

長岡技術科学大学大学院
工学研究科修士論文

題 目

主観評価を用いたH R T Fモデルの
個人適応化に関する研究

指導教員

准教授 杉田 泰則

著 者

電気電子情報工学専攻
信号処理応用研究室
17312987 楠林 透河

令和3年2月9日

ABSTRACT

A study on personal adaptation of HRTF model incorporating subjective evaluation

Author : KUSUBAYASHI Toga
Supervisor : SUGITA Yasunori

In recent years, with the development of three-dimensional video technology, sound image localization technology that gives a sense of direction to sound is expected. A method using a head-related transfer function (HRTF) has been proposed as a method for achieving high sound image localization in a sound source reproduction device such as headphones. HRTF is information about clues for humans to know the sense of direction, and is a transfer function until the sound source radiated from a certain position in free space reaches the eardrum. However, there is a problem that the acquisition of HRTF requires special equipment and devices called anechoic chambers. In addition, there are individual differences in HRTFs, and HRTFs other than the user himself have low localization of sound accuracy, making it difficult to put them into practical use. Various methods have been studied to solve these problems, but the localization accuracy of the HRTF of the person himself has not been obtained.

The purpose of this paper is to build an HRTF suitable for individual users who do not require special equipment or devices. Therefore, an HRTF model is created using a non-linear method and subject experiments are conducted. We will optimize the HRTF with a correction filter using the feedback from the subject experiment answers, and gradually aim to create an HRTF suitable for each user.

The first step is to create the HRTF model for optimization. The HRTF feature distribution is calculated using kernel principal component analysis from multiple HRTFs in one direction (not including the listener's own HRTF) existing in the database. After that, clustering into four groups with similar characteristics by the k-means method, and one HRTF model is created from each group. The HRTF to be used as a reference is the one that the subject evaluates as the best among the HRTF model.

The correction for personalizing the HRTF model is to change the notch position and peak amplitude of the HRTF. The part of the HRTF that gives people a sense of direction is known to be the peak around 4 kHz and the notch that exists between 4 kHz and 10 kHz of the HRTF, which has different amplitudes depending on the angle. An IIR notch filter and an IIR peak filter are used in combination as the correction filter to increase or decrease the

peak amplitude and adjust the notch position. The correction is made with reference to the results of the subject experiment, and the correction is made for the direction in which there was a localization error in the HRTF model in 12 directions at 30° intervals. We repeat the process until the subject perceives the corrected HRTF as approximately the correct direction. The evaluation method of the proposed method examines the effectiveness of the proposed method by comparing the sound image localization accuracy of the HRTF of the person himself and the corrected HRTF.

As a result, when compared with the case of using the HRTF of the person himself, the localization of sound accuracy was slightly inferior even if the correction was performed, but the error front and back was improved and the localization accuracy was improved as compared with the case of using the HRTF model.

目次

第 1 章	はじめに	1
1.1	研究背景	1
1.2	研究目的	2
1.3	論文構成	2
第 2 章	本研究の基礎となる知識・手法	3
2.1	音像定位技術	3
2.1.1	音像定位技術 [13]	3
2.1.2	座標系	3
2.2	頭部伝達関数	4
2.2.1	頭部伝達関数	4
2.2.2	水平方向における HRTF の特徴 [4]	5
2.2.3	HRTF 個人化の問題点	7
2.3	バイノーラル音源	8
2.4	カーネル主成分分析 [14] [15]	8
2.5	k-means 法 [16]	9
2.6	まとめ	10
第 3 章	提案手法	11
3.1	HRTF モデル作成	11
3.1.1	HRTF 特徴量算出	12
3.1.2	クラスタリング	12
3.1.3	HRTF モデルの作成手順	14
3.2	HRTF モデルの補正	15
3.2.1	補正箇所	15
3.2.2	補正フィルタ	16
3.2.3	HRTF モデルの補正手順	18

3.2.4	0°、180° の補正	19
3.2.5	各方向の補正	19
3.3	まとめ	20
第 4 章	提案手法の検証	21
4.1	本人 HRTF 測定条件	21
4.2	HRTF モデルの音像定位精度	22
4.3	補正処理後の音像定位評価	26
4.3.1	HRTF 補正結果 1	26
4.3.2	HRTF 補正結果 2	27
4.3.3	音像定位評価試験	29
4.4	まとめ	33
第 5 章	終わりに	35
謝辞	36
付録 A	付録	37
A.1	HRTF 補正結果	37
A.2	HRTF モデル被験者実験	39
A.3	補正処理後の被験者実験	41
参考文献	44

第 1 章

はじめに

本章では、本論文の研究背景および目的、本論文の構成について述べる。

1.1 研究背景

近年、バーチャルリアリティ (VR: virtual reality) に代表される 3 次元映像が一般的となっている。それに伴いより臨場感を演出するエンターテインメントが求められるようになった。臨場感を提案する一つとして、ヘッドホンやスピーカーを用いて上下左右あらゆる方向に音を定位させる音像定位技術がある。音像定位技術とは、ヘッドホンなどから再生される受聴音に対して方向感などの位置情報を付与する技術である。この技術は、映画やゲームなどのエンターテインメント分野だけでなく、視覚障害者向けの歩行訓練支援などの医療福祉分野などの様々な分野への応用が期待されている [1] [2]。

一般に、ヘッドホンで音像定位技術を再現させるには、頭部伝達関数 (HRTF: Head-Related Transfer Function) が必要となる。HRTF とは、ヒトが方向感覚を知る手がかりに関する情報で、自由空間のある位置から放射された音源が鼓膜に届くまでの伝達関数であり、頭や胴体、耳介などに寄与したものである。HRTF のインパルス応答と再生したい音源を畳み込むことで音像定位技術が実現できる。しかし、高い音像定位精度を実現するためには、受聴者本人の HRTF を使用する必要がある [3] [4]。

HRTF には複数の問題があり、音像定位技術の提供が難航している。主な要因として以下の二つのことが考えられる。一つ目は HRTF の取得である。HRTF は、受聴者に対する音の届き方から測定できる。測定には、無響室と呼ばれる特別な設備や装置が必要となり、気軽に取得することができない。また、あらゆる方向に対して音像定位を実現させるためには、必要とする全ての方向で HRTF を測定しなければならない問題点がある [4]。二つ目は、HRTF の個人性である。他人の HRTF を用いて音像定

位を行う場合、高い音像定位精度が得られることは稀であり、多くの人は方向の誤知覚を発生してしまふ。その要因は、HRTF が個人の身体情報に左右されるからである。一見似た体格の人の HRTF を用いた場合でも知覚方向の誤りを生じる問題がある [5]。この問題を解決するために、様々な HRTF の生成手法が検討されている [6]- [12]。誰もが気軽に音像定位を実現するには、上記の問題を解決する必要がある。

1.2 研究目的

特殊な設備や装置を必要としない個々のユーザに適した HRTF の構築を目的とし、受聴者本人の HRTF と同程度の音像定位技術を提案する。非線形手法を用いて HRTF モデルを作成し、HRTF モデルを用いて被験者実験を行う。実験回答からのフィードバックを利用することで HRTF の最適化を行い、次第に個々のユーザに適した HRTF を作成する。その後、HRTF との音像定位精度比較により提案法の有効性を検討する。

1.3 論文構成

本論文の構成を以下に示す。第2章では本論文で用いる基礎的知識について述べる。第3章では、被験者実験の基盤とする HRTF モデルと、提案手法である HRTF モデルの補正方法について述べる。第4章では HRTF モデルと補正した HRTF モデルそれぞれの定位精度実験を行い、HRTF との音像定位精度の比較を行う。第5章では本論文の結論を述べる。

第 2 章

本研究の基礎となる知識・手法

この章では、本論文で用いる基礎的な知識・手法について述べる。2.1 節では、音像定位技術について述べる。2.2 節では、頭部伝達関数について述べ、2.3 節ではバイノーラル音源の作成方法について述べる。また、2.4 節と 2.5 節では、本論文で用いるカーネル主成分分析と、k-means 法について述べる。

2.1 音像定位技術

2.1.1 音像定位技術 [13]

音源から発せられた音波が鼓膜に到達した際に、人が知覚した総体を音像という。また、音像の空間的位置 (音像の方向と距離) を知覚することを音像定位という。さらに、ヘッドホンなどから受聴する音に方向感や距離感などの位置情報を与え、受聴者に音像を知覚させる技術を音像定位技術という。ヘッドホンで高精度の音像定位技術を実現するためには、頭部伝達関数 (Head-Related Transfer Function : HRTF) を用いる手法が一般的である。以下の 2.2 節において頭部伝達関数について述べる。

2.1.2 座標系

次に、本論文において音源の方向を表すための座標系について述べる。本論文では図 2.1 に示す、頭部を中心とした球座標系を用いる。この座標系は次のように定義される。原点は左右の外耳道を結ぶ線分の中点である。水平面は眼窩点 (眼球の収まる頭蓋骨のくぼみ) と耳珠 (外耳道入り口のでっぱり) を含む平面であり、横断面は左右のが外耳道入り口を通り水平面に直交する面である。また、正中面は水平面と横断面の両方に直交する面である。角度は、方位角 θ° と仰角 φ° を用いる。

本論文では、音源が水平面上に存在する場合に対しての検証を行っているため、仰

角 $\varphi = 0^\circ$ であり、これ以降 φ は省略する。また、図 2.1 のように正面を 0° 方向とし、正面から時計回りに $0^\circ \sim 180^\circ$ と増えていくものとする。

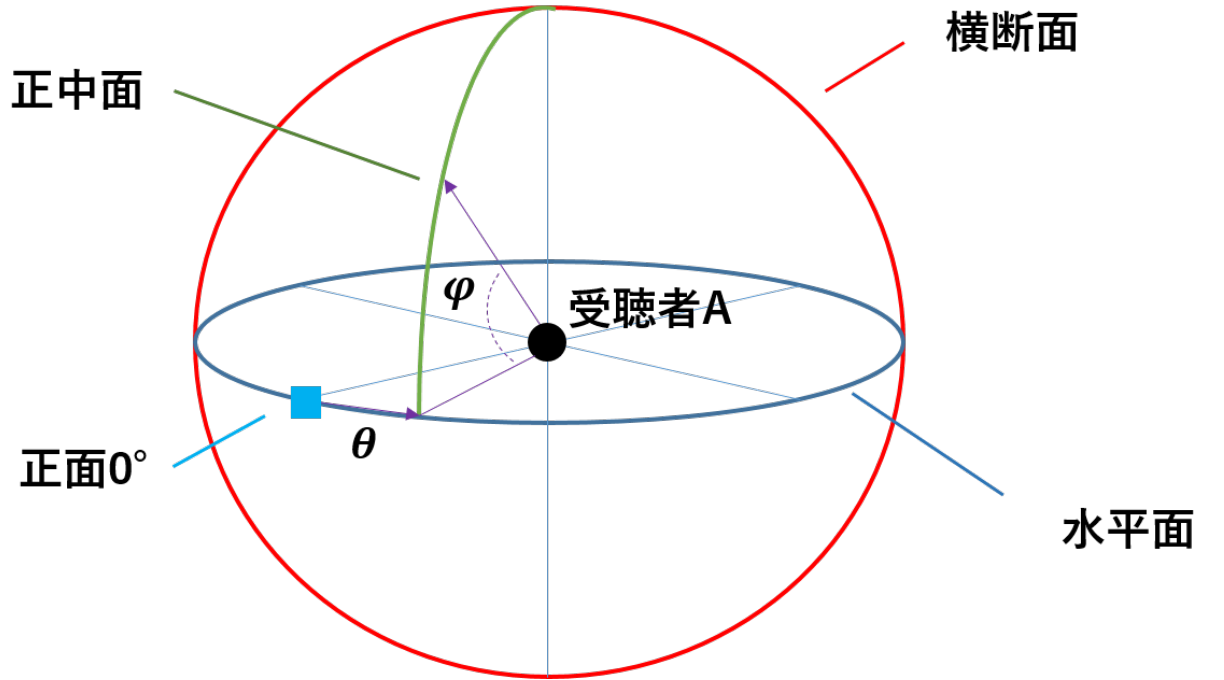


図 2.1 球座標系

2.2 頭部伝達関数

2.2.1 頭部伝達関数

頭部伝達関数の定義を説明する。ある音波が耳に到達するときに、受聴者の頭や耳介の形状、あるいは胴体等の体形状の影響を受ける。頭部伝達関数 (Head-Related Transfer Function : HRTF) とは、このような頭部周辺の影響によって生じる入射音の音圧変化を周波数の関数で表したものである。頭部伝達関数は次式で表される。

$$H_{L|R}(\theta, \omega) = \frac{G_{L|R}(\theta, \omega)}{F(\theta, \omega)} \quad (2.1)$$

ここで、 $H_{L|R}(\theta, \omega)$ は左右の HRTF であり、 θ は音源の方位角度を表し、 ω は周波数を表す。 $G_{L|R}(\theta, \omega)$ は、反射や回折のない自由音場における音源から受聴者の左右の外耳道入り口までの伝達関数である。 $F(\theta, \omega)$ は、自由音場における受聴者がいない状態での音源から受聴者の頭部中心に相当する位置までの伝達関数である。

2.2.2 水平方向における HRTF の特徴 [4]

まず、水平方向における HRTF の特徴について述べる。水平方向の音像定位の実現に対して、左右方向の知覚には HRTF の時間差、音圧のレベル差で判定しているとされている。

音像が右側にある場合、左耳を反受聴側、右耳を受聴側とする。図 2.2 と図 2.3 に同一被験者に対して、正面（方位角 $\theta = 0^\circ$ ）から右方向（方位角 $\theta = 90^\circ$ ）まで 30° 間隔で測定した HRTF の周波数振幅特性を示す。図 2.2 は左耳の反受聴側 HRTF を、図 2.3 には右耳の受聴側 HRTF を示す。図 2.2 の P1 は、HRTF の振幅が最も大きくなる点である。また、図 2.3 の N1 は、4kHz 以降に発生する複数のノッチの中で最初に存在するノッチとする。また、表 2.1 に P1 ピークの周波数と括弧内に振幅レベル (dB) を、表 2.2 に N1 ノッチの周波数を示す。図 2.2 と図 2.3 より、HRTF は受聴者に対する音源の角度によって特性が異なることがわかる。また、その特性の違いから人は音源の方向を知覚している。そのため必要な方向数の HRTF を全て取得しなければ、多方向の音像定位は実現しない。

図 2.2 より反受聴側 HRTF は、角度ごとに P1 ピークとそれ以前の振幅が上下に変化する。反受聴側 HRTF の特徴として、音像の角度が顔正面 (0°) から背面 (90°) に移動するにつれ P1 の振幅が増加する。図 2.3 より受聴側 HRTF は、角度が顔正面 (0°) から顔横 (90°) になるにつれ N1 ノッチが高周波に移動している。この時、受聴側 HRTF の角度が顔横 (90°) から背面に (180°) になる場合も、N1 ノッチは低周波に移動する。

