

修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報理工学研究科 情報・通信工学専攻 博士前期課程		
氏 名	齋藤 央	学籍番号	1131048
論 文 題 目	単発音を利用したハンドジェスチャインタラクションの提案と評価		
要 旨	<p>本研究では、ハンドジェスチャと単発音を組み合わせた“音付ハンドジェスチャ”を提案し、評価実験によりその有用性を評価した。</p> <p>近年、空中でのハンドジェスチャを利用したシステムが増えている。しかし、ユーザの動作がジェスチャとして意図されたものなのか判別しづらいといった問題がある。この問題を解決するためジェスチャ開始トリガーや終了トリガーに特別な動作を用いるものがあるが、1ステップ余計な動作が必要になるため効率的な操作とは言えない。また、即時的な操作には、直線的でシンプルなジェスチャの利用が望ましい。しかし、操作の種類数の増加に応じてジェスチャの種類数を増やすために複雑なジェスチャを導入する必要もでてくる。</p> <p>そこで、ジェスチャの明確化、シンプルなジェスチャ数の増加を図った音付ハンドジェスチャを提案した。音付ハンドジェスチャは、ハンドジェスチャと単発音を組み合わせた手法であり、単発音とハンドジェスチャを同時に認識した際にジェスチャとして認識させるようにすることで、ジェスチャの明確化を実現した。また、複数種の単発音の分類を行い、同様のハンドジェスチャを別のジェスチャとして認識させることで、シンプルなジェスチャ数の増加を図った。単発音の検出には周波数毎の振幅における特徴を用い、これによって複数種の単発音の分類も行った。ジェスチャ認識には深度カメラを用いた。</p> <p>音付ハンドジェスチャの有用性を評価するために、評価実験を行った。4種類の単発音と8種類のハンドジェスチャを組み合わせた14種類の音付ハンドジェスチャに対し、認識精度を調査した。認識精度は66%と低い値となったが、単発音の種類を2種類と少なくしたところ、92%と高い精度で認識可能であることが分かった。この場合、音付ハンドジェスチャの種類数が9となり、大幅なジェスチャ数の増加は達成できなかったが、単発音の分類方法の改善により達成可能であると考えられる。既存手法との比較実験からは、音付ハンドジェスチャが効率的に動作可能であることが分かり、有用性が示された。</p> <p>また、深度カメラ1台での広角化を実現した“回転台座”を実装した。評価実験により十分な速度でユーザの検出、追跡が可能である事が分かり、ハンドジェスチャの問題点のひとつである計測可能範囲の制限を解決することができた。</p>		

平成 24 年度修士論文

単発音を利用した
ハンドジェスチャインタラクションの提案と評価

電気通信大学大学院 情報理工学研究科
情報・通信工学専攻 コンピュータサイエンスコース

角田 H I* 研究室

指導教員：角田 博保 (*Kakuda Hiroyasu*)
岩田 茂樹 (*Shigeki Iwata*)

学籍番号：1131048 / 齋藤 央 (*Saito Hisashi*)

提出日：平成 25 年 1 月 25 日 (金)

概要

本研究では、ハンドジェスチャと単発音を組み合わせた“音付ハンドジェスチャ”を提案し、評価実験によりその有用性を評価した。

近年、空中でのハンドジェスチャを利用したシステムが増えている。しかし、ユーザの動作がジェスチャとして意図されたものなのか判別しづらいといった問題がある。この問題を解決するためジェスチャ開始トリガーや終了トリガーに特別な動作を用いるものがあるが、1ステップ余計な動作が必要になるため効率的な操作とは言えない。また、即時的な操作には、直線的でシンプルなジェスチャの利用が望ましい。しかし、操作の種類数の増加に応じてジェスチャの種類数を増やすために複雑なジェスチャを導入する必要もでてくる。

そこで、ジェスチャの明確化、シンプルなジェスチャ数の増加を図った音付ハンドジェスチャを提案した。音付ハンドジェスチャは、ハンドジェスチャと単発音を組み合わせた手法であり、単発音とハンドジェスチャを同時に認識した際にジェスチャとして認識させるようにすることで、ジェスチャの明確化を実現した。また、複数種の単発音の分類を行い、同様のハンドジェスチャを別のジェスチャとして認識させることで、シンプルなジェスチャ数の増加を図った。単発音の検出には周波数毎の振幅における特徴を用い、これによって複数種の単発音の分類も行った。ジェスチャ認識には深度カメラを用いた。

音付ハンドジェスチャの有用性を評価するために、評価実験を行った。4種類の単発音と8種類のハンドジェスチャを組み合わせた14種類の音付ハンドジェスチャに対し、認識精度を調査した。認識精度は66%と低い値となったが、単発音の種類を2種類と少なくしたところ、92%と高い精度で認識可能であることが分かった。この場合、音付ハンドジェスチャの種類数が9となり、大幅なジェスチャ数の増加は達成できなかったが、単発音の分類方法の改善により達成可能であると考えられる。既存手法との比較実験からは、音付ハンドジェスチャが効率的に動作可能であることが分かり、有用性が示された。

また、深度カメラ1台での広角化を実現した“回転台座”を実装した。評価実験により十分な速度でユーザの検出、追跡が可能である事が分かり、ハンドジェスチャの問題点のひとつである計測可能範囲の制限を解決することができた。

目次

第 1 章	はじめに	1
1.1	空中でのハンドジェスチャを利用したシステムの増加	1
1.2	ハンドジェスチャの問題点	1
1.3	本研究の位置づけ	2
1.4	本論文の構成	3
第 2 章	関連研究	4
2.1	ハンドジェスチャ	4
2.2	音の認識	5
2.3	マルチモーダル（ハンドジェスチャと音の組み合わせ）	6
2.4	カメラの広角化	6
第 3 章	提案手法“音付ジェスチャ”	8
3.1	設計方針	8
3.2	音付ジェスチャの認識方法	8
3.2.1	単発音認識方法	9
3.2.2	ハンドジェスチャ認識方法	9
3.3	音付ジェスチャの利点	10
3.4	応用先	10
第 4 章	予備調査	12
4.1	概要	12
4.1.1	目的	12
4.1.2	被験者	12
4.2	タスク	12
4.3	実験環境	13
4.4	結果：音の大きさ	13
4.5	結果：アンケート	15
4.6	考察	16
第 5 章	単発音認識方法の調査	18
5.1	単発音データの新規採集	18
5.1.1	被験者	18
5.1.2	実験システム	18
5.1.3	実験環境	19
5.2	単発音の検出方法の検証	19
5.2.1	単発音の検出方法	19
5.2.2	閾値の決定	20
5.3	単発音の分類方法の検証	21
5.3.1	単発音の分類方法	21
5.3.2	解析方法	21

5.3.3	解析結果と考察	21
5.3.4	重み付けと正規化	22
5.3.5	重み付けと正規化を行った際の分類精度	23
5.4	単発音の種類数による分類精度	24
5.4.1	単発音を3種類とした場合	24
5.4.2	単発音を2種類とした場合	26
5.5	アンケート結果	28
第6章	音付ジェスチャ認識実験	29
6.1	概要	29
6.1.1	被験者	30
6.2	実験システム	30
6.2.1	実験画面	30
6.3	実験環境	31
第7章	音付ジェスチャ認識実験結果と考察	32
7.1	音付ジェスチャ認識率	32
7.2	音付ジェスチャ失敗回数	33
7.3	アンケート結果	33
第8章	比較実験	35
8.1	概要	35
8.1.1	被験者	35
8.2	実験システム	35
8.2.1	実験画面	35
8.3	実験環境	36
8.4	ジェスチャの割り当て	37
8.4.1	タスク	38
第9章	比較実験結果と考察	40
9.1	タスク完了時間	40
9.2	アンケート結果	40
第10章	提案手法“回転台座”	42
10.1	設計方針	42
10.2	機能	42
10.2.1	ユーザの検出	42
10.2.2	ユーザの追跡	43
10.3	性能評価実験：概要	43
10.3.1	実験環境	44
10.4	結果	44
10.5	考察	45
第11章	おわりに	46
	参考文献	47
	謝辞	49

目次

1.1	SAMSUNG の Smart TV	1
1.2	東芝のてぶらナビ	1
1.3	意図しない動きの誤認識の例	1
1.4	開始ジェスチャに拳を用いた例	1
1.5	計測可能範囲の制限	2
1.6	音付ジェスチャ	3
2.1	Mistry らの研究	4
2.2	Niels らの研究	5
2.3	長谷川らの研究	5
2.4	Chris らの研究	6
2.5	尾崎らの研究	6
2.6	David らの研究	6
2.7	Wilson らの研究	7
2.8	Tomari らの研究	7
3.1	音付ジェスチャの例	8
3.2	Kinect	8
3.3	音付ジェスチャの認識アルゴリズム	9
3.4	DTW によるジェスチャ認識 (長谷川らの研究)	9
3.5	連続ジェスチャの比較	10
4.1	24 種の音付ジェスチャ	13
4.2	音付ジェスチャの最大振幅	14
4.3	音付ジェスチャと環境音の最大振幅の比較	15
4.4	単発音の最大振幅と環境音の平均振幅の比較	15
4.5	アンケート結果:動作のしやすさ (左の値), 実際の機器に使用したいもの (右の値)	16
5.1	データ収集中の画面	19
5.2	単発音の継続時間 (横軸: 時間, 縦軸: 振幅)	19
5.3	各被験者の最大振幅÷平均振幅の値	20
5.4	各被験者の最大振幅÷最小振幅の値	20
5.5	各被験者の平均振幅÷最小振幅の値	20
5.6	安定した音が出せたか	28
6.1	音付ジェスチャ (クラブ, 机タップ, 机ノック使用)	29
6.2	音付ジェスチャ (スナップ使用)	30
6.3	実験中の画面	31
6.4	実験環境 1	31

6.5	実験環境 2	31
7.1	音付ジェスチャの認識精度 (%) : 単発音を 4 種類とした場合	32
7.2	音付ジェスチャの認識精度 (%) : 単発音を 2 種類とした場合	32
7.3	単発音の種類毎の認識精度 (%)	33
7.4	音付ジェスチャの失敗回数 (回)	33
7.5	認識のされやすさ	34
7.6	動作のしやすさ	34
7.7	疲労度	34
8.1	実験画面 : 音付ジェスチャ	36
8.2	実験画面 : ジェスチャのみ	36
8.3	実験風景	37
8.4	操作方法 : 実験環境	37
8.5	操作方法 : 音付ジェスチャ	38
8.6	操作方法 : ジェスチャのみ	38
9.1	タスク完了時間	40
9.2	アンケート結果	41
10.1	Kinect を回転させる台座の構成図	42
10.2	ユーザの検出	43
10.3	ユーザの追跡	43
10.4	Kinect を回転させる台座の完成図	44
10.5	実験装置	44
10.6	Kinect の水平視野角	44
10.7	音源の位置	44

表 目 次

4.1	動作のしやすさ	16
4.2	実際の機器操作にしたいもの	16
5.1	分類の内訳 (全バンド)	22
5.2	分類の内訳 (帯域数 24)	22
5.3	分類の内訳 (帯域数 12)	22
5.4	分類の内訳 (帯域数 8)	22
5.5	正規化, 重み付け後の分類の内訳 (全バンド)	23
5.6	正規化, 重み付け後の分類の内訳 (帯域数 24)	23
5.7	正規化, 重み付け後の分類の内訳 (帯域数 12)	24
5.8	正規化, 重み付け後の分類の内訳 (帯域数 8)	24
5.9	分類の内訳 (クラブ, 机タップ, 机ノック)	25
5.10	分類の内訳 (クラブ, 机タップ, スナップ)	25
5.11	分類の内訳 (クラブ, 机ノック, スナップ)	26
5.12	分類の内訳 (机タップ, 机ノック, スナップ)	26
5.13	分類の内訳 (クラブ, 机タップ)	27
5.14	分類の内訳 (クラブ, 机ノック)	27
5.15	分類の内訳 (クラブ, スナップ)	27
5.16	分類の内訳 (机タップ, 机ノック)	27
5.17	分類の内訳 (机タップ, スナップ)	28
5.18	分類の内訳 (机ノック, スナップ)	28
10.1	ユーザ検出時間 (s)	44

第1章 はじめに

1.1 空中でのハンドジェスチャを利用したシステムの増加

近年、空中でのハンドジェスチャを利用した研究やシステムが増えている。SAMSUNG の Smart TV(図 1.1) や東芝のてぶらナビ(図 1.2) の登場によりテレビやパソコンをハンドジェスチャで操作することが一般的になりつつあり、ハンドジェスチャはより身近なものとなってきている。



図 1.1: SAMSUNG の Smart TV



図 1.2: 東芝のてぶらナビ

1.2 ハンドジェスチャの問題点

しかし、ハンドジェスチャには次のような大きく三つの問題点がある。

問題点 1 : 意図しない動きの誤認識

視線ポインティングにおける Midas Touch Problem と同様に、ハンドジェスチャにはユーザの動作がジェスチャとして意図されたものなのかそうでないのか判別しづらいといった問題がある。ハエをはらうといった動作がジェスチャとして認識されてしまう例を図 1.3 に示す。

この問題を解決するためジェスチャ開始トリガーや終了トリガーに特別な動作(例えば、拳を握る動作や数秒間静止させる動作)を用いるものがあるが、連続してジェスチャを行いたい場合に1ステップ余計な動作が必要になるため効率的な操作とは言えない(図 1.4)。



図 1.3: 意図しない動きの誤認識の例

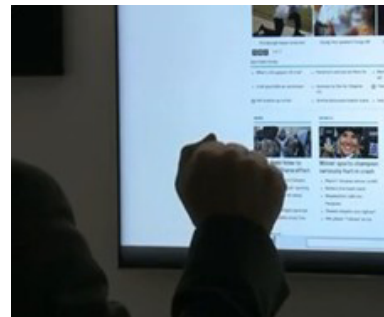


図 1.4: 開始ジェスチャに拳を用いた例

問題点 2：シンプルなハンドジェスチャの数が少ない

また、即時的な操作には、手を上げたり、左にスライドするといった直線的でシンプルなハンドジェスチャの利用が望ましい。しかし、操作の種類数の増加に応じてジェスチャの種類数を増やすために複雑なハンドジェスチャを導入しなければならない。

