), 後続のメディア基礎, 音響工学基礎・応用 7

巨木研での研究例 その1

1



この状況を
数式化できる
ように整理

数式化し、
実験プログラム
作成

受聴者の正面

発話者

θ_2

θ_1

発話者の
回転角度

受聴者

発話者の方向



仮説を検証する
そのために、実験パターン
(研究開始時に設計済) に沿った実験



統計処理、可視化して、
議論

- 仮説検証
- PDCA



巨木研での研究例 その2

空間音響技術，AIによるオブジェクト認識を活用した，仮想空間でのDAW(Digital Audio Workstation)の開発を行っています。

現実空間でカメラで撮影した動画内の複数オブジェクトをAIで認識させ，それらを仮想空間内に配置し，推定した空間情報に基づき音像をレンダリングを行い，作り手の意図を簡便に反映させることができ，臨場感を向上させることを目指しています。



既存のDAWの例
左：水平面上から
右：奥行

菅木研での研究例 その 2

仮想空間でのDAW(Digital Audio Workstation)の開発

- ・画像内のすべてのオブジェクトの音をレンダリングするには
消費電力の増加が生じる（スマホなどのバッテリー消費が速くなる）

↓ 演算量を減らし、効率的なサウンドスケープを生成したい

- ・人は見ている画面において、どのオブジェクトに注意し、注意の向いていないオブジェクトの音はサウンドスケープの構成要素として必要か否か？

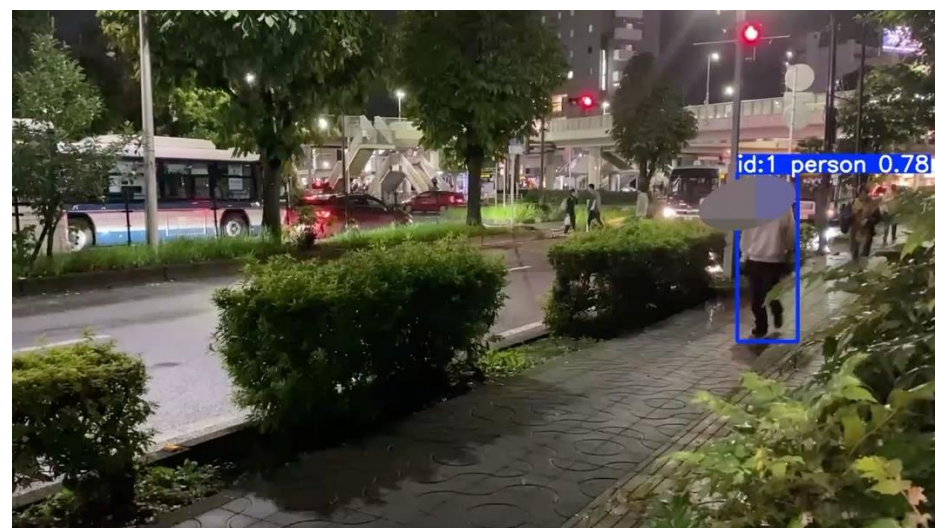
↓ 必要な音だけを複雑でも適切な処理を行い、
注意の向いていない音源の音像はシンプルな処理で済ませると演算量が減る
（先のマスキング効果の応用と同じ流れです）

人がどのような場面で、何に注意を払っているかを研究する必要がある

巨木研での研究例 その2



原動画



オブジェクト検出結果

このオブジェクト検出結果から、
相対的な位置情報を推定し、
オブジェクトに適した音を生成する

サンプル動画

このような動画を仮想空間内で編集することを目指しています。
音像の距離制御，移動を補正，不要な音の低減等を可能とし、
作り手の意思を簡便に反映させることができるシステム開発を目指
す。

