第１章　はじめに

近年、コンピュータの機能が多種多様になるにつれ、一つのコンピュータに機能が集中し、非常に複雑なものとなっている。このように、コンピュータが複雑になるにつれて、多種多様なインターフェイスが求められるようになり、またその使い勝手の向上が要求されるようになっている。その中でも、空間情報の入力インターフェイスは、作業効率に大きな影響を及ぼす。そこで、作業効率を低下させずに、人が形状のデザインなど幅広い情報を入力する為の入力インターフェイスが要求され、更に、その入力方法は理解しやすいものである必要がある。現在までに、これらの空間情報を入力するインターフェイスには、バーチャルグローブやビデオカメラを用いた身振りインターフェイスなど様々な方法が検討されてきたが、まだ十分とは言えない。

そこで本研究では、空間情報の入力インターフェイスとして、電子楽器テルミンに着目した。テルミンとは古くに発明された電子楽器であり、空間上での人の手の動作に対応して音程を変化させることができる。この手の動作で音程を変化させる楽器をインターフェイスとしての視点でとらえると、まさに空間情報を入力していることになり、また、入力時の手の位置の変化に対する操作量のフィードバックとして、音程が利用されることになる。

従来の空間情報の入力インターフェイスでは、一般に操作量のフィードバックは取り込まれた空間上での手の動きをディスプレイに表示して確認しているので、視覚にだけ頼っていることになる。より良いインターフェイスのためには視覚以外の補助的なフィードバックの利用が有効であると考えられる。具体的には、先ほど述べた音程のフィードバックを用いることにより、位置の変化が視覚と聴覚の操作量で確認することができるので、操作性が更に向上するのではないかと考えられる。

本研究ではまず、この音程のフィードバックがある場合とない場合についての特性を検証することにした。具体的には、非接触型のインターフェイスとして従来より用いられているビデオカメラを使った入力システムを作製し、音程のフィードバックの有無による操作性の比較を行った。また、実際のテルミンを利用した入力装置を作製して同様の実験を行い、新しいインターフェイスとしての可能性を探ることにした。

本稿では、空間情報を扱う 2 つの非接触型インターフェイスについて、音程による操作量のフィードバックについての操作性の比較実験を行いその結果について述べ、考察する。

第２章　電子楽器テルミンを用いたインターフェイスの可能性

この章ではテルミンを使った入力インターフェイスについての可能性について探る。そのためまず、テルミンについてその歴史的背景及びその発音原理を述べる。そして、発音原理を理解したところで実際にテルミンから発音される音程の変化と手の位置の変化についての出力特性を示す。

２．１　歴史的背景

テルミンとは 1919 年、ロシアの物理学者レフ・テルミンによって発明された世界最古の電子楽器である。図 2.1 にテルミンの外観図を示す。テルミンはあらゆる電子楽器の中でも際立った特長を持っており、その演奏方法は 2 本のアンテナに手を触れることなく音を発生させる楽器である。操作の方法は縦方向と横方向に設置された２本のアンテナと、手との間の静電容量の変化を利用して音を変化させて演奏する。

図 2.1　テルミンの外観

２．２　発音原理

２．２．１　音程制御

図 2.2 はテルミンの全構成図を示したものである。テルミンの楽音は図のビート発振器によって作られる。ビート発振器には２つの発振器があり、それぞれに共振回路が組み込まれている。そして、その２つの発振器から出力される共振周波数を検知器で合成し、0～3kHz の周波数の音をうなりとして取り出している。3kHz の変化で基準となるドの音から 3.5 オクターブ上の音が発生する。図 2.2 で固定発振器は約 285kHz で一定の共振周波数を出力し、そしてもう１つの可変発振器ではアンテナの静電容量の変化に対応して約 282kHz~285kHz の間で共振周波数が変化する。図 2.2 よりピッチアンテナ回路は可変発振器に接続されており、静電容量が増加すると可変発振器の周波数は 3kHz 減少する。この変化により音程としては約 5 オクターブ変化

させることが可能となる。

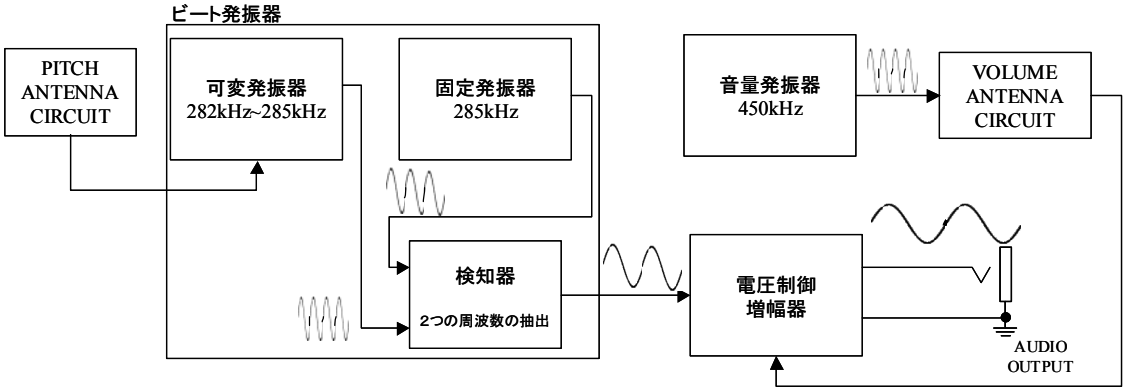


図 2.2　テルミンのブロック図

２．２．２　音量制御

音量を変化させるアンテナは音量発振器の出力に接続されている。そして図 2.2 の音量アンテナ回路から発生する共振周波数と音量発振器の周波数が等しい時、電圧が最大となる。もし、ボリュームアンテナの手の静電容量が増加、つまりアンテナに手を近づけると電圧は減少し音量は下がる。

