

外耳道入口までの耳介付近伝達関数の補正による 臨場感再現のためのヘッドホンの定位改善

Improvement of sound localization of headphones for realistic reproduction by calibrating acoustic characteristics up to the entrance of the ear canal

工藤 翔輝

Kakeru Kudo

法政大学情報科学部デジタルメディア学科
kakeru.kudo.7w@stu.hosei.ac.jp

Abstract

Due to COVID-19, the number of online live performances and the opportunities to listen to music with headphones have increased. However, with headphones, the sound image is localized in the head, losing the sense of reality. Therefore, in this research, we propose a DSP to reproduce the sense of reality and aim at improving localization. A dummy head measures the effects of reflections from the headphones to the entrance of the ear canal. Measure, estimate the transfer function, and convolve its inverse characteristic with the sound source. Here, by performing HRTF/free sound field characteristic calibration so as not to include the properties of the dummy head and free sound field, only the effect that occurs in the space between the earcup and the entrance of the ear canal is estimated as a transfer function. The evaluation method for this calibration system is the same as for stereophonic sound: the accuracy rate of sound image localization and frequency characteristics. For sounds recorded from multiple directions in advance, The inverse filter of the transfer function is convoluted, and the user evaluates the localization accuracy rate. As a result, the accuracy rate increased by 90% compared to the original sound source, and the localization in the head was improved, making it clear that the calibration was effective.

1 はじめに

近年の新型コロナウイルスのパンデミックは世界中の音楽シーンに大きな影響を与え、大規模なフェスティバルやツアーのような大型のライブは自粛されている。その一方でオンラインライブや SNS、動画配信サイトを通じた配信での活動が活発になった。しかし、オンラインライブではリスナーそれぞれが使用する音響機器が違いうえにライブのクオリティや臨場感を損なう一面がある。これはスピーカーやヘッドホンは機器ごとに周波数特性が異なる為である。また、音像定位についてはスピーカーとヘッドホンは大きく異なり、これにより聴感も大きく異なる。スピーカーは音像が頭外に定位し、ヘッドホンは音像が頭内に定位する。実際のライブでは、会場に設置されたスピーカーから出力された音を聴取する為、頭外定位となる。オンラインライブの臨場感再現に適しているのはスピーカーでの再生であるが、音響機器の普及状況や移動中や騒音問題等の再生環境を考慮した時、手軽に楽しめるという点ではやはりヘッドホンが優位である。これらの観点から、ライブの臨場感を自宅や外出先でも再現する為、臨場感の一因と考えられるヘッドホンの頭内定位を改善する必要がある。

前述した改善を行うにあたり問題となるのが、身体構造の差である。人間の外耳、中耳、内耳の構造は個人差があり、計測する事ができる。しかし、リスクやコストがかかる為、ユーザー自身が計測を行う事は考えにくい。また、外耳道から鼓膜を通過した後の耳小骨、蝸牛、半規管、聴神経を含む内耳の実際の処理に個人差があるかは明確になっていない。一方で目視で確認できる耳殻から外耳道入口までの個人差は比較的シミュレートしやすい。

本研究ではスピーカーとヘッドホンの差を無くす為、外耳道入口までの音響信号に対してイコライジングを施し、頭内定位の改善を行う。

2 従来研究

2.1 音像定位

音像 (Sound localization)[1] とは人間が音を受聴した時に知覚される音の姿のことである。特定の物体 (発音オブジェクト) から発生した音の姿を指すことがほとんどである。音像として知覚される音は、元々の音源が持つ音色や、指向性、部屋の反射・残響を含んでいる。人間は、物理的な音入力を両耳の鼓膜で受け、脳の中で処理をし、音の姿を知覚する。

普段我々が音を受聴する場合は、音像が頭外に定位している。つまり発音オブジェクトが頭外にある状態である。これを頭外 (音像) 定位と言う。音像が頭外に定位していれば、我々が普段音を受聴する自然な状態であると言える。一方でヘッドホンを装着する事により、この音像が頭内に定位してしまう。これを頭内 (音像) 定位という。頭外定位は受聴者の周囲に広がる音場で音が鳴っているが、頭内定位は頭の内側の狭い空間で音が鳴っているように知覚する。また、頭内定位によって音像の定位が判断しづらくなる。