巨木研での研究例 その1

2



実写を模倣した動画

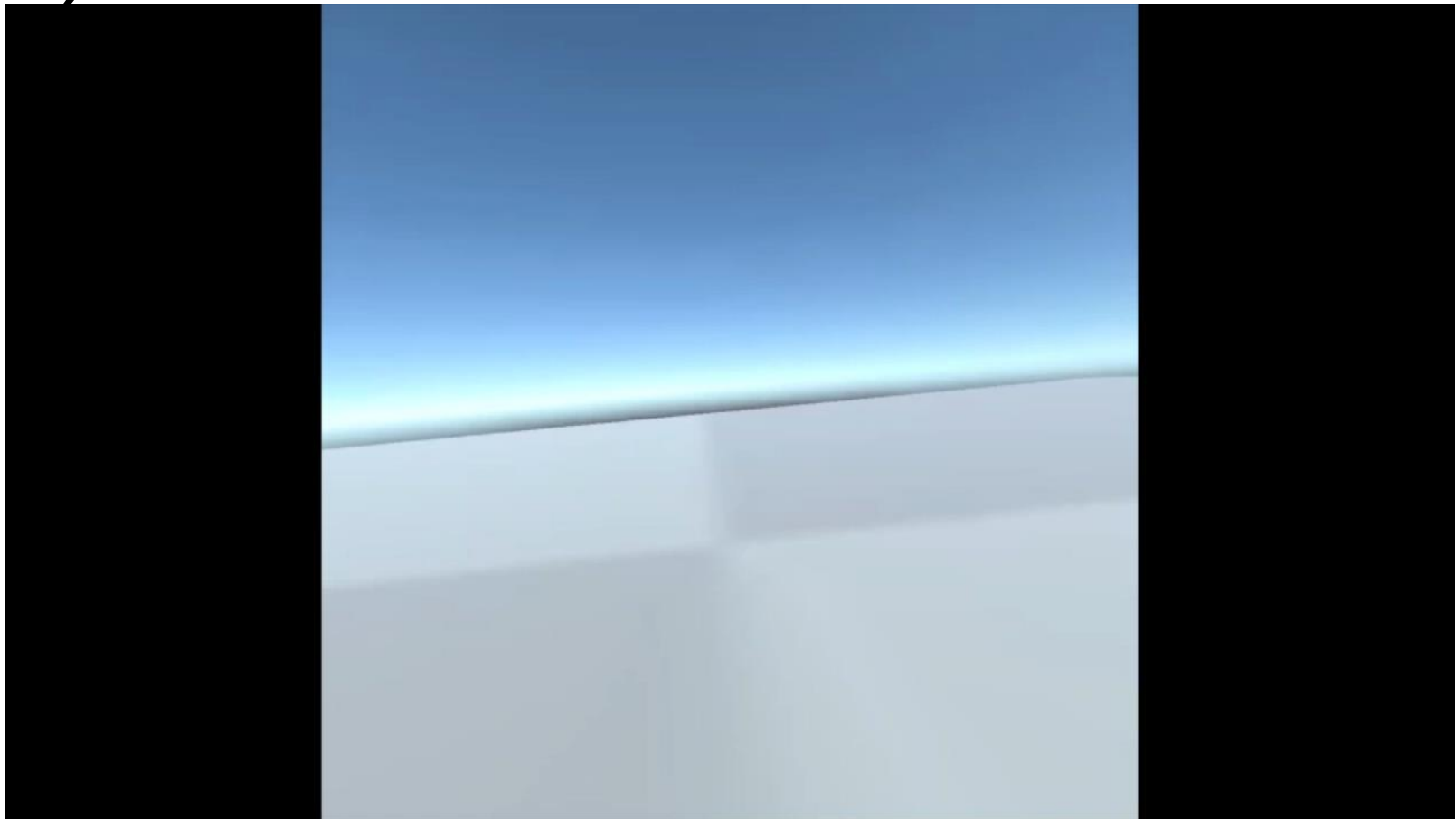
推定誤差などの議論をするために、オブジェクトの位置情報などが無いの実写ではなく、位置などを設定したシミュレーション動画を作成。



オブジェクト認識後、空間音響技術で音を付与。ヘッドフォンで聞くと、音像がオブジェクトに合わせて、左←→右に移動しています。

自動車が人の後ろに隠れても音が途切れずに生成できている！

巨木研での研究例 その2



仮想空間で音の編集をするイメージ

先程のカメラで撮影した動画を仮想空間に写像し、
推定した軌跡を編集できるようにし、
仮想空間内でポストプロダクションを目指す！

巨木研での研究例 その 3

空間認識における人の特性分析とトレーニング方法の開発

このテーマは、知能メディア工学科の学生が自らが、卒業研究から大学院修士までの3年間で研究計画を立てて、学生自身が研究テーマを自ら立ち上げて、巨木がサポートしているテーマです。