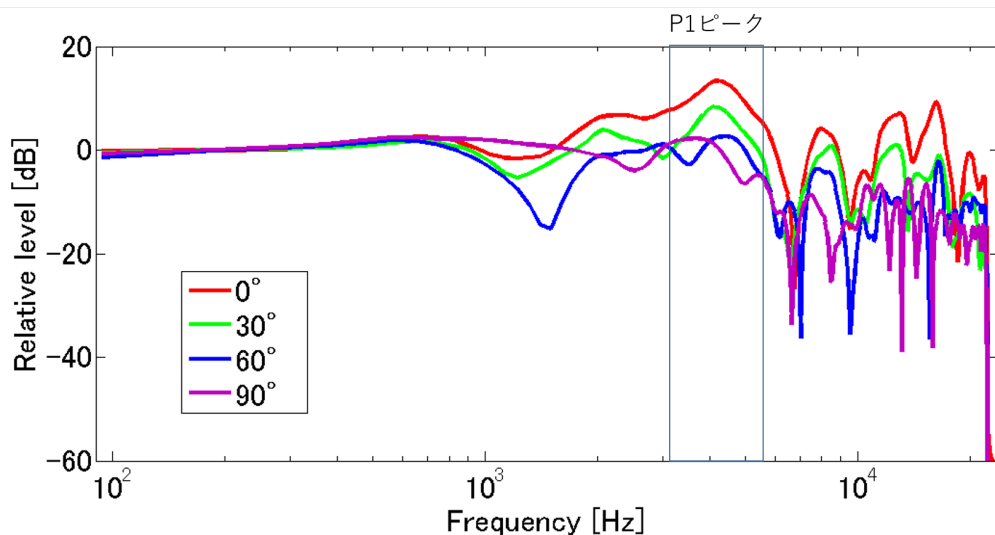


図 2.2 反受聴側 HRTF

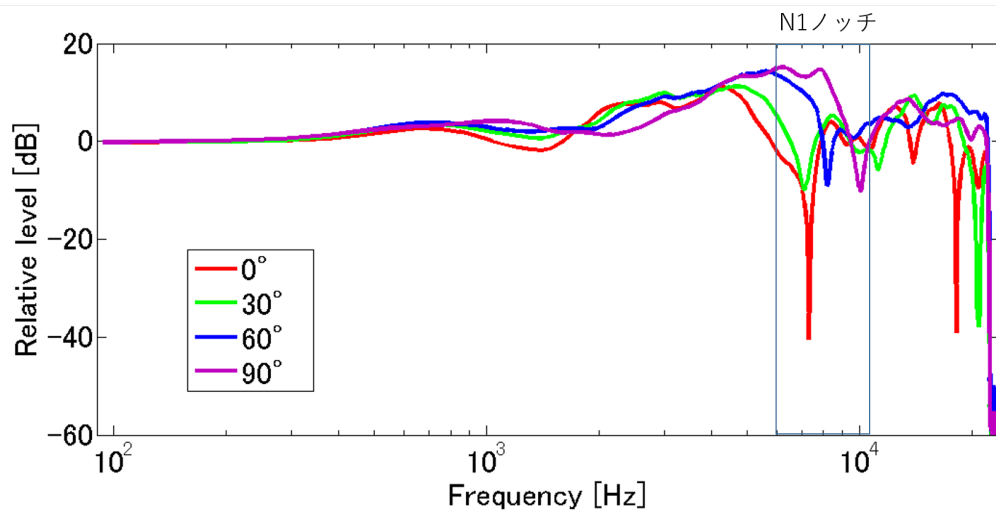


図 2.3 受聴側 HRTF

表 2.1 各方向の振幅位置例

方向 [°]	反受聴側 [Hz]	受聴側 [Hz]
0	4219(11.06)	4313(10.98)
30	4125(6.316)	4875(11.25)
60	4405(2.53)	5075(14.31)
90	3824(2.12)	7878(15.01)
120	4125(3.21)	8053(14.48)
150	4031(7.15)	4125(10.04)
180	4719(8.81)	4598(9.58)

表 2.2 各方向のノッチ位置例

方向 [°]	反受聴側 [Hz]	受聴側 [Hz]
0	6750	7313
30	6553	7125
60	6566	8156
90	7031	10032
120	7959	10974
150	8344	9938
180	9459	9788

2.2.3 HRTF 個人化の問題点

本節では、HRTF の個人差による問題を説明する。図 2.4 は、同一方向正面 (方位角 $\theta = 0^\circ$) に対する、4 名の被験者の HRTF 振幅特性である。図 2.4 からは HRTF は音源の角度が同じ場合でも 4 kHz～10 kHz 付近の N1 ノッチ位置やノッチ前の P1 振幅に個人差が生じていることがわかる。この個人差が HRTF の大きな課題点である。この個人差は、両耳間時間差やレベル差などに影響している。他者の HRTF を用いて音像定位を行う場合、定位角度のずれや前後方向について誤判定が頻繁に発生するなど、音像定位精度が低下してしまう。そのため、本人に合う音像定位を得るためには、受聴者本人の HRTF を測定する必要がある。しかし、一章で述べたように HRTF の取得には、無響室と呼ばれる特別な設備や多方向から音を再生する装置を用いるため普及が難航している。

この問題を解決するために、以下のような HRTF の個人化に関する研究が行われており、表 2.1 に設備面や定位精度の比較を示す。

- ・ 試聴により HRTF を選択する手法 [6] [7]
- ・ 体の形状が似ている人の HRTF を選択する手法 [8] [9]
- ・ 体の形状より HRTF を推定する手法 [10]
- ・ HRTF を補間する手法 (未所持角度の推定) [11] [12]

表 2.3 各手法の比較

項目	定位精度	設備・装置面	学習の有無
本人 HRTF を使用する手法	◎	×	無
試聴により HRTF を選択する手法	△	◎	無
体の形状が似ている人の HRTF を選択する手法	○	△	有
体の形状より HRTF を推定する手法	○	△	有
HRTF を補間する手法 (未所持角度の推定)	○	×	有

表 2.1 において◎は最良を×は最悪を表している。本人の HRTF を用いる手法より良いとされるのは、受聴者本人の情報を用いて学習より算出される HRTF を用いる手法であるしかし、体の形状を利用する手法や HRTF を補間する手法は、本人の体情報や HRTF の測定が必要といったデメリットなどが存在する。また、本人の HRTF を使用した場合より、高い音像定位精度を出す手法は提案されていないため問題は難航し

ている。

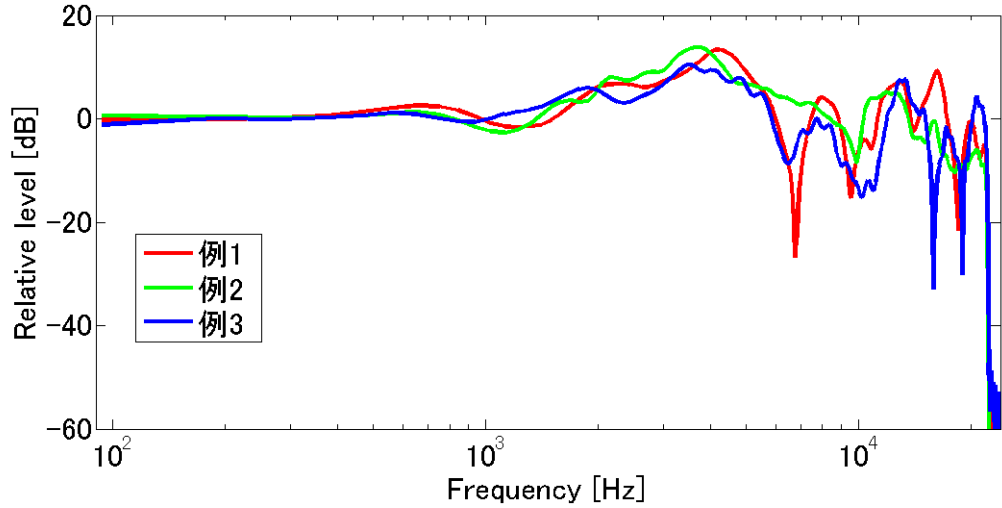


図 2.4 同角度における被験者 3 人の右耳の HRTF

2.3 バイノーラル音源

2.2.1 節で述べたように、音が人の耳に到達するとき、HRTF の影響を受ける。この HRTF の影響を受けた、信号をバイノーラル音源という。

$$X_{L|R}(\theta, \omega) = H_{L|R}(\theta, \omega)S(\omega) \quad (2.2)$$

ここで、 $X_{L|R}(\theta, \omega)$ は左右の耳におけるバイノーラル音源であり、 $S(\omega)$ は、音源である。人はこの音源を手掛かりに音源の位置情報を知覚している。すなわち、このバイノーラル音源をヘッドホンなどの音源再生機器から再生することで、実際に音源から放射された受聴者を再現することが可能であり、方向性を持った音像定位を実現することができる。

2.4 カーネル主成分分析 [14] [15]

主成分分析は、特徴量の相関関係を利用することによって次元圧縮された部分空間への射影を求める手法である。しかし、主成分分析は線形手法であり、特徴量分布が線形性を持つデータに対して有用な手法であり、特徴量分布が非線形性が強いデータでは特徴量を抽出することができず情報を損失してしまう可能性がある。HRTF の個人差は、非線形性が強いと考え本論文では、非線形手法のカーネル主成分分析を使用する。

主成分分析において、ある被験者の HRTF は、主成分と重みの積和で合成することができる。重みは HRTF の個人性を示すものである。本論文では、主成分分析を行う際に算出される固有ベクトル \mathbf{v} と固有値 λ を用いて、主成分得点算出しを HRTF 特徴量として使用する。以下にカーネル主成分分析を用いた特徴量の算出方法を示す。 i 番目の被験者の一方向の HRTF 振幅応答を要素とするベクトルを \mathbf{h}_i とし、平均 HRTF 振幅応答ベクトル $\bar{\mathbf{h}}$ を

$$\bar{\mathbf{h}} = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{h}_i \quad (2.3)$$

とすると、変量ベクトル $\hat{\mathbf{h}}_i$ は次式で表せる。

$$\hat{\mathbf{h}}_i = \mathbf{h}_i - \bar{\mathbf{h}} \quad (2.4)$$

N_S は分析に用いた変量ベクトル \mathbf{h}_i の数であり、主成分分析に用いた被験者数である。算出した変量ベクトル $\hat{\mathbf{h}}_i$ よりカーネル関数 \mathbf{K} を算出し、カーネル関数 \mathbf{K} に対して主成分分析を行う。カーネル関数を算出するために、カーネルを選択する必要がある。本論文では、最も頻繁に使用される次式のガウシアンカーネル \mathbf{K}_{ij} を使用する。

$$\mathbf{K}_{ij} = e^{-2p(\hat{\mathbf{h}}_i - \hat{\mathbf{h}}_j)} \quad (2.5)$$

$\hat{\mathbf{h}}_j$ は、 $\hat{\mathbf{h}}_i$ の転置行列であり、 p はカーネルパラメータである。本論文では、 $p = 0.25$ を使用する。ガウシアンカーネル \mathbf{K}_{ij} より、カーネル関数 \mathbf{K} は次式で表せる。

$$\mathbf{K} = \mathbf{K}_{ij} - \frac{1}{N_h} \mathbf{K}_{ij} + \frac{1}{N_h^2} \mathbf{K}_{ij} \quad (2.6)$$

N_h は、HRTF の点数でありカーネル主成分分析における次元数である。必要な方向分のカーネル関数 \mathbf{K} に対して、固有値問題を解くことで固有ベクトル \mathbf{v} と固有値 λ を算出すると、 i 番目の被験者のある方向の主成分得点 Y_i は次式で表せる。

$$Y_i = \frac{\mathbf{v}_i}{\sqrt{\lambda_i}} \mathbf{K}_{ij} \quad (2.7)$$

これを必要な角度数に対して行うことで、全被験者の角度ごとの主成分得点を算出し、HRTF 特徴量を得る。

2.5 k-means 法 [16]

本節では、クラスタリング手法の k-means 法について説明する。クラスタリングとは、データ間の類似度にもとづいて、データをグループ分けする手法である。クラスタリングによってできた、類似したデータが集まったグループのことをクラスタと呼

ぶ。k-means 法とは、クラスタリング手法の一つで、まずデータを適当なクラスタに分けた後、クラスタの平均を用いてデータがわかれるように調整させていくアルゴリズムである。任意の指定のクラスタを作成可能な特徴から本論文では、k-means 法を用いる。k-means 法のアルゴリズムを以下に示す。

- 分類する個数（クラスタ数 k ）を決定
- 各クラスタの中心点 V_k を決定（最初の一つは無作為に設置し、残りの点の一つ目の点から遠い位置に算出）
- 各データ X_i と V_k との距離を求め、 X_i を最も近い中心点 V_k のクラスタに割り当てる
- 各クラスタで平均値をとり、平均値点に中心点 V_k を移動させる
- 上記を繰り返し中心点 V_k が移動しなくなるまで繰り返す

2.6 まとめ

本章では、本研究の基礎となる知識・手法について述べた。人が音を受聴する際に、音源が外耳道入り口に到達するまでの伝達関数が HRTF である。ヘッドホンなどの外耳入り口で音源を再生する機器と HRTF 等を用いて作成されるバイノーラル音源を組み合わせることで、受聴音に位置情報を付与した音源を届けることが可能である。しかし、HRTF は個人性に大きく依存し、受聴者本人以外の HRTF を用いてバイノーラル音源を作成し音像定位を行う場合、方向のばらつきや前後方向について誤判定が頻繁に発生するなど、音像定位精度が低下してしまう問題点が発生してしまう。3 章では、受聴者本人の HRTF を用いないで音像定位を実現する提案手法として、カーネル主成分分析を用いたモデル HRTF 作成と、HRTF モデルを個人に最適化するための補正フィルタについて述べる。

第 3 章

提案手法

提案手法について説明する。まず HRTF より作成されたバイノーラル音源を用いて被験者実験を行う。被験者実験では音像定位実験を行い、被験者が知覚した音像位置の回答を得る。その後、回答をフィードバックし補正フィルタを適用することで HRTF の最適化を行う。被験者実験と最適化を繰り返すことで音像定位の精度向上を目指し次第に個々の受聴者に適した HRTF を作成する。被験者実験の際、使用する HRTF は試聴から被験者に合うものを選択するのが最も良いと考えられる。しかし、データベース等には多くの HRTF が存在し、求める方向全てを試聴することは困難である。そのため、本論文では被験者実験で基準となる HRTF（以降：HRTF モデル）を作成し使用する。

本章の 3.1 節ではまず、被験者実験のモデルとなる HRTF の作成手順を示す。次に、3.2 節では被験者実験の主観回答をフィードバックし HRTF モデルを補正する補正フィルタを示す。

3.1 HRTF モデル作成

この節では、HRTF モデル作成手順と作成結果を示す。HRTF モデルの作成方法について説明する。まず、データベースに存在する一方向の複数ある HRTF からカーネル主成分分析を用いて HRTF 特徴量分布を算出する。その後、k-means 法より特徴が類似する 4 つのグループにクラスタリングし、各グループから 1 つの HRTF モデルを作成する。HRTF モデルは片耳ずつ作成し、1 方向あたり右耳 4 通り×左耳 4 通りの計 16 通りの HRTF を作成する。ここで、本研究では受聴者本人の体情報や HRTF を用いない手法を検討するため、HRTF モデルを作成する際には本人の HRTF を含まないデータセットを利用する。