2.2 スピーカーの音響情報処理

人間がスピーカーから音を聴取する場合、HRTF (Head Related Transfer Function) [4] と呼ばれる伝達関数と外耳道の影響を受け、鼓膜を通過する。HRTF とは、自由空間における鼓膜近傍あるいは外耳道入口の音響伝達関数である。頭部の 3次元形状および音響インピーダンスによって規定され、計測にはダミーヘッドのバイノーラルマイクを使用する。外耳道のインピーダンスを Z_{ec} 、鼓膜のインピーダンスを Z_{ed} として、図 1 にモデル化したものを示す。

音像定位としては、頭外定位となる。実際にラウドスピーカーを水平面円周上に、 30° 間隔に 12 個設置した音像定位実験では、100% ではないものの、音源位置と音像位置は概ね一致する。また、1.5m で計測した HRTF と、ホワイトノイズを畳み込んだバイノーラル信号を不特定多数の受聴者に聴取してもらい、座標を問う実験では、音像定位の傾向が変化したり、頭内定位を引き起こした [9]。但し、この実験では HRTF は受聴者本人のものではない事や、クラウドソーシングによる実験の為、誤定位の要因は分かっていない。

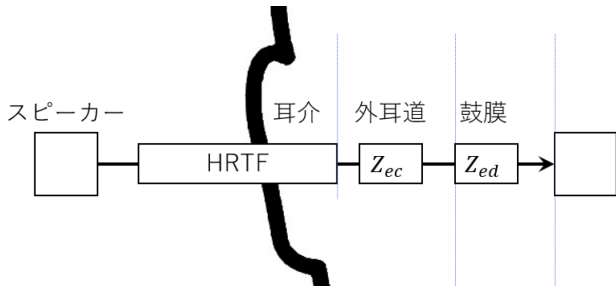


図 1. スピーカーを使用したときの人間の音響情報処理

2.3 HRTF の個人性

日本人成人 10 名の HRTF の個人差は 4kHz までは少なく、それ以上の周波数ではノッチやピークの周波数やレベルも受聴者によって異なる。波長との関係から、耳介形状が HRTF に影響を及ぼすのは約 4kHz 以上であり、HRTF の個人差は耳介形状に起因するものと推測されている [10]。

両耳時間差 (ITD: interaural time difference) においては、日本人 33 名に対し実験が行われた。水平面 12 方向についての ITD を計測した。いずれの受聴者においても ITD は真横 (90°, 180°) で最大となる。また、正中面 (0°) と後方 (180°) において ITD が 0 とならない被験者がいるが、頭部形状の左右非対称性に起因するものと推測される。

両耳間レベル差 (ILD: interaural level difference) は、周波数によって異なる事が報告されている。方位角による ILD の違いは 500Hz では ±10dB 以内であるが、8kHz では被験者によって ±30dB 程度に達し、側方 (90°, 270°) で最大となるが、周波数が高くなると 120°, 240° の方が大きくなる被験者がいる。これは耳介の影響であると考えられる。

2.4 ヘッドホンの音響情報処理

人間がヘッドホンから音を聴取する場合、ヘッドホンから外耳道、鼓膜のインピーダンスまでを伝達関数とした HpTF (Headphone-To-Ear Transfer Function) [8] を考慮する。計測にはマイクアレイ法を使用する。図 2 にヘッドホンを使用した場合の、人間の音響情報処理モデルを示す。ここで Z_s は耳介とヘッドホンのイヤークップ内のインピーダンス、 Z_{ec} は外耳道のインピーダンス、 Z_{ed} は鼓膜のインピーダンスである。HpTF の計測結果からは外耳道の断面積は信号のゲインに影響し、外耳道の長さは HpTF の構造に影響がある事が明らかになっている [2]。しかし、実際に人間の外耳道の長さを正確に測定することは難しくリスクがある。聴感としては頭内定位となりこもったような感覚を受ける。