指導教員から与えられたテーマではなく、自ら興味を持つことに取り組みということもできる実例です。

- 空間認識力は、3Dモデリング、建築関係、スポーツ選手などと幅広い職業に関連している。

- なぜ、目的地に辿り着きにくいのか？

自宅から学校/職場の間でさえも間違える人がいるのか？

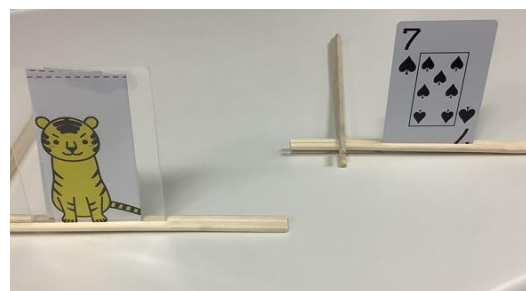
巨木研での研究例 その3



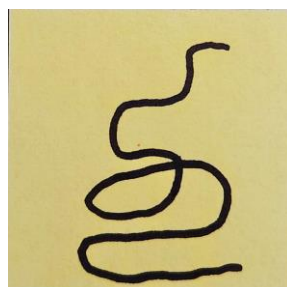
指示書に基づき、歩く。
途中、何に着目している
かビデオを記録



ランドマークに相当する各種
マーカーを設計した場所に置き、
経路記憶のヒントにさせる



覚えられない人は、認知
地図が頭の中に形成され
ないのでは？



←このような
ヒントを与えて、
再実験

統計処理、可視化して、
議論

- 仮説検証
- PDCA



菅木研での研究例 その4

防災無線 音ずれ解消

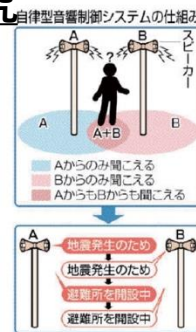
自動で文節分割順次放送

音声インタラクション・インターフェースに関する研究

2011年の当方震災時に避難指示などの情報があらゆる人に確実に届かなかったこと、特に屋外拡声は様々な問題があったため、屋外拡声について総務省プロジェクト研究費で研究をしました。

その後、科学研究費助成事業の研究費で、次のフェーズとして、屋内においても、スマートスピーカーなどの普及に伴い、新しい音声プッシュシステムと音声インタラクションの研究に発展させました。行政からの情報は緊急防災情報のみならず、日常生活に関連する情報を提供しているなかで、ユーザー個々に適応するインタラクション・インターフェースを研究しています。

2024年度の卒業研究で、アマゾンアレクサへ実装し、AIを活用して情報提示内容、方法を検討しました。



菅木 靖史 教授
熊本地震で「聞きづらい」場所では聞き手との距離に応じて音がずれる。熊本地震では、この現象をある程度「聞きづらかった」との声が、熊本県菊池市の44軒など多数寄せられた。一部の自治体は、このまま現象を生みやすい隣接地域では同時に流さない防災無線が約1650カ所ある。止策を取る。ただ、内容が数十秒に及ぶ、情報の入手で被災者間に「時間的格差」を生む欠点があった。

熊本地震で「聞きづらい」

災害時に避難などを呼び掛ける防災無線の屋外放送が、たまのようにならなくて聞き取りにくい現象を解消し、明瞭な音で伝えられる自律型音響制御システムを菅木 靖史 千葉工大教授（金澤情報学）が開発した。この現象は熊本地震でも情報伝達の障害になった。防災無線は高層ビル中心に1カ所あり、早期の製品化を目指している。

千葉工大教授開発

めた熊本六の大学院准教授 技術を用い、放送が発時代、放送が発した文章を瞬時に節で区画を細かく分け、文節単位で切り、音声データで保存。各スピーカーから順に流し、この現象が予測される場た方が時間的格差を軽減。所のアリタイで放送を始震の発生により避難所を明確められるようにした。製品設しました。を方所のス化は、既存の設備を借用でスピーカーで流す場合、1カ所当たり約10万円。製品の発生にの避難所を。を。菅木教授は「伝達に影響開閉したに文章を区する建物などの障害物も感切った上で、①を2カ所で、②を同じ放送できるような程度交互に放送した。③を同じ放送できるような程度に流すシステムを開発し、（古田 紀）と話した。