3.1.1 HRTF 特徴量算出

カーネル主成分分析に使用する HRTF は、東北大学電気通信研究所のデータベース HRTF を用いた [17]。水平面内の $0^\circ \sim 330^\circ$ まで 30° 刻みで 12 方向、1 方向につき 100 名分の HRTF をカーネル主成分分析の対象とした。 i 番目の被験者の特徴量は、カーネル関数によって拡張され、512 次元のベクトルとして、主成分得点より算出される。算出された 0° 方向の特徴量分布を図 3.1 と図 3.2 に示す。図 3.1 と図 3.2 は 100 名分の HRTF 特徴量分布を示し、図 3.1 は二次元、図 3.2 は三次元で表現している。この特徴量分布に対してクラスタリングを行う。

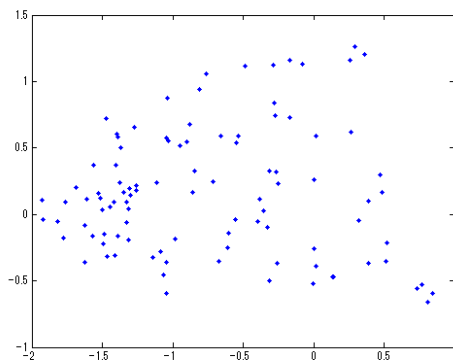


図 3.1 特徴量分布 (2 次元)

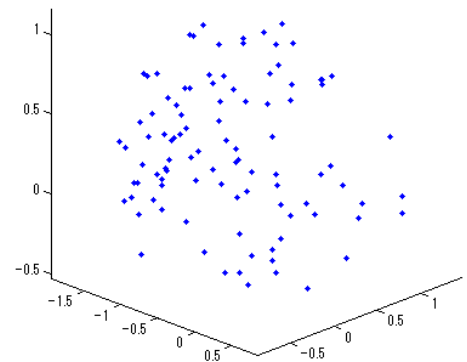


図 3.2 特徴量分布 (3 次元)

3.1.2 クラスタリング

本論文では HRTF 特徴量のクラスタリングに、タンデムクラスタリングを用いる。タンデムクラスタリングは主成分分析により少ない次元に縮約し、クラスタリングを行う 2 段階分析である。1 段階目の次元縮約にはカーネル主成分分析の結果を利用し、クラスタリングには k-means 法を利用する。

3.1.1 節で算出された 512 次元の特徴量は、HRTF の個人性、角度に寄与する共通部分、測定時のノイズ成分を含んでいる。寄与率の少ない高次元の特徴量はノイズ成分と判定し、クラスタリングに用いない。また、寄与率が大きい低次元の特徴量はより多くの被験者の共通の特徴を含んでおり、個人差が表れにくいといえる。HRTF の個人の特徴量を含む特徴量をクラスタリングするため、本論文ではカーネル主成分分

析した HRTF を復元したとき、およその外形が保たれる次元数を用いる。本論文では、全方向で 40 次元を用いた。本論文で用いるクラスタリング分析の条件を表 3.1 に示す。

表 3.1 クラスタリング条件

項目	内容
使用次元数	40 次元
クラスタ数	4

表 3.1 のクラスタリング条件を使用して、3.1.1 節で算出した HRTF 特徴量に k-means 法を適用した結果を図 3.3 と図 3.4 に示す。また、クラスタ内の HRTF 数を表 3.2 に示す。図 3.3 の + はクラスタの中心を表している。図 3.3 と図 3.4、表 3.2 より、HRTF 特徴量が 4 つのクラスタに分類されていることがわかる。

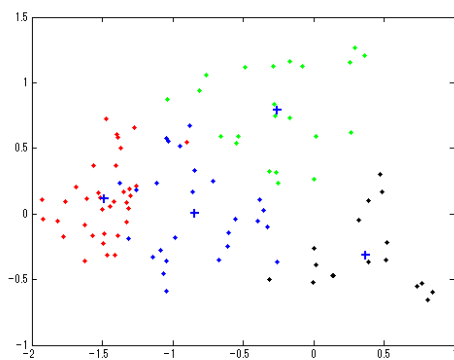


図 3.3 特徴量分布 (2 次元)

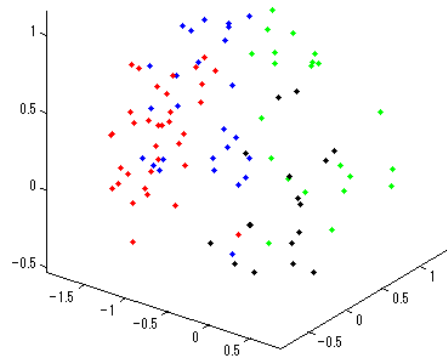


図 3.4 特徴量分布 (3 次元)

表 3.2 クラスタリング結果

グループ k	HRTF 数
グループ 1	32
グループ 2	22
グループ 3	26
グループ 4	20

3.1.3 HRTF モデルの作成手順

3.1.2 節のクラスタリング結果を用いて HRTF モデルを作成する方法について述べる。図 3.3 からわかるように kmeans 法ではクラスタリングの際、グループの中心点が算出される。この中心点は各クラスタ内での平均値を示し、グループ内での平均特徴量となる。この特徴量を利用し各グループから 1 つずつ HRTF モデルを作成する。本論文では、HRTF モデル作成に主成分分析の重みによる復元方法を用いた [10]。クラスタリング後の k グループの i 番の HRTF を一方向の HRTF 振幅応答を要素とするベクトルを \mathbf{h}_{ki} とすると、グループ内全員の HRTF \mathbf{h}_k は、

$$\mathbf{h}_k = [\mathbf{h}_{k1}, \mathbf{h}_{k2}, \mathbf{h}_{ki} \cdots \mathbf{h}_{kN_k}] \quad (3.1)$$

で表される。ここで、 N_k はグループ内の人数を表す。グループ k の平均 HRTF 振幅応答ベクトルを $\bar{\mathbf{h}}_k$ とすると変量ベクトル $\hat{\mathbf{h}}_k$ は、

$$\hat{\mathbf{h}}_k = \mathbf{h}_{ki} - \bar{\mathbf{h}}_k \quad (3.2)$$

で表される。変量ベクトル $\hat{\mathbf{h}}_k$ のすべての組み合わせに対して、共分散を計算し、分散・共分散行列を算出する。分散・共分散行列の固有ベクトルが主成分行列 \mathbf{C} であり、固有値が寄与率となる。各主成分に対応する重み係数ベクトル ω_{ki} よりクラスタリングより得られた中心点の重み係数ベクトル ω_k は、

$$\omega_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \omega_{ki} \quad (3.3)$$

となり、主成分行列 \mathbf{C} と、中心点の重み係数ベクトル ω_k を用いて HRTF モデル \mathbf{g}_k は、次式で合成できる。

$$\mathbf{g}_k = \mathbf{C}\omega_k + \bar{\mathbf{h}}_k \quad (3.4)$$

HRTF は、実数成分と虚数成分を持つ伝達関数であるが主成分分析から算出される HRTF モデルは虚数成分を失う。図 3.5 に 30° 方向の受聴側 (右耳) の HRTF モデルの作成結果を示す。図 3.5 より、4 つの HRTF は、4 kHz～10 kHz 付近のノッチ位置が異なっており、それぞれが異なる特徴を持った HRTF であることがわかる。

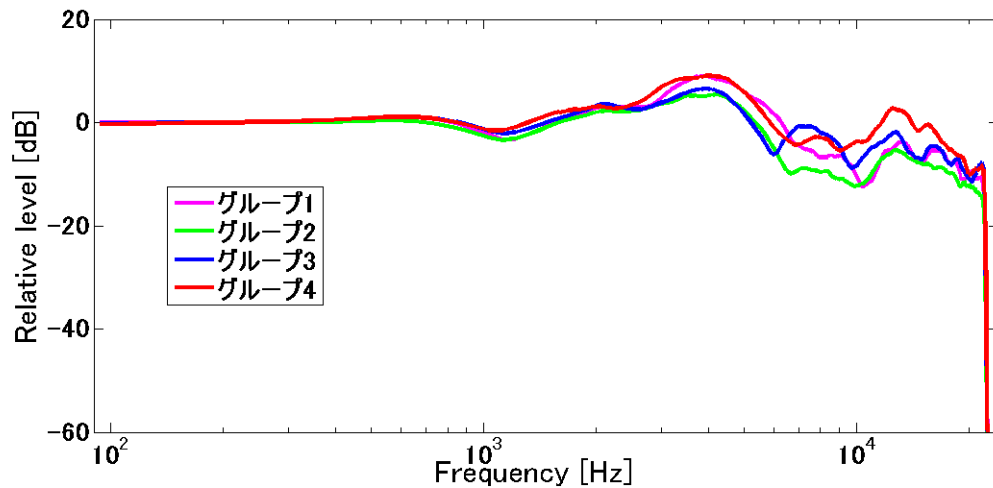


図 3.5 HRTF モデル

3.2 HRTF モデルの補正

前章の 2.3 節の HRTF の個人性に関する問題点では、4 kHz～10 kHz 付近のノッチ位置やノッチ手前の振幅に個人差があることを述べた。また、水平面の音像定位には、両耳間時間差や両耳間の音量レベル差が主に影響されることを述べた。そのため、被験者本人が最も良いと感じる HRTF モデルを選択しても定位角度のずれや前後方向について誤判定が発生すると考えられる。また、著者は両耳間の音量レベル差の発生や、個人差に HRTF の P1 ピークと N1 ノッチが起因していると考え。そのため、本論文では P1 の振幅と N1 位置を被験者実験の主観回答をフィードバックし変更することで個人に合う HRTF の作成を行う。次節ではそれぞれの変更手順について述べる。

3.2.1 補正箇所

本節では補正する箇所について述べる。2.2.3 節の個人性の問題点では、他人の HRTF を用いて音像定位を行った場合前後誤りや定位角度のずれが発生することを示した。ヘッドホン内で前後の誤りを解決するためには、音の大きさが重要と考える。図 3.6 と図 3.7 に 30° と 150°（前後）の HRTF モデルを示す。図 3.6 と図 3.7 より音の大きさに起因するのは、HRTF の P1 ピークとそれ以前の帯域であると考え、前後の誤りを補正する際は、P1 ピークとそれ以前の帯域を補正する。次に角度の誤りについて述べる。前章の図 2.2 より反受聴側 HRTF に形成される N1 ノッチの周波数位置はほぼ同じである。しかし、図 2.3 の受聴側 HRTF に形成される N1 ノッチは角度ごとに違いがある。この特徴を利用し N1 ノッチを定位させたい方向に変化させ補正する。

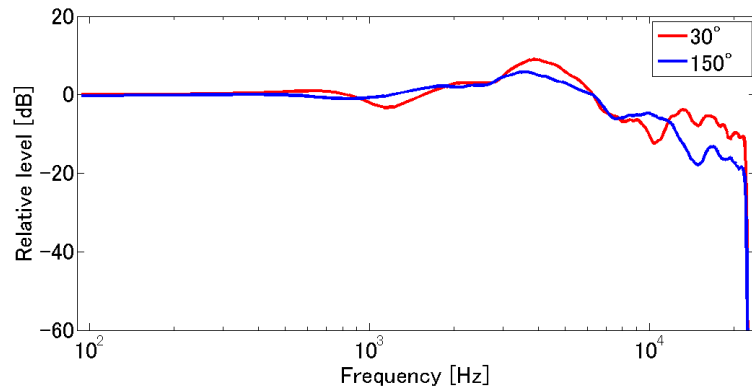


図 3.6 反受聴側（左耳）

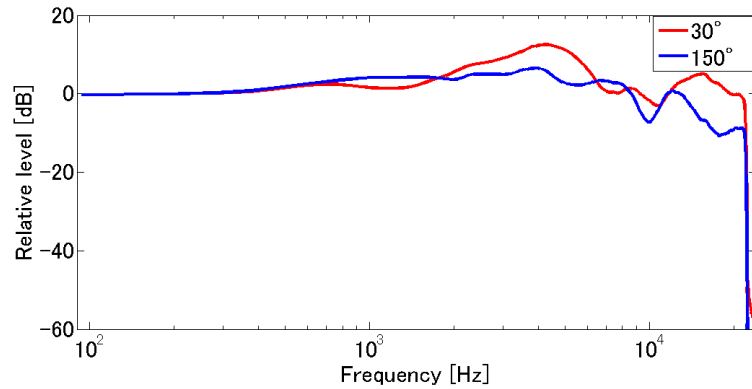


図 3.7 反受聴側（右耳）

3.2.2 補正フィルタ

本論文では、P1 ピーク振幅と N1 ノッチ位置の変更に IIR ノッチフィルタと IIR ピークフィルタを用い、これらを補正フィルタとする。ノッチフィルタとピークフィルタについて説明する。ノッチフィルタは、帯域減衰フィルタの一つで、ある範囲の周波数の信号のみ減衰させ、それ以外の周波数の信号を通過させるフィルタである。また、ピークフィルタは、ノッチフィルタと対の特性を持ち、ある範囲の周波数の信号のみ増幅させる。ノッチ周波数を f_n をノッチ幅を ζ 、ノッチの深さを d とするとノッチフィルタ $G(j\omega)$ は次式で設計できる。

$$\omega_n = 2\pi f_n \quad (3.5)$$

$$G(j\omega) = \frac{-\omega^2 + 2d\zeta\omega_n j\omega + \omega^2}{-\omega^2 + 2\zeta\omega_n j\omega + \omega^2} \quad (3.6)$$

ここで、 $d < 1.0$ に設定すると、ピークフィルタが設計できる。本論文で、N1 ノッチを変更する際は、各フィルタを組み合わせ用いる。次に N1 ノッチの変更手順を示す。

- N1 ノッチをピークフィルタで除去
- 新規のノッチをノッチフィルタで作成

また、P1 ピークとそれ以前の帯域の振幅変更する際は、ピークフィルタのみを用いる。図 3.8 にノッチフィルタフィルタの例を、図 3.9 にピークフィルタの例を示す。それぞれ任意の周波数を設定している。 ζ によってノッチ幅が変更でき、 d によって振幅が調整できていることがわかる。このフィルタを用いて、P1 ピークと N1 ノッチを調節する。図 3.10 にノッチを消した HRTF と新規にノッチを追加した HRTF について示す。ピークフィルタによって、ノッチが削除され、新規のノッチが作成できていることがわかる。

