

コンサートスコープヘッドフォン

浜中 雅俊^{*1} 李 昇姫^{*2}

Concert Scope Headphones

Masatoshi Hamanaka^{*1} and Seunghee Lee^{*2}

Abstract – We designed concert scope headphones that are equipped with a projector, an inclination sensor on the top of the headphones, and a distance sensor on the outside right headphone. We previously developed sound scope headphones that enable users to change the sound mixing depending on their head direction. However, the system could not handle images. In contrast, our headphones let the user listening to and watching a music scene scope on a particular part that he or she wants to hear and see. For example, when listening to jazz, one might want to more clearly hear and see the guitar or sax being played. The user can hear the guitar or sax sound from a frontal position by moving their head to the left or right. The user can adjust the distance sensor on the headphones and focus on a particular part they want to hear and see by simply putting their hand behind their ear.

Keywords : active music listening interfaces, headphones, projector

1 はじめに

指揮者などの音楽家は、コンサートで複数の演奏者が同じ楽器を同時に演奏していたとしても、各演奏者の音を聴き分けることができる。一方、音楽初心者はそれを聴き分けることは困難である。そこで本研究では、音楽初心者でもユーザが見たり聴いたりしたい特定の演奏者の演奏を強調して視聴することのできる音楽鑑賞インタフェース、コンサートスコープヘッドフォンを提案する。

我々は、そのようなインタフェースを設計する上で、次の二点が重要であると考える。第一の点は、インタフェースの操作に、人間が音を聴くときにする自然な動作を利用することである。そのようにすることで、初心者でも直感的にインタフェースを操作することが可能となる。コンサートスコープヘッドフォンでは、コンサートホールで自分が見たい演奏者のいる方向に顔を向ける動作や、音を耳を澄ませて聴くときに手を耳にかざす動作を用いて、ステージ全体を収めた画面の広い映像から特定の範囲や演奏者を選択して鑑賞することを可能とする。

第二の点は、一般家庭で手軽に使用できることである。我々は、近年プロジェクタが小型化されていることに着目し、小型プロジェクタと3軸角度センサ、距離センサをヘッドフォンに搭載したデバイスを構築す

る(図1)。このデバイス1つで、頭部方向の検出および耳に手をかざすポーズをしたときの手と耳の距離を検出し、検出した値に応じた映像と音を出力することが可能であるため、一般家庭をはじめ多くの環境で使用することができる。

図2は、コンサートスコープヘッドフォンの全体像である。まず、3軸角度センサによって検出した頭部方向に応じて、ステージ全体を収めた映像から切り出す映像の範囲を決定し切り出しを行う。そして、切り出した映像がプロジェクタから投影される。同時に、切り出された映像の範囲内に映っている演奏者の演奏音が強調されるように、複数の演奏音源をミキシングして音出力される。このとき、ユーザが耳を澄ますポーズをとると、ヘッドフォンの右スピーカの外側に搭載する距離センサと手との距離に応じて投影映像がズーム(拡大)され、よりピンポイントに特定の演奏者の演奏音を鑑賞できる。

構築するインタフェースは以下の3つの特長をもつ。

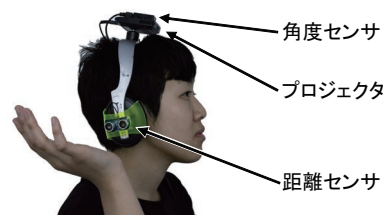


図1 コンサートスコープヘッドフォン
Fig.1 Concert scope headphones

^{*1}筑波大学 システム情報系

^{*2}筑波大学 芸術系

^{*1}Faculty of Engineering, Information and Systems, University of Tsukuba

^{*2}Faculty of Art and Design, University of Tsukuba

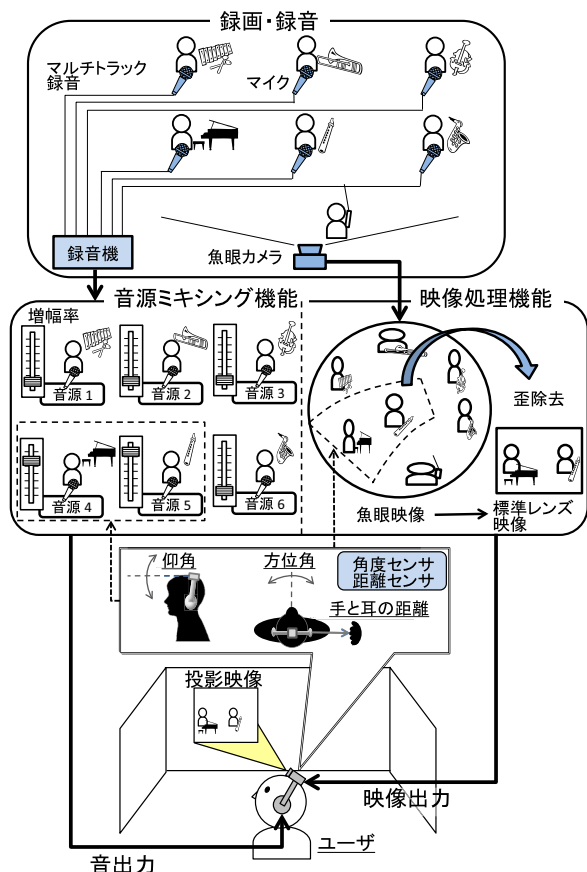


図2 コンサートスコープヘッドフォンの全体像
Fig.2 Overview of concert scope headphones

● 仮想マイクの利用

ユーザが任意に選択した映像の範囲内にいる演奏者の演奏音を強調するようにミキシングするためには、すべての演奏者の近くに1つずつマイクを設置して、それぞれ独立したトラックで録音するのが理想的である。しかし、オーケストラのように演奏者数が多い場合にはコストや労力の点からそれを実現することは困難である。このとき、マイクを設置していない演奏者が映像上に現れたときにどのようにミキシングするのが問題となる。本研究では、マイクを設置していない演奏者に対し仮想的なマイクを生成することで、全演奏者へのマイクを設置したのと同じ方法でのミキシングを可能とする。具体的には、複数のマイク有りの演奏者の音源を、目的とするマイクの無い演奏者との距離に応じてミキシングし、これを仮想マイクの音源とする。

