

修士論文

打撃音位置情報を用いる
インタラクションの実現に向けた
音源位置推定の研究

平成 28 年度

三重大学大学院
工学研究科 情報工学専攻
辻 敏和

2017 年 2 月

目次

第 1 章	序論	1
1.1	研究背景	1
1.1.1	インタラクション	1
1.1.2	タッチパネルのインタフェース	2
1.1.3	タッチパネル内で実現されたインタラクション	2
1.1.4	タッチパネル外で実現されたインタラクション	3
1.2	研究目標	4
1.3	本論文の構成	6
第 2 章	インタラクティブエリアの仕様	7
2.1	インタラクティブエリアの設計方法	7
2.2	収録環境	7
2.3	収録機器	10
2.4	収録方法	11
第 3 章	打撃音位置推定手法	13
3.1	様々な音源位置推定手法	13
3.2	音の到達時間差を利用した音源位置推定	13
第 4 章	CSP 法	17
4.1	DOA の算出方法	17
4.1.1	CC 法	17
4.1.2	GCC 法	17
4.2	CSP 法の問題点	18
4.3	CSP 加算法	22
4.4	本研究への適用における問題点	24
第 5 章	提案手法	27
5.1	音源位置推定における前提条件	27
5.2	位置シミュレートによる CSP 加算する DOA の抽出	28
5.3	本研究における CSP 加算法	31

5.4	提案手法の特徴	32
第 6 章	音源位置推定実験	33
6.1	収録 A における音源位置推定実験	33
6.2	打撃対象物の影響	35
6.3	収録 B における音源位置推定実験	37
第 7 章	結論	41
参考文献		43
謝辞		45

要旨

コンピュータは主にキーボードやマウスといった入力装置を用いて操作を行う機器である。キーボードは多くのボタンから構成される入力装置のため、コンピュータを効率よく操作するためにはキーボードの習得が必要不可欠である。一方、現在、スマートフォンやタブレットと呼ばれる小型で可搬型のコンピュータの利用が広がっている。これらの機器は、タッチパネルと呼ばれる表示装置と入力装置を組み合わせた装置を用いることで、直感的な機器操作を実現している。しかし、このような機器は持ち運びが容易な大きさを実現するために入力操作を行うタッチパネルが小さく設計される。タッチパネルが小さくなることでタッチパネル内に設計するボタンが人の指より小さくなり、ユーザがボタンを押せない、他のボタンを押してしまうといった誤入力が発生し、うまくコンピュータとのインタラクションができないという問題がある。

そこで本研究では、スマートフォンやタブレットの周囲にコンピュータとのインタラクションを行うための領域を設計することでこの問題の解決を図る。画面外にインタラクションできる領域を設計してボタン等を設置することで、画面内にボタンを設計する必要が無くなったり、画面内のボタンとの連携によるコンピュータ操作方法が可能になると考えられる。この新たなインタラクションを実現するため、本研究では画面外での打撃音の位置情報を利用することとし、この打撃音の位置推定手法について検討を行った。打撃音の位置情報の取得に向けて、少数のマイクロホンで実現可能な従来法である CSP (Cross-power Spectrum Phase analysis) 法に基づいた CSP 加算法を本研究が想定する音源位置に適応するために改良した CSP 加算法を提案し、打撃音位置推定の性能評価を行った。打撃音位置推定実験の結果、打撃の際に打撃対象物のたわみによる面全体の振動が発生しない環境で平均正解率の最大が 75.2 [%] となり、検討するインタラクションが実現できると確認した。また、実験結果からスマートフォンやタブレットの上下、左右のどちらか 1 つの方向に少なくとも 4 つ、最大 8 つの仮想ボタンが実現可能であると確認できた。今後の課題として、8 つの仮想ボタンにおける打撃音位置推定による平均正解率の更なる向上や提案した打撃音位置推定を用いたインタラクションシステムの構築と構築されたインタラクションシステムがユーザにとって使いやすい入力システムとなるか評価することが挙げられる。

第 1 章

序論

1.1 研究背景

1.1.1 インタラクション

インタラクション (Interaction) の意味は相互作用, 相互の影響, 対話などである. コンピュータ科学において, ヒューマンコンピュータインタラクション (Human-Computer Interaction : HCI) という言葉があり, HCI とは人とコンピュータのインタラクションのことである [1]. 本研究において, インタラクションを人とコンピュータのインタラクションのこととして定義する.

デスクトップ型のパソコンやノートパソコンの電源ボタンを押すことでコンピュータが起動し, マウスを使ってマウスカーソルを画面上のアイコンに合わせて検索エンジンなどのアプリケーションを立ち上げ, 調べたいものをキーボードを使って文字入力して検索すると, コンピュータが検索結果を出力してくれる. 電源ボタンというインタフェースを利用してコンピュータが起動するインタラクション, マウスや画面上のマウスカーソル, アイコンというインタフェースを利用してアプリケーションが立ち上がるというインタラクション, キーボードというインタフェースを利用して人の調べたいものを入力してコンピュータが検索結果を返すインタラクション, このようにあらゆる人とコンピュータの間にインタフェースが実現され双方によるインタラクションが可能となっている. 参考文献 [1] では, 画面内のアイコンをマウスカーソルでクリックできるグラフィカルユーザインタフェース (Graphical User Interface) は HCI の分野で最大の成果であると言及している.

人とコンピュータのインタラクションは, 機械的なボタンやスイッチ, GUI などの様々なインタフェースを利用して実現されている. HCI や UI (User Interface) は同義語 [1] として扱われるが, 様々なインタフェースを利用することでインタラクションが実現されるため, インタフェースが手段であり, インタラクションはその手段とその手段により得られる結果 (コンピュータとの対話) の一連のことを指す.

1.1.2 タッチパネルのインタフェース

コンピュータや情報端末への入力インタフェースとして、キーボード、マウスやスイッチといった入力装置が主流として利用されている。近年では、画面に直接触れることで操作できるタッチパネルによる入力インタフェースが導入されたスマートフォンやタブレットが普及している。タッチパネルによる入力インタフェースは、直感的な操作が可能となっており、画面上で文字入力やスクロールをすることでキーボードやマウスといった従来の入力装置の代替を可能としている。タッチパネルはスマートフォンやタブレットだけではなく、デスクトップ型のパソコンやノートパソコンのディスプレイ、冷蔵庫や電子レンジなどの家電製品など様々な情報機器に利用されており、我々の生活に広く浸透している。

タッチパネルが生活に広く浸透した要因としては、スマートフォンやタブレットのように持ち運びが容易な薄さや大きさであるにも関わらず、直感的な操作や画面に触れるだけで使用でき、さらにキーボードやマウスを使わなくてもパソコンやノートパソコンと同等のことができるという強みを有することが考えられる。

このような強みがある一方で、このような強みが弱点となるという問題が挙げられている。スマートフォンやタブレットのように持ち運びを容易にするために小さな画面で設計されているので、画面に表示するものには制限があり、ユーザの指にとってはボタンなどの操作インタフェースが小さくなることがあるという弱点がある。画面に設置されているボタンが小さい場合、ボタンを押す時に指でボタンが隠れ、ユーザはボタンを押せるかあいまいな状態で操作することになり、ボタンが押せなかったり、ボタンが密集していればユーザが押したいボタンと異なるボタンを押して誤操作となってコンピュータとうまくインタラクションできないという問題が発生する [2][4]。またタッチパネルにタッチする際に、指の太さや指の角度によってユーザがタッチしているタッチパネル内の座標とコンピュータの認識している座標にずれがあるとされている [2][3]。

1.1.3 タッチパネル内で実現されたインタラクション

タッチパネルの画面の制約による問題を解決するために、従来のデバイスで読み取れる入力動作から新たな入力システムを構築したり、新たなセンサを利用してユーザの入力動作に対応した新たな入力システムを構築するなど、人の入力をコンピュータが認識して動作していくインタラクションシステムが構築されている。提案されているインタラクションはタッチパネルの画面内で新たなシステムを構築している。

画面内で構築されたインタラクションとして、ユーザがタッチしている時にコンピュータが認識するタッチ座標を示すポインタが明示されている指で隠れている画面を指で隠れていない画面に表示して、画面をタッチしている状態で指を動かし指の先と線で繋がったタッチ座標を示したポインタをユーザが所望のタッチ位置に照準を合わせてボタンを押すシステム (*shift*)[2]、1回目のタップで画面が拡大され2回目のタップで押

したいボタンを押したり，ユーザのタッチしている状態でタッチしている指の位置から棒状のカーソルを画面に出力してカーソルの先を操作したいボタンに照準を合わせてボタンを押すシステム (*TapTap*, *MagStick*)[4] や画面の逆面にタッチパッドを取り付けたデバイスを製作して，逆面から指で触れている位置をタッチパッドで検出し，その位置を画面に表示して操作できるシステム (*nanoTouch*)[5] などがある．これらのインタラクションにより，指で隠れていない画面に操作位置や状況を画面に出力してユーザに指で隠れている部分を明示することで誤操作を防ぐことができ，さらにボタンを押せているかどうかのコンピュータ側のフィードバックも得られるためコンピュータとのより良いインタラクションを築くことができる．

shift や *MagStick* のようにタッチ座標となるカーソルを合わせてからボタンを押すことで，ボタンを押せる精度は上がるが通常のタッチ動作と比べると操作に時間がかかる[4]．またこれらのインタラクションはポインタやカーソルを合わせる必要があり，複雑な操作で使いこなすには時間を要すると推測される．*nanoTouch* は非常に斬新なシステムで指で画面を隠すことがないが，常にデバイスを持った状態で使用することになり机等に置いて使用できない．*TapTap* は2回のタップで操作を行うことができるためとても単純な操作であり，小さなボタンを押す際の精度も良く，*shift* や *MagStick* に比べると通常のタッチ動作の操作にかかる時間に差はない[4]．しかし *shift* や *TapTap* のように一度画面を拡大する場合，画面内に密集して配置されているリンクタグを押す場合に一部の文字だけ拡大され，それが押したいリンクタグか分からなくなる可能性がある．以上のことからより単純な操作で指に対して小さなボタンやリンクタグなどを画面状態を変えることなく押すことができるシステムが必要であると考えられる．

1.1.4 タッチパネル外で実現されたインタラクション

画面が小さいことで制約を受けるならば，デバイスの周囲である画面外で入力操作ができるように新たなインタフェース部分を設計してコンピュータとインタラクションできるインタラクティブエリアを広げようという試みがある．操作位置が画面外に移動することにより指で画面を隠すことはなく，さらには操作できる範囲も広がることとなる．

画面外で構築されたインタラクションは新たなセンサを利用してユーザの入力操作を認識することで実現している．画面外で構築されたインタラクションとしては，机などを掻いた時に発生する音から掻き方を検知して，コンピュータがその掻き方に応じた動作をするシステム (*Scratch Input*)[6] や机など叩いた時に発生する振動から叩いた位置を検知して，コンピュータがその位置に応じた動作をするシステム (*Toffee*)[7] などがある．この叩いた位置に応じたインタラクションは，本研究に最も近い研究である．

これらのインタラクションは画面外に構築されたインタラクティブエリアで音楽のアプリケーションにおける音量の調整やよく使用するアプリへの切り替えなど単純な機能を実現している．これらのインタラクションはユーザの動作を認識できる種類が増えたり，入力操作をしている位置の分解能が良くなったり，操作範囲が広がることでタッチパネル自体を代替することも期待できるため，有力なインタラクションであると考え

られる。タッチパネルを利用せずにタッチインタフェースを実現したものとして、プロジェクタで腕の上に操作画面やボタンを投影し、腕をタッチした時に発生する振動から操作位置を検出するウェアラブル型のデバイスが提案されている [8]。

1.2 研究目標

本研究では、タッチパネルを利用しているスマートフォンやタブレットの補助となるインタラクションを目指し、ユーザに対してより使いやすい入力システムが提供できるインタラクションシステムの構築を目標とする。

提案するインタラクションは、音源位置情報を用いたインタラクションである。提案するインタラクションのイメージ図を図 1.1 に示す。デバイスを机の上で利用する際、机を指で打撃した音から打撃位置を検知することでデバイスの周囲でボタン入力を実現することができ、また指で掻いた音から指の移動方向を検知することでスクロールやフリックを実現することができる。このインタラクションは操作範囲がデバイスの画面外となり指で画面が隠れることがない。またこのインタラクションが発展することで、キーボードやマウス、さらにはタッチパネル自体を代替することも期待できる。

本研究に最も近い先行研究 *Toffee*[7] は、インタラクションを構築するために机を叩いた時に発生する振動から叩いた位置を検知している。振動のみを利用することで人の声や生活音などの外来雑音の影響を受けないというメリットがあるが、使用環境は平面状に制限されるというデメリットがある。本研究のように音源の位置情報を利用する場合、指を鳴らしたり、人の声などを入力信号とすれば、入力方法としては汎用性があり、3次元空間上での利用も可能となる。

本研究では、提案するインタラクションの重要な要素となる打撃音の位置情報を用いたボタン入力のためのインタラクション実現に向けた打撃音位置推定について検討して、位置推定の精度を基に実現性を確認する。8つの仮想ボタンによるボタン入力を想定して打撃音の位置を推定する。想定している利用方法を図 1.2, 1.3 に示す。図 1.2 のように、押したいボタンやリンクタグが他のもので密集している場合、目的のボタン等を含むように範囲を指定した後に画面外に範囲内にあったボタン等に対応した仮想ボタンを設置して画面外で入力操作をすることで、[2][4] で取り上げられた指に対してボタンが小さい問題を解決することができる。さらに *shift* や *MagStick* のような複雑な操作は必要なく、*shift* や *TapTap* のような拡大による視覚情報の変化による懸念もない。また *Toffee* のように画面外に単純な操作を行うことで、図 1.3 のように画面外に移動した操作の代わりに画面内に新たな機能をつけることでスマートフォンやタブレットをより使いやすいものとしてユーザに提供できる。