問題点 3：計測可能範囲の制限

デバイスレスでのジェスチャ認識にはカメラが必要であるが、カメラの計測可能な範囲は限られている。例えば、Kinect の深度カメラの計測可能範囲は水平視野角は 57 度と限られている。そのため、ユーザは常に特定の位置でジェスチャをしなければならない（図 1.5）。機器操作のために、利用者に常に特定の位置でのジェスチャを強いることは望ましくない。

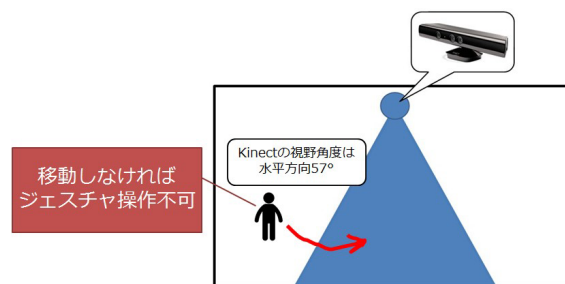


図 1.5: 計測可能範囲の制限

1.3 本研究の位置づけ

そこで本研究ではハンドジェスチャと単発音¹を組み合わせた“音付ハンドジェスチャ(図 1.6)”（以後、音付ジェスチャ）を提案し、ハンドジェスチャに関する問題点の解決を試みる。音付ジェスチャによりジェスチャの明確化、ジェスチャ数の増加を実現し、上述の問題点 1, 2 を解決する。単発音の検出には周波数毎の振幅における特徴を用い、複数種の単発音の分類も行う。ジェスチャ認識には深度カメラを用い、単発音とジェスチャを同時に検出した際に音付ハンドジェスチャとする。

また、上述の問題点 3 を解決するため、深度カメラを回転させる“回転台座”を提案し、深度カメラ 1 台での広角化を実現する。

¹手を叩くなどしてユーザが意識的に発生させる短時間のみ継続する音であり、声は含まない



図 1.6: 音付ジェスチャ

1.4 本論文の構成

以下、第2章では本研究に関連するシステムや手法を挙げる。第3章では提案手法“音付ジェスチャ”について説明する。第4章では予備調査について説明する。第5章では単発音の認識方法について説明する。第6章では音付ジェスチャ認識実験について説明する。第7章では音付ジェスチャ認識実験の結果および考察について説明する。第8章では比較実験について説明する。第9章では比較実験の結果および考察について説明する。第10章では提案手法“回転台座”について説明する。第11章では本研究の結論を述べる。

第2章 関連研究

2.1 ハンドジェスチャ

空中でのハンドジェスチャを利用した研究は多くある。

Mistry ら [1] は、カメラで指先に装着したカラーマーカーを追跡することにより、ハンドジェスチャを行った (図 2.1)。ズームインやズームアウトなどを空中で行うことが可能である。

Bailly ら [2] は、靴にカメラを取り付けることでウェアラブルなジェスチャ認識を可能とした。3種類のジェスチャを提案し、94~99%と高い認識率を実現している。

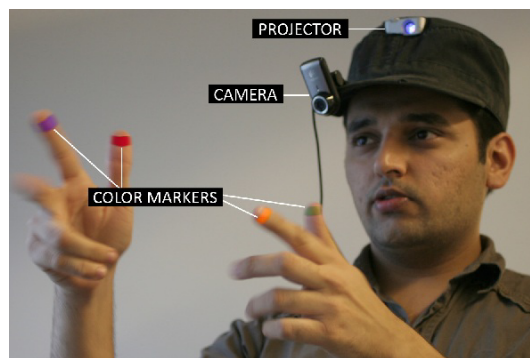


図 2.1: Mistry らの研究

池ら [3] の研究では、デバイスレスでの操作を実現している。高い手追跡技術により、手の姿勢や光の変化によらない認識技術を開発した。これによりユーザの意図した動きだけを認識することが可能となったが、画像処理に関する複雑なアルゴリズムが必要になってしまう問題がある。

Niels ら [4] の研究では音楽プレーヤーの操作に、長谷川ら [5] の研究では、動画再生プレーヤーの操作にハンドジェスチャを取り入れており、デバイスレスでの操作を実現している (図 2.2, 図 2.3)。全てシンプルなジェスチャで構成されているが、さらに操作の種類数を増やすには、より複雑なジェスチャが必要になる。また両研究ともなんらかの開始ジェスチャを必要とするため、連続して操作する場合に効率的とは言えない動作となってしまう。



図 2.2: Niels らの研究

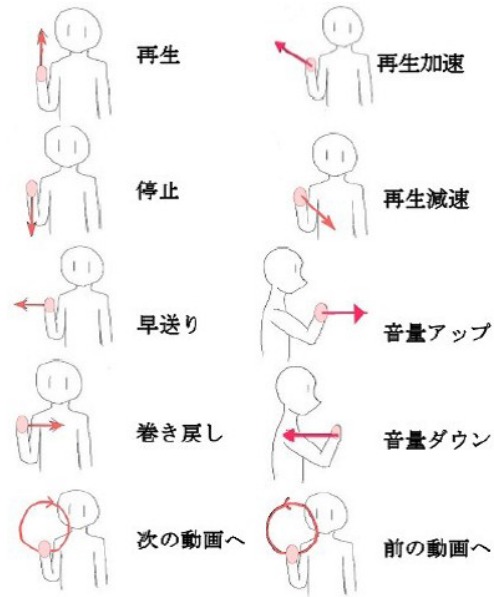


図 2.3: 長谷川らの研究

2.2 音の認識

Yinlin ら [6] では、ハンドジェスチャの選択トリガーにスナップ (通称、指パッチン) やクラップ (拍手) による単発音を取り入れた研究を行った。ジェスチャ開始トリガーとして拳を握ったり、静止させることは不自然であり、利用者にとって重荷と捉える点は本研究と一致するが、ジェスチャと音を別に認識していることやポインティングの選択での利用のみである点が本研究と異なる。

Chris ら [7] や尾崎ら [8] , 長谷川ら [9] は音の分類に関する研究を行った。

Chris らはタッチパネル上で指の腹や爪、ノックなどでタッチした際に生じる音の違いを分類し、音とタッチパネルでの操作を組み合わせることでインタラクションの拡張を実現した。図 2.4 の 4 種類 (Tip, Nail, Knuckle, Pad) の入力判別で 95% と高い精度を実現している。

尾崎らは指パッチンや拍手による単発音の違いを高い精度で分類した。音声は発話の仕方による変動が大きく、認識処理のために比較的高負荷な処理を行う必要があるが、単発音は変動が少なく波形の長さが短いため、低性能プロセッサでも判別が可能である。指パッチンの音は時間的な減衰が速く、特徴的な波形であるため、シミュレーション評価での判別率は 100% であった。

長谷川らはタブレット端末のベゼル部分や背面を指の腹や爪でタップした際に生じる音の違いを分類した。全 13 種類の操作を提案しており、平均 88.1% と高い分類精度を実現している。

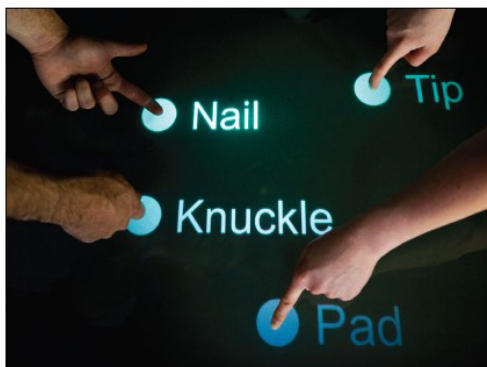


図 2.4: Chris らの研究

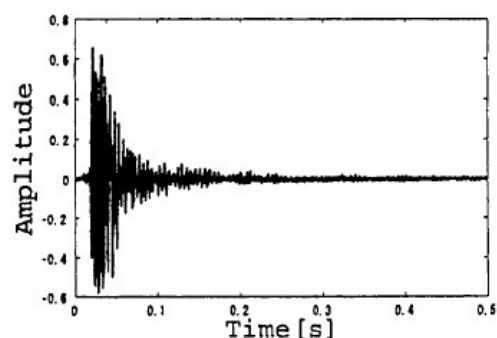


図 2.5: 尾崎らの研究

2.3 マルチモーダル (ハンドジェスチャと音の組み合わせ)

David ら [10] の研究では, 指パッチンとハンドジェスチャを組み合わせている (図 2.6). 指パッチンの音により, 操作対象の決定とジェスチャとを区別しているが, 音の種類がひとつ (指パッチン) である点や単発音を発してからジェスチャを行う点が本研究と異なる. また, ジェスチャの明確化を目的としておらず, 複数の機器の選択を目的としている.



図 2.6: David らの研究

神原ら [11] の研究では, 「キラキラ」, 「もこもこ」といったオノマトペとペン操作を組み合わせた. 声を出しながら絵を描くというマルチモーダル性により, ペイントソフトのインタラクションを拡張した. さらに, 視覚的, 聴覚的なフィードバックも同時に得ることが可能となった.

2.4 カメラの広角化

Wilson ら [12] の研究では, 装置自体を物理的に移動させ利用者を追跡するアプローチで解決を図っている (図 2.7). しかし部屋全体をカバーするために計 4 台の Kinect を使用するうえ, 機材を固定する必要があるため利用環境が制限される. また, 非常に高価な機材を利用しているため現実的ではない.

Tomari ら [13] の研究では, 魚眼レンズを用いた光学的手法により視野を広げるアプローチで解決を図っている (図 2.8). レンズを装着した Kinect を用意すればよく, 設置に関する制約はないが, 利用に先立ち, RGB 画像および深度画像に生ずる歪を補正するための変換パラメータを決定し, レーザ光を用いたデータでニューラルネットを学習させる必要があり, 複雑な処理が必要になってしまう.

複数のカメラを設置するアプローチもあるが、光学的手法と同様に複雑な処理が必要となってしまう。



図 2.7: Wilson らの研究



図 2.8: Tomari らの研究

第3章 提案手法“音付ジェスチャ”

本研究ではハンドジェスチャに単発音というモダリティを組み合わせた“音付ジェスチャ”を提案する。これにより、ジェスチャ時のユーザの意図の明確化、ジェスチャの種類数の増加を目指す。

3.1 設計方針

ハンドジェスチャとその直後に、身体の各部位(掌, 腕, ももなど)および周辺にある身体以外のもの(たとえば机や壁など)をタップして(叩いて)発生させた単発音やスナップにより発生させた単発音を組み合わせる(図 1.6)。単発音を発するハンドジェスチャを認識することで、ジェスチャの明確化を図る。

また、その単発音を区別することでジェスチャの種類数の増加を図る。具体的には、手を振り下ろすジェスチャであれば、机を叩いたりノックして発生させる単発音と組み合わせることとなる(図 3.1)。

デバイスには Microsoft の Kinect¹(図 3.2)を利用する。深度カメラ, マイクロホンアレイが搭載されており、ジェスチャと音の認識が可能である。

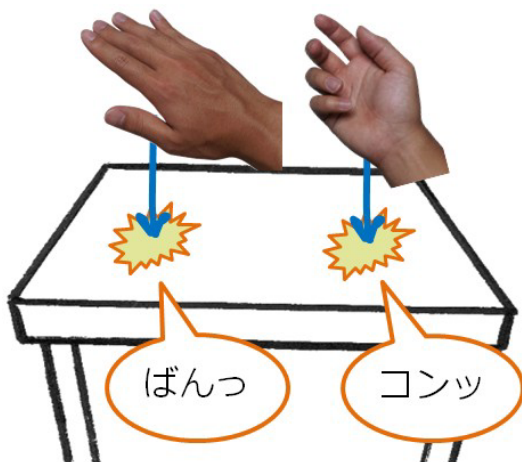


図 3.1: 音付ジェスチャの例



図 3.2: Kinect

3.2 音付ジェスチャの認識方法

音付ジェスチャの認識方法を図 3.3 に示す。マイクでの音の取得と深度カメラでの手の座標の取得をリアルタイムに行っていく。あらかじめ登録されている単発音, ハンドジェスチャかどうかをマッチング計算していき, 指定の単発音, ハンドジェスチャを一定時間以内に認識した場合に音付ジェスチャと判定する。

¹Kinect for Xbox 360

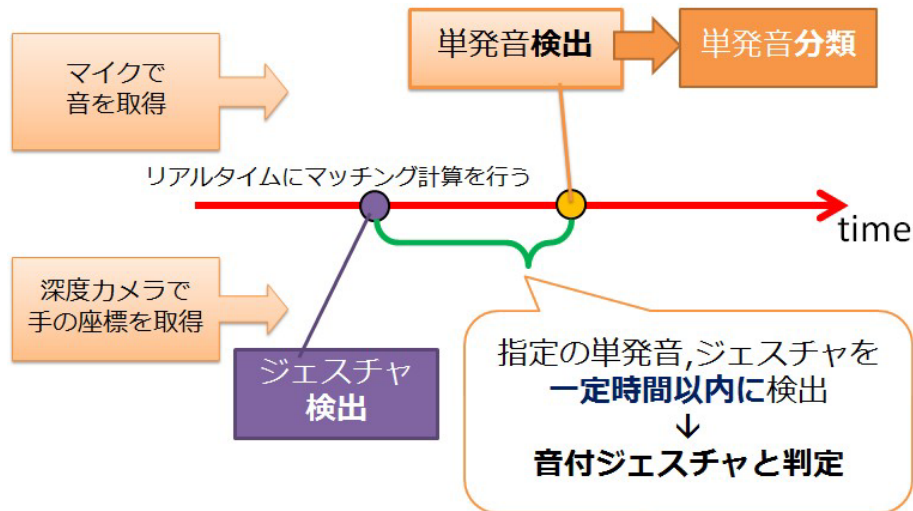


図 3.3: 音付ジェスチャの認識アルゴリズム

3.2.1 単発音認識方法

単発音の認識は2つのステップに分かれている。まず振幅の減衰により、単発音かどうかを検出する。その後、周波数毎の振幅における特徴を用いて複数種の単発音の分類を行う。詳しくは第5章で述べる。

3.2.2 ハンドジェスチャ認識方法

先行研究である長谷川らの研究 [5] を参考にし、動的時間伸縮法 (DTW²) によって実装した。これはデータ長やピーク値をとるタイミングが一致するように、2つの波形データの同期を取るアルゴリズムである。これによりジェスチャの動作時間や動作距離によらないジェスチャ認識が可能である (図 3.4)。

