

# 博士學位論文

楽器音を対象とした単一音源における  
方向推定に関する研究

近畿大学大学院  
システム工学研究科 システム工学専攻

山本華帆

# 博士學位論文

楽器音を対象とした単一音源における  
方向推定に関する研究

2023年1月6日

近畿大学大学院  
システム工学研究科 システム工学専攻

山本華帆

# 目次

第1章	序論 .....	1
第2章	基本周波数成分を用いた単一音源の方向推定 .....	3
2.1	序言 .....	3
2.2	MUSIC 法による音源方向推定 .....	3
2.3	実験 .....	7
2.3.1	実験条件 .....	8
2.3.2	実験結果 .....	9
2.4	結言 .....	9
第3章	高調波成分を用いた音源方向推定 .....	22
3.1	序言 .....	22
3.2	FFNN を用いた音源方向推定 .....	22
3.3	実験 .....	25
3.3.1	実験条件 .....	25
3.3.2	実験結果 .....	26
3.3.3	MUSIC 法による推定結果との比較 .....	51
3.4	結言 .....	52
第4章	基本周波数が高い音響信号を対象とした方向推定における推定誤差の 低減 .....	53
4.1	序言 .....	53
4.2	CNN を用いた音源方向推定 .....	53
4.3	実験 .....	56
4.3.1	実験条件 .....	56
4.3.2	MUSIC 法による推定結果 .....	58
4.3.3	CNN による結果 .....	66
4.3.4	MUSIC 法による推定結果との比較 .....	72
4.4	結言 .....	75
第5章	結論 .....	76
参考文献	.....	78

謝辞

# 第1章 序論

サラウンド音響[1]は映画館などでは古くから導入されているが、近年ではテレビ放送やDVDなどのデジタルコンテンツもサラウンド音響に対応し、家庭でも楽しめるようになりつつある。サラウンドは3つ以上のスピーカで立体的な音場を作り出すシステムのことである。特に、複数のスピーカを聴取位置の周りに等距離に配置して音場を再生するシステムはマルチチャンネル音響[2]と呼ばれ、電気通信や無線通信の分野で世界的標準の確立を図る国際電気通信連合 (ITU: International Telecommunication Union) によりスピーカの配置方法が推奨されている[3]。また、高臨場映像音響メディアの実現に向けて、22.2 マルチチャンネル音響に関する研究開発も進められている[4]。人間が音の方向を知覚するうえで特に重要な要素は、左右の耳における音の振幅差 (ILD : Interaural Level Difference) と時間差 (ITD : Interaural Time Difference) で、この差によって人間は音の方向を判断している[5,6]。ステレオでは、左右2つのスピーカで同じ信号に振幅差をつけて再生することによって音像定位 (仮想音源) を制御することができる。ステレオでは各スピーカを繋ぐ線分上での音の移動の再現は可能であったが、サラウンドはステレオよりもスピーカ数が増加するため、360度空間で直線でない音の移動の再現が可能となる。しかし、サラウンド音響システムの使用時には「使用者が所望する音響空間を構築することの困難さ」や「部屋の形状などによる音像バランスの乱れ」などの問題が発生することが多い。そこで、サラウンド音響を可視化することで音源定位の調整を一般の使用者が容易に行なうことができるシステムの構築を将来的な目標とする[7~9]。音は可視化することによって音の定位や分布を認知しやすくなるため、騒音対策や楽器の音響特性の調査などを目的として可視化についての研究が行なわれてきた[10~14]。サラウンド空間をタブレットなどの端末に表示して、定位を調整するシステムを構築することにより、サラウンド音響の使用者に適した調整を行なうことができるようになり、例えば映像と音の動きが対応するVRシステムやホームシアターを違和感なく楽しむことが可能となる。

使用者が所望する音響空間を構築するために音の定位を調整するにあたって、前述したシステムの構築には音源分離や仮想音源の位置が必要となる。しかしながら、サラウンド音源は使用する環境によって異なる反響が生じるなど、音響信号を作成する際の理想的な空間と使用者が聴取する実際の音響空間とは異なるため、サラウンド音響システムを実際に使用する空間での仮想音源位置を知る必要がある[15~17]。本研究ではサラウンド音響での仮想音源の位置推定の基礎的研究として、単一音源 (実音源) での方向推定シミュレーションを行なう。

音源方向推定は、音源探索や背景雑音の除去、話者の自動識別などを目的として研究が進められている[18~20]が、楽器音を対象とする方向推定に関する研究は進んでいない。そこで本論文では、単一の楽器音の音源方向推定について検討する。音源方向



推定としては MUSIC (Multiple Signal Classification) 法[21]や ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) 法[22]などが提案されている。MUSIC 法を用いた音源方向の推定手法として、周波数の上限下限を定めてその範囲の MUSIC スペクトルを算出する手法がすでに提案されている[23]。これに対して本研究では楽器音の倍音構造に着目して MUSIC スペクトルの算出を行なう。倍音構造が存在する楽器音は、周波数によって振幅値が異なる。音源方向推定に使用する周波数成分の違いによって推定結果が変化し、振幅値の比較的大きい高調波成分を用いた方向推定では定位方向と推定方向の差（推定誤差）は小さくなる。そこで、機械学習の一種であるニューラルネットワークを用いて、楽器や周波数ごとの誤りパターンを学習させることで推定精度の改善を図る。また、基本周波数が高くなると、推定誤差が大きくなる傾向があり、基本周波数成分及び高調波成分により MUSIC スペクトルを用いることで推定誤差の低減を図る。

本論文の構成は以下の通りである。

第2章では、DOA の一種である MUSIC 法による楽器音の方向推定手法として基本周波数成分のみを用いる手法について述べる。MUSIC 法において、倍音構造を持つ楽器音を対象とした方向推定では高調波成分の振幅値が大きいと BPF (Band Pass Filter) を適用しても高調波成分が除去されきれず、振幅値の大きい周波数成分が分析対象信号に複数存在することが原因で定位方向と推定方向に差が生じる。そこで、分析対象信号が倍音構造をもつ楽器音の場合、振幅値の比較的大きい基本周波数成分のみを用いることで基本周波数以外の成分による方向推定への影響を低減させる。シミュレーションにより推定精度を評価し、最後に今後の課題について述べる。

第3章では、楽器音の高調波成分を用いた音源方向推定手法について述べる。本研究で方向推定の対象となる楽器音源は、基本周波数の整数倍の周波数にピークを持つ倍音構造を持つ。基本波成分や高調波成分は周波数成分が他の周波数成分よりも振幅値が比較的大きく、これらの成分を用いた方向推定では推定誤差が小さくなる。そこで、最も基本的なニューラルネットワークである FFNN (Feedforward Neural Network) を用いて、楽器や周波数ごとの誤りパターンを学習させることで推定精度の向上を図る。シミュレーションにより推定精度を評価し、最後に今後の課題について述べる。

第4章では、分析対象信号の基本周波数が高い場合の音源方向推定手法について述べる。基本周波数が高い信号は MUSIC 法における方向推定において、空間エイリアシングにより正しい推定ができない傾向にある。そこで、より複雑なニューラルネットワークである CNN (Convolutional Neural Network) に MUSIC 法により得られる MUSIC スペクトルを入力として定位方向を推定する。CNN を用いた回帰分析と分類問題の2つのアプローチにより推定誤差の低減を図る。シミュレーションにより推定精度を評価し、最後に今後の課題について述べる。

最後に第5章では、本研究で得られた結果の総括を行ない、結論とする。

## 第2章 基本周波数成分を用いた単一音源の方向推定

### 2.1 序言

到来方向推定 (DOA : Direction Of Arrival) 法は無線通信の分野で使用されている技術で, 移動通信や無線 LAN など基地局の設置を効率よく行なうことや, 多重波伝搬を適切にモデル化することを目的として研究が進められており, ビームフォーマ法や線形予測 (LP) 法などが存在する[24]. MUSIC 法は DOA の一種で, 空間分解能が高く, 到来方向に鋭いピークを持つことが特徴であり, 音源方向や音源位置の推定にも用いられている[25, 26]. 本研究では, MUSIC 法を用いてモノラル環境での方向推定シミュレーションを行なう. MUSIC 法を用いる音源方向の推定手法として, 周波数の上限下限を定めてその範囲の MUSIC スペクトルを算出する手法が提案されている. 本研究では, 楽器音の倍音構造に着目して MUSIC スペクトルの算出を行なう[27, 28]. 倍音構造が存在する楽器音は, 周波数によって振幅値が異なる. 音源方向推定に使用する周波数成分の違いによって推定結果が変化し, 振幅値の比較的大きい高調波成分を用いた方向推定では定位方向と推定方向の差 (推定誤差) は小さくなる. また, 基本周波数が高くなると, 推定誤差が大きくなる傾向があり, 基本周波数成分及び高調波成分により MUSIC スペクトルを用いることで推定誤差の低減を図る.

本章の構成は次の通りである. 2.2 節では MUSIC 法による方向推定と短時間フーリエ変換における窓関数のフレーム長や移動長による影響について述べる. 2.3 節では MUSIC 法による方向推定シミュレーションの実験についての結果と考察について述べ, 2.4 節で結言とする.

### 2.2 MUSIC 法による音源方向推定

MUSIC (Multiple Signal Classification) 法は相関行列の固有値と固有ベクトルを用いる手法で, 直線状に等間隔に配置された  $K$  個のマイクロフォンにより,  $L$  個の音源の方向  $\theta_l$  ( $l = 1, 2, \dots, L$ ) を推定する.  $k$  番目のマイクロフォン  $M_k$  への入力信号を  $z_k(t)$  ( $k = 1, 2, \dots, K$ ) とすると, その短時間フーリエ変換 (STFT : Short-Time Fourier Transform) は式(1)で表される. ここで  $w(n)$  は入力信号  $z_k(t)$  を有限区間で切り取る窓関数である. 本研究では式(2)で示すハニング窓を使用する.

$$x_k(t, f) = \sum_{n=0}^{N-1} z_k(t+n)w(n)e^{-j2\pi fn/N} \quad (1)$$

$$w(n) = \begin{cases} 0.5 - 0.5\cos(2\pi n/N) & (0 \leq n \leq N-1) \\ 0 & (\text{otherwise}) \end{cases} \quad (2)$$

また, 図 2.1 に示すように $\theta$ 方向に音源があると仮定すると,  $M_k$ への入力信号に対する $M_{k-1}$ への入力信号の遅延時間 $D_k(\theta)$ は式(3)となる. ここで $\tau(\theta)$ は式(4)で示すマイクロフォン間の到来時間差であり,  $d$ はマイクロフォンの設置間隔,  $c$ は音速である.

$$D_k(\theta) = (k-1)\tau(\theta) \quad (3)$$

$$\tau(\theta) = d\sin\theta/c \quad (4)$$

ここで, 方向ベクトル $\mathbf{d}(f, \theta)$ , 入力された信号ベクトル $\mathbf{x}(t, f)$ ,  $\mathbf{x}(t, f)$ の相関行列 $\mathbf{R}(f)$ を式(5)~(7)でそれぞれ定義する. 上付き文字 H は複素共役転置を表す.

$$\mathbf{d}(f, \theta) = [e^{-j2\pi f D_1(\theta)}, e^{-j2\pi f D_2(\theta)}, \dots, e^{-j2\pi f D_K(\theta)}]^H \quad (5)$$

$$\mathbf{x}(t, f) = [x_1^*(t, f), x_2^*(t, f), \dots, x_K^*(t, f)]^H \quad (6)$$

$$\mathbf{R}(f) = \int \mathbf{x}(t, f)\mathbf{x}(t, f)^H dt \quad (7)$$

$\mathbf{R}(f)$ の小さい方から $I$  ( $I = K - L$ )個の固有値に対応する固有ベクトルを $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_I$ として,  $\mathbf{R}'(f)$ を式(8)で定義する.

$$\mathbf{R}'(f) = \sum_{i=1}^I \mathbf{v}_i \mathbf{v}_i^H \quad (8)$$

MUSIC スペクトルと呼ばれる MUSIC 法による到来信号のパワーは, 式(9)で求めることができる.

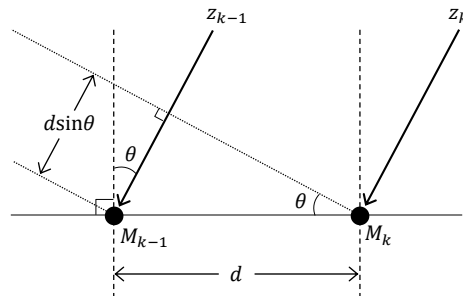


図 2.1  $M_k$ への入力信号に対する $M_{k-1}$ への入力信号の遅延時間の概略図

$$P_{MUSIC}(f, \theta) = \frac{\mathbf{d}(f, \theta)^H \mathbf{d}(f, \theta)}{\mathbf{d}(f, \theta)^H \mathbf{R}'(f) \mathbf{d}(f, \theta)} \quad (9)$$

本研究における方向推定は単一音源を対象としている ( $L = 1$ ) ため, あらかじめ定めた MUSIC スペクトルの探索範囲内における MUSIC スペクトルの最大値を探すことによって到来方向  $\theta_e$  を求める.

本研究では楽器音を対象として方向推定を行なう. 倍音構造[29]が存在する楽器音における振幅スペクトルは図 2.2, 2.3 のように基本周波数の整数倍でピークを示す. すべての周波数成分を用いると振幅値の小さい周波数成分の影響により推定誤差が大きくなる. また, 基本周波数の周辺のみを通過させる帯域通過フィルタ (BPF: Band Pass Filter) を音源に適用しても推定誤差が生じる. ここで, STFT における窓関数のフレーム長や移動長については次の通り検討した.

まず, STFT における窓関数のフレーム長  $N$  について述べる. 本研究で用いる音源については第 2.3 節で述べるが, すべての音源において標本化周波数は 44100sample である.  $N$  を 2048 としたとき, 分析される周波数の間隔は約 21.5Hz である. 通過帯域が  $f_0 \pm 20\text{Hz}$  である BPF を適用した  $f_0 = 55\text{Hz}$  の楽器音を周波数分析すると結果は図 2.4 となる. この時,  $N$  を 2 倍とすると分析される周波数の間隔は約 10.8Hz となり, 図 2.5 のように振幅値が最大となる周波数とそれ以外の差が大きくなる. このとき, 推定結果はわずかに改善する. しかし, 同様に  $N$  を音源の標本化周波数である 44100 とする, すなわち分析される周波数の間隔が 1Hz となる場合, 推定結果は  $N = 4096$  の場合と同じであるが,  $N = 44100$  の場合は  $N = 4096$  の場合と比較して相関行列の計算時間は約 400 倍となる. したがって  $N$  は一定以上の長さであれば推定結果に影響はない.

次に, STFT における窓関数の移動長について述べる. MATLAB とほぼ互換性のある数値計算ソフトウェアである GNU Octave で用意されている STFT 関数における移

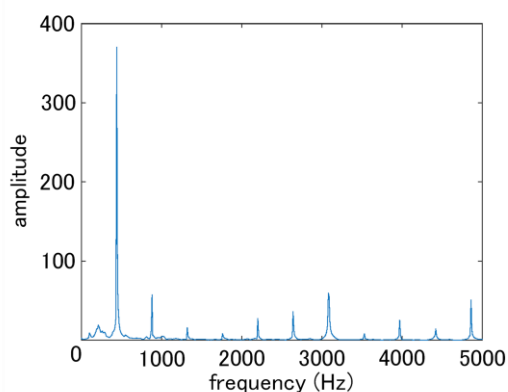


図 2.2  $f_0 = 440\text{Hz}$  である clean guitar の振幅スペクトル

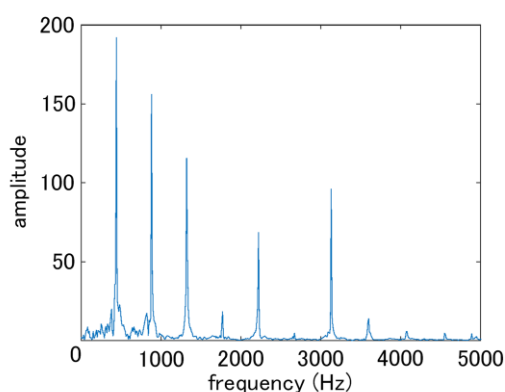


図 2.3  $f_0 = 440\text{Hz}$  である grand piano の振幅スペクトル

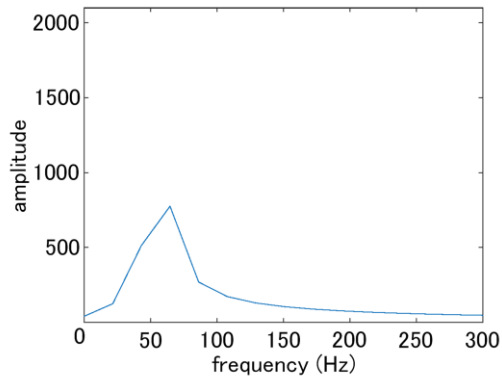


図 2.4 BPF を適用した  $f_0 = 55\text{Hz}$  である grand piano の振幅スペクトル ( $N = 2048$ )

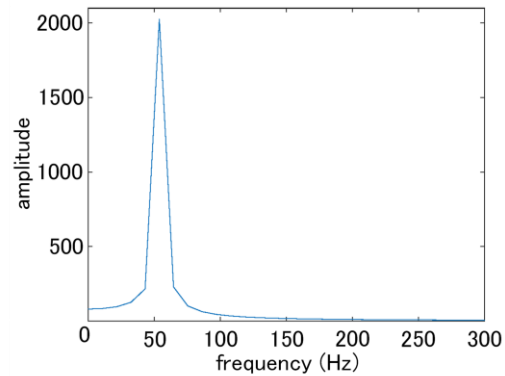


図 2.5 BPF を適用した  $f_0 = 55\text{Hz}$  である grand piano の振幅スペクトル ( $N = 4096$ )

動長の既定値は 24 である．本研究において分析対象信号の基本周波数が STFT の移動長に対して整数倍ではない．ある音響信号に対して STFT を行なう場合，まず図 2.6 のような基本周波数が STFT の移動長に対して整数倍の場合を考える．各フレームにおいて周波数分析した結果は異なるが，第 1 フレームと第 5 フレームは切り取る信号が全く同一であるため周波数分析した結果も同一となる．したがって，STFT の結果は (1 波長のサンプル数) / (STFT の移動長) ごとに同じ値となる．しかし，図 2.7 のように基本周波数が STFT の移動長に対して整数倍でない場合，第 1 フレームに対して第 2 から第 5 までの各フレームを切り取る信号が全く同一のものが存在しない．すなわち，周波数分析した結果も同一にはならない．STFT の結果が同一の値になるのは (1 波長のサンプル数) / (STFT の移動長) より大きくなるのは明らかである．したがって，相関行列の周期は基本周波数が STFT の移動長に対して整数倍の場合よりも大きくなり，推定方向が定位方向よりも外側になっていると考えられる．そこで，STFT の移動長と基本周波数が必ず整数倍の関係になるように，STFT の移動長を 1 として推定を行なった．その結果，推定結果に変化はなく，窓関数の移動長は推定結果に影響を与えない．

そこで，先に述べたように基本周波数の周辺のみを通過させる BPF を音源に適用すると基本周波数成分以外の成分が完全に除去されきれないことで推定誤差が生じることについて検討する．単一周波数信号である正弦波による方向推定を行なった結果を図 2.8, 2.9 に示す．正弦波を用いた場合の推定結果において，定位方向と推定方向の差は小さくなる．そこで本研究では倍音構造を持つ楽器音源を正弦波と同等の条件の信号となるように振幅値が比較的大きくなる基本周波数成分のみを用いて MUSIC スペクトルを求める．

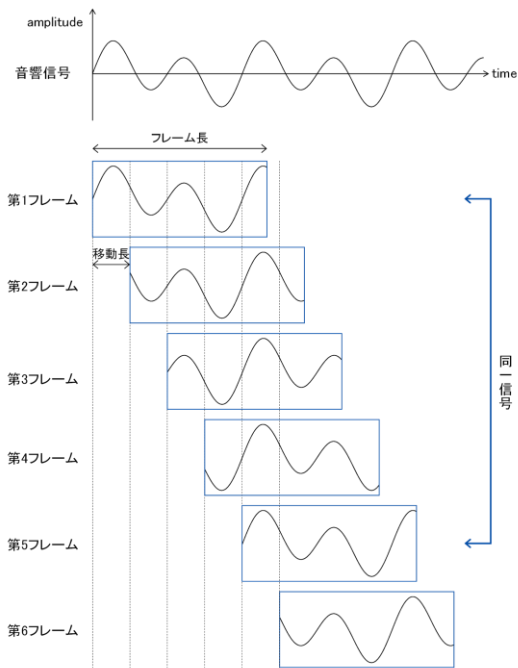


図 2.6 基本周波数が STFT の移動長に対して整数倍である場合の分析フレームの変化

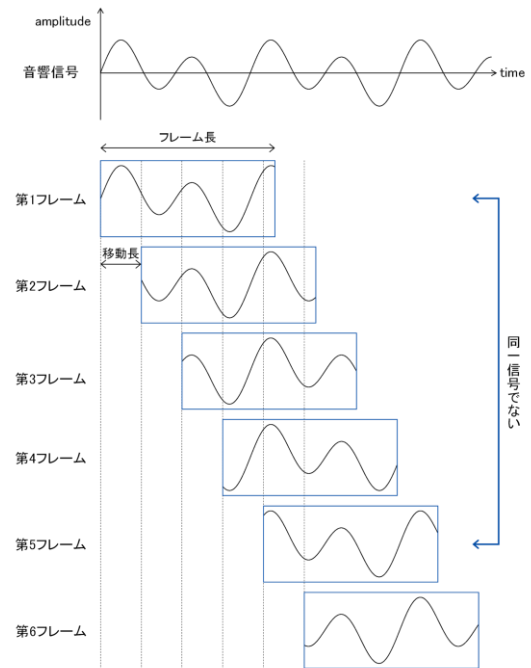


図 2.7 基本周波数が STFT の移動長に対して整数倍ではない場合の分析フレームの変化

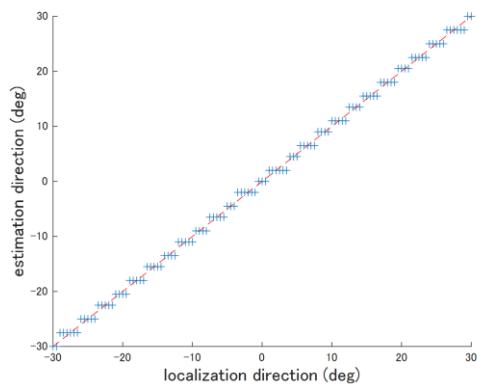


図 2.8  $f_0 = 27.5\text{Hz}$  の sine wave による方向推定結果

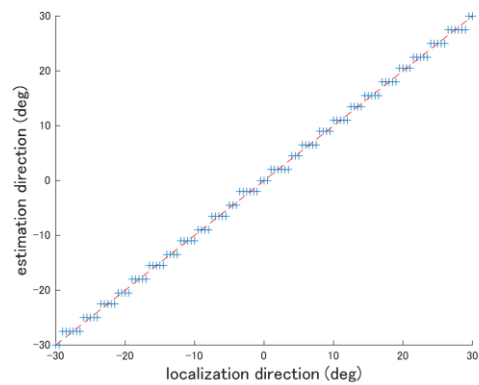


図 2.9  $f_0 = 880\text{Hz}$  の sine wave による方向推定結果

## 2.3 実験

本節では、基本周波数成分を用いた MUSIC 法による単一音源の方向推定のシミュレーションについて述べる。2.3.1 項ではシミュレーションの条件について、2.3.2 項ではシミュレーションの結果について述べる。

### 2.3.1 実験条件

本研究では $K = 3$ として単一音源における音源方向推定のシミュレーションを行なう。方向推定に用いる音源は guitar 5 種 (acoustic guitar, clean guitar, country guitar, jazz guitar, nylon guitar) , piano 5 種 (acoustic piano, electric piano, grand piano, rock piano, stage piano) の計 10 種類 (標本化周波数 = 44100Hz, 量子化ビット数 = 16bits, 基本周波数  $f_0 = 55, 110, 220, 440, 880, 1760$ Hz) で, いずれも音楽制作ソフトウェア Studio One を使用して作成したものがある。本研究では基礎的な実験として, 音源定位の方向は図 2.10 に示すように  $M_2$  正面から左右 30 度の範囲の単一音源 (0.5 度間隔の全 121 パターン) に制限する。なお, 本研究では  $M_2$  正面を 0 度とし, 左方向はマイナス, 右方向はプラスの角度で表す。また,  $d = 0.1\text{m}$  とし, 式(5)~(9)における  $f$  は基本周波数である  $f_0$ , 減衰・雑音・反響はない仮想空間上でのシングルスピーカによる放射と仮定する。なお, 分析フレーム長は 44,100 samples, STFT における窓関数のフレーム長  $N$  は 4096 samples, 移動長は 24 samples とする。また, 到来角の探索数  $p$  は 361 (-90 度から +90 度, 0.5 度間隔) である。推定精度の指標には式(10)で示す MSE (Mean Square Error) , 式(11)に示す定位方向と推定方向の完全一致率 (%) , 式(12)に示す推定誤差 (定位方向と推定方向の差) の最大 (度) を用いる。ここで,  $y_i$  は定位方向 (度),  $\hat{y}_i$  は推定方向 (度) であり,  $m$  は標本数である。

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (10)$$

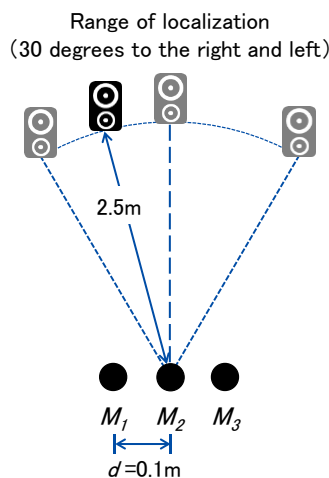


図 2.10 收音シミュレーションの概略図 1 (定位方向が  $M_2$  正面から左 15 度 (-15 度) にある場合)

$$\text{完全一致率} = \frac{\text{推定誤差が 0 となるサンプル数}}{m} \times 100 \text{ [\%]} \quad (11)$$

$$\text{maximum estimation error} = \max|\theta_l - \theta_e| \quad (12)$$

### 2.3.2 実験結果

音源方向のシミュレーション結果を表 2.1, 2.2 及び図 2.11~2.70 に示す. 方向推定の対象信号が guitar の場合の MSE の最大は 2.161 (acoustic guitar,  $f_0 = 1760\text{Hz}$ ) で, 基本周波数が 1760Hz の場合は全ての音色において MSE が 0.5 を上回った. また, 完全一致率は acoustic guitar ( $f_0 = 1760\text{Hz}$ ) を対象とした方向推定の場合に最も低く 12.40% である. この時の推定結果は図 2.16 である. 定位方向が  $M_2$  正面から離れるにつれて推定誤差が大きくなり, 最大で 3.5 度の差が生じている. また, 算出される推定方向が  $M_2$  正面と定位方向の間になる傾向にある. 対象信号が piano の場合の MSE の最大は 0.636 (electric piano,  $f_0 = 110\text{Hz}$ ) である. この時の推定結果は図 2.48 で, acoustic guitar ( $f_0 = 1760\text{Hz}$ ) を対象とした方向推定の場合とは対照的に定位方向が  $M_2$  正面から離れた方向が算出されている.  $f_0 = 1760\text{Hz}$  の場合においては acoustic piano 以外の音源で MSE が 0.5 を上回ったが, piano を対象とした場合の推定誤差は最大 2.0 度であり, 完全一致率は 20~23% となった.

全ての推定結果について, 完全一致率が 22.31%, MSE が 0.496, 推定誤差の最大が 1.5 度となることが多く, 基本周波数のみを用いた場合, 方向推定となる信号の周波数特性がほぼ同一になることから音色の違いによる推定結果の違いは小さくなると考えられる. 基本周波数によって MSE が変化し, 特に基本周波数が高くなると MSE が大きくなる傾向にあり, 楽器によっては基本周波数が低い場合にも MSE が大きくなることが示された.

## 2.4 結言

本章では, MUSIC 法を用いる音源方向推定手法について述べた. 倍音構造をもつ楽器音を対象とした音源方向推定では, 振幅値の大きい周波数成分が分析対象信号に複数存在することで推定誤差が大きくなり, MSE が大きくなるという問題があった. そこで振幅値の比較的大きい基本周波数成分のみを用いることで推定精度の向上を図った. シミュレーションにより, MSE は概ね 1.0 を下回り, 推定誤差の最大角度は 3.5 度であった. また, 特に基本周波数が高い場合に MSE が大きくなることが示された. 全体的には推定精度が向上しているが, シミュレーションにおける定位方向, 分析対象信号の音色や基本周波数などの条件が限定されているため, 実験条件に限定されない方向推定手法の検討が必要である.



表 2.1 guitar の各音色による MUSIC 法を用いた方向推定結果

tone	$f_0$	match rate	MSE	maximum estimation error(deg)
acoustic guitar	55Hz	20.66%	0.574	1.5
	110Hz	20.66%	1.029	2.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	12.40%	2.161	3.5
clean guitar	55Hz	22.31%	0.496	1.5
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
country guitar	55Hz	20.66%	0.574	1.5
	110Hz	22.31%	0.566	2.0
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
jazz guitar	55Hz	22.31%	0.496	1.5
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
nylon guitar	55Hz	22.31%	0.496	2.0
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	1.5

表 2.2 piano の各音色による MUSIC 法を用いた方向推定結果

tone	$f_0$	match rate	MSE	maximum estimation error(deg)
acoustic piano	55Hz	22.31%	0.496	1.5
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.496	1.5
electric piano	55Hz	22.31%	0.496	1.5
	110Hz	20.66%	0.636	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
grand piano	55Hz	22.31%	0.496	1.5
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
rock piano	55Hz	22.31%	0.496	1.5
	110Hz	22.31%	0.496	1.5
	220Hz	22.31%	0.496	1.5
	440Hz	22.31%	0.587	2.0
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0
stage piano	55Hz	22.31%	0.496	1.5
	110Hz	20.66%	0.574	1.5
	220Hz	22.31%	0.566	2.0
	440Hz	22.31%	0.496	1.5
	880Hz	22.31%	0.496	1.5
	1760Hz	22.31%	0.587	2.0

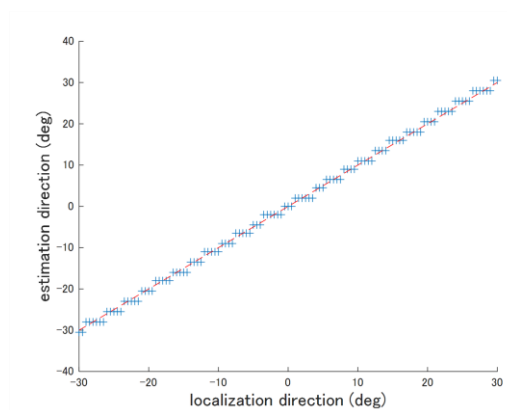


図 2.11  $f_0 = 55$  の acoustic guitar による推定結果 (only MUSIC)

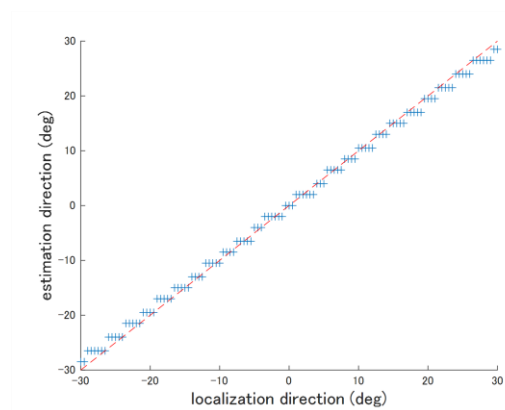


図 2.12  $f_0 = 110$  の acoustic guitar による推定結果 (only MUSIC)

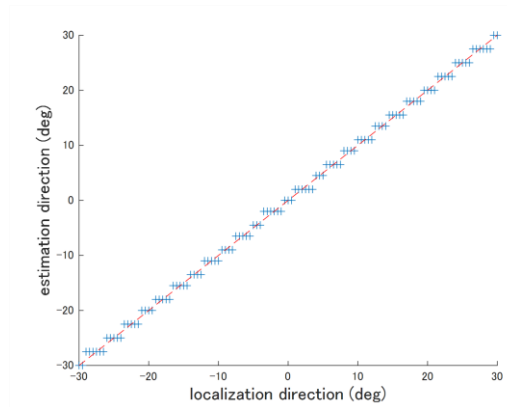


図 2.13  $f_0 = 220$  の acoustic guitar による推定結果 (only MUSIC)

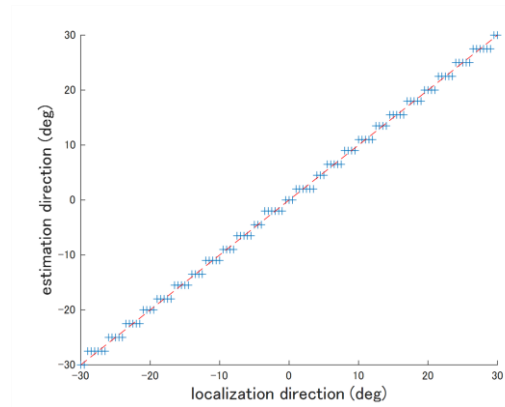


図 2.14  $f_0 = 440$  の acoustic guitar による推定結果 (only MUSIC)

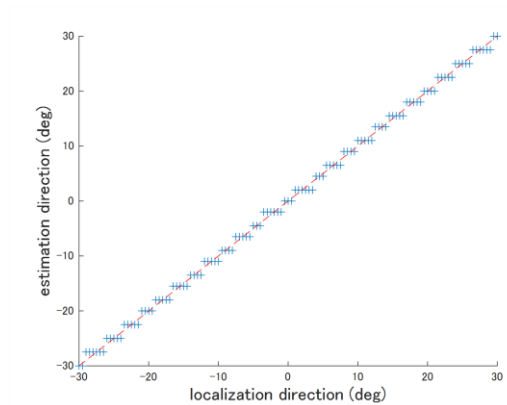


図 2.15  $f_0 = 880$  の acoustic guitar による推定結果 (only MUSIC)

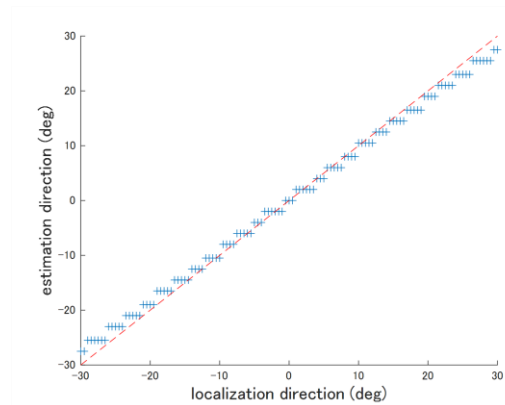


図 2.16  $f_0 = 1760$  の acoustic guitar による推定結果 (only MUSIC)

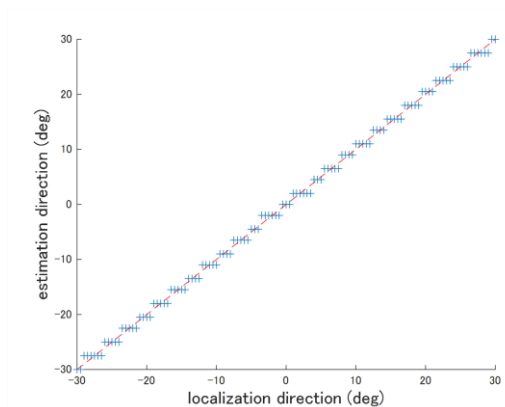


図 2.17  $f_0 = 55$  の clean guitar による推定結果 (only MUSIC)