２．２．３　手とアンテナとの静電容量について

テルミンの回路にはいくつかの共振(共鳴)回路が用いられている。図 2.3 にイメージ図と等価回路を示す。まず、空間上での手をアンテナに近づけるとこの両者の間で手が接地側、アンテナがもう一つのプレートとなり可変コンデンサができる。そして、手の位置の移動によって静電容量が変化し、手をアンテナに近づけると静電容量(1pF 以下)は大きくなる。またその他に、アンテナの大きさに依存するアンテナコンデンサ(10-15pF)、そしてアンテナコイルがある。そして可変コンデンサ、アンテナコンデンサ、アンテナコイルが共振回路を生成している。この時の共振周波数は音程を変化させる発振器で約 285kHz、そして音量を変化させる発振器で約 450kHz である。

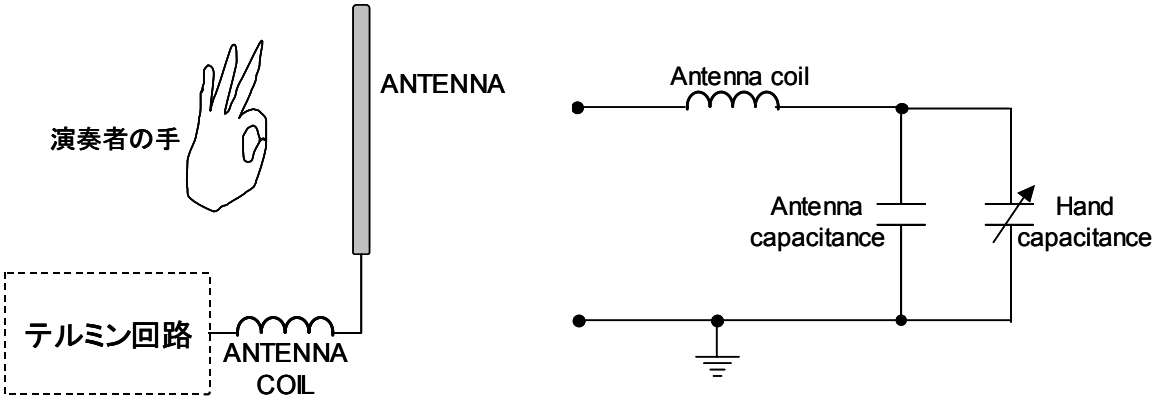


図 2.3 イメージ図と等価回路

２．３　音程の出力特性

次に手の位置と音程の出力特性につての実験を行う。実験は図 2.4 の音程を制御するアンテナの中央部から 5cm 間隔で周波数を計測する。表 2.1 にアンテナからの距離とそれに対する音の周期、および周波数を示す。尚、実験は 2 回行った。

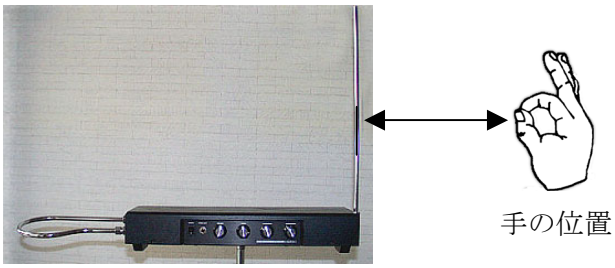


図 2.4　実験図

表 2.1　実験結果

距離 (cm)	周期1 (msec)	周期2 (msec)	周波数1 (Hz)	周波数2 (Hz)
5	0.90	0.65	1117.3	1538.5
10	1.00	1.10	1000.0	909.1
15	1.63	1.40	615.4	714.3
20	2.25	2.00	444.4	500.0
25	3.00	3.00	333.3	333.3
30	4.00	3.40	250.0	294.1
35	5.00	4.25	200.0	235.3
40	7.00	5.00	142.9	200.0
45	8.00	5.40	125.0	185.2
50	10.00	7.00	100.0	142.9
55	12.50	8.00	80.0	125.0
60	15.00	10.00	66.7	100.0
66	60.00	10.00	16.7	100.0