あらかじめ、各ハンドジェスチャの x,y,z 座標データ (今回は右手のデータとした) を記録しておき、そのデータとのマッチング計算を行う。

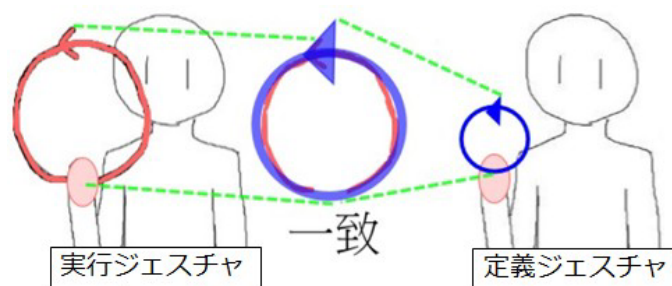


図 3.4: DTW によるジェスチャ認識 (長谷川らの研究)

²Dynamic Timewarping

3.3 音付ジェスチャの利点

ジェスチャ時のユーザの意図の明確化，シンプルなジェスチャの種類数の増加以外にも次のような利点がある。

連続ジェスチャ

システムにジェスチャの開始を明示的に伝えるために，何らかの開始ジェスチャを用いる場合，まず開始ジェスチャを行い，次いで目的のジェスチャを行う。操作を連続して行うなら，これらを繰り返す。一方音付ジェスチャでは，開始ジェスチャを行う必要がないため，すぐに次のジェスチャに移ることが可能である。これにより，効率的な操作が可能となり，疲労度の軽減も見込める。

左へスライドするジェスチャと下へスライドするジェスチャを順番に行う場合の例を図 3.5 に示す。上段は何らかの開始ジェスチャを行う既存手法の例であり，下段が今回提案する音付ジェスチャである。



図 3.5: 連続ジェスチャの比較

フィードバックの付加

ハンドジェスチャにはフィードバックがなく，実際に操作したのかどうか分かりにくい。これは，意図しない動きの誤認識の問題点にもつながる。

音付ジェスチャでは，音が鳴ったというフィードバックが得られ，実際にジェスチャしたのかどうか容易に理解可能となる。

3.4 応用先

応用先としては次のような機器の操作を想定している。

- テレビ，動画再生プレーヤー
- 照明，空調
- プレゼンテーション支援システム

- 水回り（トイレ，台所，風呂場）

通常これらの機器の操作にはリモコンが利用されるが，リモコンには，機器ごとに異なる装置を必要としたり，さらには紛失するといった問題があることが分かっている [14]．音付ジェスチャによりこれらの問題が解決できると考えている．また，プレゼンテーション中は周りの環境音が小さいため，単発音の利用に適していると言える．

第4章 予備調査

4.1 概要

シンプルな 24 種の音付ジェスチャを対象に，単発音の大きさやユーザの好みに関する調査を行った．また，典型的な利用環境の環境音下で単発音を検出できるかどうかの調査も行った．

これらの結果から音付ジェスチャとして実際に利用できるかどうかを判断する．

4.1.1 目的

次の点について調査を行う．

1. 典型的な利用環境で単発音を検出できるかどうか
2. 動作のしやすさ，ユーザの好み

1 は定量的評価，2 はアンケートによる定性的評価を行う．

4.1.2 被験者

被験者は本研究室の学生 10 名 (女性 2 名，左利き 2 名) である．

4.2 タスク

24 種の音付ジェスチャを Kinect の前で実行し，音データを取得した．今回の実験では音の解析が主なため，ジェスチャは意識しないよう指示した．24 種の音付ジェスチャを図 4.1 に示す．左右それぞれの手で指定の箇所をタップするものやクラブ，スナップがある．身体以外で単発音を発生させるために今回は机を用いた．また，前腕および上腕タップは長袖の服を着用した場合との比較も行った．身体をタップする音付ジェスチャは，則枝ら [15] の研究により腕へのタップ入力の有用性が示されているため採用した．

被験者には室内での機器操作を想定させ，自然な強さで単発音を出すよう指示した．

24 種の音付ジェスチャ各 3 回を 1 セッションとし，計 2 セッション行った．動作の指示は実験者が出した．記録された音付ジェスチャ数は，2 セッション× 24 ジェスチャ× 3 回× 10 人=1,440 となった．

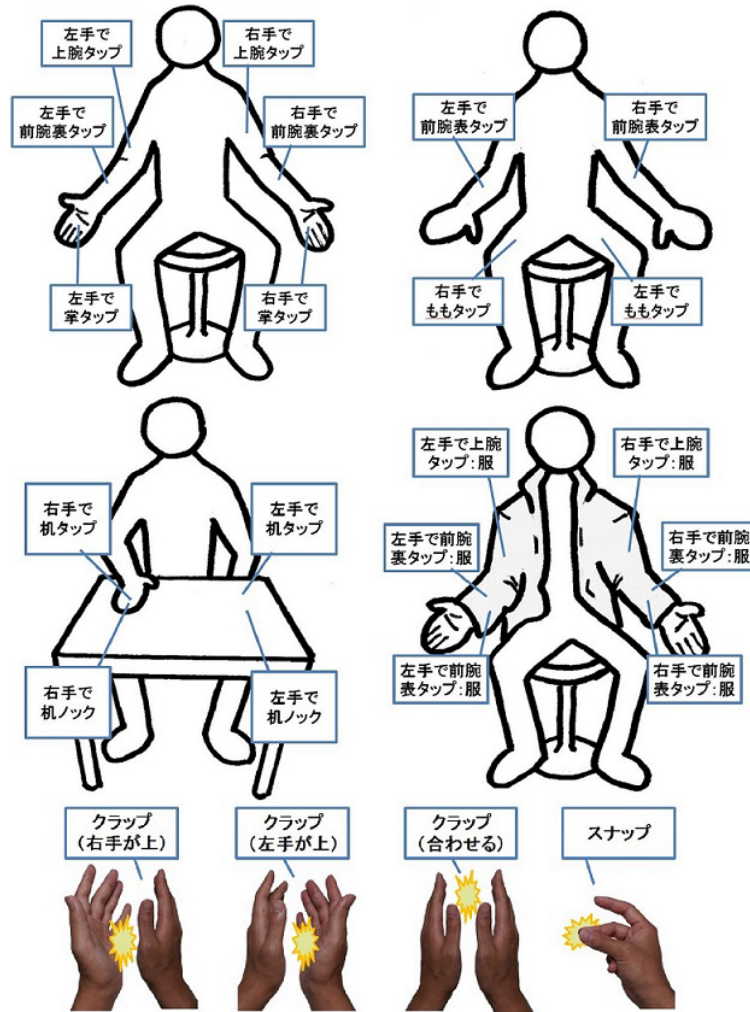


図 4.1: 24 種の音付ジェスチャ

4.3 実験環境

エアコンや人の声などの雑音が入らないよう静音環境下で行った。

実験中は深度カメラから 3m 離れた位置に椅子と机を設置し、カメラに正対して座り、行うよう指示した。長袖の服を着た場合での上腕、前腕タップも行うため、半袖の場合と長袖 (指定のパーカーを着用) の場合とに分けて行った。

実験装置には Microsoft の Kinect を使用し、録音には Kinect 内臓マイクを使用した。サンプリングレートは 16000Hz である (ハードウェアの性能上これ以上高くはできない)。使用計算機のスペックは CPU が corei5、メモリ 4G、OS は windows7 を使用した。ソフトウェアの開発には C# と Microsoft Kinect for Windows SDK ver1.5 を使用した。

4.4 結果：音の大きさ

各単発音の最大振幅について調査を行った結果を図 4.2 に示す。各音付ジェスチャの単発音の振幅を表している。

机を用いた単発音が他に比べ大きいことが分かる (t 検定, $p < .01$). 利き手と非利き手を比較した場合には有意差はなかった. 服を着た場合と着ない場合には, 利き手で前腕裏をタップ, 非利き手で前腕表をタップに有意差があった (t 検定, $p < .05$).

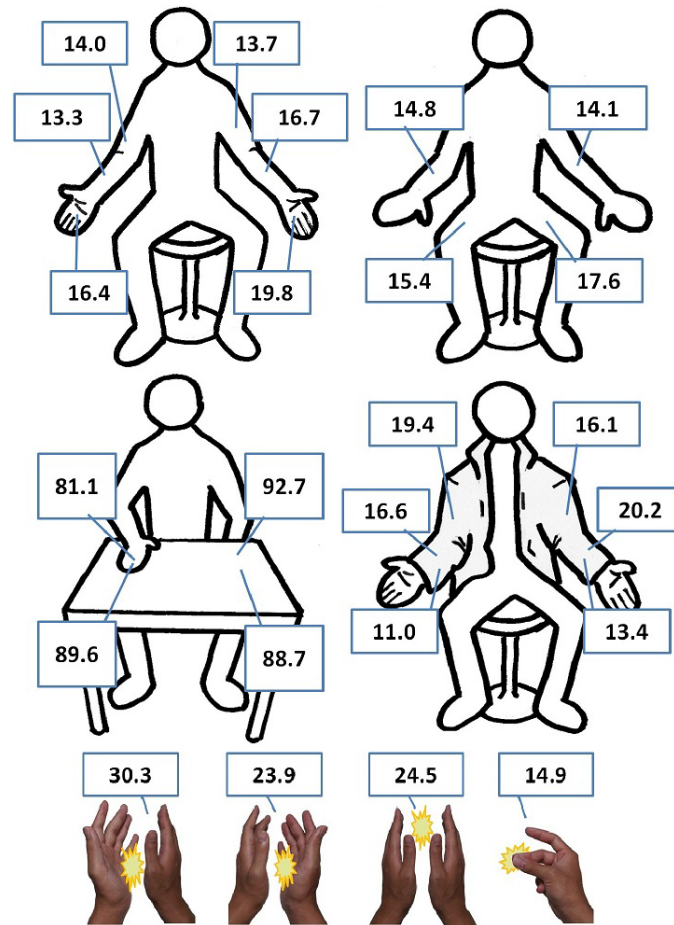


図 4.2: 音付ジェスチャの最大振幅

また, どのくらいの振幅があれば十分かを調査するため, 実際の利用環境の環境音の最大振幅を調べた. TV 番組や映画, プレゼンテーションなどの 6 種の環境について平均最大振幅を調べた結果, 平均最大振幅が 3.3, 標準偏差が 1.78 となった. この値と各音付ジェスチャとを比較したものを図 4.3 に示す. なお, 机を利用したものは振幅が比較的大きいため除いてある. 環境音の平均最大振幅は赤線で示した. 図 4.3 より, ほとんどの場合で全ての音付ジェスチャが利用できる可能性があると言える.

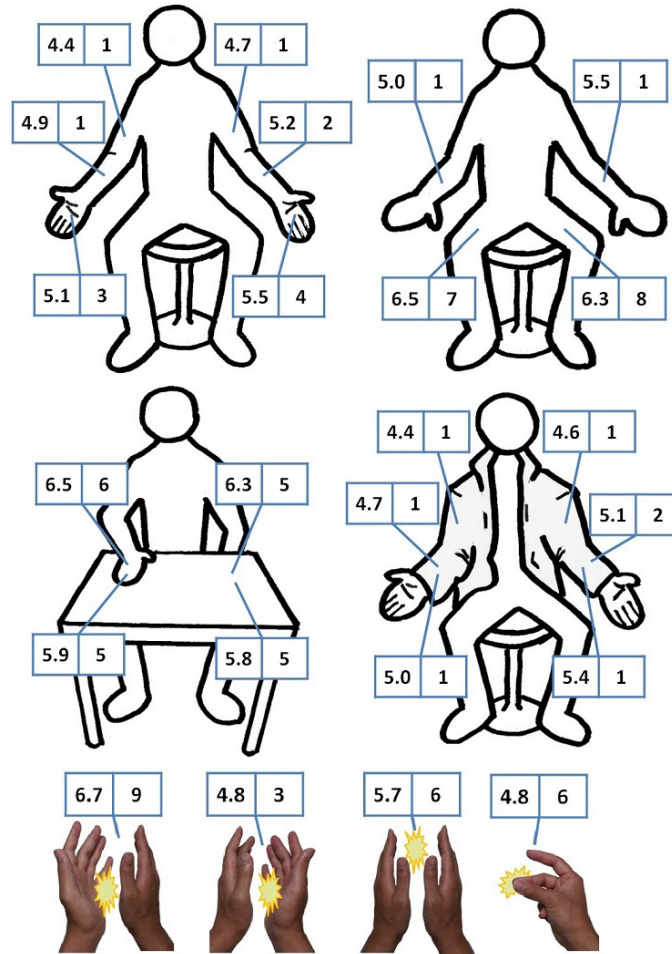


図 4.5: アンケート結果:動作のしやすさ (左の値), 実際の機器に使いたいもの (右の値)

左右の手で有意差 (t 検定, $p < .05$) のない場合に左右の平均値をとったものを表 4.1, 表 4.2 に示す. 上位 5 つの音付ジェスチャを表している. なお, スナップの結果はスナップができないと答えた 3 人を除いたデータである. 実際の機器操作に使いたいものは何パーセントの人が選んだかを表している.

表 4.1: 動作のしやすさ

動作のしやすさ	
クラップ	6.7
ももタップ	6.4
机タップ	6.4
スナップ	5.9
机ノック	5.9

表 4.2: 実際の機器操作に使いたいもの

機器操作に使いたいもの	
クラップ	90%
スナップ	86%
ももタップ	75%
机タップ	55%
机ノック	50%

4.6 考察

以上の結果より以下のことが分かった.

- 両アンケート結果から単発音の使用に抵抗がないことが分かり, 特にクラップ, 机タップ, 机ノック, スナップ, ももタップが好まれることが分かった. また, 全ての単発音の最大振幅

が環境音の平均振幅を上回ったため、ほとんどの場合で全ての単発音が検出できる可能性があることも分かった。

- 非利き手での動作はあまり好まれないが、十分な音を出すことは可能である。また、上腕、前腕タップはあまり好まれないことが分かった。3人の被験者から身体をタップする際の力加減が分からず痛い時があるという意見が得られたため、実際の利用を考えると、発せられる単発音の振幅はさらに小さくなるものと考えられる。
- スナップには個人差がある。スナップができると答えたのは10人中7人で、その中の6人が実際の機器操作で使いたいと答え、これは最も多かったクラップ(右手が上)を超える結果となる。動作のしやすさにおいても平均が5.9となるため、スナップができる人にとっては機器操作に最適の方法であると言える。

よって、本研究ではクラップ、机タップ、机ノック、スナップの4種類の単発音を採用することにした。ももタップも評価が高かったが、座っている必要があるため、今回は採用しなかった。

この4種類の単発音を用いた音付ジェスチャを実装し、HIS('12)にて対話発表・デモを行った[16].