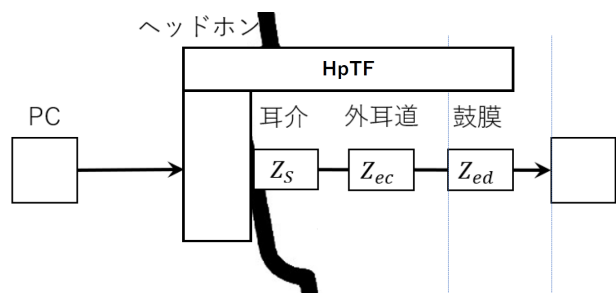


図 2. ヘッドホンを使用したときの人間の音響情報処理

従来研究では、HpTF と HRTF の逆関数で畳み込んだキャリアブレーションで仮想聴空間を作成した実験も行われている [3]。被験者に音の発生源の座標を問い、実際の音定位と比較する。結果として、キャリアブレーションが適切に行われず、不十分な結果であった。これは人体構造の個人差に起因するものであると考えられる。また、HpTF を畳み込むことにより 4-10kHz 付近の歪みを引き起こした。この歪みが誤定位を発生する要因となると考えられる。

2.5 HRTF/自由音場特性補正

HRTF/自由音場特性補正とは、ヘッドホンでもスピーカーと同じように自由音場特性を基準とした評価値を得るために補正である [5]。測定方法は、ダミーヘッドに装着されたヘッドホンからの測定信号の入力に対する周波数特性をダミーヘッド特性としてデータ化する。ダミーヘッド特性の逆特性でフィルタリングし HRTF/自由音場特性補正を行う。これによりスピーカーとヘッドホンの特性を同列に扱える。ダミーヘッド特性の概要と測定方法を図 3 に示す。

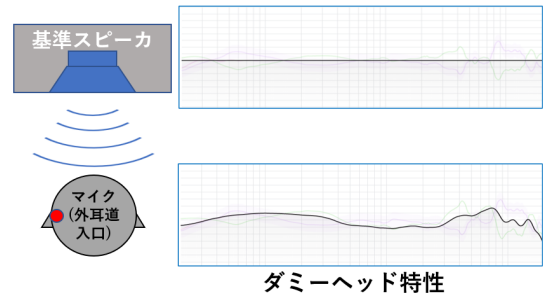


図 3. ダミーヘッド特性の概要と測定方法

3 提案手法

3.1 耳介付近伝達関数

本研究では HRTF/自由音場特性補正を施し、HRTF と外耳道入口までの耳介付近伝達関数の逆特性を畳み込むことで仮想聴覚空間を作成する。スピーカーからの信号 S_{loud} とヘッドホンからの信号 $S_{headphone}$ は、

$$S_{loud} = HRTF \times Z_{ec} \times Z_{ed} \times S \dots (1) \quad (1)$$

$$S_{headphone} = Z_s \times Z_{ec} \times Z_{ed} \times S \dots (2) \quad (2)$$

と表現できる。(1),(2) から、外耳道以降の影響は、聴取者ごとに、全ての聴取方式に共通である。また、HRTF の個人差は耳介に起因するものが多い。また、外耳道以降の計測はリスクやコストが高くユーザー自身が計測する事は考えにくい。また、HpTF を全て補正する事で歪を引き起こし誤定位の要因になる。これらの観点から外耳道入口までの影響を耳介付近伝達関数とし補正する。実際に本研究で耳介付近伝達関数 (TF) として定義する部分を図 4 に示し、数式としての処理を (3) のように表現する。また、耳介付近伝達関数を推定する一連の信号処理を図 5 に示し、実際にシステムとして適用する際の信号処理を図 6 に示す。

$$S_{headphone} = \frac{Z_s}{TF} \times Z_{ec} \times Z_{ed} \times S \dots (3)$$