表 2.1 の結果より距離と周波数を対数でとった関係について図 2.5 に示す。

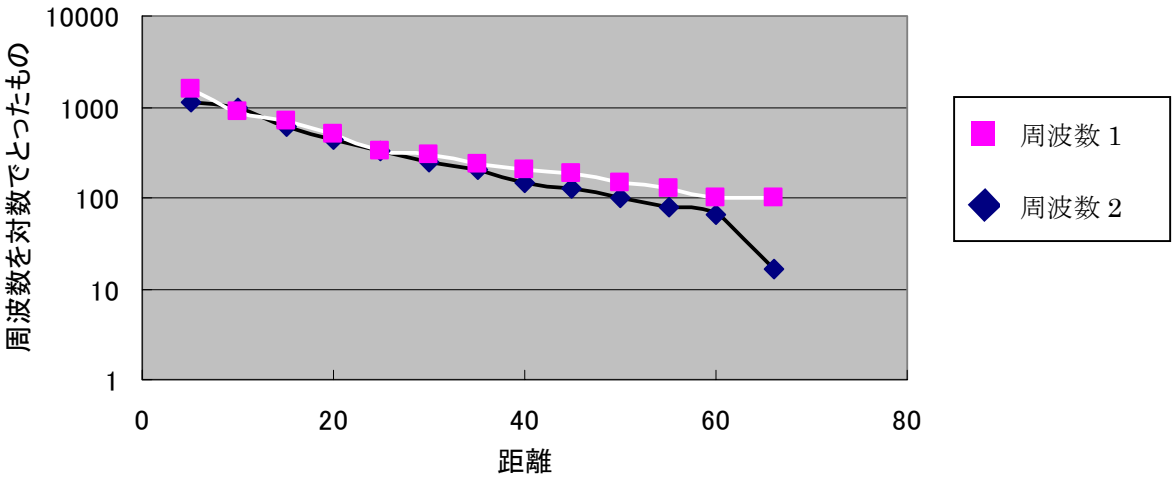


図 2.5　距離-周波数特性

２．４　考案するテルミンを用いたインターフェイス

これまでテルミンについての動作原理、およびその出力特性について調べてきたが、実際にこの特性を生かしてテルミンを入力インターフェイスとして使用することを考える。そこでまず、このテルミンを 1 つ用いて入力装置を作製し、空間を扱った入力装置について検討してみる。そして将来的にはこの空間情報を使ったテルミンを複数用いれば多次元の入力操作が可能になると考えている。具体的には図 2.6(右)のような 3 次元の入力装置を考案中である。

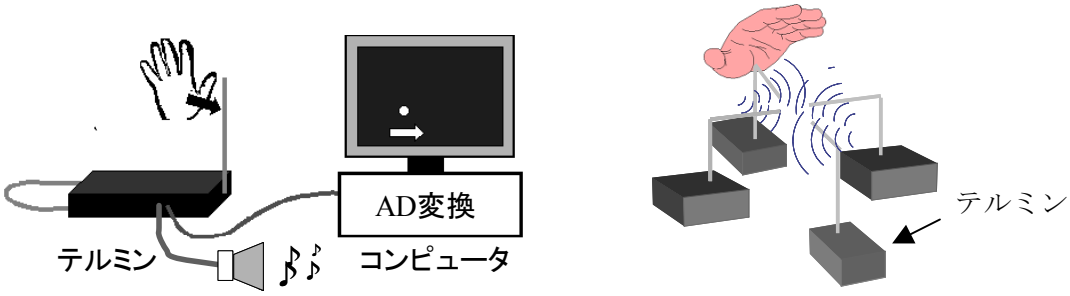
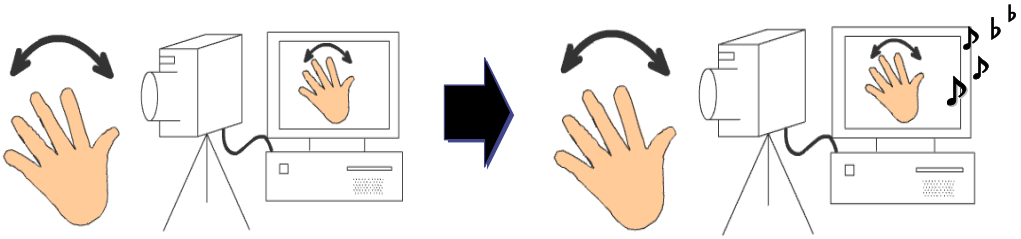


図 2.6　テルミンを用いた入力装置

第３章　音程のフィードバックを用いた動きインターフェイス(1)

電子楽器テルミンの操作方法は空間上での手とアンテナとの距離の変化が音程の変化となってフィードバックする。このことを利用して、従来からの非接触型のインターフェイスの方法であるビデオカメラを用いて、コンピュータに位置情報を与え、コンピュータから視覚と聴覚による操作量のフィードバックがなされる装置を作製する。(図 3.1)そして、音程のフィードバックがある場合とない場合に分けて、入力のしやすさにどのような影響があるのかについての調査実験を行う。なお、作製に当たってはmicrosoft Visual C++のプログラム環境を用いて行った。



(a)従来のインターフェイス　　(b)考案したインターフェイス

図 3.1　従来と考案したインターフェイス

３．１　処理手順

まず、空間上での手の動きをビデオカメラによって取り込む。そして、ビデオカメラによってキャプチャされた動画に対して閾値処理を行い、閾値を超えた点の重心を求めてその位置により発音と消音、音程を制御する。(図 3.2)また、本来はテルミンの演奏方法に近づけるためには手の認識処理を行うべきであるが、確度を上げた認識処理を組み込むと非常に大掛かりになるため、ここでは演奏にあたっては手にペンライトなどを持たせて明るさの差をはっきりさせて閾値処理で位置を割り出す。