第5章 単発音認識方法の調査

被験者から単発音の音データを採集，解析し，単発音の認識方法（単発音の検出方法，単発音の分類方法）を決定する。

まず，単発音の振幅の減衰について調査し，単発音の検出方法を決定する．次に，10-fold cross-validation にて単発音分類精度の調査を行い，単発音の分類方法を決定する．

5.1 単発音データの新規採集

4種の単発音（クラップ，机タップ，机ノック，スナップ）を対象に，単発音分類アルゴリズムの比較を行った．

録音は静穏環境下で行った．

なおデータ採集後に，指定された単発音を安定して発することができたか被験者に尋ねた．

5.1.1 被験者

調査に用いたサンプルデータは，本研究室の学生7名から集めた．各被験者から各単発音につき30サンプル計120サンプル，合計840サンプルを得た．

5.1.2 実験システム

Kinect 内臓マイクを使用した．サンプリングレートは16000Hzである（ハードウェアの性能上これ以上高くはできない）．

使用計算機のスペックはCPUがcorei5，メモリ4G，OSはwindows7を使用した．ソフトウェアの開発にはC＃とMicrosoft Kinect for Windows SDK ver1.6を使用した．



図 5.1: データ収集中の画面

5.1.3 実験環境

静穏環境下で実験を行った。Kinect から 3 m 離れた場所に座って単発音を発した。

5.2 単発音の検出方法の検証

5.2.1 単発音の検出方法

4 種類の単発音の継続時間はほぼ 200ms (1 フレーム 33ms としたため全 6 フレーム) に収まるのが分かったため、以後この値に固定することにする。例としてクラップの振幅の時間変化を図 5.2 に示す。

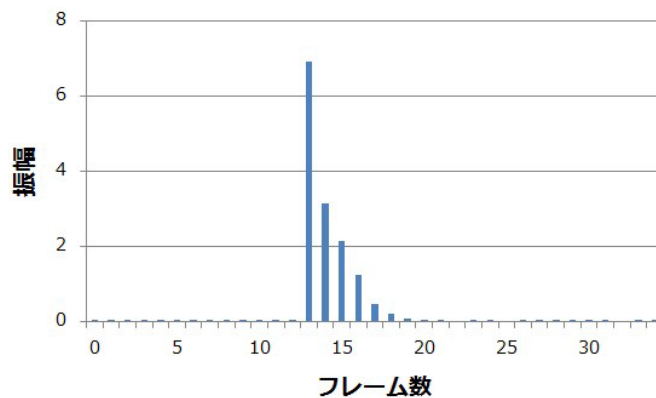


図 5.2: 単発音の継続時間 (横軸: 時間, 縦軸: 振幅)

この特徴的な振幅の減衰から単発音の検出を行う。図 5.2 から、一気に振幅が跳ね上がり、200ms 以内に減衰していくことが分かる。よって、以下の条件を満たした場合に単発音検出とした。

- 最大振幅のフレームが最初の 2 つのフレームのどちらか

- 最大振幅÷平均振幅が閾値以上
- 最大振幅÷最小振幅が閾値以上
- 平均振幅÷最小振幅が閾値以上

最大振幅、平均振幅、最小振幅は単発音の継続時間 200ms 内での値である。3 種類の閾値は今回採集した単発音データからあらかじめ求めておく。

この検出方法により、高い精度で単発音の検出が可能となった。

5.2.2 閾値の決定

採集した単発音データから閾値を決定した。各被験者毎に単発音の最大振幅、平均振幅、最小振幅の平均を求め、3 種類の倍率を求めた。各被験者の最大振幅÷平均振幅を図 5.3 に、最大振幅÷最小振幅を図 5.4 に、平均振幅÷最小振幅を図 5.5 に示す。

この中で最も低い値は、最大振幅÷平均振幅は 2.1、最大振幅÷最小振幅は 8.3、平均振幅÷最小振幅は 3.7 となった。

この結果から、3 種類の閾値は、最大振幅÷平均振幅は 2.0、最大振幅÷最小振幅は 7.0、平均振幅÷最小振幅は 3.5 とした。

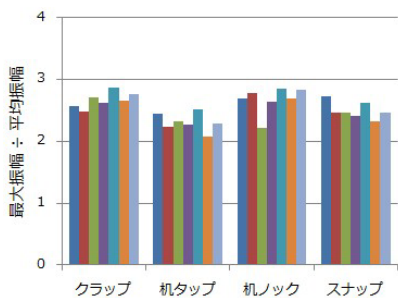


図 5.3: 各被験者の最大振幅÷平均振幅の値

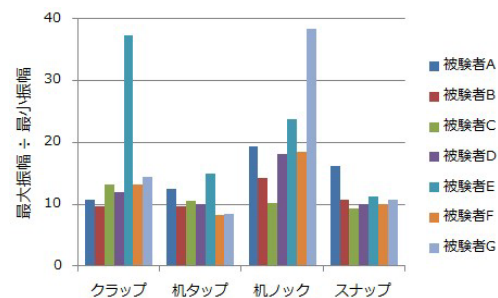


図 5.4: 各被験者の最大振幅÷最小振幅の値

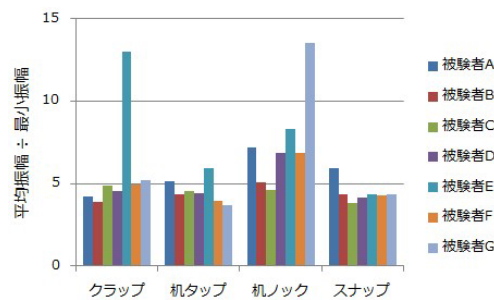


図 5.5: 各被験者の平均振幅÷最小振幅の値

5.3 単発音の分類方法の検証

5.3.1 単発音の分類方法

RMS¹エラーによるトレーニングデータとテストデータとのマッチング計算によって行うことにした。FFT を用いてあらかじめ採集しておいた複数のサンプルデータから各周波数における振幅の平均と標準偏差の組を求めておき、これをトレーニングデータとする。

サンプルデータのサンプリングレートを 16000Hz としたため、解析対象の周波数は 8000Hz (1 つのバンド幅は約 43Hz であるためバンド数は 186 個) までとした。この各バンドの RMS エラーを計算していくことになるが、今回の調査では次の 4 種類の区分分割法で分類精度を調査し、最も精度の高い区分分割法を採用することにした。

- 0 から 8000Hz までのバンド全てを用いる方法 (以後、全バンド)
- バーク尺度 [17] を用いる方法 (以後、帯域数 24)
- バーク尺度の隣り合う 2 つの帯域を統合したもの (以後、帯域数 12)
- バーク尺度の隣り合う 3 つの帯域を統合したもの (以後、帯域数 8)

バーク尺度は臨界帯域に基づいており、25 の境界があり 24 の帯域に分割される。各帯域の振幅を平均したものをデータとして用いるため、周波数が低いほど重み付けがされ、音の特徴が得やすくなる。隣り合う 3 つの帯域を統合したもの (帯域数 8) は Sampo らのスナップ検出アルゴリズム [18] にも利用されている。さらに今回は隣り合う 2 つの帯域を統合したもの (帯域数 12) での評価も行う。

なお、全バンドでは 186 (バンド数) \times 6 (フレーム) = 1116 回、帯域数 24 では 24 (バンド数) \times 6 (フレーム) = 144 回、帯域数 12 では 12 (バンド数) \times 6 (フレーム) = 72 回、帯域数 8 では 8 (バンド数) \times 6 (フレーム) = 48 回の計算を行うことになる。

この 4 種類から単発音の分類に適した区分分割法を決定する。

5.3.2 解析方法

分類精度は 10-fold cross-validation を用いて求めた。10-fold cross-validation は Chris ら [7] の音の分類に関する研究でも用いられている。

音の分類の解析は被験者毎に行った。単発音の音に個人差があることが分かっている [10] ためである。各被験者毎の分類精度を求めた後、最終的に被験者 7 人分の合計から分類精度 (%) を得た。

まず、各単発音の最初の 3 個のデータを除き、残り 27 個のデータから各単発音のトレーニングデータを作成する。除いたデータを 1 個ずつテストデータとして、4 種類のトレーニングデータとの RMS エラーを計算していく。4 つの RMS エラーのうち、最も値の小さい単発音であると判定し、カウントしていく。最初の 3 個のデータでの解析が終了したら、次の 3 個のデータで同様の計算を行っていく。

これにより、 4×4 の混合行列 (Confusion Matrix) が得られ、分類精度が求まる。

5.3.3 解析結果と考察

全バンドでの分類精度の内訳を表 5.1、帯域数 24 での分類精度の内訳を表 5.2、帯域数 12 での分類精度の内訳を表 5.3、帯域数 8 での分類精度の内訳を表 5.4 の混合行列として示す。

¹Root Mean Square

分類正答率（正しく分類された割合）をまとめると、全バンドが 86.3%、帯域数 24 が 80.0%、帯域数 12 が 80.8%、帯域数 8 が 78.1% となり、大きな違いは見られなかった。しかし、今回の解析結果は静穏環境下での結果であり、雑音環境下では結果が異なってくると考えられる。

分類精度に差が見られなかったことと、それぞれの区分分割法にはそれぞれの特徴があり、その特徴を生かすことを考え、本研究では 4 種類の区分分割法を全て採用し、それぞれの分類結果を平均することにした。

表 5.1: 分類の内訳（全バンド）

	クラップ	机タップ	机ノック	スナップ
クラップ	84.8	0.0	1.0	14.3
机タップ	1.0	82.9	16.2	0.0
机ノック	1.4	16.7	81.9	0.0
スナップ	3.3	0.0	1.0	95.7

表 5.2: 分類の内訳（帯域数 24）

	クラップ	机タップ	机ノック	スナップ
クラップ	77.1	0.0	1.0	21.9
机タップ	1.0	73.3	25.7	0.0
机ノック	1.4	25.7	72.9	0.5
スナップ	2.4	0.0	1.0	96.7

表 5.3: 分類の内訳（帯域数 12）

	クラップ	机タップ	机ノック	スナップ
クラップ	77.6	0.0	1.0	21.4
机タップ	1.0	75.2	23.8	0.0
机ノック	1.0	24.8	74.3	0.0
スナップ	2.9	0.5	0.5	96.2

表 5.4: 分類の内訳（帯域数 8）

	クラップ	机タップ	机ノック	スナップ
クラップ	77.1	0.0	1.0	21.9
机タップ	1.0	69.5	29.5	0.0
机ノック	1.4	28.6	70.0	0.0
スナップ	3.3	0.0	1.0	95.7

5.3.4 重み付けと正規化

単発音には個人差があるが、同一の被験者の単発音にもバラツキがあることが分かった。このバラツキを軽減することにより、分類精度の向上を目指した。次の 2 点を追加した。(1) 標準偏差による重み付け、(2) 音の大きさによる正規化の 2 点である。

(1) 標準偏差による重み付け

重み付けにはトレーニングデータの標準偏差を使用する。標準偏差の値が小さいほど、その周波数帯の重要度が高くなる。各周波数帯での RMS エラーを計算する際に、テストデータとトレーニングデータの差を標準偏差で割ることにより、重み付けを行った。これにより標準偏差の値が小さいほど、その周波数帯の重要度が高くなり、標準偏差の値が大きいほど、その周波数帯の重要度が低くなることになる。重み付けを行う際の RMS エラーを算出する式を (5.1) 式に示す。

$$RMS \text{ エラー} = \frac{1}{n} \sum \{ (\text{テストデータの振幅} - \text{トレーニングデータの振幅}) \div \text{標準偏差} \}^2 \quad (5.1)$$

なお標準偏差が 0 になってしまう場合を考慮し、すべての値には 0.01 のゲタをはかせてある。

(2) 音の大きさによる正規化

単発音の大きさにはばらつきがあるため、振幅の大きさに関する正規化を行った。テストデータの最大平均振幅とトレーニングデータの最大平均振幅を揃えた。この倍率を求める式を (5.1) 式とし、この値を全ての周波数に掛けた。正規化を行う際の RMS エラーを算出する式を (5.3) 式に示す。

$$x = \text{トレーニングデータの最大平均振幅} \div \text{テストデータの最大平均振幅} \quad (5.2)$$

$$RMS \text{ エラー} = \frac{1}{n} \sum (x \times \text{テストデータの振幅} - \text{トレーニングデータの振幅})^2 \quad (5.3)$$

5.3.5 重み付けと正規化を行った際の分類精度

全バンドでの分類精度の内訳を表 5.5、帯域数 24 での分類精度の内訳を表 5.6、帯域数 12 での分類精度の内訳を表 5.7、帯域数 8 での分類精度の内訳を表 5.8 の混合行列として示す。

分類正答率をまとめると、全バンドが 84.8%、帯域数 24 が 87.3%、帯域数 12 が 87.7%、帯域数 8 が 86.9% となり、全体で約 7% 精度が高くなった。

また、雑音環境下ではさらなる精度の向上が期待できるため、重み付けと正規化を採用することにした。

表 5.5: 正規化, 重み付け後の分類の内訳 (全バンド)

	クラップ	机タップ	机ノック	スナップ
クラップ	92.4	0.0	1.0	6.7
机タップ	1.0	99.0	0.0	0.0
机ノック	0.5	39.5	60.0	0.0
スナップ	11.4	0.5	0.5	87.6

表 5.6: 正規化, 重み付け後の分類の内訳 (帯域数 24)

	クラップ	机タップ	机ノック	スナップ
クラップ	91.4	2.9	1.0	4.8
机タップ	0.0	95.2	4.8	0.0
机ノック	0.5	13.3	86.2	0.0
スナップ	22.4	1.0	0.5	76.2

表 5.7: 正規化, 重み付け後の分類の内訳 (帯域数 12)

	クラップ	机タップ	机ノック	スナップ
クラップ	91.4	2.9	1.0	4.8
机タップ	0.0	96.2	3.8	0.0
机ノック	1.0	13.8	85.2	0.0
スナップ	20.5	1.0	0.5	78.1

表 5.8: 正規化, 重み付け後の分類の内訳 (帯域数 8)

	クラップ	机タップ	机ノック	スナップ
クラップ	91.0	2.4	1.0	5.7
机タップ	0.5	94.3	5.2	0.0
机ノック	1.0	13.3	85.7	0.0
スナップ	21.9	1.0	0.5	76.7

5.4 単発音の種類数による分類精度

単発音の種類数を少なくした場合の分類精度について解析した.