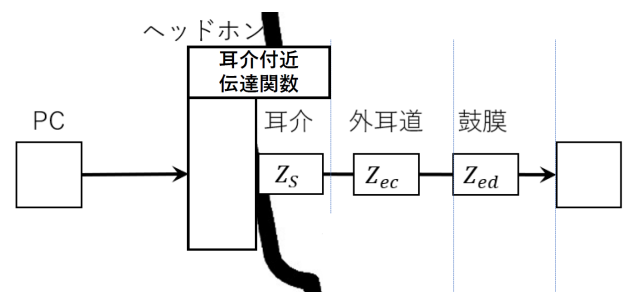


図 4. 新しく定義する伝達関数

耳介付近伝達関数を推定する為に、スピーカーを使用したダミーヘッドのバイノーラルマイクでの録音により外耳道入口までのダミーヘッド特性を測定する。その後ヘッドホン特性に対して HRTF/自由音場特性補正を施し、耳介付近伝達関数の推定を行う。任意の音源に対して耳介付近伝達関数の逆フィルタと任意の HRTF を畳み込む事で仮想聴覚空間を作成する。

実験にはオープンヘッドホン (ATH-AD900X)、クローズドバックヘッドホン (Scarlett HP60 MkIII)、インサートヘッド

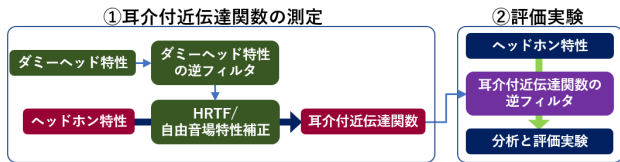


図 5. 本研究が提案する信号処理

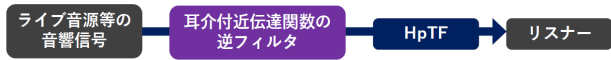


図 6. 本研究が提案するシステム概要

ホン (JABRA ELITE 4 ACTIVE) の 3 種類のヘッドホン (図 9,11 では各々 Closed,Open,Insert と表記する) を使用し, ダミーヘッドは 3Dio FS Pro2 を使用する. また,HRTF/自由音場特性補正を行うスピーカーには FOSTEX P804-S を使用した. 音響機器から線形掃引信号 (TSP 信号) を発生させ, 接続したレコーダーで録音を行う. 信号はサンプリング周波数 44.1kHz である.

3.2 逆フィルタの推定

逆フィルタの推定には RLS 適応フィルターを適用した [7].RLS 適応フィルターとは, 入力信号に関連する重み付き線形最小二乗コスト関数を最小化するフィルター係数を再帰的に求めるアルゴリズムである. コスト関数は次の方程式によって与えられる.

$$C(w_n) = \sum_{i=0}^n \lambda^{n-i} e^2(i)$$

RLS 適応フィルターは, フィルター係数 $w(n)$ を適切に選択し, 新しいデータが到着したときにフィルターを更新することにより, コスト関数 C を最小化する. ここで, $e(i)$ は時刻 i における目的の信号 d と目的の信号の推定値 d_{est} の間の誤差である. 信号 d_{est} は, RLS フィルターの出力であるので, 暗黙的に現在のフィルター係数に依存する. λ は現在より前のサンプルの重みを指数関数的に小さくする忘却係数である. $0 < \lambda \leq 1$ の範囲で設定される. また, $\lambda = 1$ のコスト関数は最小二乗法となり, 無限の記憶を有する. 今回の推定では $\lambda = 1$ として, 入力データにスピーカー, ヘッドホンのインパルス応答を設定し目的関数を TSP 信号と逆 TSP 信号のインパルス応答に設定することで,RLS 適応フィルターを設計した.RLS 適応フィルタの係数を取得し, 逆フィルタを設計した. ここで, フィルタの係数は 32 次である.