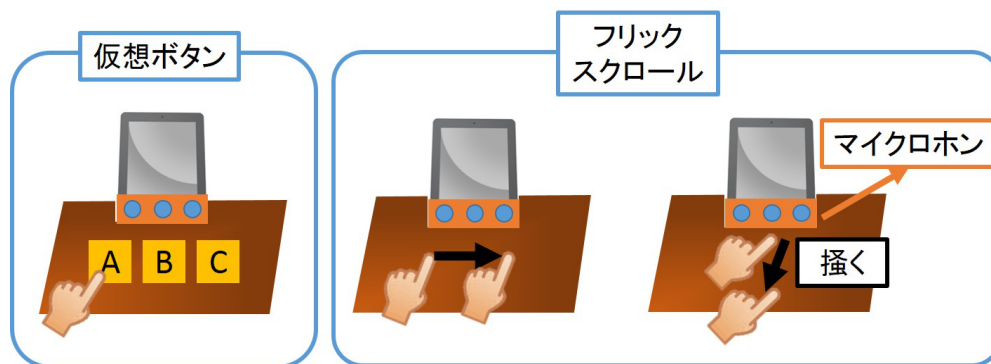


図 1.1: 提案インタラクションのイメージ図

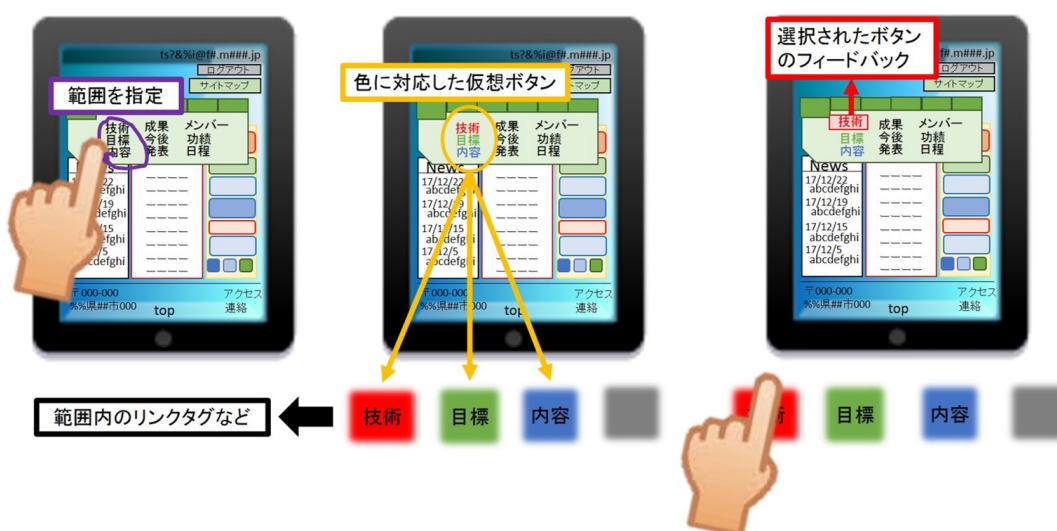


図 1.2: 画面の制約による問題の解決方法



図 1.3: インタフェース部分の拡張と利点

1.3 本論文の構成

第 2 章では，画面の周囲にインタラクティブエリアを設計する方法とその方法を評価するための収録実験について説明する．第 3 章で従来からある音源位置推定手法の紹介，第 4 章では，本研究で利用する手法の CSP 法と問題点を示す．第 5 章では，本研究に適応させた従来手法を提案手法として述べる．第 6 章では，音源位置推定実験について述べ，結果を基に提案するインタラクシヨンの実現性について述べた後，第 7 章で結論を述べる．

第 2 章

インタラクティブエリアの仕様

2.1 インタラクティブエリアの設計方法

図 1.2, 1.3 のようにスマートフォンやタブレットの周囲に仮想ボタンを設置するために画面外にインタラクティブエリアを実現する必要がある。平面状である机の上で使用することを想定して、仮想ボタンが配置された所を指で打撃した時に発生する打撃音から打撃位置を推定して、推定された仮想ボタンに対応してコンピュータが動作することでコンピュータとのインタラクションを実現する。仮想ボタンをデバイスの上下、左右の 1 つの方向に複数実現する。インタラクティブエリアに対して水平となるように複数のマイクロホンで構成される直線マイクロホンアレイを配置して、マイクロホンアレイで収録された音信号を解析することで打撃音の位置をコンピュータが認識する。この直線マイクロホンアレイを他の面方向にも配置することでデバイスの全周囲でインタラクティブエリアを実現できると考えられる。

本章では、8 つの仮想ボタンにおけるボタン入力によるインタラクション実現に向けた打撃音位置推定の精度を評価するための打撃音の収録実験を行った収録条件について説明する。2.2 節以降で収録実験の条件について説明する。

2.2 収録環境

打撃を加える対象物としてアクリル板とレンガを利用した収録環境について示す。アクリル板とレンガのサイズを表 2.1 に示す。アクリル板とレンガの 2 種類の打撃対象物による収録実験は、打撃対象物に打撃を加えた際に対象物にたわみが生じる場合と生じない場合を想定している。

打撃対象物としてアクリル板を利用した収録は情報工学科棟 5301 室で実施（収録 A）し、打撃対象物としてレンガを利用した収録は総合研究棟 I 104 室で実施（収録 B）した。収録 A の収録風景を図 2.1, 2.2 に示し、収録 B の収録風景を図 2.3, 2.4 に示す。収録 A は机の上で実施しており、防振ゴム ($10.0 \times 10.0 \times 2.0$ [cm]) を挟んでアクリル板を机の上に配置した。収録 B は図 2.3, 2.4 のように 6 個のレンガを床に並べ、マイクロホンアレイに対して収録 A と同じ位置を打撃できるようにした。

表 2.1: 打撃対象物のサイズ

	アクリル板	レンガ (1 個の大きさ)
横	45.0 [cm]	23.0 [cm]
縦	30.0 [cm]	11.5 [cm]
厚さ	0.5 [cm]	6.5 [cm]



図 2.1: 収録 A の収録風景 (1)

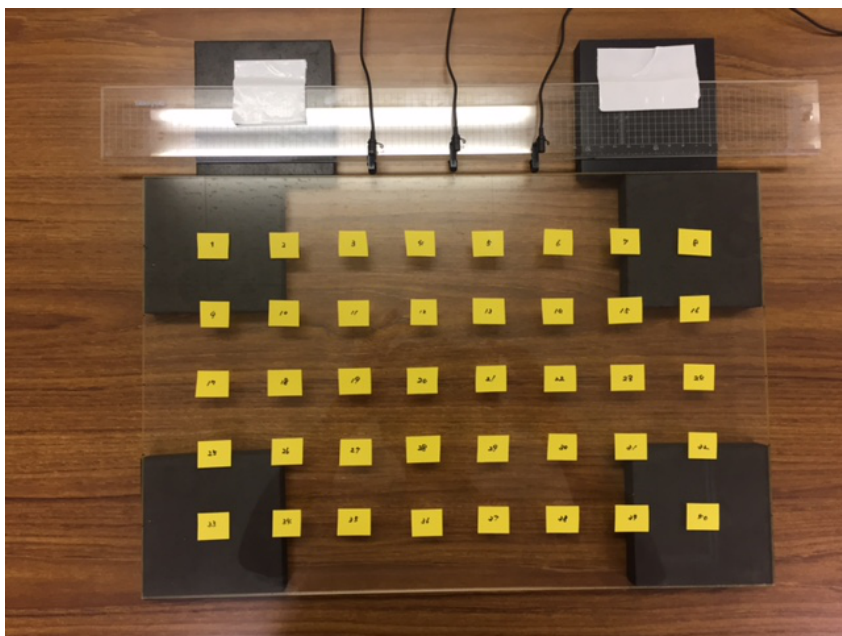


図 2.2: 収録 A の収録風景 (2)

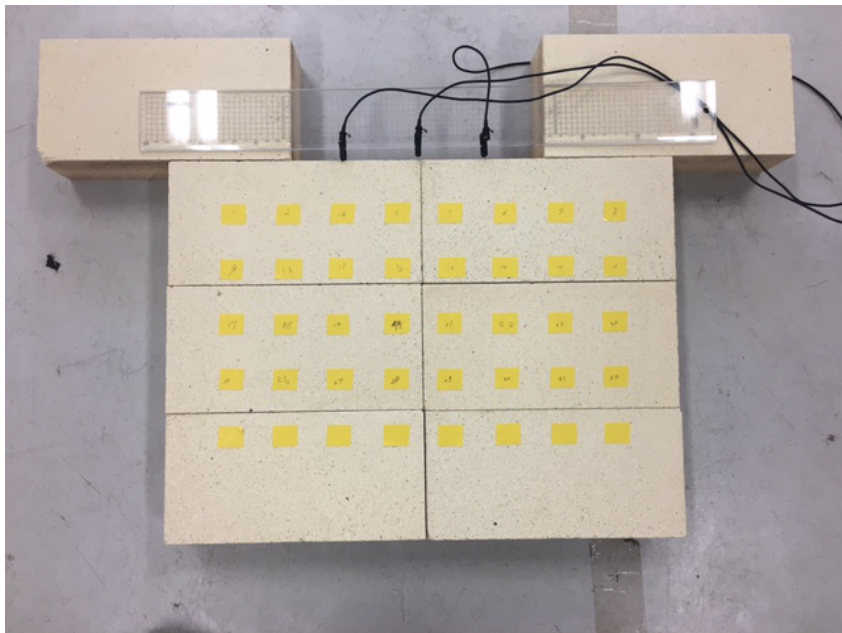


図 2.3: 収録 B の収録風景 (1)



図 2.4: 収録 B の収録風景 (2)

2.3 収録機器

実験に使用している収録機器と収録条件を表 2.2 に示す．複数のマイクロホンで同時録音をすることができるオーディオインタフェースを利用した．またこのオーディオインタフェースはファンタム電源を供給することができ，マイクロホンはこの外部電源のみを利用して動作させた．

表 2.2 に示すマイクロホン製品の付属品である縦型ホルダークリップ（図 2.5, 2.6）を使用して，50 [cm] のものさしに表 2.2 に示すマイクロホン間隔になるように 3 つのマイクロホン設置してマイクロホンアレイを構成した．よって，ペアマイクロホンのマイクロホン間隔は最大間隔が 12 [cm] となり，最小間隔が 6 [cm] となる．

表 2.2: 収録機器と収録条件

オーディオインタフェース	OCTA-CAPTURE, Roland
PC	daynabook R732/H, TOSHIBA
収録ソフトウェア	SONAR LE, Cakewalk
マイクロホン	ECM-77B, SONY
マイクロホン間隔	6.0 [cm]
サンプリング周波数	44100 [Hz]
量子化ビット数	24 [bit]



図 2.5: 縦型ホルダークリップ（正面）



図 2.6: 縦型ホルダークリップ（側面）

2.4 収録方法

収録 A, B における打撃対象物の打撃位置は、図 2.7 に示すようにマイクロホンアレイを構成している中央のマイク 2 を中心として配置した 1~32 番の黄色の四角の位置とした。打撃位置の間隔は縦に 5 [cm]、横に 5 [cm] とした。収録 A の時の防振ゴムと打撃位置の対応図を図 2.8 に示す。

収録 A, B において、図 2.9 に示す指の位置で図 2.10 のようにして打撃を加えた。図 2.7 に示す各打撃位置において、1 回の収録で 10 回の打撃音を加え、3 回収録した。

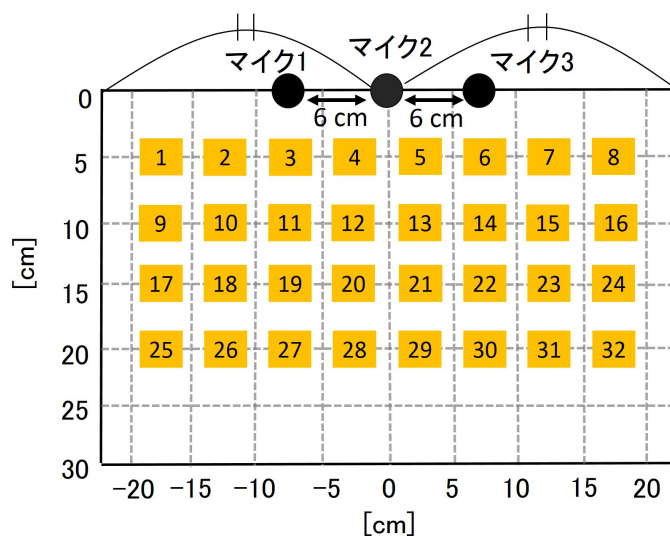


図 2.7: 打撃を加える位置

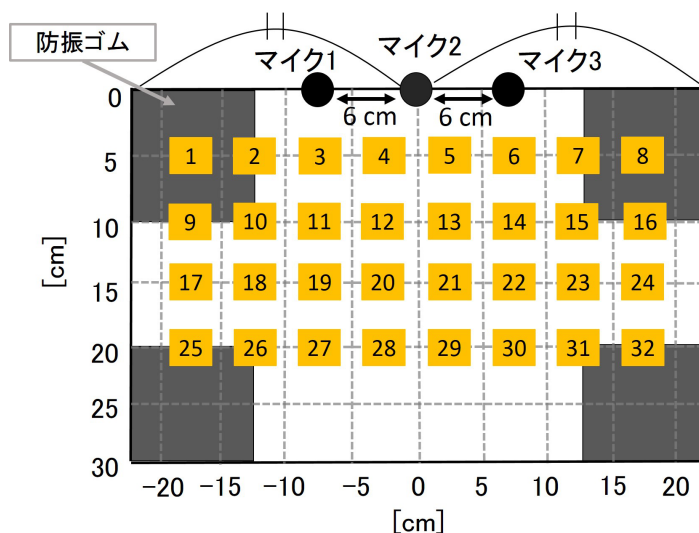


図 2.8: 実験 A における打撃位置と防振ゴムの対応

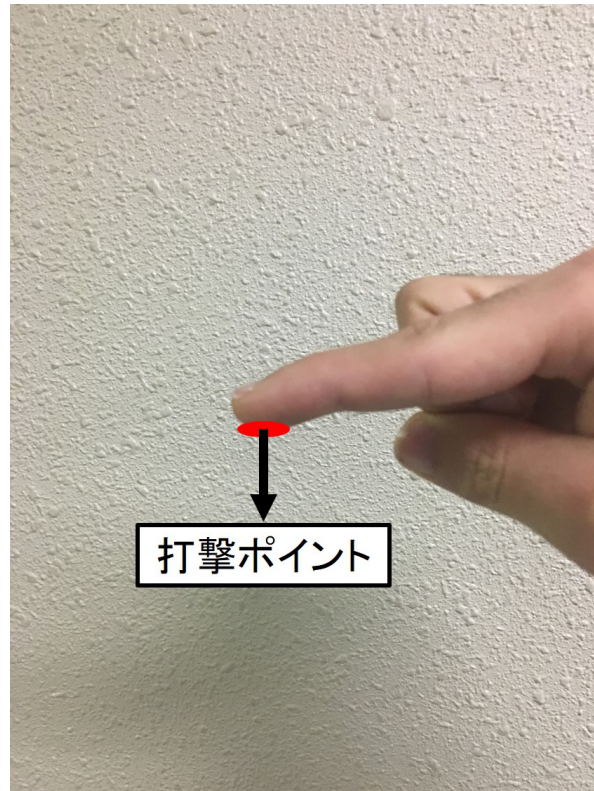


図 2.9: 打撃を加える指の位置



図 2.10: 打撃方法

第 3 章

打撃音位置推定手法

3.1 様々な音源位置推定手法

打撃対象物に打撃を加えた時に発生した音から打撃を加えた位置を推定するには、打撃音を音源と考えれば、これまで多数の研究が行われている音源位置推定手法を適用することで実現することができると思う。音源位置推定手法としては、ビームフォーミングによる音源方向推定法 [9][13]、MUSIC(MUltiple SIgnal Classification) 法 [9][11] や CSP(Cross-power Spectrum Phase analysis) 法 [10]~[16] など様々な手法が提案されている。どの手法においても音源位置を推定するには、複数のマイクロホンから構成されるマイクロホンアレイを利用している。ビームフォーミング法はマイクロホンアレイの各マイクロホンで収録された音に音源方向に対応した時間遅れを補償して足し合わせることで真の音源方向に対応する方向では音のパワーが最大となり、その方向を音源方向として推定 [9]。ビームフォーミング法は演算量が少ないという特徴がある [13] が、1 つの音源方向を頑強に推定するには複数のマイクロホンが必要とする。MUSIC 法は信号部分空間と雑音部分空間の直行性を利用して、入力信号の雑音部分空間と方向制御ベクトルを用いて音源方向を推定する [9][11]。MUSIC 法は残響によって推定精度が大きく低下するとされている [9][11]。CSP 法は 2 つのマイクロホンで収録される音の時間差を利用して音源方向を推定する。