4 種類の単発音があるため, 次の 11 通りの組み合わせがある. 4 種類が 1 つ, 3 種類が 4 つ, 2 種類が 6 つである.

5.4.1 単発音を 3 種類とした場合

クラップ, 机タップ, 机ノックでの分類精度の内訳を表 5.9, クラップ, 机タップ, スナップでの分類精度の内訳を表 5.10, クラップ, 机ノック, スナップでの分類精度の内訳を表 5.11, 机タップ, 机ノック, スナップでの分類精度の内訳を表 5.12 に示す. これらの表はすべて重み付け, 正規化を行ったものである.

この結果から, クラップとスナップ, 机タップと机ノックが互いに似ていることが分かる.

表 5.9: 分類の内訳 (クラップ, 机タップ, 机ノック)

全バンド	クラップ	机タップ	机ノック
クラップ	96.2	3.3	0.5
机タップ	0.0	99.5	0.5
机ノック	0.0	47.6	52.4
帯域数 24			
クラップ	95.7	3.8	0.5
机タップ	0.0	96.2	3.8
机ノック	0.0	25.7	74.3
帯域数 12			
クラップ	96.2	3.3	0.5
机タップ	0.0	97.6	2.4
机ノック	0.0	24.8	75.2
帯域数 8			
クラップ	96.2	3.3	0.5
机タップ	0.0	95.2	4.8
机ノック	0.0	23.8	76.2

表 5.10: 分類の内訳 (クラップ, 机タップ, スナップ)

全バンド	クラップ	机タップ	スナップ
クラップ	84.8	1.9	13.3
机タップ	0.0	100.0	0.0
スナップ	9.5	1.4	89.0
帯域数 24			
クラップ	87.1	3.3	9.5
机タップ	0.0	100.0	0.0
スナップ	18.6	1.4	80.0
帯域数 12			
クラップ	89.0	2.4	8.6
机タップ	0.0	100.0	0.0
スナップ	16.2	1.4	82.4
帯域数 8			
クラップ	86.7	2.4	11.0
机タップ	0.0	100.0	0.0
スナップ	18.6	1.4	80.0

内訳を表 5.18 に示す。これらの表はすべて重み付け、正規化を行ったものである。

この結果からも、クラップとスナップ、机タップと机ノックが互いに似ていることが分かる。クラップとスナップ、机タップと机ノックの分類精度以外は全て 100%近い精度（全て 98%以上）で分類が可能であることが分かる。

表 5.13: 分類の内訳 (クラップ, 机タップ)

全バンド	クラップ	机タップ
クラップ	96.2	3.8
机タップ	0.0	100.0
帯域数 24		
クラップ	95.7	4.3
机タップ	0.0	100.0
帯域数 12		
クラップ	96.2	3.8
机タップ	0.0	100.0
帯域数 8		
クラップ	96.2	3.8
机タップ	0.0	100.0

表 5.14: 分類の内訳 (クラップ, 机ノック)

全バンド	クラップ	机ノック
クラップ	99.0	1.0
机ノック	0.5	99.5
帯域数 24		
クラップ	99.0	1.0
机ノック	0.5	99.5
帯域数 12		
クラップ	99.0	1.0
机ノック	0.5	99.5
帯域数 8		
クラップ	99.5	0.5
机ノック	0.5	99.5

表 5.15: 分類の内訳 (クラップ, スナップ)

全バンド	クラップ	スナップ
クラップ	85.7	14.3
スナップ	11.0	89.0
帯域数 24		
クラップ	89.5	10.5
スナップ	20.0	80.0
帯域数 12		
クラップ	91.0	9.0
スナップ	17.6	82.4
帯域数 8		
クラップ	89.0	11.0
スナップ	20.0	80.0

表 5.16: 分類の内訳 (机タップ, 机ノック)

全バンド	机タップ	机ノック
机タップ	99.5	0.5
机ノック	47.6	52.4
帯域数 24		
机タップ	96.2	3.8
机ノック	25.7	74.3
帯域数 12		
机タップ	97.6	2.4
机ノック	24.8	75.2
帯域数 8		
机タップ	95.2	4.8
机ノック	23.8	76.2

表 5.17: 分類の内訳 (机タップ, スナップ)

全バンド	机タップ	スナップ
机タップ	100.0	0.0
スナップ	3.3	96.7
帯域数 24		
机タップ	100.0	0.0
スナップ	2.4	97.6
帯域数 12		
机タップ	100.0	0.0
スナップ	1.9	98.1
帯域数 8		
机タップ	100.0	0.0
スナップ	1.9	98.1

表 5.18: 分類の内訳 (机ノック, スナップ)

全バンド	机ノック	スナップ
机ノック	99.5	0.5
スナップ	1.0	99.0
帯域数 24		
机ノック	99.5	0.5
スナップ	1.4	98.6
帯域数 12		
机ノック	99.5	0.5
スナップ	1.9	98.1
帯域数 8		
机ノック	99.5	0.5
スナップ	1.9	98.1

5.5 アンケート結果

安定した音が出せたかについての7ポイントリッカートスケールのアンケート結果を図 5.6 に示す。クラップ、机タップ、机ノックは安定して音を出せたという意見が多かった。スナップの音は調整が難しいという意見や回数を重ねると疲れるなどの意見が多く、最も低い結果となった。しかし、スナップに関する主観的な結果が低いにも関わらず、上述した分類正答率は他の単発音よりも比較的高い精度となっている。これはスナップの音の周波数が特徴的であることが要因として考えられる [8][18]。

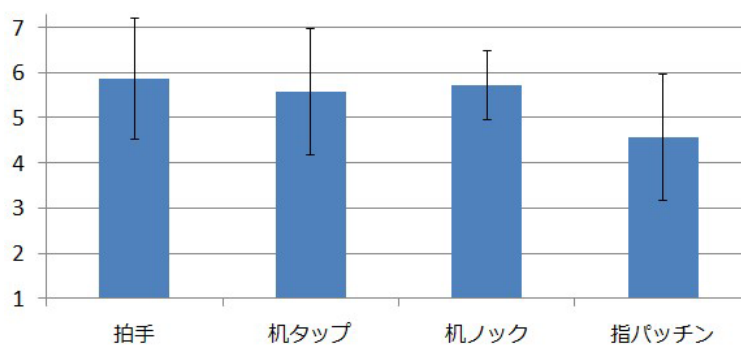


図 5.6: 安定した音が出せたか

第6章 音付ジェスチャ認識実験

6.1 概要

14種類の音付ジェスチャのデータを収集し、認識精度の調査、またユーザの好みの調査を行った。認識精度による定量的評価、アンケートによる定性的評価を行った。

今回採用した音付ジェスチャを図6.1, 図6.2に示す。図6.1はクラップ, 机タップ, 机ノックを使用したもの(6種)で, 図6.2はスナップを用いたもの(8種)である。単発音は前述した4種(クラップ, 机タップ, 机ノック, スナップ)を採用してある。単発音と組み合わせるハンドジェスチャは先行研究である長谷川らの研究[5]を参考にし, 直線的でシンプルな8種類のハンドジェスチャ(上, 下, 右, 左, 右上, 左下, 奥, 手前へスライド)を採用した。

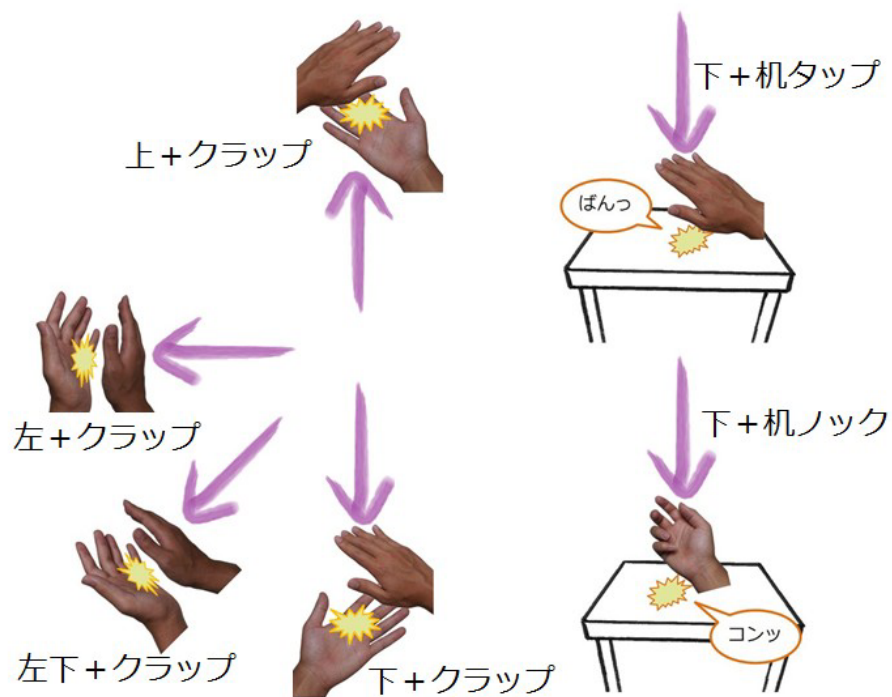


図 6.1: 音付ジェスチャ (クラップ, 机タップ, 机ノック使用)

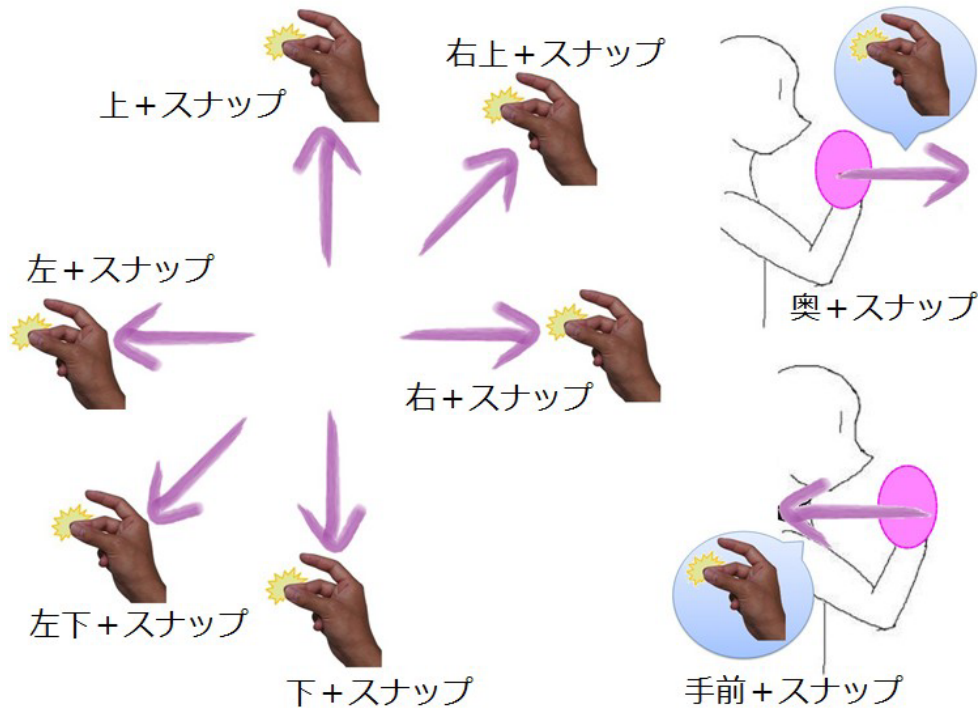


図 6.2: 音付ジェスチャ (スナップ使用)

6.1.1 被験者

本研究室の学生 5 名で、前述した単発音分類実験に参加した被験者である。単発音分類実験で採集した単発音データをもとにあらかじめトレーニングデータを作成した。

クラップ、机タップ、机ノックを使用した音付ジェスチャは 5 回ずつを 3 セッション、スナップを使用した音付ジェスチャは 2 回ずつを 2 セッション行い、各被験者から 122 サンプル計 610 サンプルを得た。スナップを使用した音付ジェスチャの種類が多いことと、スナップを繰り返し行うと指が痛くなることが予備調査より分かっているため、回数が少なくなっている。

6.2 実験システム

Kinect 内臓マイクを使用した。サンプリングレートは 16000Hz である。

使用計算機のスペックは CPU が corei5、メモリ 4G、OS は windows7 を使用した。ソフトウェアの開発には Processing, Java, SimpleOpenNI¹ を使用した。

指示通りのハンドジェスチャと何かしらの単発音が 1000ms 以内に検出された場合に音付ジェスチャ検出とし、フィードバックが得られる。フィードバックが得られたらキーボードのスペースキーを押下し、次の指示に進む。

6.2.1 実験画面

実験中の画面を図 6.3 に示す。

¹<http://code.google.com/p/simple-openni/>

ユーザはリアルタイムに認識されるハンドジェスチャの種類を確認できる．このため，ハンドジェスチャの認識に関するエラーはなく，単発音の認識エラーが音付ジェスチャの認識エラーに関わることになる．

