

# 加速度センサを活用した状況認識 技術とヒューマンインタフェースへ の応用

A Study on Context Awareness Using  
an Accelerometer and Its Applications  
to Human Interface

2017 年 3 月

大内 一成

Kazushige OUCHI



# 加速度センサを活用した状況認識 技術とヒューマンインタフェースへ の応用

A Study on Context Awareness Using  
an Accelerometer and Its Applications  
to Human Interface

2017年3月

早稲田大学大学院 先進理工学研究科

大内 一成

Kazushige OUCHI



# 概要

## ● 研究背景

ユビキタスコンピューティングという言葉が聞かれるようになって久しいが、直訳すると「遍在するコンピューティング」となることから、コンピュータを実世界の至る所に存在させることであると理解されている場合が多い。しかしながら、Mark Weiser が 1991 年に発表した *The Computer for the 21st Century* という論文で最初にこの概念を提唱した際、その副題は「Specialized elements of hardware and software, connected by wires, radio waves and infrared, will be so ubiquitous that no one will notice their presence」であり、本文は「The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.」という一文で書き始められている。すなわち、実世界の至る所にコンピュータを存在させるだけでは不十分で、それらのコンピュータの存在を意識することなく、その恩恵を享受できる仕組みを提供することがユビキタスコンピューティングの本質である。

コンピュータの存在を意識することなくその恩恵を享受するためには、コンピュータが我々人間と周囲の状況を理解し、その人のその時点の状況に応じて適切なサービスを適切なタイミングで提供することが肝要となる。本論文では、コンピュータが人間の状況を理解する技術を状況認識 (Context Awareness) 技術と呼び、その人の行動や健康状態、さらには機器を操作しようとする意図 (これらを Context と総称する) を認識し、それに基づいた適切なアクションをタイムリーに提供するサービス、機器操作の支援を提供するヒューマンインタフェースを提案した。

## ● 本論文の構成

第 1 章「序論」では、まず、ユビキタスコンピューティングおよび IoT (Internet of Things) の本来の概念を整理し、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ことが、今後のコンピューティング技術が目指すべき方向性であることであると位置づけた。その上で、Situation Awareness (状況認識) について説明した Endsley のモデルに基づき、コンピュータの存在を意識することのないサービス提供実現へのアプローチとして、誰にでも扱いやすいインタフェース、さらには操作しないインタフェースが必要であることを示した。その実現を目指し、コンピュータが人間の状況を理解する状況認識技術と、それに基づいたヒ

ユーザインタフェース応用に関する研究に取り組んだ本論文の貢献について説明した。

第2章「ウェアラブル健康管理システム」では、まず生活習慣病／メタボリックシンドローム対策の現状分析を医療従事者へのアンケート調査等により実施し、生活習慣の自己管理（行動変容）をサポートするシステムの必要性を示した。その上で、手軽で確実な生活習慣管理を実現するシステムコンセプトを検討し、腕時計型ウェアラブルセンサと外部情報端末を連携させたウェアラブル健康管理システムを提案した。現在、ウェアラブルデバイスとスマートフォンを組み合わせた活動量計サービスが多く実用化されているが、それらの草分け的な取り組みである。具体的には、加速度、脈波、発汗、皮膚温などの多様な生体情報をセンシング可能なウェアラブルセンサを開発し、それを用いた生活習慣、健康状態に関わる状況認識技術を開発した。歩行、走行などの行動と歩数、および脈拍数などの生体情報の見える化だけでなく、ウェアラブルセンサで健康管理上重要である食事イベントを精度良く検出可能にした点が特長である。食事イベントを検出することで、食事のタイミングに基づいたサービス（食事メニュー入力支援、服薬管理、血糖値計測管理など）を実現した。また、取得したユーザの状況に基づいて、生活習慣改善（行動変容）のためのアドバイスをタイムリーに提供するシステムの意義について論じた。

第3章「ウェアラブル睡眠センサとその応用」では、近年不眠に悩む人が増加し、社会問題化しているにも関わらず、睡眠状態を詳細に把握するためには、専門施設に宿泊し、脳波計など多数のセンサを身体中に装着する終夜睡眠ポリグラフ検査を行う必要がある現状から、自宅で手軽に睡眠状態をチェック可能なセンサの必要性を確認し、睡眠計測にフォーカスしたウェアラブル睡眠センサを開発した。内蔵した加速度センサと脈波センサの情報から詳細に睡眠状態を解析可能とする状況認識技術の研究を行った。まず、加速度の情報から睡眠／覚醒を判定する。次に、脈波からロバストに取得した脈波間隔の周波数解析により自律神経の状態を推定し、自律神経と睡眠状態の相関関係から REM 睡眠／NREM 睡眠と、睡眠深度などの詳細な睡眠状態を判定するアルゴリズムを提案した。手首における脈波計測に適した波長が緑色であることを確認し、ノイズに強い独自の脈波間隔計測手法と合わせて手首における安定した脈波計測を実現した点、個人差や日内変動の影響にロバストなクラスタリング手法を提案した点などが技術的なポイントである。被験者 45 名による終夜睡眠ポリグラフと本センサの同時計測実験での一致度（ポリグラフの結果を専門医が目視で判定した結果と本センサによる自動判定の結果）は 73.5%で、在宅で使える簡易モニタとしては十分な性能を達成し、自宅や外出先での手軽な睡眠状態

把握を実現した。さらに、快適に目覚められるタイミング（睡眠状態）を調査した結果、REM 睡眠時に起床させると快適に目覚められる傾向にあることがわかり、日々の生活で快適な目覚ましを実現できる可能性を確認した。

第4章「スマートフォンを用いたリアルタイム生活行動認識システム」では、普及が進んでいるスマートフォンに一般的に搭載されている加速度センサと、通話用のマイクを音センサとして積極的に活用し、特別なセンサ／デバイスを使用せずに、当該アプリケーションをインストールしたスマートフォンを持ち歩くだけで様々な日常生活行動を連続的に把握する状況認識技術を提案した。スマートフォン／携帯電話搭載センサを用いた既存研究では、主に屋外の移動状況にフォーカスした行動認識が行われていたが、屋内の様々な生活行動の高精度な認識を実現した点がポイントである。まず加速度をもとにユーザの状態を「歩行」「作業」「安静」の3状態に大まかに推定し、「作業」と推定された場合に、マイクからの音を分析することにより作業内容を細かく分類する。また、生の認識結果だけでなく、同一作業区間単位で行動の種別を再評価する事例ベース認識手法の導入により認識性能が向上し、「歯磨き」「トイレ」「掃除機がけ」「皿洗い」など、これまで加速度センサだけでは認識が難しかった様々な生活行動を精度良く認識することを可能にした。高齢者も含む21名による評価実験を行い、「皿洗い」「アイロンがけ」「掃除機がけ」「歯磨き」「ドライヤー」「トイレ水洗／手洗い」の6種類の生活行動を平均91.4%の精度で認識できることを確認した。また、ロボットインタフェースとの連携効果、離れて暮らす家族のさりげない見守りサービスなど、提案手法を応用したサービスの可能性についても述べた。

第5章「センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用」では、音声認識入力を日常的に使うために望ましいハンドヘルド型マイクにおける課題として、ヘッドセット型マイクに比べて認識精度が劣ること、音声認識の開始と終了をユーザがボタン操作で指示する必要があることを挙げ、その解決方法として加速度センサと距離センサを内蔵したハンドヘルド型マイクを開発し、それに対するユーザの発話動作を検知する状況認識技術を提案し、両課題に対する対策とその効果について論じた。また、現状非常に手間がかかるTVの番組検索操作に提案手法による音声認識入力を活用するとともに、人名と番組名の言い換えに対応する仕組みを新たに構築することで、従来のリモコン操作による番組検索よりも、操作完了までの所要時間を約40%短縮したヒューマンインタフェースについて述べた。

第6章「結論」では、上述の各研究成果を総括した結論として、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ためには、加速度センサ+ $\alpha$ によるクロスモーダル状況認識が重要であることを述べた。意図のある人物動作／行動には、その多くの場合に身体のダイナミックな動きが伴い、これを加速度センサで捉えることが状況認識の基本となる。その上で、目的・ユースケースに応じて、認識対象の動作／行動の粒度を細かくする、精度を上げる、あるいは自律神経系など人間の内部状態を把握する必要がある場合に、それらの取得に適したセンサを加速度センサと組み合わせて使用するクロスモーダルな状況認識が望ましい。さらに、現在の動作／行動の認識から、将来の動作／行動を予測することで、ユーザの意図を先読みし操作不要で所望のサービスを提供することが可能になるであろう。このためには従来の決定論的モデルによる **Reactive Computing** から、確率論的モデルによる **Proactive Computing** をクロスモーダル状況認識により実現する今後のさらなる研究の進展が期待されることを述べた。



# 目次

---

<b>1. 序論</b> .....	<b>1</b>
1. 1 研究背景 .....	1
1. 2 本研究の目的 .....	2
1. 3 本論文の構成 .....	4
<b>2. ウェアラブル健康管理システム</b> .....	<b>7</b>
2. 1 はじめに .....	7
2. 2 従来研究 .....	7
2. 3 コンセプト検討 .....	8
2. 3. 1 アンケート調査 .....	8
2. 3. 2 ウェアラブル健康管理システムのコン セプト .....	12
2. 4 システム構成 .....	13
2. 4. 1 全体概要 .....	13
2. 4. 2 腕時計型ウェアラブル生体センサ .....	14
2. 4. 3 生体センサでの状況認識 .....	16
2. 4. 4 PDA での状況認識 .....	18
2. 4. 5 タイムリーな情報提供 .....	22
2. 5 まとめ .....	25
<b>3. ウェアラブル睡眠センサとその応用</b> .....	<b>26</b>
3. 1 はじめに .....	26
3. 2 従来研究 .....	26
3. 3 ウェアラブル睡眠センサの開発 .....	27

3. 3. 1	心拍変動を用いた睡眠解析	27
3. 3. 2	手首におけるロバストな脈波計測手法	28
3. 3. 3	自律神経解析手法	33
3. 3. 4	自律神経活動からの睡眠解析手法	34
3. 3. 5	ウェアラブル睡眠センサ	35
3. 3. 6	評価実験	37
3. 4	<b>快適目覚まし機能の検討</b>	<b>39</b>
3. 4. 1	リアルタイム睡眠解析システム	39
3. 4. 2	仮説検証	41
3. 4. 3	快適目覚まし機能	42
3. 5	まとめ	44

## **4. スマートフォンを用いたリアルタイム生活**

### **行動認識システム** ..... **45**

4. 1	はじめに	45
4. 2	従来研究	46
4. 3	<b>加速度と音による状況認識</b>	<b>47</b>
4. 3. 1	認識対象とする生活行動	47
4. 3. 2	想定するサービスイメージ	49
4. 3. 3	システム構成	50
4. 3. 4	屋内生活行動推定エンジン	51
4. 3. 5	屋外移動状況推定エンジン	54
4. 3. 6	評価実験1 (基本性能評価)	56
4. 3. 7	評価実験2 (実用性能評価)	62
4. 3. 8	スマートフォン版アプリの開発	65
4. 3. 9	スマートフォン版アプリによる評価	67
4. 3. 10	装着部位の違いによる評価	69
4. 4	<b>ロボットインタフェースとの連携</b>	<b>73</b>

4. 5	まとめ	75
------	-----	----

## **5. センサ駆動ハンドヘルド型音声認識入力手**

### **法と番組検索への応用** ..... 77

5. 1	はじめに	77
5. 2	ハンドヘルド型音声認識入力の課題	78
5. 2. 1	ヘッドセット型マイクとハンドヘルド型マイクの音声認識精度の比較	79
5. 2. 2	マイクと口元間の距離と音声認識精度の関係	80
5. 2. 3	操作面での課題	80
5. 2. 4	課題のまとめと対策案	81
5. 3	センサの活用による課題解決の検討	82
5. 3. 1	距離センサによる発話動作検出	82
5. 3. 2	加速度センサによる発話動作検出	83
5. 3. 3	実験と考察	84
5. 4	センサ駆動ハンドヘルド型音声認識入力方法	86
5. 4. 1	アルゴリズム検討	86
5. 4. 2	試作デバイス仕様	88
5. 4. 3	性能評価実験	89
5. 4. 4	従来手法との比較評価	89
5. 5	番組検索タスクでの評価	92
5. 5. 1	音声認識システムの課題	92
5. 5. 2	番組検索のための認識語彙生成	93
5. 5. 3	マルチモーダル操作でのテレビ番組検索による実証実験	99
5. 6	まとめ	107

## **6. 結論 ..... 109**

6. 1 本論文の貢献 .....	109
6. 1. 1 ユビキタスコンピューティング実現の ための状況認識技術 .....	109
6. 1. 2 状況認識技術のヒューマンインタフェ ース応用 .....	114
6. 2 今後の展望 .....	115

## **謝辞 ..... 117**

## **参考文献 ..... 118**

## **著者の業績 ..... 124**

論文 .....	124
国際会議（査読有、筆頭） .....	125
国内発表（筆頭） .....	126
解説記事 .....	128
講演 .....	129
特許 .....	131
受賞 .....	131

# 1. 序論

---

## 1. 1 研究背景

ユビキタスコンピューティングという言葉が聞かれるようになって久しいが、「ユビキタス」の語源はラテン語で、直訳すると「遍在する（至る所に存在する）」という意味であることから、コンピュータを実世界の至る所に存在させることであると理解されている場合が多い。

しかしながら、Mark Weiser が 1991 年に発表した *The Computer for the 21st Century*<sup>[1]</sup> という論文で最初にこの概念を提唱した際、その副題は「Specialized elements of hardware and software, connected by wires, radio waves and infrared, will be so ubiquitous that no one will notice their presence」であり、本文は「The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.」という一文で書き始められている。すなわち、実世界の至る所にコンピュータを存在させるだけでは不十分で、それらのコンピュータの存在を意識することなく、その恩恵を享受できる仕組みを提供することがユビキタスコンピューティングの本質である。この本質の認識が広まりきっていない理由として、「ユビキタス」というネーミングが良くなかったとの鋭い指摘もある<sup>[2]</sup>。いずれにせよ、Mark Weiser が提唱した本当の意味でのユビキタスコンピューティングは、まだ広く実現されてはいない。

一方、似た世界観を示すキーワードとして、Internet of Things (IoT: モノのインターネット) という言葉も近年よく使用されている。IoT は、Kevin Ashton が 1999 年（奇しくも Mark Weiser が没した同一年）に最初に使用した<sup>[3]</sup>とされているが、「モノのインターネット」から直感的に類推される、様々なモノがインターネットに接続されることと理解されている場合が多い。しかし、ITU の定義<sup>[4]</sup>によれば、「A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.」とあり、要は、様々なモノを ICT で繋ぐことが目的ではなく、それにより先進的なサービスを実現するための手段であるとされている。

Mark Weiser が目指した「ユビキタスコンピューティング」、その後 Kevin Ashton が使い始めた「IoT」、それぞれの目指すところは共通しているが、どちらも最終的にこれらの技術によりユーザが恩恵を受けることが主眼である点に注意を払う必要がある。筆者は、これらが目指している世界観を統合し、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ことが、今後のコンピューティング技術が目指すべき方向性であると考えている。

## 1. 2 本研究の目的

Mica Endsley は、人々の意思決定プロセスを説明する考え方として、図 1-1 に示す Situation Awareness (状況認識) に基づいた Naturalistic Decision Making (NDM) モデルを 1995 年に提案した[5]。

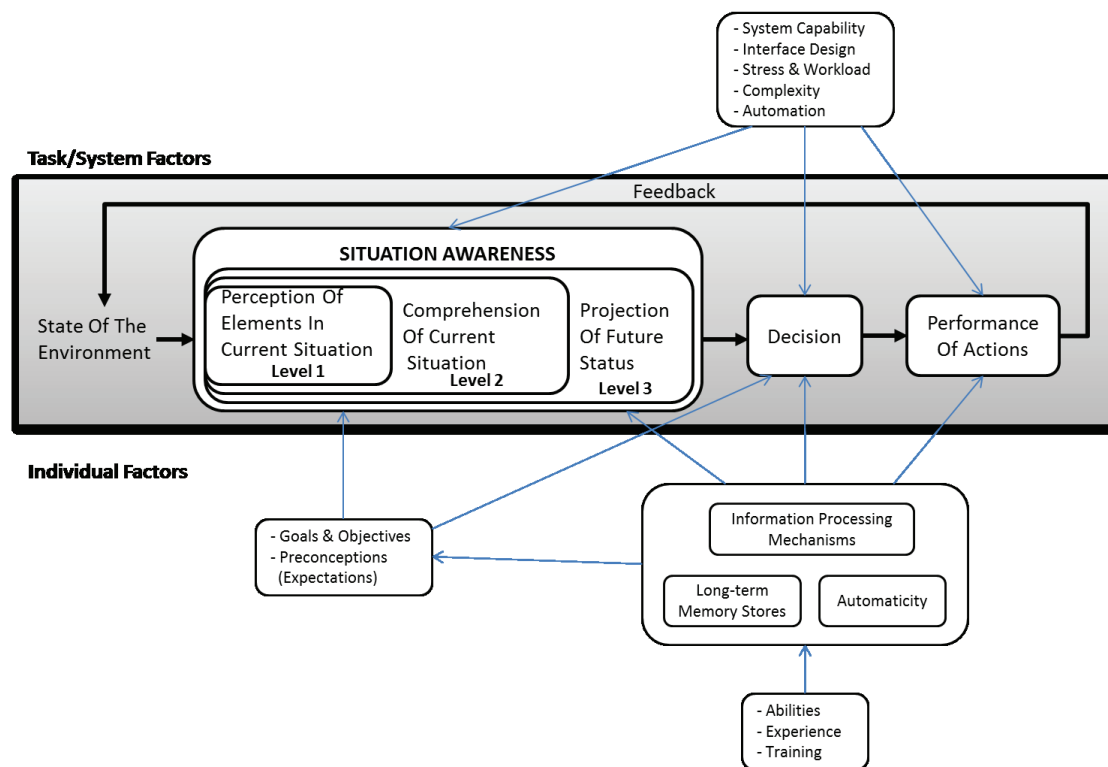


図 1-1. Endsley's model of Situation Awareness (文献[5]を参考に作成)

これによれば、意思決定プロセスには Situation Awareness (状況認識)、Decision (意思決定)、Performance of Actions (行動) の 3 段階があり、それが環境の状態に影響を及ぼし Situation Awareness に再び反映され、さらに Task/System Factors だけでなく、Individual Factors もそれぞれに影響を及ぼすとしている。すなわち、どんな熟練者であっても、状況の認識を誤ると間違った意志決定をしてしまうことが説明されるため、意志決定プロセスにおいて

Situation Awareness を正しく行うことが重要となる。また、この Situation Awareness にはレベルが 3 段階あり、

Level-1: Perception of elements in current situation (現在の状況の知覚)

Level-2: Comprehension of current situation (現在の状況の理解)

Level-3: Projection of Future Status (将来の状況の予測)

の各プロセスがある。正しい状況認識のためには、現在の状況だけでなく、将来の状況を予測することが重要であるとしている。本論文では、この人間の意思決定モデルに基づき、コンピュータの存在を意識することのないサービス提供実現へのアプローチを検討する。

図 1-2 は、Endsley のモデルをベースに、意識せずに使えるヒューマンインタフェースのターゲットを示したものである。

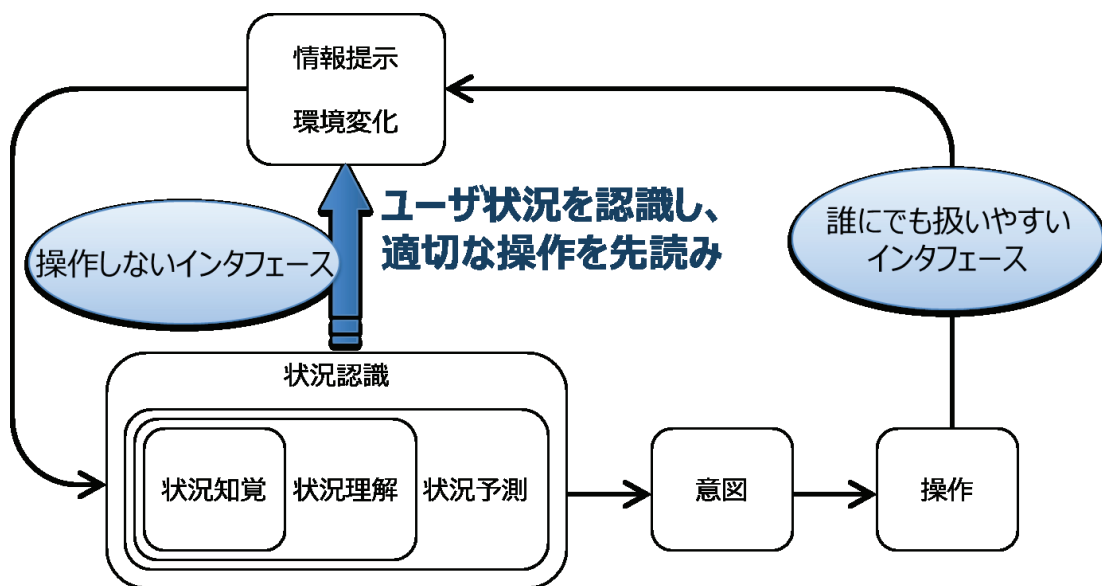


図 1-2. 意識せずに使えるヒューマンインタフェース

ユーザは、自分および周囲の状況を認識して、何らかの意図が生じ、その意図を満足させるための何らかの操作を行い、所望の情報提示／環境変化を得ようとするわけであるが、その最終段の操作方法が煩雑では意識せずにコンピュータを使っているとは言えない。ここでは、説明不要で誰にでも扱いやすいインタフェースを提供することが求められ、ボタンの数が多いリモコンの使用や、目的達成までの操作回数が多い操作をユーザに強いるものであってはならない。例えば、言葉で操作できたり、直感的なジェスチャで操作できたりするなど、機械に対しても人と接するように扱えることを目指したヒューマンインタフェースに関する研究は多く行われている<sup>[6][7]</sup>。しかしながら、すべてを音声のみ、ジェスチャのみで扱うインタフェースは、特定のユースケースによっては扱いやすい場合もあるが、必ずしも様々なユースケースで使い勝手の良さを提供できるとは限らない。あらかじめ決められた操作方法に従ってユーザが操作するアプローチは、コンピ

ュータに合わせて人間が操作する域を出ないためである。ユーザが何かをしようとした時に必然的に行う行動を察知し、適切なタイミングでそれをサポートするなど、ユーザのアクションに基づいたイベント駆動型のアプローチにより、誰にでも使いやすいインタフェースを提供できるのではないか。さらにユーザの様々な状況をコンピュータが常にさりげなく認識することにより、ユーザにとって必要な操作を先読みし、ユーザが機器を操作することなく所望の情報提示／環境変化を得ることが出来るようになるのではないかと考える。

つまり、コンピュータの存在を意識することなくその恩恵を享受するためには、コンピュータが我々人間と周囲の状況を理解し、その人のその時点の状況に応じて適切なサービスを適切なタイミングで（タイムリーに）提供することが肝要である。本論文では、コンピュータが人間の状況を理解する技術を状況認識（Context Awareness）技術と呼び、その人の行動や健康状態、さらには機器を操作しようとする意図（これらを Context と総称する）を認識し、それに基づいた適切なアクションをタイムリーに提供するサービス、機器操作の支援を提供するヒューマンインタフェースに関する研究を行った。尚、上述の Situation Awareness も状況認識と訳したが、ユーザ自身の状況とその前後の文脈も重視する点を踏まえ、本論文では Situation ではなく Context を用いることとする。

## **1. 3 本論文の構成**

第2章「ウェアラブル健康管理システム」では、まず生活習慣病／メタボリックシンドローム対策の現状分析を医療従事者へのアンケート調査等により実施し、生活習慣の自己管理（行動変容）をサポートするシステムの必要性を示す。その上で、手軽で確実な生活習慣管理を実現するシステムコンセプトを検討し、腕時計型ウェアラブルセンサと外部情報端末を連携させたウェアラブル健康管理システムを提案する。現在、ウェアラブルデバイスとスマートフォンを組み合わせた活動量計サービスが多く実用化されているが、それらの草分け的な取り組みである。具体的には、加速度、脈波、発汗、皮膚温などの多様な生体情報をセンシング可能なウェアラブルセンサを開発し、それをを用いた生活習慣、健康状態に関わる状況認識技術を開発した。歩行、走行などの行動と歩数、および脈拍数などの生体情報の見える化だけでなく、ウェアラブルセンサで健康管理上重要である食事イベントを精度良く検出可能にした点が特長である。食事イベントを検出することで、食事のタイミングに基づいたサービス（食事メニュー入力支援、服薬管理、血糖値計測管理など）を実現した。また、取得したユーザの状況に基づいて、生活習慣改善（行動変容）のためのアドバイスをタイムリーに提供するシステムの意義について論じる。



第3章「ウェアラブル睡眠センサとその応用」では、近年不眠に悩む人が増加し、社会問題化しているにも関わらず、睡眠状態を詳細に把握するためには、専門施設に宿泊し、脳波計など多数のセンサを身体中に装着する終夜睡眠ポリグラフ検査を行う必要がある現状から、自宅で手軽に睡眠状態をチェック可能なセンサの必要性を確認し、睡眠計測にフォーカスしたウェアラブル睡眠センサを開発した。内蔵した脈波センサと加速度センサの情報から詳細に睡眠状態を解析可能とする状況認識技術の研究を行った。まず、加速度の情報から睡眠／覚醒を判定する。次に、脈波からロバストに取得した脈波間隔の周波数解析により自律神経の状態を推定し、自律神経と睡眠状態の相関関係からREM睡眠／NREM睡眠と、睡眠深度などの詳細な睡眠状態を判定するアルゴリズムを提案した。手首における脈波計測に適した波長が緑色であることを確認し、ノイズに強い独自の脈波間隔計測手法と合わせて手首における安定した脈波計測を実現した点、個人差や日内変動の影響にロバストなクラスタリング手法を提案した点などが技術的なポイントである。被験者45名による終夜睡眠ポリグラフと本センサの同時計測実験での一致度（ポリグラフの結果を専門医が目視で判定した結果と本センサによる自動判定の結果）は73.5%で、在宅で使える簡易モニタとしては十分な性能を達成し、自宅や外出先での手軽な睡眠状態把握を実現した。さらに、快適に目覚められるタイミング（睡眠状態）を調査した結果、REM睡眠時に起床させると快適に目覚められる傾向にあることがわかり、日々の生活で快適な目覚ましを実現できる可能性を確認した。

第4章「スマートフォンを用いたリアルタイム生活行動認識システム」では、普及が進んでいるスマートフォンに一般的に搭載されている加速度センサと、通話用のマイクを音センサとして積極的に活用し、特別なセンサ／デバイスを使用せずに、当該アプリケーションをインストールしたスマートフォンを持ち歩くだけで様々な日常生活行動を連続的に把握する状況認識技術を提案する。スマートフォン／携帯電話搭載センサを用いた既存研究では、主に屋外の移動状況にフォーカスした行動認識が行われていたが、屋内の様々な生活行動の高精度な認識を実現した点がポイントである。まず加速度をもとにユーザの状態を「歩行」「作業」「安静」の3状態に大まかに推定し、「作業」と推定された場合に、マイクからの音を分析することにより作業内容を細かく分類する。また、生の認識結果だけでなく、同一作業区間単位で行動の種別を再評価する事例ベース認識手法の導入により認識性能が向上し、「歯磨き」「トイレ」「掃除機がけ」「皿洗い」など、これまで加速度センサだけでは認識が難しかった様々な生活行動を精度良く認識することを可能にした。高齢者も含む21名による評価実験を行い、「皿洗い」「アイロンがけ」「掃除機がけ」「歯磨き」「ドライヤー」「トイレ水洗／手洗い」の6種類の生活行動を平均91.4%の精度で認識できることを確認した。また、ロボット

インタフェースとの連携効果、離れて暮らす家族のさりげない見守りサービスなど、提案手法を応用したサービスの可能性についても述べる。

第5章「センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用」では、音声認識入力を日常的に使うために望ましいハンドヘルド型マイクにおける課題として、ヘッドセット型マイクに比べて認識精度が劣ること、音声認識の開始と終了をユーザがボタン操作で指示する必要があることを挙げ、その解決方法として加速度センサと距離センサを内蔵したハンドヘルド型マイクを開発し、それに対するユーザの発話動作を検知する状況認識技術を提案し、両課題に対する対策とその効果について論じる。また、現状非常に手間がかかるTVの番組検索操作に提案手法による音声認識入力を活用するとともに、人名と番組名の言い換えに対応する仕組みを新たに構築することで、従来のリモコン操作による番組検索よりも、操作完了までの所要時間を約40%短縮したヒューマンインタフェースについて述べる。

第6章「結論」では、上述の各研究成果を総括した結論として、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ためには、加速度センサ+ $\alpha$ による状況認識技術が重要であることを述べる。意図のある人物動作／行動には、その多くの場合に身体のダイナミックな動きが伴い、これを加速度センサで捉えることが状況認識の基本となる。その上で、目的・ユースケースに応じて、認識対象の動作／行動の粒度を細かくする、精度を向上する、あるいは自律神経系など人間の内部状態を把握する必要がある場合に、それらの取得に適したセンサを加速度センサと組み合わせて使用するクロスモーダル状況認識が望ましい。さらに、現在の動作／行動の認識から、将来の動作／行動を予測することで、ユーザの意図を先読みし操作不要で所望のサービスを提供することが可能になるであろう。このためには従来の決定論的モデルによる **Reactive Computing** から、確率論的モデルによる **Proactive Computing** をクロスモーダル状況認識により実現する今後のさらなる研究の進展が期待されることを述べる。

## 2. ウェアラブル健康管理システム

---

### 2.1 はじめに

糖尿病、高脂血症、高血圧、高尿酸血症などの生活習慣病は、死因別死亡者数の上位 3 疾病である悪性新生物（ガン）、心疾患、脳血管疾患の主な原因とされており、これら 3 疾病による死亡数の割合は全体の半数以上を占める<sup>[8]</sup>。特に、最近耳にすることが多いメタボリック症候群は、内臓脂肪型肥満に加えてこれらの生活習慣病を合併している状態であり、それぞれ単独の場合に比べて非常にリスクが高くなると言われている。これらの予防のためには、運動、食事、睡眠、ストレスなどの日頃の生活管理が重要である。

しかし、例えば糖尿病の場合、国内患者数 700 万人、予備軍を含めると 2,000 万人とも言われるなど、近年特に増加しているにも関わらず、抜本的な治療法が無く、運動療法、食事療法などを自己管理によって行い、症状を抑えるのが基本的な治療となっている。Tuomilehito<sup>[9]</sup>らは、中年の肥満度の高い被験者を、医師が体重を減らすための運動や食事に関する指導を徹底的に行う群（A 群）と、何も介入しない群（B 群）の二群に分け、3 年以上に渡ってそれぞれの健康状態を追跡調査したところ、A 群の糖尿病発症率は、B 群に対して 58%少ないことを確認した。このように、運動療法、食事療法を継続することが糖尿病のリスクを軽減するが、効果的な治療の持続には非常に強い意志が必要であり、これらの自己管理をサポートするシステムが求められている。

### 2.2 従来研究

日常における生活習慣を管理する取り組みとして、携帯電話やパソコンなどを利用し、インターネット経由で毎日の生活習慣データを記録・管理するサービス<sup>[10]</sup>、記録されたユーザの状態に応じて健康アドバイスを電子メールで配信するサービス<sup>[11]</sup>などが実運用されてきた。この種のネット利用型サービスは、新規に機器を導入する必要なくサービスを受けることが可能であるが、ユーザが別の装置で計測する、あるいは自分の行動を思い出しながら手入力するなど、データ入力に労力を要するため、健康に対する意識の高いユーザでない場合は長続きしにくく、すでにサービスが終了している場合も多い。

また、専用機器を利用して食事管理<sup>[12]</sup>、運動管理<sup>[13]</sup>をサポートする取り組みも行われてきた。これらは、いつでもデータ入力できるため、継続的なデータ収集に基づく履歴管理を手軽に実現できるものである。しかし、データ入力の手入力の場合は、入力の手間や入力のし忘れが問題となる。また、センサを利用しデータ入力の手間を省いたものも、その主眼は履歴管理にある。効果的な生活習慣管理（行動変容）のためには、システムがユーザの状況を把握し、状況に応じた適切なアドバイスをタイムリーに提示することが有効であると考えられる。

一方、様々なセンサを環境側に多数配置し、ユーザの生活状態をセンシングする試みも行われてきた<sup>[14][15][16]</sup>。ユーザには何も装着せずに環境側からセンシングを行うため、ユーザへの負担は少なく、普段の日常生活に近い状態でのデータ収集が行えるなど、実験、研究目的としては有効であるが、一般家庭における日常生活への導入にはコスト、ユーザの識別方法、プライバシーなど、課題が多い。

これらの取り組みに対し本章では、腕時計型のウェアラブル生体センサを用いた生体情報計測に基づく状況認識技術とそれを利用した生活習慣管理、生活改善支援を目指したシステムの開発を行った。ウェアラブルであることは、場所によらずにいつでもどこでもサービスを楽しむことができ、かつ、用途に応じて柔軟な構成でシステム全体を構築することが可能となる。

ウェアラブルセンサを活用した健康管理に関する取り組みとしては、心拍を連続モニタリングするもの<sup>[17][18]</sup>があるが、心臓疾患向けの機能に特化したものである。近年になって、リストバンド型活動量計とスマートフォンにより活動量を把握するサービスが多く実用化されている<sup>[19][20]</sup>が、いずれもデータの蓄積と可視化が主眼で、ユーザの状況に基づいたタイムリーなアクション（情報提供など）は未だに実現されていない。

## **2. 3 コンセプト検討**

### **2. 3. 1 アンケート調査**

ターゲットとして目指すべき健康管理システムのコンセプトを検討するため、実際に生活習慣改善指導に関わっている医療従事者 38 名（医師、看護師、薬剤師、栄養士など）に対し、運動管理、食事管理などにおける問題点に関するアンケート調査を実施した。主なアンケート項目とその結果は下記の通りである。

## 1. 運動管理

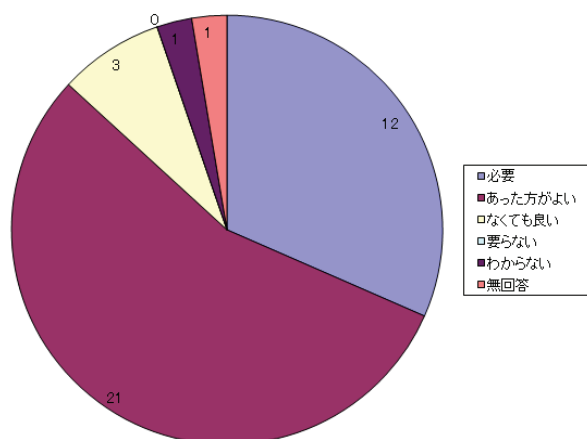


図 2-1 運動管理のために運動量（消費カロリー）の情報は必要か？

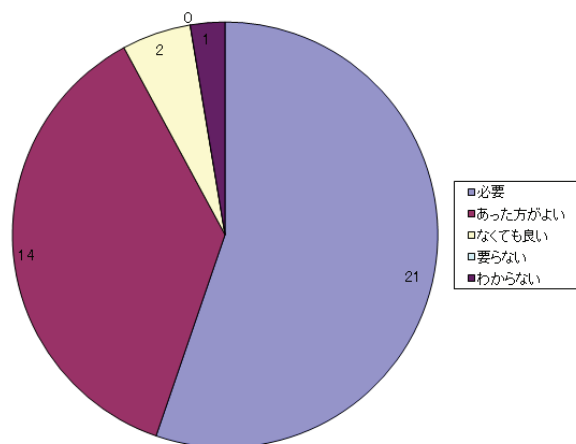


図 2-2 運動管理のために運動内容の情報は必要か？

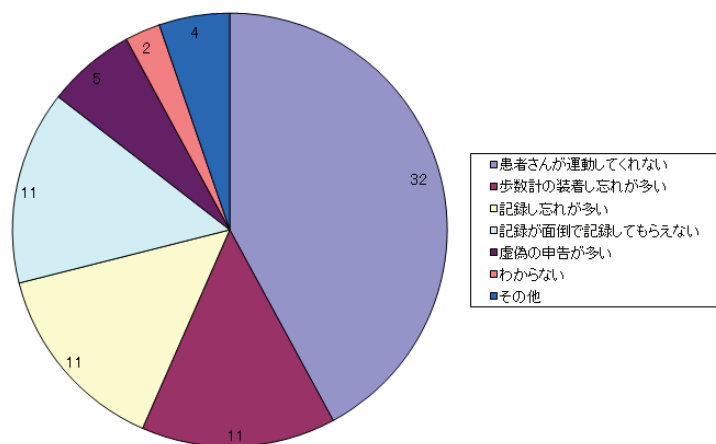


図 2-3 患者さんの運動管理において何が問題か？

## 2. 食事管理

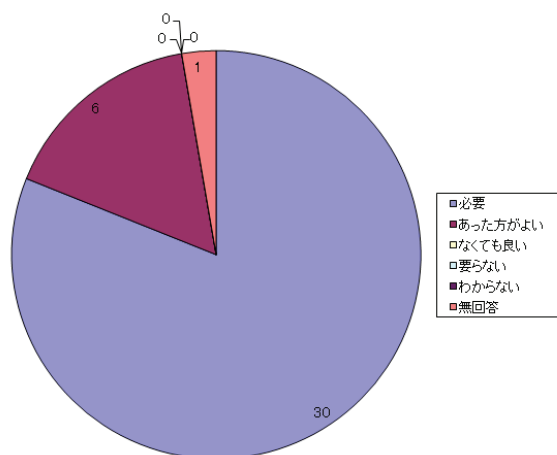


図 2-4 食事管理のために摂取カロリーの情報は必要か？

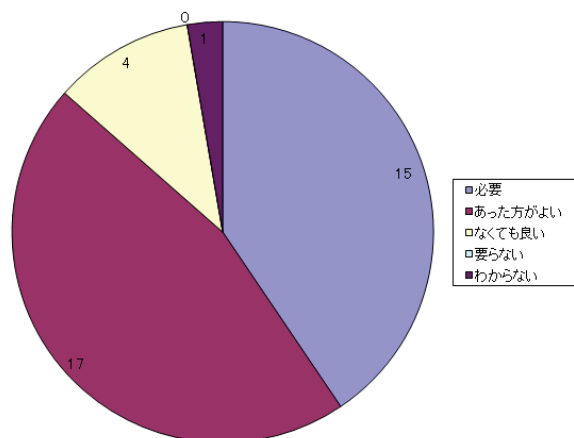


図 2-5 食事管理のために食事時間帯の情報は必要か？

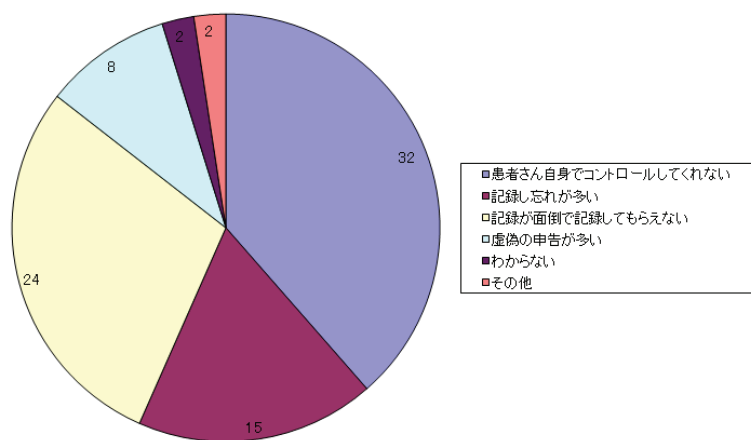


図 2-6 患者さんの食事管理において何が問題か？

### 3. 服薬管理

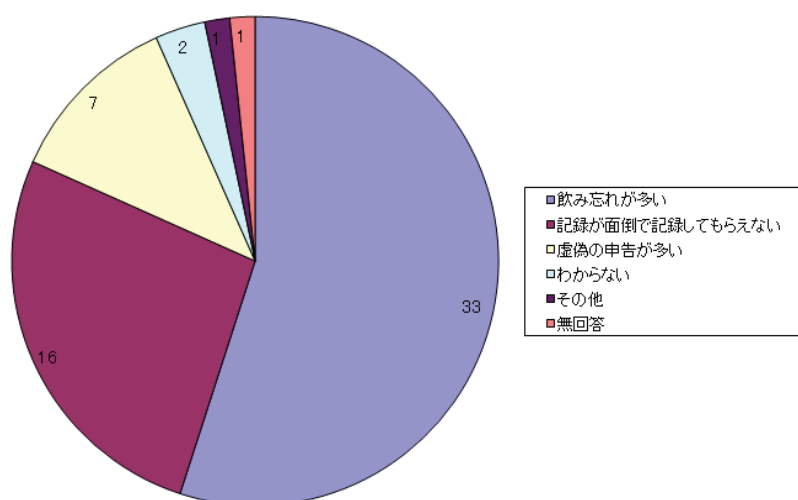


図 2-7 患者さんの服薬管理において何が問題か？

運動管理においては、運動量（消費カロリー）、運動内容の把握は重要であり、食事管理においては、摂取カロリーに加え、食事の時間帯の把握が求められていることがわかった。服薬管理も含め、それぞれの問題点をまとめると下記の通りである。