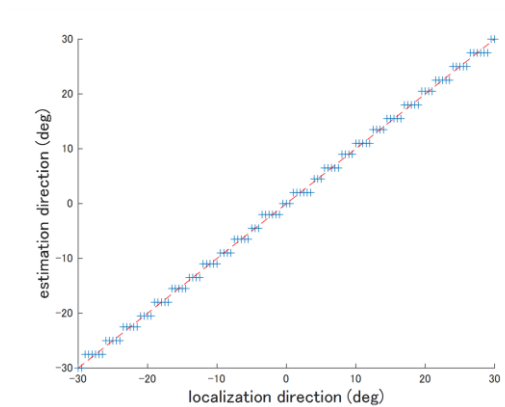


図 2.18  $f_0 = 110$  の clean guitar による推定結果 (only MUSIC)

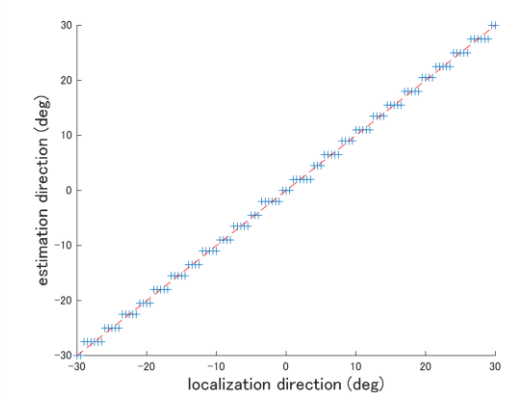


図 2.19  $f_0 = 220$  の clean guitar による推定結果 (only MUSIC)

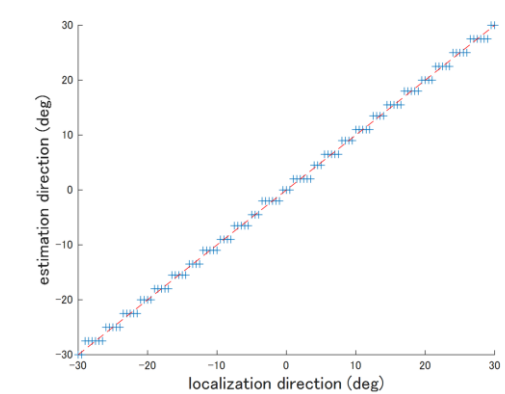


図 2.20  $f_0 = 440$  の clean guitar による推定結果 (only MUSIC)

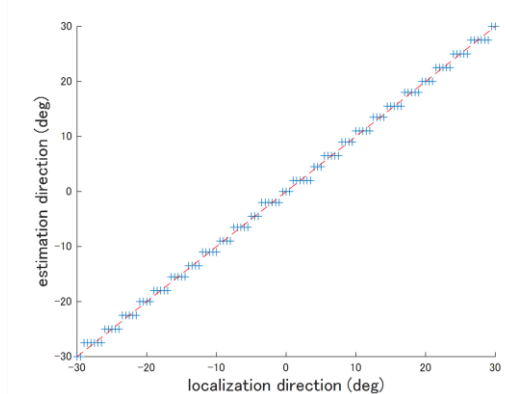


図 2.21  $f_0 = 880$  の clean guitar による推定結果 (only MUSIC)

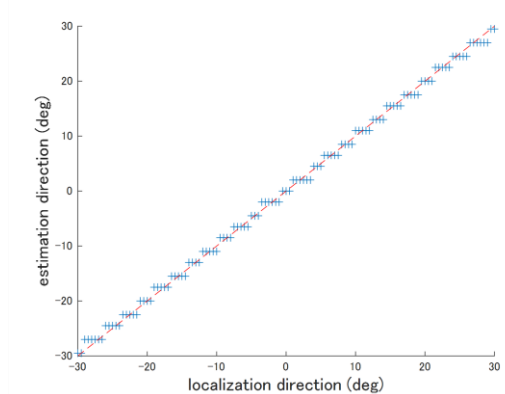


図 2.22  $f_0 = 1760$  の clean guitar による推定結果 (only MUSIC)

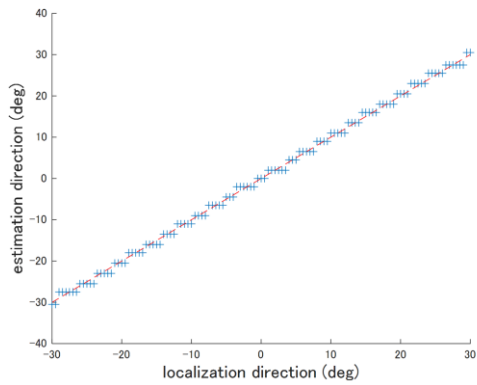


図 2.23  $f_0 = 55$  の country guitar による推定結果 (only MUSIC)

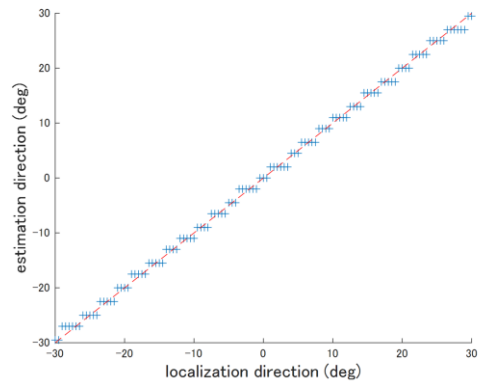


図 2.24  $f_0 = 110$  の country guitar による推定結果 (only MUSIC)

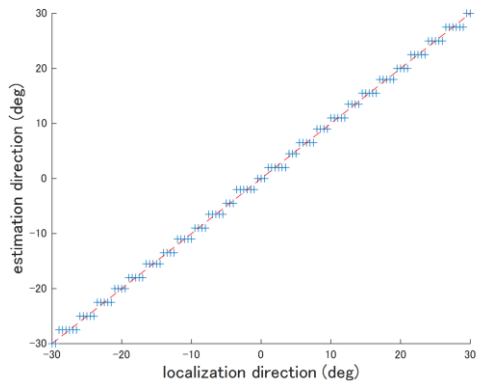


図 2.25  $f_0 = 220$  の country guitar による推定結果 (only MUSIC)

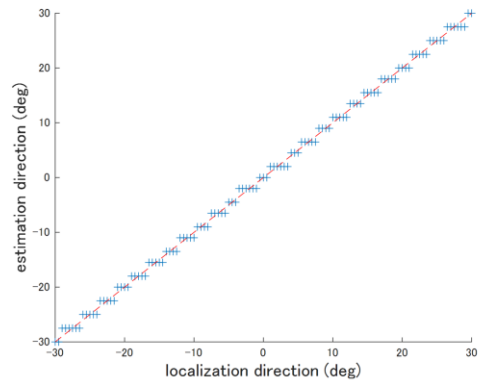


図 2.26  $f_0 = 440$  の country guitar による推定結果 (only MUSIC)

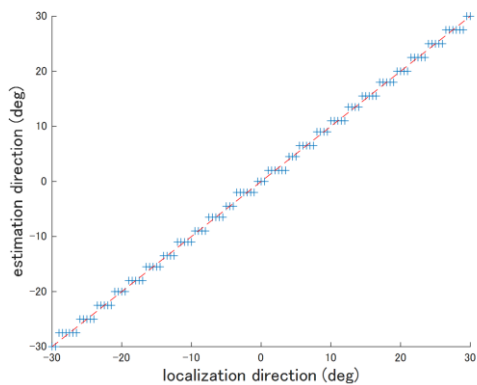


図 2.27  $f_0 = 880$  の country guitar による推定結果 (only MUSIC)

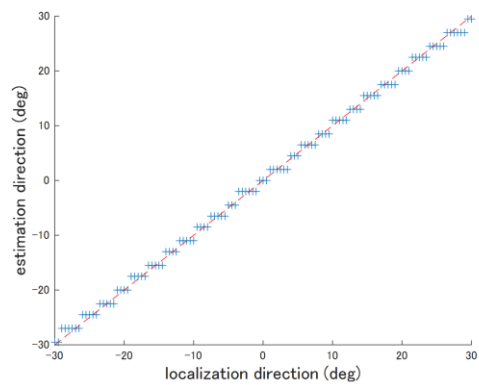


図 2.28  $f_0 = 1760$  の country guitar による推定結果 (only MUSIC)

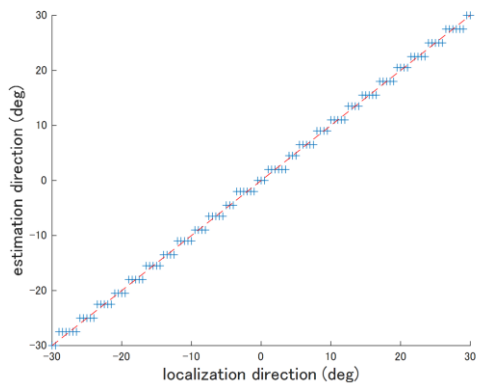


図 2.29  $f_0 = 55$  の jazz guitar による推定結果 (only MUSIC)

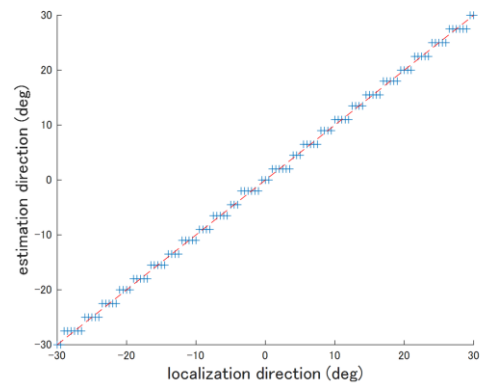


図 2.30  $f_0 = 110$  の jazz guitar による推定結果 (only MUSIC)

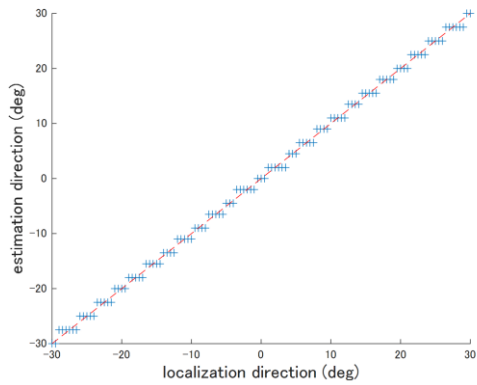


図 2.31  $f_0 = 220$  の jazz guitar による推定結果 (only MUSIC)

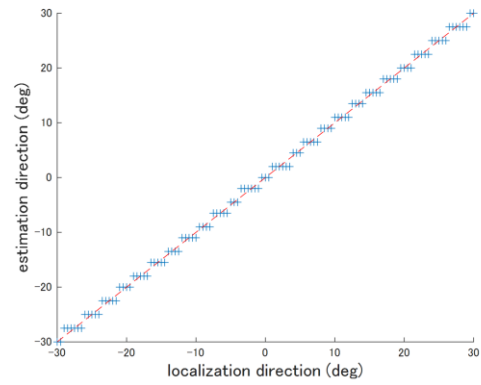


図 2.32  $f_0 = 440$  の jazz guitar による推定結果 (only MUSIC)

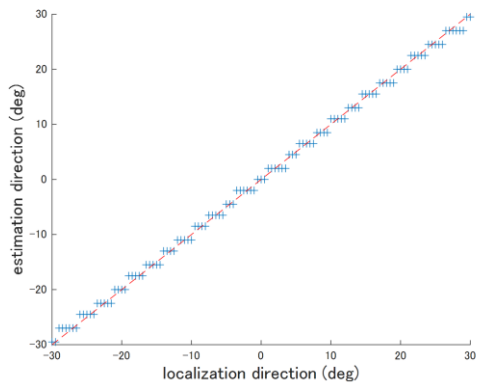


図 2.33  $f_0 = 880$  の jazz guitar による推定結果 (only MUSIC)

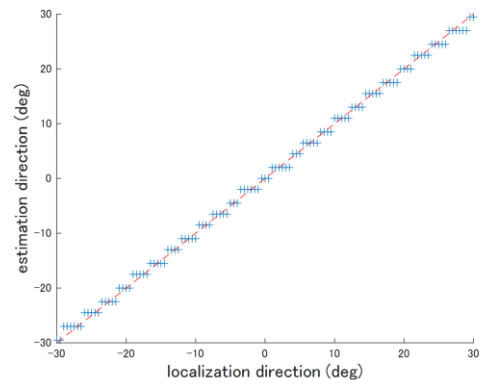


図 2.34  $f_0 = 1760$  の jazz guitar による推定結果 (only MUSIC)

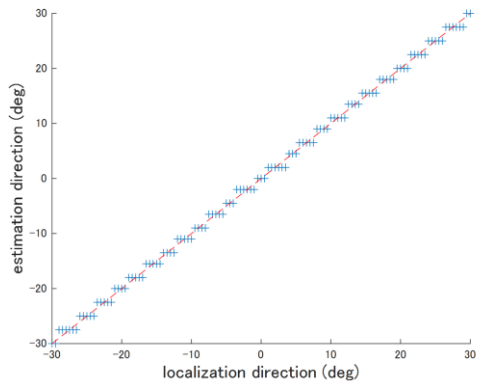


図 2.35  $f_0 = 55$  の nylon guitar による推定結果 (only MUSIC)

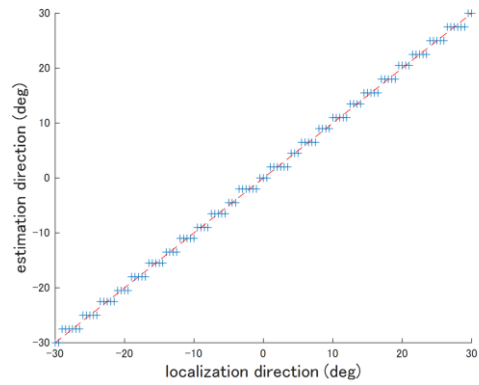


図 2.36  $f_0 = 110$  の nylon guitar による推定結果 (only MUSIC)

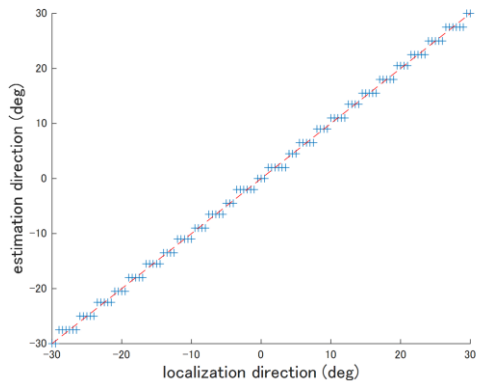


図 2.37  $f_0 = 220$  の nylon guitar による推定結果 (only MUSIC)

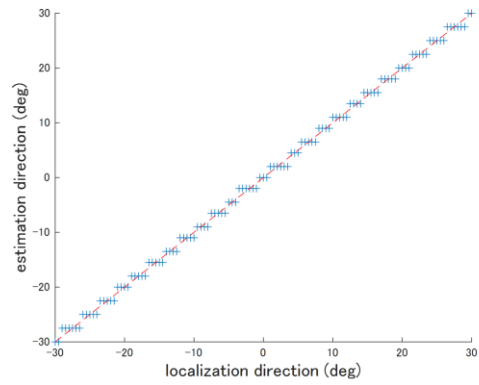


図 2.38  $f_0 = 440$  の nylon guitar による推定結果 (only MUSIC)

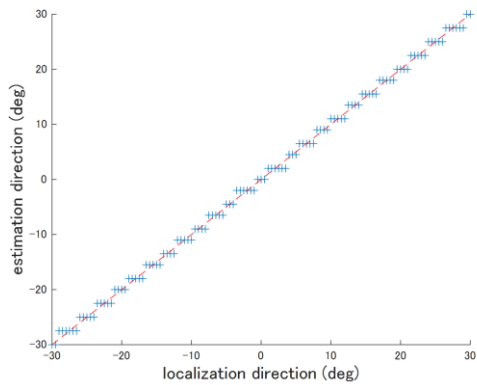


図 2.39  $f_0 = 880$  の nylon guitar による推定結果 (only MUSIC)

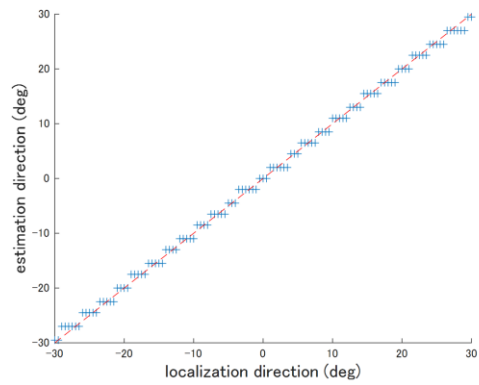


図 2.40  $f_0 = 1760$  の nylon guitar による推定結果 (only MUSIC)

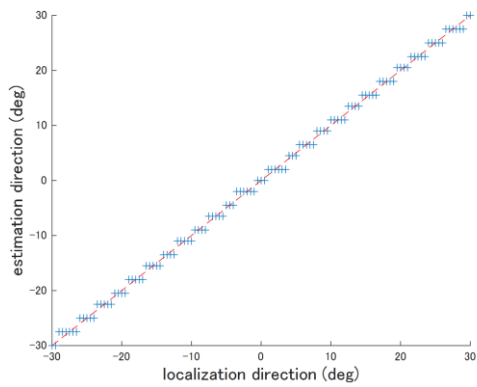


図 2.41  $f_0 = 55$  の acoustic piano による推定結果 (only MUSIC)

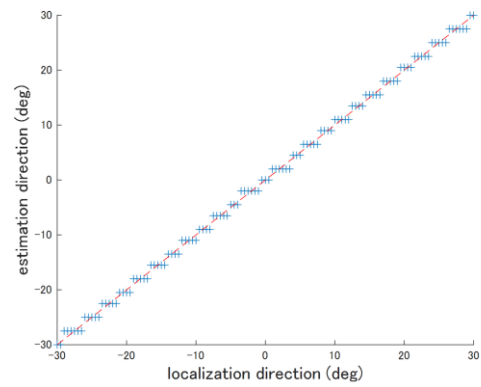


図 2.42  $f_0 = 110$  の acoustic piano による推定結果 (only MUSIC)

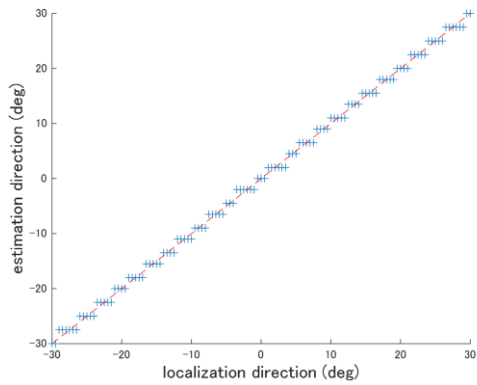


図 2.43  $f_0 = 220$  の acoustic piano による推定結果 (only MUSIC)

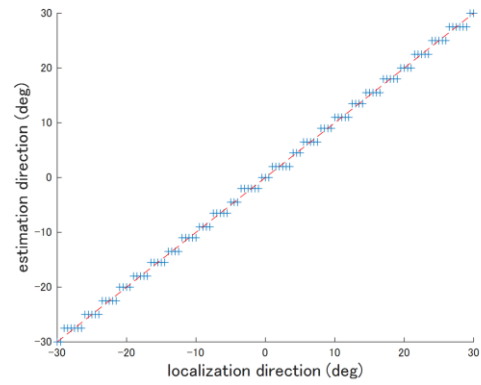


図 2.44  $f_0 = 440$  の acoustic piano による推定結果 (only MUSIC)

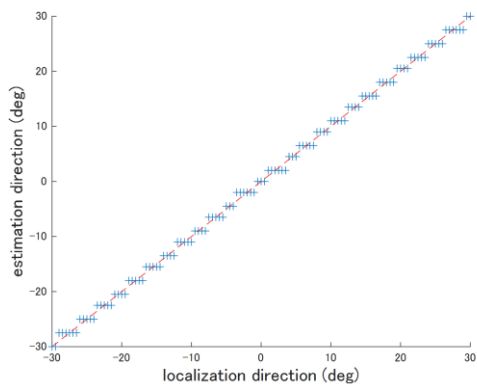


図 2.45  $f_0 = 880$  の acoustic piano による推定結果 (only MUSIC)

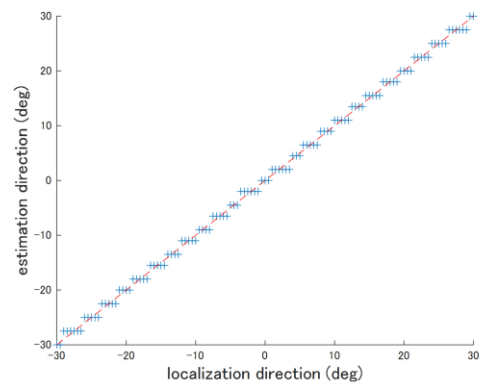


図 2.46  $f_0 = 1760$  の acoustic piano による推定結果 (only MUSIC)



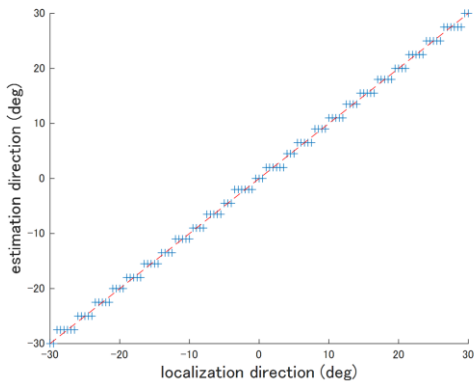


図 2.47  $f_0 = 55$  の electric piano による推定結果 (only MUSIC)

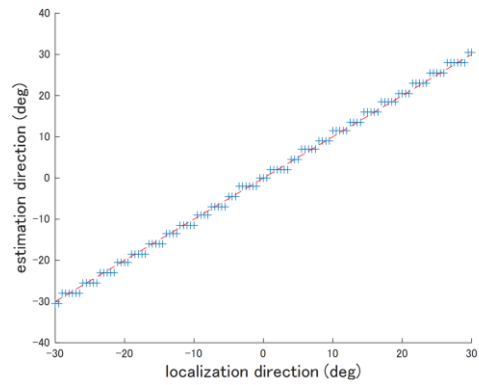


図 2.48  $f_0 = 110$  の electric piano による推定結果 (only MUSIC)

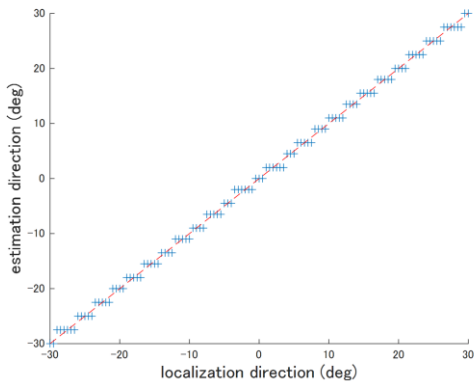


図 2.49  $f_0 = 220$  の electric piano による推定結果 (only MUSIC)

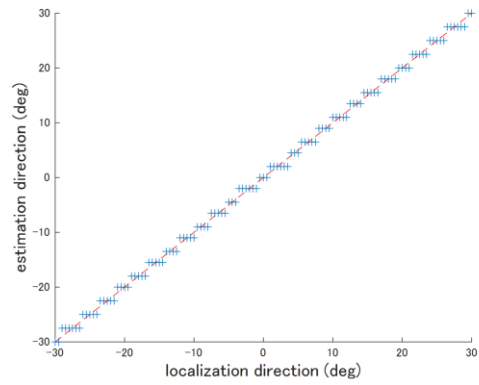


図 2.50  $f_0 = 440$  の electric piano による推定結果 (only MUSIC)

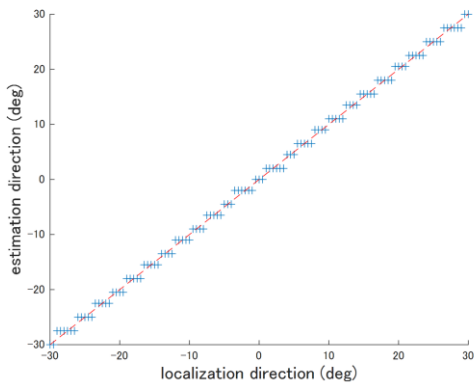


図 2.51  $f_0 = 880$  の electric piano による推定結果 (only MUSIC)

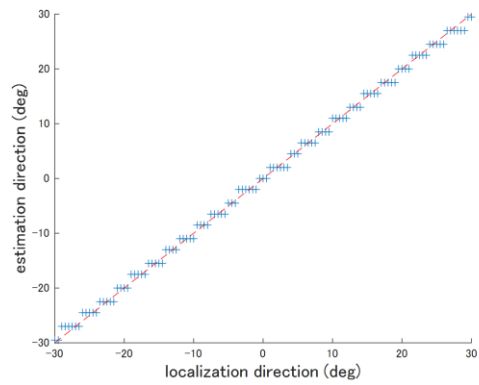


図 2.52  $f_0 = 1760$  の electric piano による推定結果 (only MUSIC)

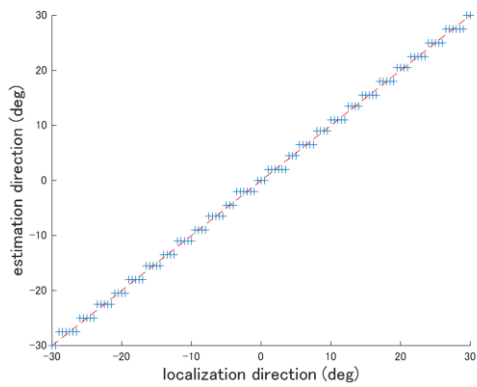


図 2.53  $f_0 = 55$  の grand piano による推定結果 (only MUSIC)

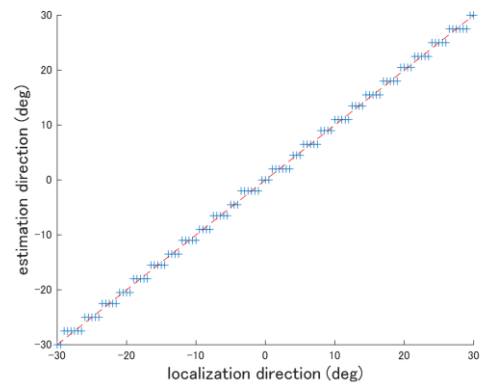


図 2.54  $f_0 = 110$  の grand piano による推定結果 (only MUSIC)

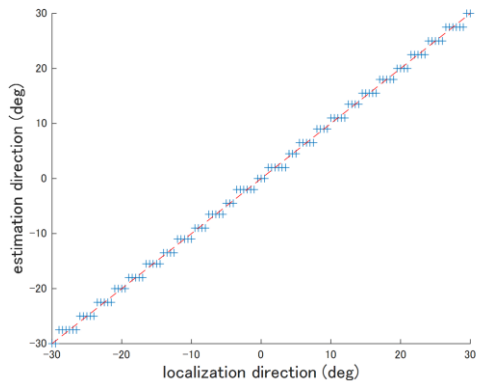


図 2.55  $f_0 = 220$  の grand piano による推定結果 (only MUSIC)

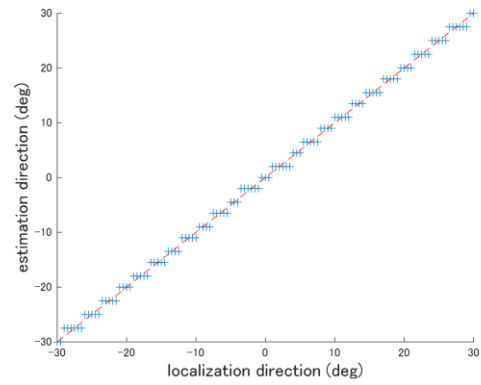


図 2.56  $f_0 = 440$  の grand piano による推定結果 (only MUSIC)

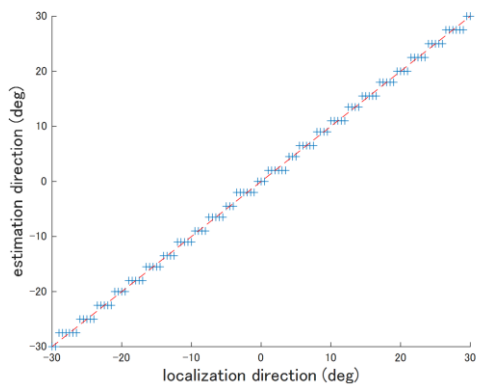


図 2.57  $f_0 = 880$  の grand piano による推定結果 (only MUSIC)

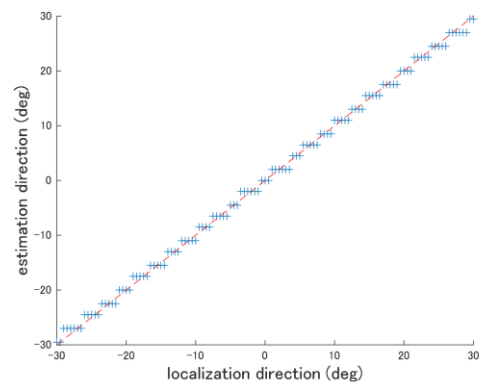


図 2.58  $f_0 = 1760$  の grand piano による推定結果 (only MUSIC)

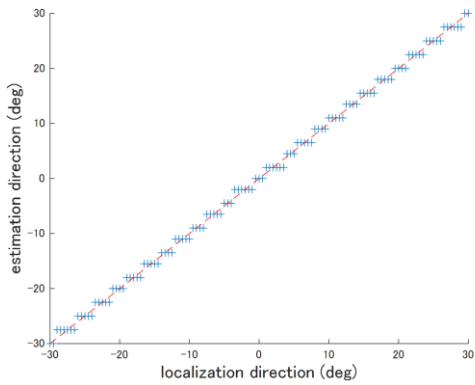


図 2.59  $f_0 = 55$  の rock piano による推定結果 (only MUSIC)

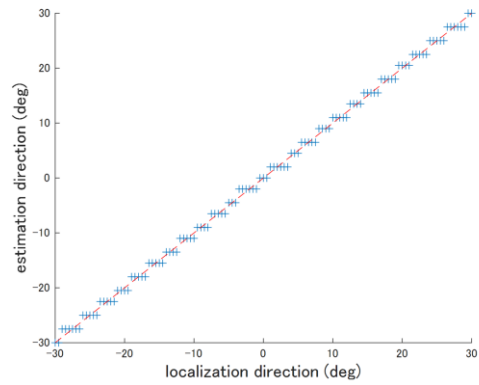


図 2.60  $f_0 = 110$  の rock piano による推定結果 (only MUSIC)

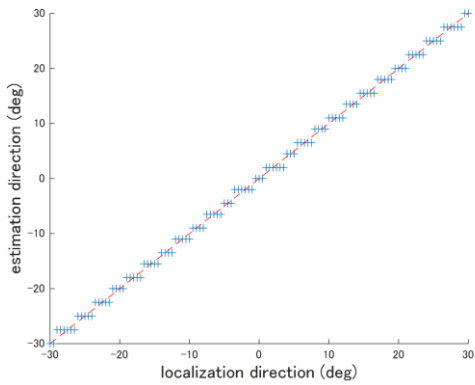


図 2.61  $f_0 = 220$  の rock piano による推定結果 (only MUSIC)

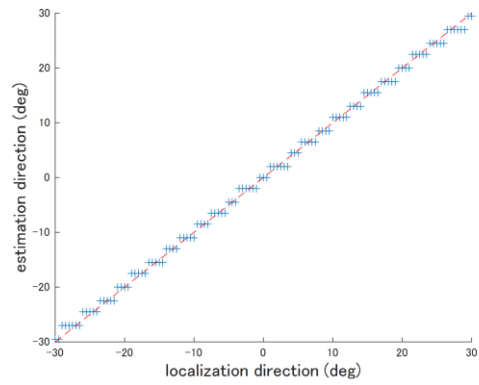


図 2.62  $f_0 = 440$  の rock piano による推定結果 (only MUSIC)

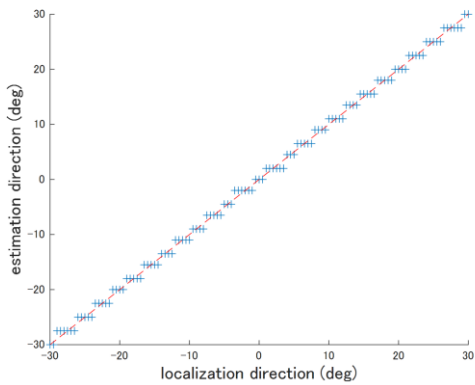


図 2.63  $f_0 = 880$  の rock piano による推定結果 (only MUSIC)

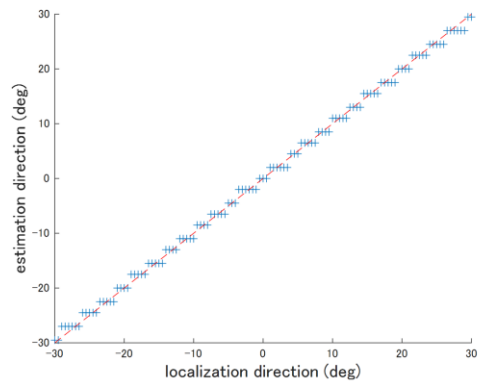


図 2.64  $f_0 = 1760$  の rock piano による推定結果 (only MUSIC)

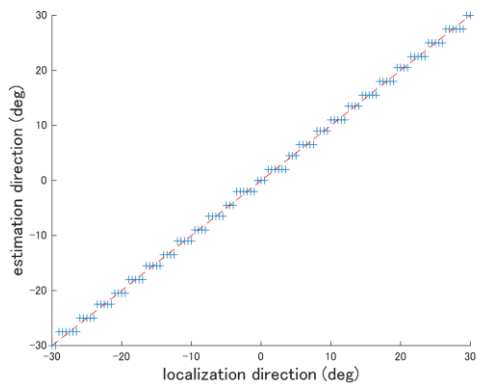


図 2.65  $f_0 = 55$  の stage piano による推定結果 (only MUSIC)

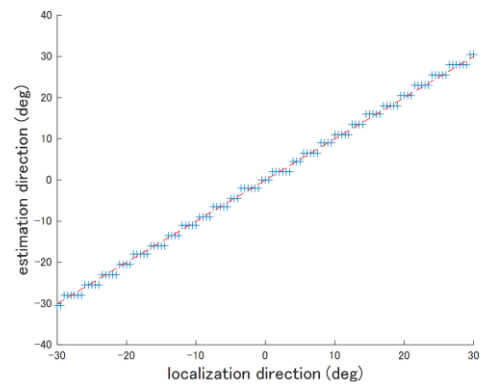


図 2.66  $f_0 = 110$  の stage piano による推定結果 (only MUSIC)

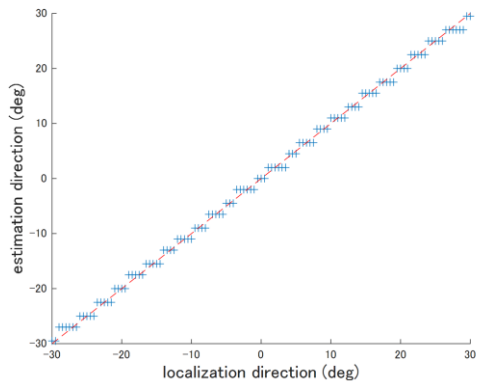


図 2.67  $f_0 = 220$  の stage piano による推定結果 (only MUSIC)