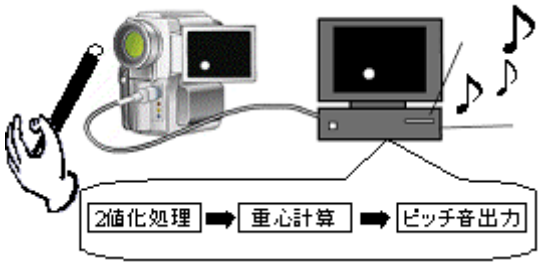


図 3.2　処理手順

３．１．１　動画の取り込み

ビデオカメラによって取り込まれる動画はリアルタイム処理を行わなければならない。そこで、このリアルタイム処理に DirectX9.0 の DirectShow を用いて動画のキャプチャを行う。DirectX とはマイクロソフト社によって開発されたマルチメディアアプリケーションを作成するためのアプリケーションプログラミングインターフェイス(API)のセットであり、グラフィックスやサウンド面においてハードウェアレベルで直接制御する事ができる。その機能の一つである DirectShow は様々なソースのファイルとストリームの高品質なキャプチャと再生を行うことができる。DirectX9 の概要を図 3.3 に、DirectShow の構成を図 3.4 に示す。

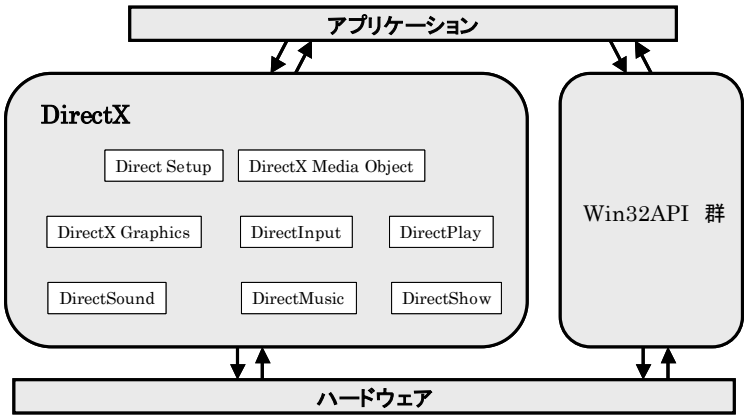


図 3.3 DirectX9 の構成

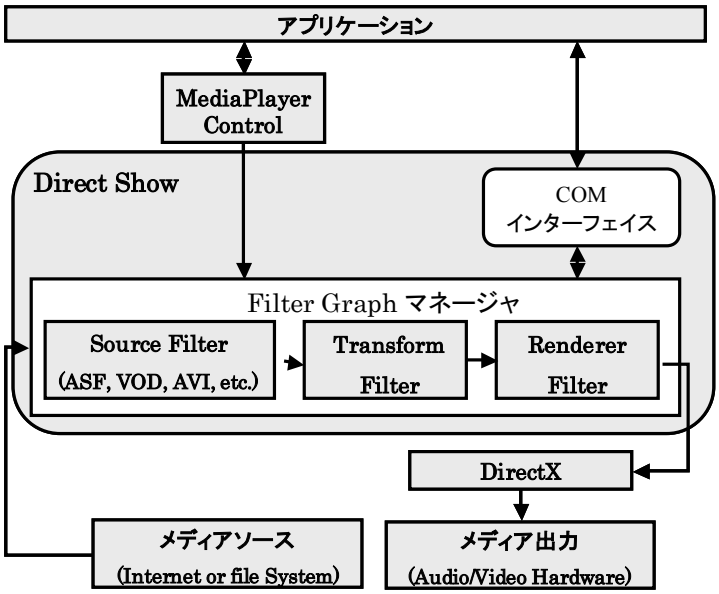
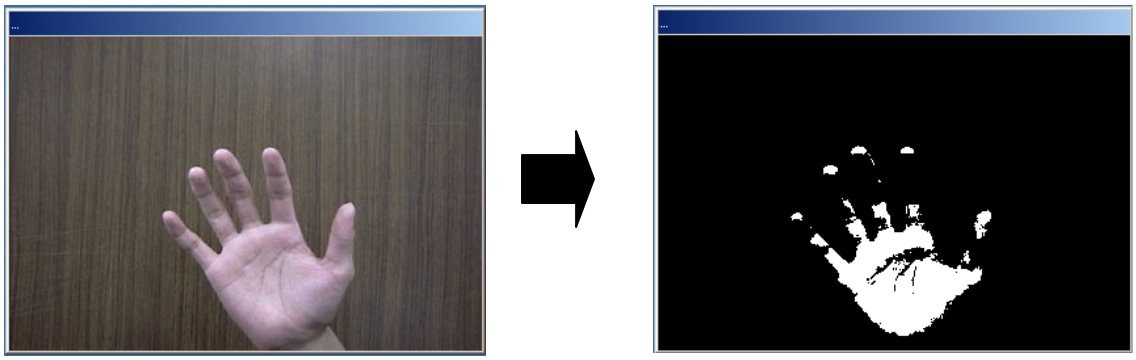


図 3.4 DirectShow の構成

３．１．２　２値化処理

閾値を設定するためビデオカメラによってキャプチャされた動画(図 3.5(a))を１フレームごとの画像で２値化処理を行う。画像の形式にはビットマップ(24bit カラー)画像を用いている。画面サイズは横 320×縦 240 ピクセルとし、動画のフレーム間隔は 1 秒間に 30 フレームとする。

２値化処理では、入力画像の各画素に対して BGR(256 階調)の平均値を求め、閾値以下あれば黒、閾値以上なら白と設定する。図 3.5 に取り込まれた動画に対して２値化処理を行った結果を示す。