- **運動管理における問題点**
  - 指導通りに運動しない
  - 歩数計を付け忘れる
  - 運動内容を記録しない（記録忘れ、面倒）
  - 嘘の報告
- **食事管理における問題点**
  - 指導内容を守らない
  - 食事内容を記録しない（記録忘れ、面倒）
  - 嘘の報告
- **服薬管理における問題点**
  - 飲み忘れ
  - 記録忘れ
  - 嘘の報告

いずれにおいても、患者が指導内容を守らない、記録を忘れる、嘘の報告をすることが主な問題点であることが改めて浮き彫りとなった。

## 2. 3. 2 ウェアラブル健康管理システムのコンセプト

### プト

上述のアンケート調査で生活習慣改善指導における問題点が確認できたが、これらは医療従事者側へのアンケートであり、患者側の立場も考慮した上で、目指すべきシステムの要件を検討する必要がある。

医療従事者は、運動管理、食事管理において、これらに関連する情報を把握したいが、関連する情報を記録し続けることは患者側の負担となってしまう点が発見された。また、医療従事者側としても、患者毎にその運動管理、食事管理の状況を把握し、その患者の状態に合わせた適切な指導を行うことは非常に負担が大きい。

すなわち、患者側の負担を軽減しつつ医療従事者側のニーズを満たす健康管理システムの必要性が示唆される。これを踏まえ、本章で開発する健康管理システムのコンセプトを「手軽で確実な生活習慣管理の実現」とし、それに必要な要件を以下の通りとした。

- **患者の負担を軽減する**

日常生活における継続的な生活習慣管理のためには、患者の負担を極力軽減する必要がある。必要なセンサは小型軽量であることが求められると共に、患者によるデータ入力などの操作は極力減らすことが望ましい。

- **必要な情報を確実に収集する**

運動管理、食事管理、服薬管理において、必要とされる情報は確実に収集する。センシング可能な情報であればセンサにより自動で入力し、センシングが困難なものについては、確実に入力する工夫が必要である。

- **データ入力の手間を軽減する**

必要な情報の全てを患者が手入力することは負担が大きく長続きしない。センシング可能な情報はセンサで自動入力し、手入力が必要な項目はその手間を極力軽減するサポート機能が必要である。

- **状況に応じてタイムリーな情報提供を行う**

状況認識技術により患者の状況を常に把握し、生活習慣改善（行動変容）、データ入力などに適したタイミングでタイムリーな情報提供（生活習慣改善



アドバイス、データ入力指示など)を行い、効果的かつ確実な健康管理を実現する。

## 2. 4 システム構成

### 2. 4. 1 全体概要

図 2-8 に示すように、腕時計型ウェアラブル生体センサと PDA (Personal Digital Assistant) で構成するウェアラブル健康管理システムを開発した。



図 2-8 ウェアラブル健康管理システム

腕時計型ウェアラブル生体センサは、3軸加速度センサ、光電脈波センサ、GSR (Galvanic Skin Response : 皮膚電気反応) センサ、温度センサ等を搭載し、ユーザの生体情報を連続計測する。また、内蔵のマイクロコントローラで脈拍数算出、動作状態推定などの状況認識をリアルタイムに行い、認識結果を Bluetooth モジュール経由で PDA へ定期的送信することとし、サンプリングの都度生データをやりとりする方式と比べて、通信データ量削減、消費電力削減とデータ処理の分散化を行った。PDA では、受信した現在のデータと過去の履歴をグラフィ

カルに表示すると共に、それらを利用して食事イベント検出など、より高度な状況認識をリアルタイムに行い、ユーザの状況に応じたタイムリーな生活習慣改善インストラクションを提示する。また、計測データを XML 形式でデータベース化し、定期的にサーバにアップロードすることで主治医とのデータ共有も可能にした。

## 2. 4. 2 腕時計型ウェアラブル生体センサ

腕時計型ウェアラブル生体センサは、小指に装着するセンサヘッドと本体で構成した。全体外観を図 2-9 に、装着時の外観を図 2-10 に、センサヘッド外観を図 2-11 に、機能ブロック図を図 2-12 に、搭載したセンサー一覧を表 2-1 に示す。



図 2-9 腕時計型ウェアラブル生体センサ外観



図 2-10 腕時計型ウェアラブル生体センサ 装着時外観

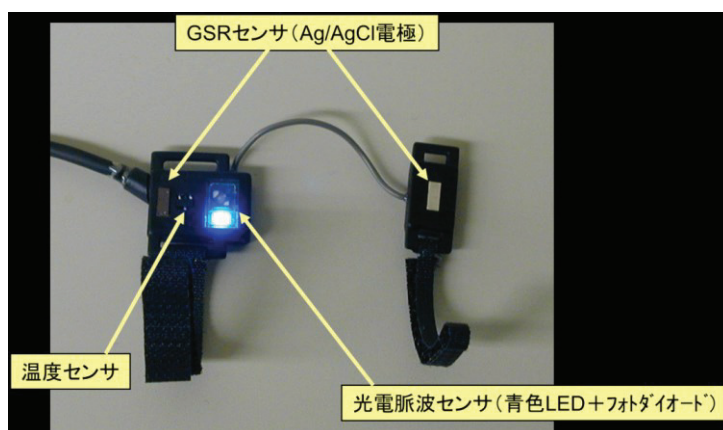


図 2-11 腕時計型ウェアラブル生体センサ センサヘッド

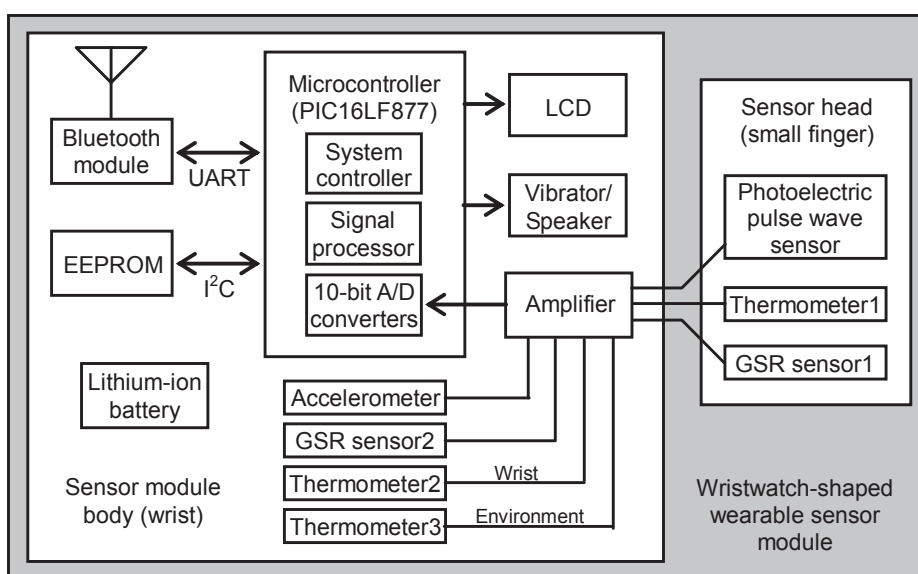


図 2-12 腕時計型ウェアラブル生体センサ 機能ブロック図

表 2-1 搭載センサー一覧

センサ	生体情報	示す指標
3 軸加速度センサ	手首加速度	動作状態、歩数
光電脈波センサ	脈波	脈拍数、自律神経活動
GSR センサ (指腹)	発汗	精神 / 心理状態
GSR センサ (手首)	発汗	温熱状態
温度センサ (指腹)	皮膚温	精神 / 心理状態
温度センサ (手首)	皮膚温	温熱状態
温度センサ (環境)	(環境温)	(環境温)

手首の本体に加え、小指に装着するセンサヘッドを使用したのは、手の平は、毛細血管が皮膚表面に多く存在するため、光電脈波センサによる脈波計測に適していることと、精神性発汗など精神状態による身体の反応が観測しやすい部位であることが理由である。

センサヘッドは、センサヘッド本体と GSR 用 Ag/AgCl 電極で構成され、それぞれ大きさは 5×17×12mm、4×8×12mm、重さは合わせて 5g と小型軽量の構成とした。センサヘッド本体には、青色 LED（ピーク波長 470nm）とフォトダイオードで構成した光電脈波センサ、温度センサ、GSR 用 Ag/AgCl 電極を内蔵した。光電脈波センサは、ヘモグロビンが吸収する波長帯の光（470nm）を LED から照射し、血流量の変化によって強弱する反射光をフォトダイオードで測定し脈波として計測する。光電脈波センサの計測部位とそれに適した LED の波長については、次章で詳しく考察する。GSR は皮膚表面の発汗状態による 2 つの Ag/AgCl 電極間の抵抗変化を計測する。上述の通り、手の平は精神性発汗など精神状態による身体の反応が顕著に表れるため、センサヘッド搭載の GSR センサ、温度センサで手の平の発汗、皮膚温を計測し、手首におけるそれらとの差分を見て精神状態等の評価も可能な構成とした。

大きさ 26×46×85mm、重さ 84g と腕時計大の腕時計型生体センサ本体には、3 軸加速度センサ、温度センサ（手首、環境）、GSR センサ（手首）を搭載した。マイクロコントローラは、50ms 間隔（20Hz）で各センサの信号を A/D 変換し、脈拍数算出、動作状態推定などの状況認識をリアルタイムに行い、認識結果および必要に応じて所望のセンサの生データを Bluetooth モジュールから PDA へ 1 秒間隔で送信する。尚、日常生活で日中を通して使用できるように 14 時間連続動作可能な設計とした。

## **2. 4. 3 生体センサでの状況認識**

内蔵のマイクロコントローラで脈拍数算出、動作状態推定などの状況認識をリアルタイムに行う。従来、ユーザの行動認識に関する研究は、主に腰<sup>[21]</sup>や膝<sup>[22]</sup>などの体幹に装着したセンサを用いたものが多かった。これらの部位は、歩行、走行などの全身運動や、姿勢の変化を捉えることに適しているが、それら以外の行動（食事、デスクワークなど）の把握には適していない。また、これらの部位に日常生活でセンサを装着して使用することは一般的でなく、歴史的に機械を身に付けることが人類に受け入れられてきた数少ない部位である手首への装着をターゲットとした。<sup>i</sup>

<sup>i</sup>腕時計は 19 世紀後半に誕生し、その後、実用性とともによりファッション性としての軸も加わって広く普及し、身に付けて使用する機器・器具としては、眼鏡とともに最も受け入れられているものと言える。

まず、日常生活における運動状態把握に有用であると考える歩行、走行、デスクワークの認識を検討した。これらの動作時の手首装着 3 軸加速度センサの出力例を図 2-13 に示す。これらの行動の波形特徴を鑑み、また消費電力等を考慮して閾値交差をベースにした計算量の少ない動作認識アルゴリズムを開発した。フローチャートを図 2-14 に示す。歩行、走行、デスクワーク、安静の 4 行動に対する性能評価を行った結果、これらの行動に限定すれば本アルゴリズムにより約 90%の精度で認識可能であることを確認した。

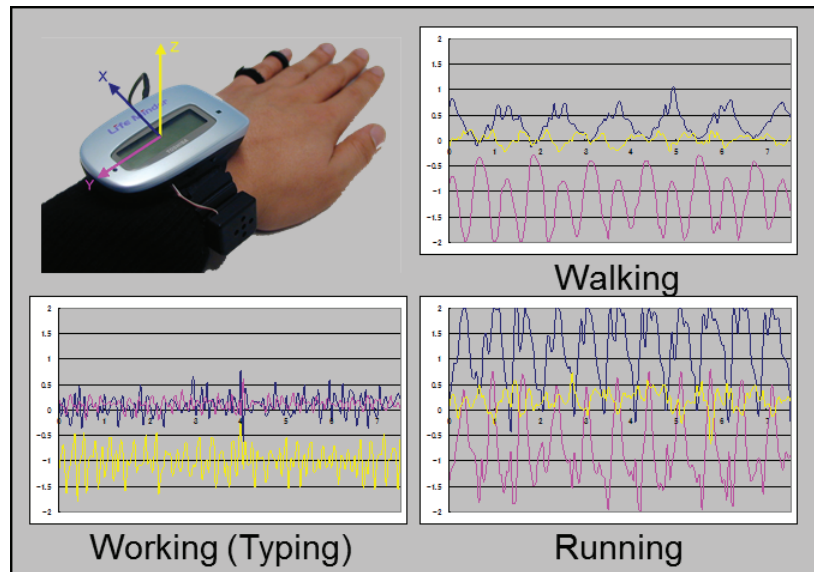


図 2-13 各動作時の 3 軸加速度センサ出力例

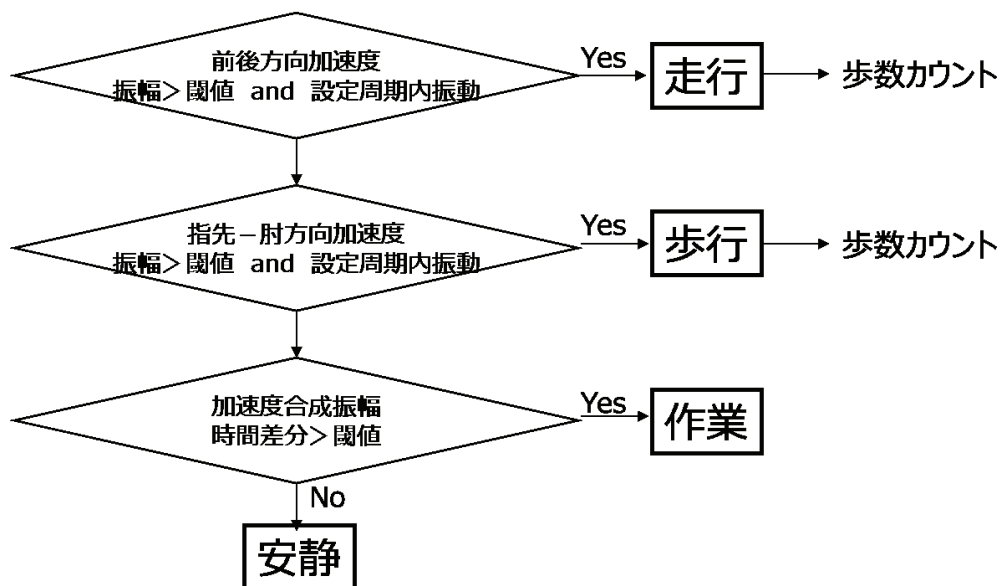


図 2-14 計算量を抑えた動作認識アルゴリズム

本アルゴリズムでは、歩行、走行時には歩数もカウントする。歩数カウントの精度は、市販の腰装着型の歩数計に対して誤差 19%程度となったため、さらなる精度改善が必要ではあるが、日々の運動管理においてはその絶対値だけでなく相対的な変化も重要な要素であり、本システムにおいては計算量が少ない本アルゴリズムによる日々の歩数の変化の把握を主眼とした。

また、光電脈波センサから得られる脈波から、脈波間隔を算出し瞬時脈拍数を得る。センサ付近の身体の動きによりノイズが発生した場合は、動きにより血流そのものが変化する場合があるため、大きな動きが発生した際は脈拍数算出を行わないなどのノイズ対策を行い、精度向上を図った。

## 2. 4. 4 PDA での状況認識

PDA は、ウェアラブル生体センサにおける状況認識の結果、および各センサの生データを 1 秒間隔で取得し、リアルタイムにグラフ表示する。表示の一例を図 2-15 に示す。

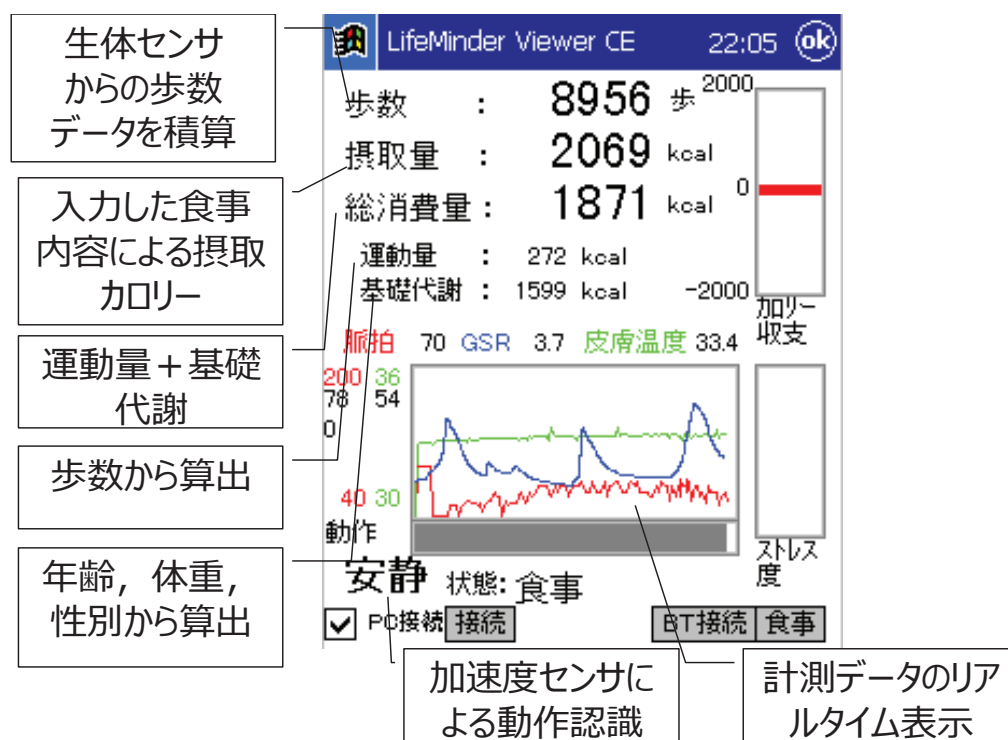


図 2-15 PDA 表示画面例

画面上に、ウェアラブル生体センサから取得した歩数、各センサデータ、動作状態をリアルタイムに表示すると共に、摂取カロリーと消費カロリーのバランス

状態などもわかりやすく表示する。ここで、歩数から換算した消費カロリーと、基礎代謝によるもの<sup>[23]</sup>を推定し、その和を総消費カロリーとして表示する。

2. 3. 1 節のアンケート調査では、食事時間帯の情報も食事管理の上では重要と結果が得られた。また、食事内容の記録も摂取カロリーや栄養バランスの把握に重要であるが、時間が経ってしまうと、入力しようとしても何を食べたか忘れてしまっていて入力できない。食事していることをセンサで検知できれば、食事時間の記録だけでなく、食後すぐに食事内容の入力を促して確実に入力させる機能の実現が期待できる。

## 食事イベント検出の検討

ウェアラブル生体センサで取得したデータを用いて、食事イベントを検出可能か評価するため、「食事」「クレペリン検査<sup>ii</sup>」「ビデオ鑑賞」「読書」「会話」「TVゲーム」の各タスク時のデータ収集を行った。被験者は20～40代の10名（男性9名、女性1名）で、ウェアラブル生体センサによりこれらの各タスク時の脈拍数、GSR、皮膚温度を収集した。収集したデータの一例を図2-16に示す。GSRからは発汗によるピーク出現のイベント発生の頻度をカウントした。結果を表2-2と図2-17に示す。

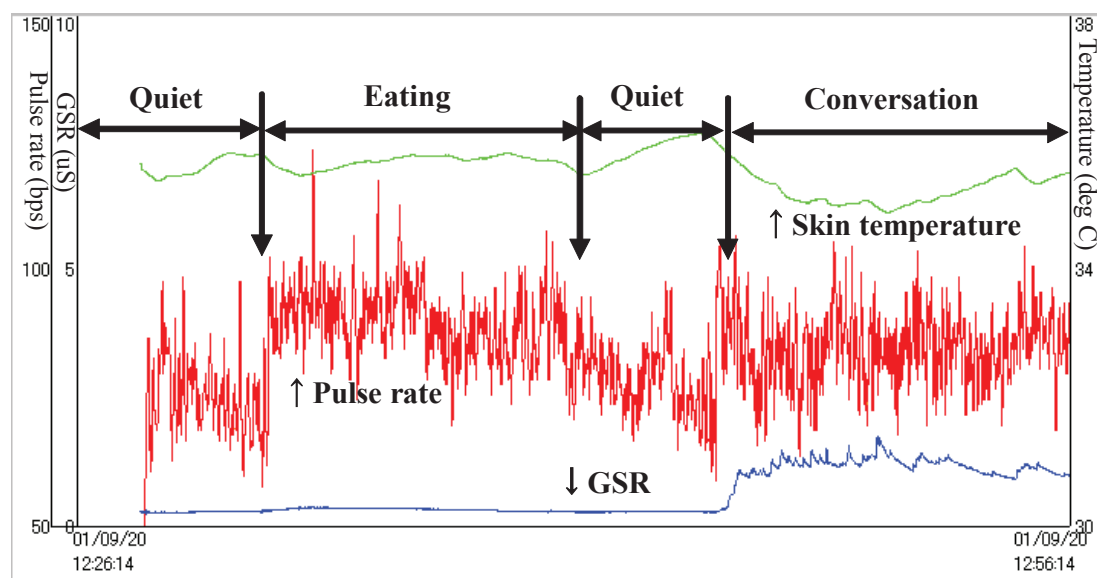


図 2-16 データ収集結果の一例

<sup>ii</sup> 正式には「内田クレペリン精神検査」。1列に並んだ1桁の数字(3～9)の足し算を繰り返し、1分経ったら次の列に移動、それを繰り返すというもの。前半15分、5分の休憩を挟んで後半15分の合計30分行う。1分ごとの計算量、経時的な計算量の変化(作業曲線)、誤答数から個々の性格・職業適性をみるのが主な目的。

表 2-2 タスク毎の脈拍数、GSR の変化

状態	脈拍 (bpm) 30秒平均	GSR出現頻度
	安静からの差分	30秒間の回数
安静	-	0
食事	+6.2	0.22
TVゲーム	+6.05	1.46
読書	+1.53	1.41
クレペリン	+4.57	0.31
デスクワーク	+3.5	0.35
会話	+4.6	1.45
TV鑑賞	+4.9	0.78

		GSR 出現頻度 (30秒間)		
		小<1	小>=1 & 大<1	大>=1
脈拍 (30秒平均)	対安静変化 <=3	安静	外部刺激による反応	
	>3		デスクワーク 家事	読書 TV 会話
	>7	食事	内発的な活動	笑い TVゲーム

注: 図中の矢印は、食事による脈拍上昇（内発的な活動）と、外部刺激によるGSR活性化を示している。

図 2-17 タスクと、脈拍数、GSR 出現頻度の変化の関係

食事の場合、安静時に対して脈拍数は大きく上昇する一方、GSRにはほとんど反応が出ない。それ以外のタスクでは脈拍数が大きく上昇する場合、GSRも一緒に反応する傾向が見られるため、脈拍数とGSRに着目することで食事イベントを検出できる可能性が示唆される結果が得られた。食事の時だけ傾向が異なることの機序は下記の通りと考察する。

- 脈拍数上昇

食事による咀嚼運動、消化器官の活性化により脈拍数は上昇する。



● GSR 出現頻度低下

GSR は主に緊張状態など交感神経が優位な状態でその出現頻度が増加する。食事の場合、まず食事前に視覚、嗅覚への刺激があり、食事中には食道、胃壁への物理的な刺激が発生する。これにより、これらを支配する迷走神経が活性化するが、臓器に行く迷走神経は自律神経、特に副交感神経としての働きが強いため、拮抗する交感神経の活動が抑制される。

上記の機序により、食事の時は脈拍数が上昇するにも関わらず、他のタスクと違って GSR 出現頻度は増加しないと考えられる。

ここで得られた知見をもとに、脈拍数と GSR 出現頻度に着目した食事イベント検出の可能性を検証したが、食事を開始したことは 2 分以内に 90%以上の精度で検出できたものの、食事の終了を、即時性を持って検知することが難しいことがわかった。これは、食事が終了しても消化器官の活動などにより、すぐに脈拍数が下がらないためである。食事イベント検出の即時性を向上させるために、図 2-18 のように食事に伴って発生する特徴的な動作を加速度センサで捉えることが効果的と考え、加速度センサの情報も加味することを検討した。

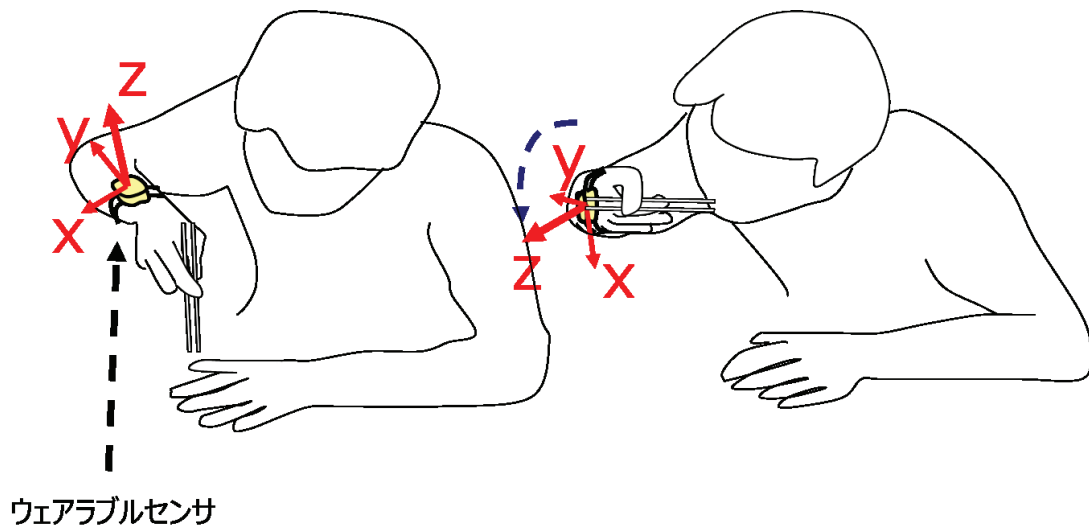


図 2-18 「食べ物を口に運ぶ」特徴的な動作を加速度センサで検出

被験者 3 名で食事タスク 8 セットと、食事以外のタスクとして「クレペリン検査」「音楽鑑賞」「新聞を読む」「PC 作業」「安静」「お茶を飲む」「電話で話す」「歯を磨く」「読書をする」「たばこを吸う」「食事のふり (実際には食べない)」のデータ収集を行い、食事、食後、食事以外の 3 クラス問題として決定木のモデル

構築を行い、発汗頻度、Z 軸加速度、Y 軸加速度、脈拍数などが食事認識に有効であることがわかった<sup>[24]</sup>。

また、加速度情報を使用した場合の方が、使用しない場合に比べて、食事検出精度が高くなることも確認した。結果を表 2-3 に示す。

表 2-3 食事、食後、食事以外の 3 クラスの認識に対するエラー率

エラー率(%)		食事	食事後	それ以外	全体
加速度なし	訓練データ	11.5	27.4	5.0	<b>6.3</b>
	テストデータ	26.2	57.1	8.6	<b>11.9</b>
	全データ	16.4	35.6	6.0	<b>7.9</b>
加速度あり	訓練データ	11.8	8.2	2.8	<b>4.0</b>
	テストデータ	20.9	26.0	5.0	<b>7.7</b>
	全データ	14.6	14.0	3.5	<b>5.2</b>

加速度情報も利用した食事検出アルゴリズムを開発し、そのアルゴリズムによる食事イベント検出精度は 92%であった。食事イベント検出結果の一例を図 2-19 に示す。

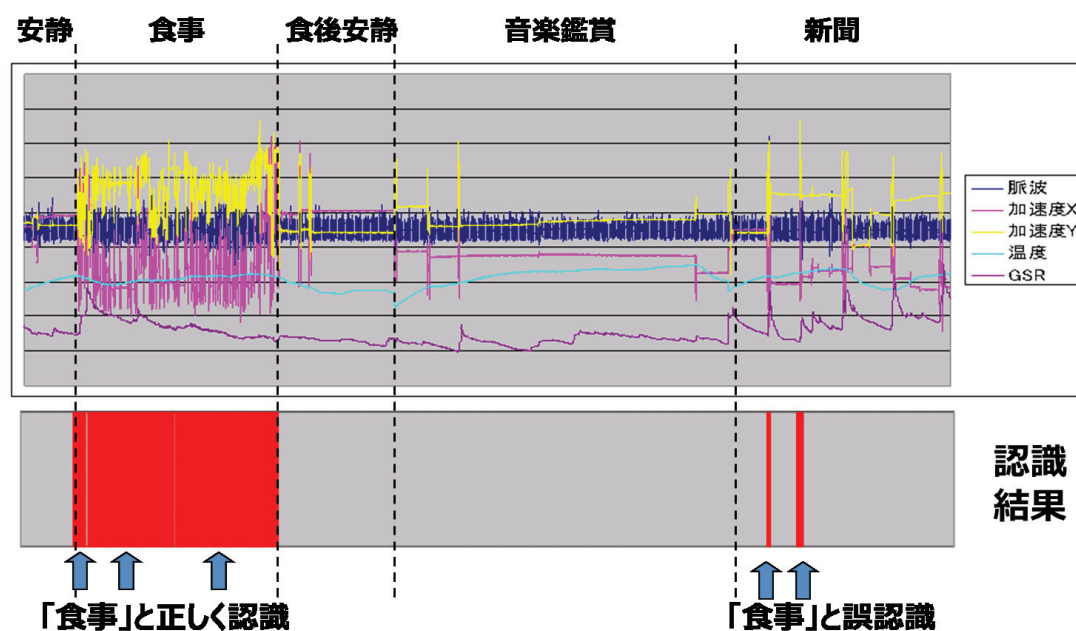


図 2-19 食事イベント検出結果の一例

## 2. 4. 5 タイムリーな情報提供

状況認識技術により、動作、歩数などの運動状況、食事イベントが認識できることを示した。これらをシステムがリアルタイムに把握することで、タイムリーな情報提供を可能にした。主な機能を以下に説明する。

## 情報入力のためのタイムリーな指示

食事イベント検出により、食事の開始、あるいは終了時にユーザに食事内容の入力を促すメッセージを表示する。前述したが、食事内容は時間が経つと忘れてしまっていて入力できないことが多くなるため、確実に入力させるには、タイムリーに入力を促す機能が有効である。

食事内容を自動的に取得することが理想的であり、食事画像を栄養士に送ってアドバイスが得られるサービス<sup>[25]</sup>や、食事画像から画像認識で食事内容を取得するサービス<sup>[26]</sup>なども行われているが、手間や精度が依然として課題である。本システムでは、食事メニューの大規模データベースを事前に用意しておき、図 2-20 のように、食事メニューを選ぶだけで目安となるカロリーを入力できるようにして手間を軽減し、確実な食事内容の記録を可能にした。



図 2-20 食事内容入力例

## 服薬・測定のためのタイムリーな指示

薬の飲み忘れも健康管理上の課題である。例えば食後に飲む薬がある場合は、事前にシステムに登録しておけば、食事イベント検出により食後を検出し、適切なタイミングで服薬確認のメッセージを表示させることで、飲み忘れを防ぎ、服薬記録を確実に残すことが可能になる。

また、糖尿病患者にとって、食後を中心として一定間隔で血糖値を測定することが重要とされているが、食後の時刻を自分で把握し、その後所定の時刻に測定することは負担が大きい。そこで、食事イベント検出により食後時刻をシステムで自動的に把握しておき、血糖値測定が必要なタイミングでタイムリーにメッセージを表示することが可能である。

それぞれ、メッセージ画面の一例を図 2-21 に示す。

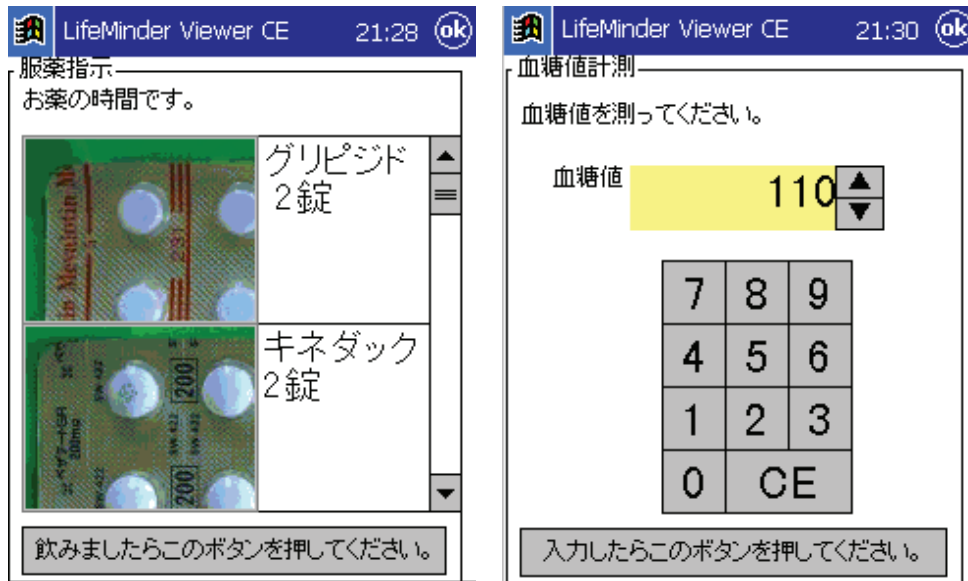


図 2-21 服薬確認、血糖値測定確認画面例

## 運動、休憩のタイムリーな提案

状況認識技術により、歩行、走行などの運動や、デスクワークなどがリアルタイムにシステムが把握できるため、例えば、スケジュール帳と連動し、運動不足のユーザが、スケジュールが特に無い休日などに運動しないでいる場合に散歩を促したり、デスクワークが長時間続いている場合に休憩を促したり、タイムリーなタイミングで健康に繋がる（行動変容）メッセージを提示することができる。

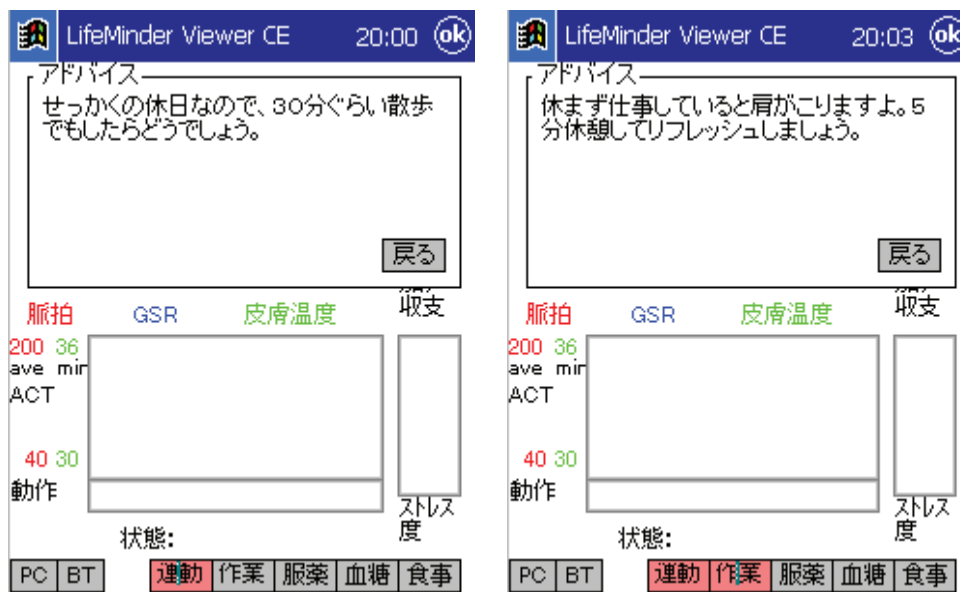


図 2-22 運動、休憩提案画面例

## 2. 5 まとめ

ユーザの手首に装着し、加速度、脈波、発汗などさまざまな生体情報をセンシング可能な腕時計型ウェアラブル生体センサを開発し、PDAと連携して日常生活における手軽な健康管理をサポートするウェアラブル健康管理システムを開発した。日常生活において、歩行、走行、作業などの主要行動と歩数、さらに健康管理上重要な食事イベントを検出する状況認識技術により、ユーザの状況に応じてタイムリーな情報提供を可能にし、生活習慣改善（行動変容）に効果的な機能を実現した。特に、食事イベントを検出可能にしたことで、負担が少なく確実な食事管理、服薬管理などを実現した点が特長である。

本章で開発した腕時計型ウェアラブル生体センサは、図 2-11 のセンサヘッドを小指に装着して、手の平で脈波、発汗、皮膚温を計測する仕様とした。これは、手の平には皮膚表面に近いところに血管が多く存在するため脈波計測に適していることと、精神性発汗が手の平に顕著に発生することから精神状態の推定に適していることなどが理由である。しかしながら、日常生活において、指にセンサを装着し続けることは、ユーザにとって負担になることが予想される。

次章では、ターゲットを睡眠計測に絞り、必要最小限のセンサを搭載し、かつ手首でロバストな脈波計測を実現して指へのセンサ装着を無くし、自宅等で日常的に睡眠状態を手軽にチェック可能なウェアラブル睡眠センサの開発を行った。

## 3. ウェアラブル睡眠センサとその応用

---

### 3. 1 はじめに

ストレスが多く、かつ 24 時間化した現代社会において、不眠症の疑いがある人が 38.1%、不眠症の疑いが少しある人が 18.4%と、半数以上の人々が睡眠に不安を持つとされている<sup>[27]</sup>。また生活習慣病患者の半数以上（54.3%）が不眠症、あるいは不眠症の疑いあり<sup>[28]</sup>とされ、生活習慣病と睡眠の関係も注目されている。このように、睡眠障害は現代病の一つとなってきたが、睡眠状態をチェックするためには設備の整った専門機関に宿泊し、脳波計など多数のセンサを身体中に装着して就寝する終夜睡眠ポリグラフ検査<sup>[29]</sup>が一般的で、患者はもちろん、検査する側にとっても非常に負担の大きいものとなっている。よって、自宅などで手軽に睡眠状態をチェックできる装置が求められている。

### 3. 2 従来研究

上述のような背景の下、在宅などでの使用を目指した睡眠モニタの研究開発が以前より行われている。

Sadeh<sup>[30]</sup>らは、睡眠／覚醒の判定を、加速度センサを用いて行ったが、加速度センサによる体動情報だけでは REM 睡眠、NREM 睡眠、睡眠深度など、詳細な睡眠状態を判定することはできていない。これに対し、心拍変動解析からの自律神経解析を利用して詳細な睡眠状態を把握しようとした取り組みもいくつか行われている<sup>[31][32][33]</sup>。これらの心拍変動解析による睡眠解析には二つの課題がある。一つは、身体の動き（寝返り等）によって発生するノイズにより、正確な心拍変動の解析が困難になる点である。もう一つは、自律神経活動は年齢、性別、身体状況などによって影響を受けるため個人差が大きく、睡眠状態も日中の活動状態や概日リズム（サーカディアンリズム）など、いろいろな要因の影響を受ける点である。

本章では、ノイズにロバストな脈波計測手法と、個人差や日内変動による影響を考慮した睡眠状態判定アルゴリズムを提案し、自宅や旅先で手軽に使用可能なウェアラブル睡眠センサを提案する。

### 3. 3 ウェアラブル睡眠センサの開発

#### 3. 3. 1 心拍変動を用いた睡眠解析

上述の通り、REM 睡眠、NREM 睡眠、睡眠深度など詳細な睡眠状態は、終夜睡眠ポリグラフによる判定を用いるが、本章では、ウェアラブルセンサによるアプローチで同様の詳細な睡眠状態判定の実現を目指す。