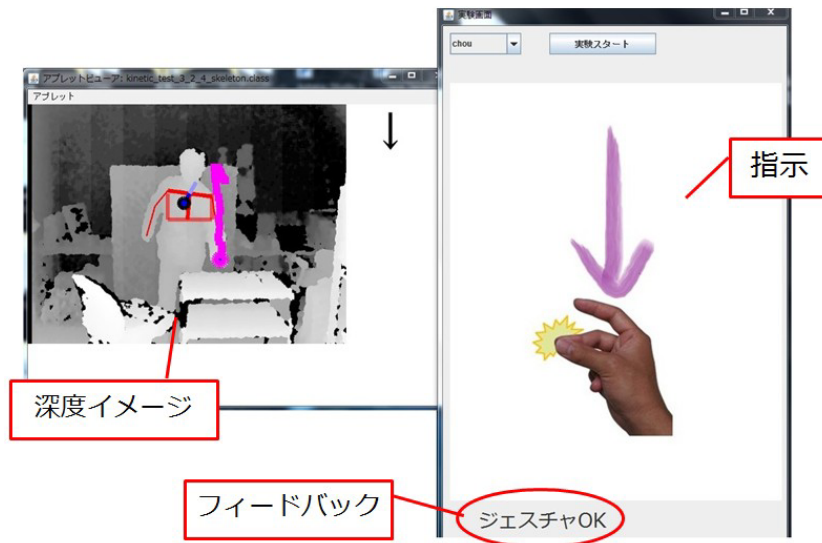


図 6.3: 実験中の画面

6.3 実験環境

静穏環境で実験を行った．Kinect から 3 m 離れた場所に立って音付ジェスチャを行わせた．実際の実験環境を図 6.4，図 6.5 に示す．

実験中はビデオカメラにより，被験者の動作を記録しておく．

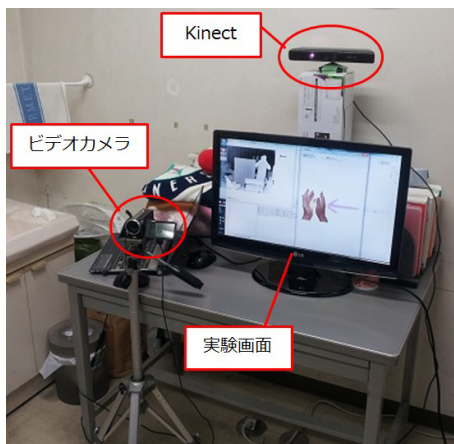


図 6.4: 実験環境 1

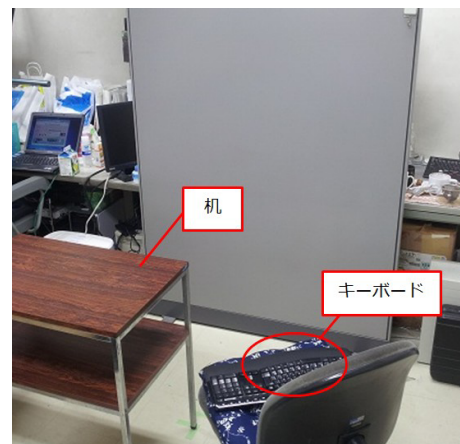


図 6.5: 実験環境 2

第7章 音付ジェスチャ認識実験結果と考察

7.1 音付ジェスチャ認識率

各音付ジェスチャの平均認識率を図 7.1 に示す。クラブ、机タップを用いた音付ジェスチャの認識率が低くなっていることが分かる。実験内容からハンドジェスチャによる誤認識は考えづらいため、これはクラブがスナップと、机タップが机ノックと誤認識されているためである。

そこで、似ている単発音を同じものとし、単発音を 2 種類とした場合の認識精度を求めた (図 7.2)。最も低い「下+クラブ」でも 78.7%と高い精度で認識可能であることが分かる。

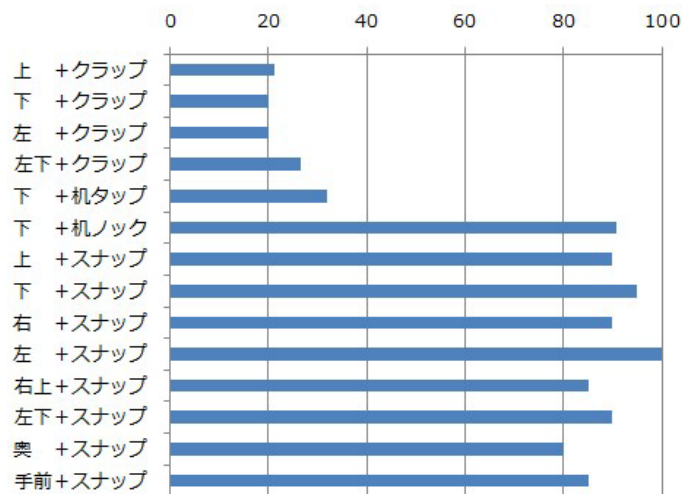


図 7.1: 音付ジェスチャの認識精度 (%) : 単発音を 4 種類とした場合

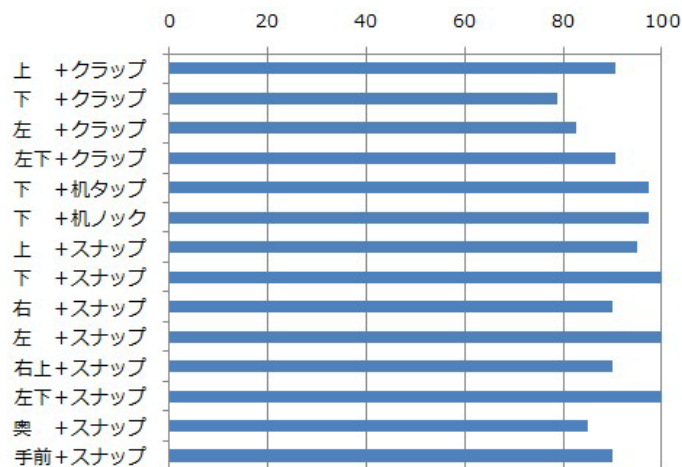


図 7.2: 音付ジェスチャの認識精度 (%) : 単発音を 2 種類とした場合

単発音の種類毎の平均認識精度を図 7.3 に示す。図の 3 種 (1) はクラブとスナップを同じものとした場合で、3 種 (2) は机タップと机ノックを同じものとした場合である。4 種類の場合は 66.1%と低いが、3 種類 (2) の場合は 71.3%、3 種類 (1) の場合は 86.6%、単発音の分類を 2 種類とした場合は平均 92.0%となった。

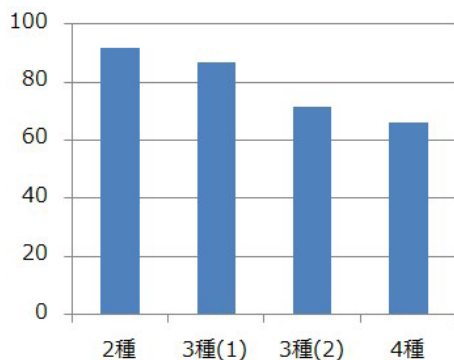


図 7.3: 単発音の種類毎の認識精度 (%)

7.2 音付ジェスチャ失敗回数

各音付ジェスチャの失敗回数を図 7.4 に示す。これは最後のセッションで音付ジェスチャを行う際に何回認識されなかったかを表している。失敗回数はビデオカメラで記録した動画を解析することで得られた。

手前+スナップが極端に多いことが分かる。これは手前にスライドするハンドジェスチャが認識されにくいことが原因として考えられる。先行研究でもこのジェスチャは最も認識率が悪い結果となっている [5]。

ビデオ解析によるとほとんどがジェスチャ認識に原因があることが分かり、ジェスチャ認識アルゴリズムの改善により、改善できると考えられる。

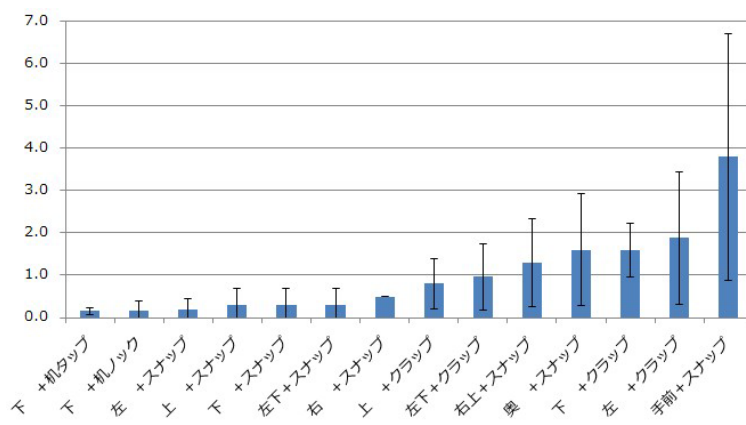


図 7.4: 音付ジェスチャの失敗回数 (回)

7.3 アンケート結果

認識のされやすさ、動作のしやすさ、疲労度についての 7 ポイントリッカートスケールのアンケート結果を示す。認識のされやすさを図 7.5, 動作のしやすさを図 7.6, 疲労度を図 7.7 に示す。

すべて7に近いほど評価が良いことを表している。

動作のしやすさ、疲労度は、「手前+スナップ」を除けば全て平均5ポイント以上の結果となった。この結果から音付ジェスチャはユーザに好まれる動作であると言え、認識精度が改善できさえすれば、ユーザに受け入れられる可能性がある事が分かった。

動作のしやすさ、疲労度の順位を見るといくつか類似点がある事が分かった。特に下へスライドするジェスチャが好まれる事が分かる。さらに机タップ、机ノックを使用したものは全てのアンケートにおいて上位に位置している。

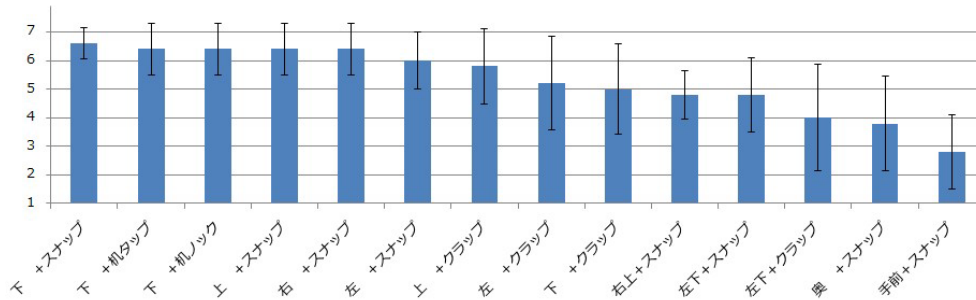


図 7.5: 認識のされやすさ

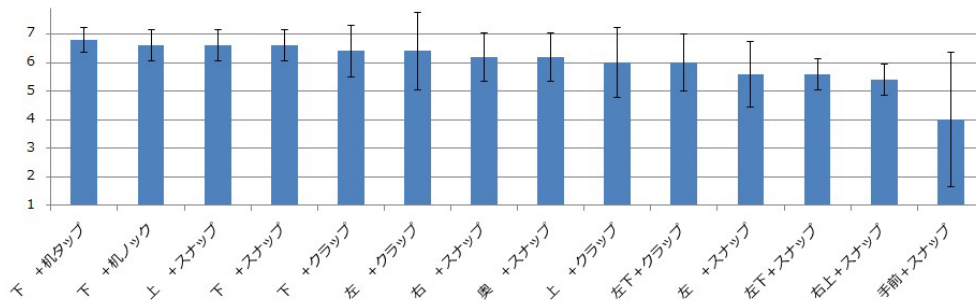


図 7.6: 動作のしやすさ

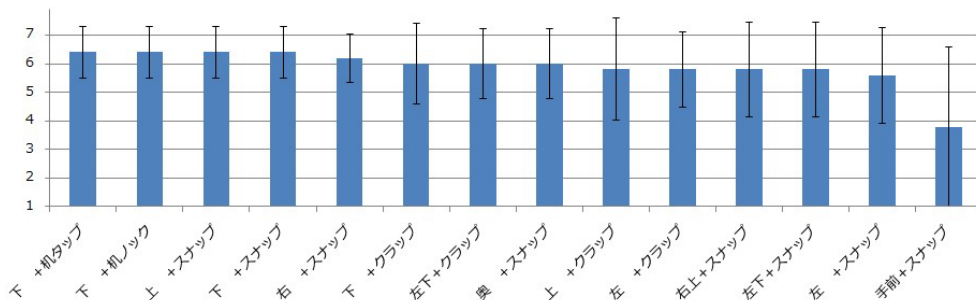


図 7.7: 疲労度

本章までの研究成果について、情報処理学会 第 151 回 HCI・第 95 回 SLP 合同研究発表会にて口頭発表を行った [19].

第8章 比較実験

8.1 概要

提案手法である音付ジェスチャと既存手法であるハンドジェスチャのみ（以後、ジェスチャのみ）の比較実験を行った。応用先のひとつである動画再生プレーヤーを対象とし、後述するタスクを行わせた。タスクの完了時間による定量的評価、アンケートによる定性的評価により有用性を評価した。

8.1.1 被験者

本研究室の学生3名で、前述した単発音分類実験に参加した被験者である。そのうち2名は前述した音付ジェスチャ認識実験にも参加した被験者である。

音付ジェスチャ認識実験と同様に単発音分類実験で採集した単発音データをもとにあらかじめトレーニングデータを作成した。

8.2 実験システム

Kinect 内臓マイクを使用した。サンプリングレートは16000Hzである。

使用計算機のスペックはCPUがcorei5、メモリ4G、OSはwindows7を使用した。ソフトウェアの開発にはProcessing、Java、SimpleOpenNIを使用した。

8.2.1 実験画面

音付ジェスチャでの実験画面を図8.1に、ジェスチャのみでの実験画面を図8.2に示す。音付ジェスチャは常にハンドジェスチャを認識するようにしてあるため、現在どのハンドジェスチャを実行しているかのフィードバックを表示するようにした。

一方、ジェスチャのみでは開始ジェスチャが実行された場合のみハンドジェスチャを認識するため、フィードバックは表示していないが、開始ジェスチャ実行後に手の軌跡の色が変化するようにになっている。



図 8.1: 実験画面：音付ジェスチャ

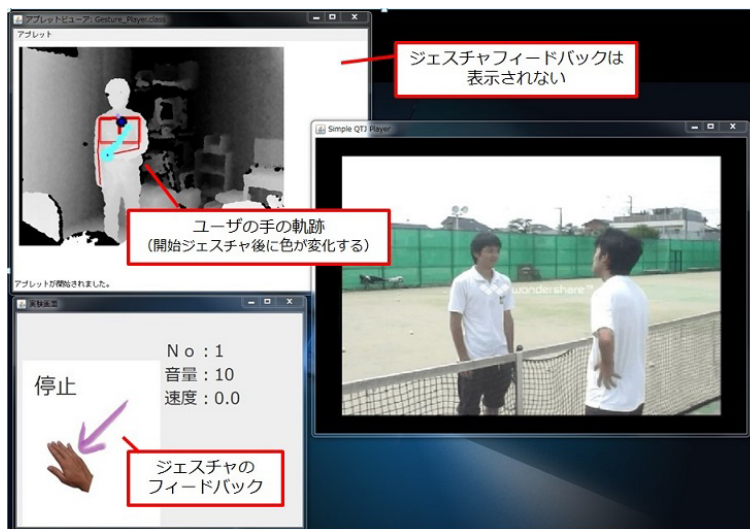


図 8.2: 実験画面：ジェスチャのみ

8.3 実験環境

Kinect から 3 m 離れた場所に立って動画再生プレイヤーを操作させた。動画の音が流れていても十分高い精度で単発音の検出が可能であることが予備調査により分かっている。実際の実験環境を図 8.3, 図 8.4 に示す。