(a)処理前　　(b)２値化処理後

図 3.5　２値化処理

３．１．３　重心計算

2 値化処理後、得られた白画素の重心を計算する。重心の計算は x 座標、 y 座標それぞれの中心を計算する。ここで挙げる重心とは位置の平均値のことであり、画面に映っている白画素の中心位置を計算できる。 n ×m サイズの画像データ B[i,j]の重心 x,y は以下の式で求めることができる。

$$\bar{x} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} iB[i,j]}{A} \quad \cdot\cdot\cdot (1)$$

$$\bar{y} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} jB[i,j]}{A} \quad \cdot\cdot\cdot (2)$$

$$A = \sum_{i=1}^{n} \sum_{j=1}^{m} B[i,j] \quad \cdot\cdot\cdot (3)$$

ここで A は画像データ B[i,j]の面積,つまり白の総画素数を示している。

３．１．４　音の出力

次に、重心の位置の移動によって音を鳴らす。音は Windows32　API の Beep 関数を用いて位置の移動ごとにコンピュータから発音される。取り込まれた画面に対して重心を左から右に動かすと 4 ピクセルごとに半音ずつ上がっていく設定にしておく。

３．２　研究方法

３．２．１　実験方法

実験の方法はコンピュータの画面に示された「0」から「5」の数字のボタンに対して指定された数字の入力を行うものとする。（図 3.6）そして、入力の操作性をみる尺度として被験者の操作時間を計測する。

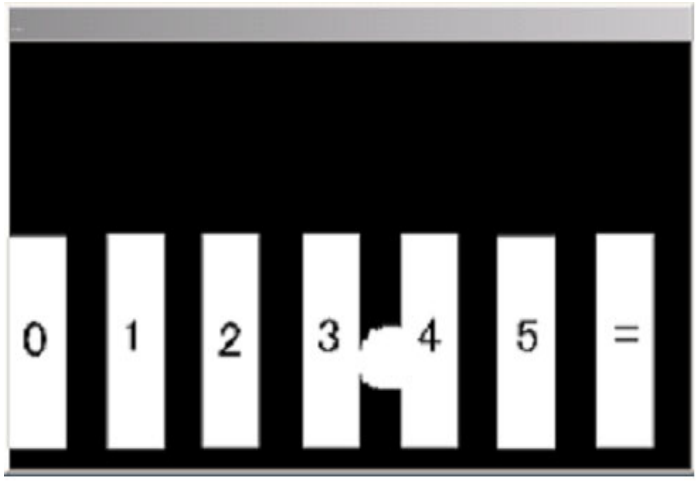


図 3.6　実験操作画面

図 3.6 はビデオカメラからキャプチャされた画面に「0」から「5」と「=」を表示させそれぞれの領域に重心ポイントが重なると、そのエリアに対応した数値が入力されるように設定する。そして、被験者はあらかじめ用意した 5 桁の数字を入力し、5 桁入力し終わると「=」に重心ポイントを重ね次の問題へといく。それを 20 問操作してもらい、その所要時間を計測する。図 3.7 は被験者が実際に行う実験の画面を示したものである。

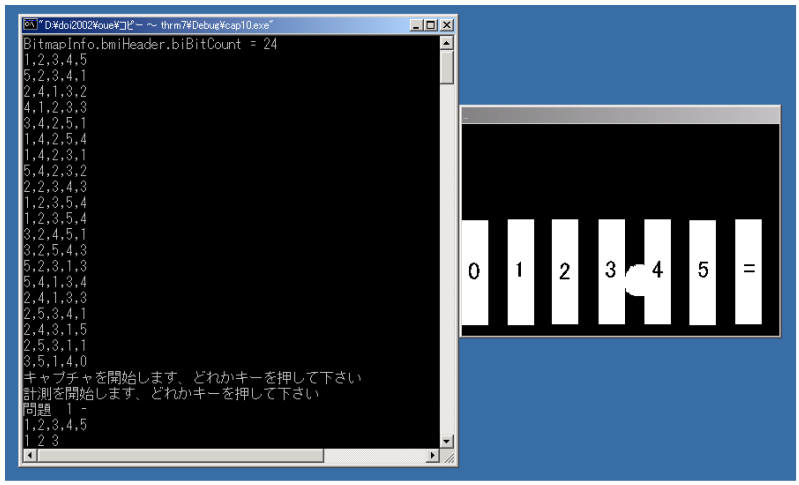


図 3.7　実験画面

３．２．２　実験条件

以下に実験を行う際の条件を述べる。

- 被験者は LED の付いた棒を持って実験を行ない、LED の発光部分を重心点とし操作する。
- 被験者は楽器演奏経験者と未経験者に分ける。これは経験者がポイントによって、音階を的確にとらえられているかを考慮したものである。
- 音あり、音なしに分けてそれぞれ 2 回ずつ実験を行う。
- 問題は 20 問とする。問題はあらかじめ作成したもので 4 回とも同じ問題を操作入力する。