● 円周魚眼レンズの利用

円周魚眼レンズを用いて、ステージ全体を収めた画角の広い映像を獲得する。この映像からユーザの頭部方向や手と耳の距離に応じて映像の一

部を切り出して投影する。円周魚眼レンズは180度の画角の映像を円形に投影するカメラレンズで、映像の中心付近は歪みが少ないが、周辺部は歪みが大きいという特徴がある。そこで、切り出しと同時に歪み補正を施し標準映像に変換してから投影する。

● 3面の壁への投影

投影はユーザを囲む3面の壁(スクリーン)に対しておこなう。平面のスクリーンに180度わたるステージ映像をすべて表示すると、ステージの端にいる演奏者が見えづらくなる。そこで、正面だけではなく左右の壁を利用した3面に映像を投影する。これにより、ステージの端にいる演奏者は左右の壁に投影されるため、見やすくなる。このとき、ユーザの頭部が壁に対して斜めになるとプロジェクタからの映像が壁に斜めに投影されるため、投影映像が台形に歪む。そこで、プロジェクタが壁に斜めに投影したときに生じる歪みと逆方向の歪みをあらかじめ映像に加えて出力することで、投影映像の台形歪みを防ぐ。

以下、2章では関連研究について議論を行い、3章、4章では、映像処理および音源のミキシングについて説明する。5章で実装について述べ、6章で評価実験について述べる。そして、7章でまとめを述べる。

2 関連研究

ヘッドフォンとプロジェクタが一体型となったデバイスによって、ステージ上の特定のエリアや特定の演奏者を選択的に鑑賞可能とする音楽鑑賞インタフェースはこれまでなかった。本研究と関連のある研究を以下に挙げ、その特徴を述べることで本研究との差異について議論する。

複数の演奏者の演奏音を聴き分けるためには、演奏者ごと複数トラックに録音した音源を用意して、音楽用ミキサーで各音源のボリュームや定位を調節する方法が考えられる。しかし、市販の音楽用ミキサーは、操作が煩雑で初心者が直感的に操作することは困難であった。たとえば、バイオリンソロを始めた瞬間に、そのバイオリン演奏者のボリュームを上げ、定位を中央に寄せて、その他の演奏者のボリュームを少しずつ下げたその定位を調節するなど、一連の動作を瞬時に行うことは初心者には困難であった。

空間音響システム [1, 2] では、聴取者のアバタや各演奏者の位置をパソコンのマウスで移動することでミキシングを変更することを可能にしていた。しかし、各演奏者のミキシングを適切に変更するためには、ソ

ロが始まった演奏者を近くに配置したり，ソロが終わり伴奏に戻った演奏者を遠くに配置しなおすなどの煩雑な操作が必要であった。

頭部の方向や位置を検出するセンサを搭載したヘッドフォンは従来から存在していたが，その目的は，仮想の音源位置を固定して臨場感を高めることであったため，本研究のように各演奏者のボリュームや定位をコントロールする目的では使われてこなかった [3, 4, 5, 6].

そこで我々は，頭部方向に応じてミキシングを変更することで特定の演奏者の演奏を強調することを可能とするヘッドフォン型音楽鑑賞インタフェースを構築してきた [7]. しかし，映像の機能は実装されていなかった。また，文献 [7] では，音源として RWC 研究用音楽データベース [8] に収められているスタジオでレコーディングされた音源を用いていたが，本研究では，実際のコンサート録音を用いる。

一方，TWISTER は，発光ダイオードと遮光板をユーザの周囲を高速に回転させることで周囲 360 度の立体視を可能としていた [9]. Ensphered Vision は，ユーザを覆う球型のスクリーンに映像を投影することで，周囲 270 度の映像を鑑賞することを可能としていた [10]. これらのシステム [9, 10] は，高い臨場感で周囲の映像を鑑賞することを可能としていた点で優れていたが，大きな装置が必要なために一般家庭への普及が困難という問題がある。

多視点映像や画角の広い映像を用いたコンテンツでは，ユーザが任意に映像を切り替えて選択的に鑑賞することを可能としていた [11, 12]. 多視点映像コンテンツ [11] は，ステージの複数箇所を設置した画角の狭い通常のカメラの映像を収録し，各映像をユーザが任意に切り替えて鑑賞するものである。画角の広い映像を用いたコンテンツ [12] は，全方位カメラや魚眼レンズを用いて 360 度のパノラマ映像を生成し，ユーザが任意に見たい部分を選択できるものである。これらの技術 [11, 12] は，広い範囲の映像を記録する点で本研究の問題意識に近いが，鑑賞方法は，映像をマウスやボタンなどのコントローラで操作するもので初心者が見たい部分を選択できるものであったために，頭部を左右に動かしたときの映像の変化と音の変化の関係に違和感を感じるユーザが多かった。そこで本論文では，各演奏者の音源のミキシングとプロジェクタにより投影された映像上の演奏者の位置との関係が自然となるよう改良を行い，その効果を評価実験により確認する。

これに対し我々は，プロジェクタとセンサを搭載し，ユーザの頭部の方向に応じて直感的に映像を切り替えることを可能とするヘッドフォンの構築を試みてきた [13]. しかし，文献 [13] ではユーザに提示する音がモノラルであったために，頭部を左右に動かしたときの映像の変化と音の変化の関係に違和感を感じるユーザが多かった。そこで本論文では，各演奏者の音源のミキシングとプロジェクタにより投影された映像上の演奏者の位置との関係が自然となるよう改良を行い，その効果を評価実験により確認する。