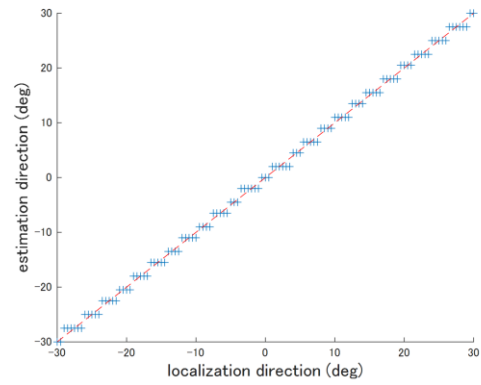


図 2.68  $f_0 = 440$  の stage piano による推定結果 (only MUSIC)

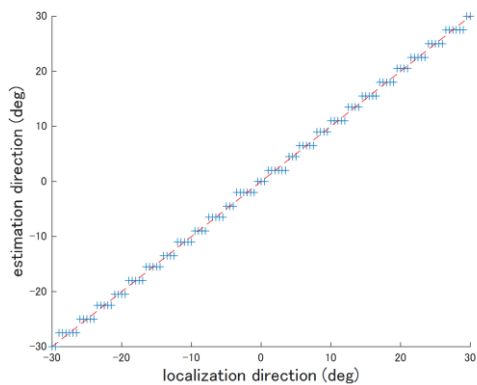


図 2.69  $f_0 = 880$  の stage piano による推定結果 (only MUSIC)

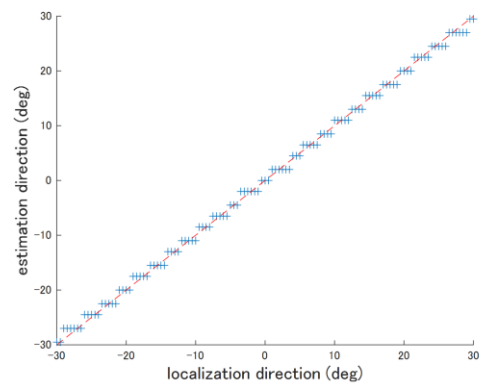


図 2.70  $f_0 = 1760$  の stage piano による推定結果 (only MUSIC)

## 第3章 高調波成分を用いた音源方向推定

### 3.1 序言

楽器音の高調波成分に着目したニューラルネットワークを用いた音源方向推定手法[30]について述べる。本研究で対象としている楽器音源は、基本周波数の整数倍の周波数にピークのある倍音構造を持つ。これらの成分はほかの周波数成分よりも振幅値が大きく、MUSIC法を用いた方向推定において推定誤差は小さくなる。そこで基本波・高調波成分を用いたMUSIC法による推定結果を基にニューラルネットワークを用いて推定精度の向上を図る。機械学習の一種であるニューラルネットワーク[31, 32]は生物の情報処理のメカニズムを計算機構として実現したもので、情報伝達を担う神経細胞であるニューロンの情報伝達の仕組みを参考としたモデルである。本研究では既知の入力データとその入力データに対する出力データ（教師データ）を用いてニューラルネットワークを学習し、未知の入力データに対して出力を推定する教師ありモデルを用いる。その中でも最も基本的なニューラルネットワークであるフィードフォワードニューラルネットワーク (FFNN : Feedforward Neural Network)を用いて、楽器や周波数ごとの誤りパターンを学習させることで推定精度の向上を図る。

本章の構成は次の通りである。3.2節では基本波・高調波成分を用いたMUSIC法による推定結果を用いたFFNNによる方向推定について述べる。3.3節ではFFNNによる方向推定シミュレーションの実験についての結果と考察について述べ、3.4節で結言とする。

### 3.2 FFNNを用いた音源方向推定

FFNNは最も単純なニューラルネットワークである。本研究で対象とする楽器音源は倍音構造を持つ。倍音構造を持つ楽器音において、図3.1や図3.2に示すように基本周波数成分や高調波成分は他の周波数成分よりも振幅値が大きい。基本周波数成分のみを用いた音源方向推定においては2章で述べたようにMSEは小さくなる。guitarやpianoにおける高調波成分は基本周波数成分よりも振幅が小さいため、高調波成分を用いるMUSIC法による方向推定では、図3.3~3.5に示すように基本周波数成分を用いた場合よりも推定精度は低くなる。しかし、基本波・高調波成分よりも振幅がより小さい成分を用いた方向推定においては図3.6~3.8のように推定精度が低下するため、高調波成分による方向推定では一定の推定精度をもつと言える。基本周波数のみを用いたMUSIC法による方向推定においては推定誤差の生じるサンプルが多いが、基本周波数成分や高調波成分を用いた方向推定において一定の推定精度をもつサンプルも多い。算出された値が定位方向であるとは限らないが、基本周波数成分や高調

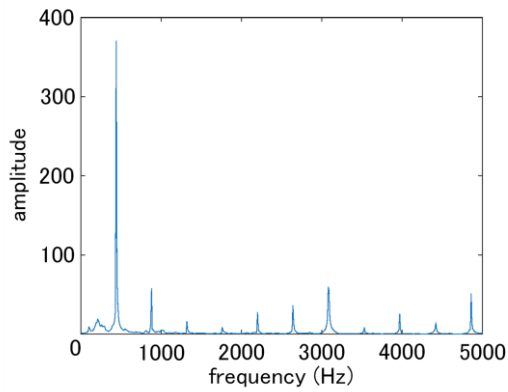


図 3.1  $f_0 = 440\text{Hz}$  である clean guitar の振幅スペクトル (図 2.2 再掲)

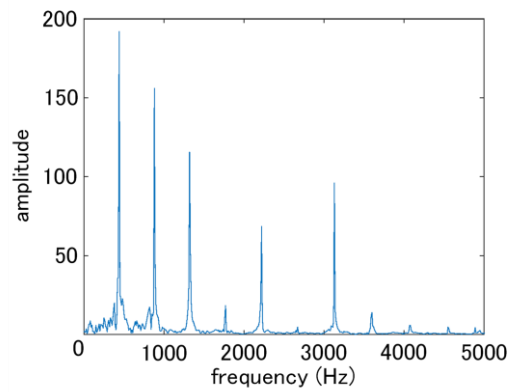


図 3.2  $f_0 = 440\text{Hz}$  である grand piano の振幅スペクトル (図 2.3 再掲)

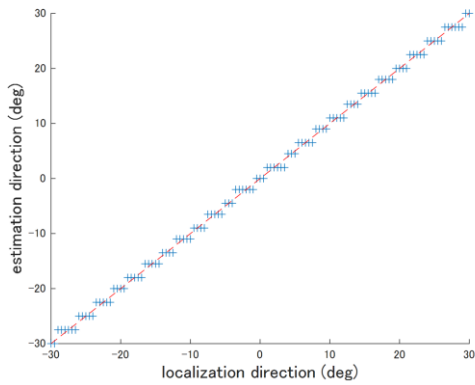


図 3.3  $f_0 = 880\text{Hz}$  である clean guitar の基本周波数成分のみを用いた音源方向推定の結果

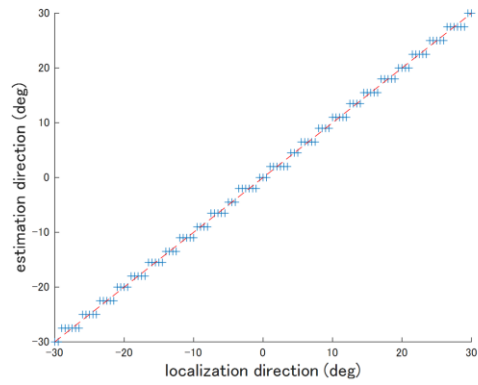


図 3.4  $f_0 = 880\text{Hz}$  である clean guitar の第 2 高調波成分のみを用いた音源方向推定の結果

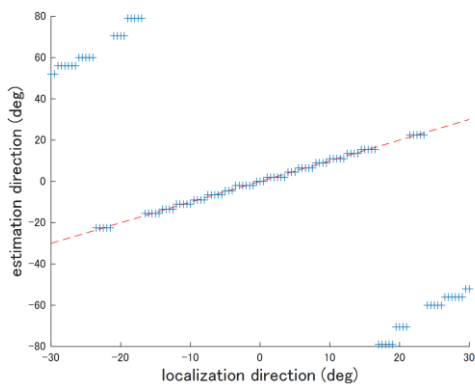


図 3.5  $f_0 = 880\text{Hz}$  である clean guitar の第 3 高調波成分のみを用いた音源方向推定の結果

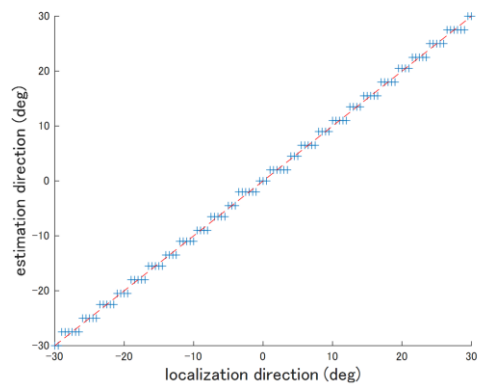


図 3.6  $f_0 = 55\text{Hz}$  である clean guitar の基本周波数成分のみを用いた音源方向推定の結果

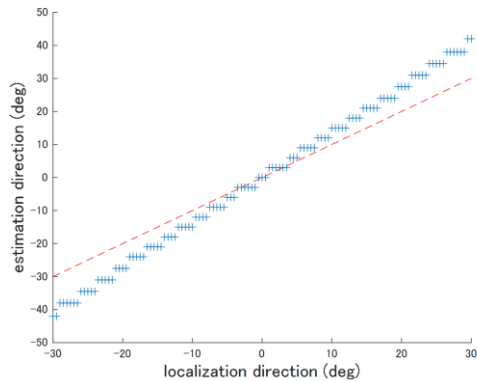


図 3.7  $f_0 = 55\text{Hz}$  である clean guitar の基本周波数の 1.5 倍成分のみを用いた音源方向推定の結果

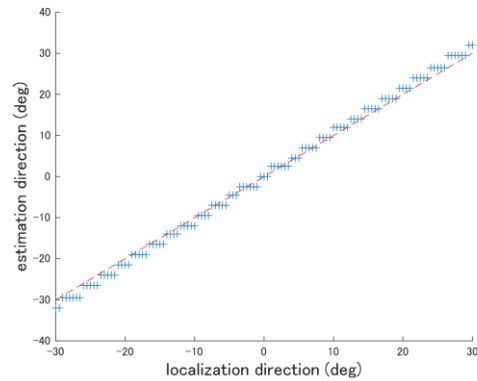


図 3.8  $f_0 = 55\text{Hz}$  である clean guitar の基本周波数の 2.5 倍成分のみを用いた音源方向推定の結果

波成分を用いた推定結果に定位方向が含まれていなくても誤りパターンに一定の傾向があるならば，この推定誤りパターンを学習することで定位方向を推定できると考えた．したがって本研究では，基本波・第 2, 3 高調波成分を用いた MUSIC 法による推定結果をニューラルネットワークに入力することで，楽器や周波数ごとに基本波・第 2, 3 高調波成分を用いた MUSIC 法による推定結果の誤りパターンに対する定位方向を学習させ，推定精度向上を図る．

本研究で使用するニューラルネットワークは，入力層・中間層・出力層がいずれも 1 層の 3 層構造の FFNN である．入力層のニューロンは 3 個とし，入力には MUSIC 法において  $f = f_0, f_1, f_2$  ( $f_1$ : 第 2 高調波,  $f_2$ : 第 3 高調波) として得られた 3 つの推定結果を使用する．ニューラルネットワークは図 3.9 のような中間層・出力層のニューロンがそれぞれ 10 個・1 個の場合 (FFNN-AO (analog output)) と図 3.10 のような中間層・出力層がそれぞれ 10 個・121 個の場合 (FFNN-DO (digital output)) の 2 通りとする．学習は guitar と piano それぞれについて，121 方向・4 種類の音色・6 種類の  $f_0$  (データ数:  $121 \times 4 \times 6$ ) の MUSIC 法 ( $f = f_0, f_1, f_2$ ) による推定結果の 70% を使用し，音色・周波数ごと (データ数: 121) で評価を行なう．

出力ニューロンが 1 個であるアナログ出力の FFNN-AO において，学習アルゴリズムには非線形最小二乗法の一つでニュートン法と最急降下法を組み合わせた手法である Levenberg-Marquardt 法[33]，検証には MSE，中間層にはシグモイド伝達関数，出力層には線形伝達関数を用いる．教師データは入力データに対する定位方向である．また，出力ニューロンが 121 個であるデジタル出力の FFNN-DO (パターン認識的アプローチ) において，学習アルゴリズムにはスケール共役勾配法，検証には予測された確率の対数と正解であるラベル(0 または 1) の積和である交差エントロピー，中間層にはシグモイド伝達関数，出力層にはソフトマックス伝達関数を用いる．教師データは入力データに対する定位方向に対応する要素のみが 1 となる  $121 \times 1$  の One-

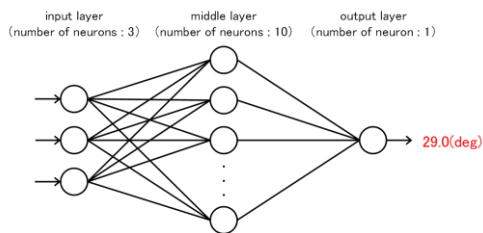


図 3.9 FFNN-AO の概略図 (推定結果が 29.0 度の場合)

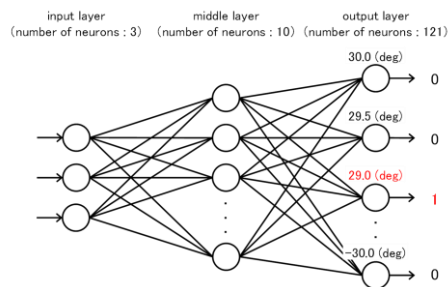


図 3.10 FFNN-DO の概略図 (推定結果が 29.0 度の場合)

hot ベクトルである。検証データには 121 方向・4 種類の音色・6 種類の  $f_0$  (データ数:  $121 \times 4 \times 6$ ) による推定結果の 15% を使用し, MSE が最小となったネットワークを最終的な学習済みネットワークとする。

### 3.3 実験

本節では, FFNN を用いた方向推定のシミュレーションについて述べる。3.3.1 項ではシミュレーションの条件について, 3.3.2 項ではシミュレーションの結果について, 3.3.3 項では MUSIC 法によるシミュレーション結果との比較について述べる。

#### 3.3.1 実験条件

FFNN を用いることで推定精度が改善されたかを確認するために, MUSIC 法のみによる方向推定精度と FFNN を用いた場合とで比較する。方向推定に用いる音源は第 2 章で述べた guitar 5 種 (acoustic guitar, clean guitar, country guitar, jazz guitar, nylon guitar), piano 5 種 (acoustic piano, electric piano, grand piano, rock piano, stage piano) の計 10 種類 (標本化周波数 = 44100Hz, 量子化ビット数 = 16bits, 基本周波数  $f_0 = 55, 110, 220, 440, 880, 1760$ Hz) で, マイクロフォンの本数  $K = 3$ , マイクロフォン距離  $d = 0.1$ m として単一音源における音源方向推定のシミュレーションを行なう。音源定位の方向は図 3.11 に示すように  $M_2$  正面から左右 30 度の範囲の単一音源 (0.5 度間隔の全 121 パターン) に制限する。それぞれの信号において基本波  $f_0$ , 第 2 高調波  $f_1 = 2f_0$ , 第 3 高調波  $f_2 = 3f_0$  の成分を用いて MUSIC 法により算出された推定結果を使用する。なお, 第 2 章と同様に分析フレーム長は 44100 samples, STFT におけるフレーム長は 4096 samples, 移動長は 24 samples とする。学習・検証に用いるデータは固定し, FFNN-AO による推定結果の評価は MSE・推定誤差, FFNN-DO による推定結果の評価は MSE・完全一致率・推定誤差, MUSIC 法との比較は MSE で行なう。



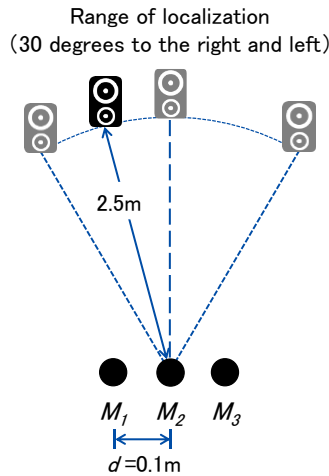


図 3.11 収音シミュレーションの概略図 1 (定位方向が $M_2$ 正面から左 15 度 (-15 度) にある場合, 図 2.10 再掲)

### 3.3.2 実験結果

アナログ出力のニューラルネットワークによる推定結果を表 3.1, 3.2 及び図 3.12~3.71 に示す. FFNN-AO を用いた場合, guitar による方向推定において最も MSE が大きくなったのは acoustic guitar ( $f_0 = 110\text{Hz}$ ) の場合で 0.995 である. 推定誤差は最大で 2.86 度 (acoustic guitar,  $f_0 = 1760\text{Hz}$ ) 生じた. piano による方向推定において最も MSE が大きくなったのは electric piano ( $f_0 = 110\text{Hz}$ ) の場合で 0.617, 推定誤差は 1.93 度 (rock piano  $f_0 = 440\text{Hz}$ , stage piano,  $f_0 = 220\text{Hz}$ ) で最大となった. FFNN-AO による全ての推定結果について MSE は 1.0 を下回り, その多くが約 0.5 となった.

デジタル出力のニューラルネットワークによる推定結果は表 3.3, 3.4 及び図 3.72~3.131 のとおりである. FFNN-DO を用いた場合, guitar による方向推定において最も MSE が大きくなったのは acoustic guitar ( $f_0 = 1760\text{Hz}$ ) の場合で 1.494 である. 推定誤差は最大で 3.5 度 (acoustic guitar,  $f_0 = 1760\text{Hz}$ ) 生じた piano による方向推定において最も MSE が大きくなったのは grand piano ( $f_0 = 880\text{Hz}$ ) と rock piano ( $f_0 = 880\text{Hz}$ ) との場合で 1.066, 推定誤差は 2.5 度で最大となった. FFNN-DO による全ての推定結果について MSE は約 1.0, 完全一致率もほぼすべてで 22.31%となった.

表 3.1 5 種類の guitar 音源による MUSIC 法を用いた方向推定結果の 70% を使用して学習させた FFNN-AO を用いて guitar の各音色・各周波数成分による推定結果で評価した MSE と推定誤差の最大角度

tone	$f_0$	MSE	maximum estimation error(deg)
acoustic guitar	55Hz	0.592	1.68
	110Hz	0.995	2.42
	220Hz	0.491	1.48
	440Hz	0.491	1.48
	880Hz	0.503	1.48
	1760Hz	0.905	2.86
clean guitar	55Hz	0.493	1.48
	110Hz	0.493	1.48
	220Hz	0.493	1.48
	440Hz	0.491	1.48
	880Hz	0.503	1.48
	1760Hz	0.540	1.77
country guitar	55Hz	0.579	1.66
	110Hz	0.560	1.92
	220Hz	0.491	1.48
	440Hz	0.491	1.48
	880Hz	0.503	1.48
	1760Hz	0.540	1.77
jazz guitar	55Hz	0.493	1.48
	110Hz	0.491	1.48
	220Hz	0.491	1.48
	440Hz	0.491	1.48
	880Hz	0.503	1.48
	1760Hz	0.540	1.77
nylon guitar	55Hz	0.491	1.48
	110Hz	0.493	1.48
	220Hz	0.493	1.48
	440Hz	0.491	1.48
	880Hz	0.503	1.48
	1760Hz	0.540	1.77

表 3.2 5 種類の piano 音源による MUSIC 法を用いた方向推定結果の 70% を使用して学習させた FFNN-AO を用いて guitar の各音色・各周波数成分による推定結果で評価した MSE と推定誤差の最大角度

tone	$f_0$	MSE	maximum estimation error(deg)
acoustic piano	55Hz	0.489	1.49
	110Hz	0.489	1.49
	220Hz	0.491	1.49
	440Hz	0.491	1.49
	880Hz	0.498	1.49
	1760Hz	0.530	1.71
electric piano	55Hz	0.491	1.49
	110Hz	0.617	1.65
	220Hz	0.491	1.49
	440Hz	0.491	1.49
	880Hz	0.498	1.49
	1760Hz	0.511	1.66
grand piano	55Hz	0.489	1.49
	110Hz	0.489	1.49
	220Hz	0.491	1.49
	440Hz	0.491	1.49
	880Hz	0.503	1.49
	1760Hz	0.511	1.66
rock piano	55Hz	0.491	1.49
	110Hz	0.489	1.49
	220Hz	0.491	1.49
	440Hz	0.585	1.93
	880Hz	0.503	1.49
	1760Hz	0.511	1.66
stage piano	55Hz	0.489	1.49
	110Hz	0.571	1.65
	220Hz	0.570	1.93
	440Hz	0.489	1.49
	880Hz	0.498	1.49
	1760Hz	0.511	1.66

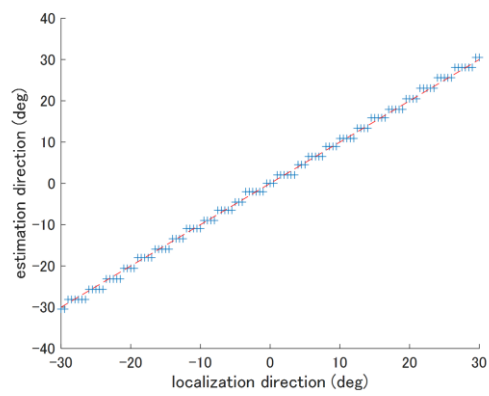


図 3.12  $f_0 = 55$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

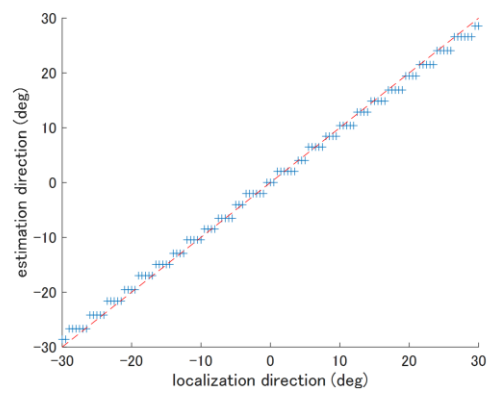


図 3.13  $f_0 = 110$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

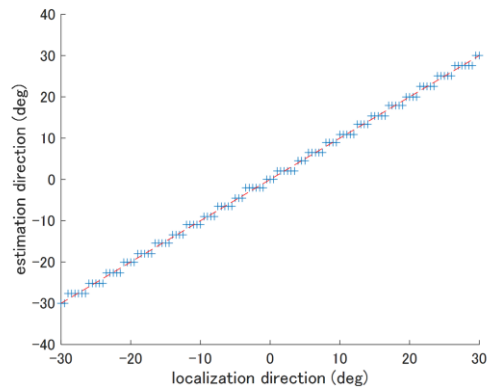


図 3.14  $f_0 = 220$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

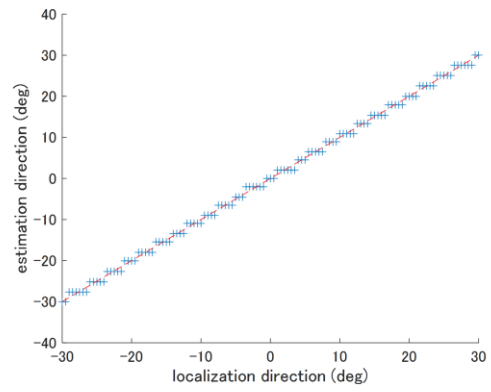


図 3.15  $f_0 = 440$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

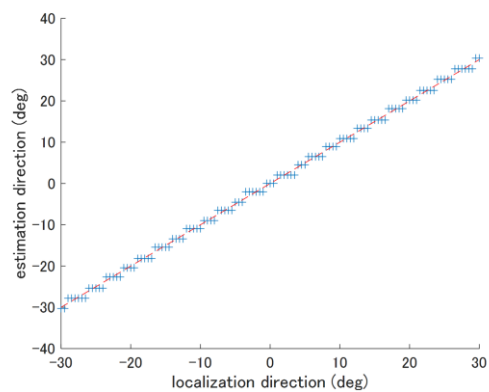


図 3.16  $f_0 = 880$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

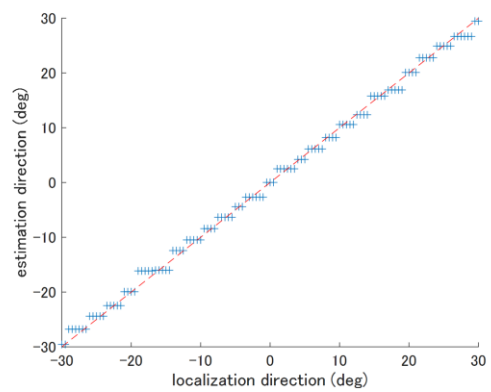


図 3.17  $f_0 = 1760$  の acoustic guitar による推定結果 (MUSIC + FFNN-AO)

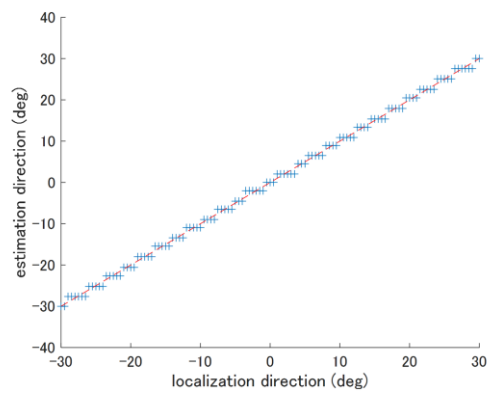


図 3.18  $f_0 = 55$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

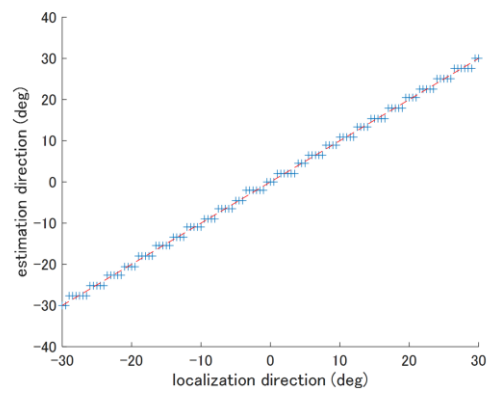


図 3.19  $f_0 = 110$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

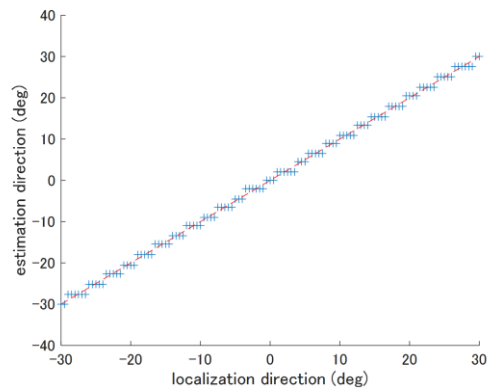


図 3.20  $f_0 = 220$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

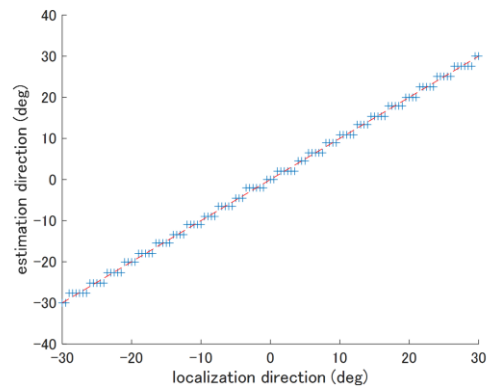


図 3.21  $f_0 = 440$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

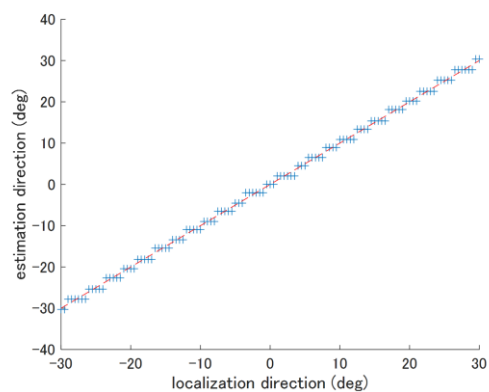


図 3.22  $f_0 = 880$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

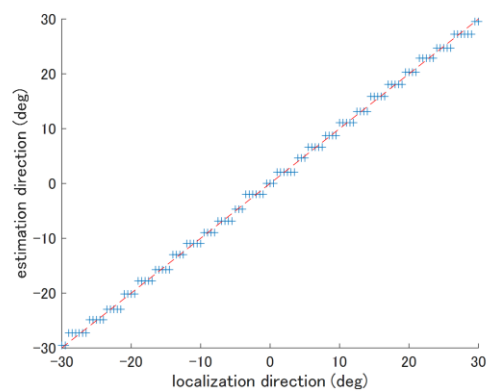


図 3.23  $f_0 = 1760$  の clean guitar による推定結果 (MUSIC + FFNN-AO)

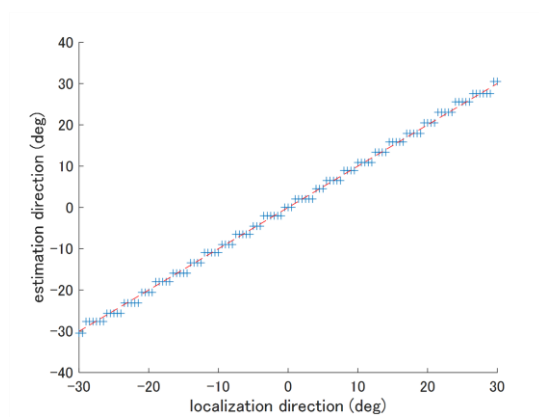


図 3.24  $f_0 = 55$  の country guitar による推定結果 (MUSIC + FFNN-AO)

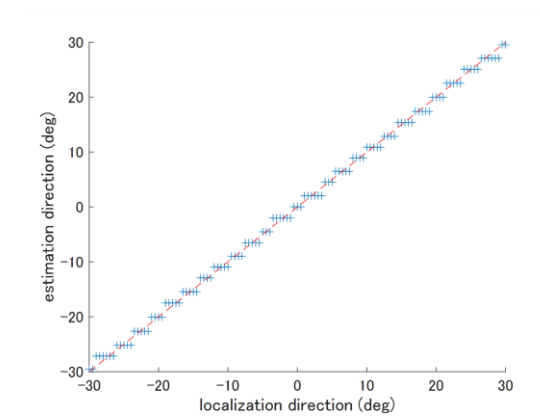


図 3.25  $f_0 = 110$  の country guitar による推定結果 (MUSIC + FFNN-AO)

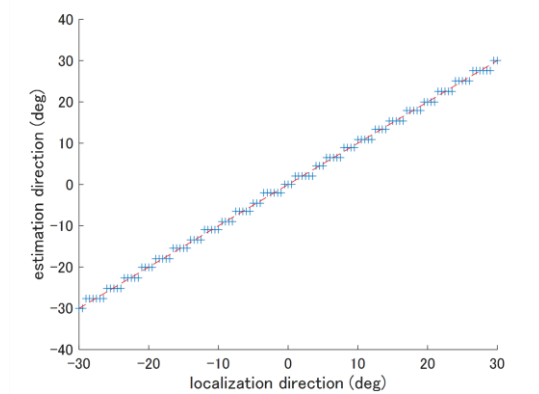


図 3.26  $f_0 = 220$  の country guitar による推定結果 (MUSIC + FFNN-AO)

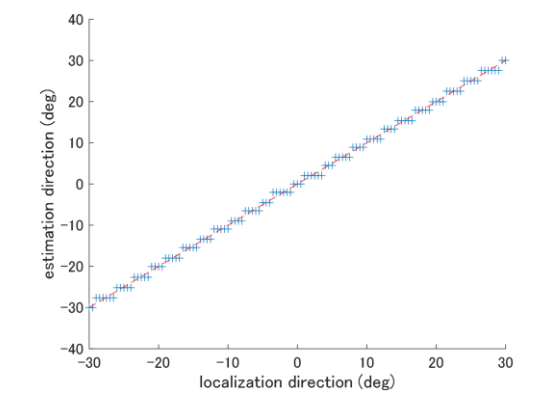


図 3.27  $f_0 = 440$  の country guitar による推定結果 (MUSIC + FFNN-AO)

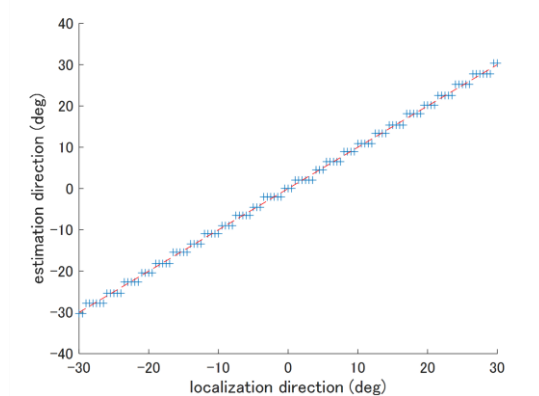


図 3.28  $f_0 = 880$  の country guitar による推定結果 (MUSIC + FFNN-AO)

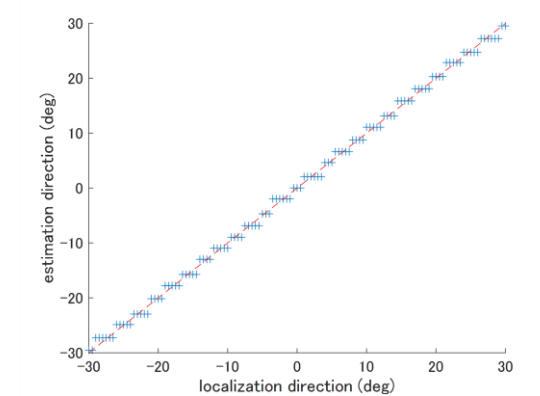


図 3.29  $f_0 = 1760$  の country guitar による推定結果 (MUSIC + FFNN-AO)