３．３　研究結果

実験はまず被験者 10 人を対照にして行った。次に被験者 40 人に対して被験者 10 人とは違う手法で実験を行った。以下に実験の結果を示す。

３．３．１　計測実験 I

まず、被験者 10 人を対照に楽器演奏経験者 5 人、未経験者 5 人に分けて実験を行う。実験の順序は

音あり　→　**音なし**　→　3 分休憩　→　**音なし**　→　音あり

で行う。これは被験者の操作の慣れを抑えるためであり 3 分の休憩の後、逆順で操作を行う。

測定結果より得られた被験者の測定例を表 3.1 に示す。尚、表 3.1 は左から順に 4 回操作実験を行なった結果を示している。

表 3.1　被験者測定例

	音なし (sec)	音あり (sec)	音なし (sec)	音あり (sec)
被験者1	143	124	119	114
被験者2	146	161	127	152
被験者3	159	155	175	157

表 3.1 を見ると被験者 1 は音あり、音なしに関係なく所要時間が段々と早くなっている。また被験者 2 では音ありに対して音なしの方が早く、被験者 3 の結果では順序に関係なく音なしよりも音が早い結果となった。また、全体の平均的な所要時間は約 2 分半であった。

これらの測定結果をもとに 10 人それぞれの平均値を求め正規化を行う。図 3.8 は正規化された値について全体、経験者、未経験者に分けて示したものである。正規化は被験者の各所要時間を 4 回の平均で割って求めたもので、1 を基準として 1 以上であれば平均時間より遅く、1 以下であれば平均時間より早いことを示している。

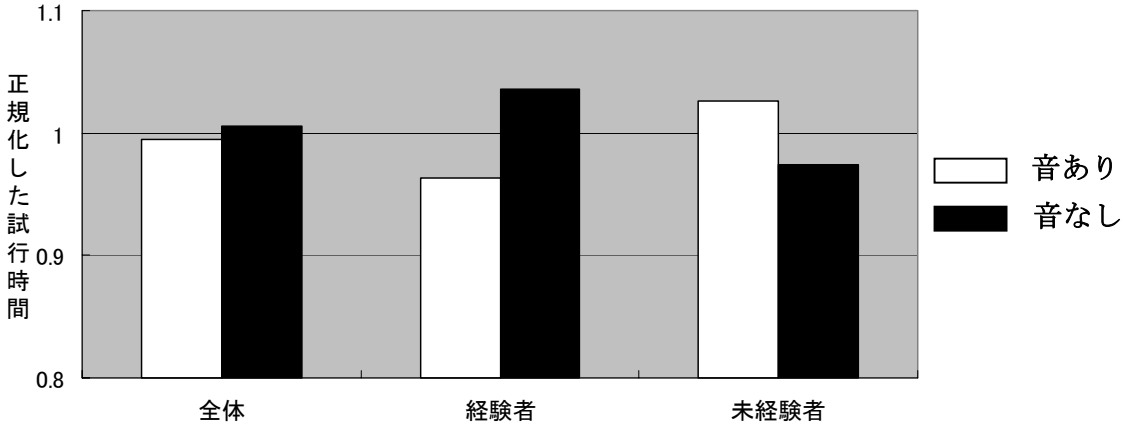


図 3.8　実験結果 I

実験の結果より経験者は音なしより音ありの方が早いのに対して、未経験者では逆に音ありより音なしの方が早い結果となった。全体をみると若干ではあるが音ありの方が早い結果となった。しかし、今回提案した実験の順序では被験者の操作の慣れは防げているといえるが確かなデータであるとはいいにくい。また、被験者の数が 10 人であった。そこで、これらを考慮に入れて次に被験者の数を増やし、実験順序の手

法を変えて実験を行った。

３．３．２　計測実験Ⅱ

実験Ⅱでは被験者の数を 40 人として測定を行う。被験者 40 人での内訳は演奏経験者を 20 人、演奏未経験者を 20 人として実験を行う。また、実験順序の手法を表 3.2 のように A~D のグループ(各 10 人ずつ)に分ける。

表 3.2　被験者の計測順序

グループ	楽器演奏経験の有無	計測順序
A	有	音あり→音なし→音あり→音なし
B	有	音なし→音あり→音なし→音あり
C	無	音あり→音なし→音あり→音なし
D	無	音なし→音あり→音なし→音あり

このように、A グループと B グループ、そして C グループと D グループをそれぞれ比較することで操作順序が相殺されると考えられる。

図 3.9 に被験者 10 人と同様に正規化をした結果を示す。

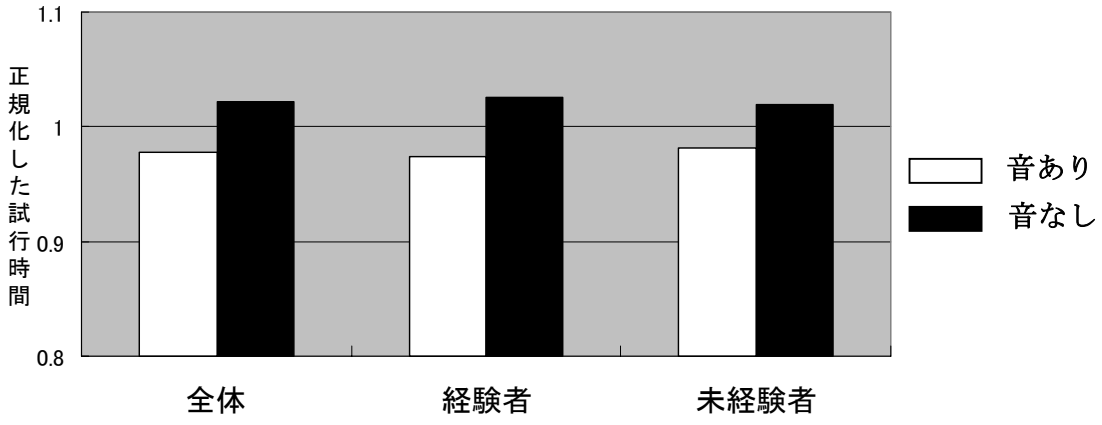


図 3.9　実験結果Ⅱ

図 3.9 より未経験者は音程のフィードバックがない場合に比べある場合の方が早くなった。また若干ではあるが経験者は音ありと音なしの差が大きかった。これは被験者が音程のフィードバックによってボタンの位置を予測できていると考察できる。