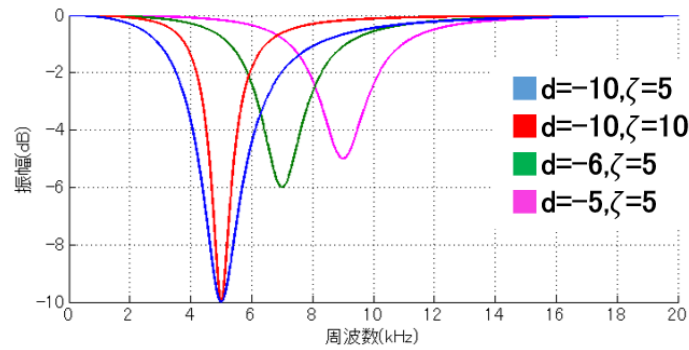


図 3.8 ノッチフィルタ

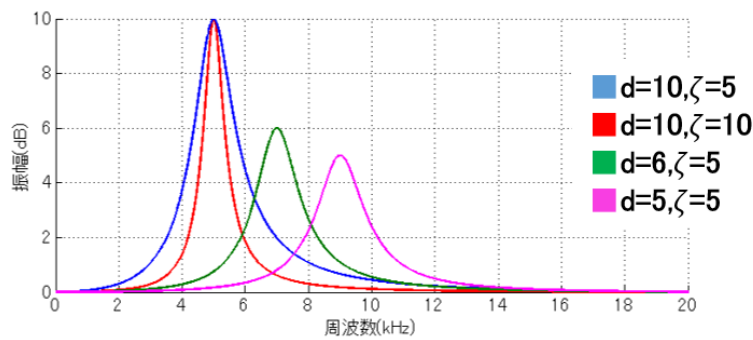


図 3.9 ピークフィルタ

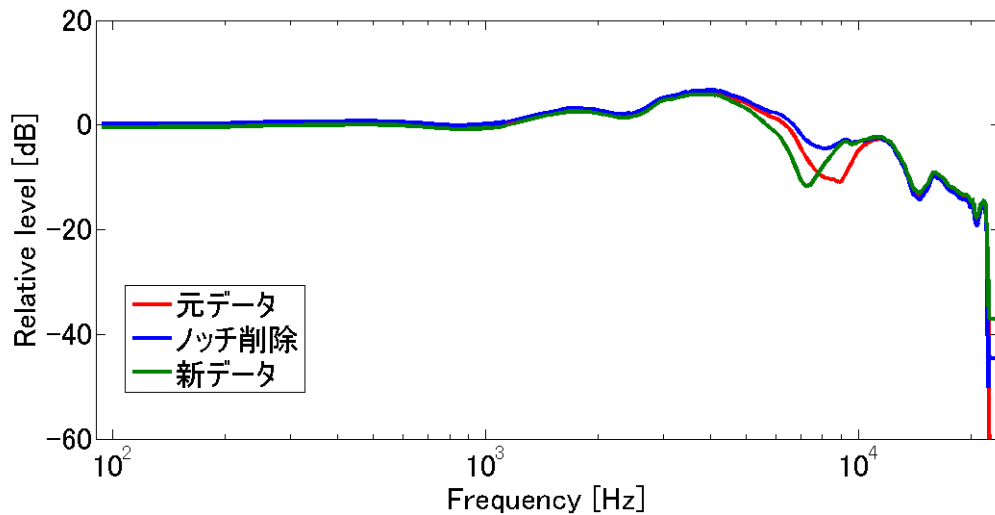


図 3.10 変更例

3.2.3 HRTF モデルの補正手順

補正は被験者実験の音像定位結果を参照することで行う。補正を行う方向は体の右側である右面 (0° 、 30° 、 60° ... 180°)、体の左側である左面 (0° 、 330° 、 300° ... 180°) のどちらかを行う。その後、補正した HRTF を左右で逆転させ、逆半面の微調整を行う。補正は主に 2 つの観点に着目する。一つは、前方を後方に知覚する (例えば、 30° を 150° 方向として知覚する)。あるいは、後方を前方に知覚する (例えば、 180° を 0° 方向として知覚する) といった前後誤りについてである。もう 1 つは、角度を誤知覚する (例えば、 30° を 60° として知覚する) 方向の誤りである。

補正を行う順番は図 3.11 に示す順番で補正する。基準角度が存在しない場合はまず 0° と 90° 、 180° の補正を行う。 0° と 180° の補正は次節に示す。 90° の補正は、 90° のノッチ位置を 200Hz ずつ移動させ定位が変化するかを検討する。 90° と知覚した場合は、基準とする角度として用いる。

方向の誤りを補正するにあたり基準とする角度 (以降：基準角度) を設けた。誤知覚が少ないとされる角度 (0° 、 90° 、 180°) と、被験者実験の音像定位結果で誤りが出なかった角度である。まず、これらの基準角度を用いて方向の誤りを補正する。その後、前後の誤りを補正する。補正箇所は、前章の 3.2.1 で示した通りである。

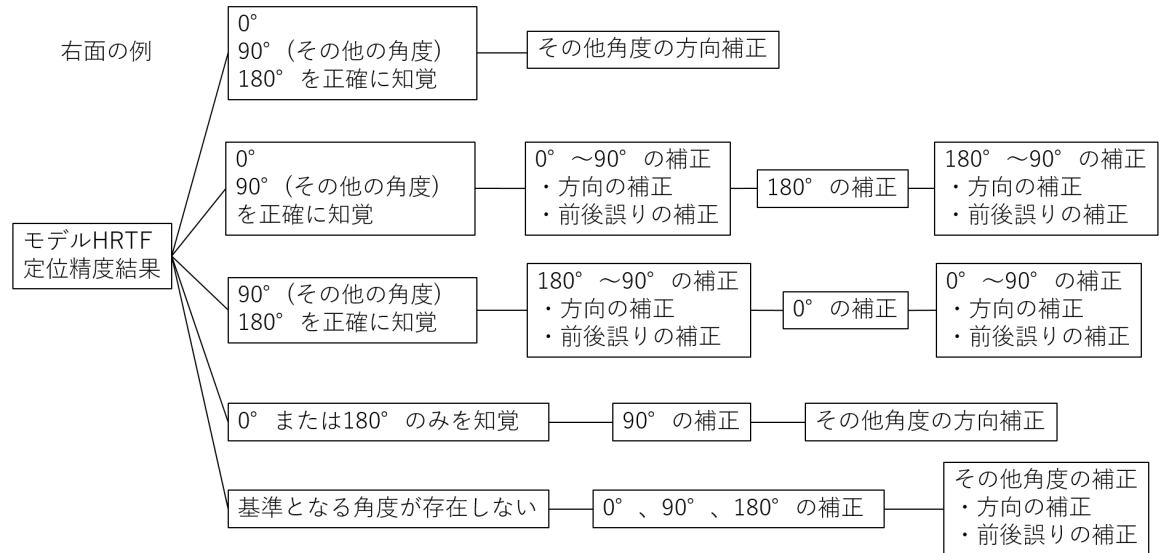


図 3.11 補正順番

3.2.4 0°、180° の補正

本節では、0° と 180° について述べる。主に 0° と 180° における誤りは、角度の誤知覚と前後の誤りであると考えられる。また、0° を 180°、180° を 0° と知覚する誤りは、水平面方向ではなく、正中面の誤りである。正中面の誤りは、と N1 ノッチ位置が影響し、音源が水平面に近づくほどノッチは低周波に移動するため N1 ノッチを変更する。変更には、基準がなく本論文では 200Hz ずつ移動させ、被験者が良いと主観するまで変更する。また、N1 ノッチ位置が P1 ピークの位置まで移動しても、定位に変化がない場合は、変更を中止した。角度を誤知覚している場合は、両耳間の振幅を補正する。0° と 180° は両耳にはいる音圧が同等であると考えられる。そのため、両耳の P1 ピーク振幅値をピークフィルタによって同等に変更する。

3.2.5 各方向の補正

本節では、0° と 180° 以外の補正について述べる。補正したい方向を θ_a° とするときの補正手順とを示す。0° と 180°、90° 270° 以外における誤りは、方向の誤りと前後の誤りであると考えられる。本節では、基準角度以外の補正方法について述べる。まず、角度の誤知覚である方向の誤り補正について述べる。方向の誤りは、受聴側の N1 ノッチを変更することで補正する。補正したい音像提示方向を θ_a° とする。 θ_a° の N1 ノッチ位置の変更決定には、提示した角度に最も近い基準角度のノッチ位置を参照し、補正したい角度と基準角度の間にノッチ位置を移動させ補正する。基準

角度を θ_b° とすると N1 ノッチ周波数 ω_{θ_a} と θ_b の N1 ノッチ周波数 ω_{θ_b} を用いて、変更するノッチ周波数位置 ω'_{θ_a} は、

$$\omega'_{\theta_a} = \frac{\omega_{\theta_a} + \omega_{\theta_b}}{2} \quad (3.7)$$

で算出される。その後、被験者に補正した HRTF を含めての音像定位試験を行い、 θ_a° の提示に対しての回答を得る。被験者の回答が θ_a° に対して、 θ_b° と逆方向へ知覚の誤りが再び生じた場合は上記式を用いて

$$\omega''_{\theta_a} = \frac{\omega'_{\theta_a} + \omega_{\theta_b}}{2} \quad (3.8)$$

ノッチ周波数を決定する。また、 θ_a° に対して、 θ_b° 方向に知覚の誤りが生じた場合は、 θ_b° のノッチ周波数位置に移動しすぎていると考えられる。そのため、 ω'_{θ_a} と ω_{θ_a} の中間点にノッチ周波数を決定する。

$$\omega''_{\theta_a} = \frac{\omega_{\theta_a} + \omega'_{\theta_a}}{2} \quad (3.9)$$

これを被験者が良いと思うまで繰り返す。次に、前後の誤りを起こしているならば、前後誤りの補正を行う。補正には、反受聴側の HRTF に存在する P1 ピークと P1 ピーク以下の周波数帯域の振幅を増加させ、前方向と知覚するまで補正を繰り返す。

3.3 まとめ

本章では、提案手法について述べた。3.1 節では、被験者実験の基準となる HRTF モデルの作成手順を示した。非線形手法のカーネル主成分分析と k-means 法を用い、結果 1 方向 16 組の HRTF モデルを作成した。また、3.2 節では、主観回答のフィードバック実験における HRTF の補正について示した。HRTF の補正には、フィルタを用いて任意の周波数帯域を変更できる IIR ノッチフィルタと IIR ピークフィルタを用いた。次章では、本章で述べた提案手法を用いてフィードバック実験を行い、本人 HRTF との音像定位精度について比較する。

第 4 章

提案手法の検証

本章では、提案手法を用いて作成した HRTF から作成されるバイノーラル音源の有用性について検証する。検証には、音像定位評価実験を行う。本人の HRTF を用いて作成するバイノーラル音源と音像定位精度について比較する。まずはじめに、4.1 節では提案手法の比較に用いるため、被験者本人の HRTF 測定について述べる。次に、4.2 節において HRTF モデルの定位精度を示す。最後に 4.3 節では HRTF モデルの補正についてと補正の結果得られた HRTF の定位精度の結果を示す。

4.1 本人 HRTF 測定条件

HRTF 測定について述べる。本論文の評価試験には本人の HRTF から作成されるバイノーラル音源と、フィードバック実験で補正された HRTF から作成されるバイノーラル音源の音像定位精度を用いる。本論文では、水平面上の音源に対して、音像定位試験を行うため、それに対応する HRTF を測定する。図 4.1 に示すように、被験者を円状に囲むよう音源を設置し、被験者の正面から 30° 間隔 12 方向の伝達関数と、被験者がいない場合の伝達関数を測定する。測定された伝達関数を用いて、HRTF を作成する。測定環境は、無響室であり、音源としては TSP 音を用いた。音源を録音するため、マイクロホンを外耳道入り口に設置する。表 4.1 に測定条件を示す。

表 4.1 測定条件

測定環境	無響室
出力信号	TSP 音
サンプリング周波数	48kHz
音源距離	150cm
角度分解能	30°
マイクロホン	Audio-technica BP896cW
スピーカー	ECLIPSE TD307MK2A

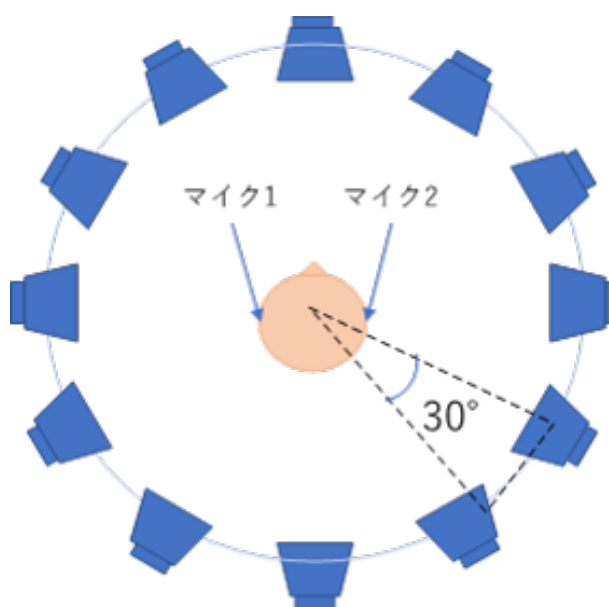


図 4.1 測定図

4.2 HRTF モデルの音像定位精度

まず、HRTF モデルを用いてバイノーラル音源を作成し、被験者実験を行う。HRTF モデル作成の際、主成分分析を用いた。主成分分析から算出される HRTF モデルは虚数成分を失っており、虚数成分は、両耳に到達する時間差の情報を含む。そのため、バイノーラル音源作成には、両耳間時間差情報の付与が必要となる。本論文では、両耳間時間差情報を付与するため遅延器を設け、時間差を再現した。表 4.3 に角度ごとの時間差を示す。角度ごとの時間差には個人差があるが、今回は文献 [17] を参考に時間差を設けた。HRTF モデルの選択には、1 方向 16 通りのバイノーラル音源を 1 方向の

みの再生から最良と考えられるものを被験者ごとに選択した。これを、12 方向全てにおいて試聴を行い被験者実験に用いる HRTF モデルを決定した。また、音像定位評価試験ではヘッドホンを使用している。ヘッドホン特性が音像定位試験に影響を及ぼさないようヘッドホン特性を測定し補正している。12 方向×2 の 24 方向を 1 セットとし、ランダムで提示される角度ごとの音源を試聴し、それぞれで知覚した方向を回答させた。表 4.2 に示す条件を用いて、音像定位試験を行った。結果を図 4.2 から図 4.5 に、それぞれの音像定位の正解率を表 4.4 に示す。

表 4.2 音像定位試験条件

信号 $x(t)$	ホワイトノイズ
ヘッドホン特性	補正
被験者数	4 人
提示回数	12 方向 6 回

表 4.3 各方向の時間差

提示方向	時間差
0°、180°	0[ms]
30°、330°	0.2[ms]
60°、300°	0.4[ms]
90°、270°	0.7[ms]
120°、240°	0.2[ms]
150°、210°	0.4[ms]

図 4.2 から図 4.5 に示すグラフは、ある角度のバイノーラル音源を受聴した際の被験者の回答した角度の回数を表すバブルチャートである。グラフの横軸は被験者に対して提示した角度、縦軸は被験者が知覚した音源の方向を表している。グラフ上の円は、回答回数を大きさで表している。すなわち、仮に被験者がすべての音源にたいして、正しい方向を知覚・回答した場合には、グラフ左下から右上にかけて対角線上に大きな円が並ぶことになる。ここで、青色の円は HRTF モデルを用いた音像定位の結果、黒い枠線のみ円は本人の HRTF を用いた結果を示している。

図 4.2 から図 4.5 より 0° (180°)、90° 270° の方向はおおむね正確に知覚していることがわかる。図 4.5 の被験者 D のみ前後の誤りが少ない結果となった。また、被験者 C は 0° 方向を全く知覚しない結果となった。被験者 A～C は前後の誤りが生じている。また、全被験者において 0°、90° 270° 以外の方向で方向知覚のずれが生じ