3 映像処理機能

コンサートスコopヘッドフォンは，大きく分けて映像処理機能と音源ミキシング機能の 2 つからなる。映像処理機能は，ユーザの頭部方向の変化に応じて映像を切り出す視野画像切り出し部，スクリーンに斜めに投影した場合に生ずる台形歪みをあらかじめ補正する台形歪み補正部，プロジェクタによってスクリーンに投影する出力部からなる。

3.1 視野画像切り出し部

本研究では，撮影のセッティングが容易に行えるよう，ステージ全体を 180 度の画角を持つ円周魚眼レンズカメラ 1 台で撮影する。円周魚眼レンズカメラは，図 3 のように，仮想半球上に映り込んだ景色を撮影面に垂直に投影することで図 4 のような映像が得られる。

視野画像切り出し部では，3 軸角度センサからユーザの頭部の仰角 ϕ と方位角 θ を 3 軸角度センサにより検出する。そして，その仰角 ϕ と方位角 θ の点を中心として，魚眼レンズの映像の歪みを補正し，標準レンズ映像に変換した画像を切り取る。また，距離センサに手をかざし，耳をすます動作をとることによって，切り取られた映像がズーム (拡大) され，さらに詳細な範囲の映像が表示される。ズームの倍率は，手と距離センサとの距離に比例する。

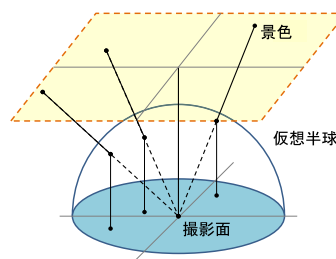


図 3 魚眼カメラの原理

Fig. 3 Principle for capturing with a circular fisheye-lens camera

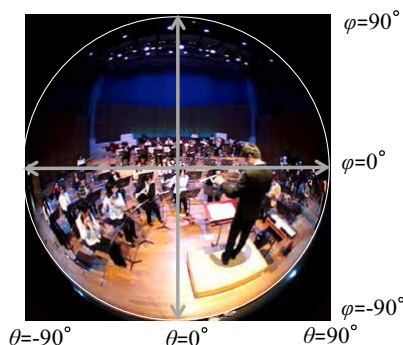


図 4 魚眼カメラの映像

Fig. 4 Image by a circular fisheye-lens camera

魚眼レンズによる映像は、円の外側に近くなるほど強い歪みとなる。この魚眼歪みを補正し、標準レンズで撮影した映像に変換するためには、投影面に映った魚眼映像を、魚眼レンズの投影過程の逆過程を行い、再度平面へ投影すればよい。具体的には、図5上部にある魚眼映像を、図下部にある半球上に投影する。次に、半球上に投影された映像を球の中心から伸びる仰角 ϕ 、方位角 θ の直線を法線に持つ平面(図5左)にさらに投影する。これらの処理は、OpenGL[14]を用いて実現した。図6は、このような処理によって歪みを補正する前後の映像の例である。

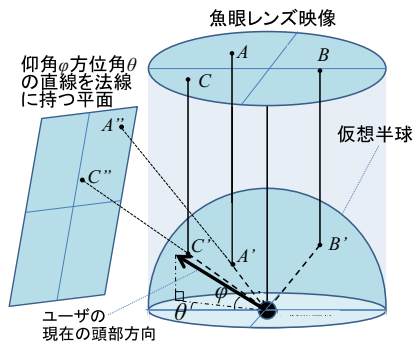


図5 魚眼歪みの補正のための投影処理
Fig. 5 Texture mapping for distortion correction

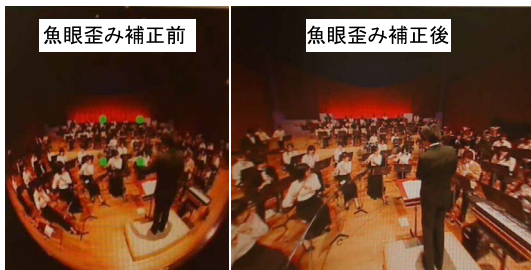


図6 補正前後の画像
Fig. 6 Distorted and corrected images.

3.2 台形歪み補正部

ヘッドフォンに装着したプロジェクタからスクリーンに投影すると、多くの場合斜めに投影される。すると、正面で投影したときに比べ画像の拡大・縮小が生じて台形となる台形歪みが生じる(図7)。台形歪みは、映像上の一部が引き伸ばされたり縮小するためユーザーは違和感を感じることになる。そこで、プロジェクタを傾けて投影した場合も映像が台形に変形しないように、その時の仰角・方位角に応じてあらかじめ台形歪みと逆方向に歪みを加えた後に投影する(図8)。この際、3軸角度センサからの情報から投影距離を算出し、正面での投影距離と比較することによって映像の拡大

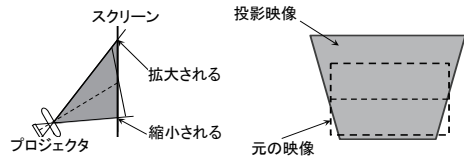


図7 映像を斜めに投影した場合
Fig. 7 Image projected at a tilt

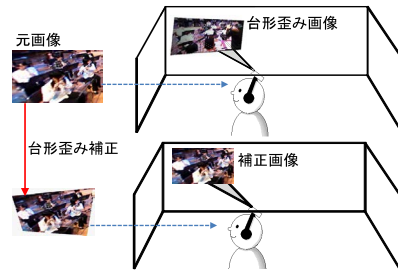


図8 台形歪み補正
Fig. 8 Keystone distortion correction

率・縮小率を決定する必要がある。以下、プロジェクタとスクリーンとの距離を A 、投影した映像の高さを H 、幅を W 、画角を α 、投影した映像の右端もしくは左端からプロジェクタまでの距離を X として説明を行う(図9)。プロジェクタが角度 β 傾いている場合、投影された映像の中心 C と左端 L との作る角度は、 $\alpha/2$ となるため、投影した画像の左端からプロジェクタまでの距離 X_1 は次式で求まる(図10)。