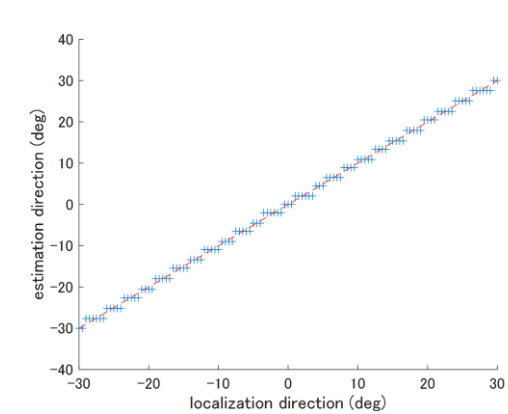


図 3.30  $f_0 = 55$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

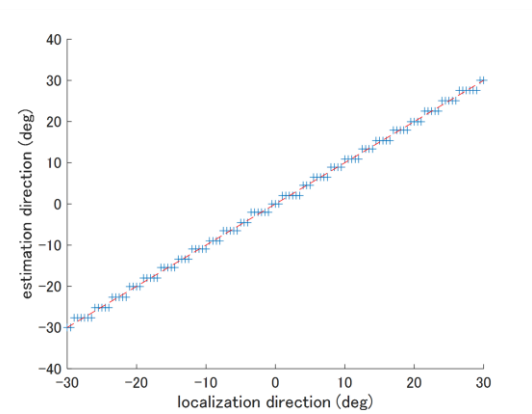


図 3.31  $f_0 = 110$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

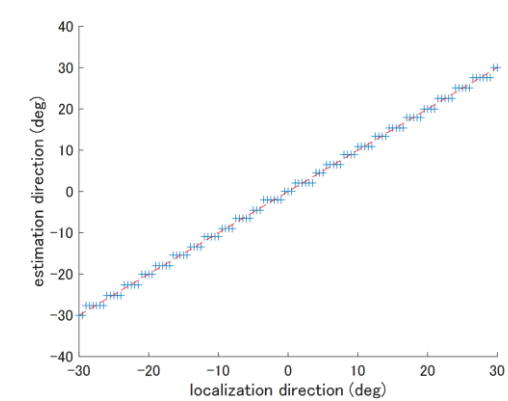


図 3.32  $f_0 = 220$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

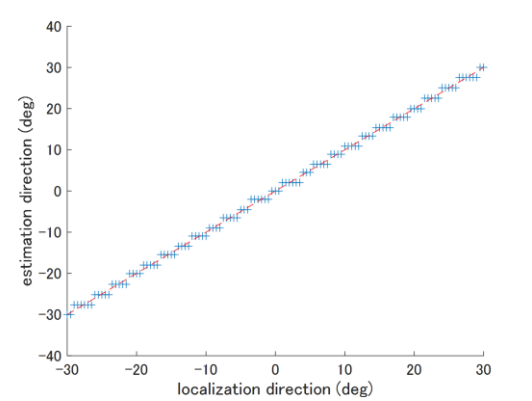


図 3.33  $f_0 = 440$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

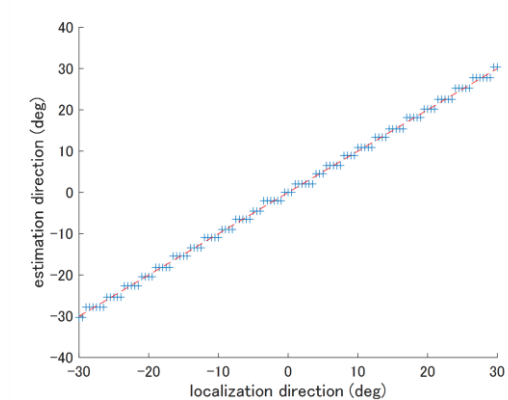


図 3.34  $f_0 = 880$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

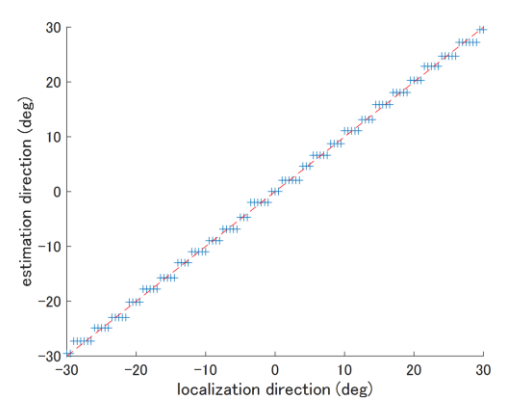


図 3.35  $f_0 = 1760$  の jazz guitar による推定結果 (MUSIC + FFNN-AO)

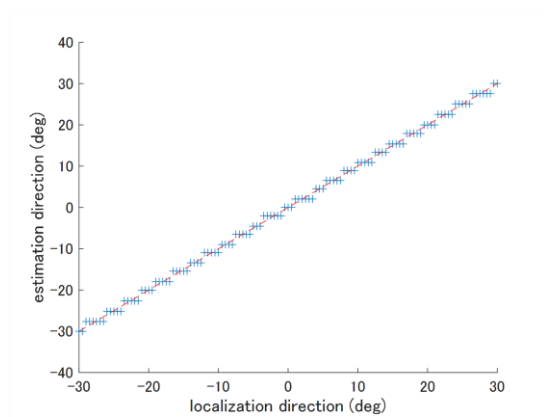


図 3.36  $f_0 = 55$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)

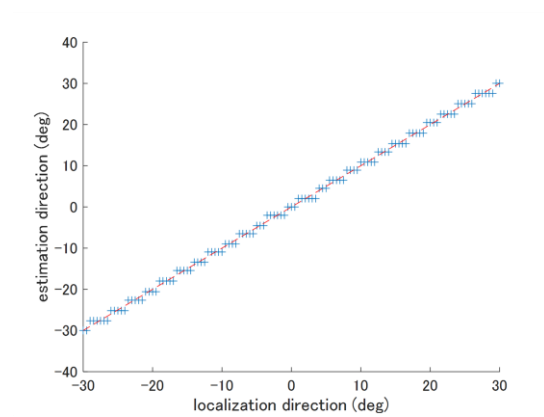


図 3.37  $f_0 = 110$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)

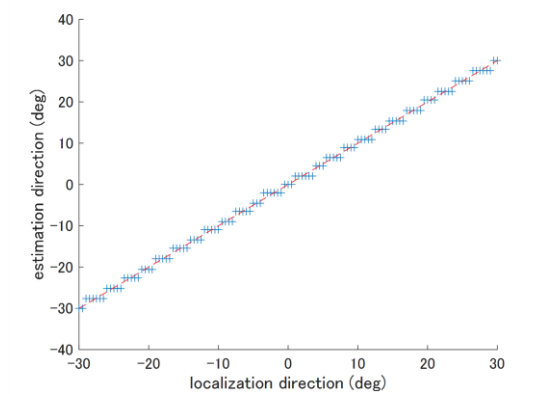


図 3.38  $f_0 = 220$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)

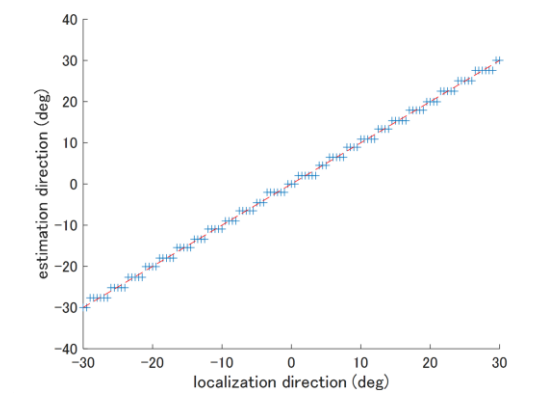


図 3.39  $f_0 = 440$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)

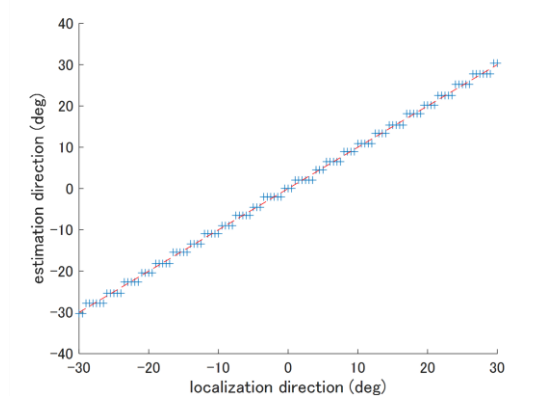


図 3.40  $f_0 = 880$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)

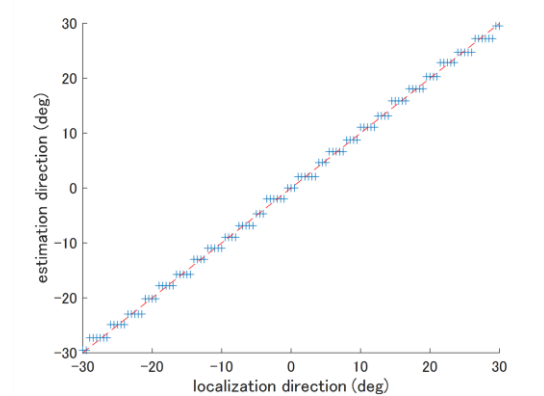


図 3.41  $f_0 = 1760$  の nylon guitar による推定結果 (MUSIC + FFNN-AO)



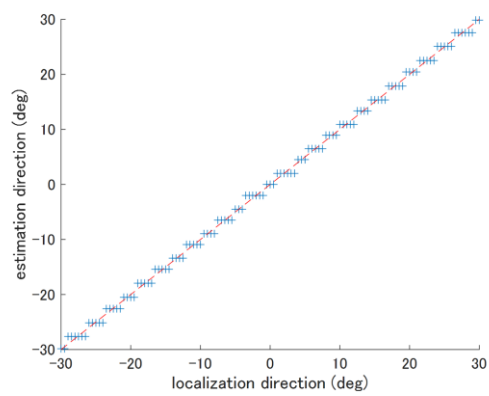


図 3.42  $f_0 = 55$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

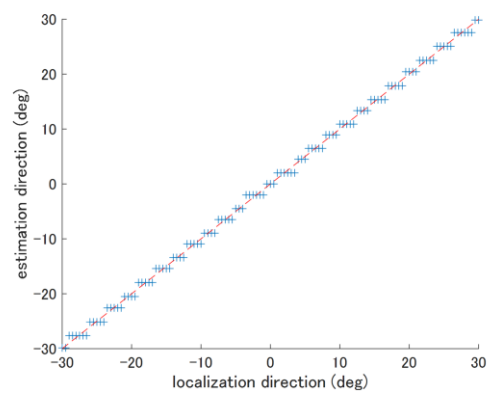


図 3.43  $f_0 = 110$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

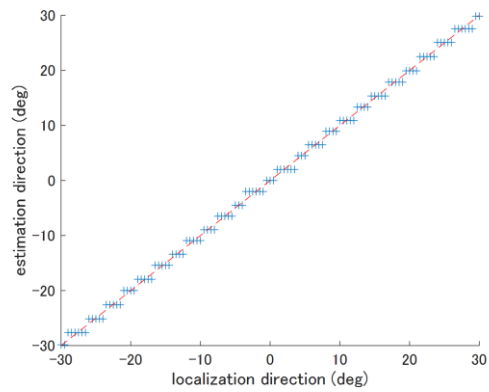


図 3.44  $f_0 = 220$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

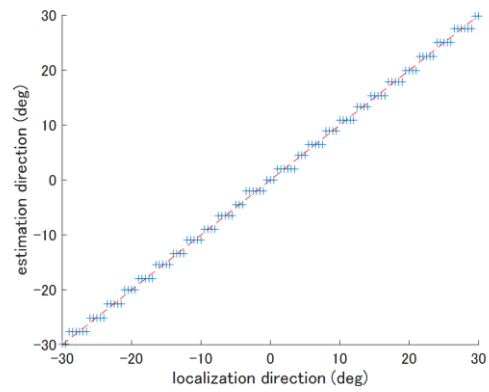


図 3.45  $f_0 = 440$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

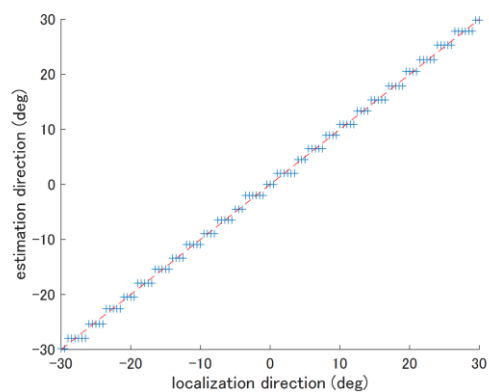


図 3.46  $f_0 = 880$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

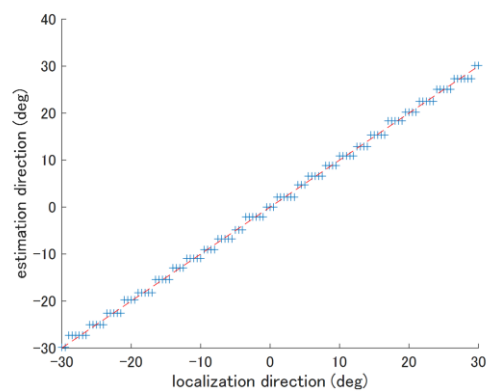


図 3.47  $f_0 = 1760$  の acoustic piano による推定結果 (MUSIC + FFNN-AO)

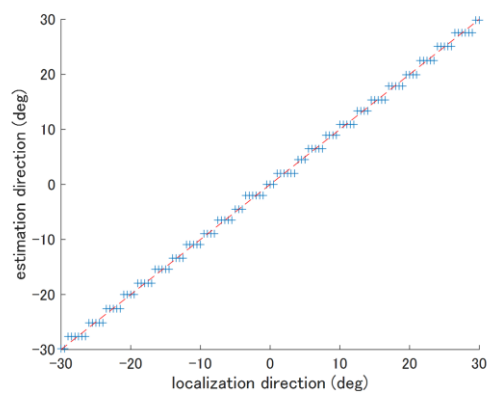


図 3.48  $f_0 = 55$  の electric piano による推定結果 (MUSIC + FFNN-AO)

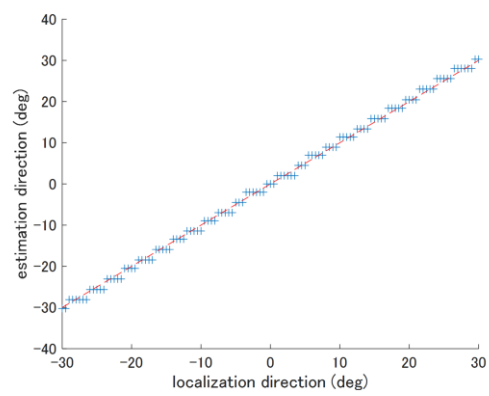


図 3.49  $f_0 = 110$  の electric piano による推定結果 (MUSIC + FFNN-AO)

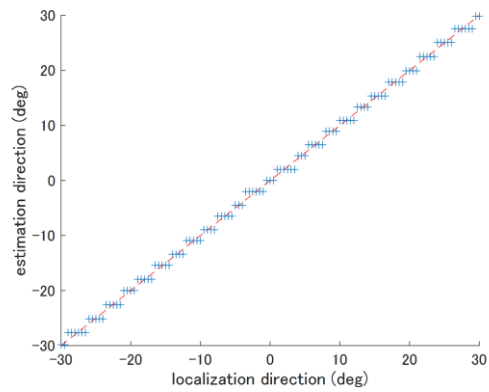


図 3.50  $f_0 = 220$  の electric piano による推定結果 (MUSIC + FFNN-AO)

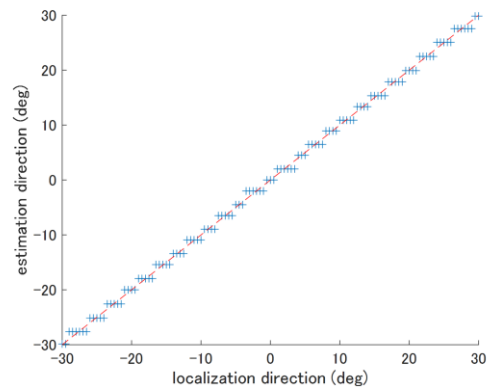


図 3.51  $f_0 = 440$  の electric piano による推定結果 (MUSIC + FFNN-AO)

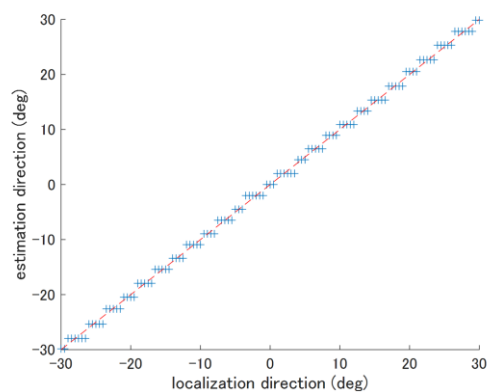


図 3.52  $f_0 = 880$  の electric piano による推定結果 (MUSIC + FFNN-AO)

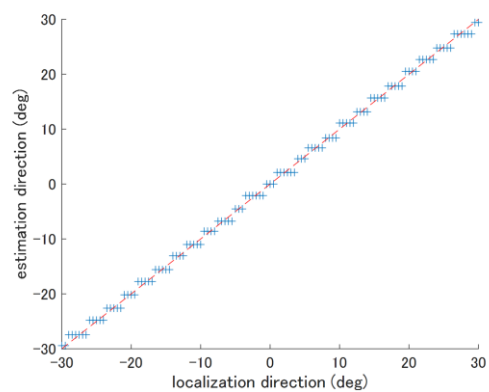


図 3.53  $f_0 = 1760$  の electric piano による推定結果 (MUSIC + FFNN-AO)

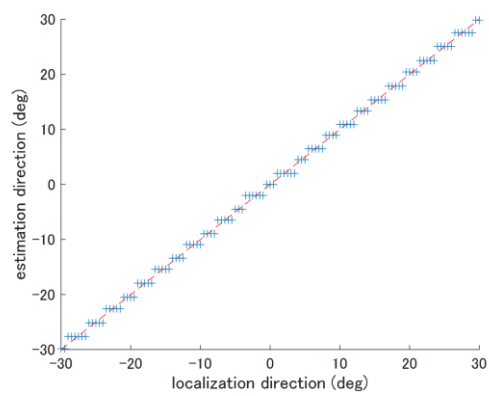


図 3.54  $f_0 = 55$  の grand piano による推定結果 (MUSIC + FFNN-AO)

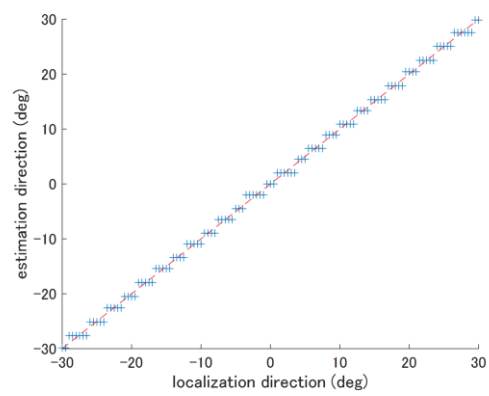


図 3.55  $f_0 = 110$  の grand piano による推定結果 (MUSIC + FFNN-AO)

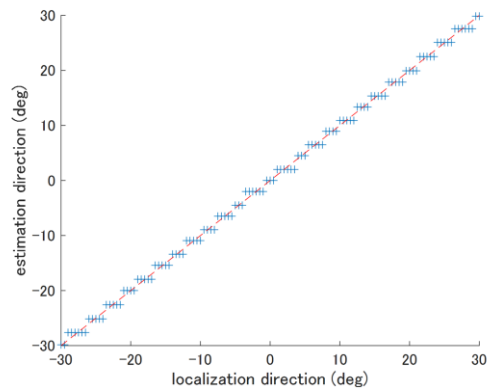


図 3.56  $f_0 = 220$  の grand piano による推定結果 (MUSIC + FFNN-AO)

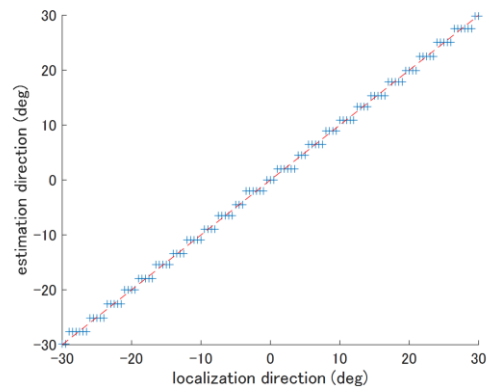


図 3.57  $f_0 = 440$  の grand piano による推定結果 (MUSIC + FFNN-AO)

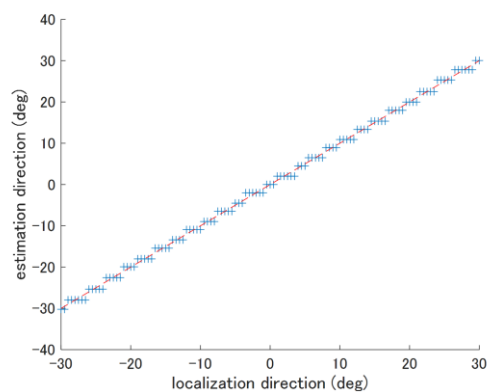


図 3.58  $f_0 = 880$  の grand piano による推定結果 (MUSIC + FFNN-AO)

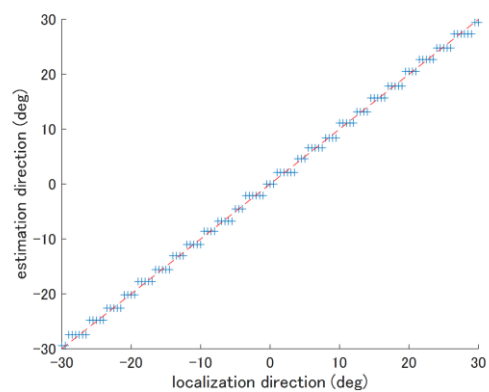


図 3.59  $f_0 = 1760$  の grand piano による推定結果 (MUSIC + FFNN-AO)

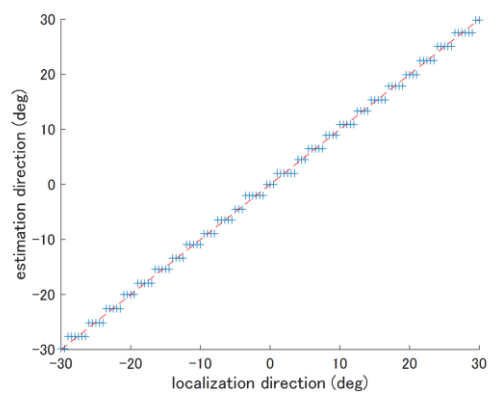


図 3.60  $f_0 = 55$  の rock piano による推定結果 (MUSIC + FFNN-AO)

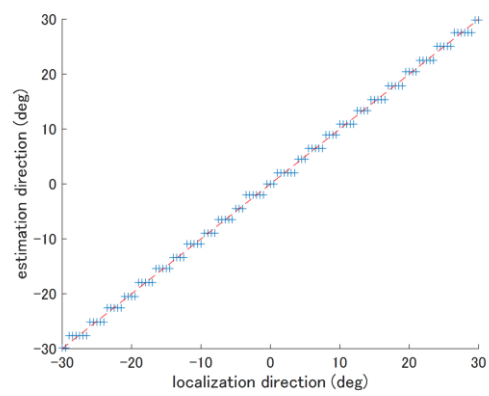


図 3.61  $f_0 = 110$  の rock piano による推定結果 (MUSIC + FFNN-AO)

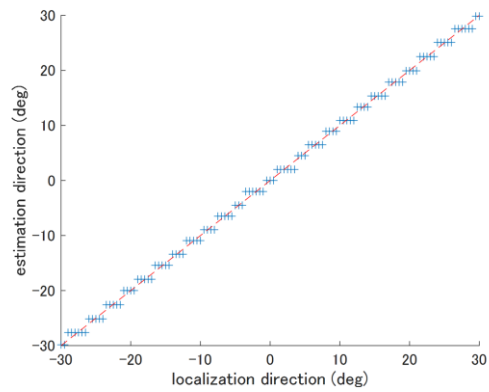


図 3.62  $f_0 = 220$  の rock piano による推定結果 (MUSIC + FFNN-AO)

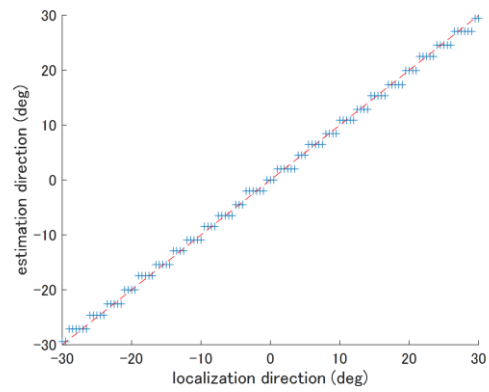


図 3.63  $f_0 = 440$  の rock piano による推定結果 (MUSIC + FFNN-AO)

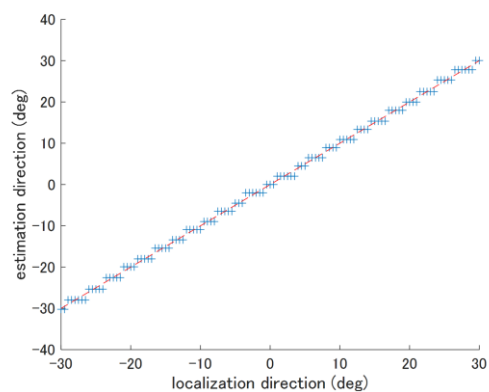


図 3.64  $f_0 = 880$  の rock piano による推定結果 (MUSIC + FFNN-AO)

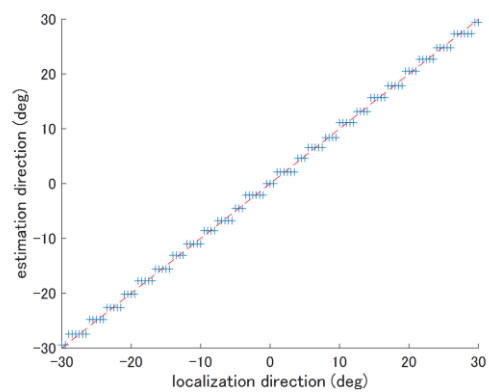


図 3.65  $f_0 = 1760$  の rock piano による推定結果 (MUSIC + FFNN-AO)

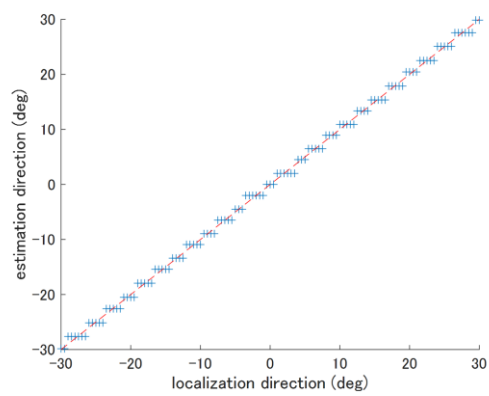


図 3.66  $f_0 = 55$  の stage piano による推定結果 (MUSIC + FFNN-AO)

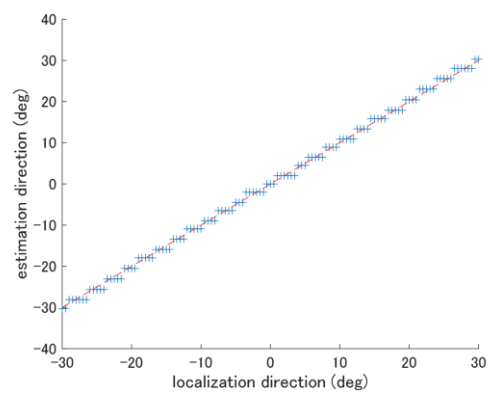


図 3.67  $f_0 = 110$  の stage piano による推定結果 (MUSIC + FFNN-AO)

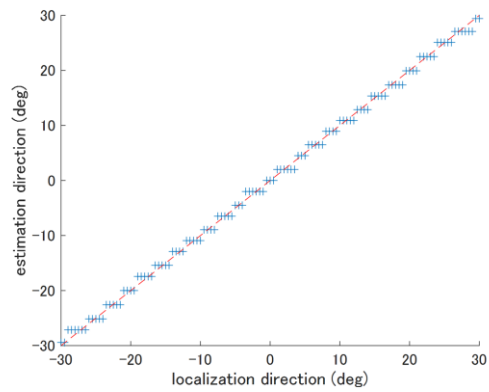


図 3.68  $f_0 = 220$  の stage piano による推定結果 (MUSIC + FFNN-AO)

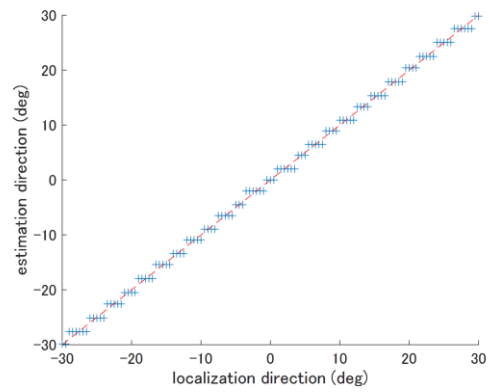


図 3.69  $f_0 = 440$  の stage piano による推定結果 (MUSIC + FFNN-AO)

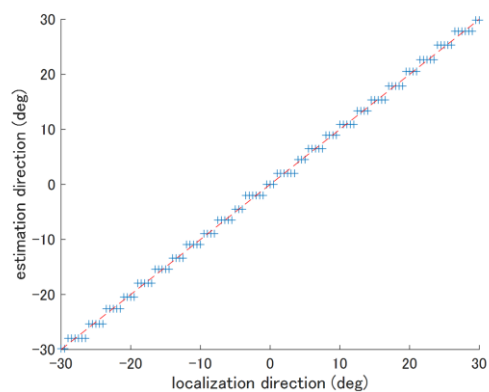


図 3.70  $f_0 = 880$  の stage piano による推定結果 (MUSIC + FFNN-AO)

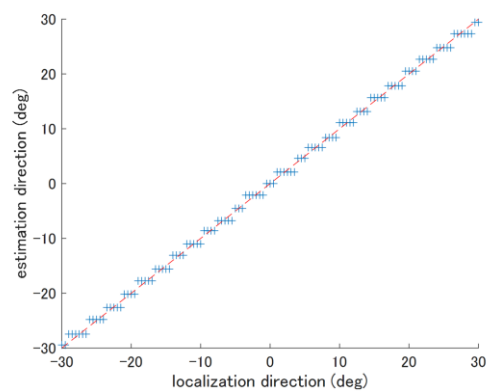


図 3.71  $f_0 = 1760$  の stage piano による推定結果 (MUSIC + FFNN-AO)

表 3.3 5 種類の guitar 音源による MUSIC 法を用いた方向推定結果の 70% を使用して学習させた FFNN-DO を用いて guitar の各音色・各周波数成分による推定結果で評価した MSE と推定誤差の最大角度

tone	$f_0$	match rate	MSE	maximum estimation error(deg)
acoustic guitar	55Hz	22.31%	0.950	2.5
	110Hz	22.31%	1.081	2.5
	220Hz	22.31%	1.033	2.5
	440Hz	22.31%	1.033	2.5
	880Hz	22.31%	1.050	2.5
	1760Hz	20.66%	1.494	3.5
clean guitar	55Hz	22.31%	1.033	2.5
	110Hz	22.31%	1.033	2.5
	220Hz	22.31%	1.033	2.5
	440Hz	22.31%	1.033	2.5
	880Hz	22.31%	1.050	2.5
	1760Hz	22.31%	1.101	2.5
country guitar	55Hz	22.31%	0.992	2.5
	110Hz	22.31%	1.033	2.5
	220Hz	22.31%	1.033	2.5
	440Hz	22.31%	1.033	2.5
	880Hz	22.31%	1.050	2.5
	1760Hz	22.31%	1.101	2.5
jazz guitar	55Hz	22.31%	1.033	2.5
	110Hz	22.31%	1.033	2.5
	220Hz	22.31%	1.033	2.5
	440Hz	22.31%	1.033	2.5
	880Hz	22.31%	1.050	2.5
	1760Hz	22.31%	1.101	2.5
nylon guitar	55Hz	22.31%	1.033	2.5
	110Hz	22.31%	1.033	2.5
	220Hz	22.31%	1.033	2.5
	440Hz	22.31%	1.033	2.5
	880Hz	22.31%	1.050	2.5
	1760Hz	22.31%	1.101	2.5

表 3.4 5 種類の piano 音源による MUSIC 法を用いた方向推定結果の 70% を使用して学習させた FFNN-DO を用いて guitar の各音色・各周波数成分による推定結果で評価した MSE と推定誤差の最大角度

tone	$f_0$	match rate	MSE	maximum estimation error(deg)
acoustic piano	55Hz	22.31%	0.967	2.0
	110Hz	22.31%	0.967	2.0
	220Hz	22.31%	0.967	2.0
	440Hz	22.31%	0.967	2.0
	880Hz	22.31%	1.035	2.5
	1760Hz	22.31%	1.072	2.5
electric piano	55Hz	22.31%	0.967	2.0
	110Hz	22.31%	1.017	2.5
	220Hz	22.31%	0.967	2.0
	440Hz	22.31%	0.967	2.0
	880Hz	22.31%	1.035	2.5
	1760Hz	22.31%	0.942	2.0
grand piano	55Hz	22.31%	0.967	2.0
	110Hz	22.31%	0.967	2.0
	220Hz	22.31%	0.967	2.0
	440Hz	22.31%	0.967	2.0
	880Hz	22.31%	1.066	2.5
	1760Hz	22.31%	0.942	2.0
rock piano	55Hz	22.31%	0.967	2.0
	110Hz	22.31%	0.967	2.0
	220Hz	22.31%	0.967	2.0
	440Hz	22.31%	0.952	2.0
	880Hz	22.31%	1.066	2.5
	1760Hz	22.31%	0.942	2.0
stage piano	55Hz	22.31%	0.967	2.0
	110Hz	22.31%	1.017	2.5
	220Hz	22.31%	0.983	2.0
	440Hz	22.31%	0.967	2.0
	880Hz	22.31%	1.035	2.5
	1760Hz	22.31%	0.942	2.0

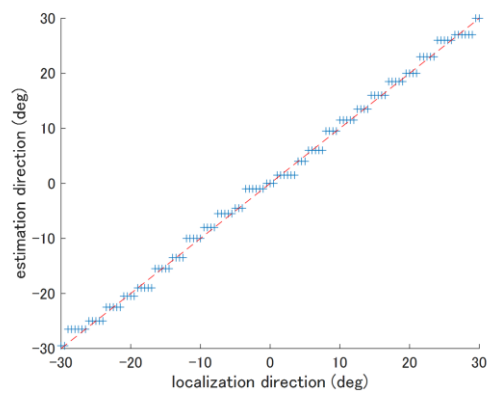


図 3.72  $f_0 = 55$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)

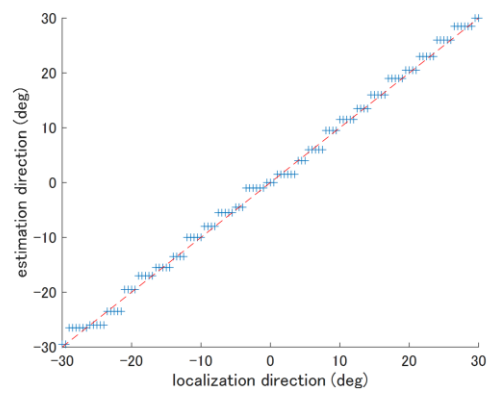


図 3.73  $f_0 = 110$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)

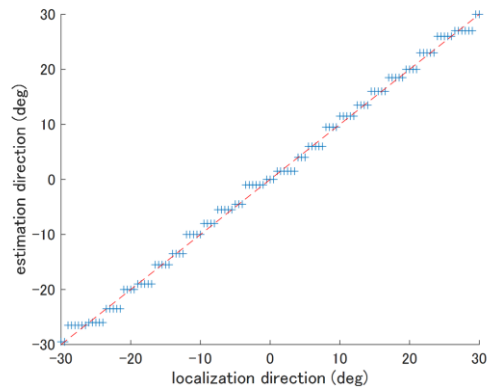


図 3.74  $f_0 = 220$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)

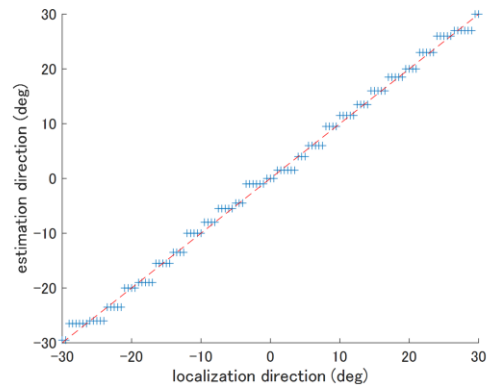


図 3.75  $f_0 = 440$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)

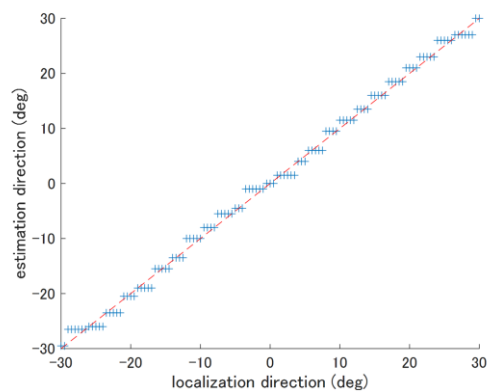


図 3.76  $f_0 = 880$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)