ている。HRTF モデルの選択には、最良と考えられるものを被験者ごとに選択したが、被験者 D 以外は正答率は 50% に満たない結果になった。個人に合う音像定位を提供するためには改善が必要である。そのため、本節の結果を参考に定位精度が悪い方向に対して補正フィルタを用いて補正を行う。

表 4.4 HRTF モデルの音像定位精度

被験者番号	本人 HRTF	HRTF モデル
A	71%	44%
B	67%	48%
C	50%	40%
D	75%	60%

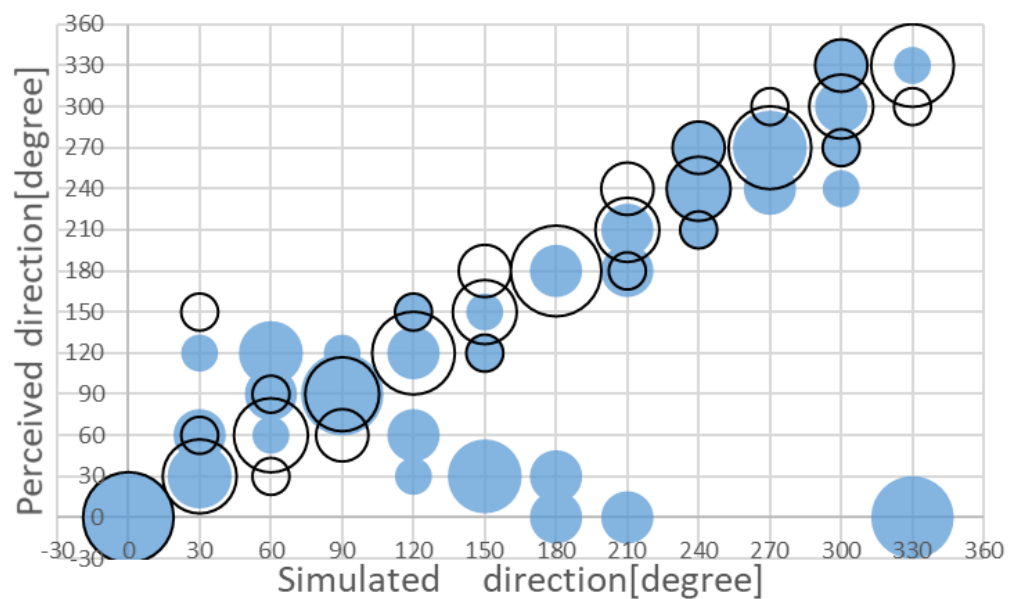


図 4.2 被験者 A の試験結果

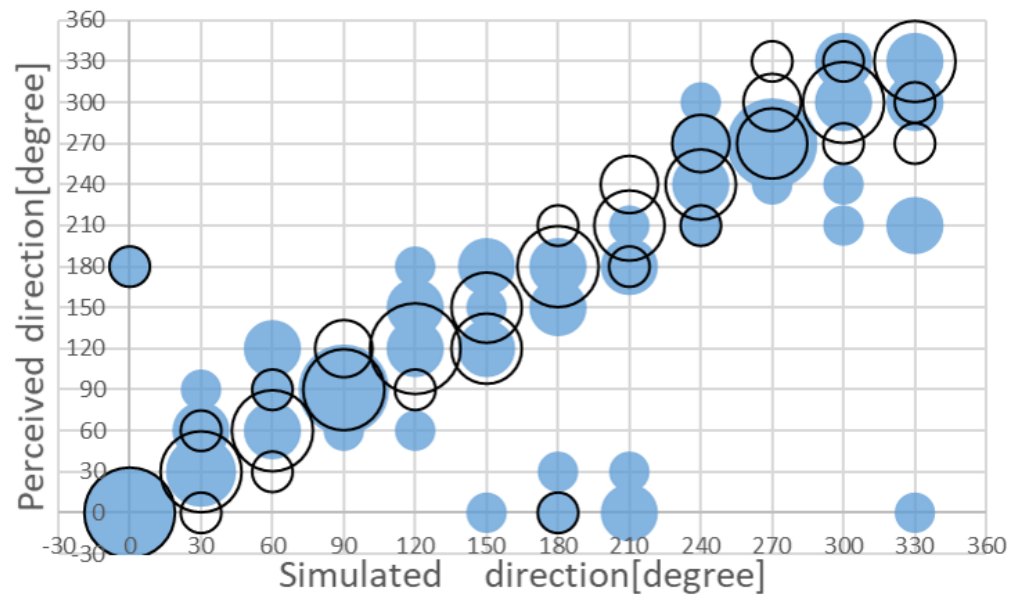


図 4.3 被験者 B の試験結果

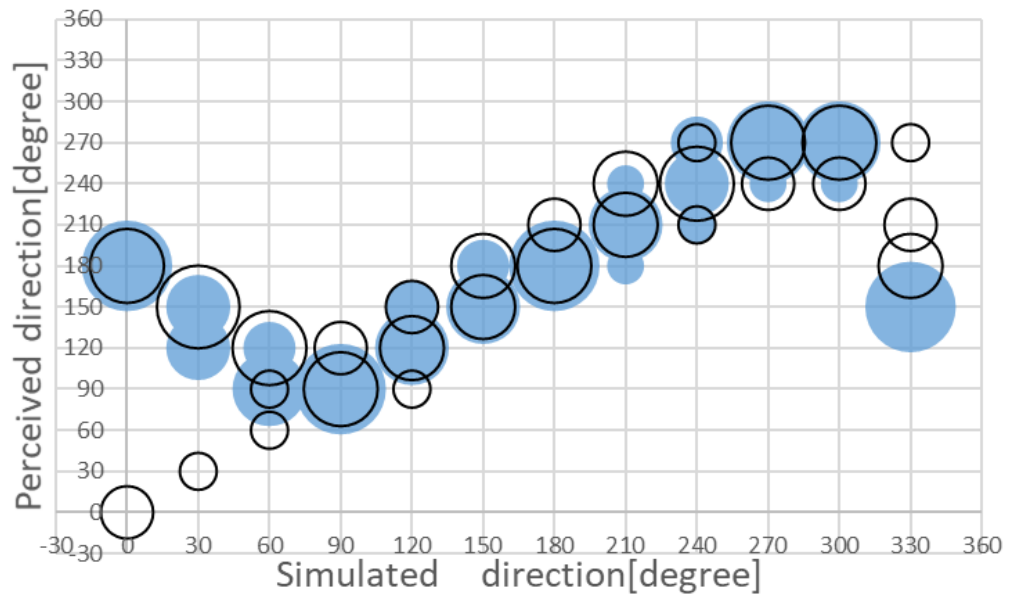


図 4.4 被験者 C の試験結果

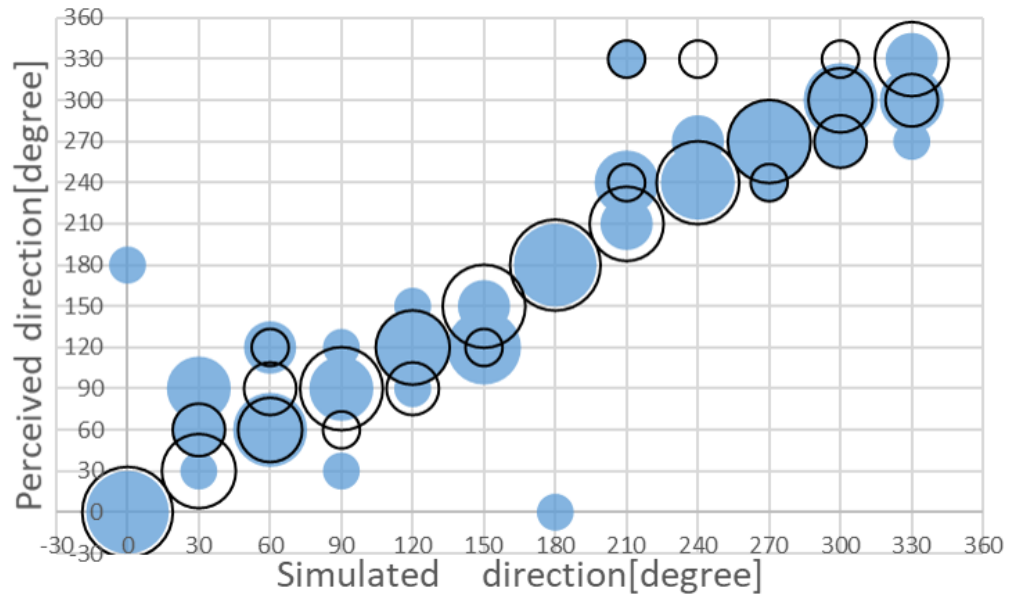


図 4.5 被験者 D の試験結果

4.3 補正処理後の音像定位評価

4.3.1 HRTF 補正結果 1

例として、被験者 A の 180° 方向の補正手順を示す。被験者 A は、 180° の音像提示に対して、 0° や右方向の 30° と知覚していた。また、被験者は、音の大きさから 0° と 180° の違いが分かるが、音源が頭部真上にあるため、分かりづらいと評価した。まず、両耳の P1 ピーク振幅値をピークフィルタによって大きい振幅と同等に変更することで、左右のずれを補正した。今回の補正では、両耳の N1 ノッチの周波数を 1000Hz 変更させたところで、被験者 A は水平面に感じると評価した。図 4.6 に受聴側である右側の補正した HRTF を、図 4.7 に補正した反受聴側である左側の補正した HRTF を示す。比較のため赤色は HRTF モデルを示し、青色は補正した HRTF、緑色は補正した HRTF を示している。補正した HRTF を被験者 A が試聴した結果、 180° 方向と知覚した。同様に被験者 B と C にも行った。被験者 C の 0° 方向の誤知覚は、同様の処理に加え振幅を増加させることで 180° と聞き比べると違いがわかると評価を得たが前方向に音像が定位するという評価はなかった。

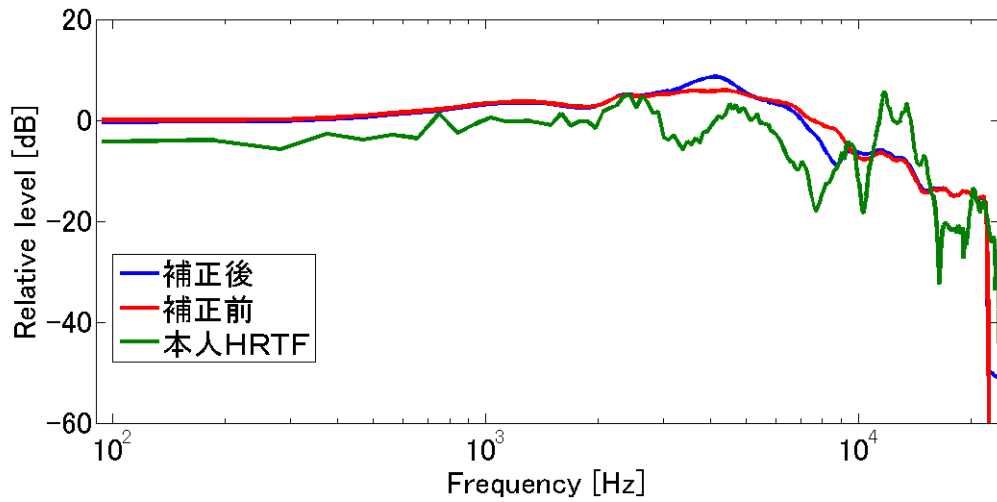


図 4.6 左側 HRTF

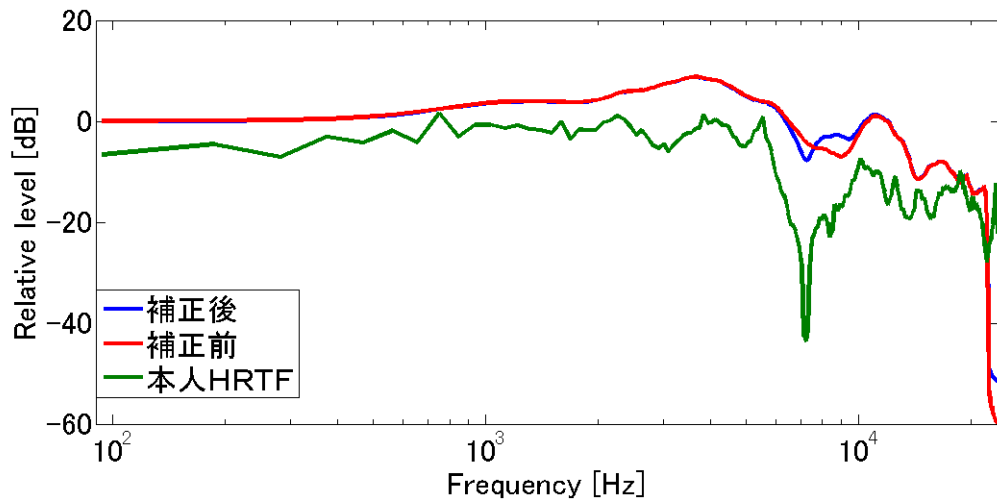


図 4.7 右側 HRTF

4.3.2 HRTF 補正結果 2

例として、被験者 A の右側 60° 方向の補正手順とを示す。被験者 A は、 60° の音像提示に対して、 60° を 120° と知覚する前後誤りや 90° に知覚する定位のずれが生じている。まず受聴側の N1 ノッチを変更することで 90° への誤知覚を補正する。被験者 A が 60° を正確に知覚するためには、N1 ノッチをより 0° 方向の低周波帯域に設定する必要がある。N1 ノッチ位置の変更決定には、正確に知覚できた 0° の N1 ノッチ周波数 ω_{A0° と 60° の N1 ノッチ周波数 ω_{A60° を用いて行った。次に、被験者

A は 60° を 120° と知覚する前後誤を起こしているため補正を行う。補正には、両側の HRTF に存在する P1 ピークと P1 ピーク以下の周波数帯域の振幅を増加させ、前方向と知覚するまで繰り返した。図 4.8 に受聴側である右側の補正した HRTF を、図 4.9 に補正した反受聴側である左側の補正した HRTF を示す。比較のため赤色は HRTF モデルを示し、青色は補正した HRTF、緑色は本人の HRTF を示している。受聴側である HRTF は N1 ノッチの位置を 1500 Hz 変更している。反受聴側は 1000Hz 帯にあったくぼみを消し P1 ピークを増幅させている。被験者 A が試聴結果、ノッチとピークを変更したことにより、音の大きさが増し前方向から聞こえると主観評価を得た。このように補正が必要な角度において調節を行う。

次節では、補正した HRTF と本人の HRTF の比較を行う。以上を全方向に適用する。なお、左右を逆転させ音像定位試験を行い補正の必要ないと被験者が主観評価した場合に補正を行わない。被験者 A から被験者 C の 3 人には、方向の補正と、前後方向の補正を行った。被験者 D は方向の補正のみを行った。