$$X_1 = \frac{A}{\cos\left(\beta - \frac{\alpha}{2}\right)} \quad (1)$$

したがって、正面に投影したのに比べて、左端の映像は X_1/X 倍されて投影される。同様に、投影した画像の右端からプロジェクタまでの距離 X_2 は次式で求まり、正面に投影したのに比べて、右端の映像は X_2/X 倍されて投影される。

$$X_2 = \frac{A}{\cos\left(\beta + \frac{\alpha}{2}\right)} \quad (2)$$

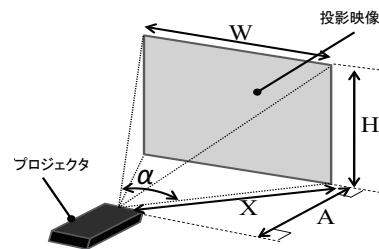


図9 投影映像の構成
Fig. 9 Configuration of projected image

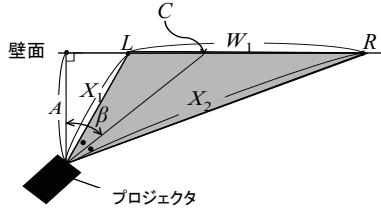


図 10 傾斜して投影した映像の構成
Fig. 10 Configuration of image projected at a tilt

そこで、あらかじめ映像の左端を X/X_1 倍、右端を X/X_2 倍した映像を投影することで、台形歪みを補正する。

3.3 出力部

映像をプロジェクタに出力する。ユーザの動作の結果、頭部が左右に傾くと、投影された映像に回転が生じる。そこで、出力前の画像を頭部の傾きと逆方向に回転することで、回転補正を行う (図 11)。

平面のスクリーンに 180 度わたる画像をすべて表示すると、視覚的に不自然となるため、正面だけではなく左右の壁を利用して映像を投影する。このとき、ユーザによる投影がスクリーンの角を超えて別のスクリーンに移動すると、そのスクリーンの方向に応じた台形歪み補正に切り替えて投影する。

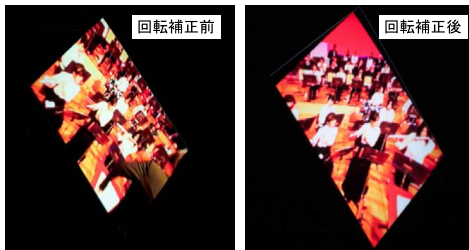


図 11 回転補正映像
Fig. 11 Rotated and corrected images

4 音源ミキシング機能

音源ミキシング機能は、仮想マイク生成部、増幅率コントロール部、左右音量比コントロール部、ミキシング部からなる。

4.1 仮想マイク生成部

各演奏パートごとに少なくとも 1 人の演奏者の譜面台にマイク (以降「実マイク」と呼ぶ) を 1 個取り付けて、実マイク音源 R_n (パートの数 $\leq n \leq$ 演奏者の数) を録音する。このとき、実マイクが取り付けられていない演奏者については、その周囲の実マイクで録音した音源を合成して仮想マイク音源 I_m ($0 \leq m < n$) を生成する。

全方向の音を集音する一般的なマイクを用いれば、仮想マイクとはなれた位置にある実マイクでの録音にも、音圧は距離の二乗に反比例して低くなるが目的とする演奏者の音が含まれている。また、多くのコンサートでは、ステージの広さは十分広く、同じ演奏パートが近くに配置され、別のパートの間には十分な距離があるため、仮想マイクの周囲にある実マイクには目的とするパートの音が大きな音圧で入ることになる。そこで本研究では、仮想マイクの周囲にある複数の実マイクの音源をその距離に応じて合成することで、仮想マイクの音源を生成する。

$$I_m = \frac{1}{n} \sum_n \frac{R_n}{\text{dist}(m, n)^2} \quad (3)$$

$\text{dist}(m, n)$ は、コンサートホールでの演奏者 m と n の距離である。

4.2 増幅率コントロール部

各演奏者の音源もしくは仮想音源の増幅率は、映像処理機能で円周魚眼カメラの映像から切り出した投影画像上での中心からの距離 d_n に基づき計算される。具体的には、 d_n は、円周魚眼映像上での演奏者の座標をあらかじめ手作業で調べておくことで算出が可能となる (図 12)。

$$d_n = \frac{\text{音源 } n \text{ の投影画像中心からの距離}}{\text{投影画像の対角線の半分の長さ}} \quad (4)$$

そして、投影画像の中心に近い演奏者の演奏音ほど強調されるように増幅率 h_n をコントロールする。距離に応じて増幅率 h_n を変化させるため、図 13 にあるような 5 種類の関数を考え、6 章の評価実験でどの関数が適切か検討する。図 13 の a は定数、b は傾きが負の一次関数、c は正規関数、d は反比例、e は距離の 2 乗に反比例する関数である。各関数は映像の中心からの距離が 0 のときミキシング比率が 1 に、0.5 のとき 0.1 になるように正規化してある。