睡眠は脳の休息であるため、直接的には脳の活動を観察する必要がある。しかし、大脳の休息を制御する脳幹には自律神経機能中枢が存在するため、脳幹の機能が正常であれば、自律神経活動と睡眠の関連は深いと考えられる。一方、自律神経活動は心拍変動解析により推定できるため、脈波から心拍変動（脈拍変動）を介して自律神経活動を捉え、睡眠の状態を間接的に推定することが可能となる。本章で提案するウェアラブルセンサによる睡眠解析手法の処理の流れを図 3-1 に示す。

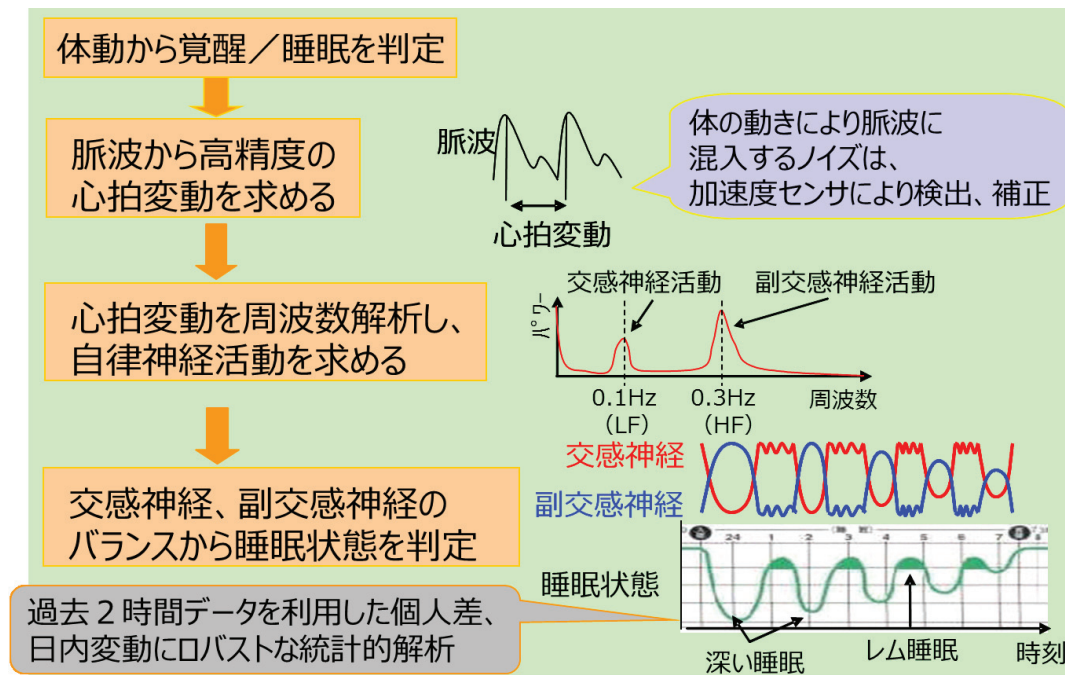


図 3-1 ウェアラブルセンサによる睡眠解析手法

まず、加速度センサからわかる体動の状態により、睡眠しているか、覚醒しているかを判定する。ここでは、Cole<sup>[34]</sup>らが提案した手首装着加速度センサで検出した一定値以上の体動発生頻度に基づいた睡眠／覚醒判定と同様のアルゴリズムを用いる。

次に、脈波から一拍一拍の脈波間隔を高精度に算出し、心拍変動を求める。心拍変動を周波数解析することにより自律神経の活動を求め、交感神経、副交感神経のバランスから最終的に睡眠状態を判定する。

### 3. 3. 2 手首におけるロバストな脈波計測手法

前章で提案したウェアラブル健康管理システムでは、小指に装着したセンサヘッドに光電脈波センサを搭載して脈波計測を行った。これを実験等で使用した被験者などから、日常生活で常時センサヘッドを装着したまま生活するのは邪魔になるとの意見が寄せられた。睡眠中であっても、それは同様であると考えられ、負担の少ない睡眠計測を実現するためには、センサヘッドでなく手首の本体で脈波計測を実現したい。

しかしながら、前章でも述べた通り、手の平は毛細血管が皮膚表面近くに多く存在するため S/N 良く脈波計測できる部位であるのに対し、手首は手の平に比べて皮膚表面近くの毛細血管が少なく、S/N が悪くなることが想定される。そこで、光電脈波センサに使用する波長から見直しを行った。

図 3-2 は、光の波長と皮膚に入り込む深さの関係を示した図である。

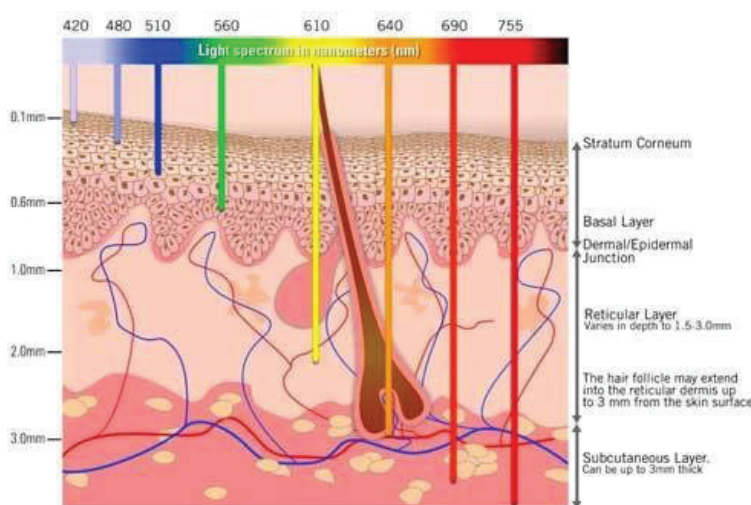


図 3-2 光の波長と皮膚へ入り込む深さの関係

※出典：SPECTRUM SCIENCE & BEAUTY

<http://www.spectrumsiencebeauty.com.au/ipl-hair-removal/>

波長が長くなるに従って皮膚の奥まで入り込むことがわかる。皮膚表面近くに毛細血管が少ない部位では青よりも長い波長帯の光で、より深部の血管の血流変化にも感度を持たせることが望ましい。しかし、あまり奥まで入り込ませると、反射光としての強度が弱まることも想定される。指先や耳たぶなど、透過させて



計測できる部位であればより長い波長帯で透過させて計測することもできるが、手首では透過は困難である。よって、適度な深さまで入り込みつつ、ヘモグロビンの吸光特性も比較的良好である緑色光（545nm 程度）が手首での反射型光電脈波計測に適していると考えられる。生体組織による光吸収・散乱の波長特性を図 3-3 に、光電脈波センサに使用する光の波長とその特性、計測に適した部位を考察した結果を表 3-1 に示す。

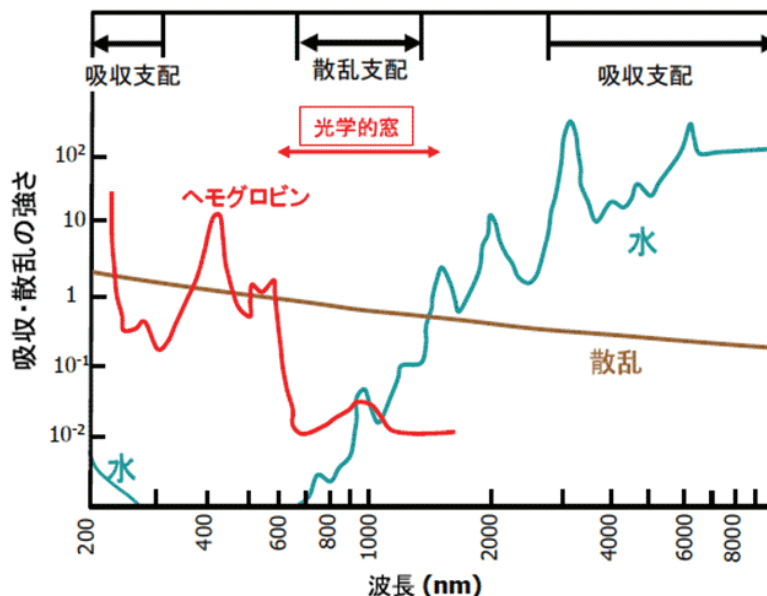


図 3-3 生体組織による光吸収・散乱の波長特性

出典：文部科学省 Web サイト

[http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu3/toushin/attach/1333543.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu3/toushin/attach/1333543.htm)

表 3-1 光電脈波センサに使用する光の波長とその特性

波長帯	入り込む深さ	反射型？透過型？ 適した測定部位は？
Blue (470nm)	表皮	反射型 手のひら、指
Green (545nm)	真皮	反射型 手のひら、手首
Red~IR (660nm~)	皮下組織	透過型 指先、耳たぶ

上記考察により、本章で開発したウェアラブル睡眠センサでは、センサ本体裏に緑色 LED とフォトダイオードで手首向けの光電脈波センサを構成した。外観を図 3-4 に示す。



図 3-4 ウェアラブル睡眠センサの光電脈波センサ部

このようにして構成した手首向け光電脈波センサであるが、手の平で計測した脈波に比べると S/N はどうしても劣る。そのため、基線揺らぎ、振幅揺らぎなどが脈波間隔検出精度に影響することが想定される。また、部位によらず光電脈波センサとしての課題であるが、体動が発生するとセンサと皮膚の接触状態が変化し、大きなノイズが混入するため、体動由来のノイズ対策も必要となる。本研究では、これらのノイズにロバストな脈波計測手法を考案した。図 3-5 にその処理フローを示す。

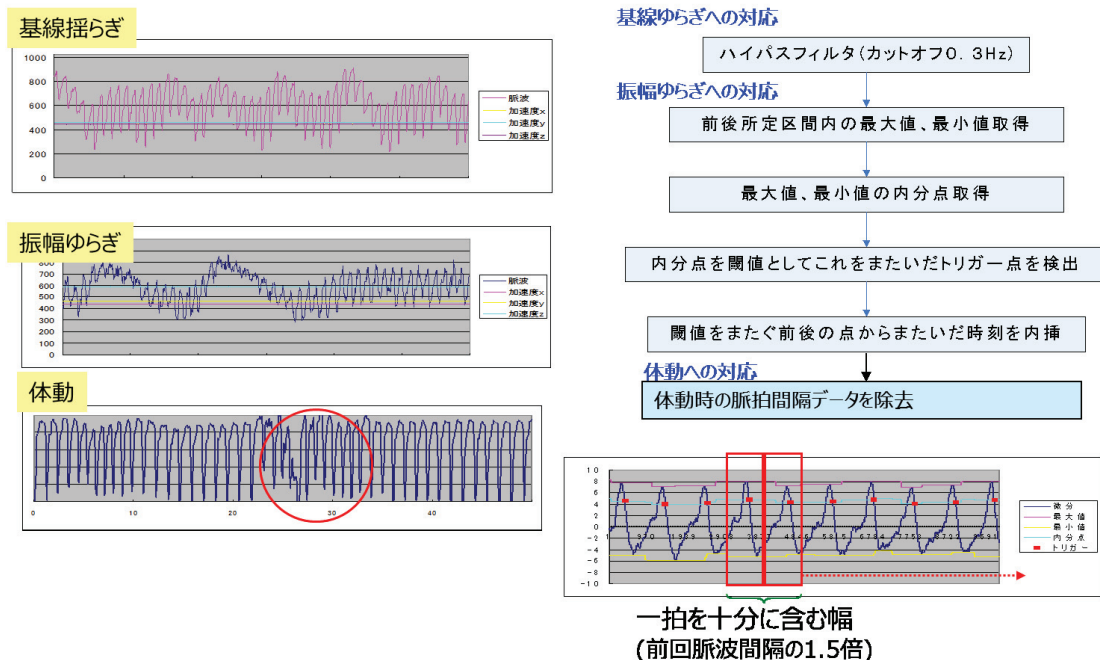


図 3-4 ノイズに強い脈波間隔検出手法

まず、ハイパスフィルタ（カットオフ周波数 0.3Hz）により基線揺らぎを除去する。次に振幅揺らぎへの対処として、脈波間隔検出のための閾値を最新の脈波振幅に応じて可変にすることとした。具体的には、脈拍一拍を十分に含む幅（前回脈波間隔の 1.5 倍）における最大値・最小値から動的に閾値（内分点）を設定し、これを跨いだトリガ点を検出することで、脈波振幅の変動に強い脈波間隔検出を実現する。

図 3-1 に示した通り、脈波間隔から心拍変動を求め、その周波数解析から自律神経活動を推定するため、脈波間隔の正確性がその後の処理の精度に大きく影響する。一方、日常での使用を想定すると、ウェアラブル睡眠センサは小型で長時間連続駆動する必要があるため、メモリサイズ、消費電力の制約がある。よって、脈波のサンプリング周期は 64Hz とすることとしたが、精度の良い自律神経解析のためにはより分解能の高い脈波間隔検出が必要となる。そこで、サンプリング周期を変更せずに、脈波間隔検出の分解能を向上させる手法を検討した。図 3-5 は提案手法を説明するための模式図である。この例では、時刻  $x_{11}$  と  $x_{12}$  の間で閾値との交差が発生しているが、ここで、 $(x_{11}, y_{11})$  から  $(x_{12}, y_{12})$  までの波形が直線であると仮定すると、三角形の相似関係から下記の式(3-1)により、真の閾値交差時刻  $x_1$  を推定することが可能となる。

$$x_1 - x_{11} = (x_{12} - x_{11}) \times \frac{y_{11} - y_{12}}{y_{11} - \text{Threshold}} \quad (3-1)$$

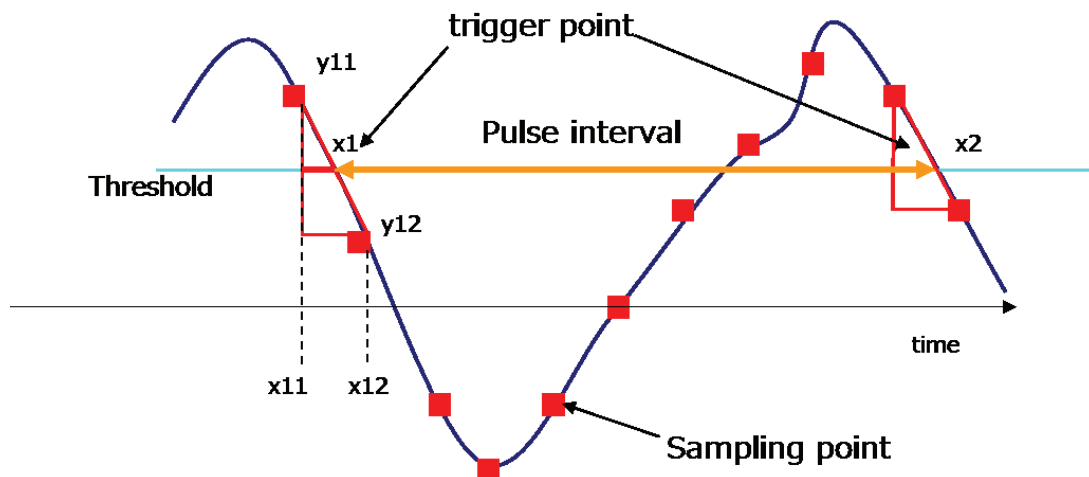


図 3-5 脈波間隔検出の分解能向上手法

各閾値交差のタイミングにおいて同様の処理を行うことで、脈波間隔検出の分解能を向上させることができる。

一方、体動が発生した場合は図 3-4 のように脈波の波形が大きく乱れるが、加速度センサで体動を検知することで、このような波形が乱れるタイミングを検知し、体動時の脈波データを除去・補間して体動ノイズの影響を軽減する。

本手法による手首脈波からの脈波間隔検出の精度を評価するため、健常者 3 名に対して通常の睡眠時に、手首装着脈波センサ（本章にて試作）、指先装着脈波センサ（市販のパルスオキシメータ）、心電図（ECG）を同時計測する実験を行った。図 3-6 は各センサの計測生波形である。手首装着脈波センサの S/N が最も悪いことがわかる。

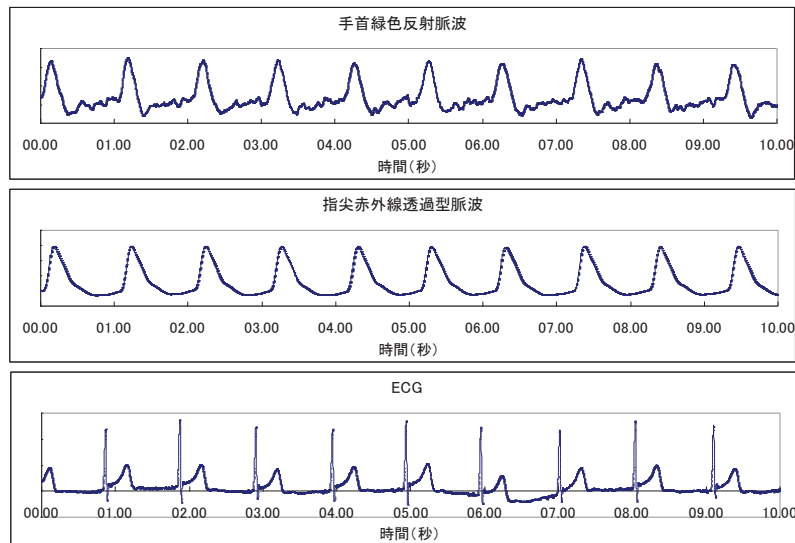


図 3-6 手首脈波、指先脈波、心電図の生波形

しかしながら、本手法により、一拍毎の心電図からの心拍間隔（RR 間隔）と手首脈波から検出した脈波間隔の相関係数は 0.96（3 名の平均）と非常に高い相関が得られた。相関図を図 3-7 に、脈波間隔と RR 間隔の平均と、その差分をプロットした Bland and Altman plot を図 3-8 に示す。

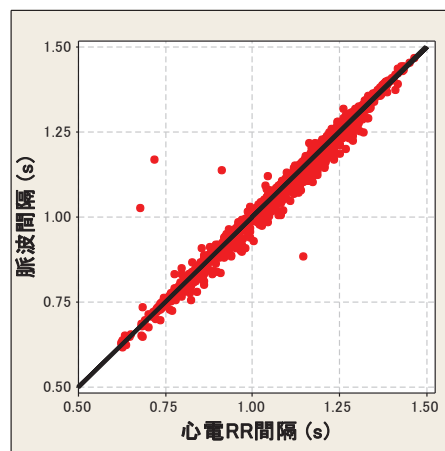


図 3-7 脈波間隔と RR 間隔の相関図

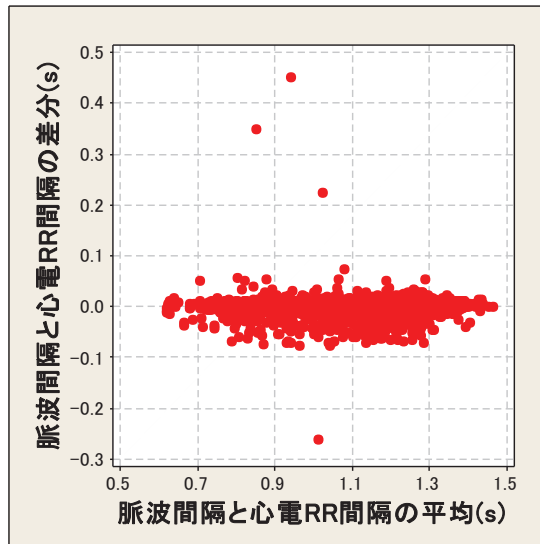


図 3-8 Bland and Altman plot

このように、もともと手の平等で計測した脈波より品質が悪い手首脈波を用いて、心電図 RR 間隔と遜色ない精度良い脈波間隔検出を実現した。

### 3. 3. 3 自律神経解析手法

脈波計測から自律神経解析までの処理の流れを図 3-9 に示す。

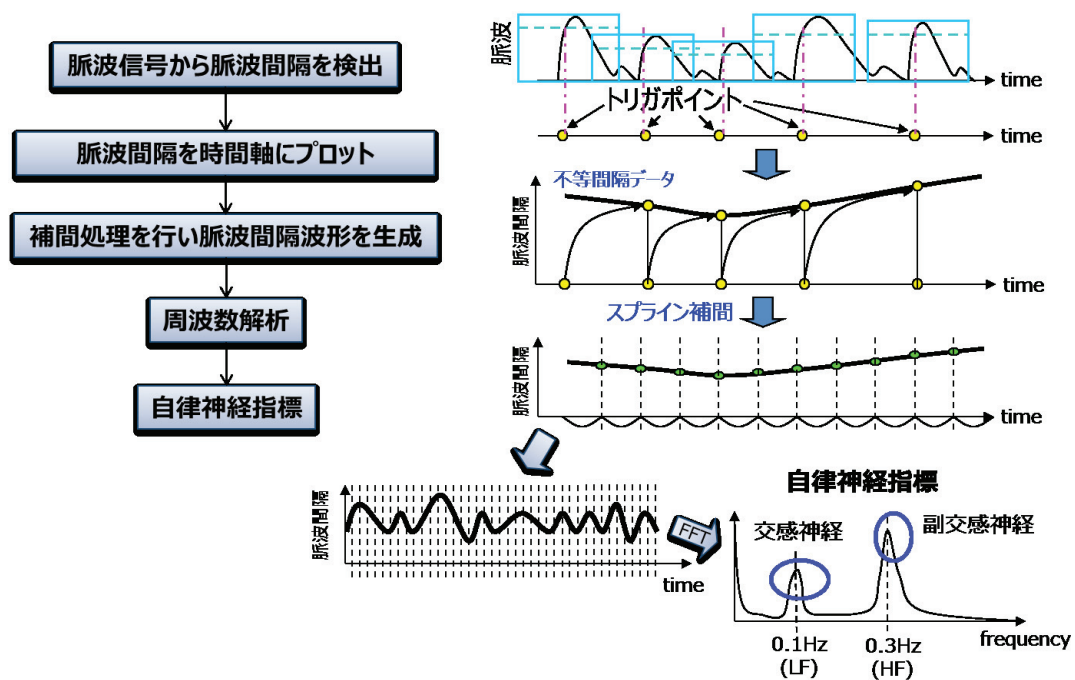


図 3-9 脈波からの自律神経解析手法

上述の通り、手首脈波から精度良く脈波間隔の検出を行った後、脈波間隔を縦軸に置き、脈波間隔の波形を生成する。この波形は不等間隔データであるため、スプライン補間を行ってから周波数解析（FFT）を行うと、0.1Hz 前後（LF: Low Frequency）と 0.3Hz 前後（HF: High Frequency）にそれぞれピークが出現する。これらは自律神経系の活動を反映しており、前者は交感神経活動（ただし副交感神経活動の修飾も受ける）を、後者は副交感神経活動を表す。このようにして、脈波から自律神経解析を行う。

### 3. 3. 4 自律神経活動からの睡眠解析手法

睡眠状態は REM 睡眠、NREM 睡眠に大別され、NREM 睡眠にはさらに睡眠の深さが定義されている。睡眠中の自律神経活動を見ると、NREM 睡眠中は副交感神経が優位、REM 睡眠中は交感神経が優位（副交感神経が抑制）で、自律神経系が乱れるといった特徴があり、これらに関連する特徴量を用いてクラスタリングする。ただし、自律神経活動は年齢、性別、身体状況などによって影響を受けるため個人差が大きく、睡眠状態も日中の活動状態や概日リズム（サーカディアンリズム）など、いろいろな要因の影響を受ける。そこで、一般的に睡眠の周期は 90 分であるとされていることから、睡眠の一周期を十分に含む 120 分のデータで k-means 法によるクラスタリングを行い、REM/NREM 睡眠を分類し、NREM 睡眠の深／浅を分類することで、これらの影響を排除した睡眠解析を行う。

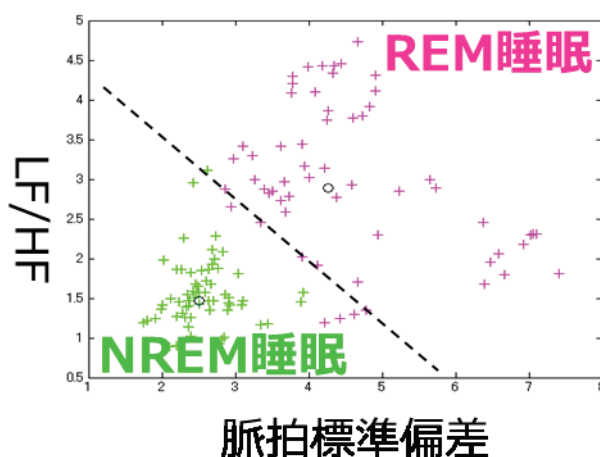


図 3-10 REM 睡眠と NREM 睡眠の分類

### 3. 3. 5 ウェアラブル睡眠センサ

前章のウェアラブル健康管理システムで用いたウェアラブル生体センサをベースにしつつ、上述の睡眠解析手法に必要な機能のみを搭載したウェアラブル睡眠センサを開発した。センサ本体の外観を図 3-11 に、PC も含めたシステム全体とそれぞれの処理内容を図 3-12 に示す。



図 3-11 ウェアラブル睡眠センサ



図 3-12 ウェアラブル睡眠センサ システム全体構成

ウェアラブル睡眠センサ本体では、脈波間隔、体動量の算出処理までを行い内蔵フラッシュメモリに記憶する。朝起きたら PC に USB で接続し、PC にデータを転送し、PC にて本章で提案したアルゴリズムを実行して自律神経解析、睡眠状態推定を行い、結果を画面に表示する。結果表示の例を図 3-13～3-15 に示す。



図 3-13 睡眠解析結果 (サマリー)

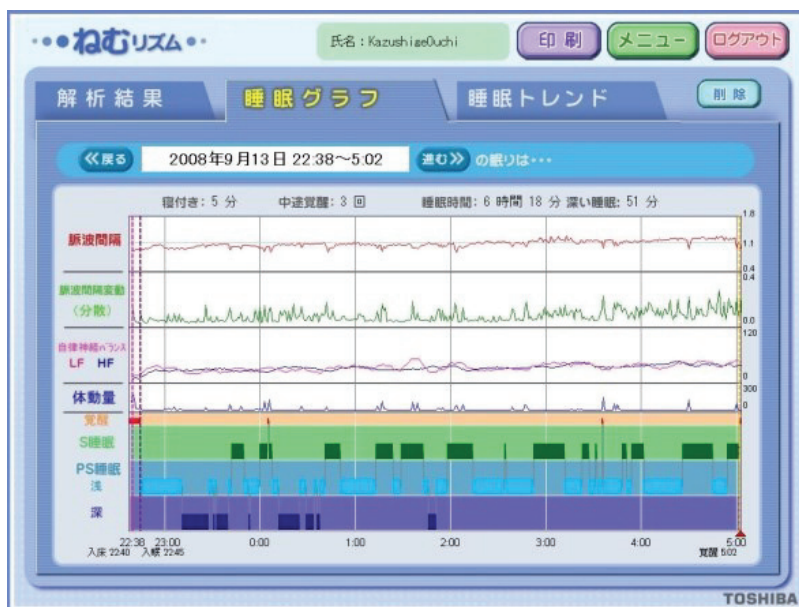


図 3-14 睡眠解析結果 (睡眠グラフ)





図 3-15 睡眠解析結果（睡眠トレンド）

### 3. 3. 6 評価実験

#### 体動量による睡眠／覚醒判定精度

図 3-1 で説明した通り、まずは体動量から睡眠／覚醒の判定を行っているため、その精度の評価として、市販の米 AMI 社の腕時計型加速度センサロガー（アクティグラフ）<sup>[35]</sup>との同時計測を行い、判定のベースとなる 1 分間あたりの体動量を比較した。被験者は 3 名で、相関係数は平均 0.95 と高い相関が得られた。うち 1 名の終夜比較データを図 3-16 に示す。体動量による睡眠／覚醒判定精度は市販品と遜色ないことを確認した。

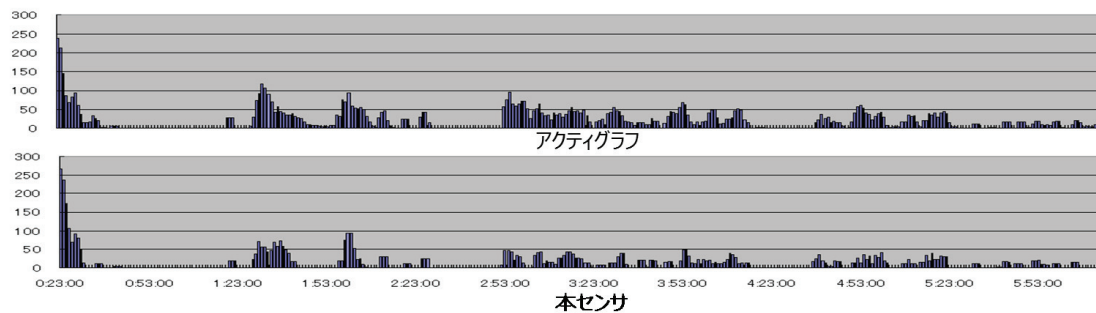


図 3-16 アクティグラフと本センサの体動量比較結果の例

## 睡眠状態判定精度

終夜睡眠ポリグラフと本センサを同時に装着し、それぞれによる解析結果の一致度を比較した。被験者は19～72歳の45名（男性30名、女性15名）で、終夜睡眠ポリグラフの検査結果は睡眠専門医が目視で判定したものを、本センサで自動判定した結果と比較した。判定結果の一例を図3-17に示す。一部判定結果が異なっている場合もあるが、睡眠周期などは概ね一致しており、睡眠状態を良好に再現できていることがわかる。また、表3-2に示すように、全被験者における全体の一致率の平均は73.5%となり、睡眠専門医から在宅で使える簡易モニタとしては十分な性能であるとの見解を得た<sup>[36]</sup>。

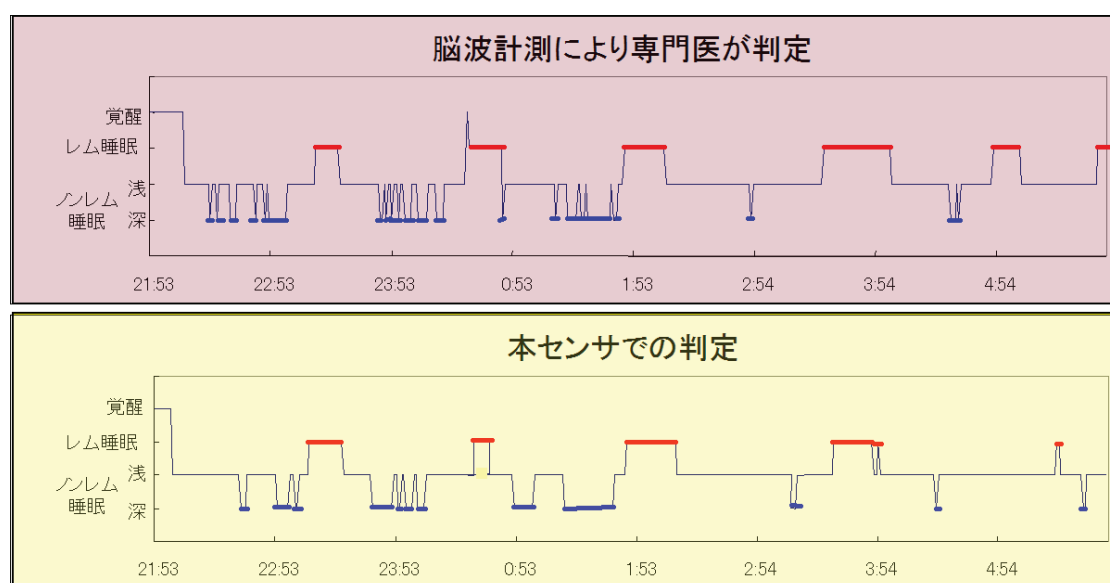


図 3-17 終夜睡眠ポリグラフと本センサの自動判定結果の比較

表 3-2 比較評価結果

	一致率				
	深睡眠	REM 睡眠	NREM 睡眠	覚醒	全体
平均	0.740	0.712	0.697	0.799	0.735
標準偏差	0.079	0.067	0.080	0.100	0.052

尚、本センサは東芝ホームアプライアンス株式会社から 2009 年 6 月に製品化した。製品としての仕様を表 3-3 に示す。

表 3-3 ウェアラブル睡眠センサ仕様

品名	東芝体動計
型名	NEM-T1
体動エポック	1分
加速度分解能	0.01G(3軸加速度の合成値)
脈波計測方式	光電式
脈波間隔計測対応範囲	35～135bpm相当
サンプリング周波数	64Hz(*)
連続計測時間	40時間
連続データ記録時間	112時間
寸法、重さ	55×57×14mm
通信方式	USB2.0(専用ケーブル付属)
使用電源	リチウムイオン充電電池(3.7V)

(\*) 脈波間隔の分解能は補間処理により 0.1ms の分解能を確保

※医療機器(薬事法)の認可取得 製造販売届出番号: 40B1X00004T00001

### 3. 4 快適目覚まし機能の検討

上述したウェアラブル睡眠センサは、図 3-12 に示すように、就寝前に手首にセンサを装着してデータ収集を開始し、起床後 PC に USB ケーブルで接続してデータを転送し、PC にて睡眠解析を行い、結果を画面上に表示することで、自宅等での手軽な睡眠モニタリングを実現するものである。本項では、無線通信機能を搭載した無線版ウェアラブル睡眠センサを開発し、リアルタイムに状況認識(睡眠解析)を行い、快適に起きられるタイミングでタイムリーにアラームを鳴らすヒューマンインタフェース(快適目覚まし機能)の検討を行った。

#### 3. 4. 1 リアルタイム睡眠解析システム

リアルタイム睡眠解析の実現に向けて、まず、無線通信機能を搭載した無線版ウェアラブル睡眠センサを開発した。外観を図 3-18 に、概略仕様を表 3-4 に示す。Bluetooth モジュールを内蔵して無線通信を可能にした以外は、前節まで述べたウェアラブル睡眠センサとほぼ同一の構成である。本センサで計測した脈波および加速度のデータを PC へ Bluetooth 経由で連続送信し、PC 上でリアルタイムに睡眠解析可能なリアルタイム睡眠解析システムを開発した。システム全体の概要を図 3-19 に示す。



図 3-18 無線版ウェアラブル睡眠センサ

表 3-4 無線版ウェアラブル睡眠センサ 概略仕様

搭載センサ	光電脈波センサ 3軸加速度センサ	
サンプリング周波数	16Hz/20Hz/64Hz	
内蔵メモリ	8MB	
通信機能	USB Bluetooth	
バッテリー	リチウムイオン電池	
連続使用時間	Clock only mode	2ヶ月以上
	オフラインモード	80時間
	オンラインモード	15時間

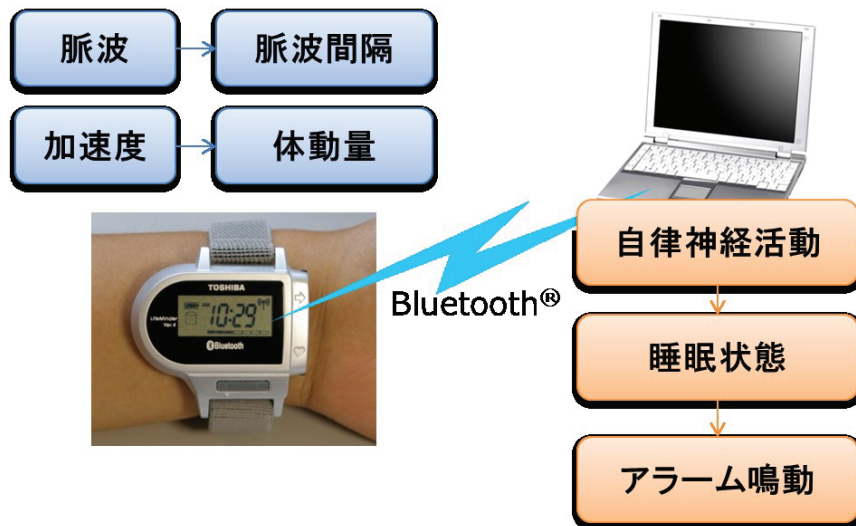


図 3-19 リアルタイム睡眠解析システム

### 3. 4. 2 仮説検証

快適目覚まし機能の実現に向けて、まずは快適に目覚められるタイミングがいつなのかを明らかにする必要がある。筆者らのこれまでの経験則に基づき、まず、以下の仮説を立て、これらの仮説が正しいかどうかを確認する予備実験を実施した。

- 仮説 1 : NREM 睡眠の深睡眠時に起きると不快に感じる
- 仮説 2 : REM 睡眠時に起きると不快感を伴いにくい

20～30 代の被験者 4 名に対し、就寝中の睡眠状態をリアルタイム睡眠解析システムでモニタリングし、NREM 睡眠中、および REM 睡眠中の各タイミングでアラームを鳴動させ、その際の目覚め感を快適 2 段階、不快 2 段階の計 4 段階で評価した。被験者にはどちらのタイミングでアラームを鳴らしたかはブラインドし、就寝からアラーム鳴動までの時間を最低 4 時間以上確保して実施した。被験者毎に実施数は異なるが、NREM 睡眠中、REM 睡眠中ともに計 9 回分実施した。それぞれのタイミングでアラーム鳴動した際の目覚め感を表 3-5、3-6 に示す。

表 3-5 NREM 睡眠中にアラーム鳴動した場合の目覚め感

被験者	とても快適	快適	不快	とても不快	計
A (30 歳、男性)	0	0	1	0	1
B (28 歳、男性)	0	1	0	0	1
C (36 歳、男性)	0	0	4	1	5
D (22 歳、男性)	0	0	2	0	2

表 3-6 REM 睡眠中にアラーム鳴動した場合の目覚め感

被験者	とても快適	快適	不快	とても不快	計
A (30 歳、男性)	1	1	0	0	2
B (28 歳、男性)	1	1	1	0	3
C (36 歳、男性)	0	2	1	0	3
D (22 歳、男性)	0	1	0	0	1

NREM 睡眠中、REM 睡眠中それぞれにおけるアラーム鳴動時の目覚め感をまとめると表 3-7 の通りとなる。REM 睡眠中にアラームを鳴動させて起床した場合の方が、NREM 睡眠中に起床した場合よりも有意に ( $P < 0.05$ ) 快適な目覚めが得られることが確認された。

表 3-7 アラーム鳴動時の目覚め感

	とても快適	快適	不快	とても不快	計
NREM 睡眠中	0	1	7	1	9
REM 睡眠中	2	5	2	0	9
計	2	6	9	1	18

### 3. 4. 3 快適目覚まし機能

仮説検証により、REM 睡眠中にアラームを鳴動させて起床させることで、快適な目覚めが得られる可能性を確認したが、実際に日常生活で使用する際にどのように設定すべきか検討が必要である。

通常の日覚まし時計は、就寝前に起床時刻を設定し、睡眠状態に関わらずその時刻にアラームを鳴動させる。これに対し快適目覚ましは、起床時刻を設定して就寝してもその時刻に REM 睡眠中である保証がないため、起床希望時間帯を設定することとする。図 3-20 は、起床希望時間帯を 5:00 から 7:00 に設定した場合のデータ解析、アラーム鳴動を説明した図である。希望起床時刻は 7:00 であるが、その前に快適に目覚められるタイミングがあれば、そのタイミングでアラームを鳴動させるという設定である。

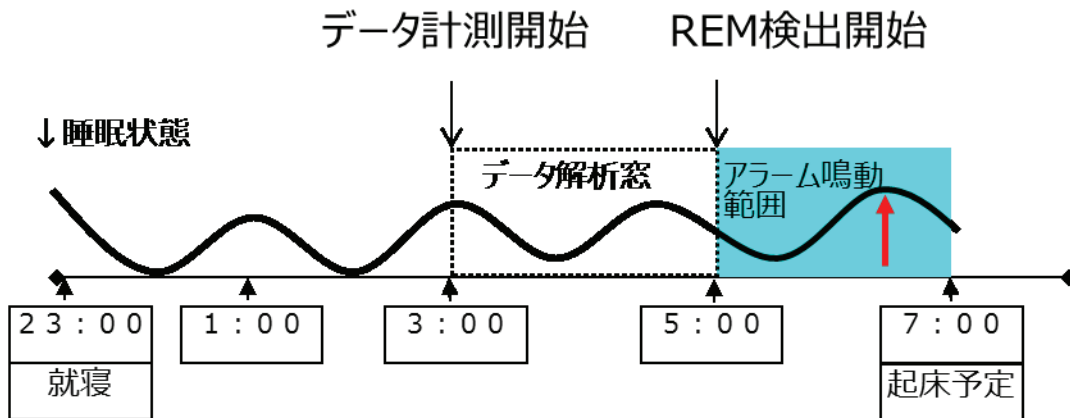


図 3-20 起床希望時間帯を設定したアラーム鳴動

前述の通り、本手法では睡眠の 1 サイクルが十分に含まれる 120 分のデータでクラスタリングを行うため、5:00 から REM 検出を始めるためにはその 120 分前の 3:00 からデータ計測を開始する。5:00 からは設定サイクル（例えば 1 分）ごとに直近の 120 分のデータでクラスタリングを行い、現在 REM 睡眠中かどうかを判定する。ここで、REM 睡眠であると検出してすぐにアラームを鳴動させると、希望起床時刻である 7:00 よりだいぶ早く起床してしまい、ユーザによってはもっと寝たいと思う場合も想定される。そこで、REM 睡眠かつ極力希望起床時