第４章　音程のフィードバックを用いた動きインターフェイス(2)

次に、実際のテルミンを用いた入力インターフェイスを作製する。その方法は、テルミンから発生する音をコンピュータに入力し、その周波数を得る。この周波数からコンピュータは位置情報を知ることができ、それを画面上に表示する。この装置を作製した上で、3 章で行った実験と同様に操作性の比較を行う。図 4.1 にテルミンを用いた入力インターフェイスの構成図を示す。

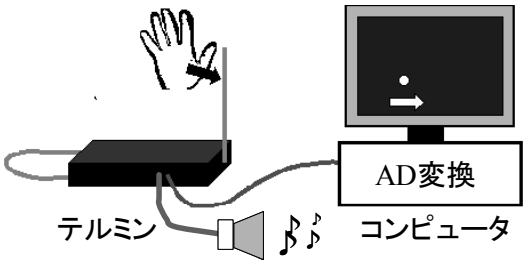


図 4.1　テルミンによる入力装置

4．１　処理手順

４．１．１　音の取得

テルミンの音をコンピュータへ取り込む際には DirectX9.0 を用いてキャプチャする。DirectX9.0 には図 3.3 より、オーディオを扱う DirectSound がある。DirectSound にはオーディオの再生だけでなく、キャプチャつまり録音の機能も含まれている。オーディオのキャプチャには DirectSound 専用の関数、DirectSoundCaptureCreate8()を使ってキャプチャオブジェクトを生成し、同時に IDirectSoundCapture8 インターフェイスを得る。この IDirectSoundCapture8 インターフェイスの CreateCaptureBuffer()を使ってキャプチャするデータのサイズ、フォーマットを設定する。

４．１．２　周波数計算

テルミンからの音をパソコンにキャプチャした後、ウィンドウに取り込んだ音の波形を表示させる。図 4.2 に取り込んだ音の波形図を示す。そして取り込まれた波形から周波数 f を求める。周波数を求める式と図を以下に示す。ここで周期 T はウィンドウに表示された波形の 0 クロスから 1 周期を求めたものである。尚サンプリング周波

数 fs は 11025Hz とする。そして、得られた周波数を重心とし、音程の変化によって重心の位置が移動するよう設定する。

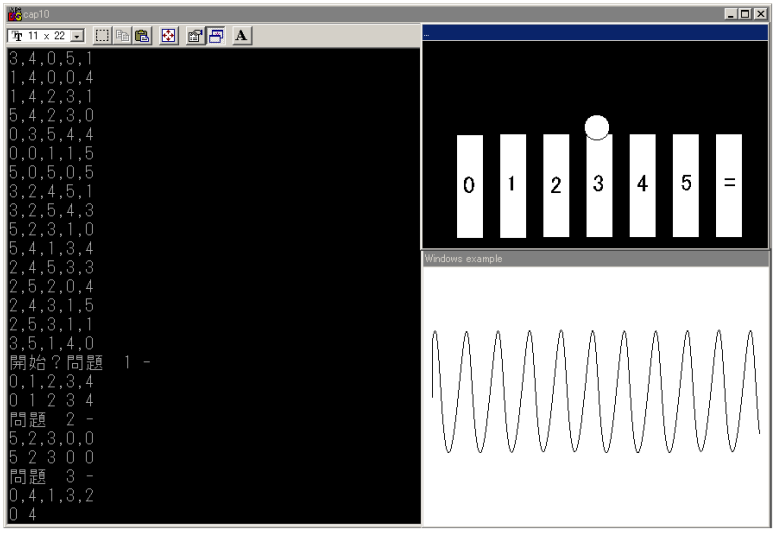


図 4.2　実験画面図

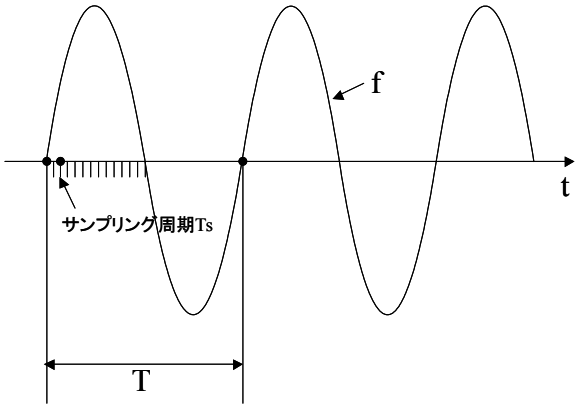


図 4.3　入力波形図

図 4.3 の周波数 f はサンプリング周期 Ts(1/fs)より

$$f = \frac{1}{\frac{1}{fs} \times T} \quad [\mathrm{Hz}]$$

で求めることができる。

４．２　研究方法

実験は 3 章と同様に、画面に表示された「0」から「5」の数字に対して重心ポイントを合わせ、指定された 5 桁の数字の入力を行う。そして 5 桁入力し終わると「=」に重心ポイントを合わせ、それを 20 問操作してもらう。そして、音がある場合とない場合に分けて入力操作を行い、被験者の所要時間を計測する。今回、被験者の人数は 20 人を対照とし、音楽経験者 10、未経験者 10 人に分けて実験を行う。また計測順序は表 4.1 のように A~D の順序に分けて実験を行う。