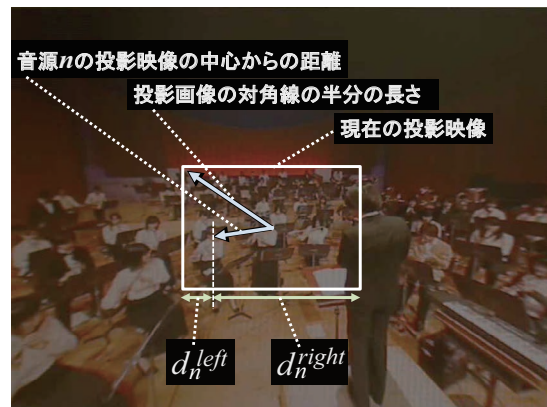


図 12 距離の計算
Fig. 12 Distance calculation

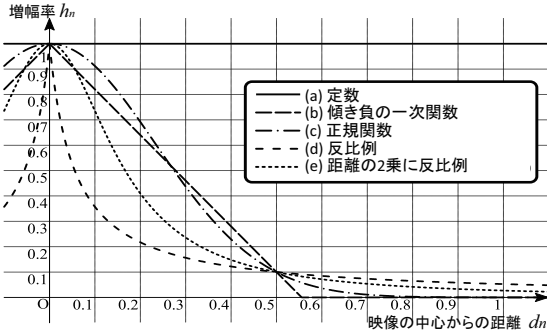


図 13 増幅率
Fig. 13 Mixing rate functions

4.3 左右音量比コントロール部

各音源の左右の音量比を調節して、映像上の各演奏者の位置と演奏音との関係が自然となるようにする。すなわち、映像の右側に投影されている演奏者の演奏音が右側のスピーカから大きな音で聞こえるようにし、左側に投影されている演奏者の演奏音が左側のスピーカから大きな音で聞こえるようにする。また、映像の中央付近に投影されている演奏者の演奏音は、左右のスピーカから同じ程度の音量で聞こえるようにする。左右それぞれの音の増幅率を次式とする。

$$h_n^{left} = \frac{d_n^{right}}{d_n^{right} + d_n^{left}} \quad (5)$$

$$h_n^{right} = \frac{d_n^{left}}{d_n^{right} + d_n^{left}} \quad (6)$$

ここで、 d_n^{left} は演奏者から投影映像の左端までの距離、 d_n^{right} は演奏者から投影映像の右端までの距離である (図 12)。

4.4 ミキシング部

すべての実マイク音源 R_n 、仮想マイク音源 I_m の信号を加算し、ヘッドフォンから出力する。その際、増幅率コントロール部、左右音量比コントロール部で求めた増幅率を掛け合わせる。 S_{left} は左スピーカの出力、 S_{right} は右スピーカの出力である。

左スピーカの出力:

$$S_{left} = \sum_n R_n \cdot h_n \cdot h_n^{left} + \sum_m I_m \cdot h_m \cdot h_m^{left} \quad (7)$$

右スピーカの出力:

$$S_{right} = \sum_n R_n \cdot h_n \cdot h_n^{right} + \sum_m I_m \cdot h_m \cdot h_m^{right} \quad (8)$$

n は実マイクの数、 m は仮想マイクの数である。

コンサートスコープヘッドフォンは、ユーザの頭部方向に応じて映像および実マイク・仮想マイクの音源のミキシングが変更される。このとき、ユーザが手を

耳にかざし、投影している映像をズームさせると、画面の端にいる演奏者は画面の外にズームアウトすることになる。したがって、中央付近にいる演奏者の演奏音は、さらに強調されることになる。

5 実装

ヘッドフォンの頭頂部に搭載する 3 軸角度センサは、小型で高精度の MicroStrain 社の 3DM-GX3-25 を用いた。3DM-GX3-25 は USB でパソコンと接続できる。

一方、ヘッドフォンの右スピーカの外側に搭載する距離センサは、Parallax 社の超音波距離センサ (Ping))) を用いた。この超音波センサは、3 メートルまでの距離を測ることができるが、手が直接センサに触れてしまうと超音波の発信や受信を妨げてしまうため、距離の測定ができなくなってしまう。そこで、本研究ではセンサに手が直接触れることがないようにアクリル板でプロテクターを作成した。距離センサの出力は、アナログ電圧であるため、SparkFun 社の Arduino Pro Micro を用いてシリアル信号に変換した。Arduino Pro Micro は、USB 端子を搭載しているため USB ケーブルでパソコンに接続することができる。

ヘッドフォンに搭載するプロジェクタとしては、RoyalTek 社の USB プロジェクタ PJU-2100 を用いた。PJU-2100 は、小型軽量で USB1 本で電源供給と映像信号の接続が可能である。またファンレスのため、放熱用モータの磁力により 3 軸角度センサ 3DM-GX3-25 が誤動作するのを防ぐことができるという特徴がある。

コンサートスコープヘッドフォンは、一般家庭で正面と左右に壁のあるスペースで使用することを想定しているが、展示スペースなどで、そのようなスペースが無い場合にも使えるようなスクリーンを作成した (図 14)。このスクリーンは、半透過スクリーンを凹型に曲げたアクリル板に貼ったもので、周囲にいる人もスクリーンの裏側に透過した映像を見てユーザがどの演奏者を見ているのか確認することが可能である。

3.2 節で述べた台形歪み補正は、プロジェクタから投影された映像が 3 面あるスクリーンのある面から別の面に移動すると、補正の方向を逆転するようになっている。そのため、使用するスペースやスクリーンを変更した場合には、使用前に台形歪み補正に必要なパラメータを再設定する必要がある。具体的には、まず、ヘッドフォンの右スピーカ外側に搭載した距離センサを正面のスクリーンに向け、ヘッドフォンと正面のスクリーンの距離を測定する。次に、ヘッドフォンを装着したユーザがスクリーンの角、すなわち、3 面あるスクリーンのうち 2 面の接合部を見て、その時の角度を測定する。このような設定を行うことで、適切に台形歪み補正が行われるようになる。



図 14 半透過凹型スクリーン
Fig. 14 Semi-permeable concavity screen

6 評価実験

評価実験では、4.2 節の増幅率コントロール部で述べた 5 種の関数のうち、いずれを用いると映像上の演奏者の位置とミキシングの変化の対応関係が自然となるのか検討する。また、実際にコンサートを収録した映像および音源を用いて、コンサートスコープヘッドフォンを試用した結果についても述べる。

6.1 コンサートの収録

つくば市のコンサートホール（Nova ホール）での筑波大学吹奏楽団の演奏会を収録した。図 15 は、Nova ホールのステージ周辺に設置したマイクとカメラの位置を示している。各パートに少なくとも 1 つのマイクが行き渡るように、パナソニック社のコンデンサマイク WM-61A を用いて作成したラベリアマイク 37 個を譜面台に取り付けた。録音は、Zoom 社の 4 トラックレコーダ H4n を 10 台使用して、37 個のマイクからの音を独立に録音した。H4n はバッテリー駆動で携帯性が高いという特徴がある。