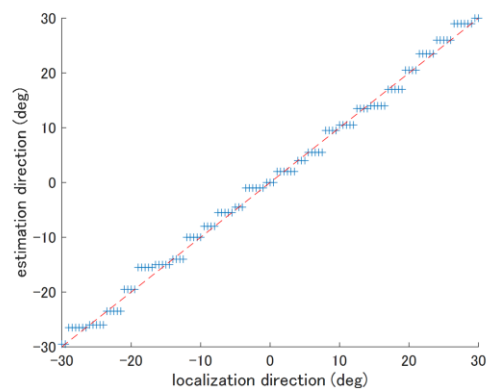


図 3.77  $f_0 = 1760$  の acoustic guitar による推定結果 (MUSIC + FFNN-DO)



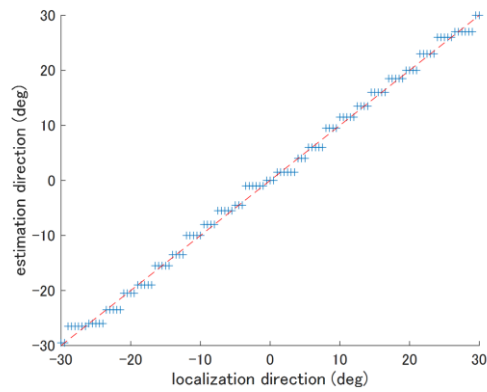


図 3.78  $f_0 = 55$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

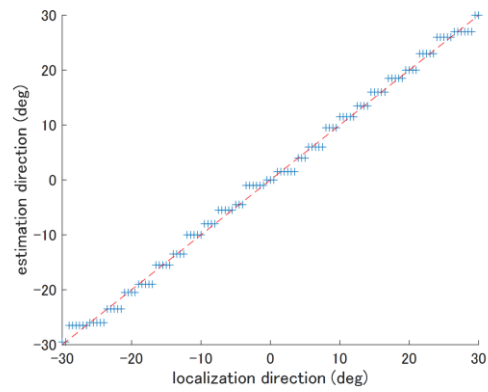


図 3.79  $f_0 = 110$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

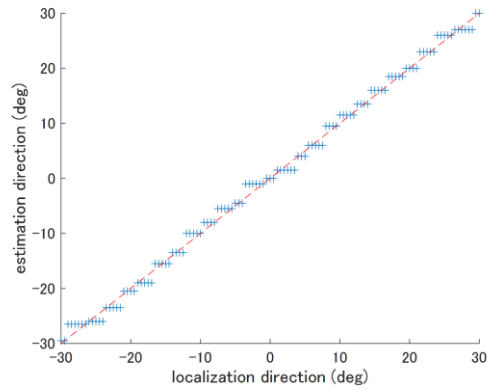


図 3.80  $f_0 = 220$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

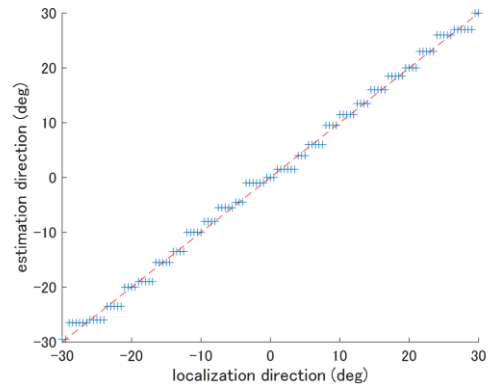


図 3.81  $f_0 = 440$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

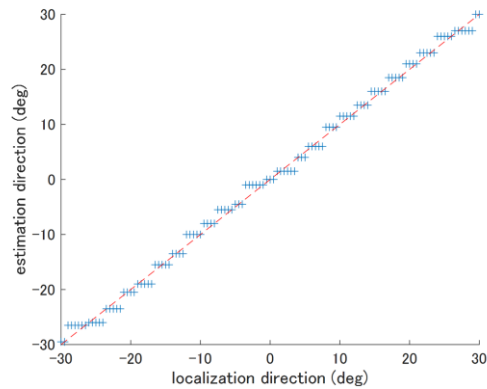


図 3.82  $f_0 = 880$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

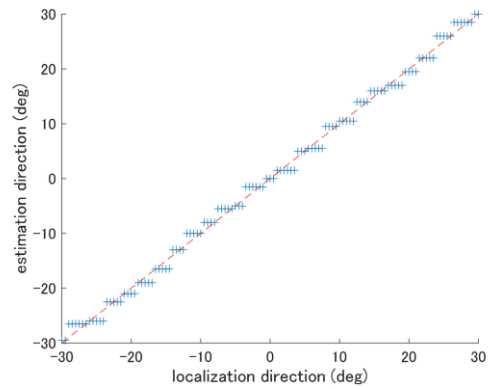


図 3.83  $f_0 = 1760$  の clean guitar による推定結果 (MUSIC + FFNN-DO)

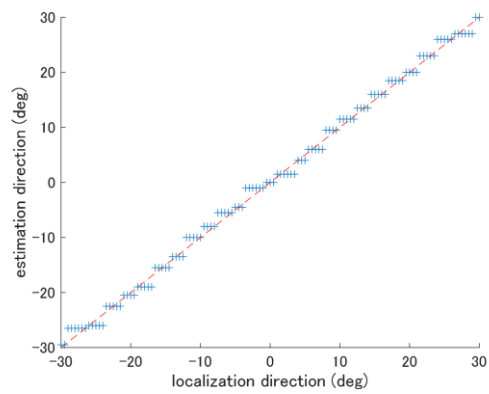


図 3.84  $f_0 = 55$  の country guitar による推定結果 (MUSIC + FFNN-DO)

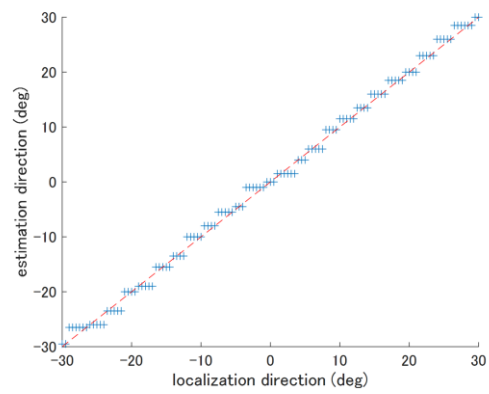


図 3.85  $f_0 = 110$  の country guitar による推定結果 (MUSIC + FFNN-DO)

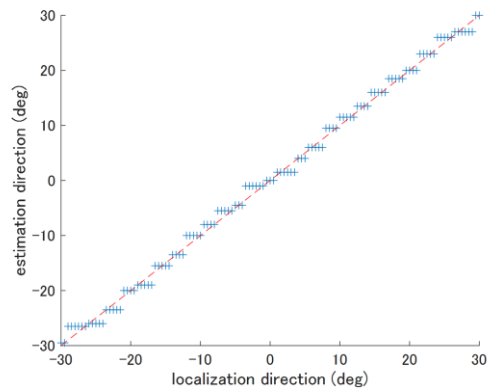


図 3.86  $f_0 = 220$  の country guitar による推定結果 (MUSIC + FFNN-DO)

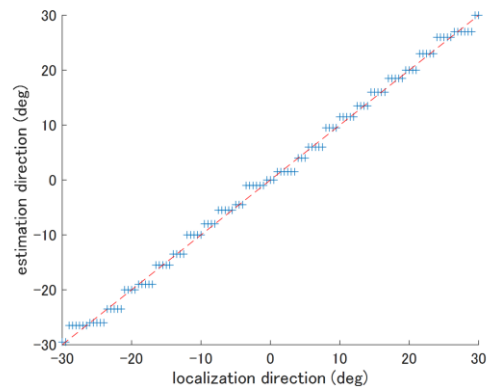


図 3.87  $f_0 = 440$  の country guitar による推定結果 (MUSIC + FFNN-DO)

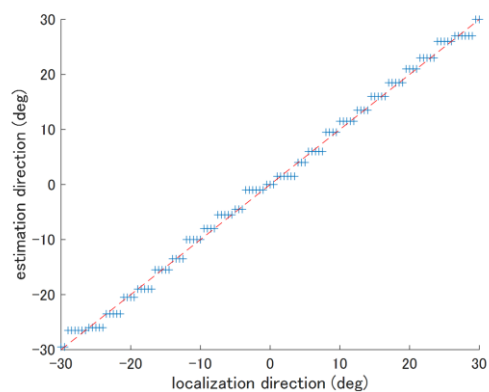


図 3.88  $f_0 = 880$  の country guitar による推定結果 (MUSIC + FFNN-DO)

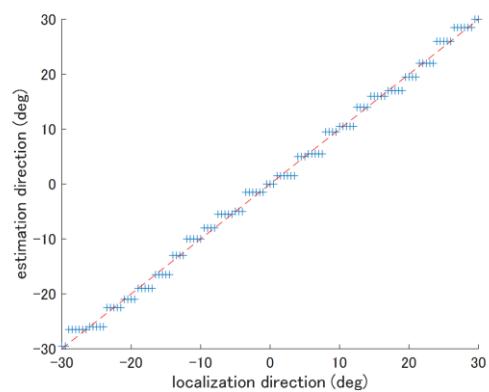


図 3.89  $f_0 = 1760$  の country guitar による推定結果 (MUSIC + FFNN-DO)

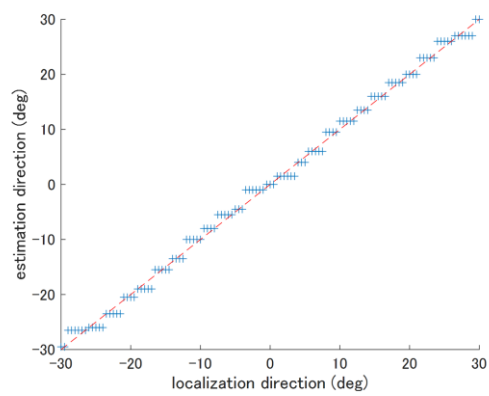


図 3.90  $f_0 = 55$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

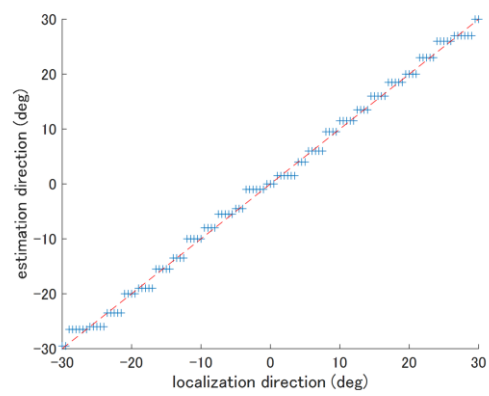


図 3.91  $f_0 = 110$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

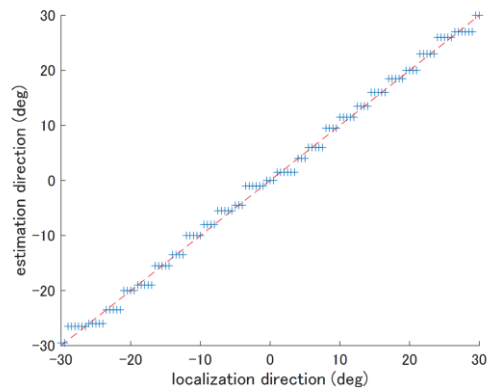


図 3.92  $f_0 = 220$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

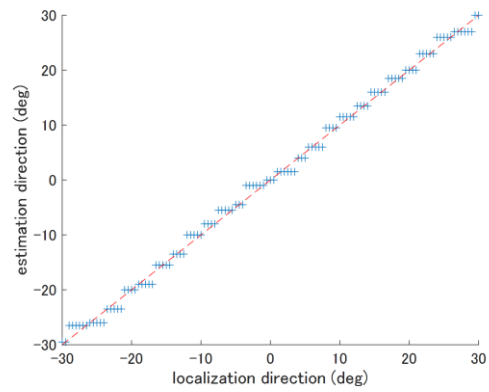


図 3.93  $f_0 = 440$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

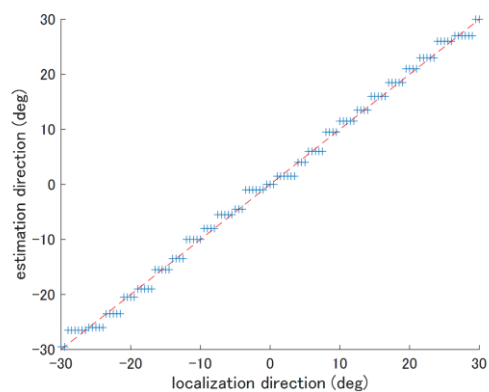


図 3.94  $f_0 = 880$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

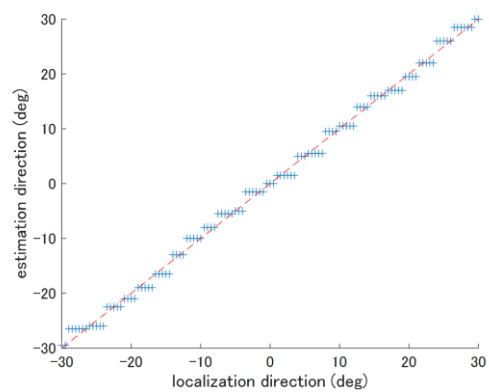


図 3.95  $f_0 = 1760$  の jazz guitar による推定結果 (MUSIC + FFNN-DO)

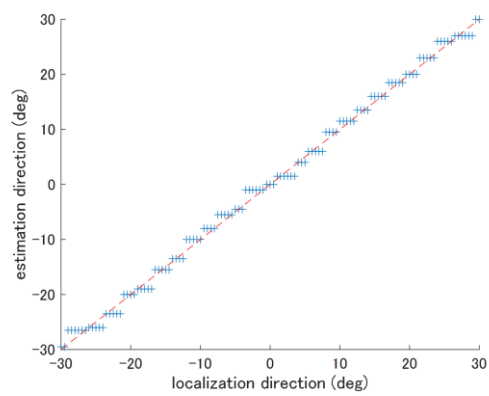


図 3.96  $f_0 = 55$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

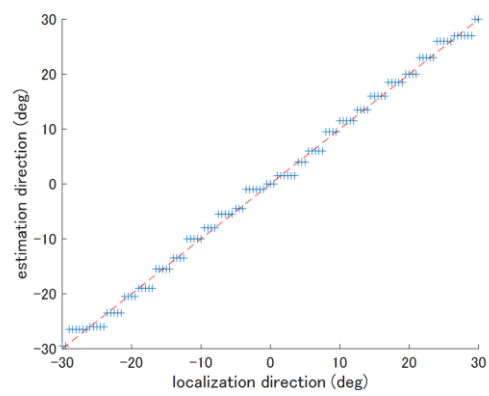


図 3.97  $f_0 = 110$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

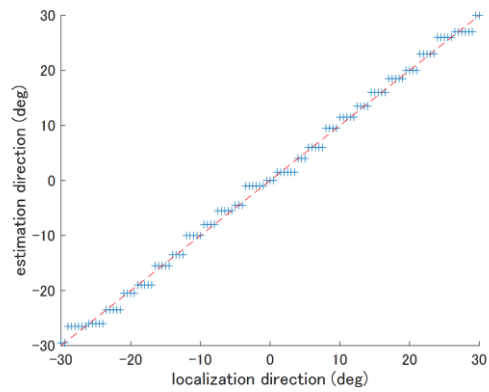


図 3.98  $f_0 = 220$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

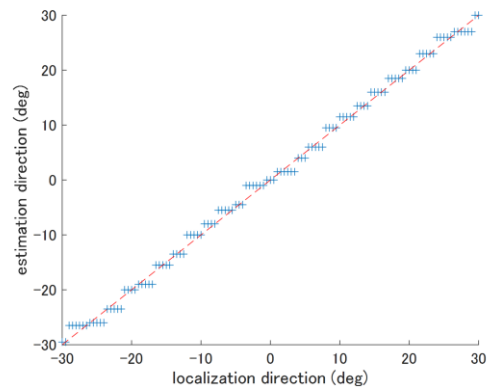


図 3.99  $f_0 = 440$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

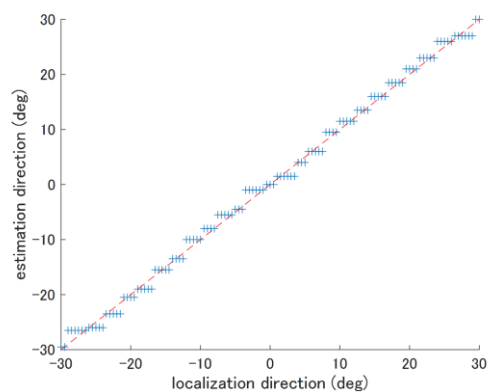


図 3.100  $f_0 = 880$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

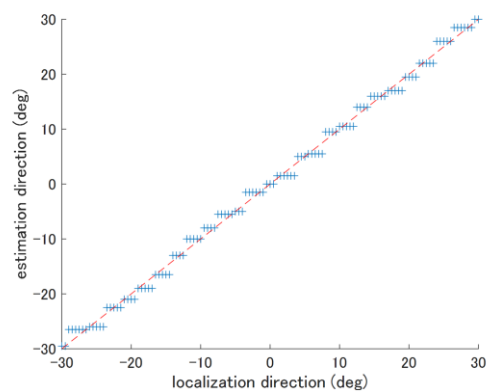


図 3.101  $f_0 = 1760$  の nylon guitar による推定結果 (MUSIC + FFNN-DO)

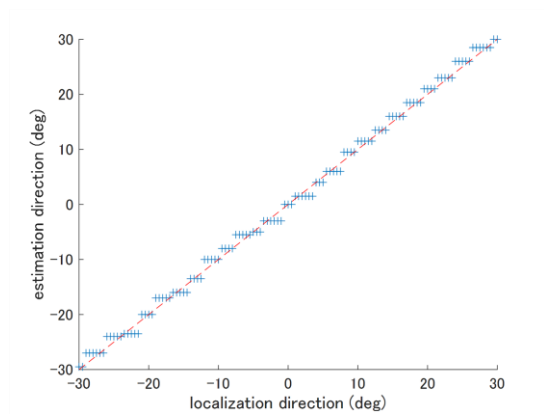


図 3.102  $f_0 = 55$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

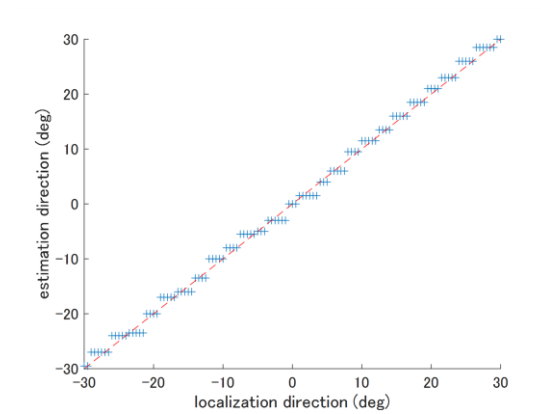


図 3.103  $f_0 = 110$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

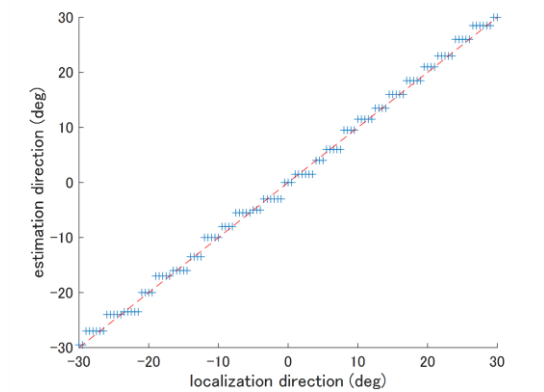


図 3.104  $f_0 = 220$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

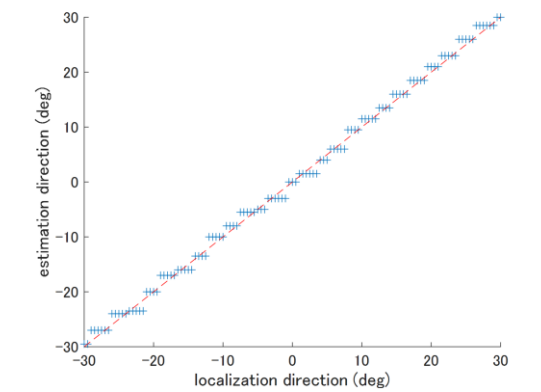


図 3.105  $f_0 = 440$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

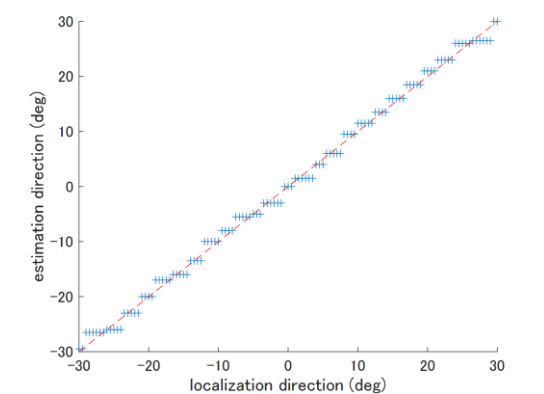


図 3.106  $f_0 = 880$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

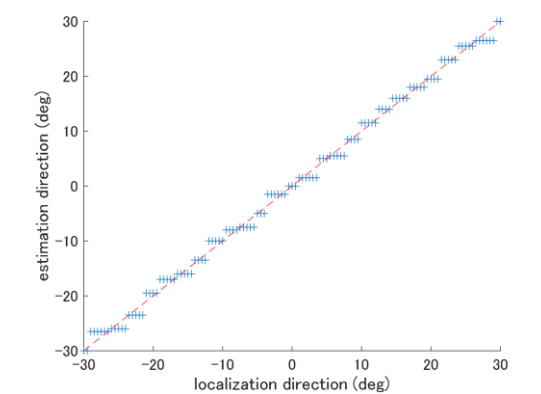


図 3.107  $f_0 = 1760$  の acoustic piano による推定結果 (MUSIC + FFNN-DO)

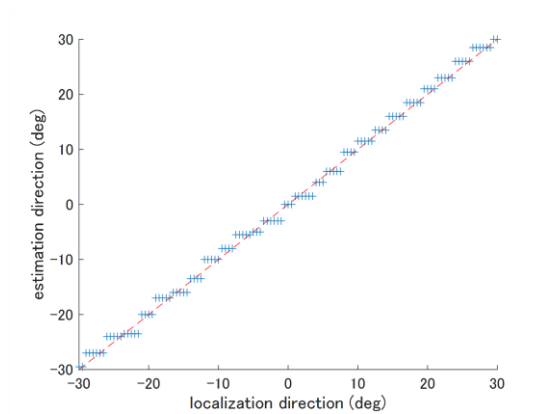


図 3.108  $f_0 = 55$  の electric piano による推定結果 (MUSIC + FFNN-DO)

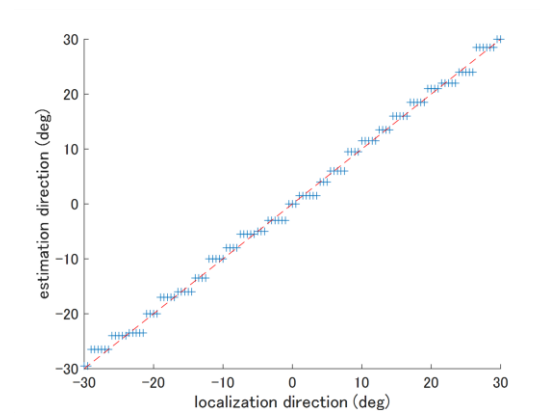


図 3.109  $f_0 = 110$  の electric piano による推定結果 (MUSIC + FFNN-DO)

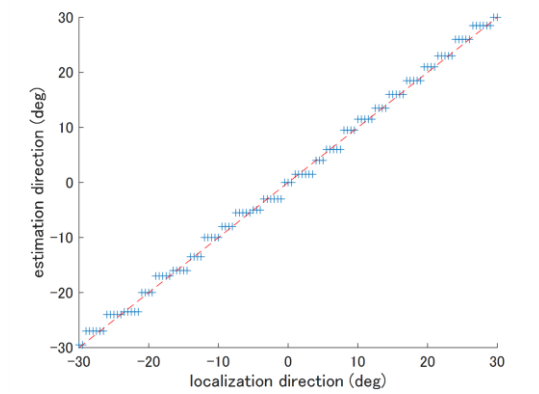


図 3.110  $f_0 = 220$  の electric piano による推定結果 (MUSIC + FFNN-DO)

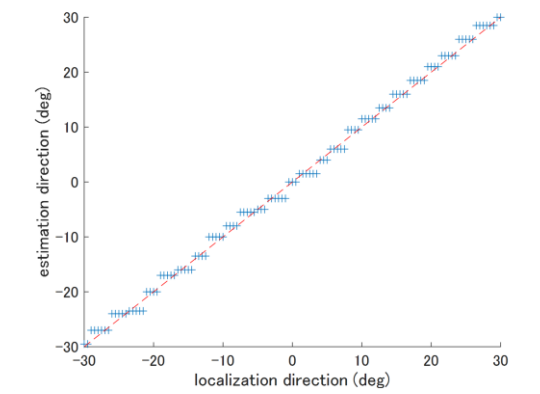


図 3.111  $f_0 = 440$  の electric piano による推定結果 (MUSIC + FFNN-DO)

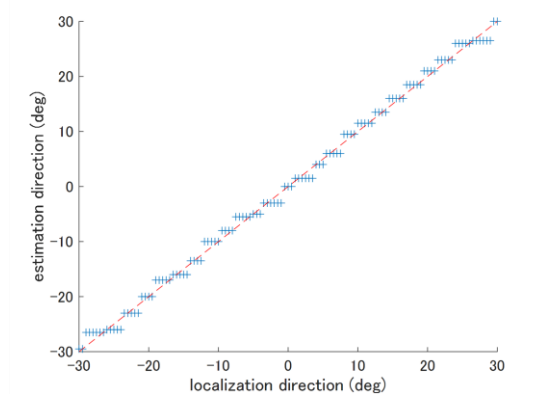


図 3.112  $f_0 = 880$  の electric piano による推定結果 (MUSIC + FFNN-DO)

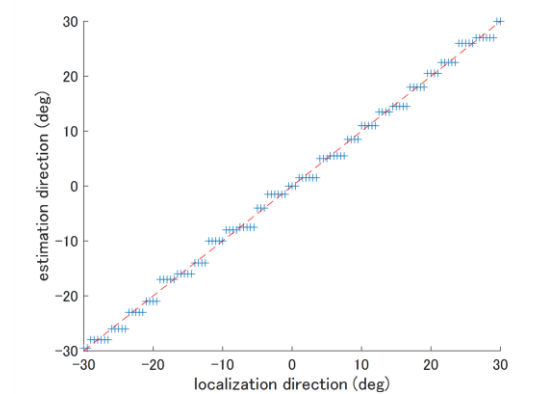


図 3.113  $f_0 = 1760$  の electric piano による推定結果 (MUSIC + FFNN-DO)

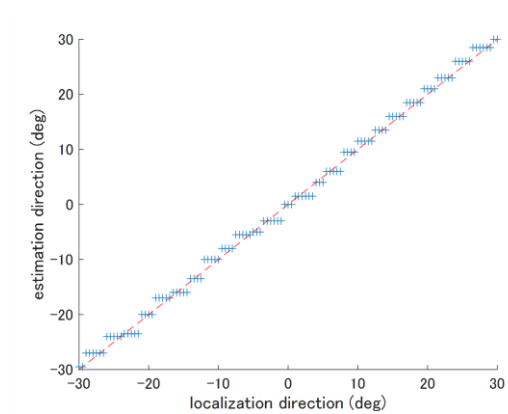


図 3.114  $f_0 = 55$  の grand piano による推定結果 (MUSIC + FFNN-DO)

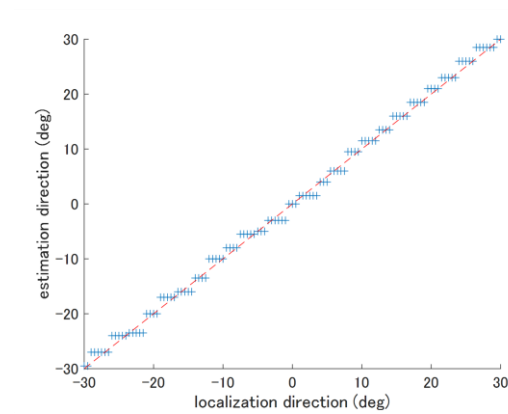


図 3.115  $f_0 = 110$  の grand piano による推定結果 (MUSIC + FFNN-DO)

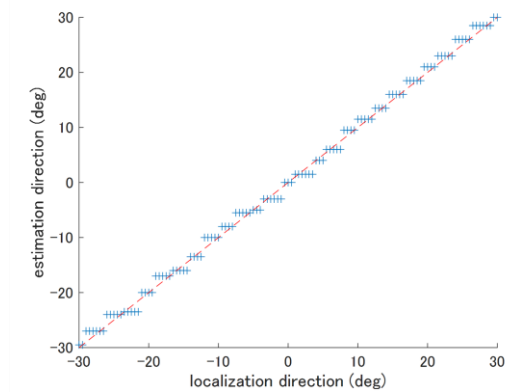


図 3.116  $f_0 = 220$  の grand piano による推定結果 (MUSIC + FFNN-DO)

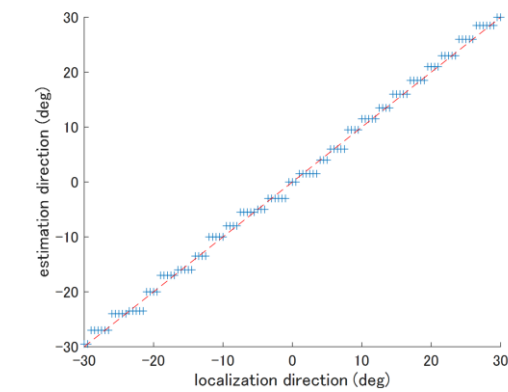


図 3.117  $f_0 = 440$  の grand piano による推定結果 (MUSIC + FFNN-DO)

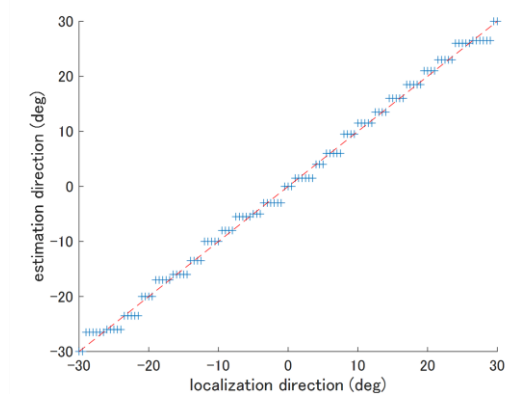


図 3.118  $f_0 = 880$  の grand piano による推定結果 (MUSIC + FFNN-DO)

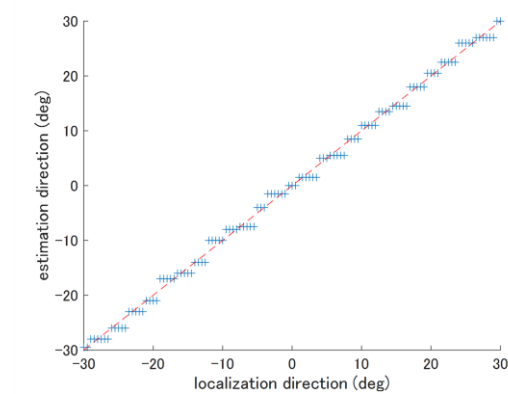


図 3.119  $f_0 = 1760$  の grand piano による推定結果 (MUSIC + FFNN-DO)

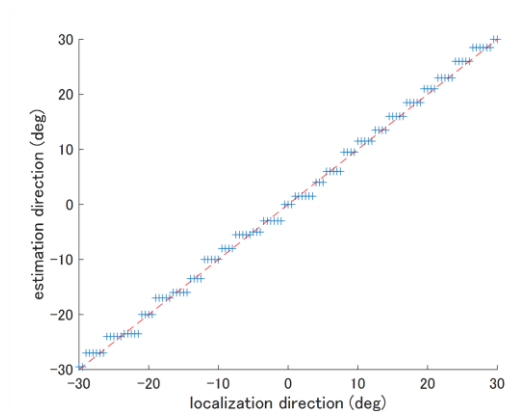


図 3.120  $f_0 = 55$  の rock piano による推定結果 (MUSIC + FFNN-DO)

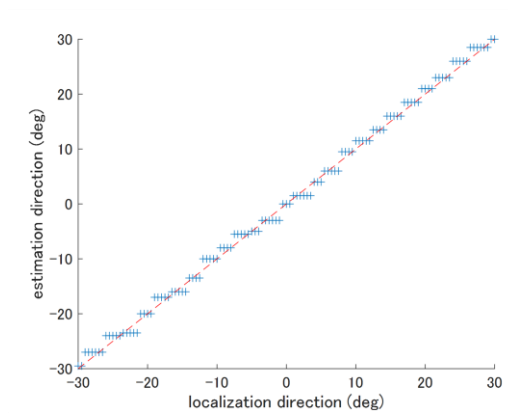


図 3.121  $f_0 = 110$  の rock piano による推定結果 (MUSIC + FFNN-DO)

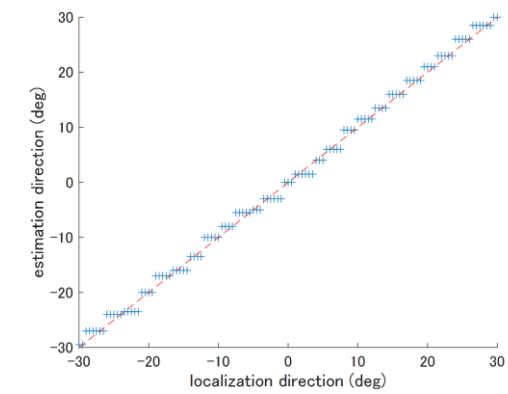


図 3.122  $f_0 = 220$  の rock piano による推定結果 (MUSIC + FFNN-DO)

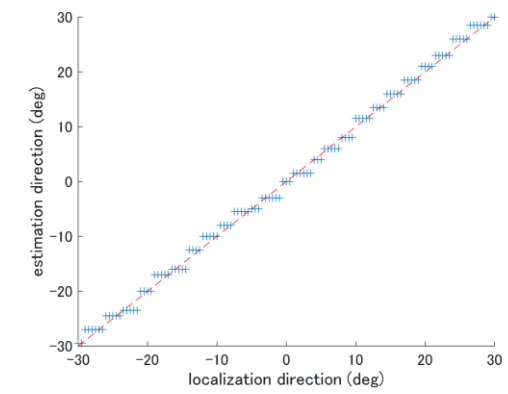


図 3.123  $f_0 = 440$  の rock piano による推定結果 (MUSIC + FFNN-DO)

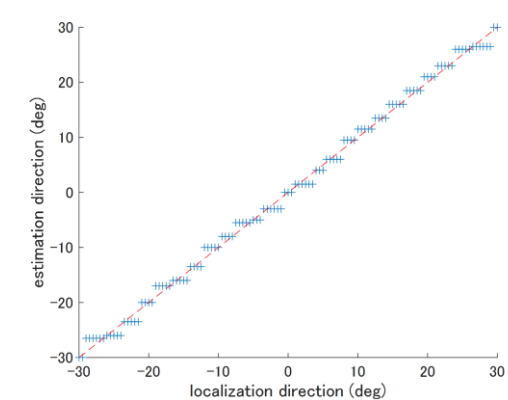


図 3.124  $f_0 = 880$  の rock piano による推定結果 (MUSIC + FFNN-DO)