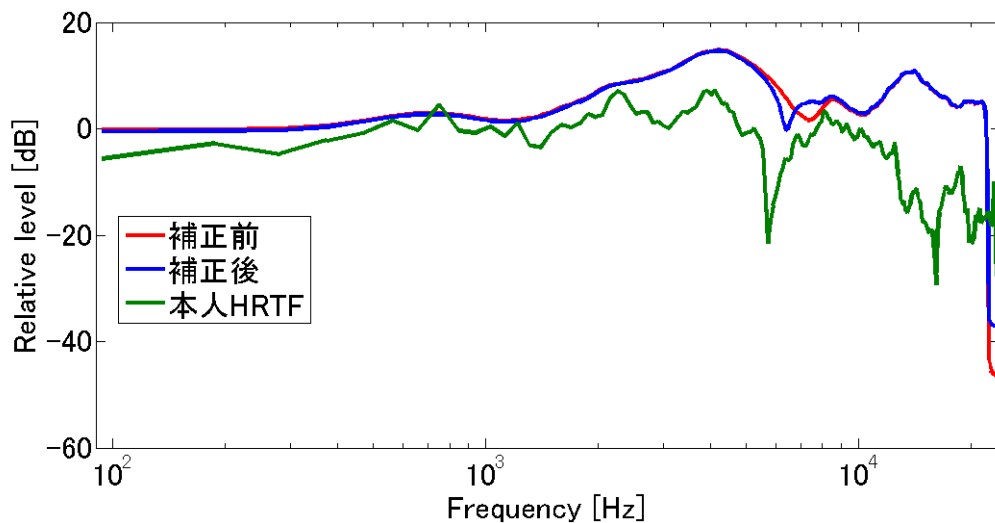


図 4.8 受聴側 HRTF

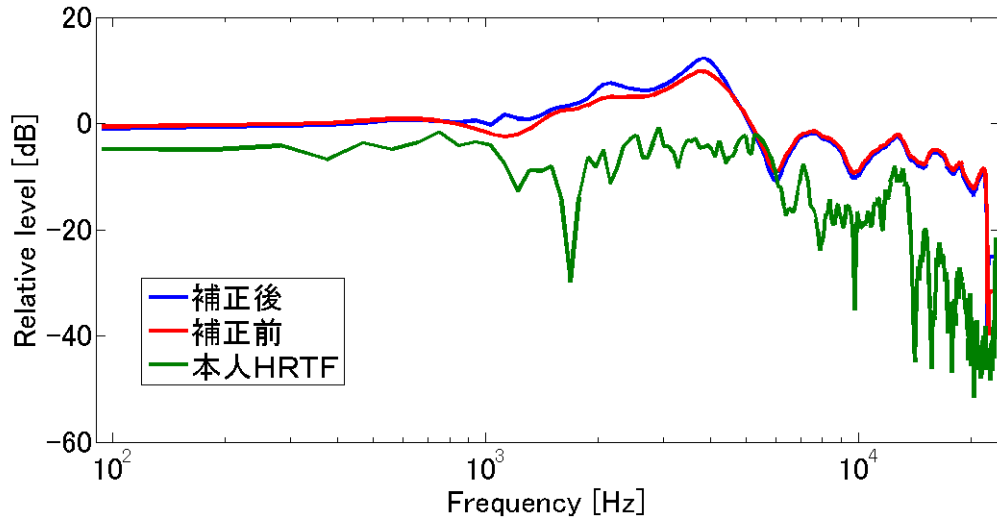


図 4.9 反受聴側

4.3.3 音像定位評価試験

本節では、HRTF モデルを補正した HRTF と本人 HRTF の音像定位評価試験比較を行った結果について述べる。前節で述べた補正フィルタを用いて作成した HRTF より各方向のバイノーラル音源を提示する。補正を行う場合のバイノーラル音源作成手順を図 4.10 に示す。また遅延器や、実験条件、音像の提示方法は 4.2 節で述べた音像定位評価試験と同様の条件で行った。

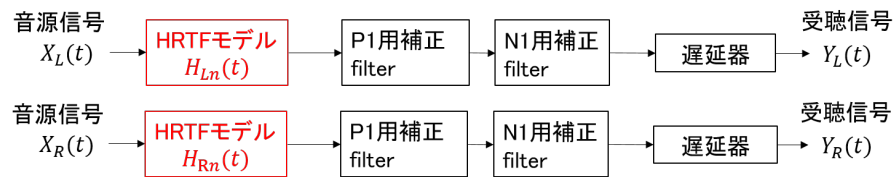


図 4.10 再生フローチャート

図 4.11 から図 4.14 にそれぞれの被験者の定位評価結果を示す。また、音像定位の正解率を表 4.4 に示す。図 4.11 から図 4.14 青色の円は HRTF モデルを用いた音像定位の結果、黒い枠線のみ円は本人の HRTF を用いた結果を示している。図 4.11 から図 4.14 より、それぞれの音像定位精度が向上する結果が得られた。

図 4.11 と図 4.12 の被験者 A、B は HRTF のピーク振幅を変更することで前後の誤りが改善した。また、被験者 C において、反受聴側のピーク振幅を増加させたところ音が大きくなり迫力が増したと主観評価を得た。これは、反受聴側の P1 ピークとピー

ク以下の周波数帯域の振幅を増減させることで音量レベルが増加し前方向と知覚したと考えられる。また、被験者 C を除く 3 人は、 0° から 330° まで 30° 刻みで音源を再生したとき、円を描くように音像が定位していると回答した。図 4.13 の被験者 C は、前後の誤りが改善されず、定位精度の向上も低い。特に前方向の音像提示が、横と後ろ方向に提示される結果となった。被験者 C は、HRTF モデル補正の際、ピークの振幅を増減させることで迫力が増し音の距離感のみが変化すると回答し、 0° から 330° まで 30° 刻みで音源を再生したとき、円を描くように音像が定位しないという評価を得た。被験者 C は、本人の HRTF を用いた場合でも前方向の知覚が低いことから、ヘッドホンで音を聞くと、音を頭の中に感じてしまう頭内定位が発生していると考えられるため、前方に全く知覚しない結果となったと考えられる。また、音像方向のずれについて検討する。音像方向のずれは全ての被験者で改善がみられた。これは、受聴側のノッチを変更したためであると考えられる。その理由としては、人は両耳間のレベル差の絶対値でも方向を知覚している。レベル差は高周波数ほど影響を及ぼし、絶対値が小さいほど正面方向に、絶対値が大きいほど横に音像は定位する。被験者実験の回答をフィードバックする際にノッチ位置の変更で音が左右に移動したとの回答が多くあり、レベル差の変化が影響したと考えられる。本論文で行った補正では、音が大きく聞こえる受聴側のノッチ位置を変更したことでレベル差に違いが生じ方向を変化させたと考えられる。

全体の結果としては、HRTF モデルを用いる場合と比較すると定位精度が向上した。しかし、本人の HRTF を用いた場合と同等の定位精度は得られなかった。その理由として両耳間での時間差が考えられる。人が音を知覚する際、主に 2 つの時間差から知覚する。耳に入力されるまでの時間差と、耳内での周波数ごとによる遅延である。

1 つ目の入力される時間差については次のように考えられる。耳と耳の距離が異なることで、人によって音の到達時間にも差が生じている。この耳に到達するまでの差は主に低周波帯域である 1600Hz 以下に影響があり、今回行った被験者実験では、耳に入力する時間差に文献にある角度ごとの平均値を使用した。時間差による個人差は $0.08[\text{ms}] \sim 0.01[\text{ms}]$ でありこの差が影響を及ぼしていると考えられ、個人に適した時間差を選択するフィードバック補正が必要であると考えられる。2 つ目は、ノッチフィルタとピークフィルタによる群遅延が考えられる。 1600Hz 以上の周波数は、耳内での入力信号の時間差が影響する。図 4.15 に 4.3.2 節で示した被験者 A の HRTF 群遅延特性を示す。青線は反受聴側、赤線は受聴側を示している。今回補正には、ノッチフィルタとピークフィルタを用いた。これらは、中心周波数を中心に遅延を発生させ、左右で異なる周波数帯にフィルタを使用しているため、異なる群遅延が発生していることがわかる。本論文で作成した HRTF モデルは、時間差がなく遅延時間は $0[\text{ms}]$ であるため音像定位試験の際に遅延器を付属している。しかし、補正を行った HRTF は、

左右それぞれに異なる補正フィルタを用いていることから遅延器を付属する前に両耳間で時間差が生じており音像定位に影響しているのではないかと考えられる。そのため、この群遅延によって音の知覚が変化し、補正した HRTF の定位精度結果に良い影響、悪い影響を及ぼしたとも検討することができる。群遅延が及ぼす影響や、フィルタによる遅延補正についても今後検討する必要があると考えられる。

表 4.5 補正後の音像定位精度

被験者番号	本人 HRTF	HRTF モデル	補正後HRTF
A	71%	44%	60%
B	69%	48%	65%
C	55%	45%	49%
D	75%	60%	66%