キャノン社のデジタル一眼レフカメラ EOS Kiss X3 の標準レンズに Raynox 社の円周魚眼コンバージョンレンズ DCR-CF187PRO を装着して動画モードで撮影を行った。ステージの全演奏者が映るようにステージ面から 2.5 メートル上方から撮影を行った。映像および音源は、複数の曲を延べ 3 時間分収録した。

6.2 音源ミキシング機能の評価

映像上の演奏者の位置とミキシングの変化の対応関係が自然となるためには、ユーザが頭部を動かすことで、映像上で正面に近づいている演奏者が、聴覚的にも正面に近づいていると感じられることが重要である。

たとえば、コンサートスコープヘッドフォンを使用しているときに、演奏音は聴こえていてユーザはその演奏音を出している演奏者の姿を見たいが、映像上ではどの演奏者であるかわからないという状況が考えられる。このとき、ユーザは頭部を動かしたり、手を耳にかざす動作をしながら、増幅率や左右の音量比の

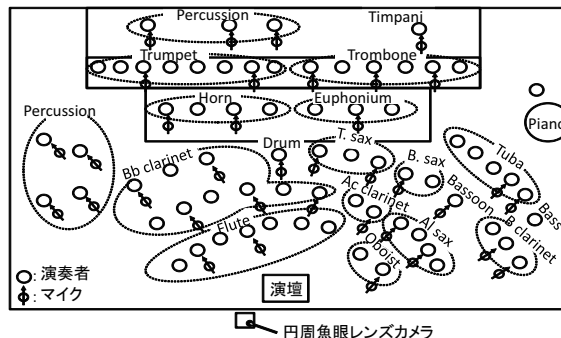


図 15 マイクとカメラの位置
Fig. 15 Microphone and camera positions

変化を手掛かりに、目的とする演奏者を探すことになる。本節では、それがなるべく早く探せることが望ましいと考え、それを実現する関数がいずれであるか検討する。

実験手順

映像上の 1 点に 440Hz の正弦波を置き、ユーザが聴覚のみでその点が正面に来るように頭部を動かし探す試行を行い、より早くその点を探しあてることのできる関数を求める実験を行った。

実験は以下の手順で行う。まず、正弦波を置く 1 点の位置をランダムで決定する。そして、ユーザが頭部を動かし、その位置を探索し、1 点を探しあてるまでの時間を測定する。具体的には、目的とする点が $d_n < 0.05$ の範囲に入った場合に、ユーザが目的とする点を探しあてたとみなすことにした。正弦波を一点に置いてから、その点を見つけるまでを一回の試行とし、試行が終わったらすぐに新たな点に正弦波を置き次の試行を開始する。このとき、増幅率コントロール部で用いる関数は、各試行ごとに 4.2 節で述べた 5 種類をランダムな順序で切り替える。

5 種類の関数をランダムな順序で行う計 5 回の試行を、1 セットとして、20 セットを連続で行った。1 人あたりのおおよその実験時間は、増幅率コントロール部の評価では 1 時間程度、左右音量比コントロール部の評価では 40 分程度であった。実験は、事前の練習時間を設けず、すぐに開始した。

実験者は視聴覚等、本実験条件において問題なく試行ができる健康な 20 代の学生 20 名で、うち 10 名が増幅率コントロール部の評価、残る 10 名が左右音量コントロール部の評価を行った。20 名のうち 8 名は、最短で 1 年、最長で 6 年の楽器演奏経験があったが、クラシック音楽の経験者はいなかった。20 名中 5 名が女性、15 名が男性であったが、実験結果に男女間の違いは見られなかった。

実験者には、正弦波を置いた 1 点をできるだけ速く

探しあてるように指示した。また、20セットのうち、最初の5セットはズーム機能なし、次の5セットはズーム機能あり、さらにその次の5セットはズーム機能なし、最後の5セットはズーム機能ありで試行するよう指示した。なお、増幅率コントロール部の関数が増幅率コントロール部のみの評価を行った。

増幅率コントロール部の評価

左右音量比コントロール部の h_n^{left} と h_n^{right} をいずれも1としてその機能を無効にすることで、増幅率コントロール部のみの評価を行った。

実験者10名の平均を求めた結果、左右音量比コントロール部を無効としてモノラルで使用した場合には、(e) 距離の2乗に反比例する関数が最も早く探すことができた(表1)。これは、(e) 距離の2乗に反比例する関数が他の関数に比べて鋭いピークを持っているため、目的とする点の付近に近づいてから探しあてるまでの時間が短かったためである。

ズーム機能を用いた場合は、用いなかった場合に比べていずれの関数でも早く探すことができた。

表1 増幅率コントロール部の評価
Table 1 Evaluation of functions for mixing

関数	ズーム機能なし [s]	ズーム機能あり [s]
(a) 定数	112.5	56.2
(b) 傾き負の一次関数	29.2	24.0
(c) 正規関数	32.6	17.8
(d) 反比例	30.9	20.8
(e) 距離の2乗に反比例	23.8	16.9

左右音量比コントロール部の評価

左右音量比コントロール部を有効にして、ステレオの場合での評価を行った。実験者10名の平均を求めた結果、いずれの関数でも探しあてるまでの時間が短縮された(表2)。また、ズーム機能を用いるといずれの関数でもさらに早く探すことができた。

左右音量比コントロール部を有効にしたことで、(b)、(c)、(d)の関数を使った場合には10秒以上時間が短縮されたのに比べ、(e)は5秒程度しか短縮されなかった。これは、鋭いピークを持っている(e)では、目的とする点の付近に近づくまで音量が小さく、実験者が左右の音量の違いを感じる事が難しかったためであると考えられる。

(c)の正規関数はズーム機能なしでは、(b)や(d)に比べて時間がかかったのに対し、ズーム機能ありでは、最も速く探すことができた。これは、正規分布の尖度が他の関数に比べて小さいためだと考えられる。つまり、ズーム機能なしでは、目的とする点の付近に近づ