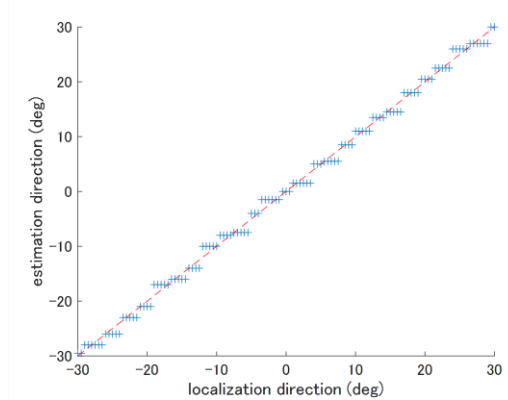


図 3.125  $f_0 = 1760$  の rock piano による推定結果 (MUSIC + FFNN-DO)



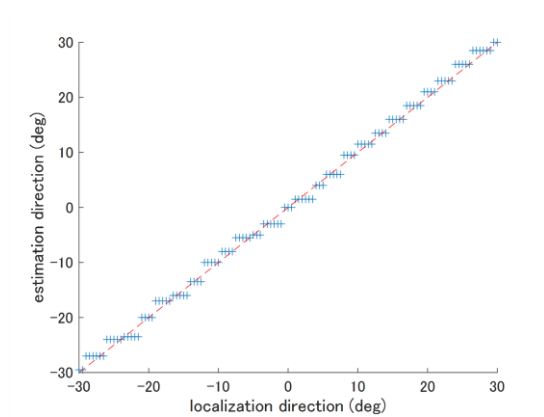


図 3.126  $f_0 = 55$  の stage piano による推定結果 (MUSIC + FFNN-DO)

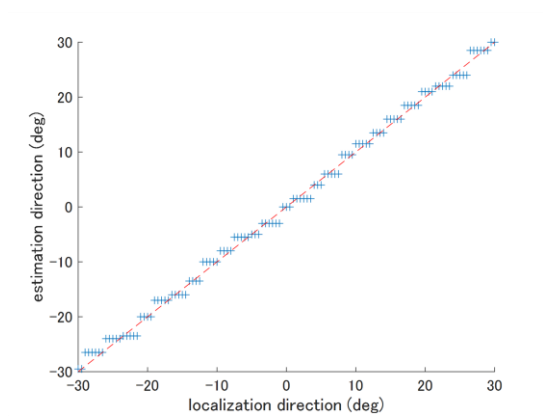


図 3.127  $f_0 = 110$  の stage piano による推定結果 (MUSIC + FFNN-DO)

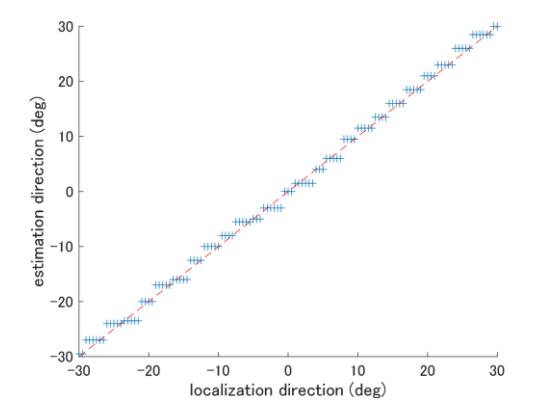


図 3.128  $f_0 = 220$  の stage piano による推定結果 (MUSIC + FFNN-DO)

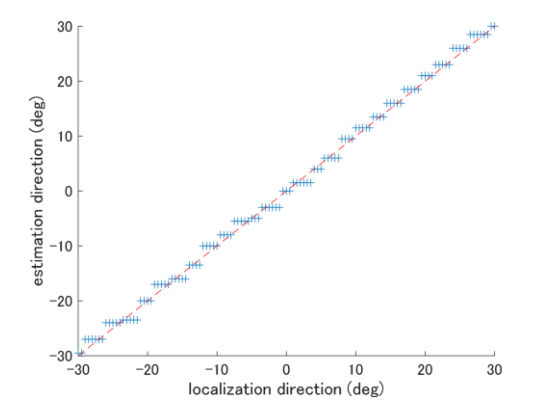


図 3.129  $f_0 = 440$  の stage piano による推定結果 (MUSIC + FFNN-DO)

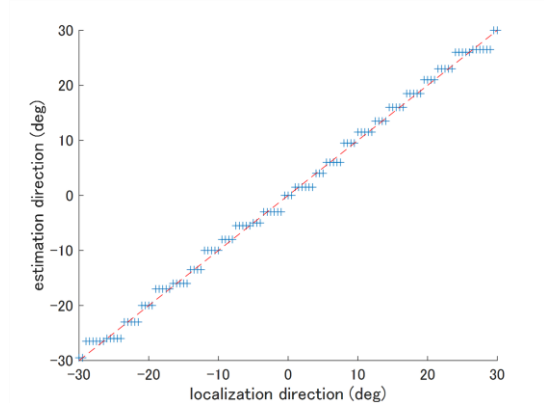


図 3.130  $f_0 = 880$  の stage piano による推定結果 (MUSIC + FFNN-DO)

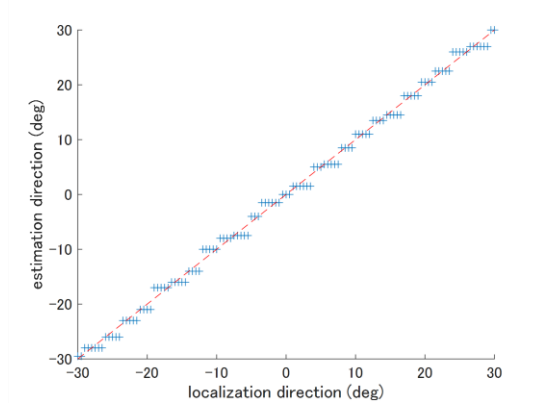


図 3.131  $f_0 = 1760$  の stage piano による推定結果 (MUSIC + FFNN-DO)

### 3.3.3 MUSIC 法による推定結果との比較

第2章で述べた MUSIC 法による推定手法と FFNN-AO / FFNN-DO を用いた推定手法について、60 種類（5×2 種類の音色・6 種類の  $f_0$ ）全ての定位方向と推定方向に対する MSE を表 3.5 に示す。3 つの手法を比較すると、最も MSE が小さくなったのは、出力ニューロンが 1 個であるアナログ出力の FFNN-AO を使用した手法となった。

FFNN-AO を用いた方向推定について、60 種類の音源のうち 81.7% の音源において MUSIC 法のみによる方向推定による MSE よりも FFNN-AO を用いた方向推定による MSE のほうが小さい値となった。acoustic guitar ( $f_0 = 1760\text{Hz}$ ) による推定結果（only MUSIC, MUSIC + FFNN-AO）を図 3.132, 3.133 に示す。MUSIC 法のみによる推定結果においては、定位方向が  $M_2$  正面から離れるにつれて推定誤差が大きくなり、最大で 3.5 度の差が生じ、MSE は 2.161 となったが、FFNN-AO を使用することで推定誤差が低減され、MSE が小さくなった。

FFNN-DO を用いた方向推定においては、acoustic guitar ( $f_0 = 1760\text{Hz}$ ) を除くすべての音源において MUSIC 法のみによる方向推定による MSE よりも FFNN-DO を用いた方向推定による MSE のほうが大きい値となった。一方で定位方向と推定方向の一致率は改悪することなく、すべての音源で 20% を超えた。図 3.134, 3.135 は electric piano ( $f_0 = 110\text{Hz}$ ) による推定結果であり、一致率は上昇したが MSE が改悪した一例

表 3.5 60 種類すべての音源における定位方向と推定方向に対する MSE

Method	MSE
only MUSIC	0.555
MUSIC + FFNN-AO	0.524
MUSIC + FFNN-DO	1.139

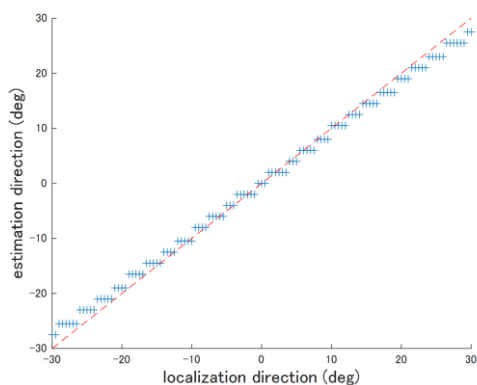


図 3.132 MUSIC 法のみによる acoustic guitar ( $f_0 = 1760\text{Hz}$ ) を対象とした推定結果（図 2.16 再掲）

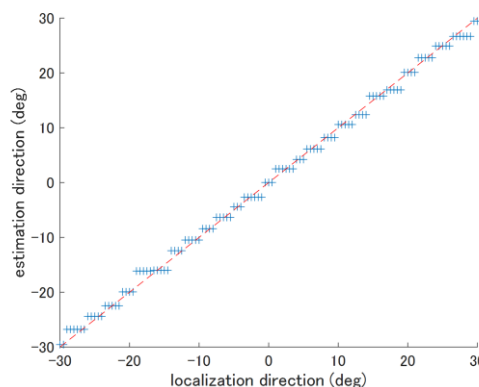


図 3.133 FFNN-AO を用いた acoustic guitar ( $f_0 = 1760\text{Hz}$ ) による推定結果（図 3.17 再掲）

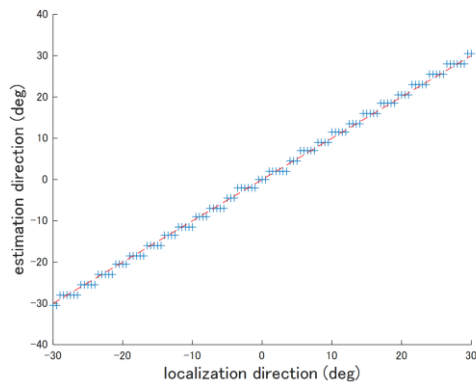


図 3.134 MUSIC 法のみによる electric piano ( $f_0 = 110\text{Hz}$ ) を対象とした推定結果 (図 2.42 再掲)

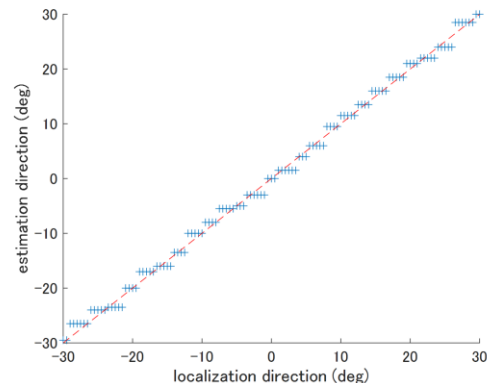


図 3.135 FFNN-DO を用いた electric piano ( $f_0 = 110\text{Hz}$ ) による推定結果 (図 3.109 再掲)

である。すなわち、推定方向が定位方向と一致するサンプル数は同一もしくは増加したが、一致しないサンプルにおける推定誤差が大きくなったと言える。

FFNN を使用する際の学習用データ数は出力ニューロン数が異なっても同じであるが、出力ニューロンが 121 個であるデジタル出力の FFNN-DO の場合はニューロン間を結合するシナプス数の増加に伴い、学習すべきパラメータが多くなることにより、より多くの学習用データが必要になると考えられる。しかし、FFNN-DO では、「60 種類すべての音源における定位方向と推定方向に対する MSE (表 3.5)」と「各音色・各周波数成分による推定結果を用いて評価した MSE (表 3.1~3.4)」との差が小さく、音源の違いによる推定結果のばらつきは最も小さかったと言える。

### 3.4 結言

楽器音の高調波成分を用いた音源方向推定として FFNN を用いた手法について述べた。楽器や周波数ごとの誤りパターンを学習させることで推定精度向上を図った。FFNN-AO を用いた場合、特に基本周波数が高いときに MSE が改善傾向であることが示された。FFNN-DO を用いた場合は、学習すべきパラメータが多くなることにより十分な学習を行なうことができず、MUSIC 法による推定結果と比較すると MSE は増加しているが、音源の違いによる推定結果のばらつきは最も小さかった。学習用データ数の増加やネットワークの構造を変更させることが今後の課題として挙げられる。

## 第4章 基本周波数が高い音響信号を対象とした方向推定における推定誤差の低減

### 4.1 序言

MUSIC 法による音源方向推定において、基本周波数の高い音源は空間エイリアシングにより正しい推定ができない傾向にある[17]. 本章では音色の基本周波数が高く、基本周波数の違いによって倍音構造が大きく変化しない piccolo を対象とした音源方向推定について述べる. 畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) は機械学習の一種である深層学習 (Deep Learning) の一つで画像認識などの画像処理など使用されており、畳み込み層・プーリング層・全結合層から構成される. 畳み込み層では畳み込み演算 (積和演算) を行なうことで画像の特徴を抽出し、プーリング層では畳み込み層で求めたデータの収縮処理として冗長な情報が削除される[34]. MUSIC 法を用いた方向推定において基本波や高調波成分を用いた場合、推定誤差は小さくなるが、基本周波数が高い場合は空間エイリアシングが生じることで推定精度が低下する. そこで推定方向算出に用いる MUSIC スペクトルにより、それぞれのスペクトルパターンを学習させることで推定誤差の低減を図る.

本章の構成は次の通りである. 4.2 節では MUSIC スペクトルと用いた CNN による方向推定について述べる. 4.3 節では CNN による方向推定シミュレーションの実験についての結果と考察について述べ、4.4 節で結言とする.

### 4.2 CNN を用いた音源方向推定

CNN は画像処理の分野で幅広く使用されているニューラルネットワークで、近年では音響信号処理である音源分離や音声分類においても使用されているネットワークである[35~39]. MUSIC 法を用いた方向推定において基本波や高調波成分を用いた場合、推定誤差は小さくなるが、基本周波数が高い場合は空間エイリアシングが生じることで推定精度が低下する. 第3章で述べた手法では基本波・高調波成分を用いて算出した推定結果を用いているため、推定方向算出のために算出される相関行列や MUSIC スペクトルの情報は一切用いられていない. 本章では推定結果より詳細な情報を持つ MUSIC スペクトルをネットワークの入力とする手法について検討する. 空間エイリアシングが生じないとき、図 4.1 のように到来角の探索範囲 (MUSIC スペクトルを算出する範囲) 内に生じる MUSIC スペクトルのピークは1つであるが、基本周波数が高い場合の方向推定では、定位方向にピークが生じるだけでなく図 4.2 のように到来角の探索範囲内に複数の MUSIC スペクトルのピークが生じる. これはよ

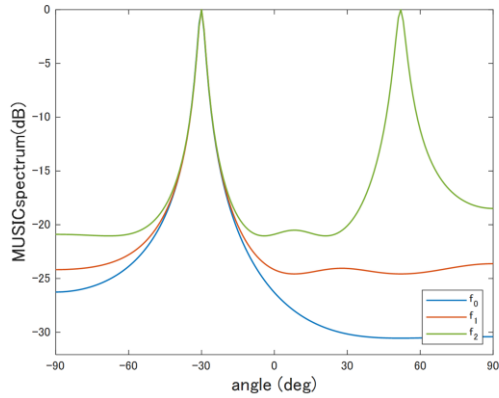


図 4.1 A5 音源 ( $f_0 = 880\text{Hz}$ ) における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向:  $-30$  度)

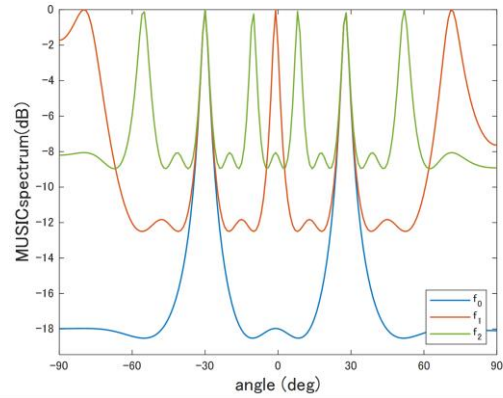


図 4.2 A7 音源 ( $f_0 = 3520\text{Hz}$ ) における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向:  $-30$  度)

り周波数の高い第 2, 3 高調波成分のみを用いたときに顕著である. こうした MUSIC スペクトルのパターンに一定の傾向があるならば, MUSIC スペクトルのパターンに対する定位方向を学習することで定位方向を推定できると考え, これにより特に推定誤差の大きい基本周波数が高い音源における推定誤差低減を図る. 本研究では基本周波数成分・第 2, 3 高調波成分による MUSIC スペクトルを用いて, あらかじめ用意された角度から選択する分類問題と角度を直接求める回帰分析として音源方向の推定を試みる. 回帰分析に用いる CNN では連続的なデータを予測するためにネットワークの最終層に回帰層を使用する. 回帰分析においては出力される推定方向の範囲を限定しない. 分類問題においてはあらかじめ用意された推定方向のうち入力信号がどの推定方向であるかを計算する. CNN の入力  $\mathbf{P}_{input}$  は  $f = f_0, f_1, f_2$  として求めたそれぞれの MUSIC スペクトルを用いて式(13)で示される. ただし,  $\mathbf{P}(f)$  は式(14)で定義され,  $p$  は到来角の探索数である.

$$\mathbf{P}_{input} = [\mathbf{P}(f_0), \mathbf{P}(f_1), \mathbf{P}(f_2)]^T, \quad (13)$$

$$\mathbf{P}(f) = [P_{MUSIC}(f, \theta_1), P_{MUSIC}(f, \theta_2), \dots, P_{MUSIC}(f, \theta_p)]^T. \quad (14)$$

ここで  $p$  は 181 ( $-90$  度から  $+90$  度, 1 度間隔) である.

次に, CNN の構造について述べる. 回帰分析における CNN (CNN-R) は図 4.3~4.5 に示す CNN-R1, CNN-R2, CNN-R3 の 3 種類, 分類問題における CNN (CNN-C) は図 4.6~4.9 に示す CNN-C1, CNN-C2, CNN-C3, CNN-C4 の 4 種類を用いる.

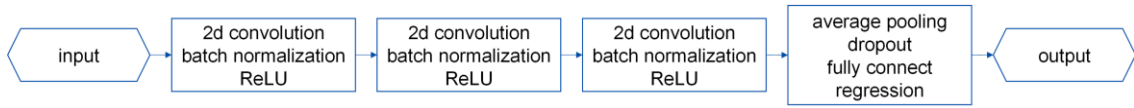


図 4.3 CNN-R1 の構造

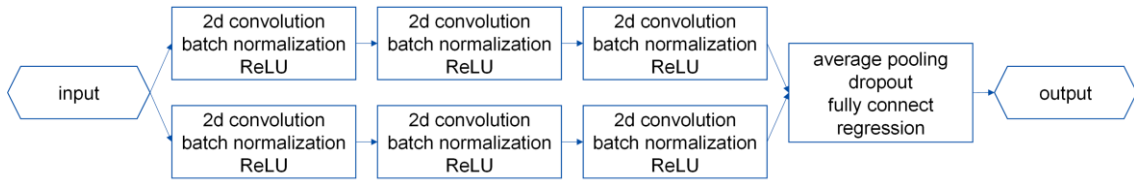


図 4.4 CNN-R2 の構造

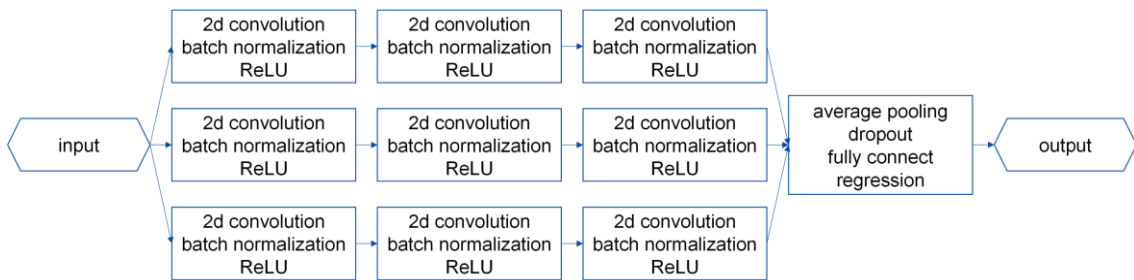


図 4.5 CNN-R3 の構造

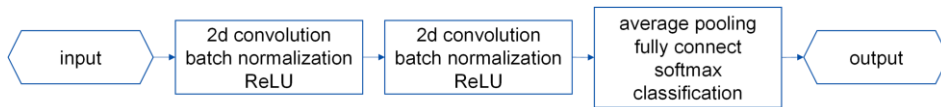


図 4.6 CNN-C1 の構造

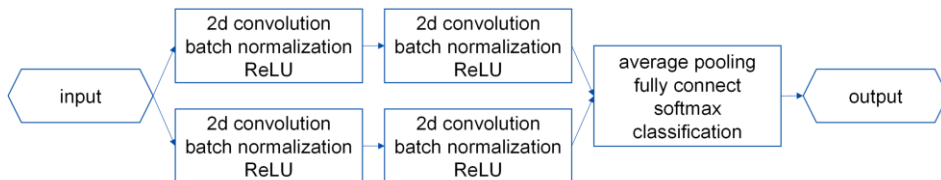


図 4.7 CNN-C2 の構造

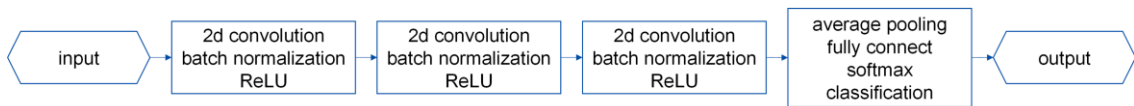


図 4.8 CNN-C3 の構造

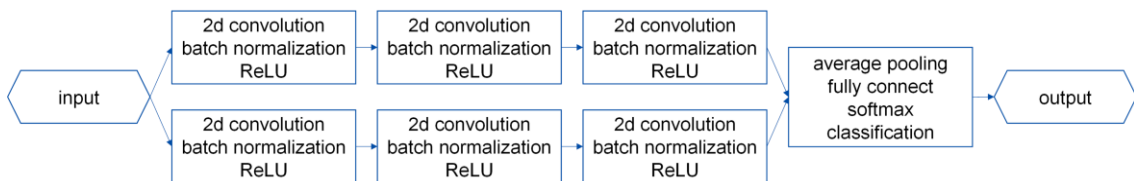


図 4.9 CNN-C4 の構造

回帰分析において CNN を使用する推定方法[34]について述べる.  $\mathbf{P}_{input}$  の 70% を学習用データ, 20% を検証用データとして使用し, 残りのデータを評価に使用した. このデータセットは無作為に 2 つ (dataset-1, dataset-2) 用意する. 学習アルゴリズムにモーメンタム項付き確率的勾配降下法 (SGDM : Momentum Stochastic Gradient Descent), 損失関数に MSE を用い, ミニバッチサイズ (学習アルゴリズムが重みを更新する際に使用するデータ数) は 128, 256, 512, 学習率 0.001, ドロップアウト率 (過学習を防ぐために入力要素をゼロに設定する確率) 10% と定義する. SGDM は確率的勾配降下法 (SGD : Stochastic Gradient Descent) において, 反復ごとに損失関数を最小化するようにネットワークのパラメータを更新する際に, 更新する方向に対してパラメータを変化させやすくする手法である. 損失勾配が負と正を繰り返す場合, パラメータの変化が小さくなり, SGD に比べて学習速度が速くなる. SGDM によるパラメータベクトル  $q_l$  の更新は式(15)で定義される.

$$q_{l+1} = q_l - \alpha \nabla E(q_l) + \gamma (q_l - q_{l-1}), \quad (15)$$

ここで,  $l$  は反復回数,  $\alpha$  は学習率 ( $\alpha > 0$ ),  $E(q)$  は損失関数,  $\gamma$  は直前のステップの寄与度 ( $0 < \gamma < 1$ ) である.

CNN を用いた分類問題としての方向推定について述べる.  $\mathbf{P}_{input}$  の 70% を学習用データ, 20% を検証用データ, 残りのデータを評価用データとして使用した. このデータセットは学習ごとに無作為に用意する. 学習アルゴリズムには SGDM, 損失関数に交差エントロピーを用い, ミニバッチサイズは 128, 学習率は 0.01, 直前のステップの寄与度を 0.9 と定義する.

## 4.3 実験

本節では, CNN を用いた方向推定のシミュレーションについて述べる. 4.3.1 項ではシミュレーションの条件について, 4.3.2 項では MUSIC 法によるシミュレーションの結果について, 4.3.3 項では CNN を用いた場合のシミュレーション結果について, 4.3.4 項では MUSIC 法による方向推定と CNN を用いた方向推定を比較した結果について述べる.

### 4.3.1 実験条件

本研究では  $K = 3$  として単一音源における音源方向推定のシミュレーションを行なう. 方向推定に用いる音源は piccolo で, 音楽制作ソフトウェア Studio One を使用して作成したものある. 音高は国際ピッチ表記法 (IPN : International Pitch Notation) における C5 ( $f_0 = 523.251\text{Hz}$ ) から C8 ( $f_0 = 4186.009\text{Hz}$ ) までの 37 種類である. piccolo

は図 4.10~4.12 のように音高が変化しても倍音構造が変化しにくく，基本周波数成分が最も多く，2 倍音，3 倍音と周波数が高くなるにつれてその成分は少なくなる．本研究では MUSIC スペクトルを学習することで推定誤差を低減できるかを検証するため，音高の変化によって周波数成分の比率が大きく変化しない piccolo を用いる．音源定位の方向は図 4.13 に示すように  $M_2$  正面から左右 90 度の範囲の単一音源（1 度間

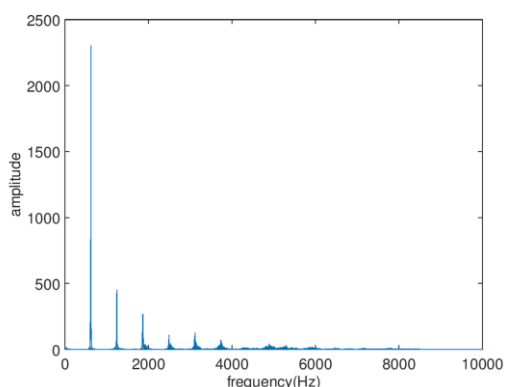


図 4.10 piccolo (D#4) の周波数特性

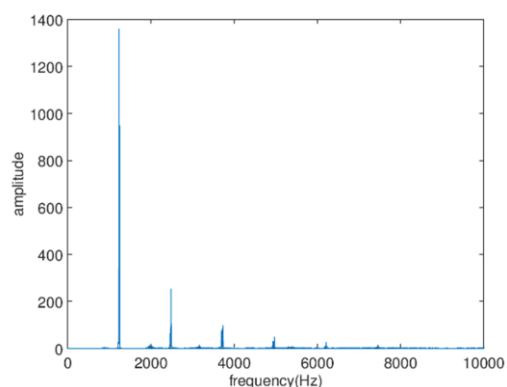


図 4.11 piccolo (D#5) の周波数特性

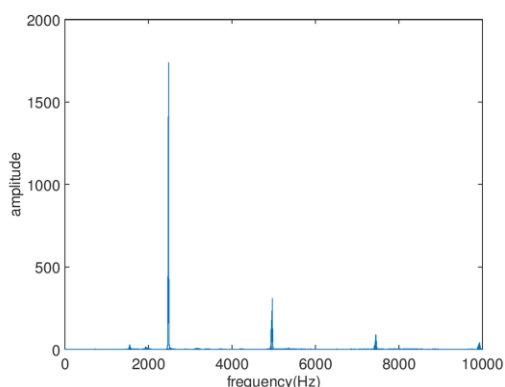


図 4.12 piccolo (D#6) の周波数特性

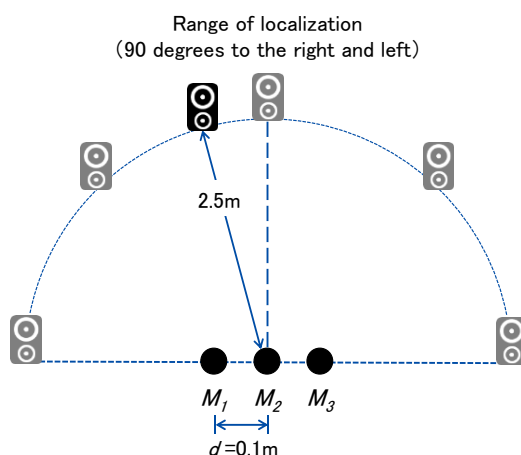


図 4.13 收音シミュレーションの概略図 2（定位方向が  $M_2$  正面から左 15 度（-15 度）にある場合）



隔の全 181 パターン) に制限する. また,  $d = 0.1\text{m}$ とし, 減衰・雑音・反響はない仮想空間上でのシングルスピーカによる放射と仮定する. CNN-R による推定結果の評価は MSE・推定誤差, CNN-C による推定結果の評価は MSE・完全一致率・推定誤差で行なう.

### 4.3.2 MUSIC 法による推定結果

MUSIC 法による推定結果 (定位方向と推定方向の一致率・MSE・推定誤差の最大) を表 4.1 及び図 4.14~4.50 に示す. 全体的な MSE は 1150.528 であり, 推定誤差の最大は 158 度であった. MSE が 20 を下回った音源では推定誤差が最大で 15 度, 定位方向と推定方向の一致率が 27~28%である. その時の推定結果の一例が図 4.48 である. また, A6 よりも高い音高では各音高での MSE が 1000 を超え, A7 を対象とした方向推定では MSE が 5141.215 となった. A7 による推定結果は図 4.17 である. 定位方向と推定方向の一致率と推定誤差の最大角も A6 よりも高い音源は著しく低下した. 収音マイクロフォン距離 $d$ が最高周波数の半波長以下とならない場合, 空間方向にサンプリング定理が満たされないため空間エイリアシング[40]が生じる. 本研究では基本周波数成分のみを用いて方向推定を行なっているため, 最高周波数は基本周波数である. 音高が G#6 のときの波長は 0.205, A6 のときの波長は 0.193 となり基本周波数が A6 よりも高い場合に空間エイリアシングが生じている. 空間エイリアシングが生じていないときの MUSIC スペクトルは図 4.51 のように複数のピークが生じていないが, 空間エイリアシングが生じる時の MUSIC スペクトルには図 4.52 に示すように複数のピークが生じる. このピークには定位方向または定位方向とほぼ等しい角度にもピークが生じているが, 定位方向とは異なる方向にあるピークのほうが大きくなることがある. 本研究においては MUSIC スペクトルが最大となる方向を推定方向としているため, MUSIC スペクトルが最も大きい方向が定位方向になるとは限らず, 推定誤差が大きくなると考えられる.