3.3 測定方法

TSP 信号を入力として, ダミーヘッドのバイノーラルマイクを使用し外耳道入口までの TSP 応答をレコーダーで測定する. 得られた TSP 応答に対して逆 TSP 信号を畳み込み, インパルス応答を得る. インパルス応答からヘッドホン特性とダミーヘッド特性を推定する. その後,HRTF/自由音場特性補正を施し, 耳介付近伝達関数を推定する. ヘッドホンの TSP 応答測定の測定環境についてのブロック図を図 7 に, スピーカーの TSP 応答測定の際のスピーカーの録音環境についてのブロック図を図 8 に示す.

3.4 測定結果

3.4.1 ヘッドホン特性

ヘッドホン特性を推定した結果を図 9 に示す.

クローズドバックヘッドホンとオープンヘッドホンは 16~17kHz 地点に大きなノッチが見られた. 特にオープンヘッドホンは 7.5kHz 地点に大きなノッチが観測された. また, オープンヘッドホンとクローズドバックヘッドホンは, インサートヘッドホンと比較して低周波数帯の音圧が大きく立ち上がり早い事が分かる. インサートヘッドホンは他の 2 形状のヘッドホンとは異なり, 大きなピークやノッチが見られず高周波帯域になるにつれて, 次第に減衰していく. 上記の結果を踏まえて, ヘッドホンの構造ごとにも大きな差がある事が分かり, それぞれの形状に対して個別のフィルターを設計するのは必要不可

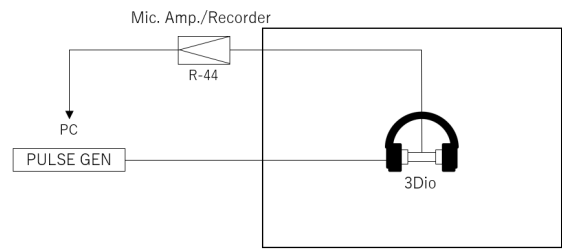


図 7. ヘッドホンの TSP 応答測定のブロック図

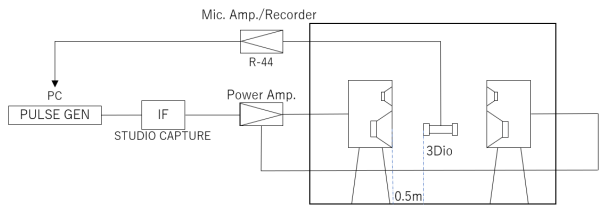


図 8. スピーカーの TSP 応答測定のブロック図

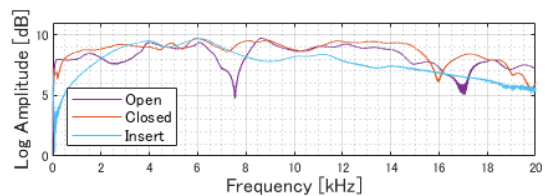


図 9. ヘッドホン特性

欠である. また, 構造に起因する特徴はメーカーが違っていても, 傾向が概ね一致する. その為本研究ではそれぞれの構造を代表した 3 種類の構造について分析, 評価を行った.

3.4.2 耳介付近伝達関数と HRTF/自由音場特性補正

HRTF/自由音場特性補正に必要なダミーヘッド特性を得る為に, スピーカーでの測定と分析を行った. ダミーヘッド特性を図 10 に示す.

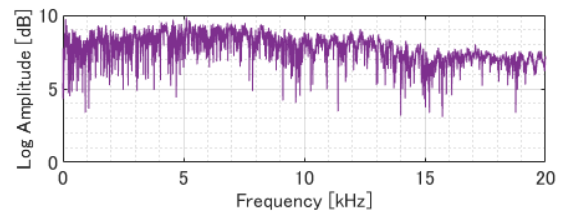


図 10. ダミーヘッド特性

図 10 から耳介付近の反射や室内の反射を含む事から, 全帯域にノイズが多く距離減衰により高周波成分になだらかな減衰が見られた. この系のインパルス応答を入力に, 目的関数を TSP 信号のインパルス応答として RLS 適応フィルターを推定し, 得られた係数で設計した HRTF/自由音場特性補正フィルターを図 11 の Calibration として示す. また, 補正し得られた耳介付近伝達関数を, ヘッドホンの構造ごとを示す.