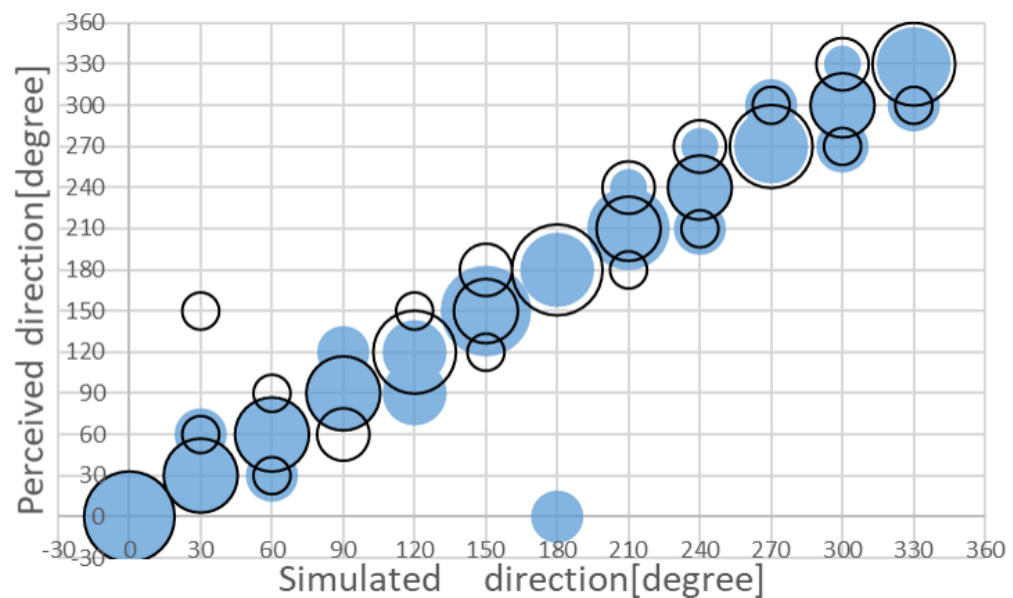


図 4.11 被験者 A の試験結果

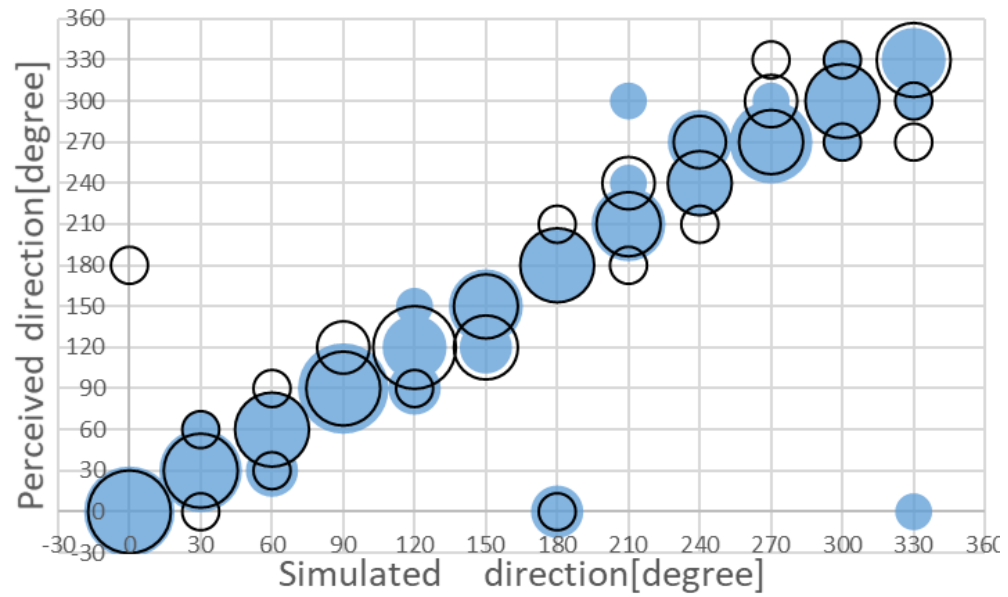


図 4.12 被験者 B の試験結果

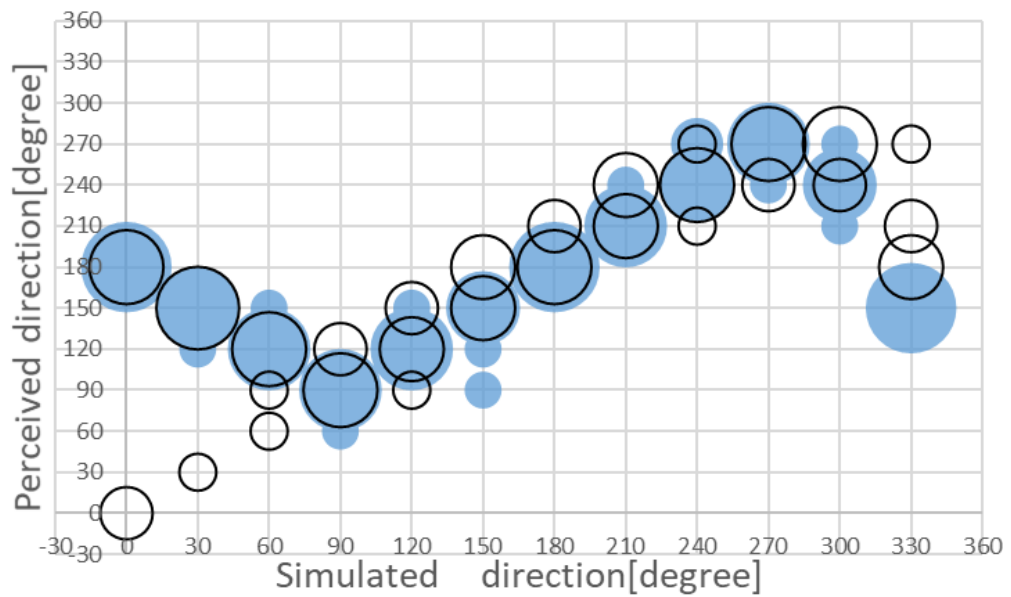


図 4.13 被験者 C の試験結果

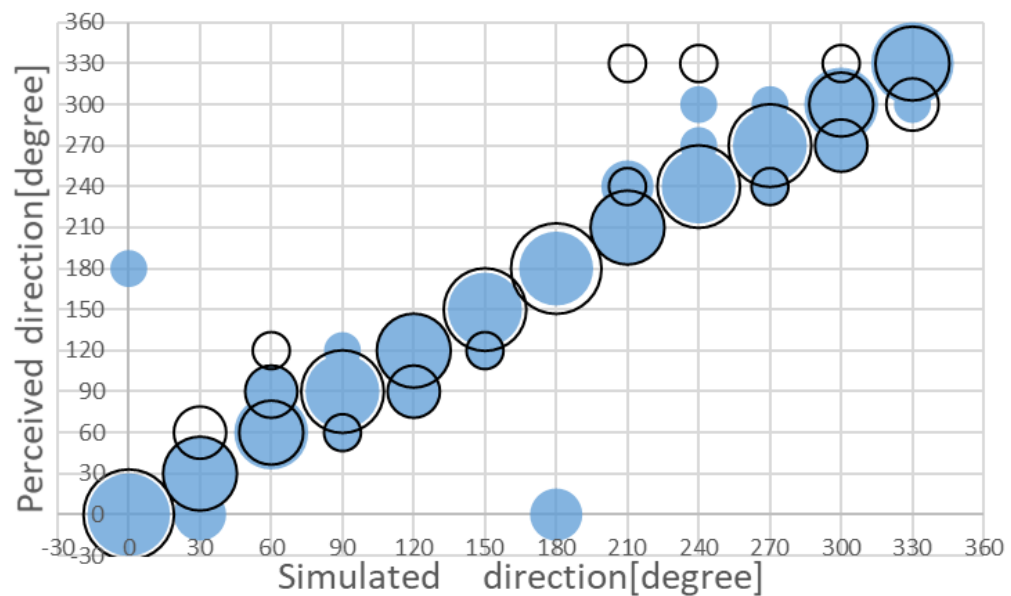


図 4.14 被験者 D の試験結果

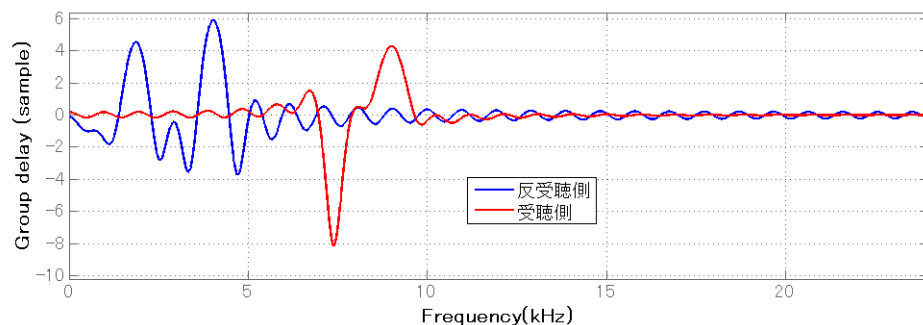


図 4.15 フィルタの群遅延

4.4 まとめ

本章では、音像定位評価実験を行い、提案手法を用いて作成した HRTF から作成されるバイノーラル音源の有用性について検証した。まず、4.2 節では HRTF モデルの音像定位評価実験を行った。被験者実験のときは、 30° 間隔の音源をランダムに提示し音像定位試験を行った。結果は、最良の HRTF モデルを選択したとしても、他人の HRTF 使用による前後誤りや定位方向のずれが生じた。そのため、個人に適応する HRTF を作成するためには、補正処理が必要なことを示した。次に、4.3 節では HRTF モデルを補正し、4.2 節で行った条件で被験者評価実験を行った。結果は前後誤りや定位方向のずれが抑えられた。しかし、本人の HRTF を使用した場合より高い定位精度

は得られなかった。この原因として、耳に入力されるまでの時間差と、耳内での周波数による遅延に影響があると考えられる。これを解決するために、個人に適した時間差を選択するフィードバック補正が必要であると考えられる。また、補正には複数のフィルタを用いていることからフィルタによる遅延が音像定位に影響しているのではないかと考えられるため、検討が必要である。

第 5 章

終わりに

本論文では、特殊な設備や装置を必要としない個々のユーザに適した HRTF の構築を目的とし、非線形手法を用いて HRTF モデルを作成し、HRTF モデルを用いて被験者実験を行った。その後、被験者実験の主観回答をフィードバックし補正を行うことで HRTF の最適化を行い、次第に個々のユーザに適した HRTF を作成する手法を提案した。第 1 章では、HRTF をもちいた音像定位について述べ、その問題点を明記し、本論文の研究目的を述べた。第 2 章では、本研究の基礎となる知識・手法と提案法で用いるカーネル主成分分析と k-means 法について述べた。また、人の音像定位と HRTF の個人性と問題点について述べた。

第 3 章では、提案法である HRTF モデルの作成方法について述べ、作成手順と作成結果を示した。また、個人化の問題を解決するための補正フィルタと補正手順を述べた。第 4 章では、被験者実験を通して、提案手法の音像定位精度について検討した。まず、HRTF モデルの被験者実験を行い、補正の必要性を示した。その結果を参照し、誤知覚している角度に対しての補正を行った。補正を行った結果、HRTF モデルの音像定位精度より精度が向上した。このことから、フィードバックを用いた補正が効果的であることを示した。しかし、本人の HRTF を用いた場合と比較した場合は、補正を行っても定位精度がやや劣る結果となった。その要因としては、耳に入力されるまでの時間差と、耳内での周波数による遅延に影響があると考えられる。これを解決するために、個人に適した時間差を選択するフィードバック補正が必要であると考えられる。また、補正には複数のフィルタを用いていることからフィルタによる遅延が音像定位に影響しているのではないかと考えられるため、検討が必要であると考えられる。

謝辞

本研究を修士論文として形にすることができたのは、なかなか結果の出ない実験を最後まで温かく見守り、適切な指導を賜った主査の杉田 泰則 准教授や、論文の審査において的確なご指示を頂きました、岩橋 政宏 教授ならびに圓道 知博 准教授、そして多くのご指摘をくださいました信号処理応用研究室の同期・後輩の皆様のおかげです。また、勉学に励む機会を与え様々な面において支えていただいた両親に感謝いたします。協力していただいた皆様へ心から感謝の気持ちと御礼を申し上げたく、謝辞にかえさせていただきます。

付録 A

付録

A.1 HRTF 補正結果

4 章では、被験者 A の結果を示した。提示していない右半面の補正結果を示す。示していない補正は、 0° （右側）、 30° 、 120° （右側）、 150° （右側）である。