くのは早かったが、最終的に点を探すまでに時間がかかるという問題があったのに対し、ズーム機能を使うことで、付近に近づいてから最終的な点を探すまでの時間も短縮されたと考えられる。

以上のことから、コンサートスコープヘッドフォンを使用しているときに、目的とする演奏音の演奏者を素早く探すためには、左右音量比コントロール部を有効にして、ズーム機能ありで(c)正規分布を用いるのが適切であるという結果となった。

表2 左右音量比コントロール部の評価
Table 2 Evaluation of panpot control

関数	ズーム機能なし [s]	ズーム機能あり [s]
(a) 定数	58.7	39.2
(b) 傾き負の一次関数	16.9	13.3
(c) 正規関数	19.3	10.6
(d) 反比例	16.7	13.8
(e) 距離の2乗に反比例	19.2	16.9

6.3 アンケートの実施

複数の演奏者の演奏音を聞いた時に、投影映像上の各演奏者の位置の変化と、その演奏音のミキシングの変化との関係が自然となることも重要である。

そこで本節では、収録したコンサート映像および音源を用いて、コンサートスコープヘッドフォンを試用しながら、4.2節の増幅率コントロール部で述べた5種の関数を切り替えることで、どの関数が最も自然であるかアンケートにより調査した。実験者は、6.1節と同じ20名の学生である。

実験者は、10分程度コンサートスコープヘッドフォンを試用した後、アンケートを行う。これを、1つの試行として、4.2節で述べた5種類の増幅率コントロール部で用いる関数をランダムな順序で切り替えて試行することで、計5回の試行を連続して行う。実験者には、あらかじめアンケートの内容を伝え、さらに、各試行ごとに、増幅率コントロール部で用いている関数の名前を伝えた。実験にかかった時間は、1人あたり約1時間である。

アンケートは以下の4問からなる。

- 問1 投影映像の動きと各演奏者の演奏音の音量変化との関係が自然であった。
- 問2 画面の中央に近づいた楽器が適切に強調されていた。
- 問3 頭部を左右に急に動かしたときの、投影映像の変化と演奏音の変化の関係は自然であった。
- 問4 5種類の関数の特徴や違いについて感じたことを述べてください。

第1問から第3問は、5段階評価によるアンケートで、実験者は、各問に(1)全く同意できない、(2)ある程度同意できない、(3)どちらとも言えない、(4)ある程度同意する、(5)強く同意する、のうちから一つを選択する。表3は、各問における20名の回答の平均値で、(b)傾き負の一次関数と(e)距離の2乗に反比例が良い結果となった。

表3 アンケート結果
Table 3 Results of questionnaire

関数	問1	問2	問3	平均
(a) 定数	2.6	2.7	3.1	2.8
(b) 傾き負の一次関数	3.9	4.1	3.9	3.9
(c) 正規関数	3.4	3.6	3.6	3.5
(d) 反比例	3.6	3.3	3.1	3.3
(e) 距離の2乗に反比例	3.7	3.7	4.0	3.8

問4は、記述式のアンケートで、5種類すべての関数を用いて試用した後に回答させた。その結果、特に(e)距離の2乗に反比例する関数について、「楽器の音色の違いを明確に感じることができた」や「多くの楽器の音を知覚できた」のように、肯定的な意見が多かった。

そこで我々は、クラシックコンサートのように多くのパートからなる曲をコンサートスコープヘッドフォンを使って鑑賞する場合には、増幅率コントロール部で(e)距離の2乗に反比例する関数を使うのが適切であると判断した。

6.4 アイマークカメラを用いた観察

収録したコンサート映像および音源を用いて、コンサートスコープヘッドフォンを試用する様子を観察した。試用したのは、6.1節および6.3節と同じ20名の学生で、自分の見たいパートや演奏者を自由に見よう指示した。

試用時には、アイマークカメラ(株式会社ナックイメージテクノロジー製 EMR-9)を装着し、投影した映像中で実験者の視線がどこにあったのが確認できるようにした。試用時間は1人約10分である。なお、増幅率コントロール部で用いる関数は、6.3節の結果を反映し、(e)距離の2乗に反比例とした。

試用の結果、多くの実験者は、ある演奏者を注視してしばらくすると、別の演奏者を見ようと探索的な動きになり、その後別の演奏者を見ようと注視するという風に、注視と探索を繰り返していた。

演奏音の音量を確認したところ、実験者は、その時音量の大きな演奏者に視線を移動する傾向があることが確認できた。このことは、実験者が、どの演奏音が投影映像上のどの演奏者からのものであるか、ある程度把握できていたことを示している。

また、注視していた演奏者を、ズーム機能を使ってさらに強調する動作もしばしば観察された。これは、

実験者が、自分の望む演奏者を強調して視聴するというコンサートスコープヘッドフォンの機能を利用してきていたことを示している。

なお、試用後に聞き取り調査を行ったところ、ほとんどの実験者が試用終了後も再びコンサートスコープヘッドフォンを体験したいと述べていた。

7 まとめ

本論文では、人間が音を聴くときに自然に行う動作をヘッドフォンに搭載した3軸角度センサ、距離センサで検出し、ヘッドフォンに搭載したプロジェクタから出力する映像や音源のミキシングを変更するコンサートスコープヘッドフォンについて述べた。

本研究の意義として、まず第一に、特定の演奏者を視覚的にも聴覚的にも強調して鑑賞するというコンサート記録の新たな鑑賞方法を提案した点があげられる。受動的に鑑賞していた従来の音楽鑑賞に比べ、より能動的な鑑賞が可能となった。第二に、音を聴くときに自然に行う動作をセンサーで検出し、その動作とプロジェクタから出力される映像や音源のミキシングとの関連づけを行った点があげられる。これにより、初心者でも直感的に使用できるインタフェースが可能となった。第三に、ヘッドフォンに複数のセンサやプロジェクタを搭載したインタフェースを構築した点である。小型で、一般家庭をはじめ多くの環境で使用することが可能となった。第四に、魚眼レンズカメラと複数のラベリアマイクを使用して、コンサートスコープヘッドフォンで使用するコンテンツの録音・録画を行い、評価実験を行ったことである。