表 4.1 MUSIC 法による推定結果 (piccolo)

pitch	fundamental frequency (Hz)	match rate	MSE	maximum estimation error(deg)
C8	4186.009	13.81%	2451.293	112
B7	3951.066	14.92%	4738.619	136
A#7	3729.31	14.92%	3624.950	138
A7	3520	13.81%	5141.215	158
G#7	3322.438	14.92%	2803.149	91
G7	3135.963	18.23%	1457.127	95
F#7	2959.955	18.23%	2168.840	93
F7	2793.826	19.34%	2211.713	94
E7	2637.02	21.55%	1388.210	98
D#7	2489.016	22.65%	1165.436	103
D7	2349.318	21.55%	1647.006	109
C#7	2217.461	21.55%	3925.414	122
C7	2093.005	23.76%	2587.160	117
B6	1975.533	22.65%	4406.862	129
A#6	1864.655	25.97%	2536.641	138
A6	1760	28.18%	8.796	11
G#6	1661.219	28.18%	8.796	11
G6	1567.982	28.18%	7.812	10
F#6	1479.978	28.18%	7.812	10
F6	1396.913	28.18%	16.508	15
E6	1318.51	28.18%	16.243	15
D#6	1244.508	28.18%	16.243	15
D6	1174.659	28.18%	11.945	13
C#6	1108.731	28.18%	16.243	15
C6	1046.502	28.18%	16.541	15
B5	987.7666	27.07%	16.254	15
A#5	932.3275	27.07%	16.254	15
A5	880	27.07%	16.254	15
G#5	830.6094	27.07%	16.254	15
G5	783.9909	27.07%	16.254	15
F#5	739.9888	28.18%	16.243	15
F5	698.4565	28.18%	10.177	12
E5	659.2551	27.07%	16.254	15
D#5	622.254	28.18%	16.243	15
D5	587.3295	27.07%	16.254	15
C#5	554.3653	27.07%	16.254	15
C5	523.2511	27.07%	16.254	15

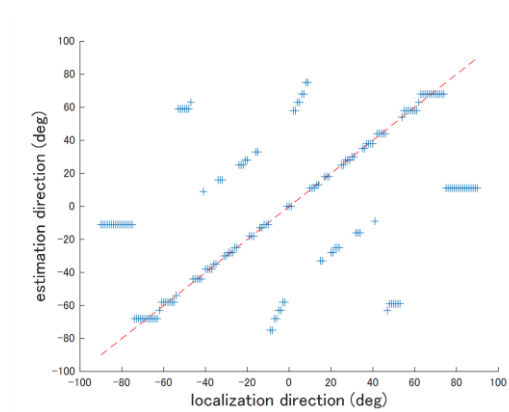


図 4.14 C8 の piccolo による推定結果 (only MUSIC)

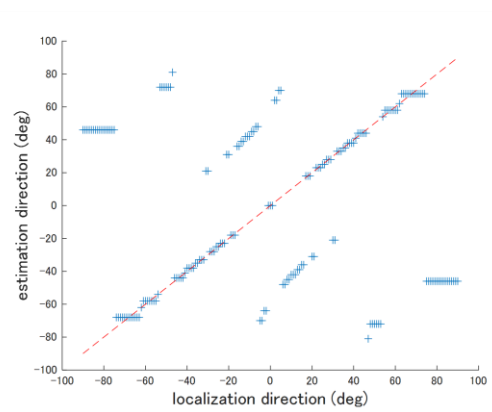


図 4.15 B7 の piccolo による推定結果 (only MUSIC)

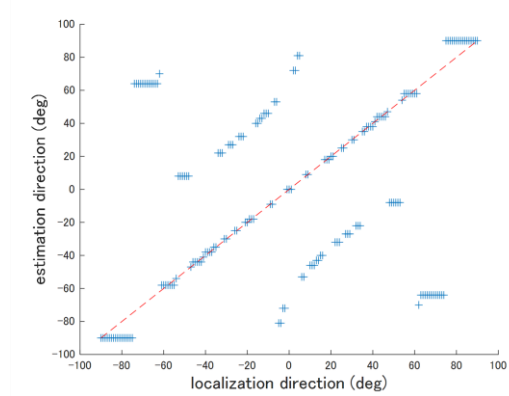


図 4.16 A#7 の piccolo による推定結果 (only MUSIC)

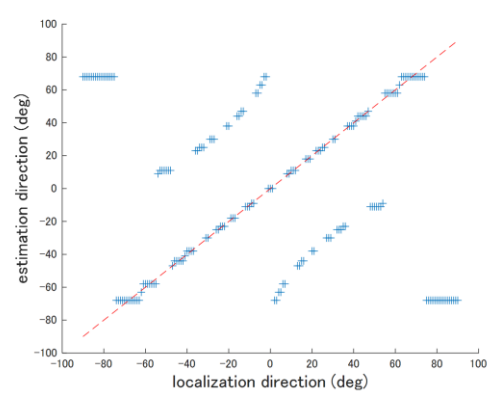


図 4.17 A7 の piccolo による推定結果 (only MUSIC)

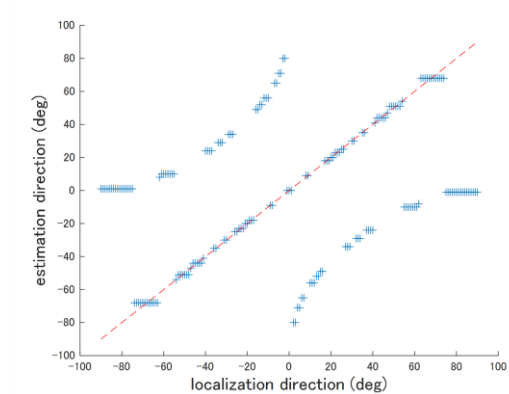


図 4.18 G#7 の piccolo による推定結果 (only MUSIC)

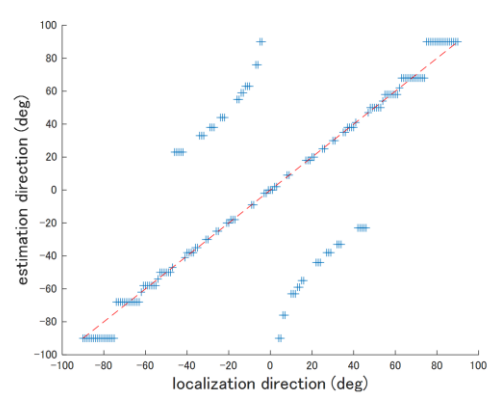


図 4.19 G7 の piccolo による推定結果 (only MUSIC)

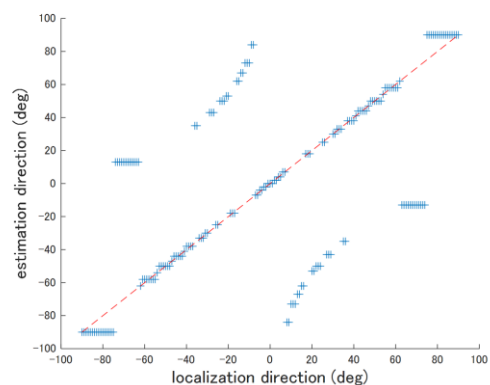


図 4.20 F#7 の piccolo による推定結果 (only MUSIC)

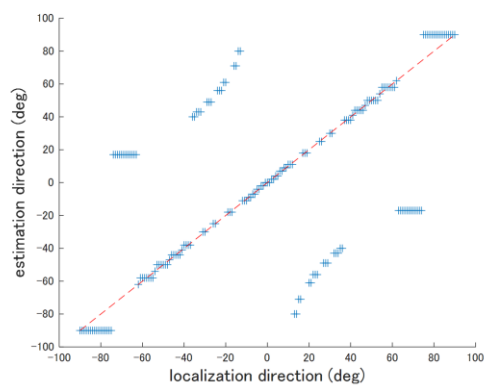


図 4.21 F7 の piccolo による推定結果 (only MUSIC)

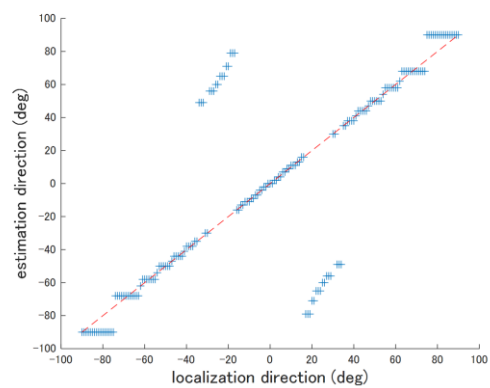


図 4.22 E7 の piccolo による推定結果 (only MUSIC)

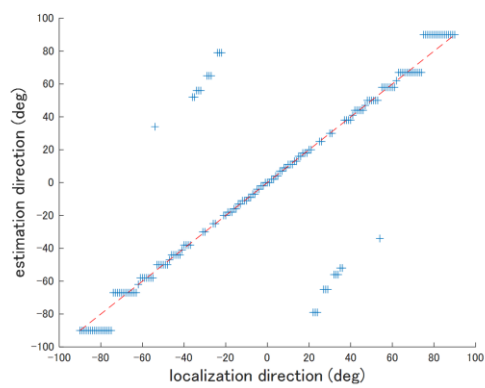


図 4.23 D#7 の piccolo による推定結果 (only MUSIC)

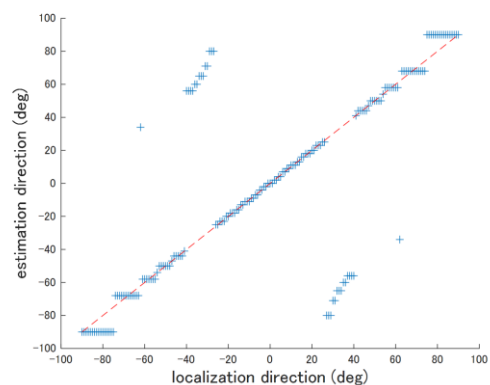


図 4.24 D7 の piccolo による推定結果 (only MUSIC)

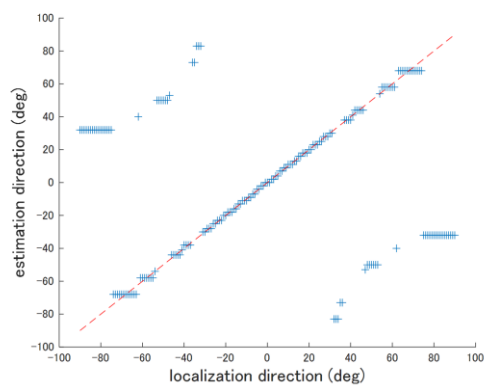


図 4.25 C#7 の piccolo による推定結果 (only MUSIC)

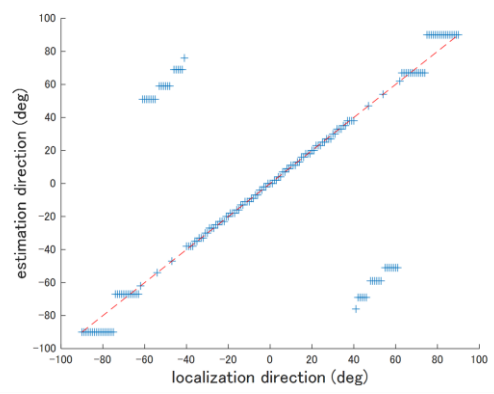


図 4.26 C7 の piccolo による推定結果 (only MUSIC)

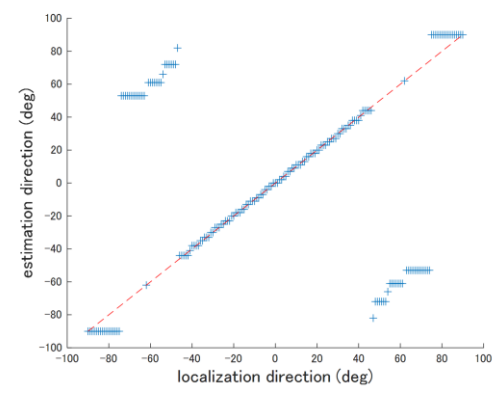


図 4.27 B6 の piccolo による推定結果 (only MUSIC)

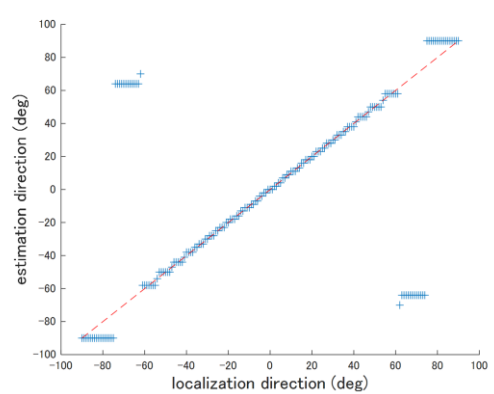


図 4.28 A#6 の piccolo による推定結果 (only MUSIC)

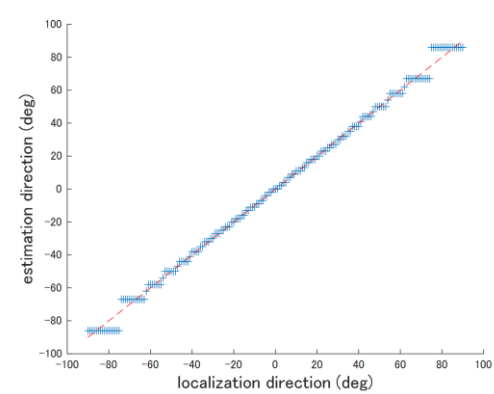


図 4.29 A6 の piccolo による推定結果 (only MUSIC)

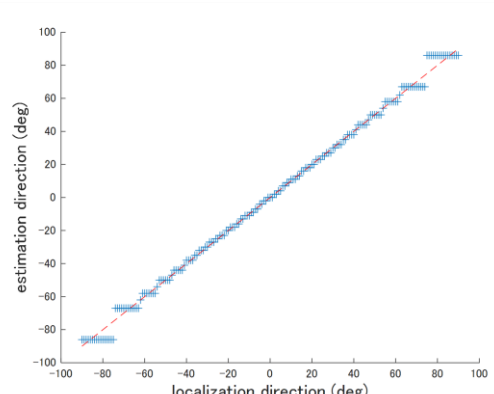


図 4.30 G#6 の piccolo による推定結果 (only MUSIC)

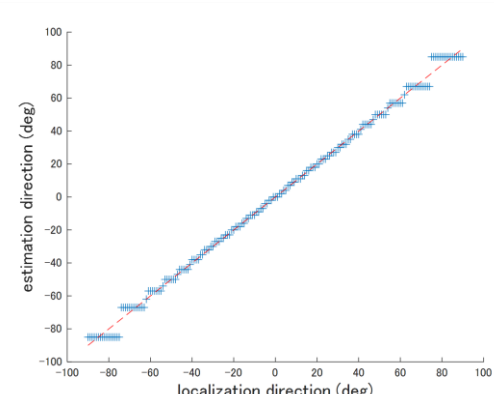


図 4.31 G6 の piccolo による推定結果 (only MUSIC)

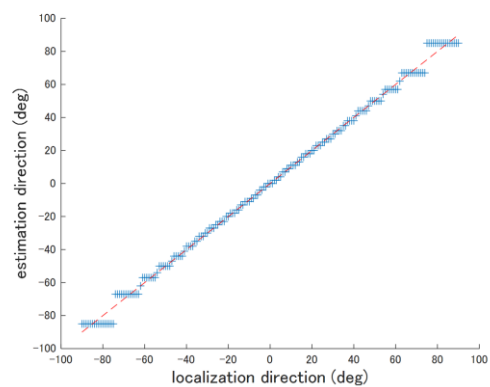


図 4.32 F#6 の piccolo による推定結果 (only MUSIC)

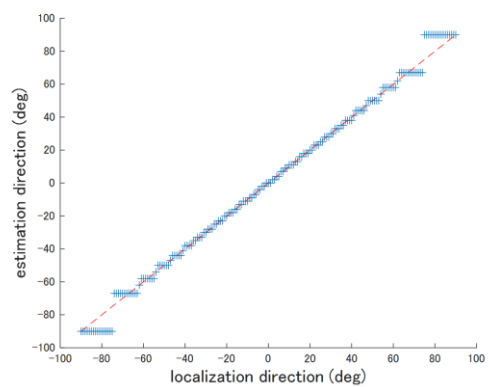


図 4.33 F6 の piccolo による推定結果 (only MUSIC)

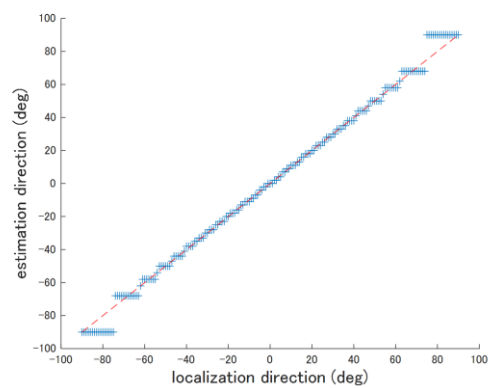


図 4.34 E6 の piccolo による推定結果 (only MUSIC)

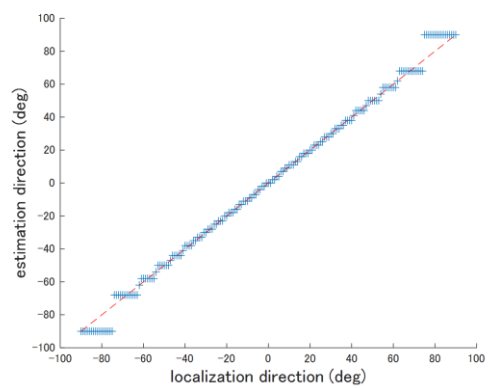


図 4.35 D#6 の piccolo による推定結果 (only MUSIC)

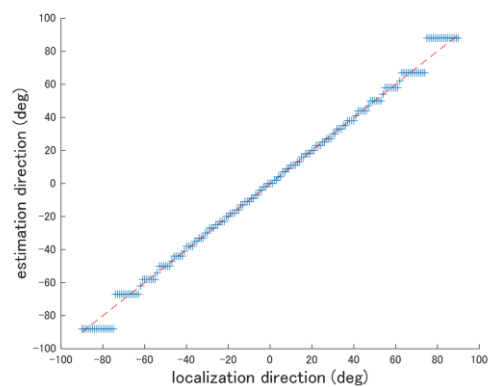


図 4.36 D6 の piccolo による推定結果 (only MUSIC)

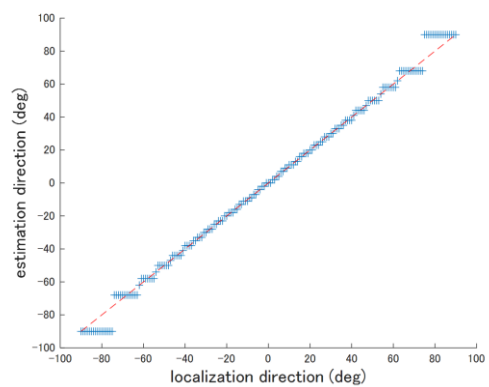


図 4.37 C#6 の piccolo による推定結果 (only MUSIC)

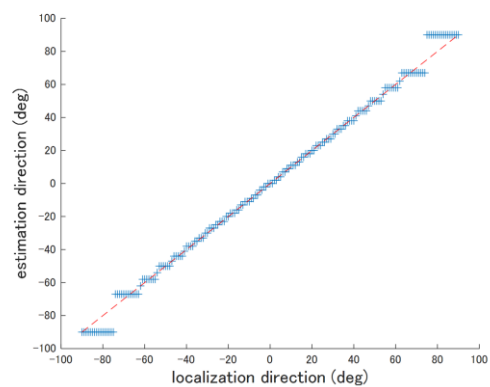


図 4.38 C6 の piccolo による推定結果 (only MUSIC)

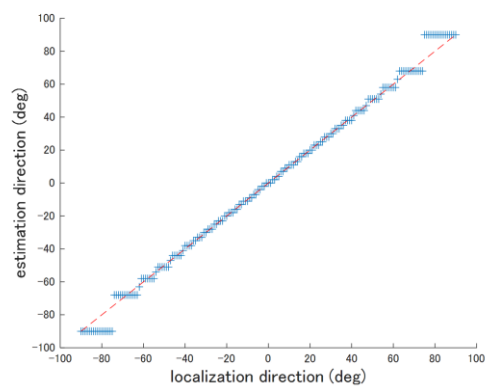


図 4.39 B5 の piccolo による推定結果 (only MUSIC)

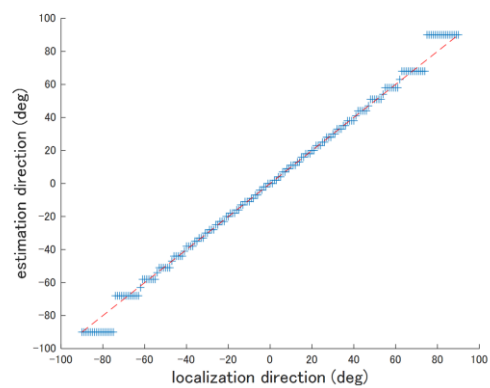


図 4.40 A#5 の piccolo による推定結果 (only MUSIC)

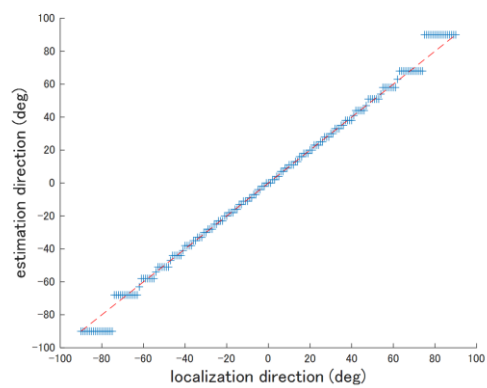


図 4.41 A5 の piccolo による推定結果 (only MUSIC)

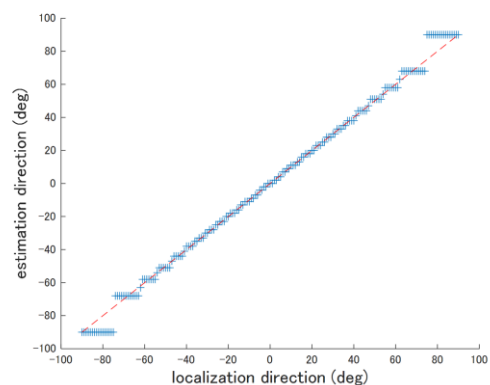


図 4.42 G#5 の piccolo による推定結果 (only MUSIC)

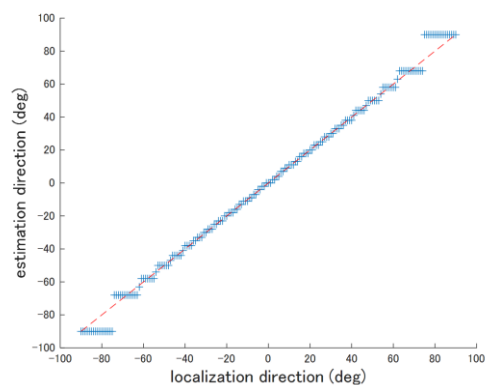


図 4.43 G5 の piccolo による推定結果 (only MUSIC)

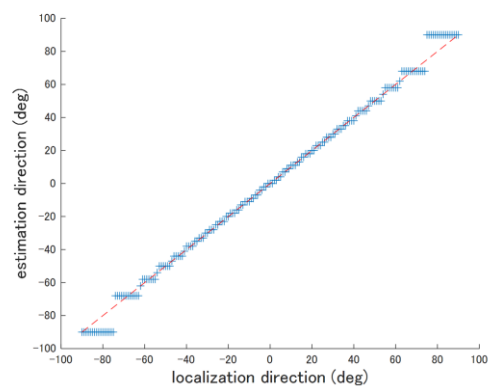


図 4.44 F#5 の piccolo による推定結果 (only MUSIC)

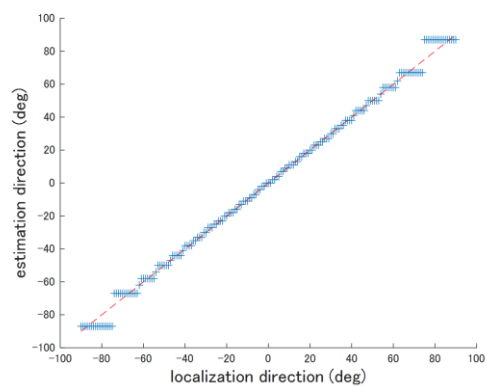


図 4.45 F5 の piccolo による推定結果 (only MUSIC)

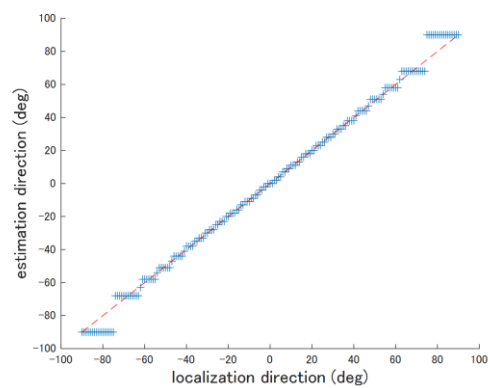


図 4.46 E5 の piccolo による推定結果 (only MUSIC)

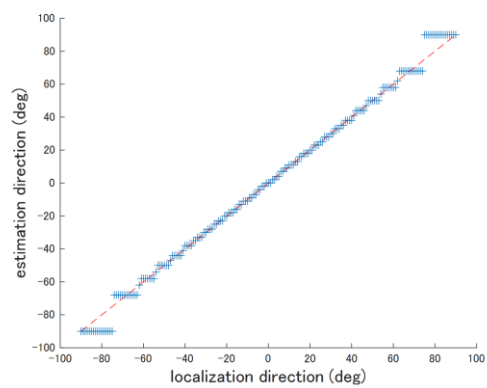


図 4.47 D#5 の piccolo による推定結果 (only MUSIC)

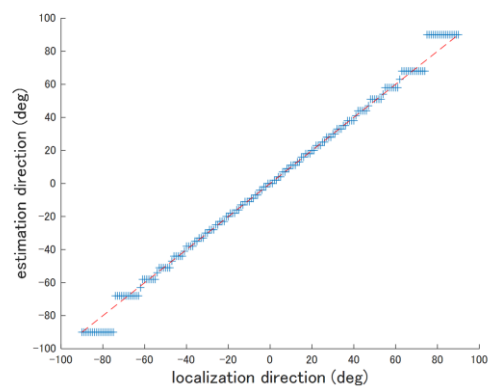


図 4.48 D5 の piccolo による推定結果 (only MUSIC)

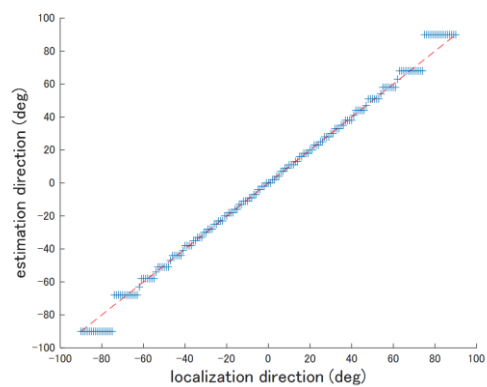


図 4.49 C#5 の piccolo による推定結果 (only MUSIC)



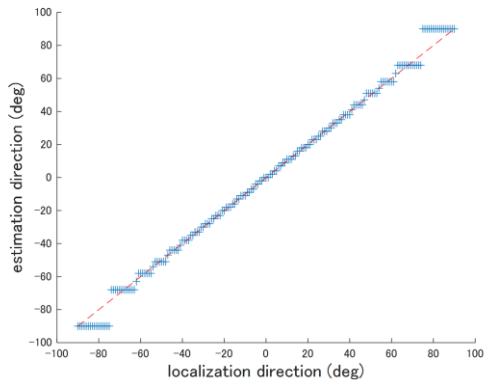


図 4.50 C5 の piccolo による推定結果 (only MUSIC)

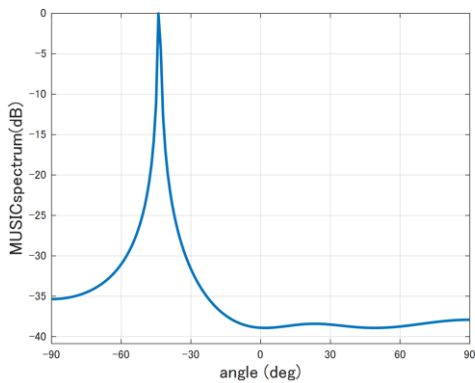


図 4.51 G6 を用いた場合の MUSIC スペクトル (定位方向: -45 度, 推定方向: -44 度)

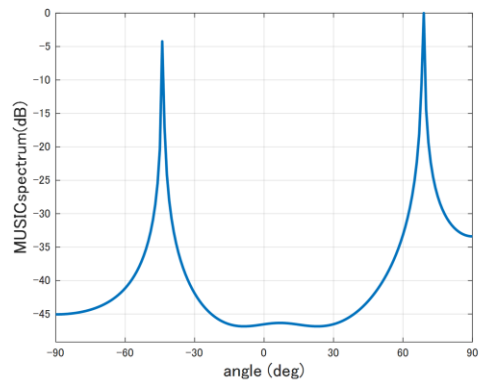


図 4.52 C7 を用いた場合の MUSIC スペクトル (定位方向: -45 度, 推定方向: 69 度)

### 4.3.3 CNN による結果

表 4.2 は回帰分析における各 CNN による定位方向と推定方向に対する MSE と推定誤差の最大値で, 図 4.53~4.70 に推定結果を示す. dataset-1 において, 最も MSE が小さい CNN の構造はミニバッチサイズ 256 の CNN-R3 で, その MSE は 208.407 である. また, 推定誤差の最大値が最も小さい CNN はミニバッチサイズ 128 の CNN-R1 で, 推定誤差は 57.864 度である. dataset-2 において, 最も MSE が小さい CNN の構造はミニバッチサイズ 256 の CNN-R2 で, その MSE は 180.400 である. また, 推定誤差の最大値が最も小さい CNN はミニバッチサイズ 128 の CNN-R2 で, 推定誤差は 57.741 度である. ミニバッチサイズが小さい場合, 高い基本周波数での推定結果の影響を受けることにより, 図 4.53 のように定位方向が  $M_2$  正面から離れるにつれて推定誤差は大きくなった. また, ミニバッチサイズが大きい場合, 学習時に特徴量が平均化され, 図 4.59 のように全体的に推定誤差が生じる. 本研究では, ミニバッチサイズ

表 4.2 回帰分析における CNN を用いた場合の MSE と推定誤差の最大値

Structure	mini-batch size	dataset-1		dataset-2	
		MSE	maximum estimation error (deg)	MSE	maximum estimation error(deg)
CNN-R1	128	295.615	57.864	330.297	60.861
CNN-R2		378.603	65.915	301.741	57.741
CNN-R3		331.985	63.771	466.929	59.325
CNN-R1	256	218.686	69.301	213.790	63.850
CNN-R2		270.225	63.413	180.400	62.253
CNN-R3		208.407	68.272	215.232	61.092
CNN-R1	512	261.361	74.511	320.728	75.192
CNN-R2		247.871	70.049	237.138	63.370
CNN-R3		240.376	70.002	235.510	66.138

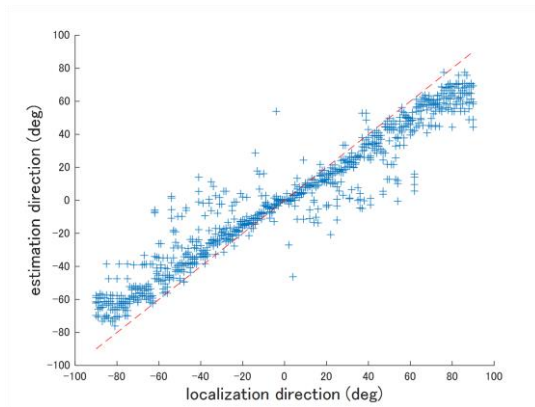


図 4.53 CNN-R1 による推定結果 (dataset-1, mini-batch size = 128)

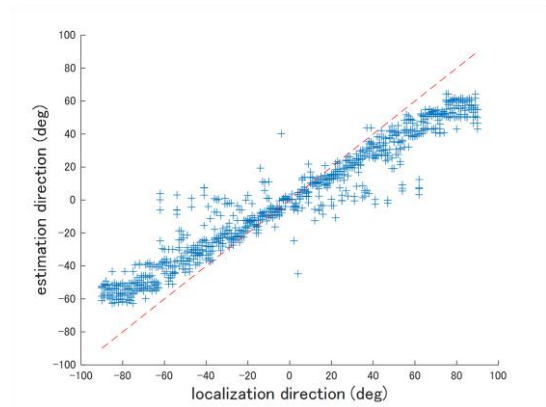


図 4.54 CNN-R2 による推定結果 (dataset-1, mini-batch size = 128)

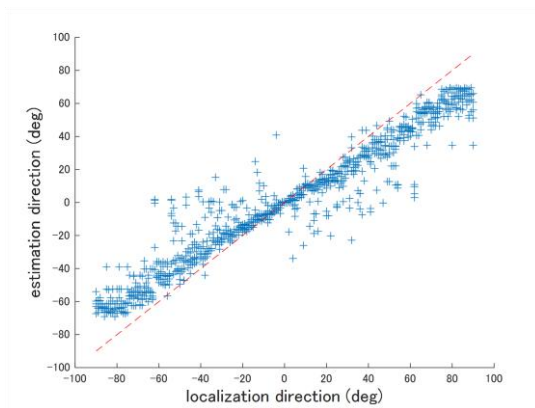


図 4.55 CNN-R3 による推定結果 (dataset-1, mini-batch size = 128)

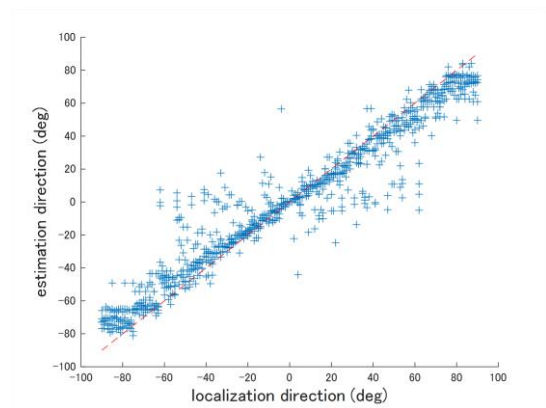


図 4.56 CNN-R1 による推定結果 (dataset-1, mini-batch size = 256)

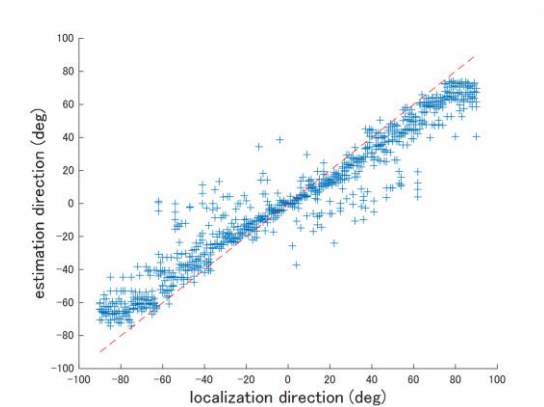


図 4.57 CNN-R2 による推定結果  
(dataset-1, mini-batch size = 256)

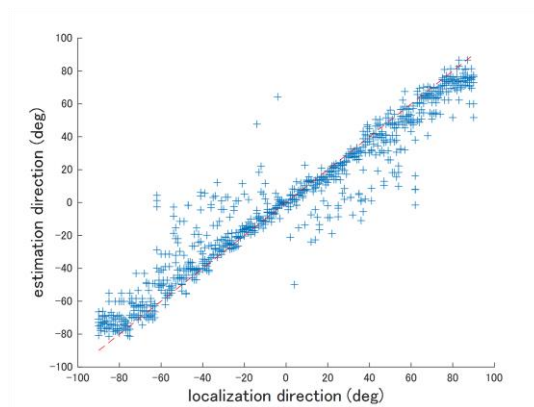


図 4.58 CNN-R3 による推定結果  
(dataset-1, mini-batch size = 256)

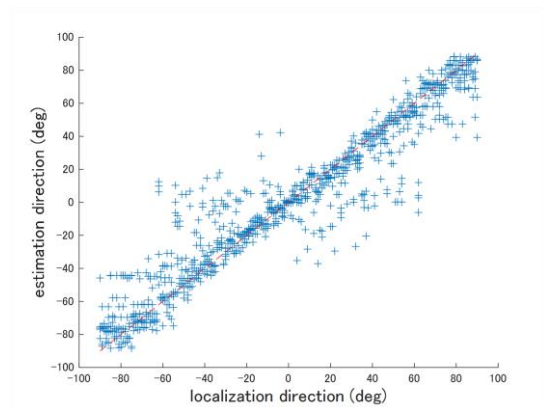


図 4.59 CNN-R1 による推定結果  
(dataset-1, mini-batch size = 512)

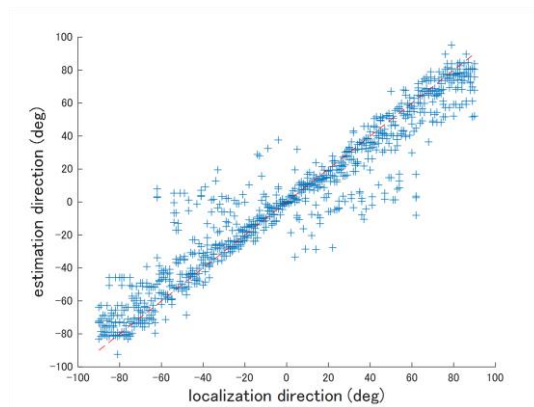


図 4.60 CNN-R2 による推定結果  
(dataset-1, mini-batch size = 512)

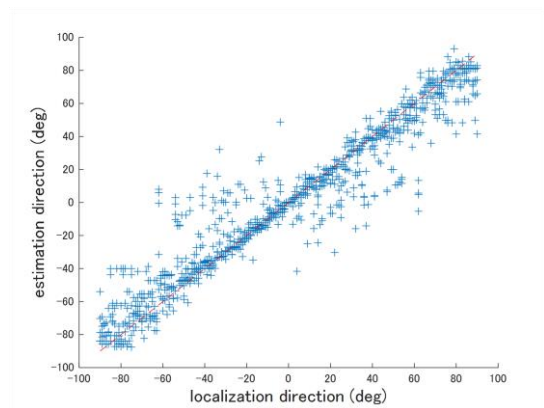


図 4.61 CNN-R3 による推定結果  
(dataset-1, mini-batch size = 512)

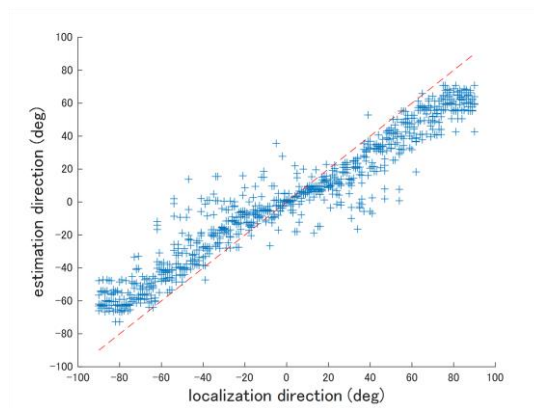


図 4.62 CNN-R1 による推定結果  
(dataset-2, mini-batch size = 128)

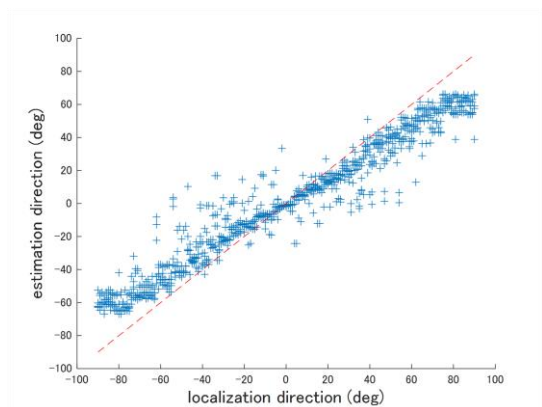


図 4.63 CNN-R2 による推定結果  
(dataset-2, mini-batch size = 128)

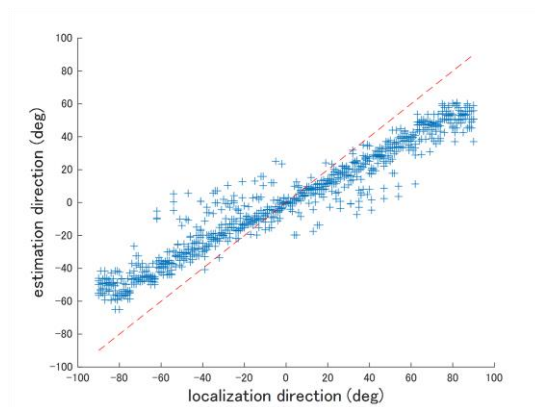


図 4.64 CNN-R3 による推定結果  
(dataset-2, mini-batch size = 128)