3.2 音の到達時間差を利用した音源位置推定

本研究では、2 つのマイクロホンから音源方向が推定可能で、計算量が少なく大変有利な手法 [12] とされている CSP 法を適用する。CSP 法はマイクロホンアレイの各マイクロホンで収録された音から発生する音の到達時間差 (Delay Of Arrival : DOA) [12] を利用した音源位置推定手法である。音がマイクロホンアレイで収録されると音源から各マイクロホンまでには伝達経路差があるため、マイクロホンアレイの各ペアマイクロホンで収録された音信号間に DOA が発生する。ここで、図 3.1 にマイクロホンアレイの 2 つのマイクロホンで DOA が発生する状態を示す。

各ペアマイクロホンから算出される DOA を利用して、音源方向の交点として、また

は音源の軌跡の交点から音源位置を推定することができる．図 3.2 のようにして音源方向の交点として音源位置が推定される．図 3.2 において， c は音速， M_i はマイクロホンのラベルを示す．また $d_{M_i M_j}$ は M_i と M_j のマイクロホン間隔， $t_{M_i M_j}$ は M_i と M_j における DOA， $\theta_{M_i M_j}$ は M_i と M_j から推定される音源方向を示す．音源方向 $\theta_{M_i M_j}$ については式 (3.2.1) から算出される．

$$\theta_{M_i M_j} = \cos^{-1} \left(\frac{c \cdot t_{M_i M_j}}{d_{M_i M_j}} \right) \quad (3.2.1)$$

各ペアマイクロホンにおける DOA から算出される音源方向の交点として音源位置が推定される．

図 3.3 に示すようにして音源の軌跡から音源位置を推定することができる．各ペアマイクロホンの DOA を利用して音源を動点，2 つのマイクロホンを焦点とした双曲線を算出して，双曲線において 2 つのペアマイクロホンで先に音が到達した側の曲線を音源の軌跡とする．各ペアマイクロホンで求めた音源の軌跡の交点として音源位置が推定される．動点となる音源座標を (x, y) とすると，双曲線は双曲線の方程式を変形した式 (3.2.2) を解くことで得ることができる．

$$x = \pm \sqrt{\left(\frac{c \cdot t_{M_1 M_2}}{2} \right)^2 \left(1 + \frac{y^2}{\left(\frac{d_{M_1 M_2}}{2} \right)^2 - \left(\frac{c \cdot t_{M_1 M_2}}{2} \right)^2} \right)} \quad (3.2.2)$$

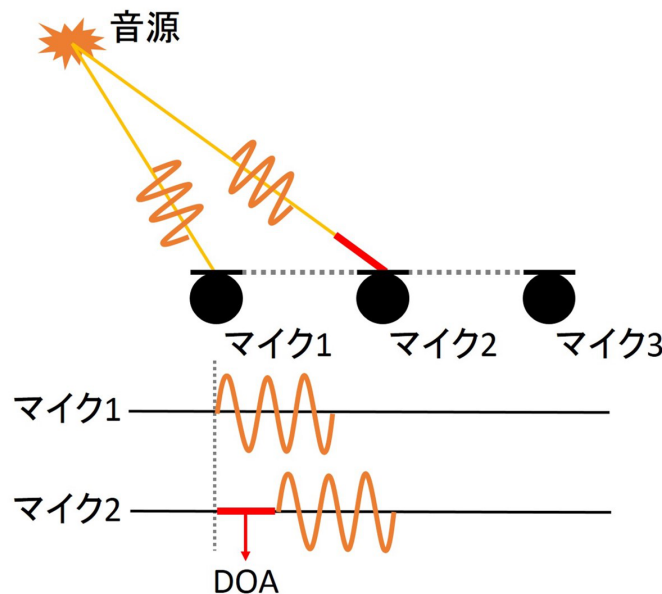


図 3.1: DOA が発生する状態図

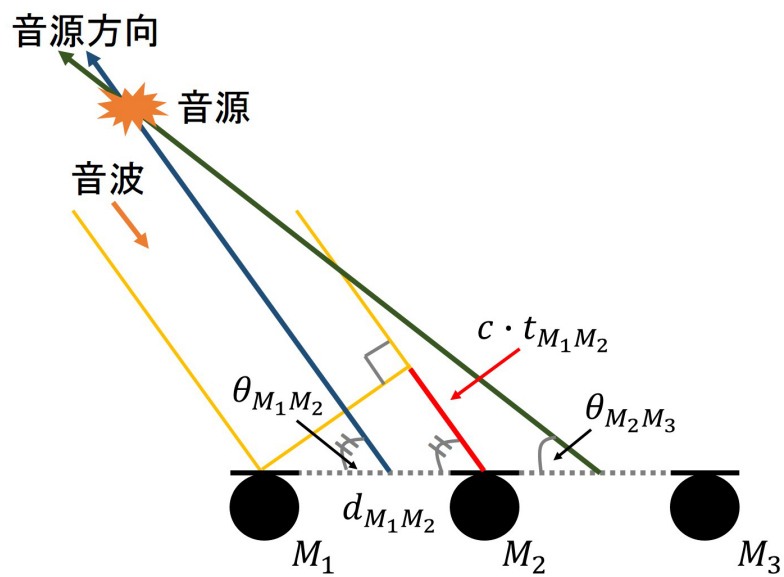


図 3.2: 音源方向を利用した音源位置推定

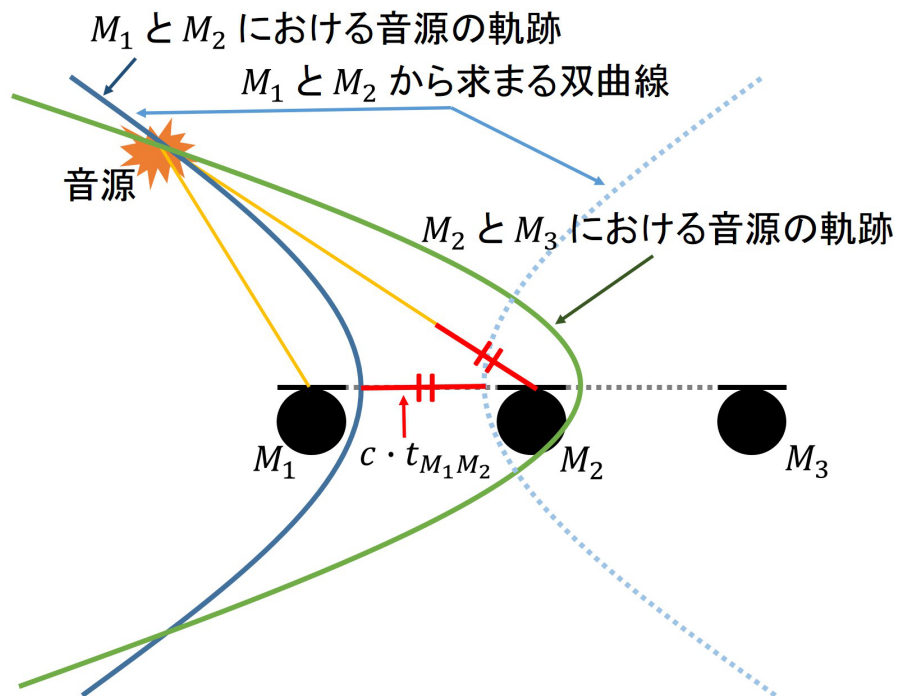


図 3.3: 音源の軌跡を利用した音源位置推定

第 4 章

CSP 法

4.1 DOA の算出方法

4.1.1 CC 法

DOA を利用した音源位置推定手法では，ペアマイクロホンの各マイクロホンで収録された音信号の相関に基づいて DOA を算出している．最も基本となる相関に基づいた DOA の算出方法は相互相関 (Cross-Correlation : CC) 法 [10][15][16] と呼ばれている．2 つのマイクロホンで収録された音信号 $x_1(i)$, $x_2(i)$ ($i = 1, 2, \dots, N$) とすると，これらの信号の相互相関関数 $R_{x_1x_2}$ は式 (4.1.1) から求まる．

$$R_{x_1x_2}(k) = \frac{1}{N} \sum_{i=1}^{N-k} x_1(i)x_2(i+k) \quad (4.1.1)$$

$R_{x_1x_2}(k)$ は $x_1(i)$ と $x_2(i)$ の類似度を示す指標であり， $R_{x_1x_2}(k)$ が最大となるサンプル時間 τ が $x_1(i)$ と $x_2(i)$ の時間ずれを表し，ペアマイクロホンの DOA に対応する． τ は式 (4.1.2) より求まる．

$$\tau = \arg \max_k [R_{x_1x_2}(k)] \quad (4.1.2)$$

CC 法から DOA を算出することは可能であるが，雑音，残響の影響を受ける環境下での音に対しては相関がとれず，相互相関関数のピーク値に広がりが生じて，DOA の算出精度が良くない [10][15]．

4.1.2 GCC 法

相互相関関数のピーク値の広がりを抑えて DOA の算出精度をあげる手法として，GCC-PHAT (Generalized Cross-Correlation with PHase Transform) 法 [10][15][16] と呼ばれる手法が利用されている．GCC-PHAT 法は一般化相互相関 (GCC) 法 [10][15][16] の一種であり，ペアマイクロホンの各マイクロホンで収録された音信号のクロススペクトルを求め，重み関数 $\psi(\omega)$ を掛けた後，逆フーリエ変換により一般化相互相関関数を求めて，式 (4.1.2) と同じようにして DOA を算出する．一般化相互相

関関数は式 (4.1.3) から求まる.

$$R_{x_1x_2}^{\text{GCC}}(k) = \sum_{\omega=-\infty}^{+\infty} \psi(\omega) S_{x_1x_2}(\omega) e^{j\omega k} \quad (4.1.3)$$

$S_{x_1x_2}(\omega)$ は, 式 (4.1.4) から求まるクロススペクトルである.

$$S_{x_1x_2}(\omega) = X_1(\omega) X_2^*(\omega) \quad (4.1.4)$$

ここで, $X_1(\omega)$ と $X_2(\omega)$ はそれぞれ $x_1(i)$ と $x_2(i)$ をフーリエ変換したものであり, $*$ は共役複素数を示す. GCC-PHAT 法は式 (4.1.5) の重み関数 $\psi_{\text{phat}}(\omega)$ を適用した手法である.

$$\psi_{\text{phat}}(\omega) = \frac{1}{|S_{x_1x_2}(\omega)|} \quad (4.1.5)$$

GCC-PHAT 法は重み関数 $\psi_{\text{phat}}(\omega)$ によって位相情報のみから DOA を算出することができるため, 音信号のパワーに依存しない.

重み関数として別の関数を適用している手法として SCOT(Smoothed COherence Transform) 法 [10][15][16] や ML(Maximum Likelihood transform) 法 [10][15][16] などがある. また重み関数が $\psi(\omega) = 1$ を満たす場合, ウィナー-ヒンチンの定理 [17] より CC 法と同等となる. すべての重み関数における手法の中で, GCC-PHAT 法が DOA の算出精度が高く [15], 雑音や残響下においても DOA の算出精度が高い手法 [16] とされている.

DOA を利用した音源位置を推定する手法として, DOA の算出に多くがこの DOA の算出精度が高い GCC-PHAT 法が適用されており, この手法が CSP 法として利用されている.

4.2 CSP 法の問題点

CSP 法は DOA の算出精度が高い GCC-PHAT 法と同等の手法であり, 位相情報のみから 2 つの信号の類似度を算出して, 類似度が最大となるサンプル時間を DOA とする手法である. CSP 法においてはこの類似度が CSP 係数と呼ばれ, CSP 係数は式 (4.2.1) から求まる.

$$\text{CSP}_{ij}(k) = \text{DFT}^{-1} \left[\frac{\text{DFT}[x_i(n)] \text{DFT}[x_j(n)]^*}{|\text{DFT}[x_i(n)]| |\text{DFT}[x_j(n)]|} \right] \quad (4.2.1)$$

ここで, M 個のマイクロホンでマイクロホンアレイが構成されている場合, $x_i(n)$ と $x_j(n)$ はそれぞれ i ($i = 1, \dots, M-1$) 番目, j ($j = i+1, \dots, M$) 番目のマイクロホンで収録された音信号を示し, $\text{CSP}_{ij}(k)$ は i 番目と j 番目のペアマイクロホンから算出される CSP 係数である. また DFT と DFT^{-1} はそれぞれフーリエ変換と逆フーリエ変換を示し, $*$ は共役複素数を示す.