刻に近いタイミングでアラーム鳴動させるため、下記の方針でアラーム鳴動させることとした。

- 希望起床時刻（図 3-20 の例では 7:00）に REM 睡眠中であればその時刻にアラーム鳴動
- 希望起床時刻前に REM 睡眠から NREM 睡眠に変化した場合はそのタイミングでアラーム鳴動
- 起床希望時間帯（図 3-20 の例では 5:00 から 7:00）に REM 睡眠が検出されなかった場合は希望起床時刻にアラーム鳴動

上記方針でアラーム鳴動させることで快適な目覚めを提供できることを検証するため、NREM 睡眠中、REM 睡眠中だけでなく、NREM→REM 変化時、REM→NREM 変化時の各タイミングでアラーム鳴動させ、その際の主観を評価した。予備実験と同様に、被験者にはどちらのタイミングでアラームを鳴らしたかはブラインドし、就寝からアラーム鳴動までの時間を最低 4 時間以上確保してアラーム鳴動し、起床時の目覚め感を快適 2 段階、不快 2 段階の計 4 段階で評価した。被験者は健常者 9 名（27～42 歳：男性 5 名、女性 4 名）で、計 70 セット実施した。それぞれのタイミングでアラーム鳴動した際の目覚め感を表 3-8 に示す。

表 3-8 アラーム鳴動時の目覚め感

	とても快適	快適	不快	とても不快	計
NREM起床	0	6	13	4	23
REM起床	3	10	5	0	18
NREM→REM起床	1	6	5	0	12
REM→NREM起床	8	6	3	0	17
計	12	28	26	4	70

表 3-8 の結果について、明確に主観が分かれている部分に差を持たせるために「とても快適」100 点、「快適」75 点、「不快」25 点、「とても」不快 0 点と点数付けを行い、 $\chi^2$  乗検定を用いて検定した。その結果、各睡眠状態変化間において有意な差があることが検証された ( $p < 0.01$ )。また、4 群における分散分析での群間差を求めるため、さらに多重比較検定を行った結果、NREM 起床に対して、REM 起床と REM→NREM 起床において差が認められた。

無線版ウェアラブル睡眠センサと PC で構成したリアルタイム睡眠解析システムでリアルタイムに睡眠状態を解析し、タームリーなタイミング（REM 睡眠中、あるいは REM 睡眠から NREM 睡眠への変化時）でアラーム鳴動させることによ

り、NREM 睡眠中に起床した場合に比べて有意に快適に目覚められる傾向があることが確認でき、状況認識技術により日々の生活で快適な目覚ましを提供するヒューマンインタフェースの実現可能性を確認した。

### 3. 5 まとめ

本章ではウェアラブル睡眠センサを開発し、自宅や旅行先での手軽な睡眠状態チェックを実現した。また、リアルタイムに睡眠解析可能なリアルタイム睡眠解析システムにより、状況認識（睡眠解析）に基づくタイムリーな快適目覚まし機能の実現可能性を確認した。本機能は睡眠状態を把握し、それに基づいたタイムリーなアクション（アラーム鳴動）を行うものであったが、今後は図 3-21 のように、空調機器、照明機器などとも連携し、睡眠自体を理想的な状態にコントロールするような、より能動的な機能の検討により、人々の不眠をさらに軽減する快適睡眠環境ソリューションが期待される。

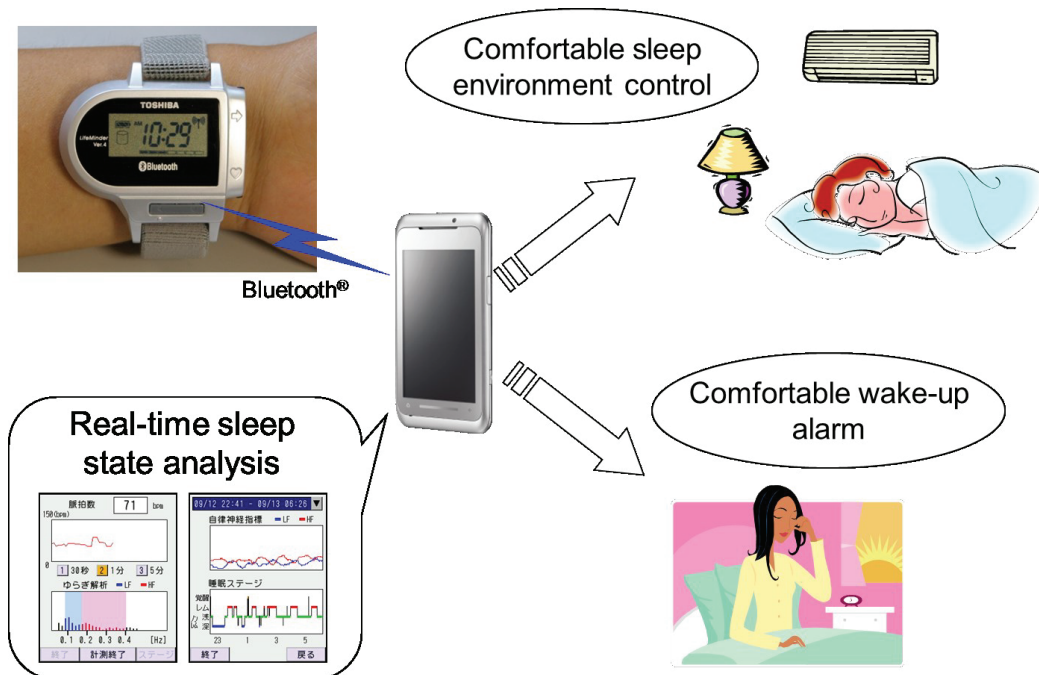


図 3-21 快適睡眠環境ソリューション



## 4. スマートフォンを用いたリアルタイム生活行動認識システム

---

### 4. 1 はじめに

加速度センサ、GPS (Global Positioning System) など、ユーザの状況、位置などの把握に活用可能なセンサが、携帯電話、スマートフォンなど小型の情報端末へ搭載され、ユーザの現在の状況に基づいて適切なサービスを提供する状況依存型サービスの普及が期待されている。一方、社会の高齢化に伴い、高齢者（特に独居の場合）の家庭内の生活行動を見守ることは、安否確認だけでなく、認知症の早期発見や、生活行動の変化に基づいた適切なタイミングでの問いかけ（離れた家族から電話をかけるなど）による高齢者の QOL (Quality of Life) 向上などの観点でも、重要性が今後更に増してくると考えられる。家庭内生活行動見守りのためには、部屋間の移動、食事、掃除、炊事、洗濯、入浴、排泄など、様々な行動の認識が必要となるが、屋内では GPS による測位が使えず、また単一の加速度センサでこれらの行動を精度良く把握することは困難であるため、上述の小型情報端末上の状況依存型サービスとは別のアプローチがこれまで試みられてきた。例えば、家庭内の至る所に各種センサを配置し、複数のセンサ情報を統合してユーザの行動を判断する取り組みがある<sup>[14][15][16]</sup>。専門の施設等ではこのような方式による実運用の可能性も考えられるが、一般の家庭に持ち込むためには、設置コスト、運用コストが大きな問題となる。これらに対し、本章では、携帯電話、スマートフォンなどの一般的な小型情報端末にすでに搭載されている加速度センサとマイクによる家庭内の生活行動見守りを検討した。簡易で安価なドア開閉センサ程度であれば、一般家庭でもコスト的に許容でき、提案する手法と環境側センサによる位置推定とを併用することによる更なる性能向上の可能性も考えられるが、環境側センサとの連携は、本章で提案する一般的な小型情報端末単体のみによる生活行動認識をベースとした上で、想定するサービス毎に必要なとされる性能、許容されるコスト等を踏まえて検討されるべきであると考えられる。

本章では、まず、小型情報端末を用いた状況認識に関する関連研究を概観し、目指す方向性を明らかにする。次に、加速度センサの情報を活用してユーザの動作状態を大まかに分類し、推定した状態に応じて環境音を分析して詳細な作業状態を分類する提案手法について説明する。続いて、スマートフォンでリアルタイ

ムに生活行動を認識するシステムを開発し、実際の家庭内生活のデータを収集して、提案手法の基本性能および実用性能を確認した結果を述べ、最後に考察する。

## 4. 2 従来研究

身体に装着した加速度センサによりユーザの動作状態を認識する研究は多い。例えば、加速度センサを左右の肩、肘、手首、腰、膝、足首に計 12 個装着した研究<sup>[37]</sup>や、上腕、手首、腰、太腿、足首に計 5 個装着した研究<sup>[38]</sup>などがあり、多いものでは 20 種類の動作状態の推定を行っている。複数の加速度センサを装着することで、詳細な動作状態の推定が可能となるが、日常生活でこれらのように多数の加速度センサを常時身につけることは、拘束性が高く受け入れがたい。

これに対し、近年、携帯端末への加速度センサの搭載が進んでいる。特にスマートフォンには画面の縦横切替などの機能向けに加速度センサが必ず搭載されている。このような背景のもと、携帯端末内蔵の単一の加速度センサを活用した動作状態認識の研究も行われている。例えば、携帯電話内蔵の 3 軸加速度センサのデータを基に、歩行、速歩、階段上り、階段下り、走行の 5 動作状態を約 80%の精度で認識する研究<sup>[39]</sup>がある。他の研究<sup>[40][41]</sup>でも歩行、走行を含む 4 種類程度の動作状態推定を行っている。本章でターゲットとしている家庭内生活行動見守りのためには、歩行、走行などの動作状態だけでなく、家事などの作業状態も認識対象とする必要があるが、単一の加速度センサだけで作業状態を推定することは困難である。

一方、携帯電話、スマートフォンには通話用のマイクがあるため、これをセンサとして活用することも可能である。環境音の分析による状況推定に関する研究としては、音特微量のみで屋内外 17 種類の生活シーンを平均 68.4%の精度で分類する取り組み<sup>[42]</sup>などがある。しかし、加速度センサのデータは例えば 10bit 20Hz 程度でサンプリングすれば良いが、文献<sup>[42]</sup>では 16bit 48kHz で音データをサンプリングして分析している。環境音分析では、加速度の分析に比べて扱うデータ量が 3 桁以上増え、また、周波数成分の抽出のためには FFT (Fast Fourier Transform) を実行するなど、ある程度の計算量を要する処理が必要となり、環境音分析を常時行うこと（音声認識を常時動かし続けることに相当する）はできれば避けたい。また、音声認識と同様に、分析すべき区間を切り出して認識させる手法が有効であると考えられる。そこで、本章では、加速度センサでユーザの動作状態を大まかに推定し、推定した状態に応じて環境音分析を実施する手法を提案する。

携帯電話に搭載された加速度センサとマイクを活用した状況認識に関する研究としては、加速度センサ、マイク、GPS を用いてユーザの移動状態を推定する

報告がある<sup>[43]</sup>。走行、歩行、停止に加えて自転車、電車、バス、自動車などの移動状態を推定する場合、加速度センサだけでは精度が低下するため、マイクからの環境音と GPS の位置情報とあわせて分析することで精度向上を図っている。具体的には、加速度センサで移動状態が電車、バス、自動車のいずれかであると判断した場合にマイクを使って自動車であるかそれ以外であるかを判断する。その後、電車とバスの判断には GPS から得られる位置情報の時間変化を使っている。他にも携帯電話に複数のセンサを搭載してユーザの状態を推定する試み<sup>[44][45]</sup>が行われているが、文献<sup>[44]</sup>では携帯電話のマナーモードへの自動切り替えなどを行うためにユーザの状態を Normal / Idle / High Activity / Uninterruptible の 4 つに分類し、文献<sup>[45]</sup>では動作状態としては歩行、走行、着席の 3 状態を推定しているなど、家庭内生活行動の推定は行われていない。これに対し、本章でのターゲットは家庭内の生活行動見守りであるため、推定対象は移動状態だけでなく家事などの生活状態も含む点が異なる。

また、携帯電話ではなくウェアラブルセンサを用いた行動認識に関する研究もある。手首、腕、胸元などに加速度センサ、マイクを複数装着して、木工作業場における 8 種類程度の作業の識別を行った研究がある<sup>[46]</sup>。また、加速度センサ、マイクに加えて、カメラが搭載されたウェアラブルセンサを手首に装着して、歯磨き、掃除機がけ、皿洗い、コーヒーを作る、ココアを作るなど、15 種類程度の生活状態の認識を行っている研究もある<sup>[47][48]</sup>。これらの取り組みでは、身につけるセンサ数を増やす、あるいはカメラで操作対象物の画像の特徴を捉えることで、細かな行動の識別を試みている点が特徴である。これらの取り組みに対し、本章では、特別なデバイスを必要とせずに、日常的に使われている携帯電話、スマートフォン単体による家庭内の主要な生活行動の認識を目指している点が異なる。また、これまでは加速度と音を同時に解析して行動を識別する手法が試みられてきたが、我々はまず加速度で大まかな動作状態を分類し、必要な場合のみ音による作業状態分類を行う点が異なる。

## **4. 3 加速度と音による状況認識**

### **4. 3. 1 認識対象とする生活行動**

一口に生活行動といってもその内容は多岐にわたる。このため、食事、着替え、移動、排泄、整容、入浴など生活を営む上で欠かせない基本的な活動を指す日常生活動作 (ADL: Activities of Daily Living) と、炊事、掃除、洗濯などの家事全般や、金銭管理、服薬管理、外出など、ADL より複雑で高次の活動を指す手段的

日常生活動作（IADL: Instrumental Activities of Daily Living）による評価が高齢者介護やリハビリテーションの分野で用いられている。高齢者・障害者の生活自立度評価の際には、バーセルインデックスに代表される ADL の評価だけでは不十分で、IADL も重要な指標であるとされている<sup>[49][50]</sup>。

また、認知症の在宅ケアではセンター方式（認知症の人のためのケアマネジメント センター方式）<sup>[51][52]</sup>という認知症の人のためのケアマネジメントシートが用いられている。認知症患者本人、家族、ケア関係者が共通のシートを使って、対話しながら効果的な支援策を見つけていくためのもので、その中の D-1（私ができること・私ができないことシート）では、ADL、IADL に相当する暮らしの場面が 22 項目あり、それぞれについて自分でできるかどうか、定期的に記入していく。これらの項目の大半あるいは一部だけでもセンサの活用で負担なく取得できたり、患者によって特に着目すべき項目について、日々の状態が連続的にモニタできたりすることは、認知症ケアにとっても重要であると考えられる。よって、我々が目指す小型情報端末による家庭内生活状態見守りでは、ADL、IADL の両方を認識対象とする。

尚、本論文では、主に高齢者の家庭内生活行動見守りのために検出すべき作業状態として、ADL、IADL の中から家事、排泄管理、整容の代表的なものについてデータ収集を行い、提案手法による生活行動認識の基本的な適用可能性を確認する。しかし、今回対象とする作業状態以外にも、家事であれば、炊事、洗濯など、排泄以外の基本的な生活行動としては食事、睡眠、入浴、着替えなど、整容であれば洗顔、化粧など、ADL、IADL には他にも分類対象として検討すべき生活行動が多く存在する。

また、総務省統計局の平成 18 年（2006 年）の調査結果<sup>[53]</sup>によれば、平成 13 年（2001 年）に比べて特に高齢者は男女ともに家事関連の生活時間が増加している。また、身だしなみ、化粧、入浴、トイレなど身の回りの用事的生活時間は、高齢になるほど長くなり、全年齢で 10 年前に比べて顕著に増加している。このことから、高齢者の家事や身の回りの用事に関する生活行動を見守ることは今後更に重要性が増すと考えられるため、分類対象とする生活行動の数を増やす方向で検討を進める方針である。

ただし、今後、様々なユースケースを検討していく上で、そのユースケースにとって分類が必要な作業状態のうち、音に基づいた作業状態分類が困難な場合も想定される。例えば、食事、睡眠など、人間の基本的な生活状態の取得は、高齢者・障害者に限らず生活習慣の管理には重要である。しかし、本章で提案した手法では、食事については、動作状態が「作業」だけでなく「安静」と判断される可能性があり、また、食事に伴う音（咀嚼音、食器の音など）に加えて、複数人で食事をしている場合は、他の人が発する音や会話の音、場合によっては TV やステレオの音なども含まれるため、本手法では分類が難しいと想定される。また、

睡眠時はほとんどの時間が「安静」と判断されることになり、時刻と合わせることである程度は推測できるが、直接的には分類が困難である。

これらについては、ターゲットとしたユースケースにおいて許容される拘束性に依存するが、2章、3章で提案したように腕時計型の生体センサを装着して、脈波、手の動き（加速度）、皮膚温度、皮膚電気反応（GSR: Galvanic Skin Response）などを計測することによる食事検出、脈波、体動を計測することによる睡眠状態判定などが可能であり、これらの計測手法と本提案手法とを組み合わせた利用形態も考えられる。

## **4. 3. 2 想定するサービスイメージ**

上述の通り、日常生活において負担なく ADL、IADL を継続的にモニタリングすることは、特に高齢者、障害者、認知症患者などにとって重要であるが、対象者によってモニタすべき ADL、IADL は異なることが考えられる。例えば、特定の部位に身体的な障害を持っている人と、認知症の人とでは、モニタすべき ADL、IADL は異なることが想定される。

また、家電機器（掃除機、電気シェーバーなど）の動作音、キッチンやトイレの水洗音などは家庭によって異なり、歯磨きに一般的な歯ブラシを使用するか／電動歯ブラシを使用するかなど、同じ生活行動において発生する音は対象者や環境によって異なる。あらゆる環境に適用可能な汎用性の高い生活行動認識の実現が最終的には理想であるが、本研究では、次のようなサービスイメージを想定した場合に有用であると考える手法を提案する。

独居の高齢者や、障害者、あるいは認知症患者（以下、対象者とする）などの生活状態をモニタリングするため、その家族あるいはケア関係者が、対象者の見守りに必要な ADL、IADL のデータを収集する。つまり、立ち会いの下で認識対象の各生活行動を行ってもらい、その際に収集したデータをシステムに学習させ、各生活行動を認識させる。その際、学習に必要な時間が長いと双方にとって負担になるため、学習はできるだけ短い時間で完了することが必要である。短い時間のデータで学習を行うことは、学習時の計算量、計算時間の短縮、および省電力の点からも望ましい。モニタする本人の実際の生活行動のデータを学習するため、家庭による音の違い、使用する機器による音の違いなどがなく、短い学習時間であっても、精度良い認識が可能になることが期待できる。認識結果は、リアルタイムにサーバへアップロードして家族あるいはケア関係者がいつでもモニタリングできるようにする、あるいは携帯電話内に記録して後から参照するなど、用途によって必要なリアルタイム性に応じた構成とする。このようなサービスが実現されると、例えば、いつも必ず 1 日 3 回歯磨きをしていた対象者が歯磨きのし忘

れが増えてきたことに（見守る側の）家族が気づいた時にさりげなく電話をかけて近況を確認するなど、認知症の予兆の早期発見や、QOL 向上などへの貢献が期待できる。

上記サービスを実現するため、本研究ではまず、ADL の移動の基本行動としての「歩行」、整容に含まれる「歯磨き」、「電気シェーバーによる髭剃り」、「ドライヤーの使用」、排泄と関連する「トイレ水洗／手洗い」、および IADL の家事に含まれる「掃除機がけ」、「皿洗い」、「アイロンがけ」を対象として、提案手法による家庭内生活行動認識の適用可能性を検証する。尚、これらの生活行動は、一般的な独居高齢者向けのモニタリングを想定して選択した。4.3.1 節で触れたセンター方式 D-1 の全項目数の約 1/3 に相当する。

尚、提案手法は、一般的な携帯電話、スマートフォンを身につけるだけで複数の家庭内生活行動を認識できる点を特長とする一方で、家庭内でも携帯電話を持ち歩くことを前提としている。この前提条件は、これを身につけることで離れた家族が自分を見守ってくれるという安心感に加え、複数の家電を簡単に操作できるマルチリモコン機能など、家庭内の様々な場所でユーザにメリットがある機能を提供することにより、許容されるとの立場で議論する。

### **4. 3. 3 システム構成**

本章で提案する屋内外生活行動センシングシステムの全体構成を図 4-1 に示す。スマートフォン上で動作する屋内外生活行動センシングエンジンは、GPS 衛星の捕捉状況を利用した屋内外判定機能により、屋内（自宅内）にいると判断した場合は屋内生活行動推定エンジンを使用し、屋外（自宅外）に出たと判断した場合は屋外移動状況推定エンジンに切り替える。この機能により、同一のスマートフォンを携帯し続けるだけで、屋内外で使用するエンジンを適切に切り替えて、シームレスな生活行動センシングを行うことが可能となる。現段階では、単純に GPS 測位が可能かどうかを屋内外の判断基準としているが、屋内でも窓際などで GPS 測位可能となる場合がある。この場合は、自宅の場所を登録しておくことで、GPS 測位可能でも自宅内と判断して屋内生活行動推定エンジンを使用するようになれる。逆に、自宅外で GPS 測位が可能でなくなった場合（建物の中にいる場合など）も、最後の GPS 測位の結果を利用することで自宅外であることがわかり、屋外移動状況推定エンジンを利用できる。

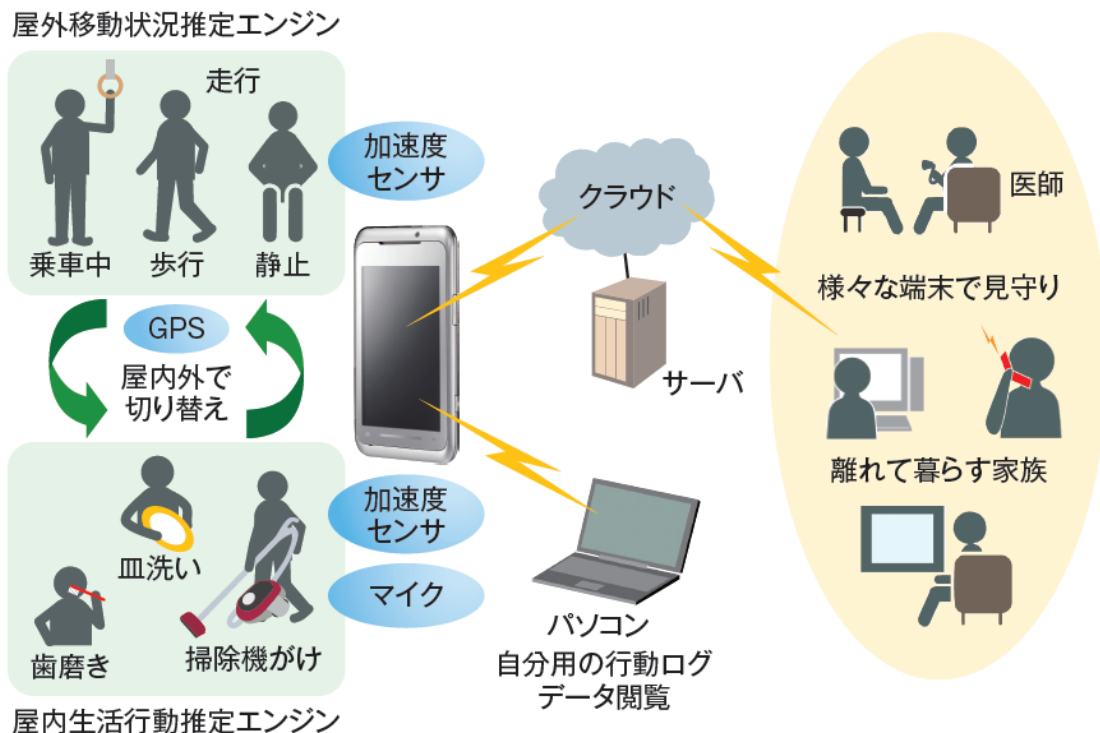


図 4-1 屋内外生活行動センシングシステム全体構成

スマートフォンで認識した生活行動の結果はスマートフォン上で閲覧することが可能であるが、認識結果を Bluetooth 経由でリアルタイムに PC など、近くの外部端末へ送信し、その画面で結果を閲覧することもできる。また、認識結果のログを定期的あるいは特定のイベント発生時などのタイミングでサーバへ送信し、医師や離れた暮らす家族などが対象ユーザの生活行動を把握することも可能である。この機能は特に独居高齢者の見守りサービスなどに有用である。

#### 4. 3. 4 屋内生活行動推定エンジン

屋内生活行動推定エンジンの処理概要を図 4-2 に示す。スマートフォンは胸ポケットに格納して携帯していることとする。まず、加速度センサのデータにより、動作を「歩行」「作業」「静止」の3つに分類する。「作業」とは、「歩行」以外に、身体的動きを伴う何らかの生活行動をしていると想定される状態のこととし、「静止」とは「歩行」も「作業」もしていない、身体を動かしていない状態のこととする。次に、「作業」と分類された場合にはマイクを起動し、作業に伴う周囲の音を分析して作業内容を推定する。

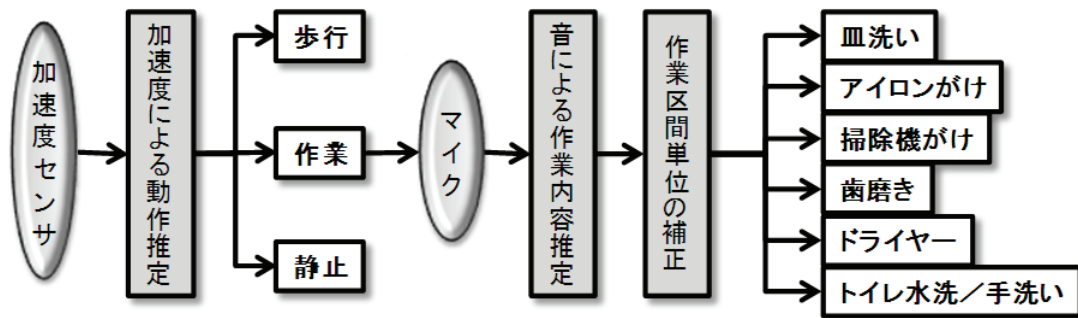


図 4-2 屋内生活行動推定処理フロー

## 加速度による動作推定

加速度による動作推定は、統計量として 1 秒間の分散のみを用いる簡便な手法を用いる。10bit 20Hz で 3 軸加速度のサンプリングを行い、まず重力加速度方向の加速度の 1 秒間の分散を算出する。これは歩行時に最も顕著に加速度変化が現れるのが重力加速度方向であるため、この分散の大きさによって「歩行」か否かを判断する。閾値は実験的に  $0.05[G^2]$  と定めた。「歩行」以外の場合には、3 軸の各加速度の 1 秒間の分散を算出し、最大の分散を持つ軸の値の大きさによって、「作業」か「静止」かを弁別する。その閾値は  $0.0001[G^2]$  とした。これは、平方根を取って標準偏差とすると  $0.01[G]$  となり、睡眠・覚醒判別の際に体動の有無を検出する加速度の大きさとして広く使われている基準値と同一である<sup>[34]</sup>。

## 音による作業内容推定

音による作業内容推定処理の流れを図 4-3 に示す。前述の通り、事前に対象作業の音データを学習しておく。ここで、事前学習に要する時間と手間をできるだけ軽減するためには、短い学習データで運用することが肝要であるため、本システムでは事前学習の時間を各作業につき 10 秒間として作業内容推定を試みている。分類器には、高い汎化性能を持つサポートベクターマシン (SVM: Support Vector Machine) を用いた。音データから抽出する特徴量としては、広く音声認識に使われているメル周波数ケプストラム係数 (MFCC: Mel Frequency Cepstral Coefficients) に加え、算出に際して計算量の比較的少ない二乗平均平方根 (RMS: Root Mean Square)、ゼロ交差率 (ZCR: Zero-Crossing Rate) を使用する。MFCC は人間の聴覚上重要な周波数成分を強調した特徴量で、文献 [41][47] など音を活用した行動認識に関する関連研究でも使用されている。また、RMS は主に音の強度を表し、ZCR は当該区間の音高 (ピッチ) に相当する特徴量で、これらは MFCC に含まれない音の特徴を表すため、これらを MFCC に追加することによる性能向上が期待できる。尚、MFCC の次数は 13 とした。



N 個（ウィンドウ幅。ここでは N=512）の音データ  $\{a_1, a_2, \dots, a_N\}$  の RMS ( $a_{rms}$ ) は、

$$a_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N a_i^2}$$

で算出する。同じく ZCR ( $a_{zcr}$ ) は、

$$a_{zcr} = \frac{1}{N-1} \sum_{i=2}^N F\{a_i a_{i-1} < 0\}$$

で算出する。ここで、関数  $F\{A\}$  は、A が TRUE の場合に 1 となり、そうでない場合に 0 となる関数である。このように、RMS、ZCR は、MFCC に比べて計算量が少なく、MFCC に含まれない特徴を表す特徴量である。抽出した合計 15 次元（MFCC：13 次元、RMS：1 次元、ZCR：1 次元）の特徴量は、それぞれでスケールが異なるため、±1 で正規化を行う。

音データは 16bit 16kHz サンプリングのデータを用いることとし、各特徴量算出の際のウィンドウ幅は 512、ウィンドウオーバーラップは 0 とし、1 秒ごとにそれぞれの平均値を出力して、その都度 SVM による作業内容推定を実施する。SVM ライブラリには LIBSVM<sup>[54]</sup> を利用し、SVM タイプは C-SVC、カーネル関数は RBF (radial basis function) を使用した。学習時には、LIBSVM が提供する最適パラメータ (C,  $\gamma$ ) をヒューリスティックに探す Grid search を用いて最適パラメータを取得し、そのパラメータで学習を行う。作業状態分類時には、1 秒ごとの特徴量平均算出のたびに、SVM による分類を実行し、作業状態の分類結果を 1 秒ごとに得る。

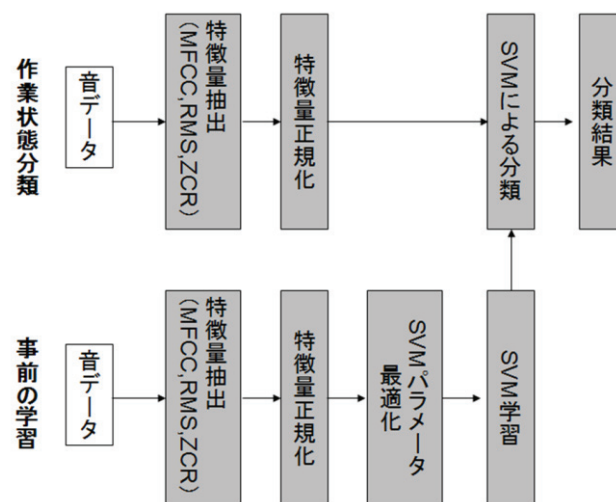


図 4-3 音による作業内容推定処理フロー

## 事例ベース認識手法

ここまでは、動作状態および作業状態を毎秒出力する提案手法の基本的な処理動作について説明した。しかし、実用を想定すると、毎秒の分類性能よりも、作業として連続した区間全体が正しい作業として分類されたかどうかの方が重要である。そこで、連続して切り出された一つの作業区間は同一の作業が継続しているものとし、切り出した区間全体の作業状態を判定する事例ベース認識手法を検討する。

事例ベース認識手法の概念図を図 4-4 に示す。この例は、加速度センサによる動作状態分類で「歩行」、「作業」、「安静」の順に分類され、「作業」の分類結果は「作業 1」、「作業 2」、「作業 3」が混在しているが、この「作業」は全体として、分類された数が最も多い「作業 1」と判断している。また、この例では「安静」中に単発的な「作業」が発生しているが、このような単発的な動作状態変化を無視することとし、動作状態変化のばたつきを軽減する。

本研究では、事例ベース認識手法の導入の妥当性を検証するため、一作業区間の中で最も出現回数の多い作業状態をその区間の正解候補とし、その出現頻度が全体の 50%以上であれば、その区間全体が正解候補の作業状態とする。50%未満の場合は、分類対象外（未学習）の「作業」が行われていることと判断する。

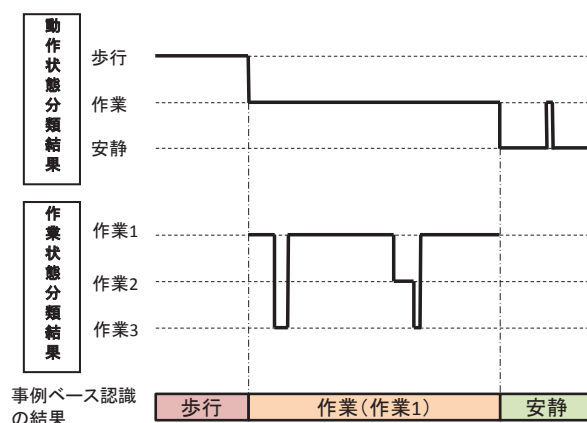


図 4-4 事例ベース認識手法の概念図

## 4. 3. 5 屋外移動状況推定エンジン

屋外の移動状況推定には、池谷<sup>[41]</sup>らの屋外移動状況推定アルゴリズムを活用する。図 4-5 に示すように、加速度センサのデータを用いて、「静止」「歩行」「走行」「乗車中（バス・電車）」の 4 種類の移動状況を推定する。認識対象を屋外移動状況に限定することで実用的な精度で認識可能となる。よって、個別の事前学習は不要とした移動状況推定処理を行う。ここで、各移動状況は排他的なもののみな

す。例えば、「乗車中」の「静止」は「乗車中」とする。また、移動状況推定に加え、GPSを利用して現在地の候補として個別に登録した POI (Point of Interest) を表示する機能も搭載した。

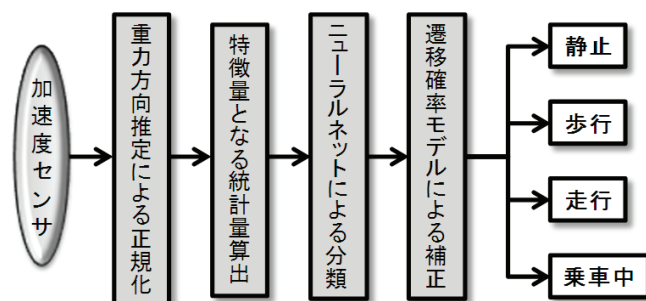


図 4-5 屋外移動状況推定処理概要

## 重力方向推定による正規化

屋外移動時にスマートフォンは上着ポケット、ズボンのポケット、鞆など、ユーザによって携帯する場所は異なり、3 軸加速度センサのデータはスマートフォンの向きに依存する。そこで、重力方向を推定し、その方向を基準にした正規化を行うことで、スマートフォンの向きに依存せずに移動状況を推定する。具体的には、最新のサンプリングから一定時間前までの 3 軸加速度の平均ベクトルを重力ベクトルとし、最新の 3 軸加速度ベクトルから重力ベクトルを減算することで、正規化した加速度ベクトルを求める。

## 特徴量となる統計量算出

正規化した加速度ベクトルから特徴量として、ベクトル長、重力ベクトルとの内積値、重力ベクトルとの外積値をサンプリングの都度、算出する。次に、現在時刻から一定時間幅を持つウィンドウ内での基本的な統計量として、各特徴量について、平均値、最大値、最小値、分散値の 4 種類の統計量を算出し、計 12 次元の統計量を用いて移動状況推定を行う。

## ニューラルネットワークによる移動状況の分類

得られた 12 次元の統計量を「静止」「歩行」「走行」「乗車中 (バス・電車)」の 4 種類の移動状況に分類する。各統計量に対して正解の確信度を 1、それ以外を 0 としてベクトルを教師信号として誤差逆伝播法によって学習した 3 層ニューラルネットワークを分類器として用いる。図 4-6 にニューラルネットワークの構成を示す。学習時のデータ入力の時系列によらないランダムな順序で行い、正解

の移動状況の入力回数の割合が均等になるように配慮して学習を行った。ニューラルネットワークは4種類の移動状況の確信度を0～1の範囲で出力する。

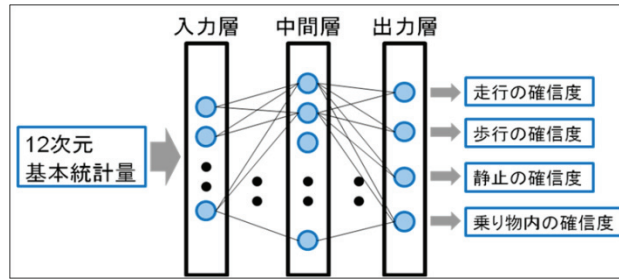


図 4-6 ニューラルネットワークの構成

### 遷移確率モデルによる補正

上述の通り、各移動状況は排他的なもののみとするため、「乗車中」の「歩行」や「静止」はすべて「乗車中」として扱う。単純にセンサの挙動から判定すると、電車やバスが一時停止した場合に「静止」と判定される可能性が高い。そこで、各移動状況の遷移確率に基づく補正を行う。例えば、乗り物から降りた直後には、通常、歩行が行われることが多い。そこで、ニューラルネットワークの出力が「乗車中」から「歩行」に変化した場合には即座に「歩行」とするが、「乗車中」から「静止」に変化した場合には即座に「静止」とせず、所定時間の「静止」の継続を確認してから遷移するような処理を行う。このような各移動状況間の変化しやすさの度合いを遷移確率モデルで表現し、ニューラルネットワークの出力に対して補正を行い、推定精度の向上を試みている。さらに、GPSと連携することで、推定性能を維持した間欠動作による省電力動作も可能である<sup>[55]</sup>。

## 4. 3. 6 評価実験 1 (基本性能評価)

### 基本性能評価用システム

4. 3. 4で述べた屋内生活行動推定エンジンの妥当性を評価するためには、実際の生活状態の加速度と音を同期して記録し、分析する必要があるが、スマートフォンでの状況認識の実現 (Android アプリ開発) に先立ち、基礎的なデータ収集を効率的に行うことを目的に、3軸加速度センサとマイクを搭載し、両センサのデータを microSD カードに同期記録できる評価用デバイスを試作した。デバイスの外観を図 4-7 に、概略仕様を表 4-1 に示す。

データ収集時には、本デバイス単体で動作して microSD カードに加速度センサとマイクのデータを同期記録する。また、Bluetooth モジュールを搭載しており、

PC などの外部機器に加速度データは SPP (Serial Port Profile) で、音データは A2DP (Advanced Audio Distribution Profile) で PC に送信し、PC で一連の処理を実行してリアルタイムに生活行動を認識するアプリを開発した。アプリ画面の一例を図 4-8 に示す。



図 4-7 基本性能評価用デバイス

表 4-1 基本性能評価用デバイスの概略仕様

3 軸加速度センサ	LIS344ALH (STMicroelectronics) ±2G/6G 10bit 20Hz サンプリング
MEMS マイク	SPM0408HE5H (Knowles) 8/16bit 8/16/32kHz サンプリング
記録メディア	microSD カード
オーディオ/ファイル制御用 CPU	ARM Cortex-M3
システム制御用 CPU	PIC18LF4685 (Microchip)
オーディオ LSI	ML2308 (OKI)
通信機能	Bluetooth™ SPP / A2DP