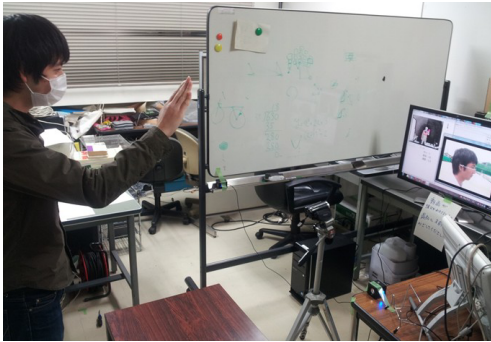


図 8.3: 実験風景

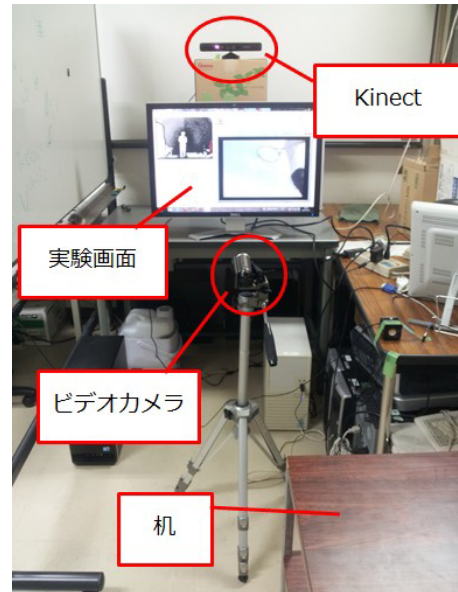


図 8.4: 操作方法：実験環境

8.4 ジェスチャの割り当て

動画再生プレーヤーの操作の種類は再生，停止，早送り（加速），巻き戻し（減速），音量 Up，音量 Down，次の動画へ，前の動画への 8 種類とした．音付ジェスチャの割り当てを図 8.5 に，ジェスチャのみの割り当てを図 8.6 に示す．

音付ジェスチャの単発音は 2 種類としたため，クラップとスナップ，机タップと机ノックは同じ音に分類される．例えば，巻き戻しをしたい場合，「左＋スナップ」と「左＋クラップ」のどちらかを行えば良い．

ジェスチャのみでは開始ジェスチャとして，胴体よりも高い位置で数秒間手を静止させるジェスチャを行う必要がある．連続してジェスチャを行う場合もジェスチャとジェスチャの間に開始ジェスチャを挟まなくてはならない．

再生	
停止	
早送り (加速)	
巻き戻し (減速)	
音量 Up	
音量 Down	
次の動画へ	
前の動画へ	

図 8.5: 操作方法 : 音付ジェスチャ

再生	
停止	
早送り (加速)	
巻き戻し (減速)	
音量 Up	
音量 Down	
次の動画へ	
前の動画へ	

図 8.6: 操作方法 : ジェスチャのみ

8.4.1 タスク

実際の動画操作の練習を両手法にて行ったあと、3つのタスクを被験者に課し、順番に行わせた。次のようなタスクを課した。

- 次のシーンを音量 MAX で見てください
- 次のシーンで3秒間一時停止してください
- 次のシーンを音量 min でみてください

タスクは8種類全ての操作を使用するように設定してある。動画は5分から15分の長さのものを7本用意した。あらかじめ被験者に動画を見せておき、どのあたりに指定のシーンがあるかが分かるようにした。

第9章 比較実験結果と考察

9.1 タスク完了時間

タスク完了時間の結果を図9.1に示す。これは3つのタスクの合計時間を表している。3人のタスク完了時間を平均すると、音付ジェスチャの使用により、12%タスク完了時間を短縮できた。また、2人の被験者が音付ジェスチャ使用時のほうが速くタスクを完了できた。被験者Cにおいては31%タスク完了時間を短縮できた。

被験者Bは若干音付ジェスチャの方が遅い結果となった。被験者Bは音付ジェスチャ認識実験に参加していない被験者であり、ビデオログを解析したところ、ジェスチャのミスが多いことが分かったため、練習により改善可能であると考えられる。

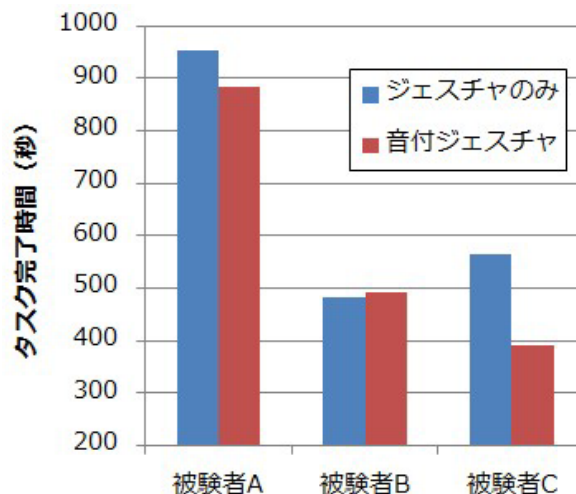


図 9.1: タスク完了時間

9.2 アンケート結果

実験後に疲れやすさ、スムーズな（連続）操作のしやすさ、誤動作のしにくさ、楽しさ、心地よさ、認識のされやすさの6つの項目に関する7ポイントリッカートスケールのアンケート調査を行った。平均した結果を図9.2に示す。

認識のされやすさを除く全ての項目において音付ジェスチャの方が良い結果となった。

誤認識のしにくさが両手法で高くなった要因としては、今回の実験では動画の操作を集中して行わせたため、ジェスチャ以外の動作が少なかった点があげられる。

スムーズな操作のしやすさに顕著な差が表れた。この結果から音付ジェスチャにより、連続ジェスチャが可能となったことが分かる。

連続ジェスチャが可能となったことで、疲労度も軽減できたと考えられる。「ジェスチャのみの場合だと手を静止しておかなければならなかったため、疲れた」という意見も得られた。

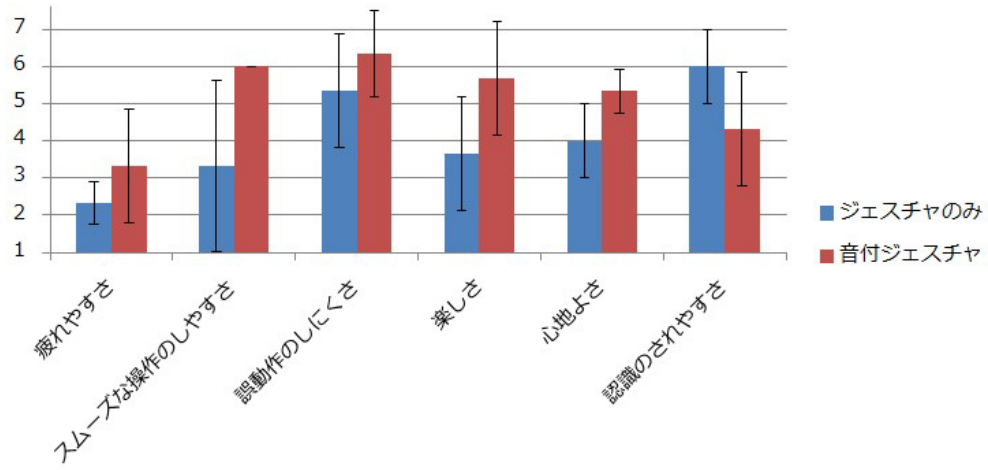


図 9.2: アンケート結果

第10章 提案手法“回転台座”

10.1 設計方針

本研究では深度情報と音情報の組み合わせに注目しており，Kinect の備える 4 つのマイクからの音情報による音源の位置推定 (ビームフォーミング) を利用することとした．具体的には，推定された位置が深度カメラの法線方向に近くなるよう Kinect を水平面で回転 (yaw) させる．そのために PC から回転角度を指定可能な回転台座を試作した (構成を図 10.1 に示す)．

台座の駆動には 12V のバイポーラ型のステッピングモータを用い，これをマイクロステップ制御可能なコントローラで制御する．PC からの制御コマンド (回転角の指定，回転や停止の指示，パラメタ設定など) はシリアル回線を介してマイコンで受け，適切なコマンドに変換後，SPI(Serial Peripheral Interface) を介してモータコントローラに送っている．

なお，まずは Kinect に対して前面の 180 度をカバーできれば良いと考え，可動範囲は約 ± 90 度とした．

また，一度の回転に対して 90~180 度/秒の角速度で動作させることとし，三段階 (fast,middle,slow) の設定を可能とした．実際の回転速度を計測したところ，fast が約 98.4 度/s，middle が約 57.9 度/s，slow が約 25.3 度/s で回転可能であることが分かっている．

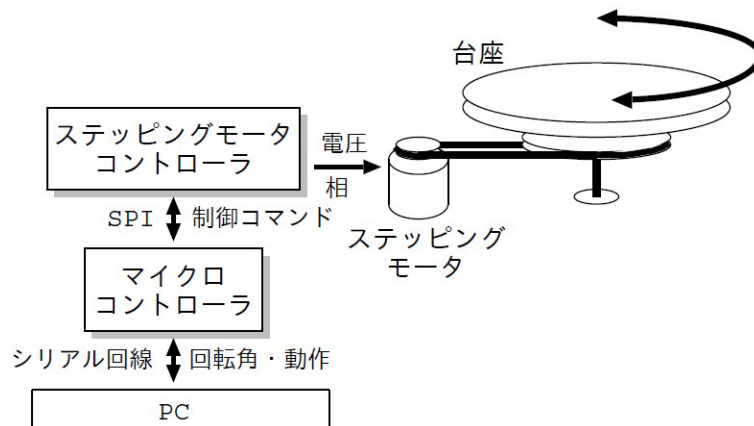


図 10.1: Kinect を回転させる台座の構成図

10.2 機能

回転台座には次の二つの機能がある．

10.2.1 ユーザの検出

Kinect の音源位置推定を用いて単発音が発生した角度を求める．単発音の認識には前述した方法を用い，音源位置推定には Kinect windows sdk のライブラリを用いる．音が発生した角度へ

Kinect を回転させ、ユーザを認識した場合に Kinect の中央がユーザの中央と一致するように回転させる。ユーザ認識には Kinect windows sdk のライブラリを用いる。

これにより、Kinect の計測可能範囲外からでもジェスチャ操作が可能となる（図 10.2）。

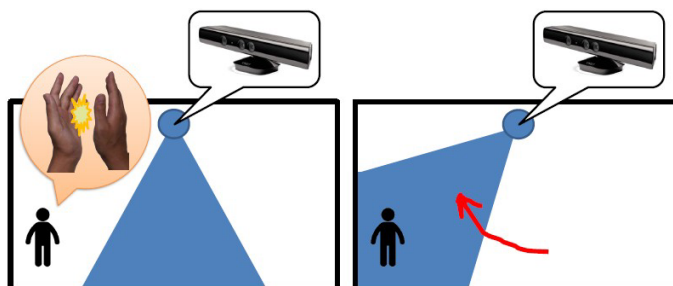


図 10.2: ユーザの検出

10.2.2 ユーザの追跡

ユーザが移動し、Kinect の中央からずれた場合、ユーザを中央に合わせるように Kinect を回転させる。これは Kinect window sdk ライブラリのユーザ認識を常に行っておくことにより、可能となる。

これにより、どこへ移動してもジェスチャ操作が可能となる（図 10.3）。

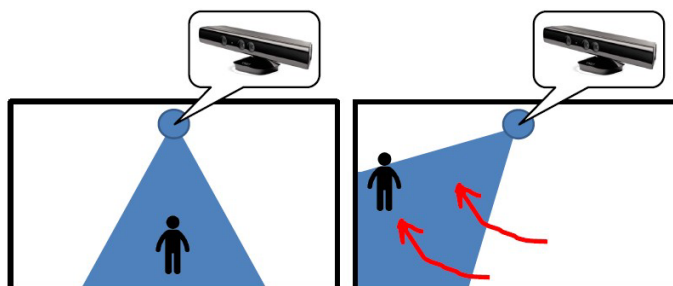


図 10.3: ユーザの追跡

10.3 性能評価実験：概要

実装したシステムの性能を評価するため、実験を行った。実際に完成した回転台座を図 10.4 に示す。

実際の使用状況を考慮し、カメラの計測可能範囲外にユーザが立っていることを想定した。

部屋の端にカメラを設置し、180° の範囲をカバーできるかを実証する。また、どの程度の時間でユーザ検出が可能かを評価する。



図 10.4: Kinect を回転させる台座の完成図

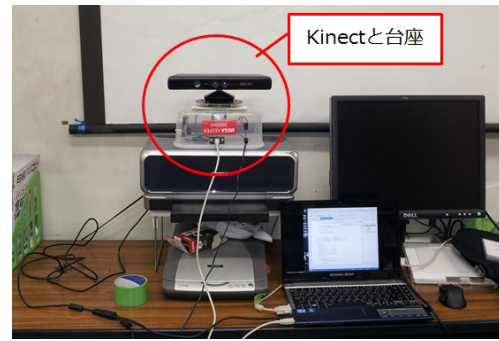


図 10.5: 実験装置

10.3.1 実験環境

研究室の一室（約 7.5m × 3.6m）で行った。

左 90°, 左 60°, 左 30°, 右 30°, 右 60°, 右 90° の 6 箇所から単発音を発し、単発音を発した時間からユーザを認識するまでの時間を求める。6 箇所から 3 種類の回転速度（fast,middle,slow）で各 3 回ずつ測定を行い、合計 54 回のデータを取得した。