各ペアマイクロホンの DOA は式 (4.2.1) から求まる CSP 係数が最大となるサンプル時間 τ_{ij} に対応し、DOA となる t_{ij} は式 (4.2.2) と式 (4.2.3) から求まる.

$$\tau_{ij} = \arg \max_k [\text{CSP}_{ij}(k)] \quad (4.2.2)$$

$$t_{ij} = \frac{\tau_{ij}}{F_s} \quad (4.2.3)$$

ここで, F_s はサンプリング周波数を示す. 各ペアマイクロホンから算出された DOA から式 (3.2.2) を利用して音源の軌跡を求め, 音源位置を推定する.

収録 A において, 図 2.7 と図 2.8 に示す打撃位置 2 番を打撃した音から式 (4.2.1) から算出した CSP 係数を図 4.1, 4.3, 4.5 に示し, 各 CSP 係数に対して各マイクロホン間隔における最大サンプル時間差の範囲に限定した CSP 係数を図 4.2, 4.4, 4.6 に示す. また算出されたサンプル時間差から位置を推定した時の結果を図 4.7 に示す. ここで, 黒い四角は図 2.7, 2.8 に示す番号付きの黄色の四角の打撃位置と対応する. $\text{CSP}_{13}(k)$ に関しては, 正しいサンプル時間差に近いサンプル時間差を算出できているが, $\text{CSP}_{12}(k)$ と $\text{CSP}_{23}(k)$ に関しては, 正しく算出できていない. 結果として正しい位置を推定することができない. しかし $\text{CSP}_{12}(k)$ と $\text{CSP}_{23}(k)$ に関して, 正しいサンプル時間差において CSP 係数値がピークとなっていないが, 大きくなっている. CSP 係数の算出精度が良くないペアマイクロホンにおいても音源位置に対応する DOA で CSP 係数値が大きくなっている. 各ペアマイクロホンにおける CSP 係数の算出精度に依存しないように, すべてのペアマイクロホンにおける CSP 係数から音源位置を決定する手法が必要である.

この章以降, 式 (4.2.2) から算出されるサンプル時間差 τ_{ij} を DOA として定義する.

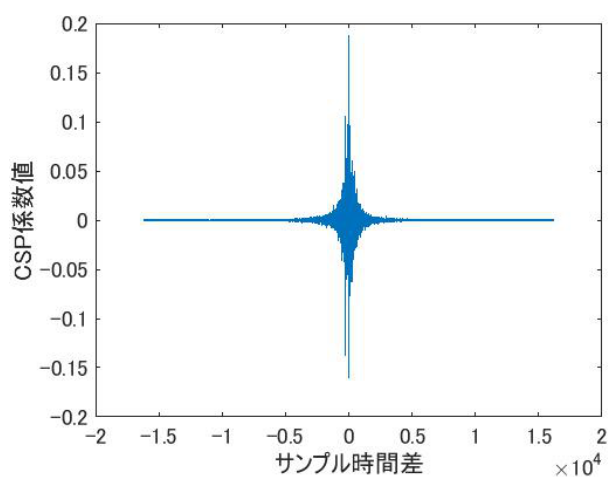


図 4.1: マイク 1 とマイク 2 の CSP 係数

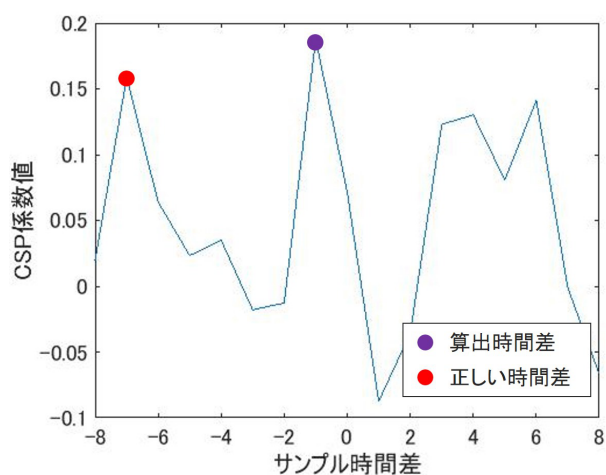


図 4.2: 図 4.1 の拡大図

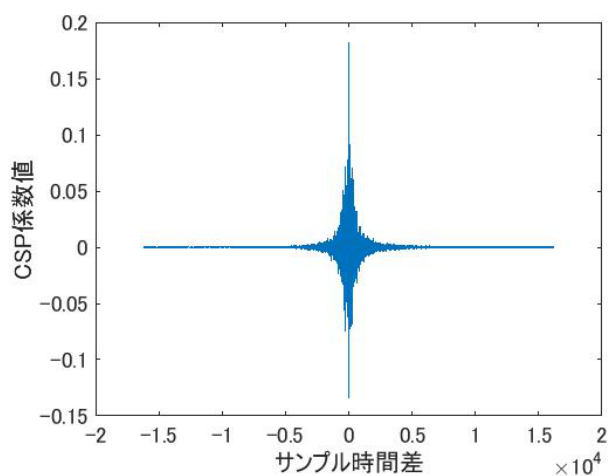


図 4.3: マイク 2 とマイク 3 の CSP 係数

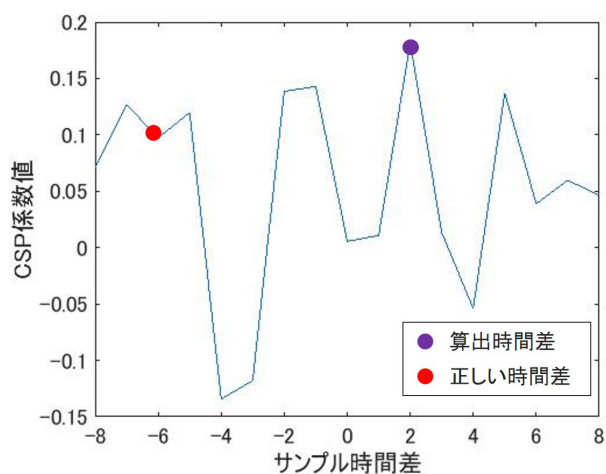


図 4.4: 図 4.3 の拡大図

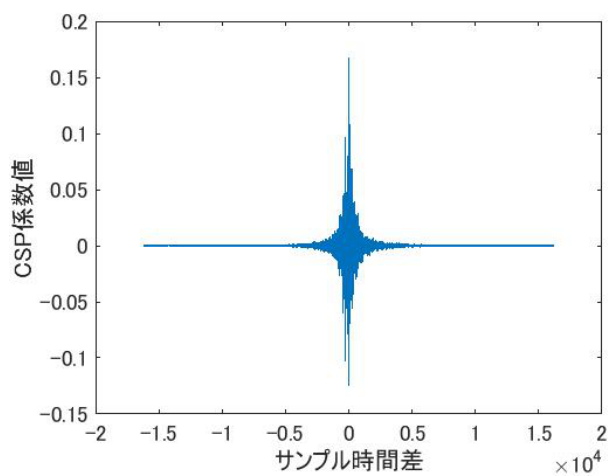


図 4.5: マイク 1 とマイク 3 の CSP 係数

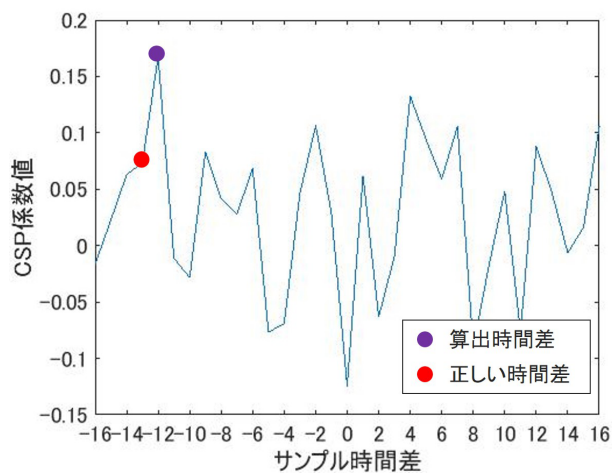


図 4.6: 図 4.5 の拡大図

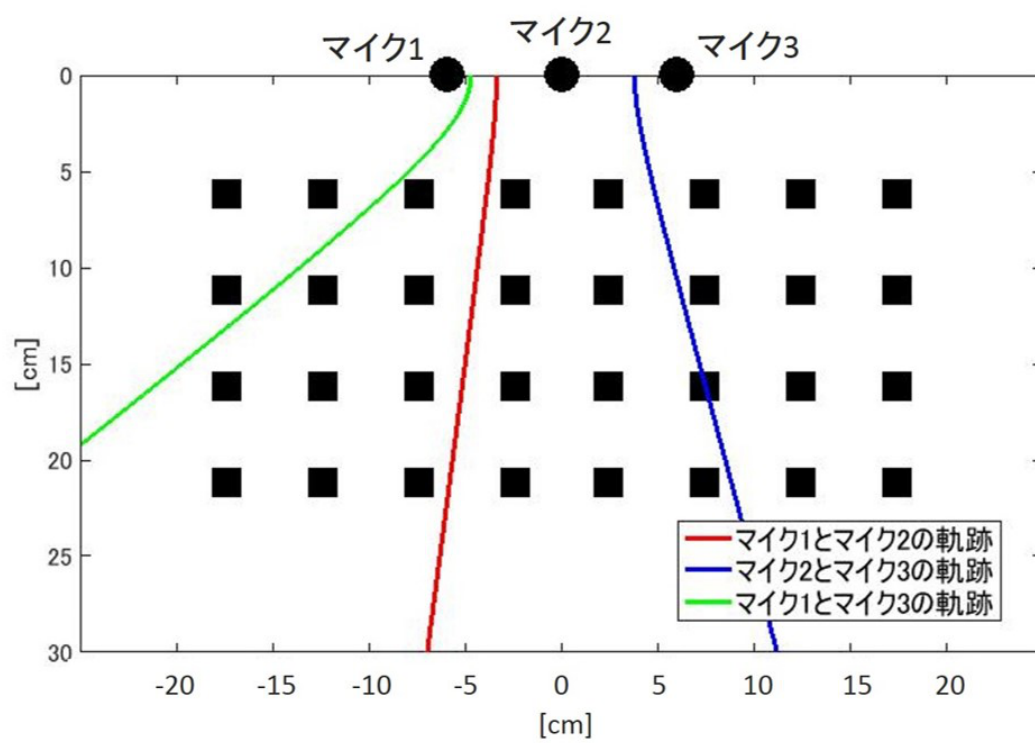


図 4.7: 位置推定結果

4.3 CSP 加算法

各ペアマイクロホンの CSP 係数の算出精度に依存せずに音源位置を推定する手法として CSP 加算法 [12][13][14] が提案されている。CSP 加算法とは、各 CSP 係数の真の音源方向や音源位置に対応する DOA の CSP 係数値を足し合わせ、音源方向や音源位置毎の相関値となるものを算出して、すべてのペアマイクロホンから音源位置や音源方向を直接推定できる手法である。先行研究 [12] では、各ペアマイクロホンの中心を合わせてマイクロホンアレイを構成して、中心が合ったペアマイクロホンの音源方向に対応する DOA における CSP 係数値を加算して音源位置を推定している。この手法では、マイクロホンアレイの配置にある程度の制限があり、多数のマイクロホンを必要とする。先行研究 [13][14] では、各ペアマイクロホンの音源位置に対応する DOA を理論的に算出して、算出された DOA における CSP 係数値を加算して音源位置を推定している。この手法では、マイクロホンの配置は自由であり、3 つのマイクロホンからマイクロホンアレイを構成可能である。

先行研究 [13][14] における各ペアマイクロホンの音源位置毎に対応する DOA は理論的に式 (4.3.1) から算出される。

$$\tau_{ijp} = \frac{\|m_i - s_p\| - \|m_j - s_p\|}{F_s \times c} \quad (4.3.1)$$

ここで、 i, j はマイクロホン番号で m_i, m_j はそれぞれマイクロホンの座標 $m_i(x_i, y_i, z_i)$, $m_j(x_j, y_j, z_j)$ を示し、 s_p は音源候補位置の座標 $s_p(x_p, y_p, z_p)$ を示す。また c は音速を示す。使用するマイクロホンの数を M として、音源候補位置 s_p における CSP 加算値 $\text{CSP}_{\text{summed}}(p)$ は式 (4.3.2) から算出される。

$$\text{CSP}_{\text{summed}}(p) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \text{CSP}_{ij}(\tau_{ijp}) \quad (4.3.2)$$

ここで、 τ_{ijp} は音源候補位置 p における i と j 番目のマイクロホンで発生する DOA を示す。式 (4.3.2) から各ペアマイクロホンの CSP 係数の算出精度に依存しない音源候補位置毎の相関配列を作成することができる。式 (4.3.3) によって、式 (4.3.2) から求まる $\text{CSP}_{\text{summed}}(p)$ が最大となる p が音源位置 \hat{p} として推定される。

$$\hat{p} = \arg \max_p [\text{CSP}_{\text{summed}}(p)] \quad (4.3.3)$$

先行研究 [13][14] において、音源位置のわずかな誤差の影響を低減するために、 τ_{ijp} における CSP 係数値を加算するのではなく、 τ_{ijp} から前後に r [sample] の範囲で式 (4.3.4) から求まる CSP 係数値が最大となる $\text{CSP}_{ij}(\tau'_{ijp})$ を加算する。

$$\text{CSP}_{ij}(\tau'_{ijp}) = \max [\text{CSP}_{ij}(n)] \quad (4.3.4)$$

ここで, n は $|n - \tau_{ijp}| \leq r, r \in N$ を満たす. よって最終的に CSP 加算値 $\text{CSP}_{\text{summed}}(p)$ は式 (4.3.5) から算出される.

$$\text{CSP}_{\text{summed}}(p) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \text{CSP}_{ij}(\tau'_{ijp}) \quad (4.3.5)$$

以上のアルゴリズムのイメージ図を図 4.8 に示す.

この節以降, 式 (4.3.2) や式 (4.3.5) の τ_{ijp} や τ'_{ijp} における CSP 係数値を加算することを CSP 加算と呼び, τ_{ijp} や τ'_{ijp} を CSP 加算する DOA と呼ぶ.