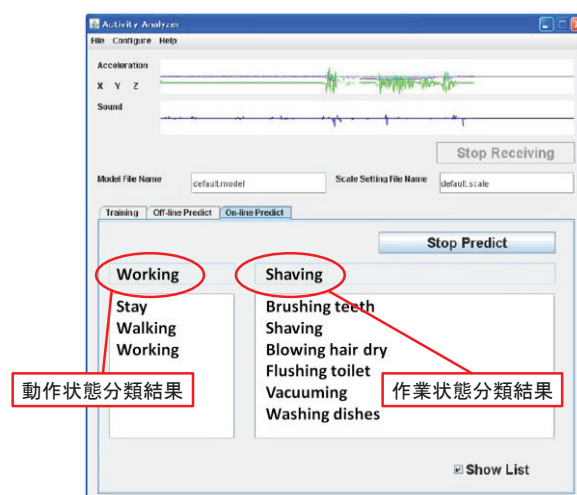


図 4-8 基本性能評価用 PC アプリ

## データ収集条件

まず、毎秒実施する動作状態分類および作業状態分類の基本性能を確認するデータ収集実験を実施した。試作した評価用デバイスを胸ポケットに入れ、一般の家庭における日常生活のデータ収集を実施した。被験者 4 名（60 代男女、30 代男女、それぞれ 1 名ずつ）に対して行動リスト（被験者毎に順番を入れ替えた）を提示し、それに従って一通り行動してもらった実験を、日を変えて 2 日分実施し、被験者 1 名あたり 2 データセット、計 8 データセットを収集した。本実験は、提案手法の基本性能を確認することを主眼としているため、被験者数は 4 名（年代の違いによる影響も確認するため、60 代と 30 代の被験者で実施した）と少ないが、10 秒間の学習データによる作業状態分類性能について、他人のデータを学習した場合の性能と、本人のデータを学習した場合の基本性能の違いを確認する目的で実施した。

分類対象とした生活状態と、それぞれの平均収集時間は表 4-2 の通りである。連続して収集したデータから、対象の生活状態に該当する部分のデータを切り出し、分析対象とした。尚、「歯磨き」は電動歯ブラシではなく一般の歯ブラシによる歯磨きのデータを収集した。「電気シェーバーによる髭剃り」は男性被験者 2 名のみ実施した。また、排泄中は「安静」と判別される可能性が高く、排泄後にトイレの水洗と手洗いをする際に「作業」となることが予想されることから、今回のデータ収集では、被験者に便座に座ってもらい、そこから立ち上がってトイレを水洗し、手を洗うように指示し、その際のデータを収集した。

4. 3. 2 で述べた通り、本研究では、できるだけ短時間の学習データで、精度良い生活行動認識を実現する点を目的としている。よって、「作業」の 7 種類の分類性能の評価方法は、収集した全データセットのうち 1 データセットをテスト用とし、残りを学習用として、全データセットが 1 回ずつテスト用となるように評価を実施する **Leave-one-out Cross-validation** とは異なる評価方法を用いることとした。つまり、収集した全 8 データセットの生活状態データのうち、1 データセットから各生活状態についてそれぞれ無作為に連続した 10 秒分のデータを学習用として抽出し、それ以外の 7 データセットの各生活状態のデータをテスト用として分類する。次に別の 1 データセットから再度、無作為に連続した 10 秒分を新規の学習用データとして抽出して同様のテストを行う。これを全データセットから 1 回ずつ学習用データを抽出するようにして評価を行った。

最適な学習時間については、実運用時の観点も含めた検討が今後必要であるが、我々が想定するサービスイメージのためには、極力短い時間で学習できるようにすべきであると考え、それぞれ 10 秒分ずつという少ない学習データで分類を試みた。分類器に汎化性能の高い SVM を選択したのはこのためでもある。

表 4-2 データ収集した生活状態とその平均時間

生活状態		平均時間（分）
歩行（階段昇降も含む）		1.15
作業	歯磨き	1.14
	電気シェーバーによる髭剃り	0.88
	ドライヤーの使用	1.48
	トイレ水洗／手洗い	0.45
	掃除機がけ	3.18
	皿洗い	1.55
	アイロンがけ	1.67
安静（TV鑑賞）		3.24

## 結果

### 加速度による動作状態分類性能

加速度による動作状態分類（「歩行」、「作業」、「安静」のいずれかに分類する）の結果を表 4-3 に示す。数値は、認識結果のうち、正しくその動作状態（「歩行」、「作業」、「安静」）に分類されたデータの割合、すなわち適合率を示す。「歯磨き」、「電気シェーバーによる髭剃り」など作業状態に含まれる各生活状態は、ここでは「作業」に分類されることが正解となる。過去 1 秒間の分散を 1 サンプルング（50ms）ごとに算出し、その都度動作状態分類を実施した。加速度の 1 秒間の分散のみを用いた分類法で、概ね 95%以上の精度で「歩行」、「作業」、「安静」を分類できていることがわかる。尚、今回は 4 名の被験者全員で 4. 3. 4 に示した同一の閾値を適用して表 4-3 の結果を得たが、対象者の身体状態（年齢、障がいの有無）等によっては、学習時に適切な閾値設定が必要になる可能性がある。ただし本実験では、30 代被験者、60 代被験者で年代の違いによる性能差は確認されなかった。

表 4-3 加速度による動作状態分類性能

生活状態		動作状態分類性能（%）	
歩行（階段昇降も含む）		95.4	
作業	歯磨き	98.6	97.4
	電気シェーバーによる髭剃り	99.5	
	ドライヤーの使用	92.9	
	トイレ水洗／手洗い	95.5	
	掃除機がけ	97.6	
	皿洗い	98.5	
	アイロンがけ	99.5	
安静（TV鑑賞）		97.5	

「掃除機がけ」の際には掃除機をかけながら移動する際の歩行も発生するが、掃除機をかけながらの歩行の大半は「作業」と分類された。本手法で、掃除中において姿勢をかがめたりしながらの移動は、部屋移動のための歩行と区別できることがわかった。

## 音による作業状態分類性能

加速度による動作状態分類で「作業」と分類された際に実行する、音による作業状態分類の結果を表 4-4 に示す。MFCC、RMS、ZCR の各特徴量を 1 秒ごとに算出し、それを用いて 1 秒単位で作業状態分類を実施し、正しくその作業状態に分類されたデータの割合（適合率）を示す。作業状態ごとに多少の性能のバラツキはあるが、任意の 10 秒間の学習データで平均 75.8%の分類性能があることが確認できた。

表 4-4 音による作業状態分類性能（全被験者データによる評価）

生活状態（作業）	作業状態分類性能（%）
歯磨き	68.3
電気シェーバーによる髭剃り	91.4
ドライヤーの使用	91.6
トイレ水洗／手洗い	71.7
掃除機がけ	70.2
皿洗い	67.2
アイロンがけ	70.3
平均	<b>75.8</b>

「電気シェーバーによる髭剃り」、「ドライヤーの使用」については、90%以上の精度で分類できているが、その他の生活状態については、いずれも 70%程度の精度であった。原因としては、「歯磨き」については、電動ではない歯ブラシによる歯磨きのデータを収集したが、歯磨きで発生する音の特徴が被験者によって異なるためであると考えられる。「トイレ水洗／手洗い」、「皿洗い」は、それぞれ水洗時の水量を指定せずに被験者任せとしたため、被験者によってバラツキが出たと考えられる。「掃除機がけ」についても同様に動作モードは被験者任せとした。極力自然な生活状態のデータを収集することを主眼に置いて実施したことによる影響ではあるが、逆に言えば、水量や、動作モードが異なるデータを学習しても、約 70%程度の精度で同一の作業状態として分類できることがわかった。

しかし、本研究では、対象者本人のデータを学習データとして用いることを当面のターゲットとして考えているため、今回収集した各被験者の 2 データセットのうち、1 データセットから任意の 10 秒間を抽出して学習データとし、別の 1 データセットをテスト用として評価し、次に逆の組み合わせで同様のテストを実施した場合とあわせて分類性能を評価した。表 4-5 にその結果（全被験者の平均）



を示す。本人の学習データを用いることによって、全体として約 10 ポイント、性能が向上することがわかった。想定したサービスイメージに対する手法として、提案手法が有用であることが確認できた。また、30 代被験者、60 代被験者で年代の違いによる性能差は確認されなかった。

目標とすべき性能は、想定するサービスによって異なってくるが、上述の結果から示唆されることは、サービス提供側が平均 75%程度の性能で十分と判断する用途であれば、汎用の学習データを用いて個人毎の事前学習は不要のサービスを提供できるが、それ以上の性能を必要とする用途に対しては、事前学習を行うことにより性能を向上することが可能であるということである。

表 4-5 音による作業状態分類性能（本人データの学習による評価）

生活状態（作業）	作業状態分類性能（%）
歯磨き	73.0
電気シェーバーによる髭剃り	91.9
ドライヤーの使用	96.1
トイレ水洗／手洗い	94.5
掃除機がけ	87.0
皿洗い	81.9
アイロンがけ	77.2
平均	85.9

ただ、「歯磨き」については、本人のデータを学習した場合でも 73.0%の精度であった。精度が上がらない原因を調べたところ、任意に選んだ学習データ 10 セットが、前歯を磨いている時のデータのみだった場合、奥歯を磨いている際の性能が著しく低下していたことがわかった（逆の場合や、口の開き具合の違いの影響なども考えられる）。

前歯磨きから奥歯磨きに変わった時点の検出結果の一例を図 7 に示す。前歯磨きの間は概ね精度良く歯磨きと分類できているが、奥歯磨きになった時点から誤検出の割合が増加している。前歯磨きと奥歯磨きで、その時に発生している音の特徴が異なることが原因である。そこで、前歯磨きの時の 5 秒間と奥歯磨きの 5 秒間を手作業で抽出した合計 10 秒間のデータを歯磨きの学習データとして再学習させたところ、73.0%だった分類性能を 80.6%に改善できることが確認できた。

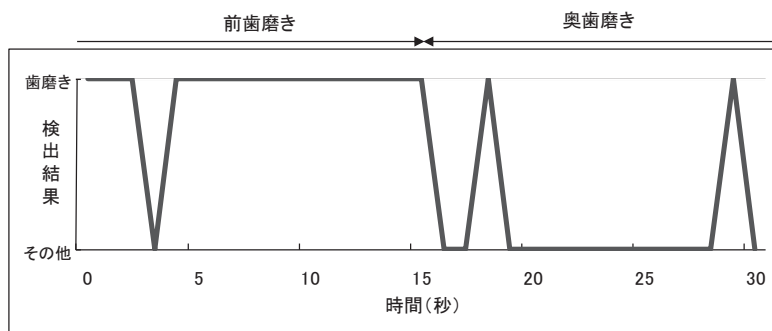


図 4-9 歯磨き時の分類結果の一例

このように、同一の作業状態において、途中で音の特徴が変化するような場合は、それぞれの場合のデータを学習データに含めるようにする、あるいは SVM にそれぞれ別のクラスとして分類させるなどの工夫が必要になることがわかった。ここで、1 データセットの歯磨きのデータ全部を学習データとして再学習させたところ、分類性能は 81.9% となった。このことから、全データの中から音の特徴をバランス良く抽出することで、10 秒間だけのデータでも、全データを学習させた場合と同等レベルの性能を示すことができることが示唆された。最適な学習区間の抽出方法については、今後検討していく。

## 4. 3. 7 評価実験 2 (実用性能評価)

### データ収集条件

上述の評価実験 1 で提案手法の基本性能の妥当性を確認できた。これを受け、次に事例ベース認識手法による実用性能を確認するデータ収集実験を行った。本実験では、提案手法のターゲットである高齢者向けのサービスを想定し、60 代の被験者 10 名 (60 代男女、それぞれ 5 名ずつ) に対してデータを収集した。評価実験 1 と同様に、評価用デバイスを胸ポケットに入れ、一般家庭のリビングを模した実験室において日常生活のデータ収集を実施した。

まず「皿洗い」、「掃除機がけ」、「アイロンがけ」、「トイレ水洗/手洗い」、「歯磨き」の各作業を 10 秒ずつ行ってもらい、これを学習用データとし、続いて上記の「作業」を順番に任意時間ずつ実行してもらおうこととし、「作業」と次の「作業」の間は 1 分間以上、実験者との会話や、ストレッチなど、自由に過ごしてもらおうこととした。このデータは分類対象外 (未学習) の「作業」となる。尚、本実験で実施した 5 種類の日常生活における作業は、リビングを模した実験室内で実施可能なものがこの 5 作業であったため選択した。このようにして、評価用デ

ータとして、被験者ごとに5種類の「作業」と「作業」間の5データを、それぞれ合計50データセット収集した。

## 結果

全テストデータの分類結果を表4-6の confusion matrix に示す。5つの学習済「作業」の分類結果は98% (49/50) の精度で正しく分類され、対象外「作業」は82% (41/50) の精度で正しく分類された。事例ベース認識手法の導入により、「作業」時に毎秒実施する作業状態分類そのものの性能を凌駕する実用性能を確認できた。しかし、対象外「作業」の9事例（身振り手振りを交えた立ち話3事例、会話しながらストレッチ4事例、ストレッチのみ2事例）については、本来分類対象外（未学習）と分類すべき所を、「皿洗い」、「アイロンがけ」、「トイレ水洗／手洗い」と誤認識している。認知症予兆の早期発見等を期待した高齢者の見守りサービスを想定すると、実際には対象外の行動をしているにも関わらず、何らかの見守り対象の行動をしているものと判断してしまうため、この種の誤認識は対象者の生活レベルの正確な把握に影響を及ぼす可能性がある。

この問題の解決には、基本的な分類性能の向上が期待されるが、学習済「作業」の判断のための、その区間で最も出現回数の多い「作業」状態の出現頻度の閾値Tを50%よりも高く設定し、50%以上T未満の場合は対象者に問い合わせる機能などによって正解率を向上させる工夫などが考えられる。高齢者であれば、「はい」「いいえ」で回答できる音声対話などが適していると考えられるが、対象者に負担をかけない提示方法が必要であり、ロボットインタフェースとの連携可能性を検討した内容については後述する。また、この際に新しく取得したデータで再学習することで、認識性能が向上する可能性がある。

表 4-6 事例ベース認識手法による分類結果の confusion matrix

	学習済「作業」					対象外「作業」
	皿洗い	掃除機がけ	アイロンがけ	トイレ水洗	歯磨き	
皿洗い	10	0	0	0	0	0
掃除機がけ	0	10	0	0	0	0
アイロンがけ	0	0	10	0	0	0
トイレ水洗／手洗い	0	0	0	10	0	0
歯磨き	0	0	0	0	9	1
対象外「作業」	3	0	4	2	0	41

## 考察

提案した加速度と音を組み合わせた家庭内生活行動況認識手法が有用であるかは、従来の加速度のみ、あるいは音のみでの行動認識手法と比較して検証する必要がある。文献[39]-[41]などの関連研究で単一の加速度センサで認識対象とする動作は、現状、歩行、走行など、移動を中心とした数種類の動作状態だけで、本研究で対象とするその他の作業状態を認識することは困難であるため、加速度センサだけによる作業状態分類についてはここでは比較の対象としない。従って、ここでは音だけで同様の生活状態分類を行う場合との比較検証を行う。

提案手法では、加速度センサで「作業」と分類した動作状態のみ、音によりさらに細かく作業状態分類を行っている。作業状態の分類は音だけを用いているので、比較すべきは、加速度センサで分類している「歩行」、「安静」となる。つまり「歩行」、「安静」の分類を音で行った時と、加速度で行った時とを比較すればよい。尚、加速度で「歩行」、「安静」の認識を行っている従来研究はあるが、装着部位、評価環境（屋内、屋外の違い）など条件が異なるため、同一条件による評価として、音による分類性能を提案手法の加速度による「歩行」、「安静」の分類性能と比較する。

提案手法における音による作業状態分類法と同じ方法で「歩行」「安静」を認識できるか確認した。一般家庭の廊下を歩行している時（歩行）と、TVを鑑賞している時（安静）の音データについて、任意の10秒分を切り出し、同一の音特微量（15次元）データをそれぞれ「歩行」「安静」クラスとして他の作業状態の各クラス（「皿洗い」、「掃除機がけ」、「アイロンがけ」、「トイレ水洗／手洗い」、「歯磨き」）に加えて学習させた。テストデータは学習データと同じ場所（一般家庭の廊下）を歩行した際、および同一のTV番組（スポーツ中継）を視聴している際の音データ（歩行：1分、安静：3分）とした。場所が異なったり、番組が異なったりすると、環境音の変化により、著しく性能が劣化することが考えられるため、今回の比較では同じ場所の歩行、同一TV番組のデータを使うことにした。加速度による「歩行」認識、「安静」認識との比較を表4-7、表4-8に示す。また、加速度の場合も、音の場合も、分類の際に処理する生データは1秒間分であるため、それぞれ処理対象である1秒間の生データのサイズも算出した。

音だけを使用して86.7%の精度で「歩行」を分類でき、88.8%の精度で「安静（TV鑑賞）」を分類できることが確認できたが、加速度だけによる分類性能よりはそれぞれ劣る。また、分類のために必要な1秒間のデータサイズは、加速度の場合は75バイトであるのに対し、音の場合は32,000バイトと3桁大きいデータの処理が必要となる。このことから、「歩行」、「安静」を認識するためには、性能とデータ量（計算量）双方の観点から、加速度による動作状態分類を行うことが望ましいと言える。

また、「安静」時にユーザ自身から発せられる音は、呼吸音など非常に小さいレベルの音のみで、周囲の環境音の影響を大きく受けることが想定され、今回は同一の TV 番組を視聴中のデータで評価したが、(番組内容に依存しない) TV 視聴、読書、音楽鑑賞、睡眠など、音のみによる複数の「安静」検出はさらに困難であると考えられる。

まとめると、単一の加速度センサのみでは本論文が対象とする様々な作業状態を認識することは従来研究から困難であると言えるが、作業状態の推定にマイクからの音を用いることで、様々な作業状態の推定が可能となった。提案手法で、加速度のみで分類している「歩行」、「安静」について、これらもすべて音により分類した場合の分類性能、および扱うデータ量の比較を行い、「歩行」、「安静」ともに音を用いるよりも加速度を用いた方が、性能、データ量とも有利であることを確認した。提案手法によるデータ量(計算量、さらには消費電力に関する)の削減効果は「作業」の発生頻度に依存するが、次に、スマートフォン上に本手法の行動認識エンジンを搭載し、より実用に近い評価を実施して検証する。

以上の観点から、加速度による動作状態分類で「歩行」、「作業」、「安静」を分類し、「作業」の場合に環境音分析を行って詳細な作業状態分類を行う提案手法の妥当性を確認することができた。

## **4. 3. 8 スマートフォン版アプリの開発**

上述の通り、提案手法の妥当性が確認できたことを受け、屋内生活行動推定エンジンと屋外移動状況推定エンジンを統合したスマートフォン(Android)版アプリを開発し、図 4-1 に示す屋内外生活行動センシングシステムを開発した。スマートフォン上の推定結果表示画面例(屋内、屋外)を図 4-10 に、サーバ側の画面表示例を図 4-11 に、さらに生体モニタリング結果との連携表示例(連続血圧モニタリングと行動モニタリング)を図 4-12 に、それぞれ示す。

図 4-11 のように、複数日の行動を比較表示することで、曜日毎の行動の傾向や、普段と違った行動への気づきが可能となる。また図 4-12 のように、他の生体モニタリング結果と連携して表示を行うことで、異常データがあった際にどのような行動をしていたのかなどの振り返りが可能になる。例えば、寒い日の朝、トイレで血圧が急上昇していることがわかった場合には、トイレを暖房することによりリスクを軽減する対策が可能になる。

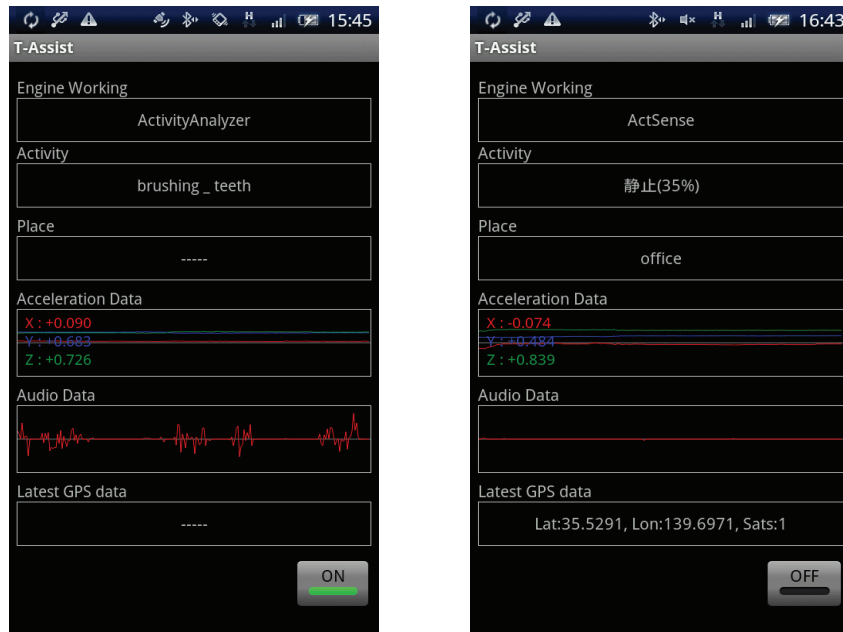


図 4-10 スマートフォン上の推定結果表示例（左：屋外，右：屋内）

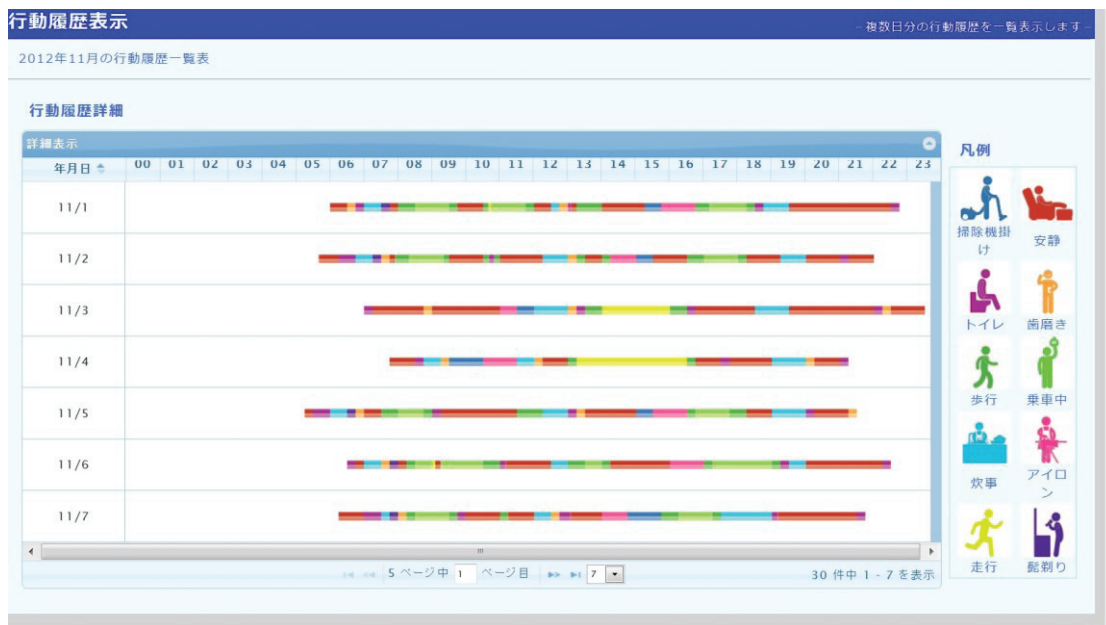


図 4-11 サーバ側結果表示例

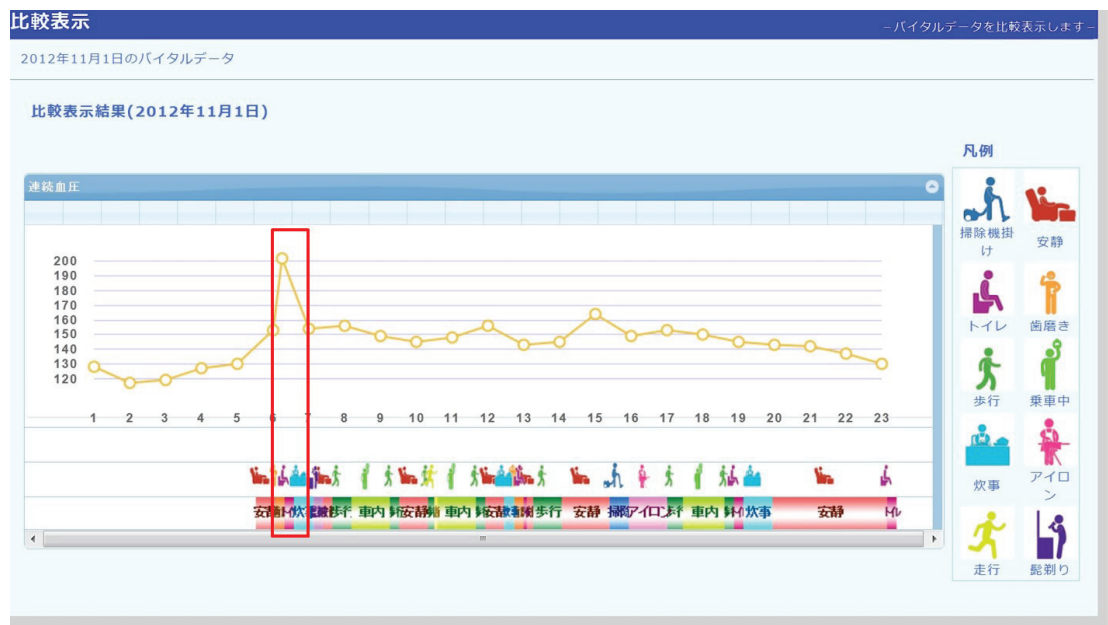


図 4-12 サーバ側結果表示例（連続血圧モニタリングとの比較表示）

## 4. 3. 9 スマートフォン版アプリによる評価

### データ収集条件

一般のリビングを模した実験室環境において、高齢者（60歳以上）12名（男性6名、女性6名）、一般成人（20～40歳代）9名（男性5名、女性4名）の計21名の被験者に、上着の胸ポケットに本アプリを起動したスマートフォンを携帯してもらい、作業リストを提示してそれに従って順に行動してもらった。実験の様子を図 4-10 に示す。



図 4-10 屋内生活行動評価実験の様子

「静止」「歩行」に加え、「作業」として「皿洗い」「アイロンがけ」「掃除機がけ」「歯磨き」「ドライヤー」「トイレ水洗／手洗い」の6作業を実施した。また、対象行動間のインターバル時の被験者の任意行動（実験者との立ち話、ストレッチなどの際に「作業」と判断された行動）については、認識対象外作業として扱った。

## 結果

加速度センサを用いた動作推定性能を表4-7に示す。「皿洗い」「アイロンがけ」「掃除機がけ」「歯磨き」「ドライヤー」「トイレ水洗／手洗い」の「作業」の各行動の場合は、ここでは「作業」と分類されるのが正解である。加速度によるこれらの動作推定はいずれも評価実験1と同様に95%以上の精度であることが確認できた。

表 4-7 加速度による動作推定性能

行動	分類性能(%)
歩行	95.2
作業	96.8
静止	96.5

次に、「作業」と分類された場合に実施する音による作業内容推定性能を表4-8に示す。ここでの性能は、事前に被験者に各作業時のデータを10秒間収集し、そのデータで作成したモデルを用いて分類した性能で、作業区間単位での補正を行う前の毎秒実施する認識処理の基本性能である。評価実験1の結果（表4-5）とほぼ同等レベルの結果（平均84.5%）が得られた。

表 4-8 音による作業内容推定性能（補正前基本性能）

生活行動	分類性能(%)
皿洗い	80.2
アイロンがけ	76.7
掃除機がけ	86.3
歯磨き	74.1
ドライヤー	96.4
トイレ水洗／手洗い	93.2

つづいて、連続して切り出された一つの作業区間は同一の作業が継続しているものとし、切り出した作業区間全体に対して作業内容を再推定する事例ベース認識手法による補正を実施した結果の confusion matrix を表4-9に示す。



表 4-9 補正処理後の confusion matrix

	認識対象作業					認識対象外作業	認識性能 (%)
	皿洗い	アイロンがけ	掃除機がけ	歯磨き	ドライヤー		
皿洗い	37			3		1	88.1
アイロンがけ		36		4			85.7
掃除機がけ			42				100
歯磨き		3		36			85.7
ドライヤー					42		100
トイレ水洗／手洗い						41	97.6
認識対象外作業	3	7		4		4	82.9

平均 91.4%と 90%以上の精度で各作業を認識できていることが確認できた。評価実験 2 の結果（表 4-6）に比べて性能が低下しているのは、認識対象作業の数が増えた（5→6）こと、被験者数が増えたこと（10→21）、より日常生活に近いプロトコルで実験を実施したことなどが挙げられるが、本人の短時間のデータを学習データとして用いることで、高い精度で多様な生活行動を認識可能であることを確認した。

### 4. 3. 10 装着部位の違いによる評価

これまでの評価実験はすべて装置（評価用デバイス、スマートフォン）を胸ポケットに格納して実施したが、スマートフォンの場合は胸ポケット以外にズボンのポケットなどにも格納されることが多い。また、スマートフォンの機能の一部を代替するスマートウォッチや、活動量計などの腕時計型デバイスにも今後マイクが内蔵される可能性がある。よって、胸ポケット以外に、ズボンのポケットと手首にそれぞれスマートフォンを同時に装着し、装着部位の違いによる性能の評価を行った。

#### データ収集条件

これまでの実験と同様に、一般のリビングを模した実験室環境において、高齢者（60歳以上）12名（男性6名、女性6名）、一般成人（20～40歳代）10名（男性5名、女性5名）の計22名の被験者に、図4-11のように、本アプリを起動したスマートフォンを、胸ポケット、ズボンポケット、および手首に装着してもら

い、まず、事前に各作業を 10 秒間実施してもらい、各部位に装着したスマートフォン毎にモデルを作成した上で、作業リストを提示してそれに従って順に行動してもらい、スマートフォン毎の認識性能を評価した。



図 4-11 装着部位の違いによる性能評価

## 結果

装着部位毎に「作業」と分類された場合に実施する音による作業内容推定性能を表 4-10 に示す。ここでは、作業区間単位での補正を行う前の毎秒実施する認識処理の基本性能である。

表 4-10 各部位における作業内容推定性能（補正前基本性能）

胸ポケット		ズボンポケット		手首	
Living activities	Accuracy (%)	Living activities	Accuracy (%)	Living activities	Accuracy (%)
Washing dishes	80.8	Washing dishes	79.2	Washing dishes	82.1
Ironing	75.3	Ironing	72.1	Ironing	73.8
Vacuuming	90.6	Vacuuming	91.2	Vacuuming	89.3
Brushing teeth	77.4	Brushing teeth	69.1	Brushing teeth	70.0
Shaving	89.2	Shaving	83.4	Shaving	82.8
Hair drier	95.6	Hair drier	95.1	Hair drier	92.6
Flushing the toilet	93.4	Flushing the toilet	91.5	Flushing the toilet	90.3
Talking	70.1	Talking	63.8	Talking	65.4
<b>Average</b>	<b>84.1</b>	<b>Average</b>	<b>80.7</b>	<b>Average</b>	<b>80.8</b>

胸ポケットに比べ、ズボンポケット、手首における性能は若干低いことを確認した。次に、各部位における補正処理後の confusion matrix を表 4-11～4-13 に、各部位の F 値を図 4-12 に示す。

表 4-11 補正処理後の confusion matrix (胸ポケット)

		Trained tasks							Untrained task	Recall (%)	
		Washing dishes	Ironing	Vacuuming	Brushing teeth	Shaving	Drying hair	Flushing the toilet			Talking
Trained tasks	Washing dishes	20			1					1	90.9
	Ironing		19						1	2	86.4
	Vacuuming			22							100.0
	Brushing teeth				19				1	2	86.4
	Shaving					20				2	90.9
	Drying hair			1			21				95.5
	Flushing the toilet							21		1	95.5
	Talking		1						18	3	81.8
Precision (%)		100.0	95.0	95.7	95.0	100.0	100.0	100.0	90.0	F-measure 93.8	

表 4-12 補正処理後の confusion matrix (ズボンポケット)

		Trained tasks							Untrained task	Recall (%)	
		Washing dishes	Ironing	Vacuuming	Brushing teeth	Shaving	Drying hair	Flushing the toilet			Talking
Trained tasks	Washing dishes	19			2					1	86.4
	Ironing		19		1				1	1	86.4
	Vacuuming			22							100.0
	Brushing teeth	1			17				2	2	77.3
	Shaving		1			18			1	2	81.8
	Drying hair			1			21				95.5
	Flushing the toilet							21		1	95.5
	Talking		2		1				16	3	72.7
Precision (%)		95.0	86.4	95.7	81.0	100.0	100.0	100.0	80.0	F-measure 89.5	

表 4-13 補正処理後の confusion matrix (手首)

		Trained tasks							Untrained task	Recall (%)	
		Washing dishes	Ironing	Vacuuming	Brushing teeth	Shaving	Drying hair	Flushing the toilet			Talking
Trained tasks	Washing dishes	19			1				1	1	86.4
	Ironing		18						2	2	81.8
	Vacuuming			22							100.0
	Brushing teeth	1			18				1	2	81.8
	Shaving					18			1	3	81.8
	Drying hair			1			21				95.5
	Flushing the toilet							22		0	100.0
	Talking		2						17	3	77.3
Precision (%)		95.0	90.0	95.7	94.7	100.0	100.0	100.0	77.3	F-measure	91.0



図 4-12 装着部位毎の認識性能 (F 値)

歯磨き、ひげ剃りなど、口元で比較的小さな音が発生する行動については、音源からの距離が遠くなるにつれて認識性能が低下するため、全体として胸ポケットが最も性能が良好であった。しかしながら、他の部位でも概ね 90%程度の認識性能が実現できており、それぞれの部位で取得したデータで事前に学習することで、さまざまな生活行動を精度良く認識可能なことが確認できた。

## 4. 4 ロボットインタフェースとの連携

上述の通り、事例ベース認識手法の導入により精度の良い生活行動認識が実現できたが、認識対象外作業であるにも関わらず認識対象の作業であるとしてしまう誤認識（過検出）は、認知症予兆の早期発見等を期待した高齢者の見守りサービスを想定すると、実際には対象外の行動をしているにも関わらず、何らかの見守り対象の行動をしているものと判断してしまうため、この種の誤認識は対象者の生活レベルの正確な把握に影響を及ぼす可能性がある。

この問題の解決には、本質的には基本的性能の更なる向上が必要であるが、学習済「作業」の判断のための、その区間で最も出現回数の多い「作業」状態の出現頻度の閾値  $T$  を 50% よりも高く設定し、50% 以上  $T$  未満の場合は、高齢者向けのロボットインタフェースを利用して対象者に問い合わせる機能などによって正解率を向上させる工夫などが考えられる。

そこで、4. 3. 7 評価実験 2（実用性能評価）の表 4-6 の結果について、最多の認識結果行動の出現頻度  $M$  に対して 2 種類の閾値を設定し、ロボットインタフェースと連携することの有用性について考察する。従来の事例ベース認識手法(1)の閾値は 50% としたが、事例ベース認識手法(2)として、70% を第一の閾値、50% を第二の閾値とし、認識結果  $A$  は第一の閾値である 70% 以上の事例、認識結果  $B$  は第二の閾値（50%）以上、第一の閾値（70%）未満の事例とする(2)。図 4-13 にそれぞれの処理フローを示す。事例ベース認識手法(2)により表 4-6 の結果を再評価すると表 4-14 の通りとなる。

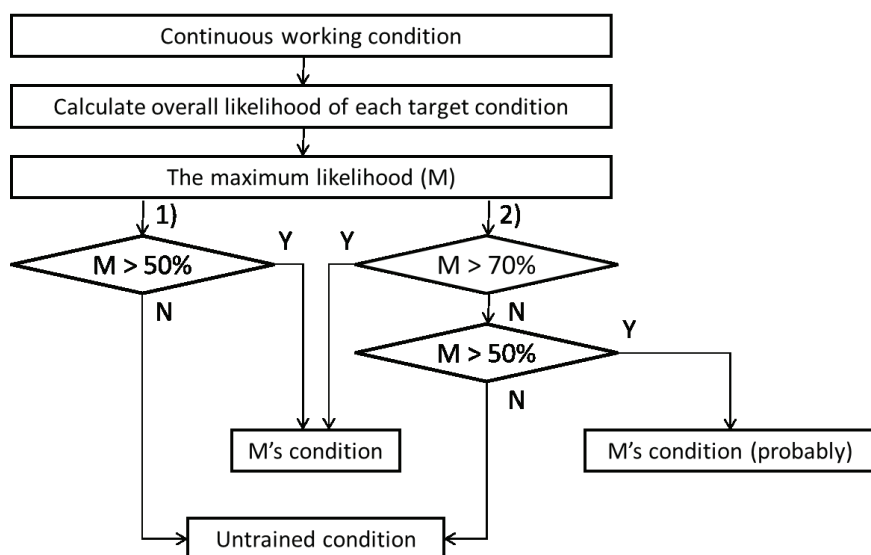


図 4-13 2 通りの事例ベース認識手法

表 4-14 2つの閾値を設定した場合の confusion matrix

	trained task										untrained task
	washing dishes		vacuuming		ironing		flushing toilet		brushing teeth		
	A	B	A	B	A	B	A	B	A	B	
washing dishes	3	7	0	0	0	0	0	0	0	0	0
vacuuming	0	0	9	1	0	0	0	0	0	0	0
ironing	0	0	0	0	5	5	0	0	0	0	0
flushing toilet	0	0	0	0	0	0	2	8	0	0	0
brushing teeth	0	0	0	0	0	0	0	0	7	2	1
untrained task	0	3	0	0	0	4	0	2	0	0	41

Aの結果については、尤度が十分に高くそのまま認識結果を正解であるとし、Bの結果については、誤認識の可能性があるためユーザにその結果が正しいか確認することで認識性能を向上させることを検討する。ここで、本研究の対象は高齢者であるため、確認のための手段がスマートフォンやPCではユーザへの負担が大きい。そこで例えば Kobayashi<sup>[56]</sup>が開発したユーザと自然な対話が可能なロボットインタフェースと連携させることで、高齢者でも負担無く（意識せずに）正しい認識結果に修正できる可能性がある。

例えば、Bと判断した場合の直後にロボットインタフェースを介して、直前にその行動をしたかどうか、さりげなく、タイムリーに問いかけ、それに対するユーザの回答によって最終的な認識結果を確定させる。問いかけの例を図 4-14 に示す。



図 4-14 ロボットインタフェースと連携した性能向上

このように、極力「はい」「いいえ」など簡単な応答で返せる問いかけを行うことで、ユーザは負担無く回答できると共に、音声認識の認識率も向上させることが期待される。仮に音声認識精度が 100%であると仮定すると表 4-14 の B で発生した誤認識 (untrained task で washing dishes, ironing, flushing toilet と判断した 9 事例) はすべて正しく修正されるため、全 100 事例のうち誤認識は brushing teeth を untrained task を判断した 1 事例のみとなり認識率 99%となる。また、この際に新しく取得したデータで再学習することで、使いながら認識性能を向上させる機能を実現できる可能性がある。今後、対話機能を持つロボットの普及と合わせて、装着型デバイス、環境側のセンサ等による状況認識と、ロボットインタフェースの連携による認識性能とサービスの付加価値向上が期待される。

## 4. 5 まとめ

本章では、環境側へ機器を設置せず、スマートフォンにすでに搭載されている加速度センサとマイクを活用した生活行動見守りのための状況認識技術として屋内生活行動推定手法を提案した。まず加速度センサでユーザの動作状態を大まかに推定し、推定した状態に応じてマイクを起動して環境音の分析を行うことで、計算量を抑えた屋内生活行動認識を実現した点が特長である。

また、屋内生活行動推定エンジンと屋外移動状況推定エンジンを一つの Android アプリ上に統合し、スマートフォン単体で屋内外の様々な生活行動をセンシング可能な、屋内外生活行動センシングシステムを開発した。GPS 衛星の捕捉状況に基づいた屋内外判定機能により屋内外で使用するエンジンを自動的に切り替え、シームレスな生活行動センシングを可能にした。また、無線通信によりサーバや外部端末へ認識結果を送信する機能も搭載し、様々な実用サービスへの適用が期待できる構成とした。

胸ポケットに格納した試作デバイス／スマートフォンで実際の家庭生活のデータを収集し、各作業状態につき任意の 10 秒間ずつの音特徴量データを SVM に学習させて評価した結果、加速度センサのみで「歩行」、「作業」、「安静」の 3 状態を概ね 95%以上の精度で分類でき、「作業」時には環境音の分析を行うことで、「歯磨き」、「電気シェーバーによる髭剃り」、「ドライヤーの使用」、「トイレ水洗／手洗い」、「掃除機がけ」、「皿洗い」、「アイロンがけ」など様々な家庭内の生活行動を高精度に分類できた。このことにより、本研究で想定する家庭内の生活状態見守りに活用可能な基本性能を持つことを確認した。また、事例ベース認識手法の導入により、実用的な性能が見込めることも確認した。尚、実際に認識が必要な作業の粒度、種類は、想定するサービス、ターゲットユーザによって異なっ