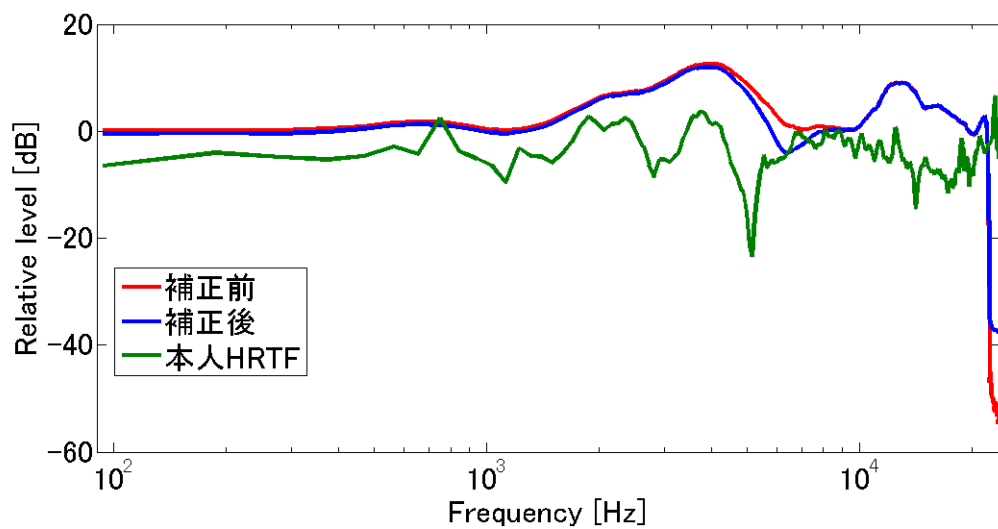


図 1.1 受聴側 0°

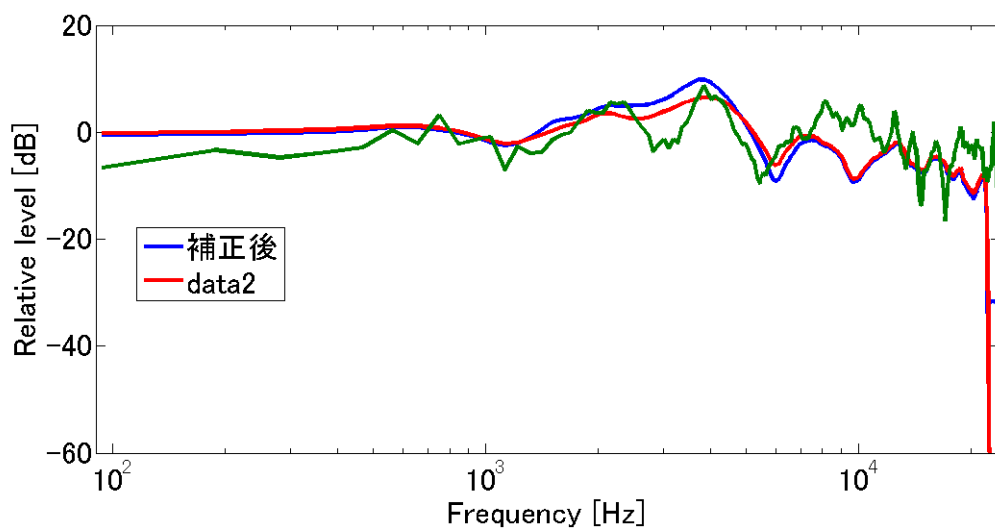


図 1.2 反受聴側 30°

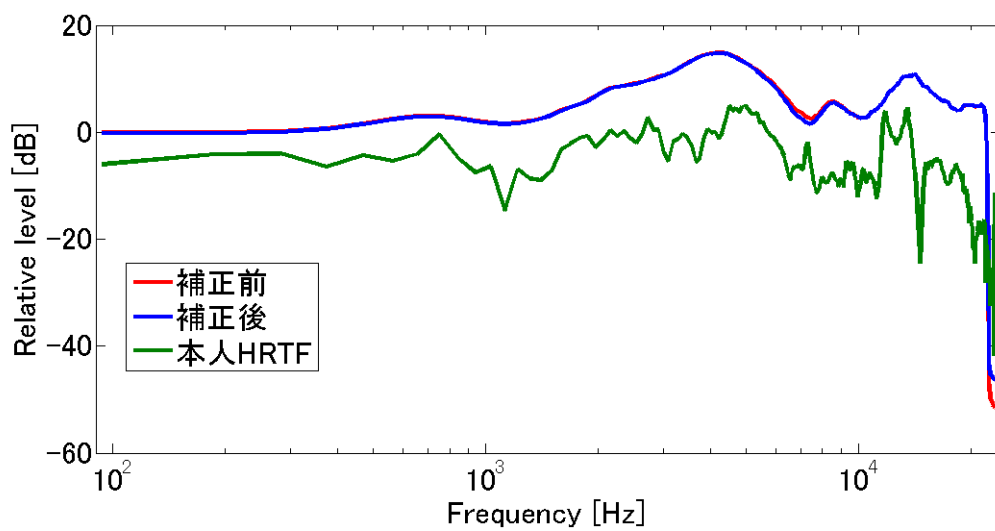


図 1.3 受聴側 30°

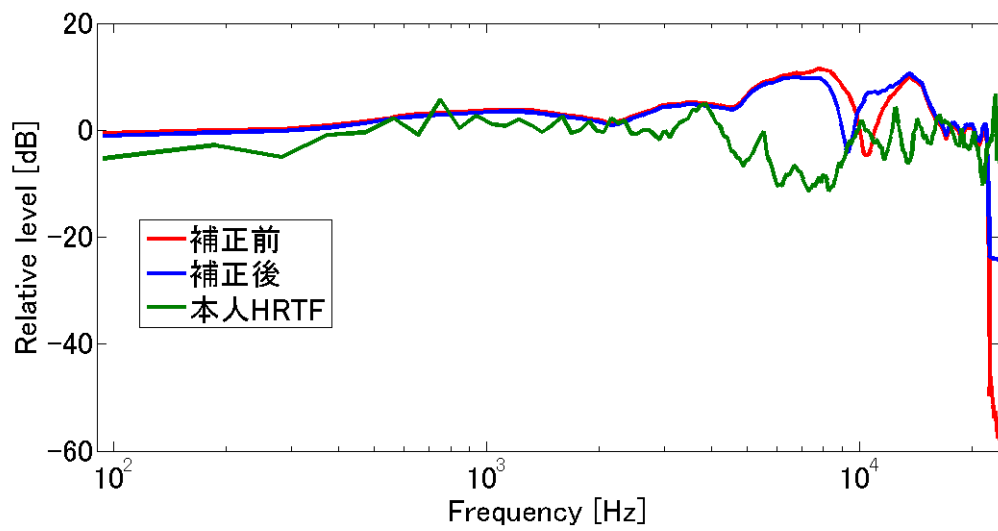


図 1.4 受聴側 120°

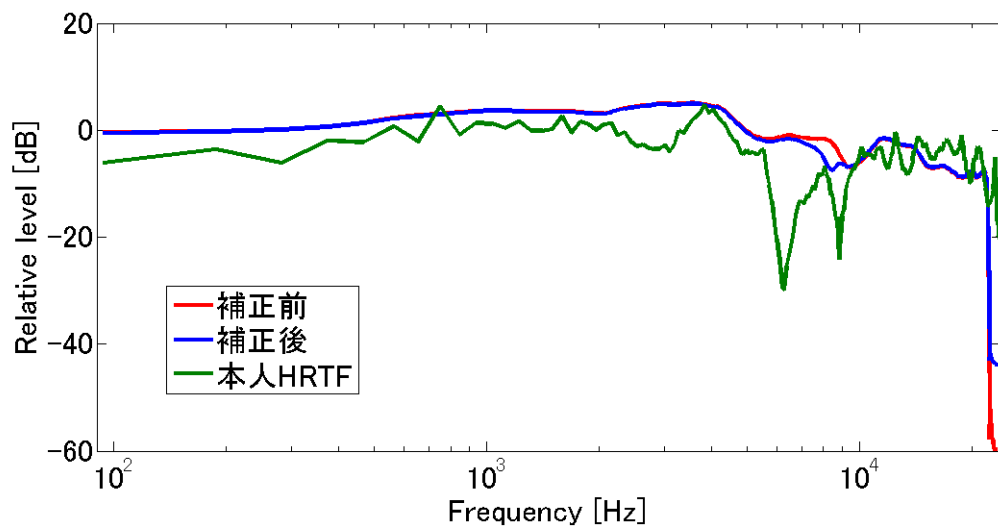


図 1.5 受聴側 150°

A.2 HRTF モデル被験者実験

4 章では、4 人の被験者の結果を示した。提示していない 3 人の被験者の実験結果を示す。

表 1.1 HRTF モデルの音像定位精度

被験者番号	本人 HRTF	HRTF モデル
E	72%	61%
F	55%	45%
G	45%	38%

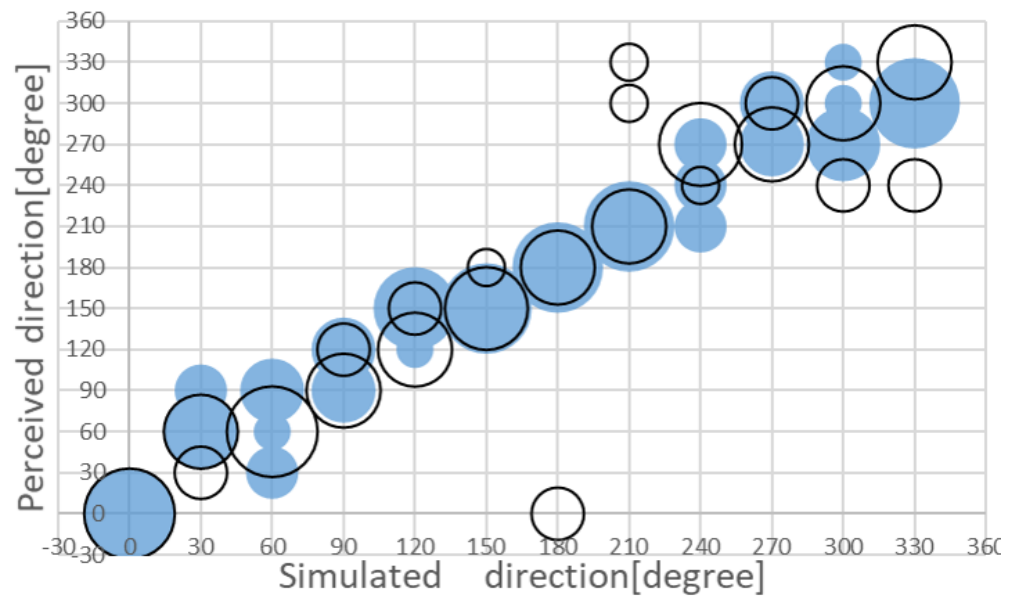


図 1.6 被験者 E の試験結果

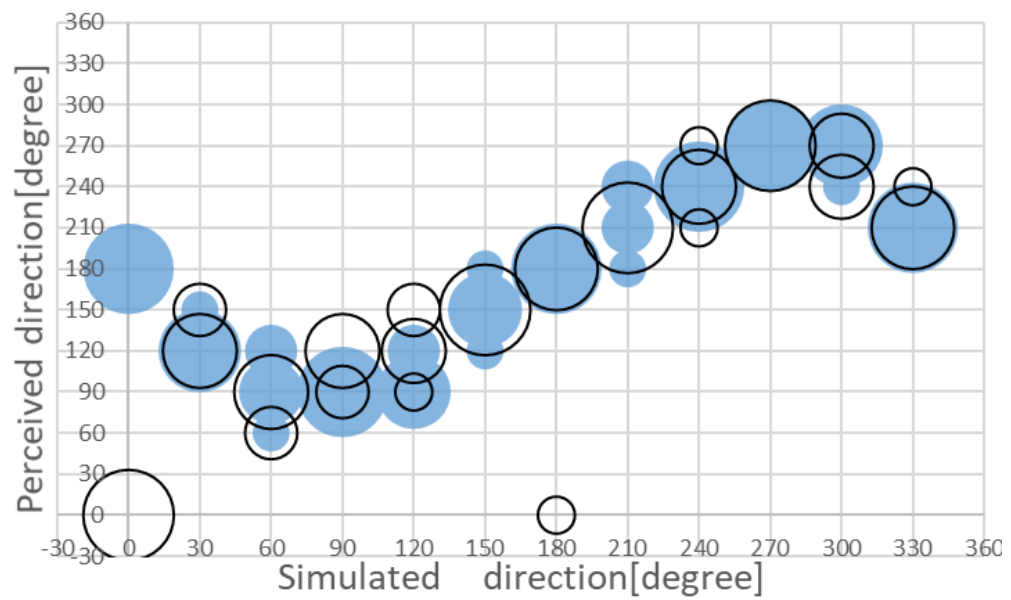


図 1.7 被験者 F の試験結果

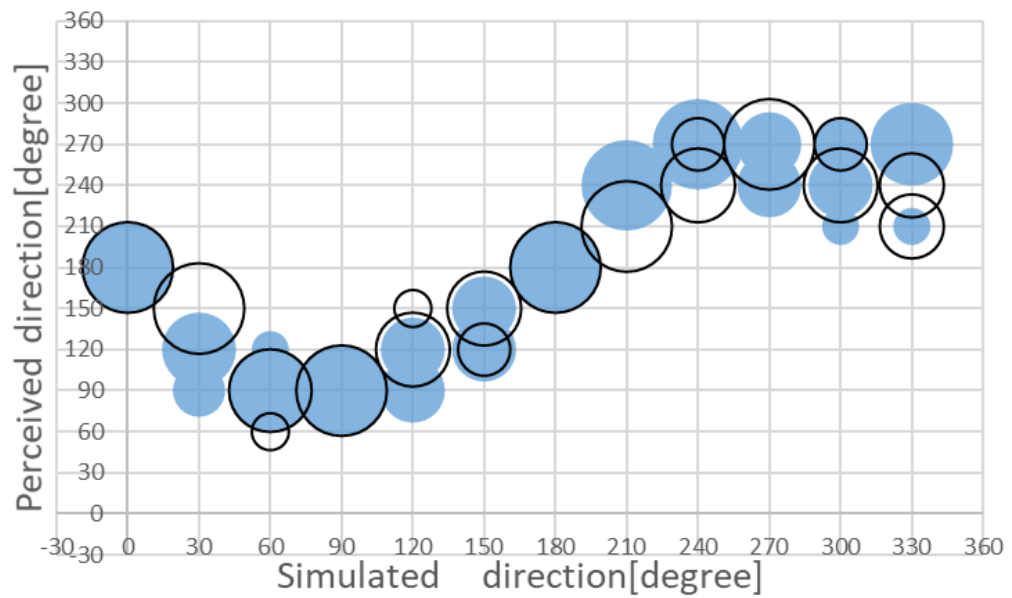


図 1.8 被験者 G の試験結果

A.3 補正処理後の被験者実験

4 章では、4 人の被験者の結果を示した。提示していない 3 人の被験者の実験結果を示す。

表 1.2 補正処理後の音像定位精度

被験者番号	本人 HRTF	補正処理後の H R T F
E	72%	69%
F	55%	57%
G	48%	43%

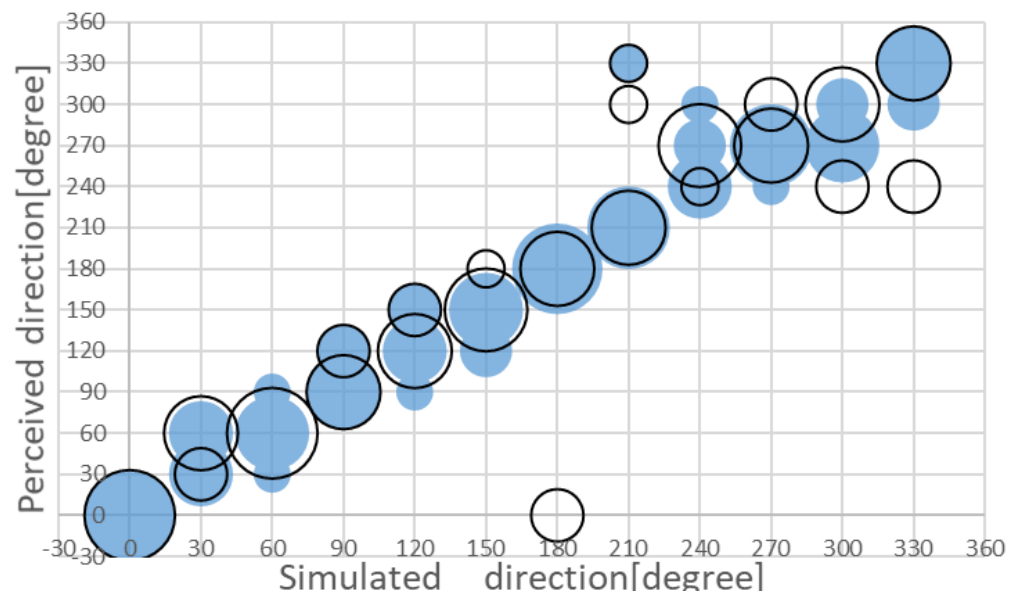


図 1.9 被験者 E の試験結果

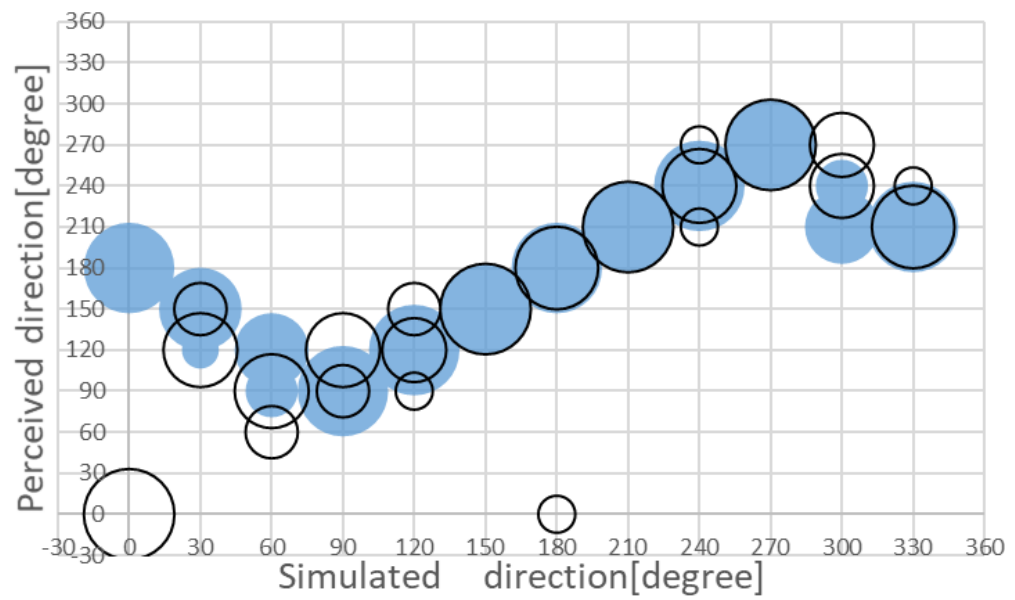


図 1.10 被験者 F の試験結果

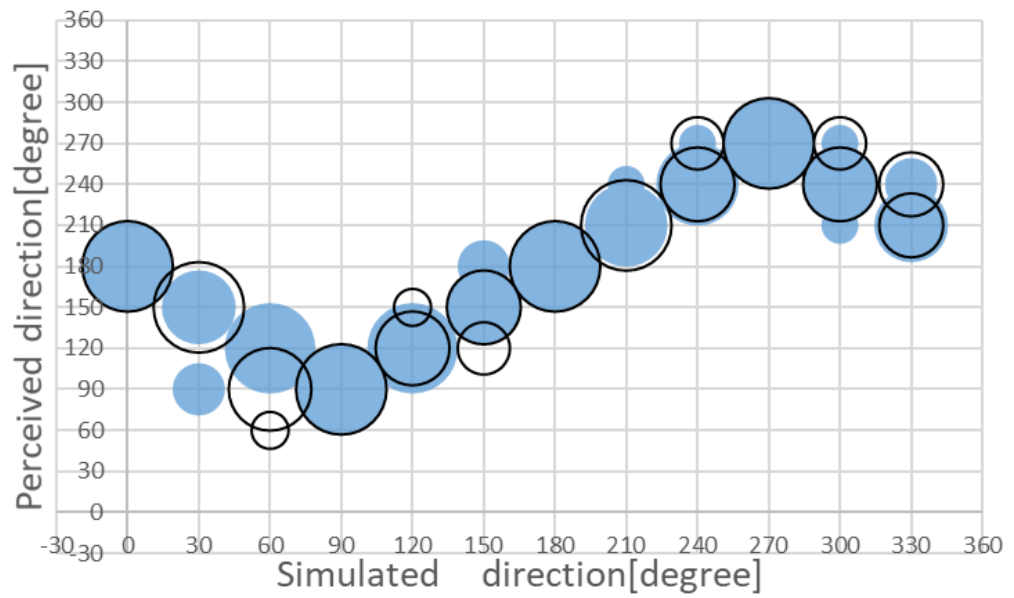


図 1.11 被験者 G の試験結果

参考文献

- [1] T.Miura,Y.Ebihara,M.Sakajiri and T.Ifukube"Utilization of auditory perceptions of sounds and silent objects for orientation and mobility by visually-impaired people,"IEEE International Conference on Systems,Man, and Cybernetics,(2011)p.1080-1087
- [2] Y.Seki and T.Sato,"A training system of orientation and mobility for blind people using acoustic virtual reality,"IEEE Transaction on Neural Systems Rehabilitation Engineering, vol.19,(2011)No.1,pp.95-104
- [3] 岩谷幸雄,"頭部伝達関数による音像定位", 日本音響学会誌 73 巻 3 号 (2017)p.173-180
- [4] 飯田一博,"頭部伝達関数の基礎と 3 次元音響システムへの応用", 音響テクロノジーシリーズ 19, コロナ社,(2017)
- [5] 飯田一博,"頭部伝達関数の個人性"日本音響学会誌 73 巻 3 号 (2017),p181-188
- [6] Y.Iwaya,"Individualization of head related transfer functions with tournamentstyle listening test",Acoust.Sci.Tech.27(2006),p.340-343
- [7] S.Yairi,Y. Iwaya and Y.Suzuki, "Individualization feature of head-related transfer functions based on subjective evaluation",ICAD(2008), p.24-27
- [8] David Schonstein,Brian F.G. Katz,"HRTF selection for binaural synthesis from a database using morphological parameters."ICA(2010),p.23-27
- [9] 飯田一博,"受聴者の耳介形状による頭部伝達関数の個人化", 電気情報通信学会技報 vol.18(2018),p.97-110
- [10] 飯田一博,"頭部受聴者の頭部形状による両耳間差キューの個人化",TVRSJ(2017),vol.22,No3,p.405-412
- [11] 松井健太郎, 安藤彰男,"主成分分析に基づく個人性頭部伝達関数の推定", 音響特集号-スーパーハイビジョン音響収音・再生と評価,NHK 技研 RD(2011)26,p49-57
- [12] 西野隆典, 中井勇祐,"重回帰分析に基づく頭部伝達関数の推定", 電子情報通信学会論文誌 vol.J84-A(2001),126,p.260-268
- [13] 飯田一博, 森本政之,"空間音響学", コロナ社,(2010)

- [14] 鷺沢嘉一, 田中正行"中心化部分カーネル主成分分析によるデノイジング", 第9回 情報化科学技術フォーラム, FIT(2010), p395-399
- [15] Laurens van der Maten "Dimensionality Reduction: A Comparative Review." TiCC TR, Netherlands, (2009)
- [16] 小野田崇, 酒井美帆 "初期値設定法による k-means 法の性能比較." Fuzzy System Symposium 27th, (2011) p.231-236
- [17] 東北大学電気通信大学先端音響情報システム研究室, : The RIEC HRTF Dataset, <http://www.riec.tohoku.ac.jp/pub/hrtf/index.html>
- [18] 飯田一博 (千葉工業大学), "頭部受聴者の頭部形状による両耳間差キューの個人化", TVRSJ(2017), vol.22, No3, p.405-412