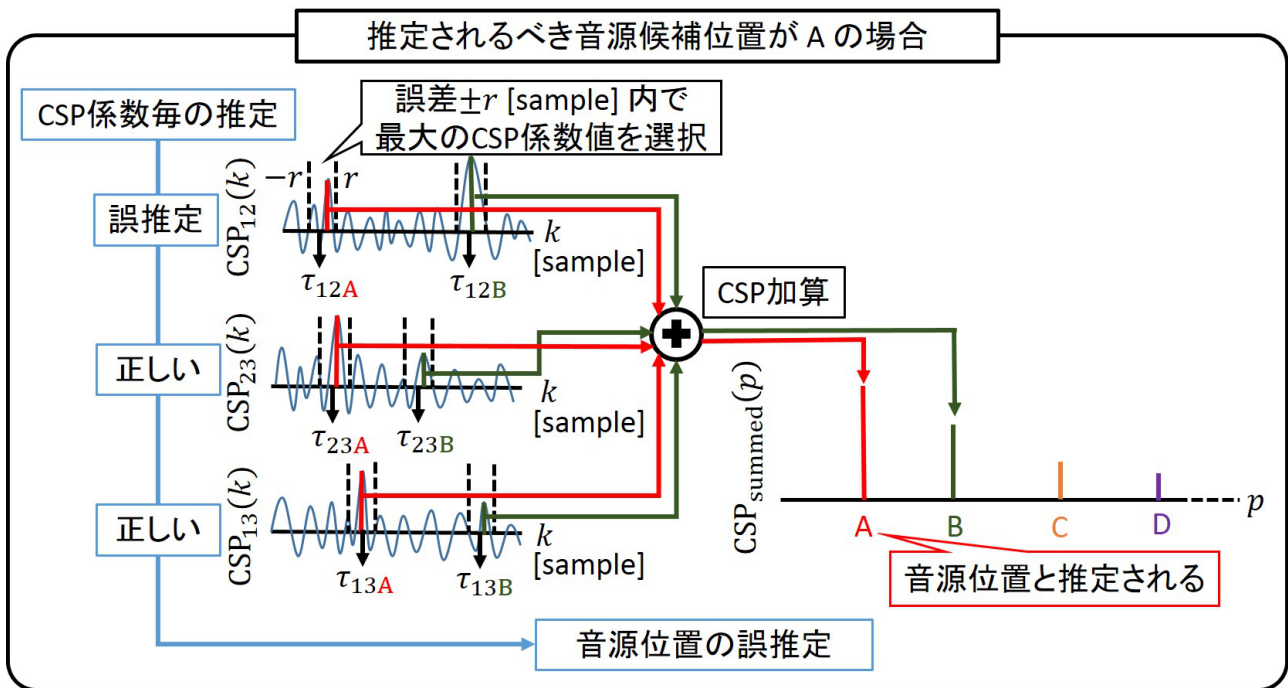


図 4.8: CSP 加算法のイメージ図

4.4 本研究への適用における問題点

先行研究 [13][14] の CSP 加算法では、1 つの音源候補位置に対して各ペアマイクロホンで理論的に算出される 1 つの DOA における CSP 係数値を加算している。また先行研究 [13][14] では音源位置の誤差を考慮するために、理論的に算出された DOA の前後数サンプルの範囲内で CSP 係数値が最大となるサンプル時間差を CSP 加算する DOA としている。先行研究 [13][14] においてマイクロホン間隔が 135 [cm] で 16 個のマイクロホンから構成される格子状のマイクロホンアレイを配置して、サンプリング周波数が 16000 [Hz] で音を収録している。またマイクロホンアレイから音源候補位置までの距離が 123 [cm] となっている。

ここで図 4.9, 4.10 にマイクロホン間隔が 6 [cm] の場合の DOA 毎の音源の軌跡を示し、図 4.11, 4.12 にマイクロホン間隔が 12 [cm] の場合の DOA 毎の音源の軌跡を示す。本研究では、マイクロホン間隔が最小で 6 [cm]、最大で 12 [cm] としているため、各ペアマイクロホンにおける DOA は $-7 \sim 7$ [sample], $-15 \sim 15$ [sample] で表現される。マイクロホン間隔が 6 [cm] の場合、図 4.9, 4.10 から DOA が 1, 2 [sample] 変化すると異なる音源候補位置の DOA となる。マイクロホン間隔が 12 [cm] の場合、図 4.11, 4.12 からマイクロホン間隔が 6 [cm] の場合と比べて 1, 2 [sample] の変化による軌跡の間隔は狭まったが 1, 2 [sample] の変化の影響は大きい。また本研究ではマイクロホンアレイから限りなく近い音源位置における音源位置推定を想定しており、図 4.9~4.12 よりマイクロホンアレイから近い所に位置する音源候補位置よりも遠い所に位置する音源候補位置で 1, 2 [sample] の変化の影響が大きくなることが分かる。

本研究において誤差を考慮して音源位置を推定する際、すべての音源候補位置での誤差を数サンプルの範囲と単純に決定するとある真の音源候補位置 p の相関値となる $\text{CSP}_{\text{summed}}(p)$ に他の音源候補位置 p' の相関値 $\text{CSP}_{ij}(\tau_{ijp'})$ を加算する可能性がある。また先行研究 [13][14] では、各ペアマイクロホンで 1 つの DOA における CSP 係数値を加算しており、音源方向が近似する 2 つの音源候補位置 p_a, p_b がある場合、1 番目と 2 番目のマイクロホンで算出される DOA において $\tau_{12p_a} = \tau_{12p_b}$ が成り立ち、 $\text{CSP}_{12}(\tau_{12p_a}) = \text{CSP}_{12}(\tau_{12p_b})$ となり、それぞれの音源候補位置の相関値 $\text{CSP}_{\text{summed}}(p_a)$ と $\text{CSP}_{\text{summed}}(p_b)$ に差がない可能性がある。音源候補位置の相関値に差がなく真の音源がその 2 つの音源候補位置のどちらかであった場合、推定結果が 2 つのどちらかに定まらなかったり、他の音源候補位置が音源位置と推定されることとなる。本研究において、音源方向が近似する 2 つの音源候補位置はマイクロホンアレイの中央からのある方向におけるマイクロホンアレイから近い所と遠い所に位置する音源候補位置である。

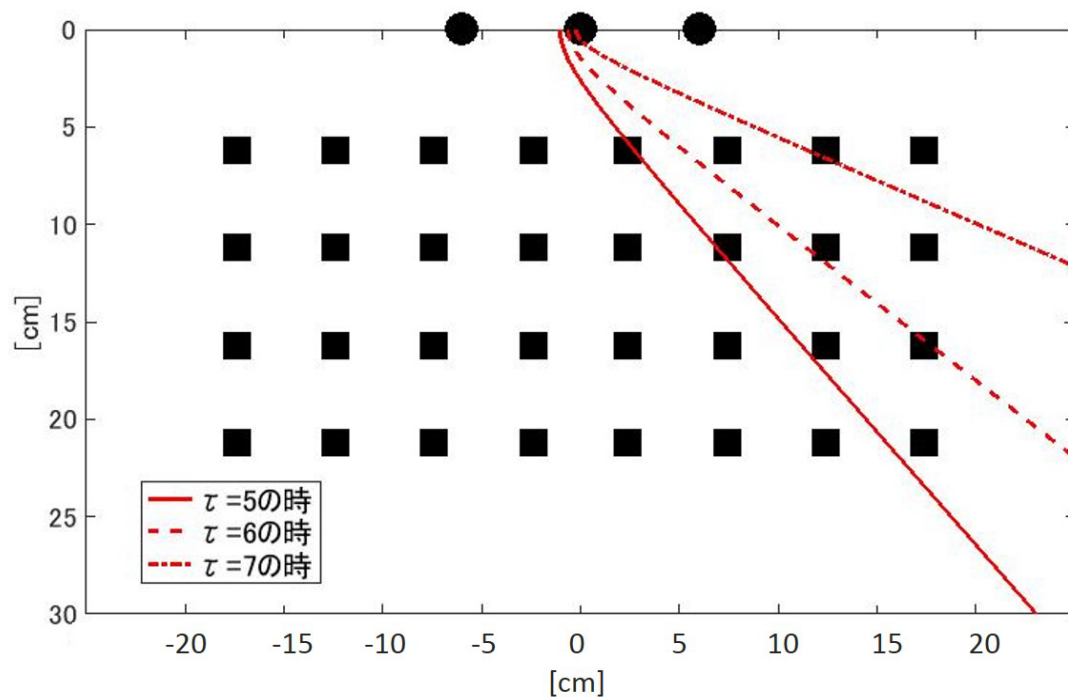


図 4.9: マイクロホン間隔 6 [cm] の DOA 毎の音源の軌跡 (1)

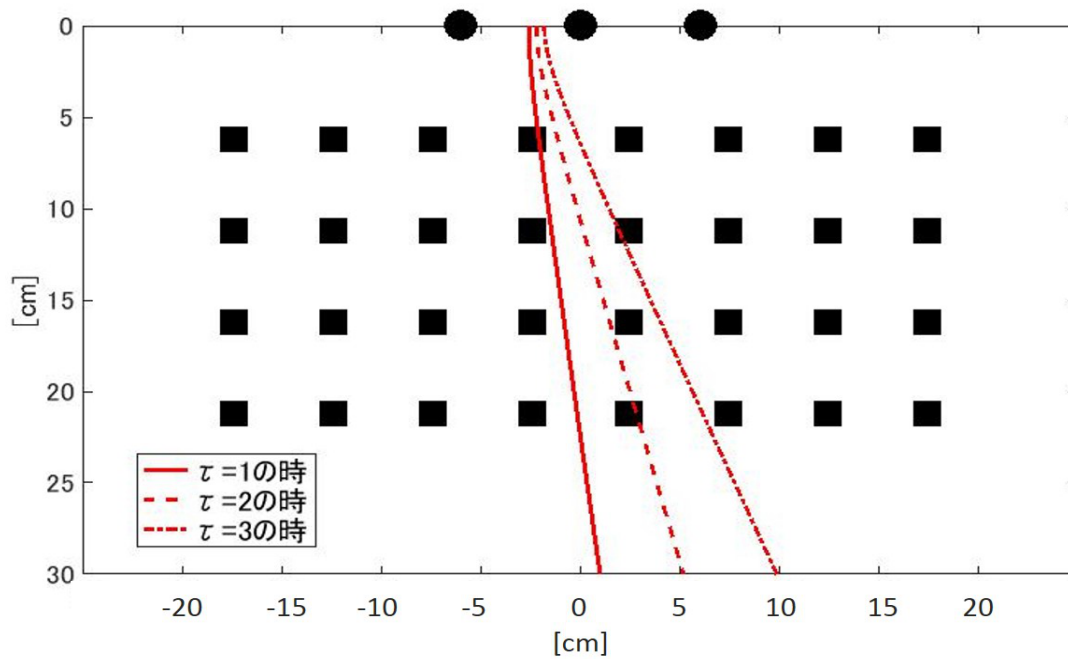


図 4.10: マイクロホン間隔 6 [cm] の DOA 毎の音源の軌跡 (2)

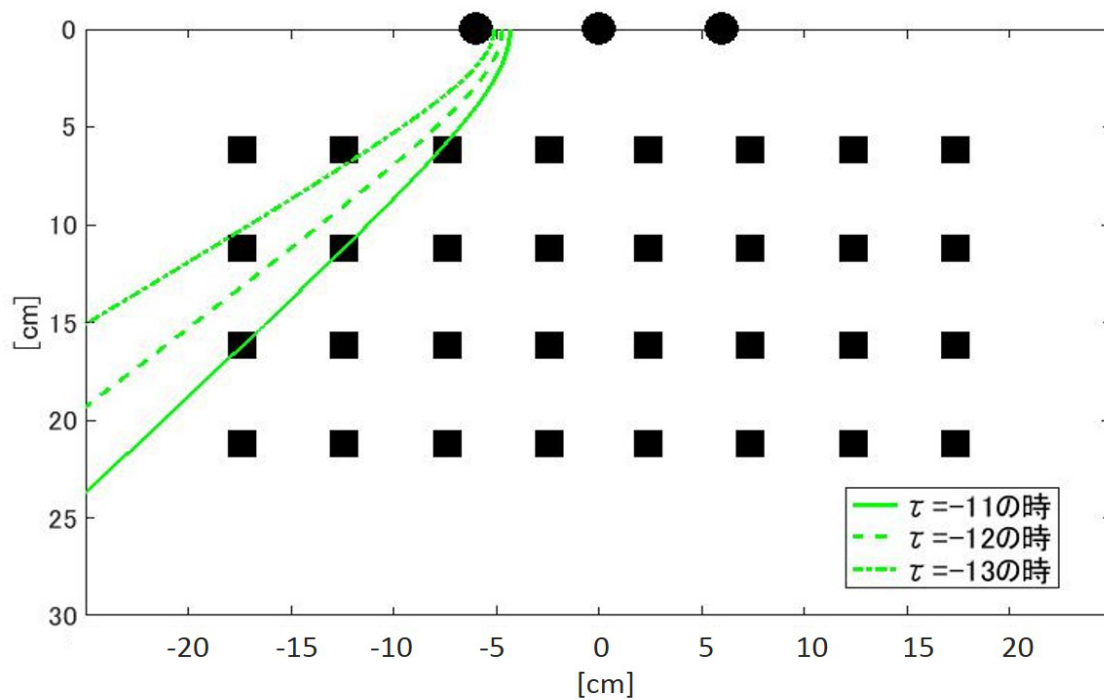


図 4.11: マイクロホン間隔 12 [cm] の DOA 毎の音源の軌跡 (1)