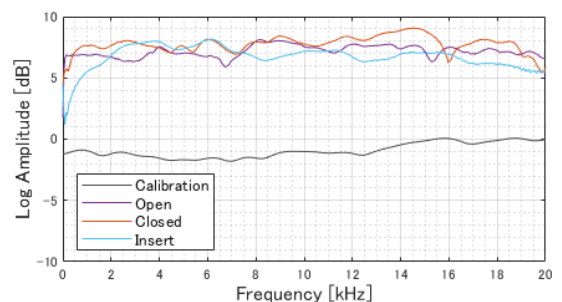


図 11. HRTF/自由音場特性補正と耳介付近伝達関数

得られた耳介付近伝達関数の逆フィルタを任意の音源に畳み込み評価実験を行うが, 今回は任意の音源に対して, 直接ダミー

ヘッド特性の逆フィルタとヘッドホン特性の逆フィルタを順次畳み込むことで、耳介付近伝達関数の補正を実装している。

4 評価実験

4.1 評価方法

評価方法は立体音響と同等に評価をする。これは疑似的にHRTF/自由音場特性補正を施し、仮想聴覚空間を作成している為である。受聴者に音源の座標を問い、実際の音像定位との正解率を確かめる。この時、HRIRのみを畳み込んだ音源と、耳介付近伝達関数補正とHRIRを畳み込んだ音源を聴取してもらう。頭内定位が改善されれば、音像が頭外に定位し、受聴者が方向を認識できる。

4.2 実験方法

受聴者を中心とした水平方向(仰角 0°)の 45° から 90° 置きに4箇所のHRIRを畳み込んだ音源に対して、耳介付近伝達関数の逆フィルタを畳み込んだ音源を、ヘッドホンから再生し受聴してもらう。実験の開始時には、比較対象としてHRIRのみを畳み込んだ音源と何も補正をしていない音源も聴いてもらう。HRIRは東北大学電気通信研究所(RIEC)のデータベース[6]のものを使用する。受聴者には音源の座標を問い水平面のどこに定位しているかをプロットしてもらう。使用する音源は、ホワイトノイズと楽曲を選択した。

【被験者】20代の健聴者2名

5 結果及び考察

表1に実際に畳み込んだHRTFの音像定位との正解率を示す。音源がホワイトノイズである時を1、楽曲である時を2として示す。

表1. 音像定位の正解率

補正	タイプ	A1	A2	B1	B2	平均
なし	Open	0.5	0.25	0.5	0	0.31
	Closed	0.25	0.25	0	0	0.13
	Insert	0.5	0	0	0.25	0.19
	平均	0.42	0.17	0.17	0.08	0.21
あり	Open	0.5	0.5	0	0.25	0.32
	Closed	0.5	0.5	0	0.5	0.38
	Insert	0.25	0.75	0.75	0.25	0.5
	平均	0.42	0.58	0.25	0.33	0.4

耳介付近伝達関数補正によって、音像定位の正解率は90%向上し、補正の有意性が認められた。しかし、前方に定位する 45° と 315° の音源についての正解率が悪かった。また、左側に位置する 45° と 135° の定位の正解率が悪かった。これは、HRTF個人性に起因するものと考えられる。一方で距離感では70.8%が補正なしの音源より遠くに聴こえ、有意性が見られた。

音色によって、定位感や聴感が大きく変化する事が確認でき、その差を補正により補間することが可能である。補正なしの場合に着目すると、ホワイトノイズは高周波成分を多く含み音像定位に関連する周波数成分が大きい為、定位がはっきりとしやすい傾向にある。一方で楽曲はホワイトノイズと比較し、高周波成分があまり含まれていない為、定位を知覚しにくい。その為補正ありの場合では、補正により楽曲の高周波数成分が増幅し、定位の正解率が向上した。