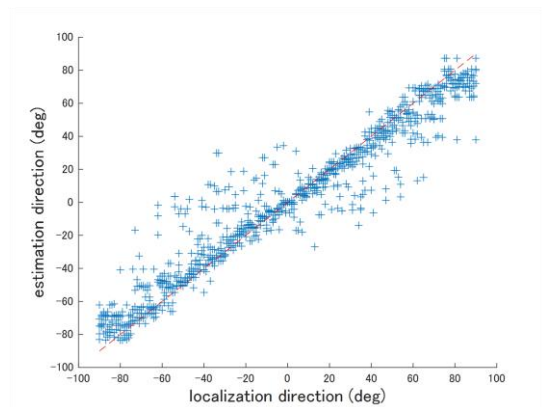


図 4.65 CNN-R1 による推定結果  
(dataset-2, mini-batch size = 256)

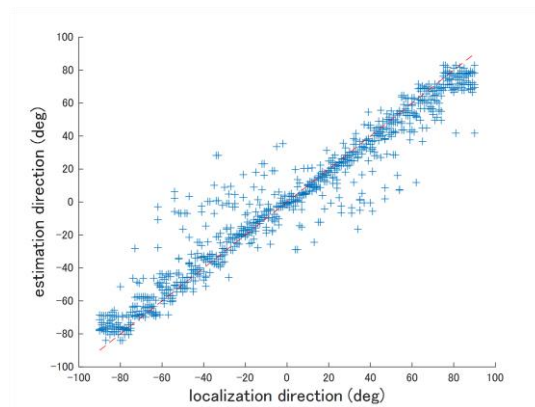


図 4.66 CNN-R2 による推定結果  
(dataset-2, mini-batch size = 256)

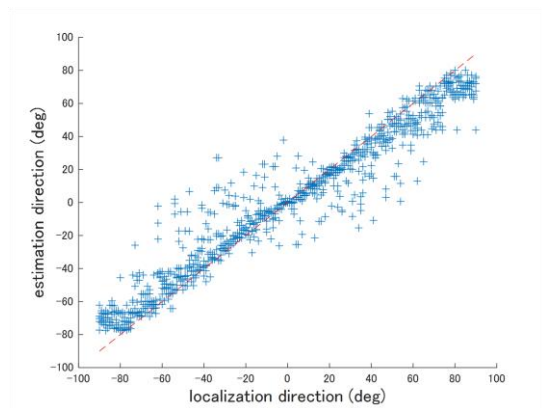


図 4.67 CNN-R3 による推定結果  
(dataset-2, mini-batch size = 256)

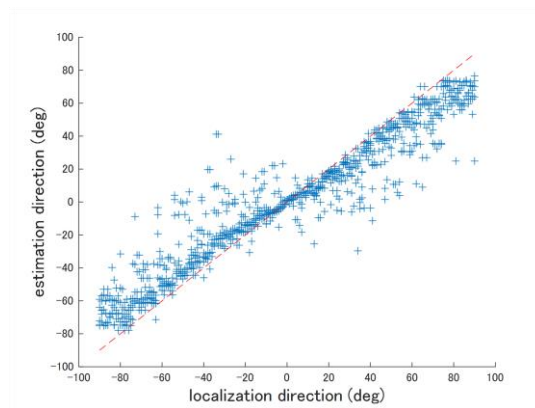


図 4.68 CNN-R1 による推定結果  
(dataset-2, mini-batch size = 512)

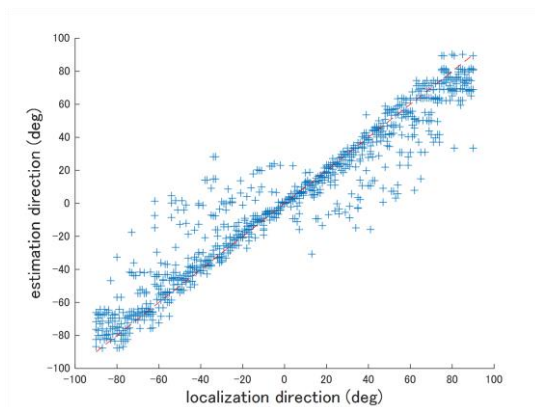


図 4.69 CNN-R2 による推定結果  
(dataset-2, mini-batch size = 512)

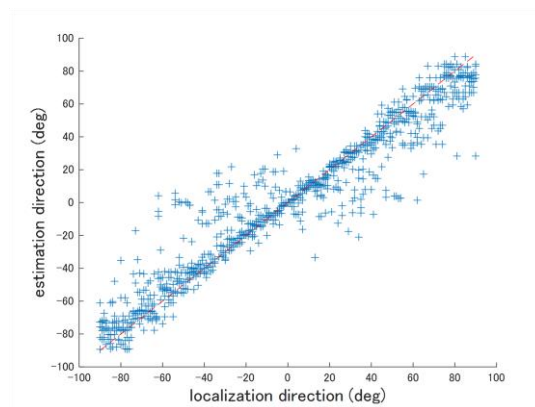


図 4.70 CNN-R3 による推定結果  
(dataset-2, mini-batch size = 512)

表 4.3 CNN を用いた分類問題における MSE と推定誤差の最大値

Structure	match rate	MSE	maximum estimation error (deg)
CNN-C1	20.99%	168.273	121
CNN-C2	17.02%	295.422	120
CNN-C3	19.67%	158.361	108
CNN-C4	18.79%	121.946	101

が 512 の場合, 全体的に推定誤差が生じ, 推定誤差が他のミニバッチサイズよりも大きくなった. また, ミニバッチサイズが 128 の場合, 推定誤差が小さくなる傾向があり, ミニバッチサイズが 256 の場合, MSE が小さくなる傾向となった.

表 4.3 は CNN を用いた分類問題における定位方向と推定方向の一致率, 定位方向と推定方向に対する MSE, 推定誤差の最大値である. 4 種類の CNN の中で最も MSE が小さくなった CNN-C4 の場合, 全評価データの 2.0% で推定誤差が 20 度よりも大きくなった. 例えば, D#7 音源において, 定位方向が 62 度であるとき, CNN を用いた場合の推定方向は -30 度であった. このとき, 入力として用いた MUSIC スペクトルは図 4.71 である. この MUSIC スペクトルはほかの音源の定位方向が 60 度の時の MUSIC スペクトルよりも D#7 音源を含めたほかの音源の -30 度の時の MUSIC スペクトルにほぼ等しい (図 4.72~4.74). したがって定位方向と異なる方向を推定方向と算出したと考えられる.

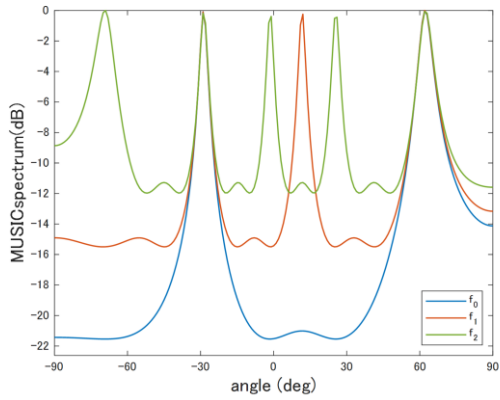


図 4.71 D#7 音源における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向 : 62 度)

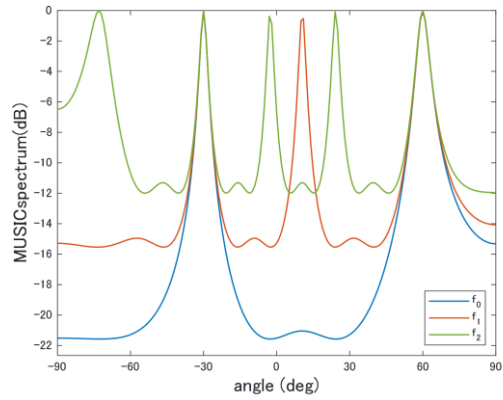


図 4.72 D#7 音源における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向 : -30 度)

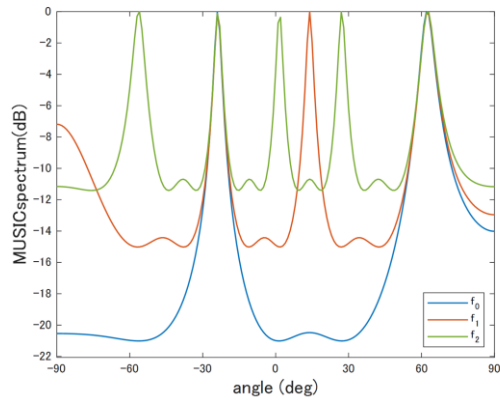


図 4.73 E7 音源における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向 : 62 度)

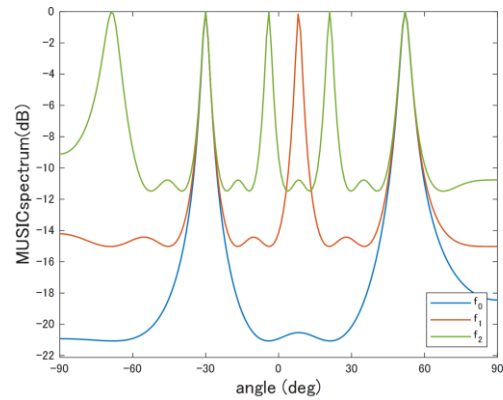


図 4.74 E7 音源における基本周波数, 第 2 高調波, 第 3 高調波成分による MUSIC スペクトル (定位方向 : -30 度)



#### 4.3.4 MUSIC 法による推定結果との比較

まず，回帰分析における CNN を用いた推定結果と MUSIC 法を用いた推定結果を比較した結果について述べる．表 4.4 に CNN（回帰）と MUSIC 法の MSE と推定誤差の最大値，図 4.75, 4.76 に CNN（回帰）と MUSIC 法の推定誤差を示す．なお，比較対象となる CNN はミニバッチサイズが 128 と 256 の CNN-R1, CNN-R2, CNN-R3 である．全ての CNN において MSE は改善された．また，推定誤差の最大角度は 60 度であり，これも MUSIC 法のみによる推定結果と比較しても改善がみられる．図 4.75 は dataset-1 を CNN-R1 により学習・評価したときの推定誤差であるが，推定誤差が 30 度以上のサンプルの割合は，MUSIC 法のみの場合には 13.6%，CNN-R1 を用いた場合には 8.5%となった．また，推定誤差が 60 度以上の割合は MUSIC 法のみの場合で 11.6%，CNN-R1 を用いた場合で 0%となった．同様に，図 4.76 に示した dataset-2 を

表 4.4 回帰分析における CNN を用いた推定結果と MUSIC 法を用いた推定結果比較

Structure	mini-batch size	dataset-1		dataset-2	
		MSE	maximum estimation error (deg)	MSE	maximum estimation error (deg)
MUSIC only		1096.343	158	1230.896	158
CNN-R1	128	295.615	57.864	330.297	60.861
CNN-R2		378.603	65.915	301.741	57.741
CNN-R3		331.985	63.771	466.929	59.325
CNN-R1	256	266.451	65.224	213.790	63.850
CNN-R2		270.225	63.413	180.400	62.253
CNN-R3		208.407	68.272	215.232	61.092

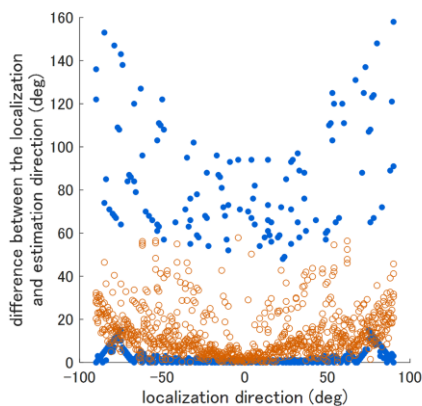


図 4.75 CNN-R1 (mini-batch size = 128) と MUSIC 法による推定誤差 (dataset-1)

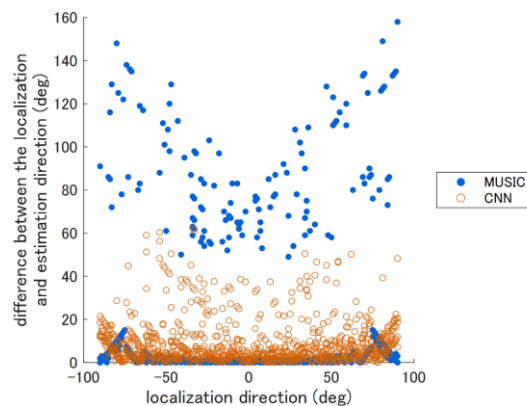


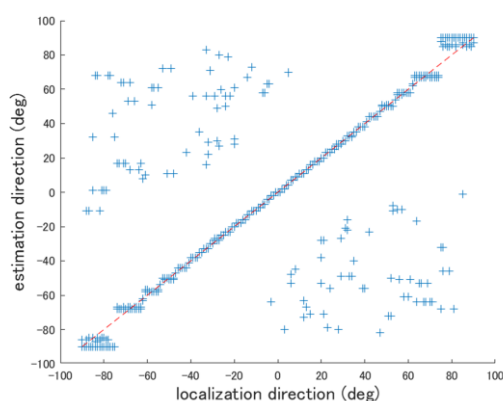
図 4.76 CNN-R2 (mini-batch size = 256) と MUSIC 法による推定誤差 (dataset-2)

CNN-R2 により学習・評価した場合、推定誤差が 30 度以上となる割合は MUSIC 法のみの場合には 14.4%，CNN-R2 を用いた場合には 5.9%であり、推定誤差が 60 度以上となる割合は MUSIC 法のみの場合には 12.9%，CNN-R2 を用いた場合には 0.3%となった。MUSIC 法による推定誤差の大きいサンプルが CNN を用いることにより減少していることがわかる。

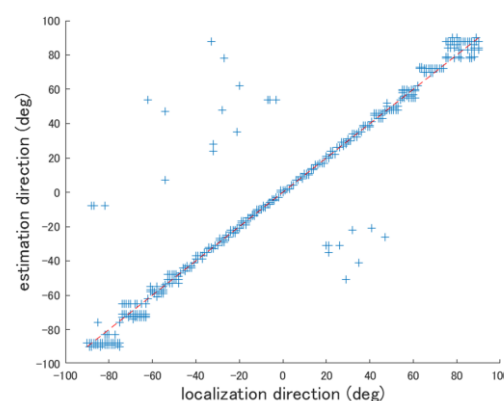
次に、CNN を用いた分類問題における推定結果と MUSIC 法を用いた推定結果を比較した結果について述べる。表 4.5 に CNN（分類）と MUSIC 法の MSE と推定誤差の最大値，図 4.77~4.80 に CNN-C と MUSIC 法の推定結果を示す。すべての CNN において MSE が改善された。また、CNN を用いた場合の推定誤差の最大は 121 度であり、MUSIC 法のみの場合よりも改善が見られる。例えば、CNN-C4 では、推定誤差が 10 度以上であるサンプルの割合は、MUSIC 法のみの場合には 17.5%，CNN-C4 を用いた場合は 5.3%であり、推定誤差が 20 度以上の割合は MUSIC 法のみの場合には 12.9%，CNN-C4 を用いた場合には 2.0%である。しかし、MUSIC 法を用いて推定した方向が正確な場合でも CNN を用いることで誤った方向を算出することがあり、MUSIC 法に

表 4.5 CNN を用いた分類問題による推定結果と MUSIC 法を用いた推定結果の比較

Structure	only MUSIC		CNN	
	MSE	maximum estimation error (deg)	MSE	maximum estimation error (deg)
CNN-C1	1230.256	152	168.273	121
CNN-C2	1317.546	152	295.422	120
CNN-C3	1218.034	158	151.361	108
CNN-C4	1136.196	157	121.946	101



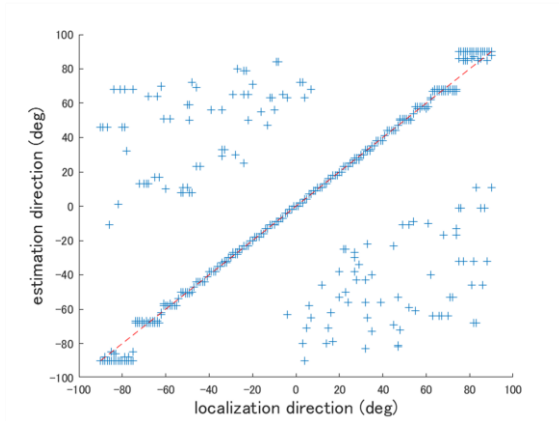
(a) MUSIC 法による推定結果



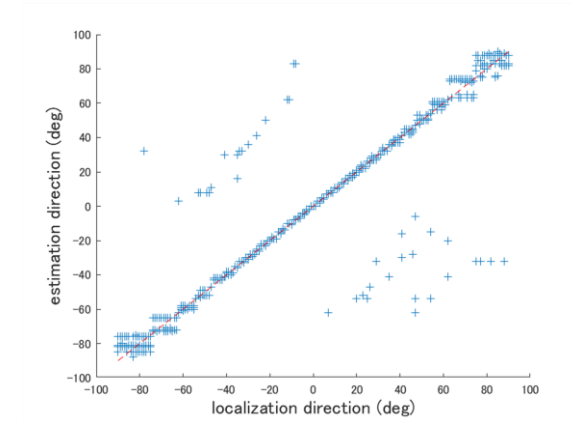
(b) CNN-C1 を用いた場合の推定結果

図 4.77 CNN-C1 を用いた場合の推定結果とその評価用データにおける MUSIC 法による推定結果



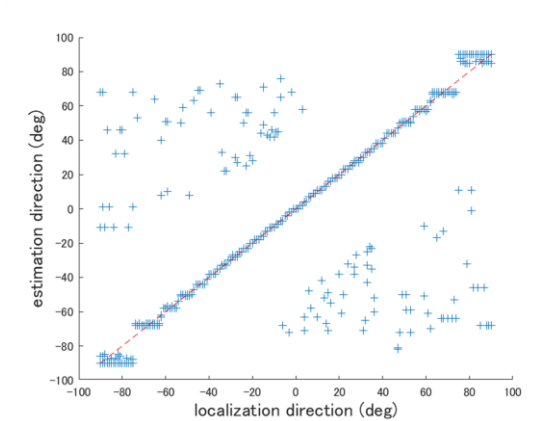


(a) MUSIC 法による推定結果

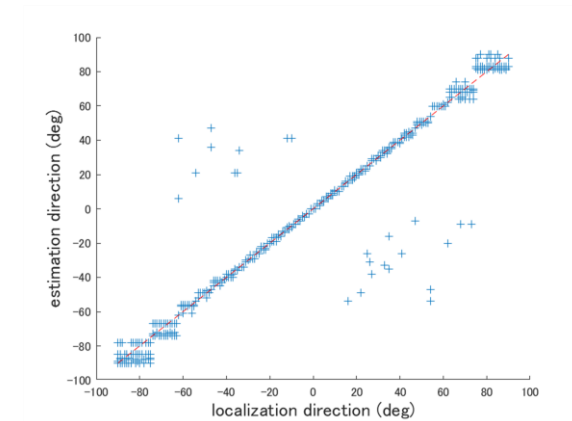


(b) CNN-C2 を用いた場合の推定結果

図 4.78 CNN-C2 を用いた場合の推定結果とその評価用データにおける MUSIC 法による推定結果

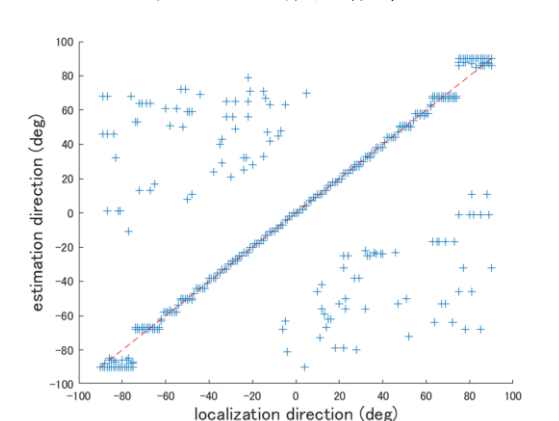


(a) MUSIC 法による推定結果

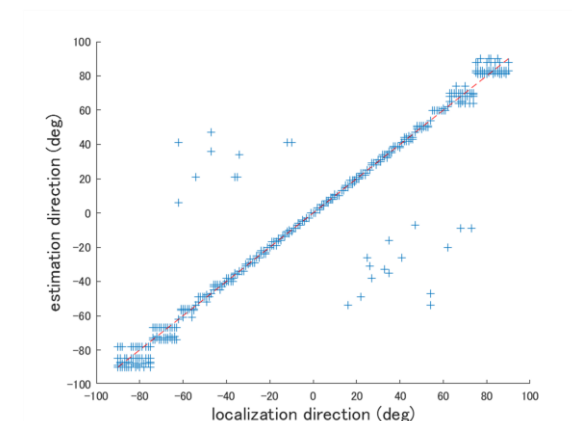


(b) CNN-C3 を用いた場合の推定結果

図 4.79 CNN-C3 を用いた場合の推定結果とその評価用データにおける MUSIC 法による推定結果



(a) MUSIC 法による推定結果



(b) CNN-C4 を用いた場合の推定結果

図 4.80 CNN-C4 を用いた場合の推定結果とその評価用データにおける MUSIC 法による推定結果

における推定方向の信頼度を用いることで MUSIC 法と CNN を組み合わせた手法[39]によってより精度の高い方向推定が可能になると考えられる。

## 4.4 結言

本章では、基本周波数が高い音響信号を対象とした方向推定における推定誤差の低減を目的とした MUSIC スペクトルを入力とする CNN を用いた方向推定手法について述べた。シミュレーションの結果、回帰・分類どちらにおいても CNN を用いることで MSE と推定誤差が改善したことが示された。しかし、大きな推定誤差が生じる場合や MUSIC スペクトルが類似することにより MUSIC 法を用いて推定した方向が正確な場合でも CNN を用いることで誤った方向を算出する場合がある。したがって MUSIC 法と CNN の複合手法によってより精度の高い方向推定が可能であると考えられる。また、本研究においては基本周波数の違いによって倍音構造が大きく変化しない信号を対象としたため、本研究で用いた piccolo 以外の楽器音を対象とした場合での方向推定についても検討する必要がある。また、時系列データを扱っているため、回帰型ニューラルネットワーク (RNN: Recurrent Neural Network) や LSTM: Long Short Term Memory などのネットワークの利用が考えられる。

## 第5章 結論

サラウンド音響は映画館などでは古くから導入されているが、近年ではテレビ放送やDVDなどのデジタルコンテンツもサラウンド音響に対応し、家庭でも楽しめるようになりつつあるが、サラウンド音響システムの使用時には「使用者が所望する音響空間を構築することの困難さ」や「部屋の形状などによる音像バランスの乱れ」などの問題が発生する場合が多い。音は可視化することによって音の定位や分布を認知しやすくなり、サラウンド空間をタブレットなどの端末に表示して定位を調整するシステムを構築することにより、サラウンド音響の使用者に適した調整を行なうことができるようになる。そこで、サラウンド音響を可視化することで音源定位の調整を一般の使用者が容易に行なうことができるシステムの構築を目標とする。しかし、サラウンド音源は使用する環境によって異なる反響が生じ、使用者が所望する音響空間を構築するためには使用者が聴取する実際の音響空間での仮想音源の位置推定が必要となる。本論文では、仮想音源の位置推定の基礎的研究として単一の楽器音源における音源方向推定のシミュレーションについて述べた。

第2章ではDOAの一種であるMUSIC法による楽器音の方向推定手法として基本周波数成分のみを用いる手法について述べた。倍音構造をもつ楽器音を対象とした音源方向推定では、振幅値の大きい周波数成分が分析対象信号に複数存在することで推定誤差が大きくなり、MSEが大きくなる。そこで振幅値の比較的大きい基本周波数成分のみを用いることで推定精度の向上を図った。シミュレーションにより、MSEは概ね1.0を下回った。また、特に基本周波数が高い場合にMSEが大きくなることが示された。全体的には推定精度が向上しているが、シミュレーションにおける定位方向、分析対象信号の音色や基本周波数などの条件が限定されているため、実験条件に限定されない方向推定手法の検討が必要である。

第3章では楽器音の高調波成分を用いた音源方向推定としてFFNNを用いた手法について述べ、楽器や周波数ごとの誤りパターンを学習させることで推定精度向上を図った。アナログ出力であるFFNN-AOを用いた場合、特に基本周波数が高いときにMSEが改善傾向であることが示された。デジタル出力であるFFNN-DOを用いた場合は、学習すべきパラメータが多くなることにより十分な学習を行なうことができず、MUSIC法による推定結果と比較するとMSEは増加しているが、音源の違いによる推定結果のばらつきは最も小さかった。学習用データ数の増加やネットワークの構造を変更させることが今後の課題として挙げられる。

第4章では分析対象信号の基本周波数が高い場合の音源方向推定手法について述べた。基本周波数が高い信号を対象としたMUSIC法による方向推定では推定精度が大きく低下する。MUSIC法により得られるMUSICスペクトルを入力とするCNN(回帰・分類)を用いることで推定誤差の低減を図った。シミュレーションの結果、回帰・

分類どちらにおいても CNN を用いることで MSE と推定誤差が改善したことが示されたが、大きな推定誤差が生じる場合や MUSIC スペクトルが類似することにより MUSIC 法を用いて推定した方向が正確な場合でも CNN を用いることで誤った方向を算出する場合がある。したがって MUSIC 法と CNN の複合手法によってより精度の高い方向推定が可能であると考えられる。また、本研究においては基本周波数の違いによって倍音構造が大きく変化しない信号を対象としたため、本研究で用いた piccolo 以外の楽器音を対象とした場合での方向推定についても検討する必要がある。

## 参考文献

- [1] 濱崎公男, “マルチチャンネル音声収録とデジタル放送,” 日本音響学会誌, vol.57, no.9, pp.610-616, 2001.
- [2] Recommendation ITU-R BS.775-1, International Telecommunication Union, 1994.
- [3] 濱崎公男, “マルチチャンネル音響,” 日本音響学会誌, vol.67, no.11, pp.526-531, 2011.
- [4] 森千晶, 西口敏行, 小野一穂, “22.2 マルチチャンネル音響における三次元残響負荷装置の開発,” 映像情報メディア学会誌, vol.73, no.1, pp.155-160, 2019.
- [5] 小野一穂, “マルチチャンネルオーディオ,” 映像情報メディア学会誌, vol.68, no.8, pp.604-607, 2014.
- [6] 安藤彰男, “音響の高臨場感技術,” 映像情報メディア学会誌, vol.66, no.8, pp.671-677, 2012.
- [7] 荻原昭夫, 村田晴美, 岡田名留美, 村本翔, 藤原礼奈, 山本華帆, “サラウンド音響の視覚化に関する検討,” 2018 年電子情報通信学会総合大会講演論文集, D-21-3, 2018.
- [8] 山本華帆, 荻原昭夫, 村田晴美, “サラウンド音響における色を用いた可視化の検討,” 平成 30 年度(第 69 回)電気・情報関連学会中国支部連合大会講演論文集, R18-08-04, 2018.
- [9] 貞廣諒, 山本華帆, 荻原昭夫, “深層クラスタリング法を用いた楽器音の音源分離,” 2022 年度(第 73 回)電気・情報関連学会中国支部連合大会講演論文集, R22-08-04, 2022.
- [10] 橘秀樹, 日高新人, “チェロの音響放射特性の可視化,” 日本音響学会誌, vol.46, no.10, pp.864-866, 1990.
- [11] 平賀瑠美, 五十嵐滋, 松浦陽平, “総合演奏視覚化システム,” 情報処理学会論文誌, vol.38, no.11, pp.2391-2397, 1997.
- [12] 金子弘美, 高島翠, 野口薫, “音刺激と視覚的図形に関する基礎的研究,” 日本色彩学会誌, vol.29, pp.138-139, 2005.
- [13] 飯田一嘉, “音場の可視化とその動向ーマイクロホンスイープ法からビームフォーミング法までー,” 可視化情報学会誌, vol.27, no.104, pp.3-8, 2007.
- [14] 森大星, 荻原昭夫, 井田真人, 山本華帆, 宮西宏和, 村田晴美, “非接触バイタルセンシングを用いた高揚度の推定及び可視化,” 2020 年度(第 71 回)電気・情報関連学会中国支部連合大会講演論文集, R20-08-01-02, 24 2020.
- [15] 山本華帆, 荻原昭夫, 村田晴美, “サラウンド音響の可視化手法と仮想音源方向推定の検討,” 2019 年電子情報通信学会総合大会講演論文集, D-21-4, 2019.
- [16] 山本華帆, 荻原昭夫, 村田晴美, “サラウンド音響における MUSIC 法を用いた仮想

- 音源位置の方向推定に関する検討,” 第 18 回情報科学技術フォーラム講演論文集, K-010, 3 2019.
- [17]山本華帆, 荻原昭夫, 村田晴美, “サラウンド音響の可視化を目的とした MUSIC 法による仮想音源の方向推定,” 令和元年度(第 70 回)電気・情報関連学会中国支部連合大会講演論文集, R19-08-01-05, 2019.
- [18]山本潔, 浅野太, 原功, 緒方淳, 麻生英樹, 山田武志, 北脇信彦, “ヒューマノイドロボット HRP-2 における音響情報と画像情報を統合したリアルタイム音声インタフェース,” 日本音響学会誌, vol.62, no.3, pp.161-172, 2006.
- [19]畝田道雄, 近藤容章, 石川憲一, 大西修, 黒河周平, 土肥俊郎, “空間移動平均法と MUSIC 法の併用による強相関複数騒音源の位置同定に関する研究 -小型マイクロホンアレーシステムを指向した高精度位置同定法-,” 精密工学会誌, vol.77, no.12, pp.1158-1164, 2011.
- [20]鈴木敬, 金田豊, “サブバンドピークホールド処理を用いた音源方向推定法,” 日本音響学会誌, vol.65, no.10, pp.513-522, 2009.
- [21]R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Transactions on Antennas and Propagation, vol.34, no.3, pp.276-280, 1986.
- [22]R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” IEEE Trans. Acoustics, Speech, and Signal Processing, vol.ASSP-37, no.7, pp.984-995, 1989.
- [23]大塚琢馬, 中臺一博, 尾形哲也, 奥乃博, “MUSIC 法を用いた音源定位のベイズ拡張,” 第 29 回日本ロボット学会学術講演論文集, RSJ2011AC33-2, 2011.
- [24]菊間信良, “アレーアンテナによる適応信号処理,” 科学技術出版, 1998.
- [25]森和義, 横山智樹, 長谷川秋雄, “周囲雑音を用いた無音物体探知における高分解能アレイ信号処理手法の比較,” 海洋音響学会誌, vol.32, no.2, pp.89-97, 2005.
- [26]畝田道雄, 石川憲一, “MUSIC による近距離音源の高分解能位置推定法に関する研究,” 精密工学会誌, vol.70, no.8, pp.1111-1116, 2004.
- [27]山本華帆, 荻原昭夫, 村田晴美, “楽器音における倍音構造に着目した MUSIC 法と FFNN による音源方向の推定,” 電子情報通信学会論文誌 D, vol.J104-D, no.10, pp.780-783, 2021.
- [28]山本華帆, 荻原昭夫, 村田晴美, “楽器音源を対象とした方向推定手法の検討および比較,” 信学技報, SP2022-3, pp.7-9, 2022.
- [29]青木直史, “ディジタル・サウンド処理入門,” CQ 出版, 2006.
- [30]K. Yamamoto, A. Ogihara and H. Murata, “Direction estimation of sound source using neural network in consideration of the overtone structure of instrument sound,” The 36th International Technical Conference on Circuits/Systems, Computers and Communications, pp.259-262, 2021.

- [31] 荒木雅弘, “フリーソフトでつくる音声認識システム,” 森北出版, 2007.
- [32] 馬杉正男, 信号解析, 森北出版, 2013.
- [33] D.W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” J. of Society for Industrial and Applied Mathematics, vol.11, no.2, pp.431-441, 1963.
- [34] 藤原毅夫, “MATLAB クイックスタート,” 東京大学出版会, 2021.
- [35] Y.N. Dauphin, A. Fan, M. Auli and D. Grangier, “Language modeling with gated convolutional networks,” Proc. the 34th International Conference on Machine Learning, pp.933-941, 2017.
- [36] M. Kitahashi and H. Handa, “Estimating classroom situations by using CNN with environmental sound spectrograms,” J. of Advanced Computational Intelligence and Intelligent Informatics, vol.22, no.2, pp.242-248, 2018.
- [37] W. Zhu, M. Zhang, C. Wu and L. Zeng, “Broadband Direction of Arrival Estimation Based on Convolutional Neural Network,” IEICE Trans. Commun., vol.E103-B, no.3, pp.148-154, 2020.
- [38] N. Yalta, K. Nakadai and T. Ogata, “Sound Source Localization Using Deep Learning Models,” J. of Robotics and Mechatronics, vol.29, no.1, 2017.
- [39] K. Yamamoto, A. Ogihara and H. Murata, “Direction estimation of sound source by MUSIC method and CNN considering overtone structure,” Proc. 2022 International Technical Conference on Circuits/Systems, Computers and Communications, pp.671-674, 2022.
- [40] 宝珠山治, “音源分離,” 電子情報通信学会知識ベース「知識の森」, 2 群 (画像・音・言語) 6 編 (音響信号処理) 2 章, [https://www.ieice-hbkb.org/files/02/02gun\\_06hen\\_02.pdf](https://www.ieice-hbkb.org/files/02/02gun_06hen_02.pdf), 参照 Jul. 6, 2022.

## 謝辞

本研究の遂行ならびに本論文の執筆にあたり，研究環境の提供や研究の細部に至るまで，常日頃より丁寧で多大なる御指導，御鞭撻を賜りました，近畿大学工学部 情報学科 荻原 昭夫 教授に深謝の意を表します。

日頃より多くの御指導と御鞭撻を賜り，主査として本論文を査読していただきました，近畿大学工学部 情報学科 田中 一基 教授に深く感謝いたします。

本論文をまとめるにあたり，御教示ならびに御配慮を賜り，副主査として本論文を査読していただきました，近畿大学工学部 電子情報工学科 中島 弘之 教授に厚く御礼申し上げます。

本研究の遂行にあたり，様々な御指導，ご鞭撻を賜りました，中京大学工学部 情報工学科 村田 晴美 准教授に心より御礼申し上げます。

大学入学時から常に気にかけていただき，多数の御助言をいただきました，近畿大学工学部 情報学科 木村 有寿 准教授に厚く御礼申し上げます。

ティーチングアシスタントや日常生活を通して，多数の御指導を賜りました，近畿大学工学部 情報学科 谷崎 隆士 教授，徐 丙鉄 教授，片岡 隆之 教授，松岡 敬興 教授，古川 亮 教授，富永 徳幸 准教授，西尾 美由紀 准教授，阪口 龍彦 准教授，加島 智子 准教授，大谷 崇 講師，中村 一美 講師，近畿大学情報学部 情報学科 山元 翔 講師に厚く御礼申し上げます。

本研究の遂行に際して，日々の研究生生活を支えていただきました近畿大学大学院 システム工学研究科の先輩，後輩である安原 尚吾 氏，赤尾 大樹 氏，平林 慧太 氏，宮西 宏和 氏，井田 真人 氏，森 大星 氏，貞廣 諒 氏および近畿大学工学部情報学科 知覚情報システム研究室の学部生の皆様に御礼申し上げます。

最後に，学部時代から学生生活を支えていただき，多くの貴重な機会をいただきました，宇根 辰徳 氏，如木 徳和 氏，小笠原 洋介 氏，長尾 明来士 氏，長谷川 晴香 氏に厚く御礼申し上げます。

令和5年1月6日  
山本 華帆