くるため、このあたりを具体化した評価は今後、各用途を想定した実用システムの開発において実施する必要がある。

分類器に SVM を用いることで、各生活状態につき任意の 10 秒間という少ない学習データでも良好な分類性能と応答性が得られたが、目的とする生活状態によっては学習させるデータの区間を適切に選択する必要がある。その区間の効果的な選び方については今後検討が必要である。また、運用前の事前学習が必要な前提で議論してきたが、大規模なデータ収集を実施することによって各作業状態の標準的な学習データを作成することにより、事前学習を不要にすることができる可能性がある。また、運用開始後の自動的な再学習の枠組みの検討も期待される。

また、現在はスマートフォンを常に携帯し続ける前提で検討しているが、家庭内ではスマートフォンを携帯していない場合も多く、その際にも継続してセンシングするための検討（置いたスマートフォンの活用や、他のセンサデバイスとの連携など）も今後の実用化検討においては議論が必要である。



# 5. センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用

---

## 5. 1 はじめに

音声認識入力とは、自動車運転中など両手が使えない状況での機器操作<sup>[57]</sup>や、パソコンを用いた特定業務での作業効率向上<sup>[61][58]</sup>などで活用されている。

一方、機能数の増加、ネットワーク化などにより、日常的に扱う一般的な家電機器の操作が複雑になってきている。特に高齢者など、機器操作を得意としないユーザ層にとってその影響は大きい<sup>[59][60]</sup>。

また、情報家電はその役割が変わりつつある。例えば、地上波デジタル放送への全面移行、インターネット接続機能や録画機能を搭載したデジタルテレビの普及などに伴い、テレビの役割は、従来の映像コンテンツ表示装置としての役割だけでなく、インターネットの検索も行うリビングでの情報端末としての役割が増えていくと考えられる。目的の情報／コンテンツをうまく探し出すための検索機能の重要性が増すほど、直接文字を入力する機会が増える。従来のテレビ画面上のスクリーンキーボードから、十字ボタンでカーソルを移動して文字を選択して入力する、あるいは携帯電話方式のキー操作で一文字ずつ入力するという方式では、不十分であるし、大変使い勝手が悪い。この使い勝手の悪さを改善するためにも、日常生活における音声認識による文字入力の重要性が増すと考えられる。音声認識入力を日常的に使うとする場合、その入力方法としては、現状で次の3種類の方法が挙げられる。

- A) ヘッドセット型マイクを装着
- B) 操作対象機器にマイクを搭載  
(機器搭載型マイク)
- C) リモコンにマイクを搭載  
(ハンドヘルド型マイク)

A) のヘッドセット型マイクは、現状では最も認識精度が高い。ヘッドセット本体に音声認識エンジンを内蔵し、直接外部機器を操作する装置なども開発されている<sup>[61]</sup>。ヘッドセットの利用は、ハンズフリーであるので、両手が使えない状況では有効である。また、パソコンを用いた特定業務などにおいても、ヘッドセットを介した音声認識入力とマウス／キーボードなどによる操作を組み合わせるUI（ユーザインタフェース）<sup>[62]</sup>により、作業効率の向上が期待される。しかし、日常生活で常時ヘッドセットを装着しつづけるのは、拘束性が強く、受け入れがたい。

B) の機器側にマイクを搭載する方法（機器搭載型マイク）は、ユーザは身体に何も装着する必要がないという利点がある。しかし、機器からの距離、つまりマイクからの距離が遠くなることにより環境ノイズの影響を受けやすくなり、認識精度が悪化するという問題がある。この問題を解決するため、マイクロフォンアレイを用いて話者の方向を推定し、ノイズの影響を軽減し音声認識精度を改善するといった試み<sup>[63][64]</sup>もなされている。しかし、日常生活環境下で実際に使われるレベルの性能を確保するために、マイクの数を増やす必要があるなど現状ではまだ難しい。

これらに対し、C) のハンドヘルド型マイクは、ユーザの手元で音声を入力できるので、機器搭載型マイクよりは音声認識の精度が高く、ヘッドセット型より拘束感を少なくできる。また、テレビ視聴など、部分的に手による操作が可能な日常生活のシーンにおいては、面倒な文字入力などを音声認識で行い、ボタンによる操作と適切に組み合わせるのが現実的な使い方と考える。その面でも、一つのデバイスで音声認識入力とボタン操作を行うことができるハンドヘルド型が望ましい。

本研究では、まずハンドヘルド型音声認識入力の課題を、実験を基に明らかにする。次に、それを解決するための方法として、ユーザがハンドヘルド型マイクに音声入力する際の自然な動作を認識する状況認識技術（センサ駆動ハンドヘルド型音声認識入力方法）を提案し、その有用性を評価する。

## **5. 2 ハンドヘルド型音声認識入力の課題**

ハンドヘルド型マイクはヘッドセット型マイクに比較し、拘束性が少ない反面、音声認識精度に問題がある。その問題を明らかにするために、まずヘッドセット型マイクと音声認識精度を比較する予備実験を行った。その結果を基に、ハンドヘルド型マイクでの口元距離と音声認識精度との関係を調べる予備実験を行った。また、音声認識精度向上に必要なと考えられる発話開始／終了指示のボタン押し忘れについても予備実験を行った。

## 5. 2. 1 ヘッドセット型マイクとハンドヘルド型

### マイクの音声認識精度の比較

まず、ヘッドセット型マイクとハンドヘルド型マイクとの、音声認識精度の違いを評価した。本論文の全実験で使用した音声認識エンジンは、番組名、出演者名など 7,000 語彙が登録された孤立単語認識エンジンである。音響信号のサンプリング周波数は 16kHz で、フレーム幅 25ms・シフト幅 8ms で分析した。音響特徴量には 0-12 次の MFCC(Mel-Frequency Cepstral Coefficient)とその  $\Delta$  および  $\Delta\Delta$  で構成される計 39 次元の特徴ベクトルを用いた。尚、前処理に雑音除去は適用していない。音響モデルには 3 状態 20 混合の left-to-right 型の HMM(Hidden Markov Model)を用いた。

被験者 5 名に、同一の 100 種類の単語（番組名および出演者名）をヘッドセット型マイク、ハンドヘルド型マイクのそれぞれを用いて連続で発話してもらい、録音した。全発話の録音完了後、手作業にて発話区間を切り出して音声認識処理を施し、正しい音声認識結果が尤度上位 5 件に入るかどうかを評価した。尚、一般的な使用環境での特徴を把握するため、本論文の全実験は、家庭のリビングを模した実験室内で、被験者は椅子に座った状態で実施した。ドアの外は廊下となっているため、通行人の足音や話し声が聞こえる環境である。実際の家庭のリビングとは異なるが、雑音環境的にはほぼ同一レベルである。実験結果を図 5-1 に示す。

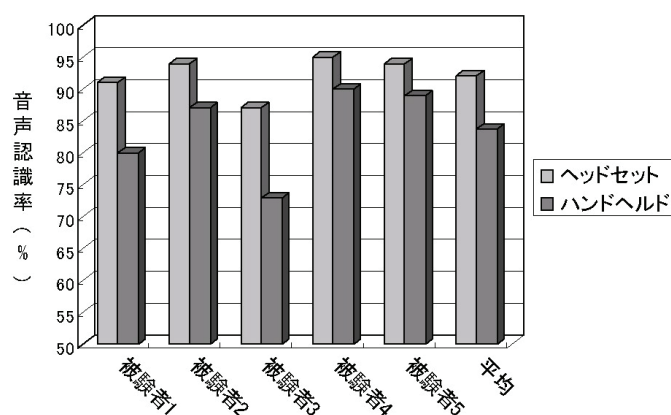


図 5-1 ヘッドセット型マイクとハンドヘルド型マイクの音声認識精度

5 名すべての被験者で、ハンドヘルド型マイクはヘッドセット型マイクよりも音声認識精度が低かった。全被験者の平均値は、ヘッドセット型マイクが 92.2%、ハンドヘルド型マイクが 83.8%であった ( $p < 0.01$ )。この差は、実験者の観察では、ハンドヘルド型マイクは発話時のマイクと口元の距離がヘッドセット型マイ

クよりも遠く、かつ一定となっていない点が原因と考えられる。次に、その影響について調査した。

## 5. 2. 2 マイクー口元間の距離と音声認識精度の関係

上述の実験と同一の実験環境において、ハンドヘルド型マイクを被験者の口元から 5cm、10cm、15cm、20cm の距離に順に固定し、それぞれの距離で同一の 100 単語の発話を録音し、後から手作業で発話区間を切り出して音声認識率を確認した。被験者は 1 名で実施した。結果を図 5-2 に示す。

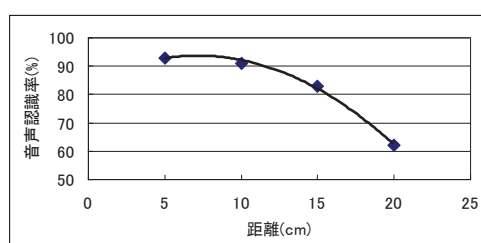


図 5-2 マイクー口元間の距離と音声認識精度

10cm 以内の距離で 90%以上（前節の同一環境のヘッドセットによる性能とほぼ同一レベル）の精度が確保できている。しかし、10cm を超えると距離に応じて精度が劣化していく。これは、距離が遠くなるとマイクに入力される音声の強度が弱くなり、SNR（Signal to Noise Ratio）が低下することが原因と考えられる。よって、一定レベルの音声認識率を確保するためには、ハンドヘルド型マイクで音声認識を行う際に、ユーザがマイクに対して適切な距離で発話するための支援が必要であることが示唆される。

## 5. 2. 3 操作面での課題

ハンドヘルド型マイクでの音声認識入力に限らず、従来の音声認識入力システムでは、認識精度を高めるために、発話開始と終了の始末端の切り出しが重要である。このため、ユーザが自らボタン操作などで音声認識の開始、あるいは終了を指示する方法が使われている。音声認識入力中に操作ボタンを押下し続ける方式（以下、プレストークと呼ぶ）や、音声認識の開始だけボタン押下で指示し、音声認識の終了は認識エンジンによる無音区間検出により自動的に行う方式（以下、プッシュトークと呼ぶ）などがある。

プレストークやプッシュトークでは、ユーザが明示的に始端（プレストークでは終端も）を指定してくれるのが利点である。しかし、筆者が過去に実施したプッシュトークによる音声認識を用いる実験では、始端でのボタン押下を忘れたまま発話してしまう事例が多く見受けられた。この傾向は、高齢の被験者など、機器の扱いに不慣れな被験者で特に顕著であった。

その発生頻度を調査するために、60歳以上の高齢被験者6名（男性3名、女性3名）に対して次の実験を実施した、前節での連続して発話する方法は操作面での課題が明確に観測しづらいため、本実験ではテレビの番組検索を音声認識およびボタン操作を組み合わせて実施するアプリケーションに変更した。そのアプリケーション使用時のボタン押し忘れ頻度を観測した。ここで被験者に課したタスクは、課題（番組名あるいは出演者名）を提示し、それに基づいて任意の検索キーワードを音声認識で入力して番組検索を行い、ボタン操作で提示された番組あるいは提示された出演者が出演している任意の番組の詳細情報を得るというもので、各被験者につき合計10タスクずつ行った。その結果、全発話における押し忘れ発話の発生頻度は15.3%であった。また、適切なタイミングでボタンを操作できず音声認識が正しく入力できない事例も観測できた。

これは操作の習熟により多少なりとも改善できる可能性はあるが、使い始めの段階から誰にでも使いやすいインタフェースとするためには、看過できない発生頻度であると考えられる。

## **5. 2. 4 課題のまとめと対策案**

以上のように、現状のハンドヘルド型マイクによる音声認識入力における課題として、以下の2点が挙げられる。

### **<ハンドヘルド型マイクによる音声認識入力の課題>**

- A) マイクと口元間の距離がユーザ任せになってしまい、ヘッドセット型マイクに比べて音声認識精度が低い
- B) 音声入力時にボタン操作で音声認識開始／終了を指示させる方式（プレストーク、プッシュトーク）では、ユーザによっては押し忘れが発生するなど、適切なタイミングでの操作が難しい

これらの課題の対策としては、以下の機能の必要性が示唆される。

## ＜課題を解決するための機能＞

- A) 発話時にマイクと口元を音声認識にとって適切な位置にユーザを誘導する機能
- B) ユーザの発話動作を検知し、ボタン操作なしで適切なタイミングで発話の開始と終了を切り出し、音声認識する機能

これらの機能を実現するために、それぞれ距離センサ、加速度センサを活用することを検討した。その検討内容について以降で詳しく説明する。

## 5. 3 センサの活用による課題解決の検討

### 5. 3. 1 距離センサによる発話動作検出

発話時のマイクと口元を音声認識にとって適切な位置にユーザを誘導する機能を実現するために、マイク近傍に距離センサを設置した実験用ハンドヘルド型マイクデバイスを試作した。音声認識にとって適切な距離は、図 5-2 の実験において 90%以上の認識精度が確保できる 10cm 以内とした。そこで、10cm を境に出力がデジタル的に変化する PSD (Position Sensitive Device) センサ GP2Y0D810Z0F (シャープ製) を距離センサとして採用した。

試作したハンドヘルド型マイクを使って自由に音声入力を行い、その際の距離センサの出力(口元が 10cm 以内にあるかどうかの検出/非検出)を測定した。図 5-3 は 4 回連続して発話した時の一例である。発話の際にハンドヘルド型マイクを口元に近づけることで概ね発話を検出できている。しかし閾値である 10cm 前後の距離で発話した図 5-3 の 1 番目と 4 番目の発話では、出力が不安定になってしまっている(手のぶれ等により 10cm 前後を行き来するため)。これを改善するため、口元がマイクから 10cm 以内であることを検出した後は、少なくとも一定時間(例えば 1 秒間)音声認識入力を継続するようにした。つまり 10cm を閾値として、10cm 以内になった時に音声認識入力を開始、一定時間音声認識入力を継続した後、距離が 10cm 以上になったときに認識終了とした。また 10cm 以上になる以前に、無音区間が続くと(ハンドヘルド型マイクを持ったまま沈黙した状態になった場合)、音声認識エンジン自身が音声認識を終了する。ここで、無音区間の検出は、音声認識開始から実際の発話開始までの雑音レベルの平均値を測定し、発話終了後、音声の入力レベルが発話開始前の雑音レベルに戻ることを基準として検出を行っている。よって、発話開始時が静かで、発話中あるいは発

話終了後に雑音が増加した場合などは、この無音区間検出は正しく動作しない場合がある。この点は今後の課題である。

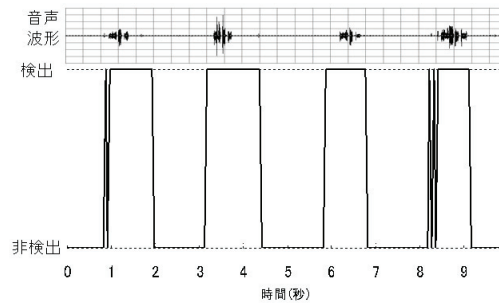


図 5-3 発話動作と距離センサの出力

### 5. 3. 2 加速度センサによる発話動作検出

ボタン操作なしで適切なタイミングで音声認識の開始と終了を切り替える機能を実現するために、3 軸加速度センサによりユーザの発話動作を検出する実験用ハンドヘルド型マイクデバイスを試作した。加速度センサの各軸の方向は図 5-4 の通りとした。図 5-3 で 4 回連続して発話した時に同時に測定した加速度センサの出力を図 5-5 に示す。

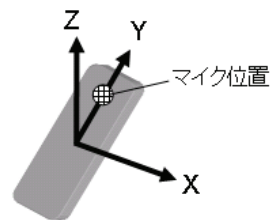


図 5-4 内蔵 3 軸加速度センサの軸方向

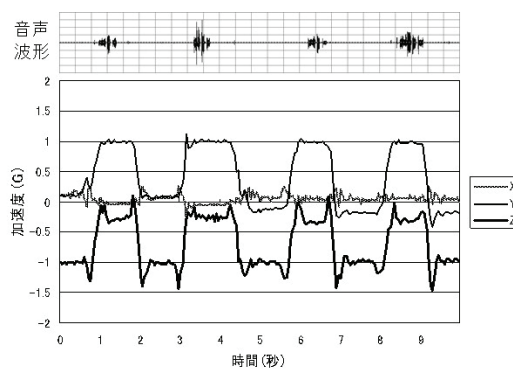


図 5-5 発話動作と加速度センサの出力

発話動作は鉛直方向の動作であるため、その Y 軸および Z 軸に特徴的な加速度の変化が発生することがわかる。これは、重力加速度に対するデバイスの傾きが変化することに起因する。

発話動作の波形全体の特徴量により、発話動作を検出する認識手法も考えられる。ただし、音声認識の開始をユーザの発話動作の開始に応じて遅滞なく切り替えるためには、発話動作の波形全体を考慮する手法では発話動作開始の検出に遅れが発生することが想定される。

そこで、立ち上がり／立ち下がりが明瞭である Z 軸に着目して、 $-0.7G$  を発話開始と発話終了の閾値とする単純な検出方法を採用した。つまり、放置状態では重力加速度が Z 軸のマイナス方向に加わることで $-1.0G$  を出力しているが、ユーザの把持・発話動作による傾き変化で $-0.7G$  を上回ると発話開始とし、 $-0.7G$  を下回ると発話終了とした。ただし、距離センサの場合と同様に、発話開始から発話終了まで最低 1 秒間は音声認識を継続する。また、 $-0.7G$  を下回る（発話動作を終了する）前に無音区間が続くと（ハンドヘルド型マイクを持ったまま沈黙した状態になった場合）、同様に音声認識エンジン自身が音声認識を終了する。突発的に大きな加速度変化があった場合に誤検出しないようにするため、一定時間以上連続で閾値を超えるかどうかを判定するなどのノイズ対策を施した上で、この閾値によって発話開始と終了を検知し、自動的に音声認識区間を検出できるようにした。

### **5. 3. 3 実験と考察**

距離センサを活用しユーザの口元がマイクに近づいた時に音声認識を開始する機能と、加速度センサを活用しユーザがハンドヘルド型マイクを構えて発話する動作を検知して音声認識を開始終了する機能について、それぞれの特徴を明らかにするための実験を行った。

被験者は、20~30 代の男女 7 名で、5. 2 の実験と同一の番組名、出演者名など 7,000 語彙が登録された孤立単語認識エンジンを使用した。それぞれ 20 人分の人名を提示して、順にハンドヘルド型マイクに発話してもらった。センサが発話動作を検知し音声認識が開始されると、パソコンの画面に音声認識開始と表示され、被験者はその表示を確認して発話を行う。尚、それぞれの検出方式の違いについては、被験者には説明していない。

まず、距離センサと加速度センサのそれぞれについて被験者の発話開始動作検出時に、音声認識開始が正しく認識されなかったエラーについて検討する。ここでは、実験者が観測した被験者の発話動作回数に対して、発話した際に音声認識開始とならない検出漏れと、発話していないのに音声認識開始となる誤検出の両



方をエラーとしてカウントし、式(5-1)で検出エラー率を算出した。その結果を図5-6に示す。

$$\text{検出エラー率(\%)} = \frac{\text{検出漏れ数} + \text{誤検出数}}{\text{発話動作回数}} \quad (5-1)$$

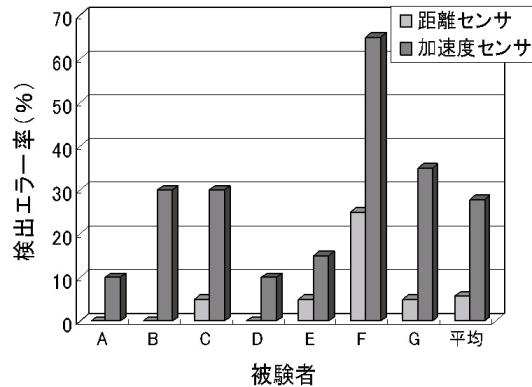


図 5-6 発話動作開始の検出エラー率

いずれの被験者も、加速度センサで発話動作を検出するよりも距離センサで口元とマイクの接近を検出して音声認識開始とした方が、エラー率が低いことがわかった。全被験者のエラー率の平均値は、距離センサの場合が 5.7%、加速度センサの場合が 27.9%であった。これは、加速度センサを用いた発話動作検出では重力加速度に対する傾きに閾値を設定して検出を行ったが、閾値として設定した角度に近い状態で把持し続けることが多い被験者でエラー率が高かった。

一方、それぞれの音声認識精度を図 5-7 に示す。尚、ここでは発話動作の検出エラーにより音声認識が正しく動作しなかった場合は除外している。すべての被験者で距離センサを活用した場合の方が、加速度センサを活用した場合に比べて同等かそれ以上の音声認識率を示した。平均値では、距離センサの場合が 81.4%、加速度センサの場合が 73.6%であった。これはマイクと口元の距離を 10cm 以内に確保して音声認識入力させる距離センサ活用方式が音声認識精度に良い影響を及ぼしていると考えられる。

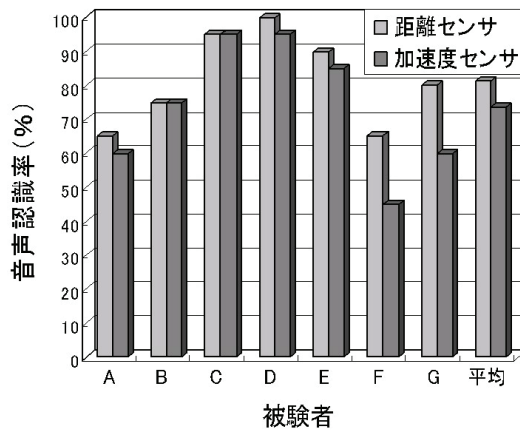


図 5-7 音声認識精度（距離センサー-加速度センサ）

以上により、発話動作検出および音声認識精度において、距離センサの使用が適していることがわかった。距離センサによる発話動作検出により、距離センサがない場合よりも平均値で 7.8 ポイント精度の高い音声認識を実現できた。ただし、距離センサによる発話動作検出は、ハンドヘルド型マイクが使用されずに机上などに置かれている場合でも、距離センサ付近に手などをかざすだけで音声認識開始としてしまう誤検出が実験準備中などに発生した。また、距離センサは消費電力が大きいため、常にアクティブにしておくのは実運用上望ましくない。そこで、距離センサに比べると低消費電力で動作する加速度センサと組み合わせ、低消費電力化するとともに、ロバストな音声認識開始/終了検出を行うセンサ駆動ハンドヘルド型音声認識入力方法を提案する。

## 5. 4 センサ駆動ハンドヘルド型音声認識入力方法

### 5. 4. 1 アルゴリズム検討

距離センサと加速度センサを活用してユーザの自然な発話動作を検出し、適切なタイミングで音声認識の開始、終了を自動的に切り替えるセンサ駆動ハンドヘルド型音声認識入力方法を開発した。その処理の概要を図 5-8 のフローチャートに示す。

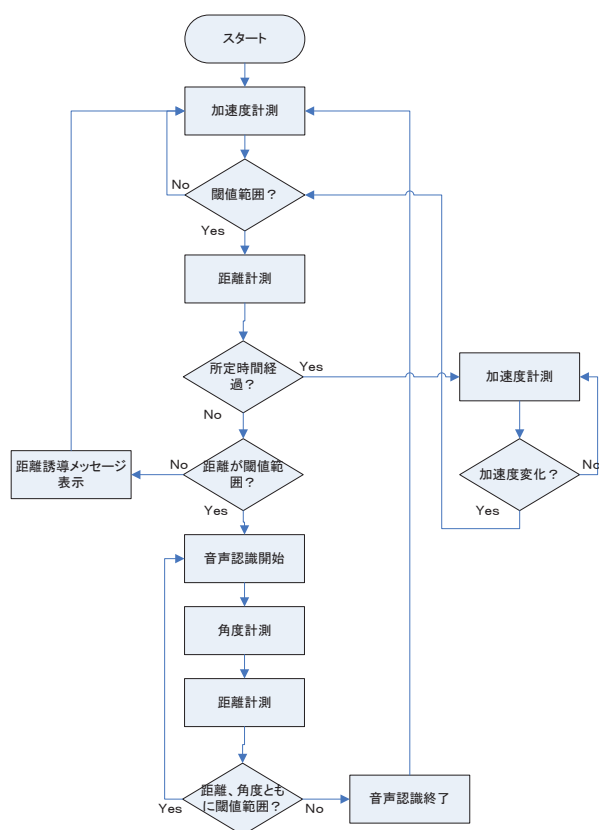


図 5-8 処理動作概要

まず、消費電力が小さい加速度センサ (typ. 0.36mA at 3V) でデバイスがユーザに把持されるかどうかを重力加速度に対するデバイスの傾きをもとに判断する。ユーザに把持され、発話動作を検知した時点で初めて距離センサを起動し、ユーザの口元とマイクの距離を計測する。加速度センサで発話動作を検知したにも関わらず、マイク-口元間距離が閾値 (10cm) よりも遠い場合は、音声認識にとっては適切な距離でないため、音声認識を開始とせず、口元をマイクに近づけるようにアプリケーション画面にメッセージ (例: 口元をマイクに近づけて下さい) を表示する。メッセージが表示できる画面を搭載していない機器の場合は、ハンドヘルド型マイクに LED を搭載してその光で通知する方法や、音声合成の活用も考えられる。加速度センサの出力が発話動作を検出した状態で、かつ距離センサによりマイク-口元間距離が 10cm 以内であることを検出した場合に、音声認識開始と判断し、アプリケーションへ音声認識開始コマンドを送信する。発話動作終了を検知した際には、音声認識終了コマンドを送信する。また、発話動作終了前に無音区間が続くと (ハンドヘルド型マイクを持ったまま沈黙した状態になった場合)、音声認識エンジン自身が音声認識を終了する。

尚、発話動作に似た角度でデバイスが置かれた状態になった場合は、常に距離センサがアクティブになってしまうため、加速度変化がない状態が所定時間を超

えた場合は一旦距離計測を終了し、再びユーザがデバイスを把持して加速度が変化するまで距離計測を行わないこととした。

以上の処理により、ハンドヘルド型マイクが使われていない時に距離センサ付近にユーザの手などが近づいた場合の誤検出をなくことができ、かつ消費電力の大きい距離センサを必要最小限に起動することで、消費電力を抑えることが可能となる。また、ユーザに対して適切な距離での音声認識入力を支援し、ユーザの自然な発話動作に応じて適切に音声認識開始／終了検出を行うため、ユーザに負担をかけずに高い音声認識精度を実現できるものと期待できる。

## 5. 4. 2 試作デバイス仕様

上述の機能を実現するために、距離センサ、加速度センサ、およびセンサの出力から発話動作、マイクロ元間距離の検知を行うマイクロコントローラ(MCU)などを搭載したセンサ駆動ハンドヘルド型マイクを試作した、外観を図 5-9 に、概略仕様を表 5-1 に示す。

MCU は、センサの出力から発話動作、マイクロ元間距離を検知し、それらに基づいて音声認識アプリケーションの処理動作を変更するコマンドをそれぞれ発行し、RS-232C 経由で外部機器へ送信する。今回の試作システムでは、パソコン上の音声認識アプリケーションに各コマンドを送信して音声認識エンジンの動作を切り替える。



図 5-9 センサ駆動ハンドヘルド型マイク

表 5-1 試作デバイス仕様

距離センサ	PSD センサ GP2Y0D810Z0F (シャープ製)
加速度センサ	H34C (日立金属製)
MCU	PIC18LF1320-ML (Microchip Technology Inc. 製)
マイク	モノラルマイク AT9642 (オーディオテクニカ製)
外部入出力	マイク出力, RS-232C
操作 / 表示	操作ボタン : 1, 状態表示用 LED : 2, 電源スイッチ

### 5. 4. 3 性能評価実験

試作したセンサ駆動ハンドヘルド型音声認識入力方法の性能を評価するため、5.3の実験と同一の被験者7名（20~30代男女）に、同様にそれぞれ20人分の人名を提示して、順にセンサ駆動ハンドヘルド型マイクに発話してもらった。結果（全被験者平均）を、距離センサのみ、加速度センサのみで実施した際の結果とあわせて表5-2に示す。距離センサ、加速度センサの両方を活用することで、特に発話動作の検出エラー率を大きく改善できることが確認できた。

表 5-2 性能評価実験結果

	距離センサのみ	加速度センサのみ	距離センサと加速度センサ
検出エラー率	5.7%	27.9%	0.7%
音声認識率	81.4%	73.6%	84.3%

### 考察

音声認識率については、距離センサにより音声認識に適したマイクと口元の位置が良好な状態に確保できている。距離センサと加速度センサの両方を活用した場合は、距離センサのみの場合と同様に、加速度センサのみの場合よりも精度が良いことが確認できた。

一方、発話動作検出エラー率については、二つのセンサの特徴を適切に活用した提案手法により、性能改善が確認できた。ただし、今回の被験者は比較的機器操作に慣れている20~30代の被験者であったため、高齢者も含む一般的な被験者に対して、提案手法の性能を改めて検証する必要がある。

### 5. 4. 4 従来手法との比較評価

開発したセンサ駆動ハンドヘルド型音声認識入力方法について、高齢者も含む一般的な（技術系の職種でない）被験者に対して、音声認識入力時にボタンを押下する従来手法とあわせて、比較評価を実施した。

被験者は21名で、内訳は20~30代が9名（男性4名、女性5名）、60代が12名（男性6名、女性6名）である。実験条件はこれまでと同一で、リビングを模した実験室内で、番組名、出演者名など7,000語彙が登録された孤立単語認識エンジンを使用した。音声入力には、試作した図5-9のセンサ駆動ハンドヘルド型マイクを用い、センサ駆動ハンドヘルド型音声認識入力方法と、従来手法であるプレストーク、プッシュトークによる音声認識入力が同一デバイスで実施できるように機能を実装し、同一タスク（人名20人分発話）に対する音声認識入力を

行った。センサ駆動、プレストーク、プッシュトークの実施順は被験者毎に入れ替えを行い、実施順による影響を排除した。

被験者には、各入力方法によるタスク実施直前に使用方法を記載した説明資料のみを提示し、練習なしで直接タスクを実施してもらった。プレストーク、プッシュトークについては、エラー率として実験者が観測した被験者の発話動作回数に対する、ボタン押し忘れ数の割合を算出した。センサ駆動については 5.3 の実験と同様に式(5-1)にて算出した。実験結果として、全被験者平均と高齢被験者の平均（括弧内）を表 5-3 に示す。

表 5-3 従来手法との比較実験結果

	プレストーク	プッシュトーク	センサ駆動
エラー率 (高齢者)	1.9% (2.9%)	8.1% (13.8%)	4.8% (5.0%)
音声認識率 (高齢者)	75.5% (62.1%)	81.9% (71.3%)	82.4% (77.3%)

プッシュトークのエラー率（押し忘れ率）は、5.2.3 で計測した 15.3%より低く 8.1%であった。これは、高齢者以外に 20~30 代の被験者もあわせた平均となっているため、高齢被験者だけの平均は 13.8%と、やはりかなり高い頻度で押し忘れが発生していることが確認できた。

音声認識率については、センサ駆動が全被験者の平均値としては最も良い数値となった。ただ、センサ駆動でのタスクを実施した後に、プレストーク、あるいはプッシュトークを実施した場合は、センサ駆動の後もセンサ駆動で指示された位置（マイクから 10cm 以内）での発話となる場合が多く、距離センサによる適切な口元位置支援機能の効果を確認しづらかった。しかし、全タスクの最初の音声認識率だけピックアップすると、プレストークが 74.3%、プッシュトークが 76.4%、センサ駆動が 80.0%である。つまり距離センサによる適切な口元位置支援機能の効果が読み取れる。更に、高齢被験者だけの音声認識率を比較すると、プレストークが 62.1%、プッシュトークが 71.3%、センサ駆動が 77.3%と、センサ駆動による音声認識率改善の効果が大きいことがわかる。

以上を踏まえ、プレストーク、プッシュトーク、センサ駆動のそれぞれの音声認識入力方法について、本実験の結果からわかった特徴を以下に示す。

#### ● プレストーク

エラー率（押し忘れ率）は最も低かった。これは、発話している間だけボタンを押し続けるため、操作に対する負荷が他の二つより高く、被験者にはボタン操作に対する意識が強く働き、押し忘れが少ないものと推測できる。実際に実験後の主観アンケートでもそのような感想を持つ被験者が多かった。

一方、音声認識率は最も悪い。これは、特に高齢被験者で、発話し始めてからボタンを押下したり、発話の途中でボタンを離してしまったり、ボタンを押し始めてからタスクを確認してしばらくしてから発話したりと、適切にボタンを操作できないため、誤認識となったケースが目立った。実際、高齢被験者（12名）のみのプレストークでの音声認識率は62.1%であった。逆に機器操作に抵抗のない被験者にとっては、音声認識の開始と終了を自分の意思ではっきり切り替えられるので操作は少々面倒でも使いやすいといった感想も少数であるがあった。

#### ● プッシュトーク

プッシュトークでは、音声認識の開始だけ被験者がボタンを押下し、音声認識終了は認識エンジンが無音区間を検出することで自動的に行うこととした。そのため、プレストークに比べると操作に対する負荷は低く、うまく扱えないことに起因する誤認識は少なかった。一方で、操作に対する負荷が少ないことが、操作を忘れがちにさせる傾向があり、エラー率（押し忘れ率）は全体で8.1%、高齢被験者だけだと13.8%と、プレストークに比べて高い結果となった。

#### ● センサ駆動

センサ駆動は、自然な発話動作をセンサが検出することで、ボタン操作をすることなく適切なタイミングで自動的に音声認識の開始／終了を切り替えることができる。そのため、プレストーク、プッシュトークによる入力を先に実施した被験者にとっては本当に操作なしで入力できるのか少々不安に感じる様子もあったが、一旦慣れてしまうと非常に使いやすいとの感想が多数寄せられた。

発話動作の検出は、距離センサと加速度センサを併用することで頑健性が増し、高齢被験者に使い方を丁寧に説明する必要なく音声認識入力を使ってもらうことができた。発話動作の検出漏れと誤検出をあわせた検出エラー率は4.8%（高齢被験者だけで集計しても5.0%）とプッシュトークのエラー率に比べて優位な性能を実現できた。

これらの特徴をまとめると、次のように結論づけることができる。プレストークは機器操作が不得手なユーザにとっては適切に扱うのが難しく、プッシュトークによる音声認識入力の場合は、ボタンを押し忘れたまま発話してしまう頻度が高い。本研究で提案したセンサ駆動は、双方の欠点を補う特長があり、習熟なしで音声認識入力を適切に扱うことができる方法として有用であると考えられる。

## **5. 5 番組検索タスクでの評価**

センサ駆動ハンドヘルド型音声認識入力方法により、高齢者でもトータルでの音声認識精度が向上できることを確認した。次に、実際にテレビの番組検索を行うシステムの開発を行い、日常生活における番組検索シーンでハンドヘルド型音声認識入力を利用可能か検証した。開発に際して、まず、上述した音声入力方法以外の音声認識システムの課題について検討した。

### **5. 5. 1 音声認識システムの課題**

音声認識（孤立単語認識、連続音声認識）を用いたシステムでは、音声認識により受理可能な語彙は事前に辞書登録されていなければならない。ユーザが辞書登録されていない語（未知語）を発話した場合、音声認識システムは、事前に登録された語彙の中から発話された音声に近い語彙を候補とする。そのため、未知語の存在は誤認識やそれに起因するシステムの誤作動の原因となる。

未知語発話による誤認識を減らすためには、発話が想定される語彙を事前に幅広く辞書登録しておく必要がある。しかし、一般に音声認識システムが受理可能な語彙数と音声認識率はトレードオフの関係にある。語彙数が多くなると類似候補が多くなるため、音声認識率は下がる（認識に要する時間も増加する）。逆に、語彙数が少ないと、類似候補が少ないので、大語彙の場合に比べて高い精度で音声認識入力を行うことができる（認識に要する時間も短い）。しかし発話語彙に対するカバー率（発話語彙カバー率）が低下するため、未知語による誤認識が多くなる。つまり、極力未知語を減らし、最適な語彙セットで辞書を構成できるかどうか、実運用では非常に重要である。

発話語彙カバー率向上の一つの方法として考えられるのが、場面に応じて辞書を切り替える方法である。音声認識による操作が可能なカーナビでは、階層化したメニューを音声で選択し、メニューごとに認識語彙を切り替えるが、基本的には正式名称の固定語彙である。しかし実際には、「二子玉川（ふたこたまがわ）」は「にこたま」、「溝の口（みぞのくち）」は「のくち」のように、普段から愛称や略称で呼ばれる地名は多い。つまり、正式名称は人口に膾炙すると愛称や略称で呼ばれるようになり、固定語彙ではこのような語彙の変動に追従できない。愛称や略称のような語彙変動への追従は、特に AV 機器の利用において重要であり、語彙変動に追従できるように、検索対象コンテンツ変動に合わせて認識語彙を生成する必要がある。例えば、テレビにおける情報検索は Web 検索と番組検索が主な対象となる。音声認識を用いた Web 検索については、検索対象テキストから言語モデルを作成することで認識率の改善を行う取り組みなどがある<sup>[65]</sup>、一方、番



組検索については、検索対象は配信されている EPG (Electronic Program Guide) データである。EPG データに基づいて辞書を再構成すれば、未知語を削減できる。しかし、EPG データだけでは、正式名称以外の愛称や略称などの言い換え表現が含まれないため、これらが未知語として誤認識の原因となる。これまでに、登録語彙に対する略語の推定を行い、辞書へ追加する取り組み<sup>[66][67]</sup>はあるが、番組検索においては、略語以外に人名の愛称が使われることも想定される。

また、従来音声認識では、操作も含めてすべて音声認識で行う方法が採られてきた。特にカーナビでは、運転中はすべての操作を音声で行う必要があり、各メニューで非常に限られた認識語彙数にして切り替えている。しかし、階層ごとに受入可能な語彙をユーザが意識しなければ、適切な語彙を入力できず、また、目的遂行まで複数回の入力操作が必要となり、煩雑になってくる。これに対しテレビでは、すべての操作を音声で行う必要はなく、手元のリモコンを併用したマルチモーダルな入力体系が適している<sup>[68]</sup>。この点は車載用途との大きな違いである(もちろん、車載用途においても、停車中であればマルチモーダルな入力は利用可能ではある)。

以上まとめると、発話語彙カバー率向上のために、検索対象コンテンツごとに愛称等を加えた認識語彙を生成することが必要である。更に、音声認識インタフェースとして使い勝手を向上するために、すべて音声認識で操作する方法ではなく、リモコン操作と音声認識を融合したマルチモーダルな操作が必要である。本研究では、この 2 つの提案手法の有用性を、テレビの番組検索機能を題材として実証する。尚、愛称等の言い換え表現の推定方法は若木らの手法<sup>[69]</sup>を活用している。この手法では、推定手法の有用性を Web に出現する愛称数に対するカバー率(言語的語彙カバー率)により明らかにしている。これに対し、本研究は音声認識インタフェースを扱っているので、実際に発話された語彙に対するカバー率(発話語彙カバー率)により有用性を明らかにしている点が異なる。

## **5. 5. 2 番組検索のための認識語彙生成**

### **言い換え表現対応の必要性**

EPG には、正式な出演者名、正式な番組名のみが記載されており、出演者の愛称や番組の略称の記載はない。ユーザが自然な発話で所望の番組を検索するためには、音声認識エンジンがこれらの言い換え表現にも対応することが必要と考える。そこで、まず言い換え表現対応の必要性を確認するため、

- テレビに出ている有名人を普段呼んでいる呼び方

● 番組名を普段呼んでいる呼び方

について、アンケート調査を実施した。被験者は 20 代から 50 代の計 32 名で、それぞれの設問に対し、思いついた人または番組について、自由に 10 個ずつ記述方式で回答してもらった。収集できたデータ数は、人名の呼び方が 326 個、番組名の呼び方が 317 個であった。人名の呼び方の内訳を図 5-10 に、番組名の呼び方の内訳を図 5-11 に示す。