表 4.1　計測順序

グ　ル	楽器演奏経験の有無	計測順序
A	有	音あり→音　・→　音 ・ 音　・
B	有	音なし→音あり→音なし→音あり
C	無	音あり→音なし→音あり→音なし
D	無	音なし→音あり→音なし→音あり

４．３　研究結果

図 4.4 に得られた測定データの正規化したグラフを示す。

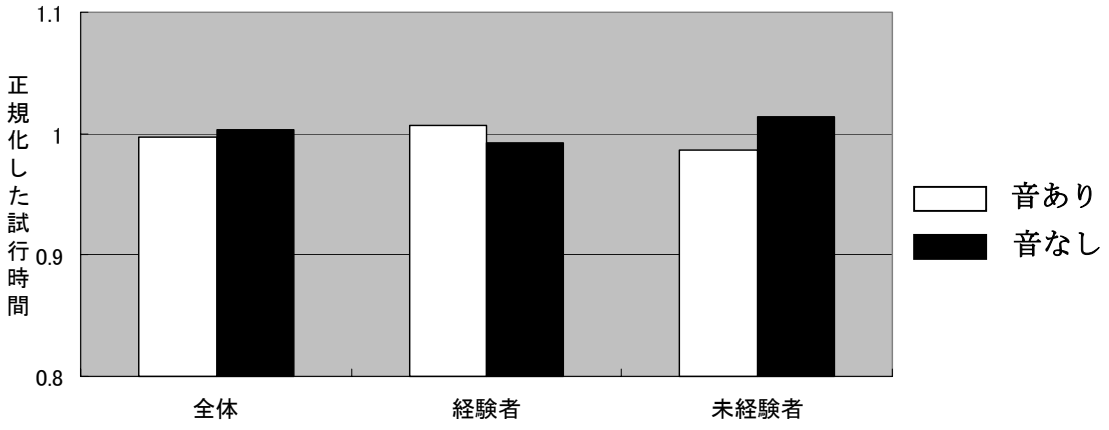


図 4.4　実験結果

実験の結果より、経験者は音ありよりも音なしの方が早く、未経験者は音なしより音ありの方が早い結果となった。また、全体で観ると若干ではあるが音ありの方が早いという結果となった。今回の結果は本来予想していた経験者の方が音階を的確にとらえられているという推測とは逆の結果となった。

５章 おわりに

本研究は最初に述べた通り、非接触型の入力インターフェイスに音程のフィードバックを付け加えた装置を作製し、その操作性の比較を行ってきた。結果としては、ビデオを用いた入力インターフェイスの実験では被験者 40 人に対して、音ありの方が音なしの場合より所要時間が早くなり、明確な差として表れた。しかし、ビデオを用いての手の認識、具体的には 2 値化や重心の計算などは、それに要する計算時間のタイムラグが生じた。そのため入力装置としては不備の残る点である。

テルミンを使っての音程フィードバックを利用した実験では、被験者 20 人に対して音あり、音なしの差がほぼなかった。この原因として考えられるのは、まずビデオによる実験の時よりもフィードバックされる音が連続的であり被験者に音程の違いが分かりづらかったと考えられる。また、ビデオを用いた装置では位置の変化によって音程が変化するのに対して、テルミンを用いた装置は周波数の変化、つまり音程の変化によって位置が変化するのでそのあたりが今回音ありと音なしの差が縮まった原因と考えられる。そこで今後はテルミンからの直接の音をフィードバックとせず、位置の変化によって音が出力されるように実験を行う必要があると考える。

これまで、音程のフィードバックを利用した入力インターフェイスの基礎実験において操作性の比較を行ってきて、入力を行う際に視覚のフィードバックに加え、聴覚のフィードバックがある方がインターフェイスとして優れているということがわかった。しかし、今回の実験だけでは音程のフィードバックの影響のみで操作性が向上しているとは一概にはいえない。そこで今後はより被験者の人数を増やし、引き続き実験を行い、検討してく必要があると考える。

参考文献

[1] 桐島，佐藤，千原，「身振り動画像によるユーザインターフェース」，
http://chihara.aist-nara.ac.jp/people/95/tosiyu-k/research-j.html

[2] 大植，土井，「音程フィードバックを用いた動きインターフェイスの基礎実験」
平成 14 年電気関係学会関西支部連合大会講演論文集 G14-3

[3] 大植，土井，「電子楽器テルミンに着目した動きインターフェイスの基礎研究」
平成 15 年電子情報通信学会学生発表会講演論文集 D-31

[4] 「ETHERWAVE THEREMIN」，Big Briar，アメリカ

[5] 土井滋貴，「動画処理と Windows プログラミング」，CQ 出版，東京，2003