実験装置を図 10.5 に示す。Kinect と各ユーザの配置を図 10.7 に示す。

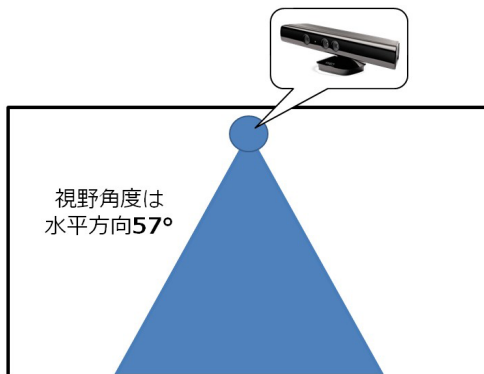


図 10.6: Kinect の水平視野角

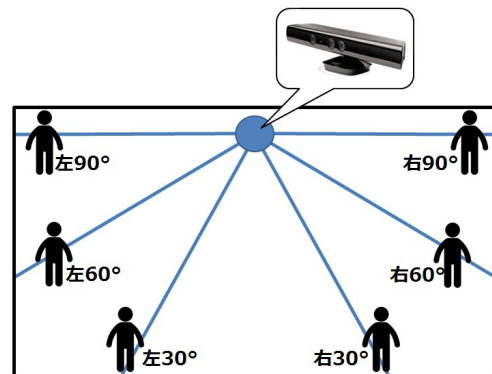


図 10.7: 音源の位置

10.4 結果

各音源の位置から 3 回ずつ音を発生させ、ユーザを認識するまでの時間を平均したものを表 10.1 に示す。

表 10.1: ユーザ検出時間 (s)

	左 90°	左 60°	左 30°	右 30°	右 60°	右 90°
回転速度 : fast	2.36	0.81	0.66	0.87	0.74	1.98
回転速度 : middle	1.90	1.19	0.98	1.06	1.27	1.83
回転速度 : slow	3.37	1.88	1.09	1.52	2.00	3.06

10.5 考察

角度が大きくなるにつれて時間も長くなっていることが分かる。また、回転速度が速くなるにつれて時間も長くなっていることが分かる。各ユーザ検出時間の平均は、fast が 1.24s, middle が 1.37s, slow が 2.15s である。

この結果から最も時間のかかる slow : 左 90° でも 3.37s でユーザ検出が可能である事が分かる。これは先行研究 [12] の約 4s という結果よりも良い値である。

ユーザ検出時間からユーザ追跡可能な速度を計算した。slow : 左 30° と slow : 左 60° のユーザ検出時間の差 0.79s を 30° の回転にかかる時間とする。半径 3m, 30° の扇形の弧の長さは約 1.57m であるため、slow でも約 2.0m/s の速さでユーザが動いても追跡できることになる。一般的な人の歩く速度は 1.1m/s と言われているため、これは十分実用に耐える回転速度であると言える。

なお、同様の計算をしたところ、fast だと 10.3m/s, middle だと 7.4m/s で追跡可能であることが分かった。

第11章 おわりに

本稿では、音付ジェスチャの提案と実装を行った。まず、24種の音付ジェスチャに対する、単発音の大きさや好みなどの個人差についての調査を行った。典型的な利用場面での環境音と比較したところ、どの単発音も識別できる可能性があることが分かった。しかし、ユーザに好まれないものもあり、取捨選択が必要である。この結果を受けて、拍手、机タップ、机ノック、指パッチンのシンプルな4種の音付ジェスチャが認識可能な試作システムを実装した。

また、以下の二点を改良した。一つ目は単発音分類アルゴリズムの改善であり、4種類のアルゴリズムに対する調査を行った。被験者から採集した単発音データをもとに単発音分類実験を行ったところ、高い精度で音の種類を分類することができた。二つ目はシステムのカバー範囲の拡大であり、Kinectを水平面で回転させ、広角化を実現するシステムを試作した。ユーザがKinectのカバー範囲外から操作したとしても、数秒でユーザを認識できることが分かった。

音付ハンドジェスチャの有用性を評価するために、評価実験を行った。4種類の単発音と8種類のジェスチャを組み合わせた14種類の音付ハンドジェスチャに対し、認識精度に関する評価実験を行った。実験の結果、認識精度は66%と低い値となってしまった。しかし、単発音の種類を2種類と少なくしたところ、92%と高い精度で認識可能であることが分かった。この場合、音付ハンドジェスチャの種類数が9となり、大幅なジェスチャ数の増加は達成することができなかったが、単発音の分類方法の改善により達成可能であると考えられる。既存手法との比較実験からは、音付ジェスチャが効率的に動作可能であることが分かり、有用性が示された。

また、深度カメラ1台での広角化を実現した回転台座を実装した。評価実験により十分な速度でユーザの検出、追跡が可能である事が分かり、ハンドジェスチャの問題点のひとつである計測可能範囲の制限を解決することができた。

今後の課題としては、単発音の分類方法の見直し、雑音環境下での音付ジェスチャの認識精度の調査があげられる。

参考文献

- [1] Mistry, P., Maes, P. and Chang, L.: WUW - Wear UrWorld - A Wearable Gestural Interface, CHI, pp. 4111-4116 (2009).
- [2] Bailly, G., Muller, J., Rohs, M., Wigdor, D., and Kratz, S.: ShoeSense: A New Perspective on Hand Gestures and Wearable Applications, CHI, pp. 1239-1248 (2012).
- [3] 池司, 中州俊信, 岡田隆三: 自然な手振りによるハンドジェスチャーユーザーインタフェース, pp. 36-39 (2012).
- [4] Henze, N. and Hesselmann, T.: Free-Hand Gestures for Music Playback: Deriving Gestures with a User-Centred Process, MUM, No. 16 (2010).
- [5] 長谷川秀太, 赤池英夫, 角田博保: 姿勢を考慮したハンドジェスチャーを利用する機器操作の提案・評価, 情報処理学会 HCI 研究会, Vol. 2011, No. 24, pp. 1-6 (2012).
- [6] Yinlin, L., Christoph, G., Jochen, D., Wolfgang, S. and Monika, F.: An acoustic interface for triggering actions in virtual environments, SPIE, Vol. 5444, pp. 246-251 (2004).
- [7] Harrison, C., Schwarz, J. and Hudson, S. E.: TapSense: Enhancing Finger Interaction on Touch Surfaces, UIST, pp. 627-634 (2011).
- [8] 尾崎晃, 宮島千代美, 西野隆典, 北岡教英, 武田一哉: マイクロコンピュータを用いた単発音入力インタフェースの開発, 情報処理学会研究報告, Vol. 2007, pp. 1-4 (2007).
- [9] 長谷川伸吾, 赤池英夫, 角田博保: 両手で把持したタブレットのための入力手法の提案と評価, 情報処理学会 HCI 研究会, Vol. 2011, No. 30, pp. 1-5 (2012).
- [10] Fleeer, D. and Leichsenring, C.: MISO: A Context-Sensitive Multimodal Interface for Smart Objects Based on Hand Gestures and Finger Snaps, UIST, pp. 93-94 (2012).
- [11] 神原啓介, 塚田浩二: オノマトペン, コンピュータソフトウェア, Vol.27, No.1, pp.48-55 (2010).
- [12] Wilson, A. D., Benko, H., Izadi, S. and Hilliges, O.: Steerable Augmented Reality with the Beamatron, UIST, pp. 413-422 (2012).
- [13] Tomari, R., Kobayashi, Y. and Kuno, Y.: Wide Field of View Kinect Undistortion for Social Navigation Implementation, ISVC, pp. 526-535 (2012).
- [14] 株式会社ヒューマンインタフェース: あふれるリモコン (2008).
- [15] 則枝 真, 三橋秀男: ArmKeypad:腕へのタップ入力による機器操作, 情報処理学会インタラクション (2011).
- [16] 齋藤央, 赤池英夫, 角田博保: 単発音を利用したハンドジェスチャーインタラクションの提案と評価, ヒューマンインタフェースシンポジウム, pp. 355-358 (2012).

- [17] Zwicker, E.: Subdivision of the Audible Frequency Range into Critical Bands(Frequenzgruppen), J.Acoust.Soc.Am, Vol. 33, No. 2, p. 248 (1961).
- [18] Vesa, S. and Lokki, T.: AN EYES-FREE USER INTERFACE CONTROLLED BY FINGER SNAPS, DAFx, pp. 262-265 (2005).
- [19] 齋藤央, 赤池英夫, 角田博保: 単発音を利用したハンドジェスチャインタラクション, 情報処理学会 HCI 研究会, Vol. 2012, No. 5, pp. 1-6 (2013).

謝辞

二年間に渡り御指導いただきました角田先生，赤池先生を始め，実験に快く協力していただいた研究室の方々に心から感謝を申し上げます．

付録 実験アンケート

実験後に回答してもらったアンケート用紙を添付する。アンケートは予備調査，単発音データ採集実験，音付ジェスチャ認識実験，比較実験の4種類ある。

齋藤央：予備調査アンケート

実験にご協力いただきありがとうございました。引き続きアンケートをお願いします。アンケートに関して質問がある場合は遠慮せずお聞きください。

お名前：

問1.以下の1~24番のジェスチャ（実験でやってもらった動作のこと）について動作のしやすさを7段階で教えてください

動作しにくい ← 1 2 3 4 5 6 7 → 動作しやすい

※以下の「：」の後に番号を記入

～拍手系～

1. 拍手（右手が上 ver）：
2. 拍手（左手が上 ver）：
3. 拍手（合わせる ver）：

～右手でタップ系～

4. 右手で掌をタップ：
5. 右手で前腕裏をタップ：
6. 右手で前腕表をタップ：
7. 右手で上腕をタップ：
8. 服を着て右手で前腕裏をタップ：
9. 服を着て右手で前腕表をタップ：
10. 服を着て右手で上腕をタップ：
11. 右手でももをタップ：
12. 右手でテーブルをタップ：
13. 右手でテーブルをノック：

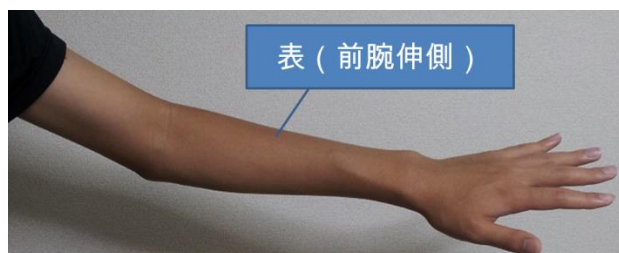
～左手でタップ系～

14. 左手で掌をタップ：
15. 左手で前腕裏をタップ：
16. 左手で前腕表をタップ：
17. 左手で上腕をタップ：
18. 服を着て左手で前腕裏をタップ：
19. 服を着て左手で前腕表をタップ：
20. 服を着て左手で上腕をタップ：
21. 左手でももをタップ：
22. 左手でテーブルをタップ：
23. 左手でテーブルをノック：

～その他～

24. 指パッチン：

※前腕裏が前腕屈側で、前腕表が前腕伸側です



問2.実際に機器操作として使える場合、使いたいジェスチャ番号とその理由を教えてください（問 1 の 1~24 の番号で回答）。複数可。

問3.今回やってもらったジェスチャで機器を操作するとしたら何か問題がありますか？（例：家の中で机をバンバン叩きたくない など）複数可。

問4.他に単発音を利用したジェスチャで使ってみたい動作はありますか？（例：ほっぺを叩くなど）複数可。

問5.その他気づいた点があればお書きください

ご協力ありがとうございました。

齋藤央：単発音データ採集実験アンケート

実験にご協力いただきありがとうございました。引き続きアンケートをお願いします。アンケートに関して質問がある場合は遠慮せずお聞きください。

お名前：

問1.安定した音が出せましたか？

安定して出せたかどうかを7段階で教えてください。

また、各単発音に対してコメントがあればお書きください。

不安定 ← 1 2 3 4 5 6 7 → 安定

※以下の「：」の後に番号を記入

1. 拍手：

コメント ()

2. 机タップ：

コメント ()

3. 机ノック：

コメント ()

4. 指パッチン：

コメント ()

問2.その他気づいた点があればお書きください

ご協力ありがとうございました。

齋藤央：音付ジェスチャ認識実験アンケート

実験にご協力いただきありがとうございました。引き続きアンケートをお願いします。

お名前：

問1.以下の 1~14 番の音付ジェスチャについて認識のされやすさ，動作のしやすさ（認識率とは関係なく），疲労度を 7段階で教えてください。

認識しにくい ← 1 2 3 4 5 6 7 → 認識されやすい

動作しにくい ← 1 2 3 4 5 6 7 → 動作しやすい

疲れた ← 1 2 3 4 5 6 7 → 疲れない

※以下（認識のされやすさ，動作のしやすさ，疲労度）の用に番号を記入

～拍手系～

1. ↑ +拍手：(, ,)

2. ↓ +拍手：(, ,)

3. ← +拍手：(, ,)

4. 左下+拍手：(, ,)

～テーブル系～

5. ↓ +テーブルタップ：(, ,)

6. ↓ +テーブルノック：(, ,)

～指パッチン系～

7. ↑ +指パッチン：(, ,)

8. ↓ +指パッチン：(, ,)

9. → +指パッチン：(, ,)

10. ← +指パッチン：(, ,)

11. 右上+指パッチン：(, ,)

12. 左下+指パッチン：(, ,)

13. 奥 +指パッチン：(, ,)

14. 手前+指パッチン：(, ,)

問2.音付ジェスチャを以下の環境で使える場合、使いたい音付ジェスチャ番号（問1の番号）と具体的にどの操作に使いたいかを教えてください。（例：5で部屋の照明をつけたい）

環境例：動画再生プレーヤー，照明・空調操作，キッチン（電子レンジなど），プレゼン

問3.その他気づいた点，改善案があればお書きください

ご協力ありがとうございました。

齋藤央：比較実験アンケート

実験にご協力いただきありがとうございます。引き続きアンケートをお願いします。

お名前：

問1. 疲労度

疲れた←1 2 3 4 5 6 7→疲れない

- ジェスチャのみ：
- 音付ジェスチャ：

問2. スムーズな（連続）操作のしやすさ

連続操作しづらい←1 2 3 4 5 6 7→しやすい

- ジェスチャのみ：
- 音付ジェスチャ：

問3. 認識のされやすさ

認識されにくい←1 2 3 4 5 6 7→しやすい

- ジェスチャのみ：
- 音付ジェスチャ：

問4. 誤動作のしにくさ（誤動作：意図していない動きがジェスチャとして認識されてしまうこと）

誤動作しやすい←1 2 3 4 5 6 7→しにくい

- ジェスチャのみ：
- 音付ジェスチャ：

問5. 楽しさ

楽しくない←1 2 3 4 5 6 7→楽しい

- ジェスチャのみ：
- 音付ジェスチャ：

問6. 心地よさ

心地よくない←1 2 3 4 5 6 7→心地よい

- ジェスチャのみ：
- 音付ジェスチャ：

問7.その他気づいた点，改善案があればお書きください