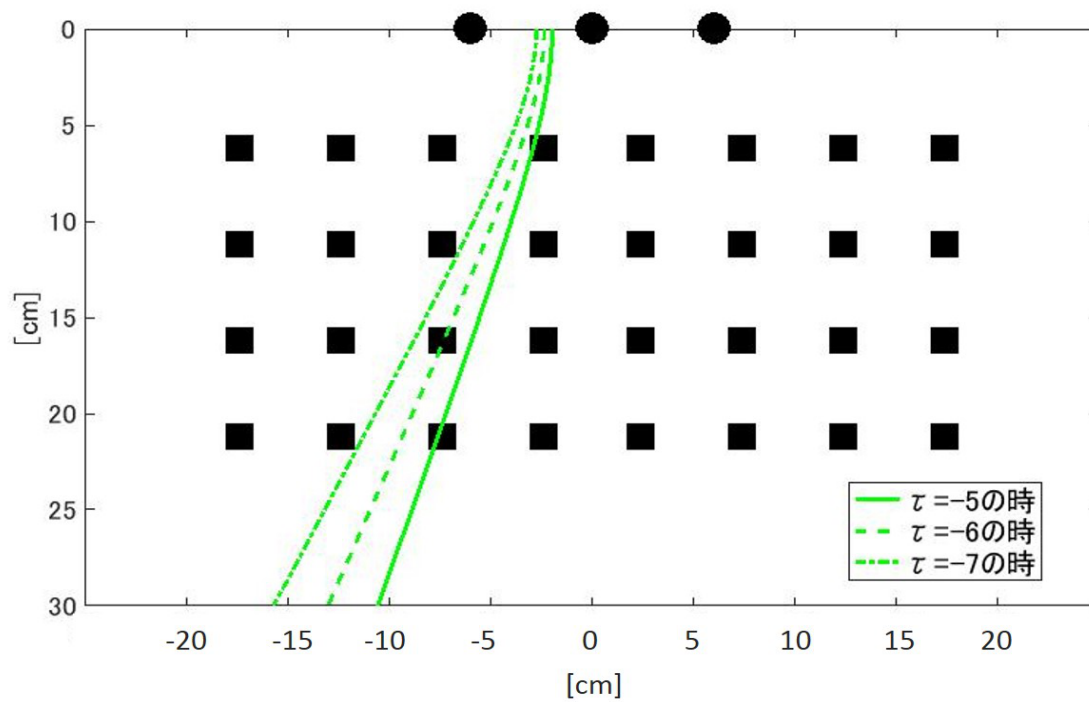


図 4.12: マイクロホン間隔 12 [cm] の DOA 毎の音源の軌跡 (2)

第 5 章

提案手法

5.1 音源位置推定における前提条件

8つのボタン入力を実現することを想定して、音源候補位置として8つの位置を設定する。8つの音源候補位置 A~H の位置関係を図 5.1 に示す。1つの音源候補位置に対して $10\text{ [cm]} \times 10\text{ [cm]}$ の許容範囲を許し、音源候補位置の間隔が横 5 [cm] 、縦 5 [cm] の8つの音源候補位置で位置推定を行う。収録した打撃音データの位置は図 5.1 に示す各音源候補位置の4隅である。各音源候補位置で収録された打撃音データが同じ音源候補位置に推定された場合、正しい位置推定が行われたとする。

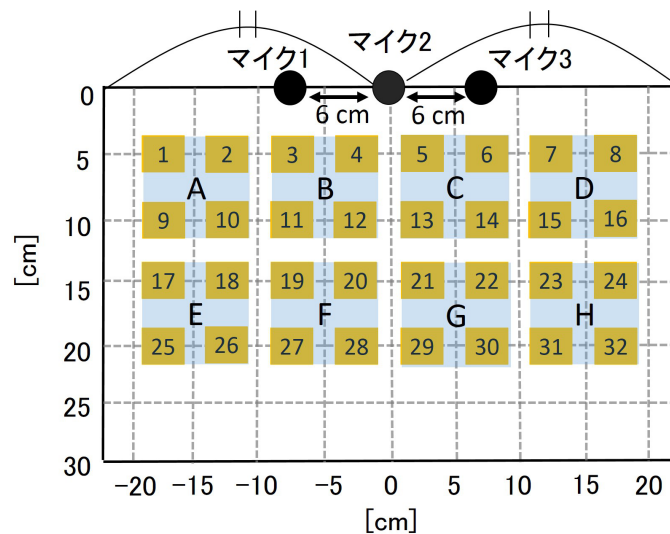


図 5.1: 8つの音源候補位置

5.2 位置シミュレートによる CSP 加算する DOA の抽出

マイクロホンの数を $3 (= M)$ 個として、 i ($i = 1, \dots, M - 1$) 番目と j ($j = i + 1, \dots, M$) 番目のペアマイクロホンから算出できる DOA を τ_{ij} ($-\tau_{max_{ij}} \leq \tau_{ij} \leq \tau_{max_{ij}}$) とした時、マイクロホンアレイにおいて 1 つの DOA の組み合わせ $(\tau_{12}, \tau_{23}, \tau_{13})$ を利用して式 (3.2.2) により求まる音源の軌跡から音源位置をシミュレートする。1 つの DOA の組み合わせ $(\tau_{12}, \tau_{23}, \tau_{13})$ において一意の音源候補位置を決定できる時、その DOA の組み合わせを各ペアマイクロホンにおける CSP 加算する DOA として抽出する。ここで、 $\tau_{max_{ij}}$ は i 番目と j 番目のマイクロホンアレイにおける最大の DOA である。CSP 加算する DOA の抽出アルゴリズムのフローチャートを図 5.2 に示す。

図 5.2 において一意の音源候補位置に決定されるかどうか判定する際に必要となる音源候補位置の選択アルゴリズムを図 5.3 に示すフローチャートとともに説明する。

まず、1 つの DOA の組み合わせ $(\tau_{12}, \tau_{23}, \tau_{13})$ からそれぞれの DOA を利用して音源の軌跡 T_{12}, T_{23}, T_{13} を算出した後、各音源候補位置 p の位置判定点 p' との距離差 $d_{p'}$ を算出する。ここで $p' (= 1, \dots, 32)$ は図 5.1 における音源候補位置 $p (\in A, \dots, H)$ のそれぞれの 4 隅であり、打撃音の収録位置となる図 5.1 における 1~32 番の黄色マーカの位置に等しくなる。ここで、図 5.3 の $X_{p'} (x_{p'}, y_{p'})$ は p' の座標を示す。

次に、距離差 $d_{p'}$ の許容誤差を d_{error} とした時、式 (5.2.1) を満たすならある 1 つの軌跡が $X_{p'}$ を通るとする。

$$d_{p'} \leq d_{error} \quad (5.2.1)$$

1 つの $X_{p'}$ に対して 3 つの軌跡が通る (図 5.3 における $N(p') = 3$) のであれば、その p' が属する p を推定されうる位置とする (図 5.3 における $L(p) = L(p) + 1$ で、 $L(p) = 1$ となった状態)。

最後に、すべての p において 3 つの軌跡が通る p' が 1 つ以上あり、かつすべての p において 3 つの軌跡が通る p' の個数 (図 5.3 における $L(p)$) が最大でその個数が最大となる p がたった 1 個の場合、その p が一意に決定される音源候補位置とする。

ここで図 5.4 に一意に決定される音源候補位置の選択アルゴリズムにおいて一意の音源候補位置が決定されるイメージ図を示す。図 5.4 において一意に決定される音源候補位置は桃色に色付けされた位置判定点 $p' = 26$ が属する音源候補位置 $p = E$ である。桃色に色付けされている状態は図 5.3 において $N(p') = 3$ の状態であり、3 つのペアマイクロホンにおける音源の軌跡が通っている状態である。また橙色に色付けされている状態が $N(p') = 2$ の状態であり、黄色の場合は $N(p') = 1$ である。すべての音源候補位置において桃色に色付けされているのは音源候補位置 E だけ ($L(E) = 1$, その他は 0) であるので、E が一意の音源候補位置と決定される。

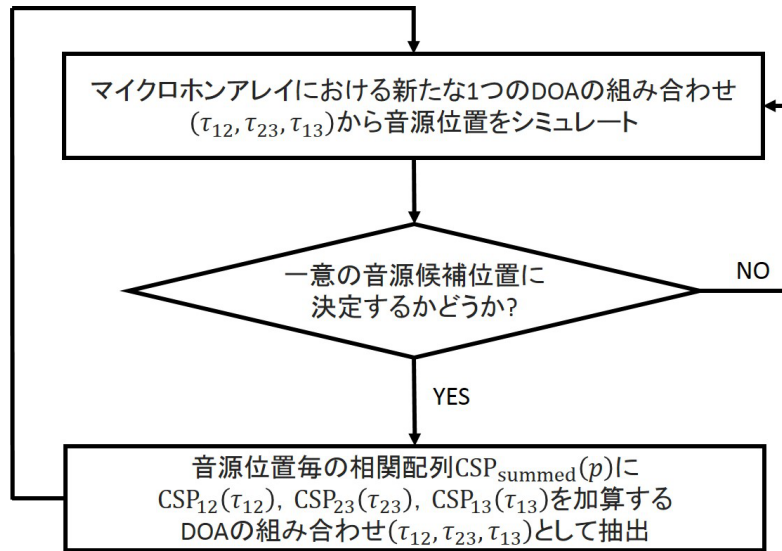


図 5.2: CSP 加算する DOA の抽出アルゴリズム

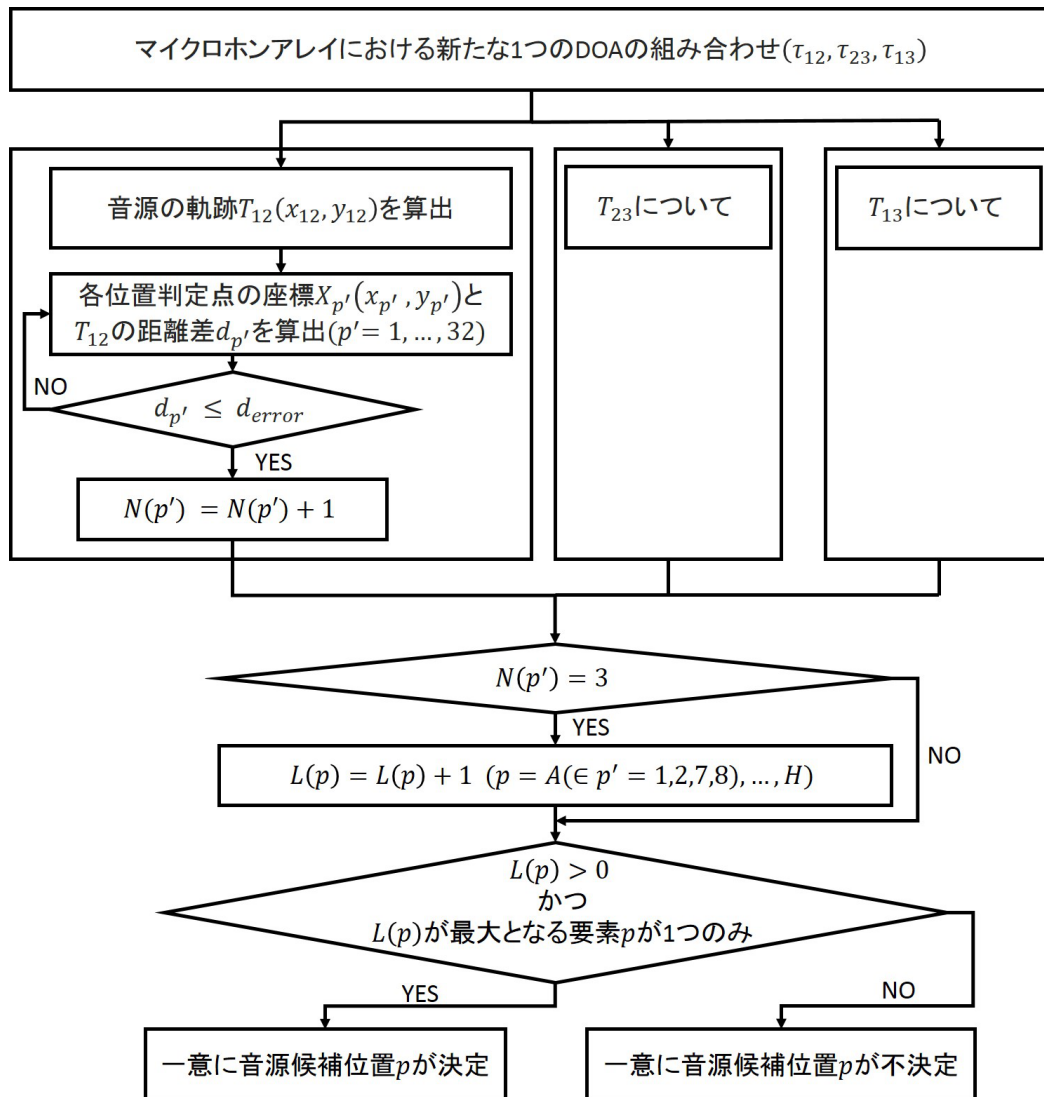


図 5.3: 一意に決定される音源位置候補位置の選択アルゴリズム

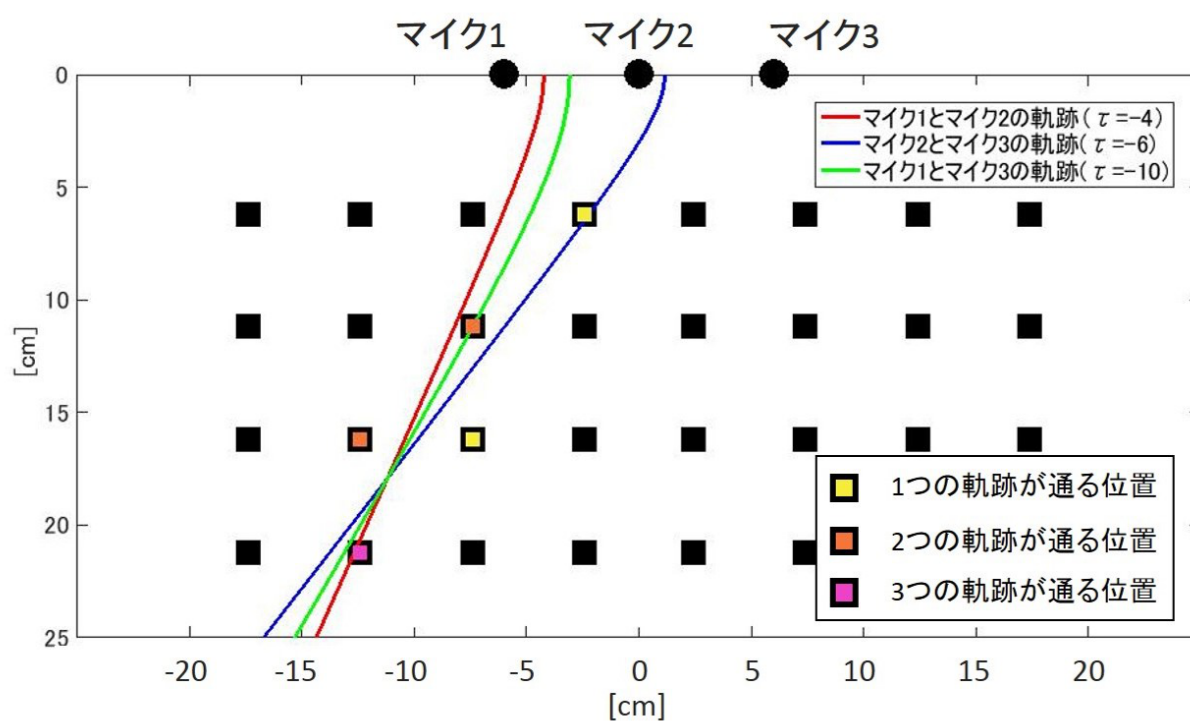


図 5.4: 一意の音源候補位置が選択されるイメージ図

5.3 本研究における CSP 加算法

本研究において、音源候補位置が一意に決まる DOA の組み合わせ $(\tau_{12}, \tau_{23}, \tau_{13})$ に対応する CSP 係数値が加算される。音源候補位置が一意に決まる DOA の組み合わせ $(\tau_{12}, \tau_{23}, \tau_{13})$ が N 組抽出された場合、 N 組の DOA の組み合わせ $(\tau_{n12}, \tau_{n23}, \tau_{n13})$ における CSP 係数値をすべて足し合わせる。 N 組の DOA の組み合わせにおける CSP 係数値が足し合わされた時、加算回数 N として平均をとり $\text{CSP}_{\text{summed}}(p)$ を算出した。よって $\text{CSP}_{\text{summed}}(p)$ は式 (5.3.1) から求めた。

$$\text{CSP}_{\text{summed}}(p) = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{M-1} \sum_{j=i+1}^M \text{CSP}_{ij}(\tau_{n_{ijp}}) \quad (5.3.1)$$