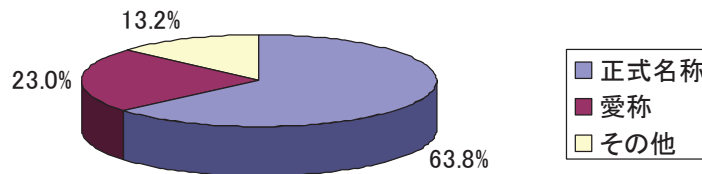


図 5-10 出演者名の呼び方の内訳

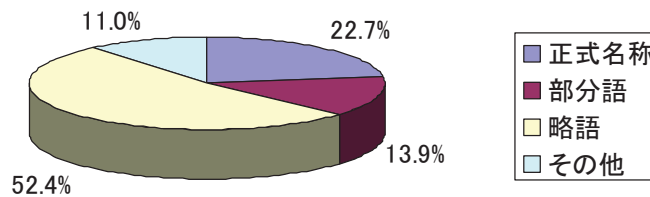


図 5-11 番組名の呼び方の内訳

人名の呼び方については、正式名称またはそれに準じる表現が 63.8%、愛称が 23.0%、その他の呼び方が 13.2%であった。ここで、正式名称またはそれに準じる表現とは、本名（例えば「森田一義」）、芸名（例えば「タモリ」）、姓、名、あるいは姓+「さん」（例えば正式名称が「木村拓哉」の場合の「木村」「拓哉」「木村さん」）などを指す。愛称とは、特定の人物に対して定型的に使用される表現であり、一般に通用する呼び名であるとする（例えば「木村拓哉」の愛称としての「キムタク」）。これら正式名称、愛称に分類できない表現をその他の呼び方とする。その他の呼び方としては、グループ名と組み合わせた表現（例：爆笑の太田）、役職や職業名をつけた呼び名（例：三谷監督）、説明的表現（例：麒麟の田村じゃない方）などがあつた。つまり、EPGに記載されている正式名称およびそれから機械的に生成できるそれに準じた表現だけの対応では、ユーザが使う人物表現に対し 6 割程度の言語的語彙カバー率しか見込めないこととなる。

一方、番組名の呼び方については、正式名称（例：パネルクイズ アタック 25）および正式名称を明示的な区切り文字（記号、空白、句読点など）で分割し

て生成できる部分語表現（例：アタック 2 5）を合わせた呼び方は、全体の約 3 分の 1 にあたる 36.6%であった。番組名の正式名称を省略して作られた 1 語（例えば「ズームイン！！サタデー」を略した「ズムサタ」、「ミュージックステーション」を略した「M ステ」など）は、52.4%と全体の半数以上を占めることがわかった。

以上のように、出演者名、番組名の双方について、愛称、略称への対応の必要性が示唆される結果となった。本研究では、まず、人名の愛称への対応を行い、番組名の部分語対応と合わせて音声認識語彙を拡張し、その効果を確認する。

## 処理概要

音声認識による番組検索を実現するために、テレビ放送で実際に配信されている EPG から、出演者名、番組名を抽出すると共に、前節で必要性が示唆された出演者名の愛称を追加した音声認識用の語彙セットを生成する。一連の処理の流れを図 5-12 に示す。

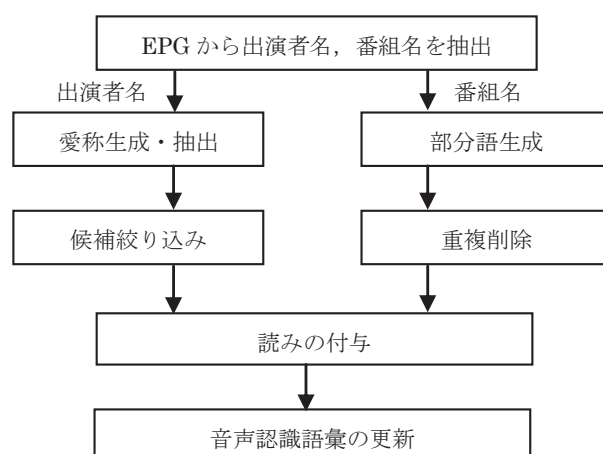


図 5-12 EPG からの音声認識語彙生成処理概要

まず、配信されている EPG（概ね 8 日分）に記述されている正式な番組名と正式な出演者名を抽出する。EPG はタグによって構造化されており、番組名として <TITLE>タグに囲まれた文字列を抽出する。出演者名は長い番組内容記述を表す <LONG\_DESC>タグに囲まれた記述内容のうち、<ITEM>タグに「出演者」と記載された直後の <TEXT>タグに記述されており、それらを抽出する。尚、記述内容には、カンマ、括弧、「〇〇ほか」など人名以外の記号や記述も多く存在するため、それらを除外して出演者名のみを抽出する。

このようにして抽出した出演者名に対する愛称推定、番組名に対する部分語生成は、それぞれ別プロセスで実行する。その処理内容については後述する。最後に、EPG から抽出した出演者名、番組名、および出演者名の愛称、番組名の部分語に含まれる漢字やアルファベット表記に対して読みの付与を行い、音声認識エ

エンジンの認識語彙を更新する。読みの付与には、音声合成で用いられている汎用の読み生成ツールを活用した。

## 言い換え表現による語彙拡張

### 出演者の愛称

EPG から抽出した出演者名について、以下の二通りの手順で愛称を推定する。

#### ① ルールに基づいた（人名由来の）愛称候補生成

人名に由来した愛称候補を生成するにあたり、まず、人名とその愛称の組を事前に学習データとして与え、人間が愛称をつけるルールを学習させる。ここで、正式名称（姓／名）、その読みのひらがな、その読みのカタカナ、愛称の 4 つを学習データとして与える。各入力に対し、次に示すルールで番号を与え、1 文字ごとに 3 桁の数字へ置換する。まず、100 の位の値は文字種を表し、「1」が正式名、「2」がひらがな表記、「3」がカタカナ表記に対応する。10 の位の値はその文字が姓名のいずれに含まれるかを表し、姓ならば「1」、名ならば「2」とする。1 の位の値はその文字の各単語中での先頭からの位置を表す。表 5-4 は、正式名称「滝沢秀明」のカタカナ表記に対して上記ルールで番号を付与した例である。「滝沢秀明」の愛称「タッキー」を学習する場合、「タッキー」は、表 5-4 の記号に置き換えると「311 ッ 312ー」となり、このルールを学習し保存する。

表 5-4 人名「滝沢秀明」の場合の記号への変換例（カタカナ表記）

	文字種	姓名	文字位置	記号
タ	3	1	1	311
キ	3	1	2	312
:	:	:	:	:
ア	3	2	3	323
キ	3	2	4	324

このようにして得られた愛称生成ルールを元に、新たな人名（正式名称、ひらがな表記、カタカナ表記）に対応する候補を生成し、これらの各候補について、Web 上で正式名称を検索語とする検索結果上位の文書中で、正式名称 (fullname) の前後に各候補の記述があるかどうかを調べ、該当するものを愛称候補 (nick) とする。更に、Web 上で「nick こと fullname」の連語を検索語として検索を行い、検索結果が一定数以下の nick は削除する。このようにして絞り込みを行い、人名由来の愛称候補を生成する。

## ② Web上の知識を用いた人名に由来しない愛称の抽出

愛称には、人名に由来せず上述のルールでは生成できないものも存在する（例えば、「斎藤佑樹選手」の愛称である「ハンカチ王子」）。これらは単純なルールでは生成することが難しいため、Web上の知識を活用した以下の手順で対応を試みる。

まず、「こと fullname」で Web 検索をし、検索結果上位のページ（Page）のうち、「こと fullname」の前の文字列を集める（この文字列集合を Str とする）。「こと」の前の 10 文字に対する全接尾辞 10 個に対し、Page と Str 中での直前の異なり数と先頭の異なり数を数える。各接尾辞について、直前の異なり数>1 かつ先頭の異なり数=1 の時、その接尾辞を Data として登録し、また、直前の異なり数=1 かつ先頭の異なり数>1 の時、その接尾辞の先頭の 1 文字を除いた残りの文字列を Data として登録する。Data のうち、Page 中で 1 回も出現しないものは削除する。残った Data について、末尾の共通する部分文字列は Str 中でのそれぞれの出現頻度を用いて絞り込みを行い、nick とする。最後に、①と同様に Web 上で「nick こと fullname」の連語を検索語として検索を行い、検索結果が一定数以下の nick は削除する。

上述の二つの方法で抽出した愛称が、実際に使われている有名人の呼称表現をどのくらいカバーできるかを調査したところ、本手法で抽出した愛称の言語的語彙カバー率は 81.5%であった。つまり、図 5-10 中の愛称 23.0%のうち、81.5%分に相当する 18.7%は提案手法によってカバーできるものと推定できる。すなわち、図 5-10 のうち 82.5% (=63.8+18.7) がカバーできる計算となる。尚、出演者の愛称抽出については、文献[69]に詳しいので、そちらを参照されたい。

## 番組名の部分語

EPG から抽出した番組名については、特徴的な記号や空白を区切りとして以下の手順で部分語を生成する。

### ① 不要文字・不要区間の削除

[二]（ニカ国語放送）、[再]（再放送）、[午前 5：00 まで放送を休止します] など、全角“[]”、および半角“[]”には番組名以外の放送に関する属性情報が記述されているため、それらは削除する。

また、“[字]あしたをつかめ～平成若者仕事図鑑▽笑顔とセンスでお客様をつかめ～デパート販売員”のように、番組名の中の“▽”以降に番組内容に関する記述がある場合があるため、“▽”以降の記述も削除する。

## ② メインタイトルとサブタイトルの分割

次に、括弧（「」『』など）をサブタイトルと判断して、括弧前までをメインタイトル、括弧の中をサブタイトルとして分割する。

## ③ メインタイトルの分割

メインタイトル中に、空白、感嘆符“！”、記号“～”、“－”などがあれば、それらの前後でタイトルを分割し、それぞれの組み合わせをメインタイトルのバリエーションとして登録する。

上記手順で生成した部分語のそれぞれについて読みを付与することで、例えば下記のような部分語読みを生成する。

・ EPG 記載番組名：NHK 高校講座 理科総合 A・B 「ひんやり・ほかほかの物理」

・ 生成した部分語の読み：

えぬえいちけいこうこうこうざ

りかそうごうえーびー

ひんやりほかほかのぶつり

えぬえいちけいこうこうこうざりかそうごうえーびー

えぬえいちけいこうこうこうざりかそうごうえーびーひんやりほかほかのぶつり

尚、“NHK 高校講座 日本史「……」”などの他の教科の番組名からも“えぬえいちけいこうこうこうざ”の部分語が生成されるため、同一の部分語が生成された場合は重複登録しないように削除する。また、部分語の読みに対応する表記は、その読みに対応した表記だけを登録しておく。よって、当該部分語を検索キーワードとした部分一致検索を実施することになる。

また、上記の例では、“えぬえいちけいこうこうこうざりかそうごうえーびーひんやりほかほかのぶつり”など、EPG 記載の番組名全部に対する読みも今回は登録したが、実際にこのように長い番組タイトルをそのまま全部発話することはあまりないと考えられる。一定文字数以上の読みは削除するなど、読みの長さを基準とした語彙の選別も今後検討が必要である。

## 5. 5. 3 マルチモーダル操作でのテレビ番組検索

### による実証実験

上述の手法により実際の EPG から出演者名の愛称、番組名の部分語による音声認識語彙の拡張を行い、音声認識を活用した番組検索の効果を検証するための実証実験システムを試作した。本システムにより、従来のリモコン操作のみによる番組検索に比べて、操作負担（操作時間）の軽減（短縮）と使い勝手の向上を目標としている。

検索対象とするのは、実際の地上波テレビ放送における EPG 8 日分のデータで、上述した手法で出演者名とそれらの愛称、番組名とそれらの部分語を抽出し、読みを付与して音声認識語彙を生成した。8 日分の EPG から抽出した正式な出演者名 3,573 語に、提案手法で抽出した 992 語の愛称を追加した計 4,565 語と、正式な番組名と提案手法で生成した部分語を追加した 2,498 語、両者を合わせ音声認識語彙として登録した語彙数は、計 7,063 語とした。

追加した愛称 922 語の選定は次のようにした。図 5-13 は愛称数と Web 上の出現頻度との関係を示す概略図である。ある出現頻度（今回は目視で決定した。その値を T とする。T の決定方法については今後の課題である。）を境に、それ以下では愛称数が一気に増加する。これらには滅多に使われていないもの、一般に知られていないものが多数存在する。Web に存在する愛称全体に対する言語的語彙カバー率を高くするには、T はなるべく小さく選定すべきである。しかし、出現頻度が少ないものは、実際にユーザが知らない場合も多く、発話されない可能性も高い。つまり

#### 言語的語彙カバー率 ≠ 発話語彙カバー率

となっており、音声認識では発話語彙カバー率を優先すべきである。言語的語彙カバー率は Web に出現する全語彙数と選択した語彙数から単純に算出できるが、発話語彙カバー率は実際にユーザに発話してもらわないとわからない。よって、今回の実験では、この出現頻度 T より多い出現頻度の愛称 922 語を採用した。この場合の言語的語彙カバー率は 7.6% であるが、発話語彙カバー率は後述するように 93.7% となっている。ただし、実用化に向けて実験を積み重ねることで、発話語彙カバー率と言語的語彙カバー率との関係についても明確にしていく必要がある。また、発話語彙カバー率を上げるために語彙数（特に愛称）を増やし続ければ良い訳ではなく、類似した発音が増え、音声認識率が悪化する可能性もある。発話語彙カバー率と音声認識率との関係も今後検討の必要がある。

ちなみに、本システムで使用する大語彙孤立単語認識エンジンにおいて、語彙数約 7,000 語、SNR が 20dB の環境で、発話区間を手動で切り出した場合、音声

認識率が 88.5%と実用的な性能が確保できることを、システム構築前の調査で確認した。尚、音響信号のサンプリング周波数は 16kHz で、フレーム幅 25ms・シフト幅 8ms で分析した。音響特徴量には 0-12 次の MFCC (Mel-Frequency Cepstral Coefficient) とその  $\Delta$  および  $\Delta \Delta$  で構成される計 39 次元の特徴ベクトルを用いた。尚、前処理に雑音除去は適用していない。音響モデルには 3 状態 20 混合の left-to-right 型の HMM (Hidden Markov Model) を用いた。

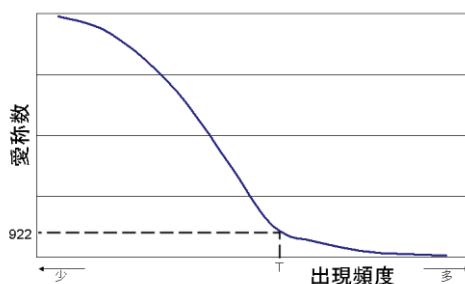


図 5-13 Web 上の出現頻度と愛称数の関係

## マルチモーダル操作

実験用システムはノート PC 上に構築した。操作デバイスは、本章で開発したセンサ駆動ハンドヘルド型マイクには操作ボタンが 1 つのみ搭載されておらずメニュー操作が困難であるため、市販のマイク内蔵型コントローラ (Philips 社製 SpeechMike Pro) を用い、音声認識トリガにはプッシュトークを使用することとした。ただし、5.5 の実験結果から、プッシュトークを用いたマルチモーダル操作の有用性が確認できれば、センサ駆動ハンドヘルド型マイクを使用したマルチモーダル操作では、さらに入力の負担を軽減することが可能であると言える。

出演者名、番組名およびその言い換えを含めて音声入力すると、音声認識候補リストが画面上に表示される。その中からの正解候補や目的の番組の選択などは、コントローラを用いたボタン操作にて実施する構成とした。日常生活においてテレビの番組検索を行うシーンを想定すると、自動車の運転中などとは異なり、リモコンによるボタン操作が可能であるため、音声認識入力とボタン操作をそれぞれ組み合わせて使用するマルチモーダル操作が適切であると考え、上記のような操作体系とした。

尚、8 日分の EPG データは、XML データベースに格納しておき、キーワード検索が実行された際に XQuery を発行してデータベース内を検索する。



## 機能概要

ここでは、音声認識による検索キーワード入力から、入力した検索キーワードによる EPG 検索および結果の表示までの機能概要について説明する。

まず、ユーザは目的の出演者名、あるいは番組名を正式名称、愛称、部分語など、好みの表現で音声入力を行う。音声入力の開始時に、マイク内蔵コントローラのボタン押下により音声認識開始をシステム側に通知し、音声認識の終了は、音声認識エンジンが無音区間を検出することにより自動的に行うプッシュトークを採用した。出演者名の愛称（この例では「そのまんま東」）を音声入力した際の一例、番組名の部分語（この例では「日本人テスト」）を音声入力した際の一例を図 5-14 に示す。音声認識の結果は、尤度順に上位の 8 件がリスト表示される。それぞれ正式出演者名「東国原英夫」に対する愛称「そのまんま東」、正式番組名「全国一斉！日本人テスト[字][S]」の部分語「日本人テスト」が音声認識エンジンに受理されたことがわかる。尚、それぞれの候補がどんな読みに基づいて候補としてリストアップされたかわかるように、正式出演者名、番組名（部分語）に対応した読みを括弧内に記載しておくこととした。もし、正解がリスト中不在の場合は、再度音声入力を行う。

図 5-14 の音声認識結果のリストから目的の人名あるいは番組名を選択すると、その正式名称で EPG が格納された XML データベースを検索する。検索結果の例を図 5-15 に示す。出演者名の場合はその出演者が出演している番組が、番組名の場合はその検索キーワードの文字列が含まれる番組が検索されてリスト表示される。あとは、これらのリストの中から所望の番組を選択すると、EPG の詳細情報を確認することができ、簡単に録画予約や視聴ができる。

尚、検索した出演者の当該期間での出演番組数が非常に多い場合、あるいは、検索キーワードの部分語文字列にヒットする番組数が非常に多い場合は、現状では番組候補リストを上下にスクロールして多数の候補の中から目的の番組を探し出す必要がある。図 5-16 に当該出演者の出演番組数が多い場合、部分語文字列にヒットする番組数が多い場合の例をそれぞれ示す。このように多数の候補の中から、リモコンのボタンだけで目的の番組を探し出すことは負荷が高い。また、別のキーワードによる再検索を行うにしても、どのようなキーワードで効果的な検索が可能かをユーザが判断することは難しい。よって、検索結果に応じて絞り込み検索用の認識語彙を生成し効率良く音声で絞り込みを行うことができる機能、更には、絞り込みに適したキーワードを提示する機能などの必要性が示唆される。

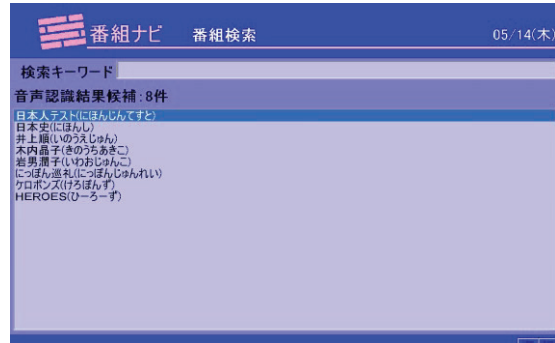
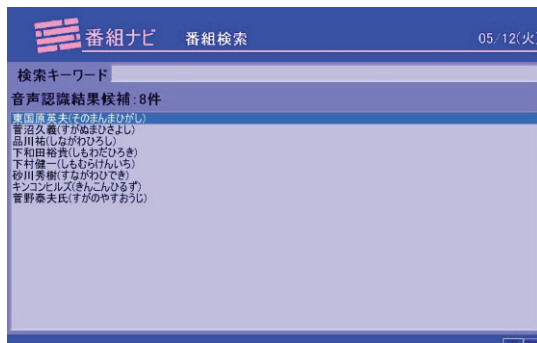


図 5-14 音声認識結果例

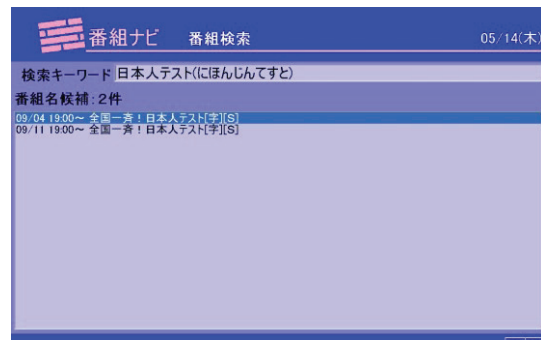
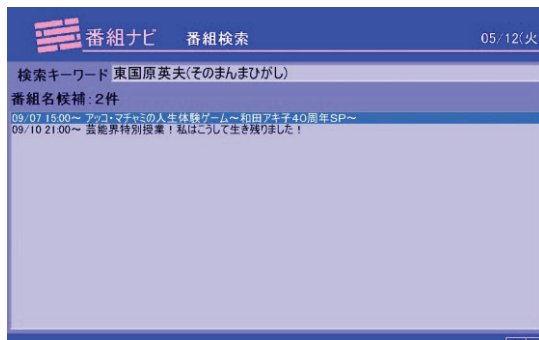


図 5-15 番組検索結果例 1

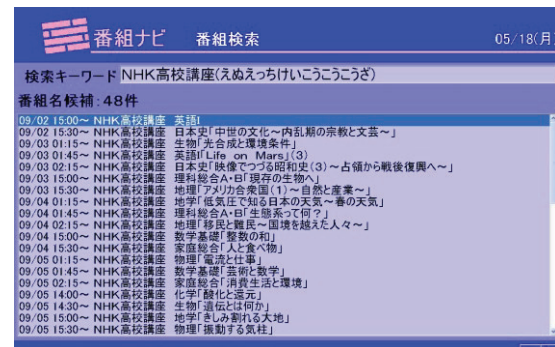
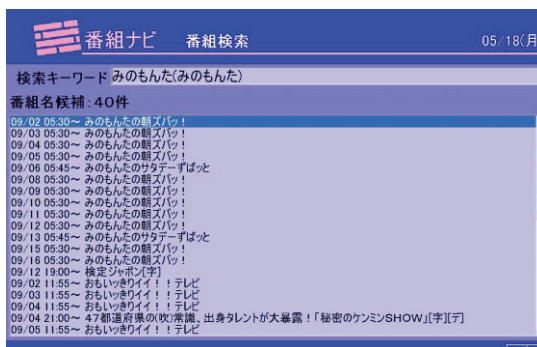


図 5-16 番組検索結果例 2

## 実験

### 実験概要

提案システムの有用性を評価するため、本システムを用いた番組検索の際の検索キーワードの入力を、下記の3種類の入力方法で行い、それぞれの特徴を比較評価する実験を実施した。

- ①キーボード入力（PC環境を想定）
- ②スクリーンキーボード入力（既存のテレビ／レコーダでの入力を想定）
- ③音声認識入力（提案手法であるマルチモーダル操作）

①では、検索キーワード入力をノート PC のキーボードで行い、その他の番組選択などの操作はマウスを用いた。②では、本システム上に既存のテレビ／レコーダで使用されているものを模して作成したスクリーンキーボードを用意し、すべてマイク内蔵型コントローラで（音声認識入力を使わずに）操作を行った。尚、テレビ／レコーダでは専用のリモコンによる十字キーあるいは携帯電話式の数字キー複数回押下により文字入力を行うのに対し、マイク内蔵型コントローラではトラックボールとボタンによる操作となる。事前に実際のレコーダ＋専用リモコンでの文字入力と、本システム＋マイク内蔵型コントローラでの文字入力を同一タスクで実施し、後者が前者より有意差（ $p < 0.01$ ）を持って短時間でタスク完了できたことが確認できたため、既存手法相当の入力として本システム＋マイク内蔵型コントローラを使用することとした。これに対して提案手法が有利であることが確認できれば、提案手法は既存手法に対しても有利であることができる。③では、マイク内蔵型コントローラのマイクに向かって音声認識入力を行い、音声認識候補や番組の選択はマイク内蔵型コントローラのトラックボールとボタンにより行う。

尚、①、②でも出演者名の愛称、番組名の部分語による番組検索を実施できるよう、入力されたテキストについて、音声認識語彙を参照することで正式名称を検索できるようにした。特に、入力した任意の文字列に対して部分一致検索を行えるようにしたため、③よりも①、②の方が入力の自由度は高く、有利な条件で実験を実施した。

## タスク

実験でのタスクは、被験者に録画予約する番組名あるいは出演者名を検索タスクとして画面に提示し、被験者は提示された検索タスクに基づいて検索キーワードを自由に入力して目的の番組を検索する。尚、被験者には本システムが愛称や部分語に対応していることは事前に説明しておいた。

検索タスクが番組名の場合は、その番組を録画予約できてタスク完了とする。レギュラー番組など同一名の番組が複数ある場合は日付も指定して提示することとした。出演者名が提示された場合は、その出演者が出演している任意の番組が録画予約されればタスク完了とする。一つのタスクが完了すると、続いて次の課題が提示され、次々にタスクを実施する。実験で使用したタスクは、図 5-17 の通りである。尚、これらの番組名、出演者名に関連する音声認識語彙の読みがすべて正しく付与されていることを確認した上で実験を実施した。

- |                              |
|------------------------------|
| (1) 風林火山(40)「三国同盟」           |
| (2) 「みのもんた」が出ている番組           |
| (3) 爆笑問題の検索ちゃん (10月13日放送)    |
| (4) 「森繁久彌」(もりしげ ひさや)が出ている番組  |
| (5) 日曜洋画劇場「アルマゲドン」           |
| (6) 「木村拓哉」(きむら たくや)が出ている番組   |
| (7) タモリ倶楽部(10月13日放送)         |
| (8) 「坂下千里子」(さかした ちりこ)が出ている番組 |

図 5-17 実験タスク

**実験条件**

家庭のリビングを模した実験室において、合計 34 名の被験者に対して、図 5-17 のタスクを①キーボード入力、②スクリーンキーボード入力、③音声認識入力のそれぞれについて実施した。尚、各入力の実施順は被験者ごとに入れ替えを行い、実施順による影響を排除した。被験者 34 名の内訳は以下の通りである。

- IT系研究者 (20～50代) : 12名 (男性7名, 女性5名)
- 非研究職一般者 : 22名
  - 若者 (20代) : 10名 (男性5名, 女性5名)
  - 高齢者 (60～70代) : 12名 (男性6名, 女性6名)

**実験結果**

各入力方法別に図 5-17 のタスクが全て完了するまでの平均所要時間を表 5-5 に示す。

表 5-5 入力方法別 全タスク完了までの平均所要時間 (秒)

キーワード入力方法	非研究職一般者			IT系研究者	全被験者
	若者	高齢者			
①キーボード	119.6	323.8	224.7	101.1	171.4
②スクリーンキーボード	217.2	426.6	323.8	243.3	277.0
③音声認識	123.8	228.9	180.8	139.9	165.8

ここで、高齢者 4 名は、①キーボードと②スクリーンキーボードでのタスクを途中で断念 (タスクごとの制限時間 10 分以内に完了しなかったか、被験者がギブアップした) したため、表 5-5 の集計には含んでいない。ちなみに、タスクを途中で断念した被験者は、①、②とも同一の 4 名であった。

全被験者の平均所要時間では、従来手法としての②スクリーンキーボードの場合は 277.0 秒、提案手法の③音声認識によりキーワード入力をした場合は 165.8 秒と、提案手法により所要時間を約 40%短縮することができた ( $p < 0.01$ )。

高齢者に対してはさらに顕著で、提案手法により所要時間を約半減でき、PC 環境を想定した①キーボードに対しても約 30%短縮することができた。しかも、上述の通りうち 4 名の被験者は①と②のタスクを途中で断念しており。提案手法の有用性を明らかに確認することができた。

また、非研究職一般者の若者と高齢者について、入力方法別の所要時間を図 5-18 に示す。この集計にも①キーボードと②スクリーンキーボードにおけるタスク途中断念 4 名のデータは除外している。高齢者は若者に比べて①キーボードと②スクリーンキーボードの分散が大きく、個人差が大きいことがわかる。その一方で、③音声認識の分散は小さく、機器操作の得意／不得意に関わらず、音声認識による検索キーワード入力は個人差が少ないことが読み取れる。

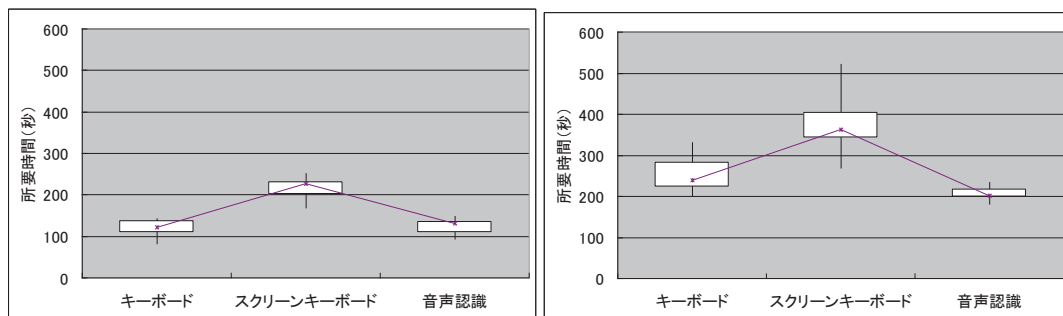


図 5-18 入力方法別 所要時間（左：若者，右：高齢者）

一方、発話に対する音声認識語彙の発話語彙カバー率は 93.7%であった。このうち愛称表現対応を行っていないとした場合の発話語彙カバー率は 85.1%で、愛称表現対応により発話語彙カバー率を約 9 ポイント向上できていることも確認できた。

音声認識語彙に登録されていない未知語発話のうち最も多かったものは、検索タスク「爆笑問題の検索ちゃん（10月13日放送）」に対する「けんさくちゃん」という発話で、全被験者で 8 回の発話が観測された。形態素解析を用いた部分語生成、更には、番組名の略称対応の必要性が示唆される結果となった。ただし、「けんさくちゃん」が受理されないことは少なくとも 2 度目の認識失敗で理解され、「ばくしょうもんだい」「ばくしょうもんだいのけんさくちゃん」など、別の受理可能な表現で再入力され、目的の番組を探し出すことができた。

ちなみに、1 タスクあたりの平均発話回数は 1.38 回で、これは誤認識発生の場合、あるいは未知語を発話した場合でも、たいていの場合は多くとも 3 回目の発話までに（未知語発話の場合は受理可能な表現を言い直して）正しく入力できるレベルである。尚、音声認識語彙に対応した発話における音声認識率は平均で

77.7%であった。事前に確認した認識率より低い結果となったが、ボタン押下による音声認識開始指示の後、目的のキーワード発話の前に「えーと」など、入力を意図しない発話に起因する誤認識が特に高齢者で多く発生したためであると考えられる。この点については今後の課題である。しかし、逆の見方をすれば、音声認識率がこの程度であっても、従来手法よりも所要時間を大幅に短縮できたことは、本手法の大きな特長であるとも言える。

各入力方法について、主観アンケートも実施した。①～③のそれぞれについて、使いやすいかどうか、面白いかどうか、実際に使ってみたいかどうかを、1 (Negative) ～5 (Positive) の5段階で評価してもらった。結果を平均値で表5-6に示す。IT系研究者や若者は、普段からキーボードに慣れているため、使いやすさに関しては①キーボードが最も高い結果となったが、実際に使ってみたいかどうかについては③音声認識が最も高い結果となった。日常生活におけるテレビ視聴時には、わざわざキーボードで文字入力を行うよりも、手元のリモコンで手軽に入力できることが求められていると考えられる。

表 5-6 主観アンケートの結果

キーワード入力方法		非研究職一般者			IT系研究者	全被験者
		若者	高齢者			
①	使いやすい	3.6	4.7	4.1	4.8	4.4
	面白い	3.3	3.1	3.2	2.8	3.0
	使ってみたい	3.1	4.3	3.6	3.7	3.7
②	使いやすい	1.8	2.1	1.9	1.4	1.7
	面白い	3.5	3.9	3.7	2.2	3.2
	使ってみたい	2.4	2.6	2.5	1.7	2.2
③	使いやすい	4.4	4.3	4.4	3.9	4.2
	面白い	4.6	4.3	4.5	4.4	4.4
	使ってみたい	4.4	3.9	4.2	4.2	4.2

また、高齢被験者に対しては、テレビ番組の録画予約を自分でできるかどうかを質問した。自分でできると回答した被験者は12名中3名のみで、その他の9名は同居している若い人をお願いして録画してもらっているとのことであった。これらの被験者からも、提案手法の音声認識を用いた番組検索が実際に使えるのであれば、番組予約を自分でやってみたいとのコメントが数多く得られた。

## まとめ

テレビの番組検索を題材に音声認識システムとしての問題解決を図り、音声認識とボタン操作を用いたマルチモーダル操作と言い換え表現に対応した音声認識語彙の生成により、システム全体としての使いやすさを大きく改善し、従来手法

(スクリーンキーボード+リモコン) に比べて番組検索タスクの所要時間を約40%短縮可能であることを確認した。特に高齢者でその効果が顕著であった。

## 5. 6 まとめ

本章では、音声入力インタフェースを日常生活で使用するためには、音声認識エンジンそのものの性能向上はもちろん重要であるが、それ以外に入力方式や音声認識語彙についてもターゲットに応じた検討が必要であることを問題提起した。

具体的には、使い方の側面から音声認識の性能向上を図る取り組みとして、加速度センサと距離センサを用いてユーザの自然な発話動作を認識する状況認識技術により、簡単かつ高精度な音声認識入力を実現するセンサ駆動ハンドヘルド型音声認識入力方法を提案した。まず、実験により、従来のハンドヘルド型マイクによる音声認識入力の課題は口元とマイクとの距離、発話開始と終了の検出にあることを明らかにした。その解決策として発話距離については距離センサ、発話動作検出については加速度センサの活用を検討した。実験の結果、距離センサの方が、音声認識精度、発話動作の検出に有効であることがわかった。しかし、距離センサ単独では、発話意図のない状況での誤検出が想定されると共に、常に動作させ続けるには消費電力が大きいという問題がある。そこで、両センサを組み合わせさせたセンサ駆動ハンドヘルド型音声認識により、発話動作検出のロバスト化と低消費電力動作を実現した。実験により、センサ駆動ハンドヘルド型音声認識は、プッシュトークでのエラー率(8.1%)より低いエラー率(4.8%)で発話開始を検出できること、プレストークより高い音声認識精度(プレストーク:75.5%、センサ駆動:82.4%)を得られることが明らかになった。これは、ハンドヘルド型マイクを使った音声認識入力において、ユーザへの負荷を少なくし、かつ高精度の音声認識を実現する手法として期待できるものである。

また、音声認識システムの課題である発話語彙カバー率を向上させることを目的とし、検索対象コンテンツごとに愛称等を含めた認識語彙を生成する手法を用いたシステム構築を行い、マルチモーダル操作との組み合わせによるユーザビリティ向上を検討した。出演者名の愛称に対応することにより、実用的な音声認識精度を確保できる語彙数の範囲において、正式名称だけの場合より、発話語彙カバー率が約9ポイント向上していることを確認した。また、音声認識インタフェースの使い勝手を向上させるために、すべての操作を音声認識で行うのではなく、リモコン操作と音声認識を融合したマルチモーダル操作を用いることの有用性を示した。テレビ番組検索による実証実験では、スクリーンキーボードを用いた従来手法による番組検索に比べて、提案手法の方がタスク完了までの所要時間を約

40%短縮することができた。また、提案手法は、高齢者などで機器操作が得意でないユーザにとって特に効果的であることもわかった。



## 6. 結論

---

### 6. 1 本論文の貢献

本論文では、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ことを目指し、各章に述べた研究に取り組んだ。これらによる成果を総括すると本論文の貢献は次の2点である。

- ① ユビキタスコンピューティング実現のための状況認識技術
- ② 状況認識技術のヒューマンインタフェース応用

以下に、それぞれについて具体的な貢献の内容をまとめる。

#### 6. 1. 1 ユビキタスコンピューティング実現のための状況認識技術

第2章「ウェアラブル健康管理システム」では、加速度、脈波、発汗、皮膚温などの多様な生体情報をセンシング可能なウェアラブルセンサを開発し、それを用いた生活習慣、健康状態に関わる状況認識技術の研究を行った。歩行、走行などの行動と歩数を加速度センサのデータから認識する状況認識技術を開発すると共に、脈拍数など他の生体情報についても単なるセンシング結果の見える化だけでなく、これらのデータを用いて健康管理上重要である食事イベントを精度良く検出可能にする状況認識技術を提案した点が特長である。

第3章「ウェアラブル睡眠センサとその応用」では、睡眠計測にフォーカスしたウェアラブル睡眠センサを開発し、内蔵した脈波センサと加速度センサのデータから詳細に睡眠状態を解析可能とする状況認識技術の研究を行った。加速度センサのデータから睡眠／覚醒を判定し、次に、脈波からロバストに取得した脈波間隔の周波数解析により自律神経の状態を推定し、自律神経と睡眠状態の相関関係からREM睡眠／NREM睡眠と、睡眠深度などの詳細な睡眠状態を判定する状況認識技術を提案した。手首における脈波計測に適した波長が545nm程度（緑色光）であることを確認し、ノイズに強い独自の脈波間隔計測手法と合わせて手

首における安定した脈波計測を実現した点、個人差や日内変動の影響にロバストなクラスタリング手法を提案した点などが技術的なポイントである。

第4章「スマートフォンを用いたリアルタイム生活行動認識システム」では、スマートフォンに一般的に搭載されている加速度センサと、通話用のマイクを音センサとして積極的に活用し、特別なセンサ／デバイスを使用せずに、当該アプリケーションをインストールしたスマートフォンを持ち歩くだけで様々な日常生活行動を連続的に把握するための状況認識技術の研究を行った。まず加速度をもとにユーザの状態を「歩行」「作業」「安静」の3状態に大まかに推定し、「作業」と推定された場合に、マイクからの音を分析することにより作業内容を細かく分類する。また、生の認識結果だけでなく、同一作業区間単位で行動の種別を再評価する事例ベース認識手法の導入により認識性能が向上し、「歯磨き」「トイレ」「掃除機がけ」「皿洗い」など、これまで単一の加速度センサだけでは認識が難しかった様々な生活行動を精度良く認識することを可能にした点が特長である。

第5章「センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用」では、加速度センサと距離センサを内蔵したハンドヘルド型マイクを開発し、それに対するユーザの発話動作を検知する状況認識技術の研究を行った。両センサを組み合わせることにより、発話動作検出のロバスト化と低消費電力動作を実現し、使い方の側面からの音声認識精度向上を実現した。

いずれの研究においても、その目的に適したセンサを搭載したデバイスを開発／使用したが、表 6-1 に示す通り、すべての研究で加速度センサを使用した。加速度センサは状況認識技術のキーコンポーネントであると考えられるためである。

表 6-1 各研究におけるデバイス／アルゴリズムと使用センサ

章	デバイス／アルゴリズム	使用センサ
2	ウェアラブル健康管理システム	加速度、脈波、発汗、皮膚温
3	ウェアラブル睡眠センサ	加速度、脈波
4	スマートフォンを用いた生活行動認識	加速度、マイク、GPS
5	センサ駆動ハンドヘルド型音声認識入力	加速度、距離

## 状況認識技術のキーコンポーネント：加速度センサ

1章で述べた通り、本論文で研究した状況認識技術が認識の対象とする「状況」について、ユーザ自身の状況とその前後の文脈も重視する点を踏まえ、英訳には「Context」を用いた。高橋<sup>[70]</sup>による人物動作の分類を参考に、「Context」を人物の意図との関連性の観点で図 6-1 のように整理した。

「Context」は、人物の動作を伴うものと、動作とは独立した生体内部の生理状態に大別される。人物動作は、意図の強い意図的動作と、意図の弱い非意図的動作に分類できる。意図的動作を、図 1-2 を用いて説明すると、人物が状況を知覚、理解し、将来を予測する状況認識の結果生じるある明確な意図に基づいて行う、

ジェスチャ、機器操作、運動などの身体的な動作であると言える。非意図的動作は、状況認識の結果、意図的動作ほどは明確な意図が生じない場合にもしばしば発生する情動（一時的な感情の動き）に伴う動作（頭をかく／首をかしげるなどの動作、表情変化、発話など）や、条件反射的な動作である。一方、生理状態は、状況認識の結果に影響を受けるが、自分の意思でコントロールすることが難しい。