今後、コンサートスコープヘッドフォンを一般家庭で利用する際に生じる以下のような問題について検討を行っていききたい。第一に、一般家庭でコンサートスコープヘッドフォンを利用するためには、正面と左右に壁があるスペースが必要で、しかも、壁が白一色など、プロジェクターの照射に適している必要があるが、そのようなスペースが家に無い場合の対処法について検討していききたい。第二に、映像の提示方法としてヘッドマウントディスプレイが利用可能かについて検討していく。ヘッドマウントディスプレイを映像の提示方法として用いれば、スクリーンが不要となり、システムの小型化が期待できる。しかしその一方で、ヘッドマウントディスプレイは、激しく映像が変化すると「映像酔い」のような症状が生じたり、強い疲労を感じるなどの問題が指摘されている。実際に我々がコンサートスコープヘッドフォンの映像提示にヘッドマウントディスプレイの利用を試みた場合にも、そのような問題が生じていた。映像酔いや疲労感を軽減する方法について検討していく。第三に、魚眼レンズカ

メラや複数のマイクを設置して、コンサートスコープヘッドフォンのコンテンツを収録することに、現状では大きな労力がかかるが、その収録を容易にするための手法について検討していく。そして、様々なコンテンツの録画・録音を行い、音楽鑑賞以外での利用法についても検討していきたい。

謝辞

演奏の録画・録音にご協力いただいた筑波大学吹奏楽団の皆様にご感謝いたします。

参考文献

- [1] Pachet, F., and Delerue, O., A mixed 2d/3d interface for music spatialization, In Lecture Notes in Computer Science (no. 1434), pp. 298-307(1998).
- [2] Pachet, F., and Delerue, O., On-the-fly multi-track mixing, Proceedings of AES 109th Convention, 10 pages (2000).
- [3] Wu, J., Duh, C., Ouhyoung, M., and Wu, J., Head motion and latency compensation on localization of 3d sound in virtual reality, Proceedings of the ACM symposium on Virtual reality software and technology, pp. 15-20 (1997).
- [4] Goudeseune, C., and Kaczmariski, H., Composing outdoor augmented-reality sound environments, Proceedings of the International Computer Music Conference, pp. 83-86 (2001).
- [5] Warusfel, O., and Eckel, G., Listen. augmenting everyday environments through interactive soundscapes, Proceedings of IEEE Workshop on VR for public consumption, pp. 268-275 (2004).
- [6] 佐藤光一, デジタルコードレスサラウンドヘッドフォンの開発, PIONEER 技術情報誌, 14(2), pp. 66-73 (2004).
- [7] 浜中雅俊, 李昇姫: サウンドスコープヘッドフォン, 日本バーチャルリアリティ学会論文誌, 12(3), pp.295-304 (2007).
- [8] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, 45(3), pp.728-738 (2004).
- [9] 國田 豊, 尾川 順子, 佐久間 敦志, 稲見 昌彦, 前田 太郎, 館 : 没入形裸眼立体ディスプレイ TWISTER の設計と試作, 映像情報メディア学会論文誌, 55(5), pp.671-677 (2001).
- [10] 橋本 渉, 岩田 洋夫: 凸面鏡を用いた球面没入型ディスプレイ: Ensphered Vision, 日本バーチャルリアリティ学会論文誌, 4(3), pp.479-486 (1999).
- [11] immersive media, <http://www.immersivemedia.com/demos/>
- [12] Google Maps with Street View, <http://maps.google.com/intl/en/help/maps/streetview/>
- [13] Miyashita, S., Hamanaka, M., and Lee, S. H., Concert viewing headphones, Proceedings of International Conference on Advances in Computer Entertainment Technology, pp. 108-109 (2010).
- [14] OpenGL: The Industry's Foundation for High Performance Graphics, <http://www.opengl.org/>.

(2013年3月19日受付)

[著者紹介]

浜中 雅俊 (正会員)



2003年筑波大学大学院工学研究科電子・情報工学専攻博士課程修了。2003-2004年日本学術振興会特別研究員P.D., 2004年より2006年まで科学技術振興機構さきがけ研究員(専任)として独立行政法人産業技術総合研究所で研究に従事。現在, 筑波大学大学院システム情報系講師。音楽情報処理の研究に従事。2004-2005年オランダ・ナイメヘン情報認知研究所(NICI)客員研究員。2010-2013年科学技術振興機構さきがけ研究員(兼任)。博士(工学)。2001年情報処理学会山下記念研究賞, 2001年SCI (5th World Multiconference on Systemics Cybernetics and Informatics) in Art 優秀論文賞, 2003年筑波大学大学院優秀論文賞(博士課程長賞), 2005年ICMC2005 Best Paper Award (Journal of New Music Research Distinguished Paper Award) 各賞受賞。

李 昇姫



1999年筑波大学大学院芸術学研究科博士課程修了。博士(デザイン学)。同年, オランダデルフト工科大学デザイン工学部主任研究員。2000年筑波大学芸術学系講師。2001年より筑波大学大学院人間総合科学研究科感性認知脳科学専攻講師。2007年同専攻准教授として現在に至る。2004年オランダアイントホーベン工科大学客員研究員。1997年アジアデザイン学会最優秀論文賞、2001年WMSCI最優秀論文賞各賞受賞。2003年度21世紀COE拠点形成プログラム「こころを解明する感性科学の推進」の事業推進担当委員。2009-2010年総務省SCOPEプロジェクトリーダーとして子どもの安全を守るウェアラブルデバイスを開発。現在は, 感性情報評価およびTangible技術を用いた感性インタラクションの研究に取り組む。