ここで $\tau_{n_{ijp}}$ は音源候補位置 p に一意に決まる n 組目の DOA の組み合わせの中の i 番目と j 番目のマイクロホンの DOA を示す。尚、CSP 係数は式 (4.2.1) から算出して、推定される音源候補位置 \hat{p} は式 (4.3.3) から算出した。

さらに本研究では、式 (4.2.1) から算出された CSP 係数に対して絶対値をとる処理を施している。相関を示す指標である CSP 係数は-1~1 の範囲の数値をとり、1 に近づくほど 2 つの音信号に正の相関があり、-1 に近づくほど負の相関があるといえる。CSP 係数は位相情報のみから 2 つの信号の相関を求めているので、CSP 係数が 1 に近づくほど 2 つの音信号は同位相の関係にあり、-1 に近づくほど逆位相の関係にあるといえる。図 4.2, 4.4, 4.6 のように、CSP 係数は同位相のより強い関係になる時間差の前後で、逆位相の強い関係が見られる。本研究では、CSP 加算する DOA に誤差を考慮するのに位置シミュレーションから抽出された複数の DOA をすべて足し合わせ平均をとるため、加算平均をとる時に CSP 係数の負値を加算することは $\text{CSP}_{\text{summed}}(p)$ の算出精度に悪い影響を与える可能性もあり、その影響を抑えるために CSP 係数の絶対値をとっている。最終的に本研究における $\text{CSP}_{\text{summed}}(p)$ は式 (5.3.2) から求めた。

$$\text{CSP}_{\text{summed}}(p) = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{M-1} \sum_{j=i+1}^M |\text{CSP}_{ij}(\tau_{n_{ijp}})| \quad (5.3.2)$$

5.4 提案手法の特徴

ここで、提案する CSP 加算法が従来法と異なっている点を以下に挙げる．

- CSP 加算する DOA は理論値ではなく、位置シミュレーション後に抽出．
- 1 つの音源候補位置において加算回数が 1 回ではなく、複数回ある．
- 絶対値をとった CSP 係数値を加算する．

1 つ目については、本研究において従来法のように DOA の誤差を許容する場合、CSP 加算される CSP 係数値に対応する DOA が他の音源候補位置の DOA となる可能性があるが、位置シミュレーション後に DOA を抽出するため他の音源候補位置の DOA における CSP 係数値を加算することはない．2 つ目については、音源方向が近似しているマイクロホンアレイから近い所と遠い所に位置する音源候補位置間で加算回数が異なり、それらの音源候補位置間の相関値に差を生じさせることで、真の音源がどちらかの音源候補位置であっても正しく推定できることが期待できる．3 つ目については、CSP 係数が最大となる時間差の前後で負の強い相関が見られ、本研究では CSP 加算回数で平均をとっており負値の加算が悪い影響を与えるため、CSP 係数の絶対値をとることでこの影響を抑える役目がある．また、1 つの音源候補位置において加算回数が複数回あり、連続した DOA における CSP 係数値が加算されることが大いにありうるため、CSP 加算の際に良い影響を与えることも期待できる．

以上のことから、本手法の特徴はマイクロホンアレイに限りなく近い音源位置において、音源方向が近似している音源候補位置に音源がある場合にも対応していることである．

第 6 章

音源位置推定実験

6.1 収録 A における音源位置推定実験

一意に決定される音源候補位置を選択するアルゴリズム (図 5.3) 内の許容誤差 d_{error} の条件を以下のように音源候補位置 (A~H) 毎で変化させて、音源位置推定実験を行った。

条件 a A~H に属する p' での許容誤差を 2.5 [cm].

条件 b B, C に属する p' での許容誤差を 1.5 [cm], その他に属する p' で 2.5 [cm].

条件 c B, C に属する p' での許容誤差を 1.0 [cm], その他に属する p' で 2.0 [cm].

入力データは図 5.1 の黄色の四角の位置を打撃して収録した打撃音である。これらの位置で収録した打撃音がその打撃音が属する図 5.1 に示す音源候補位置に推定された時、その推定が正解であるとした。各打撃位置において 1 回の収録で 10 回打撃音を収録して、その収録を 3 回行ったため、音源候補位置毎で入力データとして 120 データある。正解率は式 (6.1.1) から算出した。

$$\text{正解率 [\%]} = \frac{\text{正しい音源候補位置が推定された回数}}{120} \times 100 \quad (6.1.1)$$

ここで、それぞれの条件における 8 つの音源候補位置での正解率を図 6.1 に示す。提案法との比較のために、8 つの音源候補位置の中心座標が音源と仮定した時のマイクロホンアレイにおける理論的に算出される 1 組の DOA を理論値として、その理論値を利用した CSP 加算による 8 つの音源候補位置を推定した結果を従来法 (理論値, 収録 A) とした。

図 6.1 より、3 つの提案法は音源候補位置 B, C 以外では従来法 (理論値, 収録 A) の正解率よりも比較的高くなるという結果が得られ、提案法 (条件 c, 収録 A) が本研究に最も適した有用な手法であると考えられる。提案法 (条件 a, 収録 A), 提案法 (条件 b, 収録 A) において、音源候補位置 B, C での正解率が低くなったのは、加算回数に影響していると考えられる。

表 6.1 に条件毎のそれぞれの音源候補位置 A~H の CSP 加算回数を示す。音源候補位置 B, C において条件 a での加算回数が他の音源候補位置に比べ大きく異なる。音源

候補位置 B, C に属する p' での許容誤差を小さくすることで、すべての音源候補位置における加算回数が平坦化されている。加算回数が多いと音源候補位置 B, C の相関値が高くなる DOA の組み合わせはあるが、相関値が低くなる DOA の組み合わせが多くなり、音源候補位置毎の相関配列における B, C($=p$) の相関値 $\text{CSP}_{\text{summed}}(p)$ が平均をとることで低く抑えられたと考えらる。

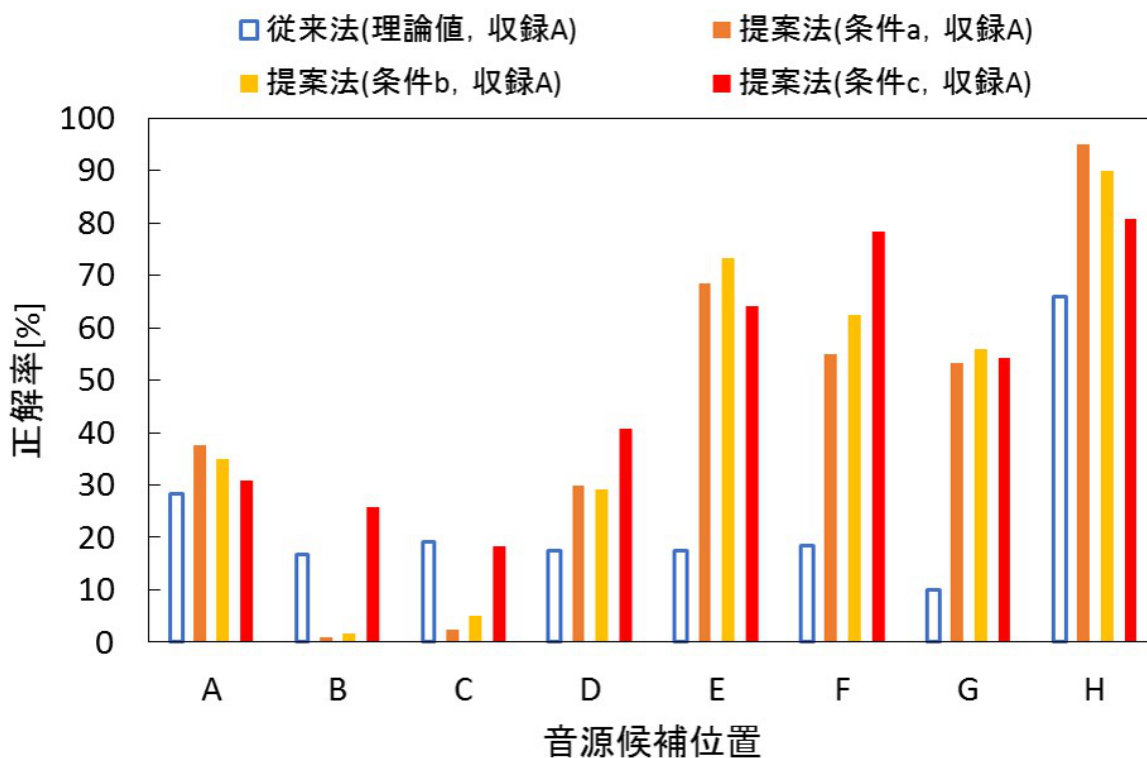


図 6.1: 収録 A における音源位置推定の正解率

表 6.1: 条件毎の CSP 加算回数

	A	B	C	D	E	F	G	H
条件 a	6	273	273	6	4	13	13	4
条件 b	6	46	46	6	6	29	29	6
条件 c	4	10	10	4	5	17	17	5

6.2 打撃対象物の影響

収録 A における音源位置推定実験で音源候補位置全体として正解率が高くならなかったのは、打撃を加えたアクリル板の厚みが 0.5 [cm] と薄いのが原因であると考えられる。アクリル板に打撃を加えるとアクリル板全体がたわみ、アクリル板が振動してアクリル板の面全体が音源となっている可能性がある。

ここで各音源候補位置の打撃音がどの音源候補位置に推定されたかを示す混同行列を表 6.2, 6.3 に示す。表 6.2, 6.3 において行ラベルが入力データとなる打撃音が収録された位置を示す音源候補位置であり、列ラベルが推定された音源候補位置を示す。表 6.2 は条件 b における混同行列であり、表 6.3 は条件 c における混同行列である。表 6.2, 6.3 から音源候補位置 B での打撃音がマイクロホンアレイの中心に対して逆の方向 D に、C での打撃音が逆の方向 A に、F での打撃音が逆の方向 D などに推定されている。また表 6.2, 6.3 において全体として、1 つの音源候補位置における推定位置にばらつきが見られる。このため収録 A における音源位置推定実験で音源候補位置全体として正解率が低くなった。このような結果になった原因が打撃をする際に打撃対象物がたわみが生じ、振動することで面全体が音源となり、真の音源候補位置の相関値と他の音源候補位置の相関値の差が多くの場合で小さくなったり、他の音源候補位置の相関値の方が大きくなったことであると考えられる。

表 6.2: 入力データ (A~H) が推定された音源候補位置 (提案法 (条件 b, 収録 A))

	A	B	C	D	E	F	G	H	計
A	42	13	3	7	16	21	14	4	120
B	15	2	0	10	57	34	0	2	120
C	50	0	6	14	1	1	25	23	120
D	1	3	9	35	7	0	31	34	120
E	6	7	0	8	88	8	0	3	120
F	9	0	0	31	0	75	0	5	120
G	6	0	0	46	0	1	67	0	120
H	1	0	0	2	0	1	8	108	120

表 6.3: 入力データ (A~H) が推定された音源候補位置 (提案法 (条件 c, 収録 A))

	A	B	C	D	E	F	G	H	計
A	37	11	4	23	4	23	18	0	120
B	12	31	0	7	39	29	1	1	120
C	46	1	22	4	0	0	36	11	120
D	3	10	19	49	0	0	24	15	120
E	15	7	0	11	77	10	0	0	120
F	1	8	0	15	0	94	0	2	120
G	3	14	26	12	0	0	65	0	120
H	12	0	0	2	1	0	8	97	120

6.3 収録 B における音源位置推定実験

打撃対象物としてレンガを利用した収録 B で音源位置推定実験を行った。この収録実験は対象物にたわみが生じない場合を想定して行った実験である。図 6.2 に 4 つの場合の正解率を示す。4 つの場合は、収録 B における CSP 加算する DOA に理論値を利用した従来法、収録 B における条件 b の提案法、収録 B における条件 c の提案法、収録 A における条件 c の提案法である。ここで、収録 B の音源位置推定実験は、CSP 加算回数が平坦化されている条件 b, c の 2 つの条件で行った。

提案法（条件 b, 収録 B）、提案法（条件 c, 収録 B）において、提案法（条件 c, 収録 A）よりも全体として良い結果が得られた。提案法（条件 b, 収録 B）の音源候補位置 B, C に関しては、提案法（条件 c, 収録 A）と正解率に差が見られないが、許容誤差の条件が同じである提案法（条件 b, 収録 A）の音源候補位置 B, C における正解率は 5 [%] 以下であり、収録 B においてより良い結果となっている。また、提案法（条件 b, 収録 B）、提案法（条件 c, 収録 B）において、80 [%] 以上の認識率となる音源候補位置が複数あり、提案法（条件 c, 収録 B）に関しては最大で 100 [%] の正解率となり条件 c が最も良い条件と考えられる。