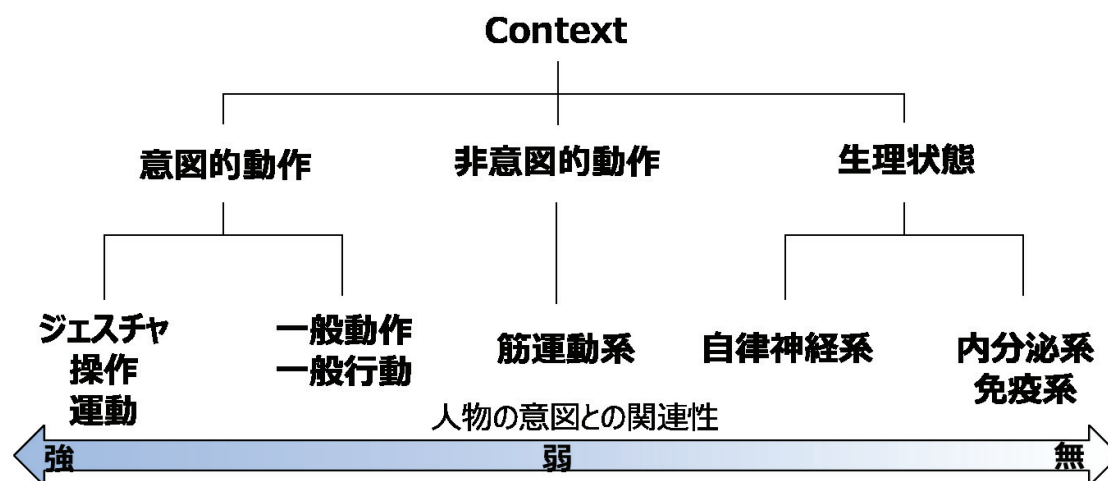


図 6-1 Context の分類（文献[70]を参考に作成）

状況認識の結果、強弱に関わらず何らかの意図が発生した場合、意図は力となって現れる。すなわち、その多くに身体的動作が伴う。加速度センサは文字通り加速度をセンシングするセンサである。装着部位のダイナミックな動作を加速度センサでセンシングすることにより、その動作を認識し、動作から人物の意図を推定できる可能性がある。

ただし、動作を認識する手法としては、カメラを用いた画像認識や、加速度センサ以外の物理量をセンシングするセンサの活用も可能である。ここで、加速度センサで動作を認識する利点について整理する。

## 加速度：場所の制約／リファレンスなしで計測可能な物理量

監視カメラの普及に伴い、カメラを用いた画像認識による人物動作認識に関する研究が多く行われている。センサを装着せずに、環境側に設置したカメラで認識可能な点が利点であるが、下記のような課題もある。

### <カメラ映像を用いた人物動作認識の課題>

#### ● 場所

カメラに対象人物が映らなければ認識できない。広範囲で認識可能にするためには複数のカメラの設置が必要になり、システムが大規模になる。データ量、計算量も課題となる。

- オクルージョン

カメラと対象人物の間に障害物があると認識できない。

- キャリブレーション

カメラ設置位置、画角／絞りなどの設定等、精度良く認識するためにはカメラと人物との距離なども考慮し、適切なキャリブレーションが必要になる。

- プライバシー

顔や服装から個人を特定される可能性があるため、プライバシーに十分に配慮したソリューションが求められる。

これに対し、身体に装着／携帯した装置に内蔵した加速度センサを用いるアプローチは、リファレンスなしで直接センシング可能な物理量である加速度を利用することで、場所によらず、どこでも対象人物の行動を把握することが可能である。ただし、日常生活で常に装置（センサ）を装着／携帯することは、少なからずユーザの負担になるため、その大きさ、重さ、使い勝手などとともに、ファッション性なども含めて検討する必要がある。特に、常に装置を装着／携帯してもらうためには、その負担を上回るユーザメリットを提供する必要がある点に留意しなければならない。ここで、加速度センサの進化も含めた概要を以下にまとめる<sup>[71]</sup>。

### ▶ 加速度センサ概要

加速度センサは、加速度（速度の時間微分）そのものの値を測定するものであるが、動きによる加速度の変化だけでなく、傾きによる重力加速度の変化も検出可能なセンサである。すなわち、動的な加速度から静的な加速度まで、検出可能な加速度の周波数帯域も幅広いため、様々な用途に利用されている。

加速度センサ黎明期（1980年代）は、科学実験用、地震計など、厳密な精度が要求される用途向けに開発されたが、その頃は機械式あるいは電気機械式などが主流で、小型化が困難であった。

その後、1990年代になり MEMS（Microelectromechanical Systems）技術の進化により MEMS 加速度センサが登場し、乗用車のエアバッグシステム用に MEMS 加速度センサが使われたことで普及が進んだ。ただし、当時は 1 つの MEMS 加速度センサでは、1 軸～2 軸しか検出できなかったため、他の用途向けに多軸方向の加速度検出が必要な場合は、複数の加速度センサを組み合わせる必要があった。実際に筆者がウェアラブル健康管理システム向けの最初のデバイス開発を行った 1990 年代後半は小型の MEMS 加速度センサは 2 軸までしか無かったため、3 軸化するために 2 軸の加速度センサを 2 個、向きを変えて基板に実装したこともあった。2000 年前半から 3 軸 MEMS 加速度センサが複数メーカーから発売されるようになると、エアバッグシステム以外の用途にも広く加速度センサ

が使用され始め、さらにスマートフォンの普及により一気に出荷数が増え、超小型低消費電力化、低価格化が加速した。現在では、サイズ：3mm×3mm×1.45mm、消費電力：300 $\mu$ A(typ)の3軸MEMS加速度センサ<sup>[72]</sup>が1.5米ドル程度の低価格で市販されている。

## 加速度センサ+ $\alpha$ による状況認識：クロスモーダル状況認識

加速度センサは、装着部位の動きによる加速度と、重力加速度をセンシング可能であるため、動作に加え、傾きなどの姿勢も把握可能であるが、1個の加速度センサのみでは、認識可能な対象も限定的である。よって、4章にも述べた通り、加速度センサを複数用いた研究として、例えば、加速度センサを左右の肩、肘、手首、腰、膝、足首に計12個装着した研究<sup>[37]</sup>や、上腕、手首、腰、太腿、足首に計5個装着した研究<sup>[38]</sup>などがあり、多いものでは20種類の動作状態の推定を行っている。複数の加速度センサを装着することで、詳細な動作状態の推定が可能となるが、日常生活でこれらのように多数の加速度センサを常時身につけることは、拘束性が高く受け入れがたい。よって、サービスの目的やユースケースに応じて、加速度センサに加えて適切なセンサ(+ $\alpha$ )を組み合わせるアプローチが有効である。ここで、複数のセンサ(様式)を使用することはマルチモーダル<sup>[63]</sup>やセンサフュージョン<sup>[73]</sup>と呼ばれるが、本論文では単に複数のセンサを使用するだけでなく、それらをより密に連携させることによる相乗効果に主眼を置いたアプローチとして「クロスモーダル状況認識」と呼ぶこととする。クロスモーダル状況認識により実現できる相乗効果として、次の2点が挙げられる。

### (1) 状況認識対象の詳細化、高精度化

1個の加速度センサのみでは認識が困難な動作や行動も、適切な+ $\alpha$ のセンサを組み合わせるクロスモーダル状況認識により認識可能となる。また、加速度センサのみの場合よりも状況認識精度、さらにはシステム全体としての精度を高めることができる。

第4章「スマートフォンを用いたリアルタイム生活行動認識システム」では、加速度センサに+ $\alpha$ としてマイクとGPSを用いることにより、加速度センサ単体の場合に比べて、生活行動を数多く、かつ精度良く認識することを可能にした。

第5章「センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用」では、加速度センサに+ $\alpha$ として距離センサを用いることで、発話動作の認識精度を向上させるとともに低消費電力動作を実現した。また、システム全体としての音声認識性能の向上に大きく寄与した。

## (2) 生理状態も合わせた状況認識

図 6-1 の動作とは独立した生体内部の生理状態に分類した自律神経系、内分泌系／免疫系などは、動きとしては捉えにくいいため、加速度センサで直接センシングすることは難しい（ただし、加速度センサの高感度化と信号処理技術の発展に伴い、手首に装着したデバイス内蔵の加速度センサで心拍と呼吸をセンシングすることも可能になりつつある<sup>[74]</sup>）。よって、サービスの目的やユースケースに応じて、加速度センサに生理状態をセンシング可能なセンサを組み合わせることによるクロスモーダル状況認識により、動作／行動と生理状態の両方を同時にセンシングすることが可能となる。

第 2 章「ウェアラブル健康管理システム」では、加速度センサに  $+ \alpha$  として、脈波センサ、発汗センサ、温度センサを組み合わせることにより、歩行、走行、歩数などの動作／行動と、脈拍数や精神状態などの生理状態の両方をセンシング可能にした。また、加速度センサ単体の場合や、加速度センサを用いない場合に精度良く認識することが困難であった食事イベントを、高精度かつリアルタイムに検出可能にした。

第 3 章「ウェアラブル睡眠センサとその応用」では、加速度センサに  $+ \alpha$  として脈波センサを組み合わせることにより、加速度センサ単体では困難であった REM 睡眠、NREM 睡眠、睡眠深度などの詳細な睡眠解析を実現した。

## 6. 1. 2 状況認識技術のヒューマンインタフェース

### ス応用

「コンピュータの存在を意識することなく、先進的なサービスを実現する」ために、上述したクロスモーダル状況認識によりコンピュータが我々人間と周囲の状況をリアルタイムに理解することを可能にした。さらに、その人のその時点の状況に応じて適切なサービスをタイムリーに提供するヒューマンインタフェース応用に取り組んだ。

第 2 章「ウェアラブル健康管理システム」では、リアルタイムに認識したユーザの動作／行動（特に食事の検出が特長）に基づき、食事のタイミングに基づいたタイムリーなサービス（食事メニュー入力支援、服薬管理、血糖値計測管理など）、生活習慣改善（行動変容）のためのアドバイスをタイムリーに提供するサービスなどの健康管理サービスを実現した。

第 3 章「ウェアラブル睡眠センサとその応用」では、リアルタイム睡眠解析システムを開発し、リアルタイムに認識した睡眠状態に基づき、快適に目覚められ

るタイムリーなタイミングでアラームを鳴動させる快適目覚まし機能を提案した。

第4章「スマートフォンを用いたリアルタイム生活行動認識システム」では、リアルタイムに様々な生活行動を認識することを可能にし、普段と違う生活行動を検知したタイムリーなタイミングでのさりげない問いかけや、ユーザの行動に基づき必要な情報をタイムリーに提供する機能の可能性を示した。

第5章「センサ駆動ハンドヘルド型音声認識入力手法と番組検索への応用」では、ユーザの発話動作をリアルタイムに認識し、音声認識トリガを自動でタイムリーに ON/OFF する機能を実現したことにより、使いやすさと音声認識精度を向上させた。

## **6. 2 今後の展望**

加速度センサと他のセンサを組み合わせることでユーザの日常生活を捉えようとする取り組みとしてライフ顕微鏡<sup>[75]</sup>がある。加速度センサ、脈波センサ、温度センサを搭載した腕時計型ウェアラブルセンサを常時装着し、これらのセンサデータを長期にわたって収集し可視化することで、ユーザ個人の生活リズムなどを把握可能にするものである。また、複数人で長期間データ収集したデータを分析することでさまざまな知見が得られ、ビジネス現場の分析などにも効果を上げている注目すべき取り組みである。

これに対し、本論文では、蓄積したデータを分析してさまざまな知見を得るだけでなく、「コンピュータの存在を意識することなく、先進的なサービスを実現する」ために、リアルタイムにユーザの状況を把握し、その状況に基づいて適切なサービスをタイムリーに提供するサービスの実現を目指して各章の研究に取り組んだ。つまり、加速度センサ+ $\alpha$ によるクロスモーダル状況認識により、加速度センサ単体では認識困難な状況の認識や認識精度の向上と、タイムリーなサービス提供を実現した点が本研究のポイントである。ただし、文字通り「コンピュータの存在を意識しない」レベルには未だ到達してはいない。

「コンピュータの存在を意識しない」ためには、装着型デバイスを用いるアプローチでは、デバイスをユーザが意識しないレベルに小型化し、「さりげないセンシング」を実現する必要がある。しかし、センシング精度を向上させたり、センシング対象を増やしたりしようとするすると、従来は大きさ、消費電力などの「さりげなさ」に影響する点が犠牲にならざるを得なかった。よって、図 6-2 のように、「さりげなさ」を維持しつつ、これらを解決する新たな技術が必要となる。

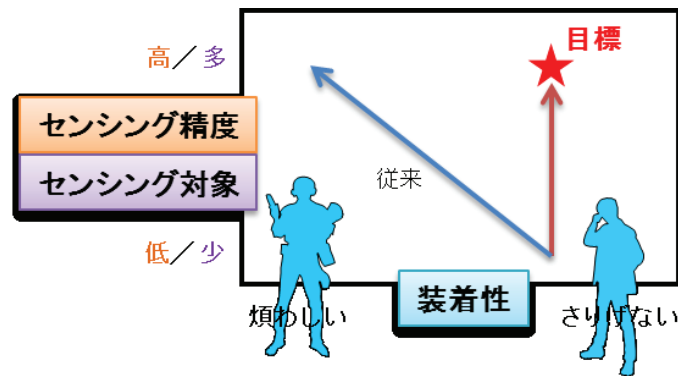


図 6-2 「さりげないセンシング」実現の課題

部品の超小型化、超低消費電力化、エネルギーハーベスティングなどのハードウェアに関する基盤技術の更なる進化が期待されるが、本論文で提案したクロスモーダル状況認識もそのための一つのアプローチであると考えられる。

さらに、使い方の面から「コンピュータの存在を意識しない」ためには、図 1-2 で述べた通り、現在の状況認識から、ユーザの意図を先読みし操作不要で所望のサービスを提供する先進的なサービスの実現が期待される。このためには従来の決定論的モデルによる **Reactive Computing**（起きたことに対するコンピューティング）から、確率論的モデルによる **Proactive Computing**（これから起きることを先読みするコンピューティング）への転換が必要になると考える。具体的には、ユーザの現在の状況だけでなく、過去のデータ、周辺環境、さらには社会全体の状況も含めた大量のデータに基づいたクロスモーダル状況認識により、将来予測の確度を向上させる今後のさらなる研究の進展が期待される。

引き続き「コンピュータの存在を意識することなく、先進的なサービスを実現する」ための研究開発とその成果の実用化に取り組み、社会に貢献していく所存である。



## 謝辞

---

本論文を作成するにあたり、主査の橋本周司教授から、本論文の構成、主張すべきポイントなどについて、さまざまな視点からの丁寧なご指導を賜りました。ここに感謝の意を表します。また、副査として論文内容に関するご指導をいただいた、中島啓幾教授、森島繁生教授、竹内淳教授に感謝いたします。

また、入社以来、一貫して本論文に関連する研究活動に従事することを支援して下さった株式会社東芝 研究開発センターの関係各位、一緒に研究開発、製品開発をさせていただいた先輩、同僚各位に深く感謝いたします。特に、本論文の主要業績として挙げた論文 5 編全てで、共著者として常に親身にご指導いただいた土井美和子氏に心より感謝申し上げます。

最後に、本論文の執筆を応援し、支えてくれた家族に感謝します。

## 参考文献

---

- [1] M. Weiser, The Computer for the 21st Century, Scientific American, 265 (3), pp.94-104, 1991.
- [2] 石井裕, ユビキタスの混迷の未来, ヒューマンインタフェース学会誌, Vol.4 No.3, pp.129-130, 2002.
- [3] K. Ashton, That 'Internet of Things' Thing, RFID Journal, 22 July 2009.
- [4] Overview of the Internet of Things, Recommendation ITU-T Y.2060, 06/2012.  
<http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060>
- [5] M. R. Endsley, Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64, 1995.
- [6] R. Pausch, and J. H. Leatherby, An empirical study: Adding voice input to a graphical editor, Journal of the American Voice Input/Output Society, Vol.9, pp.55-66, 1991.
- [7] 中洲俊信, 大平英貴, 池司, 大内一成, 岡田隆三, 自然な手振りによる直感的なハンドジェスチャ UI, IS3-06, 第 18 回画像センシングシンポジウム (SSII2012), 2012.
- [8] 厚生労働省 平成 20 年人口動態統計月報年計.
- [9] J. Tuomilehto, et al., Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance, New England Journal of Medicine, 344, pp.1343-1350, 2001.
- [10] マイライフ手帳,  
[http://www.humandatalab.com/hdl\\_service\\_mylife.html](http://www.humandatalab.com/hdl_service_mylife.html)
- [11] 三健人 2,  
<http://www.nttdata.com/jp/ja/news/release/2004/032400.html>
- [12] ボイス de 栄養士,  
<http://www.g-mark.org/award/describe/27461>
- [13] ライフコーダ,  
[http://www.suzuken.co.jp/product/healthcare/equip\\_product7.html](http://www.suzuken.co.jp/product/healthcare/equip_product7.html)
- [14] 美濃導彦, 家庭におけるユビキタス環境の構築ーゆかりプロジェクトー, 電子情報通信学会 2004 総合大会, No.A-16-8, pp.317, 2004.

- [15] 森武俊, 野口博史, 佐藤知正, センシングルームー部屋型日常行動計測蓄積環境 第 2 世代ロボティックルームー, 日本ロボット学会誌 Vol.23 No.6, pp.665-669, 2005.
- [16] 松岡克典, 宅内ネットワークを用いた生活見守り技術, 第 49 回自動制御連合講演会論文集 SU1-2-2, pp.1-2, 2006.
- [17] T. Martin, Issues in Wearable Computing for Medical Monitoring Applications: A Case Study of a Wearable ECG Monitoring Device, Proc. of ISWC2000, pp.43-49, 2000.
- [18] B. H. Yang, S. Rhee, and H. H. Asada, A Twenty-Four Hour Tele-Nursing System Using a Ring Sensor, Proc. of the 1998 IEEE International Conference on Robotics & Automation, pp.387-392, 1998.
- [19] Fitbit, <https://www.fitbit.com/jp>
- [20] Jawbone, <https://jawbone.com/>
- [21] J. Farrington, A. J. Moore, N. Tilbury, J. Church, and P. D. Biemond, Wearable Sensor Badge & Sensor Jacket for Context Awareness, Proc. of ISWC'99, pp.107-113, 1999.
- [22] K. V. Laerhoven, and O. Cakmakci, What Shall We Teach Our Pants?, Proc. of ISWC2000, pp.77-83, 2000.
- [23] J. Harris, and F. Benedict, A Biometric Study of Basal Metabolism in Man, Washington, Carnegie Institute, Publ. 279, pp.266, 1919.
- [24] 佐藤誠, 森田千絵, 土井美和子, 生体データと加速度センサを用いた行動認識, 情報処理学会第 65 回全国大会講演論文集, 3T5B-2, No.5, pp.239-242, 2004.
- [25] あすけん, <http://www.asken.jp/>
- [26] フードログ, <http://www.foodlog.jp/>
- [27] 2014 年 MSD 社調査,  
[https://www.msd.co.jp/static/pdf/product\\_20141106.pdf](https://www.msd.co.jp/static/pdf/product_20141106.pdf)
- [28] 2015 年 MSD 社調査,  
[https://www.msd.co.jp/static/pdf/product\\_20150831.pdf](https://www.msd.co.jp/static/pdf/product_20150831.pdf)
- [29] A. Rechtschaffen, and A. Kales, A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, Washington D.C. : Public Health Service, U.S. Government Printing Office, 1968.
- [30] A. Sadeh, J. Alster, D. Urbach, and P. Lavie, Actigraphically based automatic bedtime sleep-wake monitor scoring: validity and clinical applications, J Ambul Monit 2, pp. 209-216, 1989.

- [31] 道盛章弘, 福島薫, 萩原啓, 心拍変動解析による睡眠モニタリングシステム (特集 高齢化対応技術), 松下電工技報 (82), pp.29-33, 2003.
- [32] T. Watanabe, and K. Watanabe, Noncontact Method for Sleep Stage Estimation, IEEE Transactions on Biomedical Engineering, Vol.51, No.10, pp.1735-1748, 2004.
- [33] Y. Wakuda, A. Noda, Y. Hasegawa, F. Arai, T. Fukuda, and M. Kawaguchi, Biological Rhythm Based Wearable Sleep State Observer, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.11, No.2, pp.232-241, 2007.
- [34] R. J. Cole, et al., Automatic sleep/wake identification from wrist actigraphy, Sleep 1992, 15 461-469, 1992.
- [35] アクティグラフ, <http://www.sanita.co.jp/actigraph/>
- [36] T. Suzuki, K. Ouchi, K. Kameyama, and M. Takahashi, Development of a Sleep Monitoring System with Wearable Vital Sensor for Home Use, Proceedings of the International Conference on Biomedical Electronics and Devices (BIODEVICES 2009), pp.326-331, 2009.
- [37] N. Kern, B. Schiele, and A. Schmidt, Multi-sensor Activity Context Detection for Wearable Computing, Proceedings of the First European Symposium on Ambient Intelligence (EUSAI 2003), LNCS 2875, pp.220-232, 2003.
- [38] L. Bao, and S. S. Intille, Activity Recognition from User-Annotated Acceleration Data, Proceedings of the Second International Conference on Pervasive Computing (PERVASIVE 2004), LNCS 3001, pp.1-17, 2004.
- [39] T. Iso, and K. Yamazaki, Gait Analyzer based on a Cell Phone with a Single Three-axis Accelerometer, Proceedings of the 8th conference on Human-computer interaction with mobile devices and services (MobileHCI2006), pp.141-144, 2006.
- [40] 倉沢央, 川原圭博, 森川博之, 青山友紀, センサ装着場所を考慮した 3 軸加速度センサを用いた姿勢推定手法, 情報処理学会 研究報告, 2006-UBI-11 (3), pp.15-22, 2006.
- [41] 池谷直紀, 菊池匡晃, 長健太, 服部正典, 3 軸加速度センサを用いた移動状況推定方式, 情報処理学会 研究報告, 2008-UBI-19 (14), pp.75-80, 2008.
- [42] V. Peltonen, J. Tuomi, A. Klapuri, J. Huopaniemi, and T. Sorsa: Computational auditory scene recognition. Proceedings of

- Acoustics, Speech, and Signal Processing (ICASSP2002), pp.1941–1944, 2002.
- [43] 小林重令, 岩本健嗣, 西山智, 釈迦: 携帯電話を用いたユーザ移動状態推定方式, 情報処理学会論文誌 Vol.50 No.1, pp.193-208 (2009).
- [44] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger, J. Shaffer, and F. Wong, SenSay: A Context-Aware Mobile Phone, Poster of 7th IEEE International Symposium on Wearable Computers (ISWC 2003), 2003.
- [45] T. Iso, N. Kawasaki, and S. Kurakake, Personal context extractor with multiple sensor on a cell phone, 7th IFIP International Conference on Mobile and Wireless Communications Networks (MWCN2005), D. 2 C200525, 2005.
- [46] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner, Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers, Proceedings of the Eighth International Conference on Pervasive Computing (PERVASIVE 2004), LNCS 2001/2004, pp.18-32, 2004.
- [47] 前川卓也, 柳沢豊, 岸野泰恵, 石黒勝彦, 亀井剛次, 櫻井保志, 岡留剛, ウェアラブルセンサによるモノを用いた行動の認識について, 情報処理学会 研究報告, 2010-UBI-25 (57), pp.1-8, 2010.
- [48] T. Maekawa, Y. Yanagisawa, Y. Kishino, K. Ishiguro, K. Kamei, Y. Sakurai, and T. Okadome, Object-Based Activity Recognition with Heterogeneous Sensors on Wrist, Proceedings of the Eighth International Conference on Pervasive Computing (PERVASIVE 2010), LNCS 6030/2010, pp.246-264, 2010.
- [49] S. Katz, Assessing self-maintenance, activities of daily living, mobility, and instrumental activities of daily living, Journal of the American Geriatrics Society 31(12), pp.721-727, 1983.
- [50] G. G. Fillenbaum, Screening the elderly. A brief instrumental activities of daily living measure, Journal of the American Geriatrics Society 33(10), pp.698-706, 1985.
- [51] 永田久美子, 認知症ケアをもっと“楽”に!, 認知症介護研究・研修東京センター監修・発行.
- [52] 認知症介護研究・研修東京センター ケアマネジメント推進室: 「いつでもどこ」ネット, <http://itsu-doko.net/index.html>
- [53] 総務省統計局 平成 18 年社会生活基本調査, <http://www.stat.go.jp/data/shakai/2006/index.htm>

- [54] C-C Chang, and C-J Lin, LIBSVM: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [55] K. Cho, N. Iketani, H. Setoguchi, and M. Hattori, Human Activity Recognizer for Mobile Devices with Multiple Sensors, uic-atc, Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp.114-119, 2009.
- [56] Y. Kobayashi, et al., Design Targeting Voice Interface Robot Capable of Active Listening, Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010), pp.161-162, 2010.
- [57] 北岡教英, 角谷直子, 中川聖一, 音声対話システムの誤認識に対するユーザの繰返し訂正発話の検出と認識, 電子情報通信学会論文誌 D-II, Vol. J87-D-II, No.7, pp.1441-1450, 2004.
- [58] L. R. Karl, M. Pettey, and B. Shneiderman, Speech versus mouse commands for word processing: an empirical evaluation, International Journal of Man-Machine Studies, Vol.39, Issue 4, pp.667-687, 1993.
- [59] 財団法人共用品推進機構, 高齢者の家庭内での不便さ調査報告書, 1999.
- [60] 原紀代, 志田武彦, 中俊弥, 南部美砂子, 原田悦子, 家電操作における高齢者の認知特性の研究, Matsushita Technical Journal, Vol.51, No.4, pp.29-33, 2005.
- [61] 金澤博史, 友田一郎, 高島由彰, 竹林洋一, ユビキタス社会に向けた Bluetooth ヘッドセットの開発, 日本音響学会 2002 年春季研究発表会, pp.219-220. 2002.
- [62] 新田恒雄, GUI からマルチモーダル UI (MUI) に向けて, 情報処理, Vol.36, No.11, pp.1039-1046, 1995.
- [63] M. Kawamoto, F. Asano, H. Asoh, and K. Yamamoto, Particle Filtering Algorithm for Tracking Sound Sources Using Microphone Arrays, Proceedings of ICASSP 2007, Vol. I, pp.129-132, 2007.
- [64] W. Herbordt, T. Horiuchi, M. Fujimoto, T. Jitsuhiro, and S. Nakamura, Noise-robust hands-free speech recognition and communication on PDAs using microphone array technology, IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 307-310, 2005.

- [65] 伊藤克亘, 藤井敦, NTCIR-3 ワークショップにおける音声入力型ウェブ検索タスク, 情報処理学会研究報告－音声言語情報処理 (2002-SLP-043), pp.25-32, 2002.
- [66] 榎将功, 皇甫美華, 大田健紘, 柳田益造, 日本語における略語自動生成法の検討とその音声インタフェースへの応用, 情報処理学会研究報告－音声言語情報処理 (2007-SLP-069), pp.313-318, 2007.
- [67] 勝丸真樹, 駒谷和範, 尾形哲也, 奥乃博, 音声対話システムにおける簡略表現認識のための誤認識増加を抑制する自動語彙拡張, 情報処理学会研究報告－音声言語情報処理 (2008-SLP-071), pp.71-76, 2008.
- [68] 新田恒雄, GUI からマルチモーダル UI (MUI) に向けて, 情報処理, Vol.36, No.11, pp.1039-1046, 1995.
- [69] 若木裕美, 藤井寛子, 福井美佳, 住田一男, Web 情報を用いた人物の愛称抽出, 日本データベース学会論文誌 Vol.7, No.1, pp.169-174, 2008.
- [70] 高橋正樹, 映像解析による人物動作理解に関する研究, 総合研究大学院大学平成 23 年度博士論文, 2011.
- [71] 平成 21 年度特許出願技術動向調査報告書 加速度センサ (要約版), 特許庁, 2010.
- [72] ADXL337, Analog Devices Inc., <http://www.analog.com/jp/products/mems/accelerometers/adxl337.html>
- [73] 石川正俊, センサフュージョンの課題, 日本ロボット学会誌, 8 巻 6 号, pp.95-102, 1990.
- [74] J. Hernandez, D. McDuff, and R. Picard, BioWatch: Estimation of Heart and Breathing Rates from Wrist Motions, Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '15), pp.169-176, 2015.
- [75] 鈴木敬, 栗山裕之, 山下春造, 矢野和男, ライフ顕微鏡: 20 人のライフタペストリーが語る人とセンサと IT の未来, 日立評論 Vol.89 No.12 914-915, pp.30-35, 2007.

# 著者の業績

---

## 論文

K. Ouchi, T. Suzuki, and M. Doi

LifeMinder: A wearable healthcare support system with timely instruction based on the user's context, IEICE Transactions on Information and Systems, Vol.E87-D, No.6, pp.1361-1369, June 2004.

大内一成, 土井美和子

センサ駆動ハンドヘルド型音声認識入力方法の提案 – センサを用いた発話動作検出手法, 情報処理学会論文誌 Vol.51 No.2, pp.324-333, 2010年2月.

大内一成, 若木裕美, 屋野武秀, 住田一男, 土井美和子

人名と番組名の言い換えに対応する音声認識インタフェース, 情報処理学会論文誌 Vol.51 No.3, pp.846-855, 2010年3月.

K. Ouchi, M. Doi

Living activity recognition using off-the-shelf sensors on mobile phones, annals of telecommunications, Vol.67, pp.387-395, June 2012.

大内一成, 土井美和子

携帯電話搭載センサによるリアルタイム生活行動認識システム, 情報処理学会論文誌 Vol.53 No.7, pp.1675-1689, 2012年7月.

(共著論文) 森屋彰久, 鈴木琢治, 大内一成, 亀山研一

脈波情報を用いた睡眠状態推定とその応用, ヒューマンインタフェース学会論文誌 Vol.10, No.2, pp.207-214, 2008年5月. (受賞4)



## 国際会議（査読有、筆頭）

K. Ouchi, Y. Ohmori, S. Matsushita, and M. Doi

A Wearable Authoring System Using Organized Multimedia Data, Extended Abstracts of CHI 2000, pp. 109-110, April 2000.

K. Ouchi, T. Suzuki, and M. Doi

LifeMinder: A wearable healthcare support system using user's context, Proceedings of ICDCS2002 Workshops (International Workshop on Smart Appliances and Wearable Computing, IWSAWC2002), pp.791-792, 2002.

K. Ouchi, T. Suzuki, and M. Doi

LifeMinder: A Wearable Healthcare Support System with Timely Instruction Based on the User's Context, Proceedings of the 8th IEEE International Workshop on Advanced Motion Control (AMC2004), pp.445-450, 2004.

K. Ouchi, N. Esaka, Y. Tamura, M. Hirahara, and M. Doi

MagicWand: An Intuitive Gesture Remote Control for Home Appliances, Proceedings of the 2005 International Conference on Active Media Technology (AMT2005), pp.274, 2005.

K. Ouchi, T. Suzuki, and K. Kameyama

A Wearable Healthcare Support System Combined with a Cell Phone for Use in Daily Life, The 3rd Annual International Conference on Mobile and Ubiquitous Systems: Networks and Services (MOBIQUITOUS 2006), 2006.

K. Ouchi, T. Suzuki, and K. Kameyama

Development Of A Less Restrictive PPI Measuring System For Monitoring Of Autonomic Nervous Activity, Proceedings of BMES2007 Fall Meeting, 2007.

K. Ouchi, T. Suzuki, and K. Kameyama

Development of a wearable sleep sensor to monitor sleep states, Adjunct Proceedings of Ubicomp 2008, pp.43-44, 2008.

K. Ouchi, K. Imoto, and M. Doi

A Feasibility Study on Indoor Location Awareness by Collaborating with a Microphone and an Accelerometer, Proceedings of the International Symposium on Ubiquitous Computing Systems (UCS2009), 2009.

K. Ouchi, M. Doi

A Real-Time Living Activity Recognition System Using Off-the-Shelf Sensors on a Mobile Phone, Proceedings of CONTEXT 2011, LNAI 6967, pp.226-232, 2011.

K. Ouchi, M. Doi

Indoor-Outdoor Activity Recognition by a Smartphone, Adjunct Proceedings of Ubicomp 2012, pp.600-601, 2012.

K. Ouchi, M. Doi

T-SARCAS: Two-Step Activity Recognition Combined with Acceleration and Sound for Smartphones and its Practical Evaluation, Proceedings of the 2nd International Workshop on Mobile Computing Platforms and Technologies (MCPT-2013), 2013.

K. Ouchi, M. Doi

Smartphone-based Monitoring System for Activities of Daily Living for Elderly People and Their Relatives Etc., Adjunct Proceedings of Ubicomp 2013, pp.103-106, 2013.

## 国内発表（筆頭）

大内一成, 大盛善啓, 松下宗一郎, 土井美和子

情報獲得機能を持つウェアラブルオーサリングシステム, インタラクション 2000 論文集, pp.45-46, February 2000.

大内一成，大盛善啓，松下宗一郎，土井美和子  
ウェアラブルオーサリングシステム，第 11 回ヒューマンインタフェース研  
究会「ウェアラブル／アウトドアコンピューティング」研究報告集，  
pp.41-46, February 2001.

大内一成，大盛善啓，松下宗一郎，土井美和子  
ウェアラブルオーサリングシステム，マルチメディア，分散，協調とモバ  
イル(DICOMO2001)シンポジウム 論文集，2001. (受賞 1)

大内一成，鈴木琢治，土井美和子  
LifeMinder: 状況認識を用いたウェアラブル健康管理システム，第 3 回シ  
ステムインテグレーション部門講演会 (SI2002), 1P91-02, 2002. (受賞 2)

大内一成，鈴木琢治，土井美和子  
LifeMinder: ウェアラブル健康管理システムにおける腕時計型生体センサ  
の高機能化／低消費電力化，インタラクシオン 2003 論文集，pp.241-242,  
2003.

大内一成，鈴木琢治，亀山研一  
生体情報を利用したウェアラブル家電リモコン，ヒューマンインタフェー  
スシンポジウム 2003, 2003.

大内一成，鈴木琢治，森屋彰久，亀山研一  
ウェアラブル機器を用いたヘルスケアサービス，情報処理学会研究報告  
2007-UBI-13, pp.29-36, 2007. (受賞 3)

大内一成，若木裕美，屋野武秀，住田一男，土井美和子  
複数のセンサと自然言語処理技術による使いやすい音声入力インタフェー  
ス，マルチメディア，分散，協調とモバイル(DICOMO2009)シンポジウム  
論文集，pp.1804-1814, 2009. (受賞 5)

大内一成，土井美和子  
加速度と音による家庭内ユーザ状況認識の可能性検討，マルチメディア，  
分散，協調とモバイル(DICOMO2010)シンポジウム 論文集，pp.508-515,  
2010.

大内一成，土井美和子

加速度と音で日々の生活行動を認識する ActivityAnalyzer，インタラクシ  
ョン 2011 論文集，1CR3-7，2011. (受賞 6)

大内一成，土井美和子

携帯電話搭載センサによるリアルタイム生活行動認識システム，情報処理  
学会研究報告，Vol.2011-UBI-30 No.3, pp.1-8, 2011. (受賞 8, 受賞 9)

大内一成，土井美和子

スマートフォンによる屋内外生活行動センシング，マルチメディア，分散，  
協調とモバイル(DICOMO2012)シンポジウム 論文集，pp.173-179，2012.  
(受賞 7)

大内一成

ウェアラブルセンサ開発と産業応用に向けた実証実験，情報処理学会高齢  
社会デザイン研究会，Vol.2016-ASD-4 No.8, pp.1-4, 2016.

## **解説記事**

大内一成，若木裕美

日常的使用を目指した音声入力インタフェース，情報処理 Vol.51 No.1,  
pp.22-25, 2010 年 1 月.

大内一成

実世界に広がる装着型センサを用いた行動センシングとその応用：4. スマ  
ートフォンを用いた生活行動認識 -家の中も外もスマホで行動認識-，情報  
処理 Vol.54 No.6, pp.578-581, 2013 年 6 月.

大内一成

ウェアラブル/ユビキタス環境におけるパターン計測技術，計測と制御  
53(7), 579-585, 2014 年 7 月.

大内一成

介護予防のためのウェアラブルセンシング技術，IT VISION No.32 ウェア  
ラブルで変わるヘルスケアの現場，2015 年 7 月.

## 講演

人間行動センシングに関する東芝のこれまでの取り組みについて，第1回人間行動センシングシンポジウム (HASC01)，名古屋大学ベンチャービジネスラボラトリ，2009年10月7日，

人間情報計測による日常生活見守り技術，東京大学山上会館，人間情報学会講演会，2010年6月16日．

行動センシングと健康支援，電気四学会 関西支部 専門講習会「行動センシングとその応用」，中央電気倶楽部，2011年11月11日．

日常生活見守り技術，みんなのエネルギーフェスタ 2012「エネルギーを着る、そしてコンピュータを着る」，秋葉原 UDX，2012年8月12日．

マルチセンサを活用した日常生活見守り技術，第38回センサ&アクチュエータ技術シンポジウム「センサ技術最前線 2012」，化学会館，2012年9月21日．

人間行動センシングとそれを活用した UI，精密工学会 画像応用技術専門委員会 2012年第4回定例講演会「人の計測」，東京電機大学，2012年11月16日．

スマートフォンを用いた生活行動認識技術，経団連 21世紀政策研究所「ビッグデータビジネスが描く未来」研究会，経団連会館，2013年10月28日．

ウェアラブルセンシング技術のヘルスケア応用，情報処理学会連続セミナー2014，第2回：ウェアラブルが切り開く IT 新潮流，化学会館，2014年7月15日．

日常人間ドックを実現するウェアラブルセンサとライフログモニタリング，第3回生命医薬情報学連合大会 (IIBMP2014)，COI TOHOKU「センサー・オブ・イノベーション東北拠点の目指す社会」，仙台国際センター，2014年10月3日．

東芝のウェアラブルセンシング技術，情報処理学会ユビキタスコンピューティングシステム研究会特別講演，沖縄科学技術大学院大学，2014年10月13日．

東芝のウェアラブルセンシング技術，特定非営利活動法人ウェアラブルコンピュータ研究開発機構 HMD ミーティング，早稲田大学，2014年10月29日．

ウェアラブルにおけるデータ分析の必要性和今後の課題，WebDB Forum 2014 特別セッション，芝浦工業大学 豊洲キャンパス，2014年11月20日．

東芝の健康センシングに関する取り組み，ビジョン技術の実利用ワークショップ 2014(ViEW2014)，介護だけじゃない！健康センシング（特別企画セッション），パシフィコ横浜，2014年12月5日．

東芝のウェアラブルセンシング技術とヘルスケアサービス，国際モダンホスピタルショウ 2015，コーナー展示者セミナー，東京ビッグサイト，2015年7月17日．

ヘルスケア向けウェアラブルデバイス開発，北陸経済連合会講演会「次世代ヘルスケア産業の創出」，金沢都ホテル，2015年12月21日．

東芝のヘルスケア向けウェアラブルデバイス開発，ウェアラブル EXPO2016 専門セッション「医療現場 ヘルスケアで活用されるウェアラブル」東京ビッグサイト，2016年1月15日．

東芝のヘルスケア向けウェアラブルデバイス，第2回ウェルネスライフサポートフォーラム，ソラシティ カンファレンスセンター，2016年2月18日．

ヘルスケアの30年，電子情報通信学会 2016年総合大会，特別企画「2046年からの招待講演者」，九州大学，2016年3月16日．

## 特許

特許第 5646146 号など国内出願 88 件． 米国出願 48 件．

## 受賞

(受賞 1) 情報処理学会 DICOMO2001 ベストプレゼンテーション賞，2001 年 6 月 29 日．

(受賞 2) 計測自動制御学会 SI2002 ベストセッション賞，2002 年 12 月 19 日．

(受賞 3) 情報処理学会 ユビキタスコンピューティングシステム研究会優秀論文賞，2007 年 11 月 30 日．

(受賞 4) ヒューマンインタフェース学会 論文賞，2009 年 3 月 2 日．

(受賞 5) 情報処理学会 DICOMO2009 シンポジウム優秀論文賞，2009 年 8 月 27 日．

(受賞 6) 情報処理学会 インタラクション 2011 インタラクティブ発表賞 (PC 推薦)，2011 年 3 月 10 日．

(受賞 7) 情報処理学会 DICOMO2012 シンポジウム最優秀プレゼンテーション賞，2012 年 7 月 6 日．

(受賞 8) 情報処理学会 ユビキタスコンピューティングシステム研究会優秀論文賞，2012 年 11 月 1 日．

(受賞 9) 情報処理学会 山下記念研究賞，2013 年 3 月 6 日．

(受賞 10) 情報処理学会 長尾真記念特別賞，2013 年 6 月 5 日．