ヘッドホンの構造に起因する正解率の差は耳介付近伝達関数の補正を施す事により無くなった。補正なしの場合では、オープンヘッドホンの正解率が最も高く、頭外定位しやすいと言える。補正ありの場合ではインサートヘッドホンはHRTF/自由音場補正により約4kHz以上の高周波数成分が増加する事で、音像制御を行う事が可能になった。

実験の際に受聴者に音源についての聴感の感想をもらった。耳介付近伝達関数補正ありの音源はHRIRのみを畳み込んだ音源と比較し音質が良いという感想が得られた。ここでの音質は音色を意味していると考えられる。HRIRのみを畳み込んだ音

源は音像定位に関連する高周波数帯域が増幅する。受聴者に適切なHRTFが選択されていない場合、HRIRのみの音源は過剰に高周波数成分が再生され不快感を与える。これが耳介付近伝達関数補正によって過剰な高周波数成分が緩和され、音像制御を保ちつつ、受聴者にとって聞き取りやすい音源となったと推測される。本研究はライブ音源を再生する事を前提としている為、受聴者の聴感重要な観点であり、有用性の高い補正だと言える。

6 まとめ・今後の展望

本研究では、オンラインライブ等の臨場感再現の為に、外耳道入口までの耳介付近伝達関数を補正する事で音像が正しく定位するかの評価実験を行った。音像定位の正解率は90%向上し、優位性が認められた。また、耳介付近伝達関数の補正によって、ヘッドホンの構造に起因する聴感の差や音色による差を補正する事が可能である。

本研究では特定のヘッドホンを使用し実験を行っているが、ヘッドホンの測定が可能なバイノーラルマイクさえあれば、手軽に仮想聴覚空間を作成する事が可能である。また、HRTF/自由音場特性補正に関しても、無響室での測定とフラットな特性を持つスピーカーを使用できれば、より音場再現性の高い補正が可能になる。

HRTFについてはダミーヘッドのHRIRを使用し実験を行っているが、あらかじめユーザーに適したHRIRを選択することによって音像がより正しく定位する可能性がある。また、現在の研究ではユーザーが適したHRTFを所持するには、手当たり次第に確かめる事が唯一の方法である。標準HRTFがどのような基準で適用できるのかが明らかになれば、ヘッドホンのバイノーラル再生についての同等な評価や起因する影響を確かめる事が容易になる。また身体構造の差について補正ができていないかを確認する為には、標準HRTFを用いた方法あるいは、ロバストな頭部伝達関数を選択できているという条件を満たしている必要がある為、更なる研究が待たれる。

ヘッドホンやスピーカー等の機器それぞれの特性や測定方法について、実際に補正ソフトとして製品化されている為データを入手するのが難しい場合が多かった。これがデータベースとして提供されるようになれば、より測定や聴感に関する実験が行いやすくなる。

参考文献

- [1] 岩谷幸雄, “頭部伝達関数による音像定位,” 音響学会誌, 73, 173-180 (2017).
- [2] J. Liu, et. al., “Headphone-To- 761 Ear Transfer Function Estimation Using Measured 762 Acoustic Parameters”, Applied Sciences, vol. 8, 763 no. 6, 2018
- [3] Pralong, et. al., “The role of individualized headphone calibration for the generation of high fidelity virtual auditory space.” Journal of the Acoustical Society of America, 100(6): December 1996.
- [4] 平原達也, 他, “頭部伝達関数の計測とバイノーラル再生にかかわる諸問題,” 信学FR誌, vol.2, no.4, pp.68-85, April 2008.
- [5] 電子情報技術産業協会 (JEITA) 規格 RC-8140C 「ヘッドホン及びイヤホン」 (2020)
- [6] <http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/index.html>
- [7] Haykin Simon, 適応フィルタ理論, 科学技術出版, pp.630誌-657, 2001.
- [8] Møller, H., “Fundamentals of Binaural Technology,” Applied Acoustics, Vol.36, 171-218 (1992).
- [9] 森川大輔, “はじめての音像定位実験,” 日本音響学会誌, 74, 547-554 (2018)
- [10] 飯田一博, “頭部伝達関数の個人性,” 日本音響学会誌, 73, 181-188 (2017)