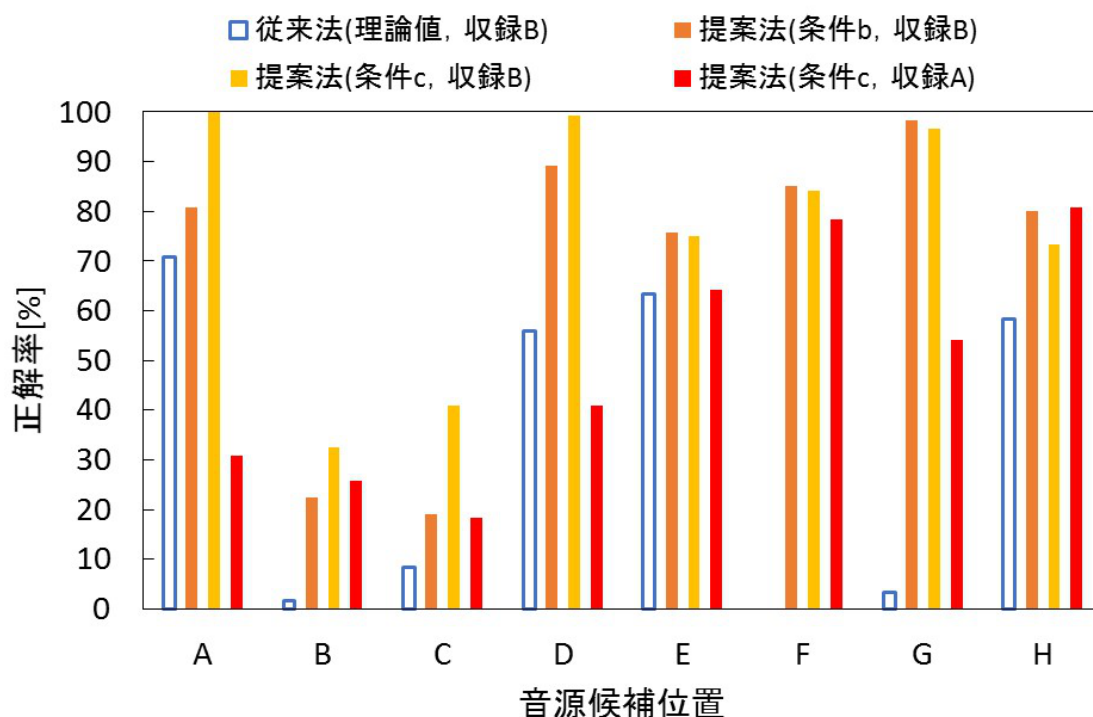


図 6.2: 収録 B における音源位置推定の正解率

ここで、提案法 (条件 b, 収録 B) の混同行列を表 6.4 に、提案法 (条件 c, 収録 B) の混同行列を表 6.5 に示す. 表 6.4, 6.5 からある音源候補位置での打撃音が逆の方向の音源候補位置に推定されることがほとんどなくなり, この結果から収録 A での音源位置推定実験は打撃対象物の影響を受けていることが確認できる.

収録 B において最も良い結果となった提案法 (条件 c, 収録 B) で図 5.1 などの黄色の四角の打撃位置 1~32 番のそれぞれの打撃音から推定された音源候補位置を示した混同行列を表 6.6 と表 6.7 に示す. 表 6.6 がマイクロホンアレイから近い打撃位置 1~16 番, 表 6.7 がマイクロホンアレイから遠い打撃位置 17~32 番の結果である. 表 6.6 において, 音源候補位置 B に属する打撃音の内 4 番のみの正解率は 100 [%] となり, 4 番の後ろの位置である 12 番も正解となる音源候補位置 B にある程度推定された. 3, 11 番については, 音源方向が似た方向で位置関係が近い音源候補位置 A, E に推定された. 対象の位置関係である音源候補位置 C についても同じような結果となっている. 表 6.7 において, 音源候補位置 E に属する打撃音の内 17 番が音源方向が似た方向で位置関係が近い音源候補位置 A に推定され, その他の打撃位置での打撃音は正解となる音源候補位置 E に推定された. 対象の位置関係である音源候補位置 H についても同じ結果となっている. また音源候補位置 F に属する打撃音の内 19 番が音源方向が似た方向で位置関係が近い音源候補位置 E に推定され, その他の打撃位置での打撃音は正解となる音源候補位置 F に推定された. 対象の位置関係である音源候補位置 G についても同じ結果となっている.

以上のことから, 提案するインタラクションにおいて打撃対象物は打撃の際にたわみが生じない所で使用するという条件が必要であると確認できた. また音源候補位置内で高い推定精度の打撃位置と低い推定精度の打撃位置が存在していることが確認できた. 高い推定精度の打撃位置については, 限りなく 100 [%] に近い精度となっている. したがって適切な音源候補位置となる許容範囲を設定することで音源候補位置毎の正解率が高くなることが期待できる. 適切な音源候補位置となる許容範囲を設定することは, 提案するインタラクションの仮想ボタンを設計する位置を決定することと等しい意味を持つ. 現状の結果であれば, 音源候補位置 B において 4 番近く, 音源候補位置 C において 5 番近くのみで設計することで 8 つの仮想ボタンを実現できると考えられる. また音源候補位置 B, F を統合し, 音源候補位置 C, G を統合した 6 つの仮想ボタンを実現でき, さらに 1 つの仮想ボタンの範囲を広げた 4 つの仮想ボタンであればより良い精度で実現できると考えられる. また現在の直線マイクロホンアレイに対して垂直方向に新たな直線マイクロホンアレイを配置することで, 正解率を向上することが期待でき, さらに音源候補位置を細分化して, 設計できる仮想ボタンの数を増やすことも期待できる.

表 6.4: 入力データ (A～H) が推定された音源候補位置 (提案法 (条件 b, 収録 B))

	A	B	C	D	E	F	G	H	計
A	97	12	0	0	11	0	0	0	120
B	11	27	0	0	52	30	0	0	120
C	0	0	23	15	0	0	31	51	120
D	1	1	7	107	1	1	0	2	120
E	22	7	0	0	91	0	0	0	120
F	0	7	0	0	11	102	0	0	120
G	0	0	0	0	0	0	118	2	120
H	0	0	5	18	0	0	1	96	120

表 6.5: 入力データ (A～H) が推定された音源候補位置 (提案法 (条件 c, 収録 B))

	A	B	C	D	E	F	G	H	計
A	120	0	0	0	0	0	0	0	120
B	15	39	0	0	45	21	0	0	120
C	0	0	49	17	0	0	12	42	120
D	0	0	0	119	0	0	0	1	120
E	30	0	0	0	90	0	0	0	120
F	0	0	0	0	19	101	0	0	120
G	0	0	0	0	0	0	116	4	120
H	0	0	1	31	0	0	0	88	120

表 6.6: 入力データ (1～16 番) が推定された音源候補位置 (提案法 (条件 c, 収録 B))

	A	B	C	D	E	F	G	H	計
1	30	0	0	0	0	0	0	0	30
2	30	0	0	0	0	0	0	0	30
3	15	0	0	0	15	0	0	0	30
4	0	30	0	0	0	0	0	0	30
5	0	0	30	0	0	0	0	0	30
6	0	0	0	17	0	0	0	13	30
7	0	0	0	29	0	0	0	1	30
8	0	0	0	30	0	0	0	0	30
9	30	0	0	0	0	0	0	0	30
10	30	0	0	0	0	0	0	0	30
11	0	0	0	0	30	0	0	0	30
12	0	9	0	0	0	21	0	0	30
13	0	0	18	0	0	0	12	0	30
14	0	0	1	0	0	0	0	29	30
15	0	0	0	30	0	0	0	0	30
16	0	0	0	30	0	0	0	0	30

表 6.7: 入力データ (17～32 番) が推定された音源候補位置 (提案法 (条件 c, 収録 B))

	A	B	C	D	E	F	G	H	計
17	30	0	0	0	0	0	0	0	30
18	0	0	0	0	30	0	0	0	30
19	0	0	0	0	19	11	0	0	30
20	0	0	0	0	0	30	0	0	30
21	0	0	0	0	0	0	30	0	30
22	0	0	0	0	0	0	26	4	30
23	0	0	0	0	0	0	0	30	30
24	0	0	0	30	0	0	0	0	30
25	0	0	0	0	30	0	0	0	30
26	0	0	0	0	30	0	0	0	30
27	0	0	0	0	0	30	0	0	30
28	0	0	0	0	0	30	0	0	30
29	0	0	0	0	0	0	30	0	30
30	0	0	0	0	0	0	30	0	30
31	0	0	0	0	0	0	0	30	30
32	0	0	1	1	0	0	0	28	30

第7章

結論

タッチパネルの補助となる音源位置情報を用いたインタラクションシステムを構築するために、重要な要素となる打撃音の位置情報を用いた仮想ボタンによるボタン入力の実現に向けた打撃位置推定手法について検討した。打撃位置推定手法として CSP 加算法を適応した。CSP 加算法による音源位置推定の先行研究でのマイクロホン間隔やマイクロホンアレイから音源までの距離には、本研究が想定しているそれらの条件と大きな差があるため、先行研究の CSP 加算法を適用することが困難であると考えた。そこで本研究に適した CSP 加算法を提案し、打撃音の収録、音源位置推定の実験から仮想ボタンによるボタン入力の実現性を調査した。

打撃を加えた際に打撃対象物の面全体が振動する厚みが薄いアクリル板と振動しない厚みがあるレンガを利用して打撃音を収録した。アクリル板よりもレンガによる収録実験の音源位置推定の正解率が大きく向上することを確認した。よって打撃を加えた際に打撃対象物のたわみによる面全体の振動が発生しない環境で提案するインタラクションを利用する必要があると確認でき、机の上などで利用できると考えられる。レンガによる収録実験における 8 つの音源候補位置での音源位置推定実験から提案法で 6 つの音源候補位置で高い正解率を確認できた。現状では 6 つの仮想ボタン、少なくとも 4 つの仮想ボタンによるボタン入力の実現できる。また音源候補位置の許容範囲を適切に設定できれば、つまりは仮想ボタンの設計位置を工夫できれば 8 つの仮想ボタンによるボタン入力の実現できると考えられ、タッチパネルの補助となるインタラクションシステムを構築できると確認した。また指を鳴らしたり、人の声などの入力方法で平面以外でも利用できることを音源位置情報を用いるインタラクションのメリットとしたが、3次元空間上では振動の影響を受けず、打撃対象物の比較と正解率から実現できると考える。

今後の課題として、音源位置推定実験から各音源候補位置における正解率の平均の最大が 75.2 [%] となり、この平均正解率の更なる向上のために推定精度が低いマイクロホンアレイに近い音源候補位置 B, C での推定精度の向上が挙げられる。またユーザの入力となる打撃音の収録、位置推定の処理、ユーザの入力に対応したコンピュータの動作や出力をするインタラクションシステムの構築、さらに構築されたインタラクションシステムがユーザにとって使いやすい入力システムとなるかの評価などが挙げられる。

参考文献

- [1] 椎尾 一郎, ” ヒューマンコンピュータインタラクション入門, ” サイエンス社, 2010.
- [2] D. Vogel and P. Baudisch, “Shift: a technique for operating pen-based interfaces using touch,” Proc. CHI’07, 657-666, 2007.
- [3] C. Holz and P. Baudisch, “The generalized perceived input point model and how to double touch accuracy by extracting fingerprints,” Proc. CHI’10, 581-590, 2010.
- [4] A. Roudaut, S. Huot, and E. Lecolinet, “TapTap and MagStick: improving one-handed target acquisition on small touch-screens,” Proc. AVI’08, 146-153, 2008.
- [5] P. Baudisch and G. Chu, “Back-of-device interaction allows creating very small touch devices,” Proc. CHI’09, 1923-1932, 2009.
- [6] C. Harrison and S. E. Hudson, “Scratch input: creating large, inexpensive, un-powered and mobile finger input surfaces,” Proc. UIST’08, 205-208, 2008.
- [7] R. Xiao, G. Lew, J. Marsanico, D. Hariharan, S. E. Hudson, and C. Harrison, “Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation,” Proc. MobileHCI’14, 67-76, 2014.
- [8] C. Harrison, D. Tan, and D. Morris, “Skinput: appropriating the body as an input surface,” Proc. CHI’10, 453-462, 2010.
- [9] 浅野 太, ” 音のアレイ信号処理 -音源の定位・追跡と分離-, ” コロナ社, 2011.
- [10] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay,” IEEE Transactions on Acoustics, Speech, and Signal, Processing, vol.24, no.4, 320-327, 1976.
- [11] 林田 亘平, 森勢 将雅, 西浦 敬信, ” 近接音源位置推定のための 2D-CSP 法の評価,” 電気情報通信学会技術研究報告. EA, 応用音響, 110(54), 49-54, 2010
- [12] 西浦 敬信, 山田 武志, 中村 哲, 鹿野 清宏, ” マイクロホンアレーを用いた C S P 法に基づく複数音源位置推定,” 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J83-D-II(8), 1713-1721, 2000.
- [13] 奥村 元, 趙 國, 西浦 敬信, 山下 洋一, ” 分散マイクロホンシステムを用いた音源位置同定,” 電子情報通信学会技術研究報告. TL, 思考と言語 106(485), 61-66, 2007 .
- [14] 趙 國, 奥村 元, 西浦 敬信, 山下 洋一, ” 実環境における分散マイクロホンシステ

- ムを用いた CSP 法に基づく音源位置同定,” 日本音響学会春季講演論文集, 3-Q-1, 629-630, 2006.
- [15] P. Marmaroli, X. Falourd, and H Lissek, ” A comparative study of time delay estimation techniques for road vehicle tracking,” Proceedings of the 11th French Congress of Acoustics and 2012 Annual IOA Meeting, 4134-4140, 2012.
- [16] 上杉 信敏, 金田 豊, “音源方向検出法の室内残響下での性能評価について,” 日本音響学会春季講演論文集, 3-Q-4, 635-636, 2006.
- [17] 城戸 健一, ” デジタルフーリエ解析 (II) -上級編-, ” コロナ社, 2007.

謝辞

本研究を進めるにあたり，終始適切なご指導とご助言をいただきました，本学工学研究科准教授 西野 隆典 先生に深く感謝します．本研究に対し，貴重な助言を頂きました，本学工学研究科 教授 成瀬 央 先生に厚くお礼申し上げます．また，学生生活を共にした，パターン情報処理研究室の皆様に感謝します．