西日本新聞, 2017年

菅木研での研究例 その5

音響信号処理に関する研究

知能メディア工学科の授業では十分に取り扱いえていない、リアルタイム信号処理の研究も行っています。

直近では、みなさんの日常利用が普及しつつあるアクティブノイズコントロール技術が一例です。

大学のカリキュラムにあるデジタル信号処理以上の内容を自ら深く学ぶ必要がありますが、菅木は指導が可能です。

2024年度に卒業研究でリアルタイム信号処理，デジタルフィルタの卒業論文を書いた学生がいます。

苮木研での研究例 その6

教育工学に関する研究

2024年度に卒業研究で、LMS(Learning Management System)のデータを活用して、授業改善、躓いている学生の早期検出などを検討し、卒業論文を書いた学生がいます。

必要とされる言語は、SQL, PHP, JavaScript, Pythonなど

対面で説明します 研究室配属後

【3年後期 6S】

実質11週

(短すぎて、ポイントだけしか教えられない、自主勉強が必須)

●ゼミナール1

- 卒業研究に向けて必要な文献調査能力や読解力を高め、基礎知識と技術を習得する。(論文を読む)
- プログラミング能力の向上

●卒業研究テーマ決め

【4年前期 7S, 4年後期 8S】

実質25週

(研究というには短すぎる。院進学が望ましい。いずれにせよ、自発的行動が必須)

4月～7月, 9月中間発表, 10月～12月卒論

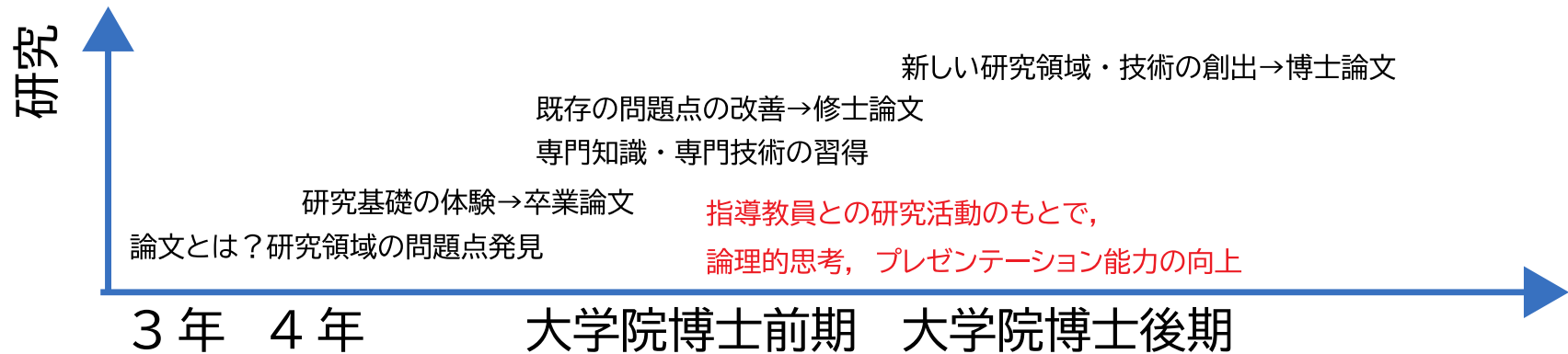
提出

●卒業論文

ゼミナール2, 3

- 卒業研究を進めていくために必要な基礎知識を習得する。(論文を読む)
- + 就職活動, 中間発表

対面で説明します 研究室配属後



世界中、どこの研究室でも求められること

- 自発的活動
- 自ら調査し、思考し、実践する
- 指導教員とのディスカッション
- 成し遂げようとする熱意



まとめ

音を中心としたマルチモーダル研究室で研究を遂行するには、
知能メディア工学科の複数の領域の科目履修が必要です。

- ・音響に関する科目
音の物理の理解，自分で実験音を作成・加工することが必須
- ・AIに関する科目
オブジェクト認識のみならず，雑音低減，音源分離にも必要
- ・UX/UIに関する科目
すべてにおいて，研究成果の利用者は人です。音や画像を提示する，
加工する際にインターフェースが適切であるか否かを検討するために
必要
- ・プログラミングに関する科目
卒業研究を始める時点では，自分でアルゴリズム開発，実装，デバッ
グの
基礎知識があり，必要に応じてChatGPTなどを活用して
プログラミングできることが必要
研究室ではプログラミングを教えません（教えている時間がありませ
ん）
- ・英語
どの研究においても，英語の論文を読む場面があります。
AIなどを活用して良いので，読み書きができることが必須
- ・統計学
実験結果を適切に評価する力が必要。あわせて，分析力も必
要