

サウンドスコープパッド

浜中雅俊¹

概要 : 360 度映像上で演奏している複数の奏者を見ながら、頭部方向やハンドジェスチャーで聴きたい奏者の演奏音を強調して聴くことのできる iPad アプリケーション、サウンドスコープパッドについて述べる。従来は頭部方向の検出に、方位センサなどを搭載した特別なヘッドフォンを必要としていた。一方、我々の提案したサウンドスコープフォンでは、スマートフォンのフロントカメラをからの画像を用いたフェーストラッキング情報と iPad に内蔵の加速度・ジャイロセンサの情報を統合することで頭部方向の検出を可能としていた。しかし、演奏者の映像は表示できていなかった。本稿で提案するサウンドスコープパッドでは、タブレットの方向に応じて表示範囲が変化する 360 度映像上に演奏者映像を重畳して表示することで、あたかも目の前に演奏者がいるかのような VR 体験を提供する。構築したアプリケーションは SoundScopePad というアプリ名で Apple App Store 上で公開している。

キーワード : SoundScopePad, Appstore, フェーストラッキング, ハンドジェスチャー, 空間音響, 360 度映像

1. はじめに

サウンドスコープパッドは、複数の奏者が演奏している映像を見ながら、頭部方向やハンドジェスチャーで聴きたい奏者の演奏音を強調して聴くことのできる iPad アプリケーションである。音楽初心者は、複数のパートからなる曲の特定のパートを注聴したいと思っても、そのパートのメロディを聴き取ることが難しい場合がある。また、音楽初心者でなくても、特定のパートの音を、よりはっきりと聴きたい場合がある。

これまで我々は、ヘッドフォンに顔の向きを検出する方位センサや、手と耳との距離を測定する距離センサにより上記の機能を実現するサウンドスコープヘッドフォンを提案し、日本科学未来館や SIGGRAPH2009 Emerging Technologies 等で展示を行ってきた (図 1) [1, 2].



図 1 サウンドスコープヘッドフォンの展示風景

一方、対面での展示が社会情勢により現在困難となっていることから、同様の機能をスマートフォンアプリ「サウンドスコープフォン」で実装し、多くの人が体験できるようアプリの提供を開始した(図 2)[3]. サウンドスコープフォンで曲を再生しながら、ユーザが右を向けば、右から聞こえていた音が正面で、左を向けば左から聞こえていた音を正面で聴くことができる。ユーザが耳に手のひらを近づけ「耳を澄ますようなポーズ」をすると、正面から聞こえていた楽器の音によりフォーカスして聴くことができる。サウンドスコープフォンのコンテンツは公開当初は Miracle Garden の 1 曲のみであったが、新たに Wild Birds および鋼鉄魂 2022 が追加された [4].

Wild Birds は、国立環境研究所 生物多様性領域 福島地域協働研究拠点で、2015 年から 2018 年に渡って開催してきた福島県の帰還困難区域とその周辺で野鳥の声のモニタリング音声から種を判別するイベント「バードデータチャレンジ」(現在は新しい形での開催を模索している) で使用したデータの提供を受けたものである(図 3) [5].



図 2 サウンドスコープフォンの使用風景

¹ 理化学研究所 革新知能統合研究センター 音楽情報知能チーム
Music Information Intelligence Team, Center for Advanced Intelligence
Project (AIP), RIKEN

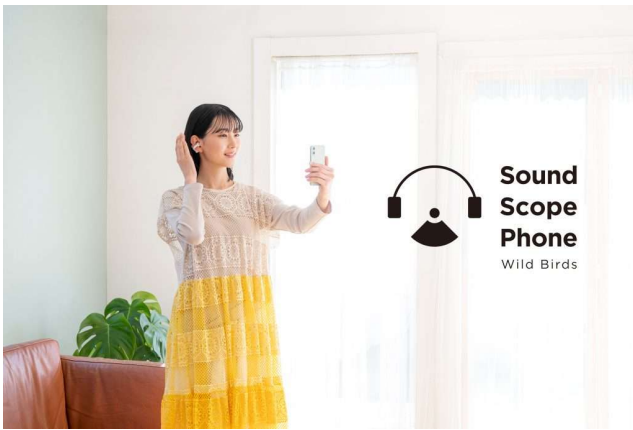


図 3 Wild Birds の聴取風景

コンテンツ Wild Birds では、9 種の野鳥の鳴き声の聴き分け体験ができる。

本稿で提案するサウンドスコープパッドは、サウンドスコープフォンの機能に加えて、iPad を向ける方向に応じて画面上にコンサートホールの 360 度映像が表示され、360 度映像上で複数の奏者が演奏する様子を観ることができる(図 4)。アプリ上でユーザと奏者の距離を設定すると、360 度映像上で奏者との距離および空間音響上での距離のいずれもが変化する。

以下本稿では 2 節で関連研究について述べ、3 節では、サウンドスコープパッドにより可能となる音楽鑑賞体験について述べる。そして 4 節では実装について述べ、5 節ではまとめと今後の課題について述べる。

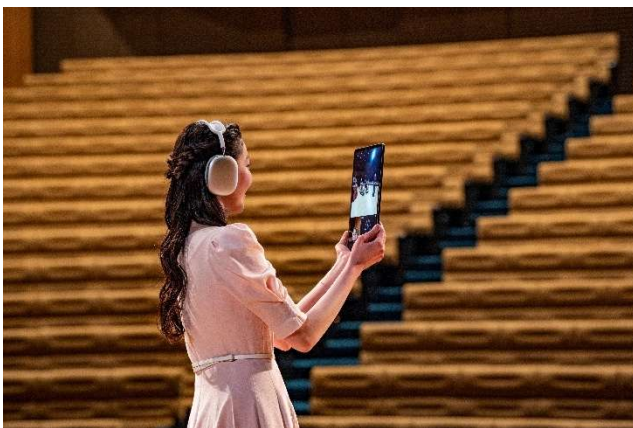


図 4 サウンドスコープパッドの使用風景

2. 関連研究

360 度映像や空間音響などの VR 関連技術が普及しつつある。本研究では、VR 体験に加えて特定のパートの強調やユーザと奏者の距離の調整などを可能とすることで、ユーザごとあるいは鑑賞するたびに新たな音楽体験を提供することを目指している。

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[6, 7]。近年では、Dolby Atmos (ドルビーアトモス)[8]を用いたアップルの空間オーディオ機能や、MPEG-H 3D オーディオを用いたソニーの

360 Reality Audio サービス[9]は、いずれも、専用のイヤフォンを用いて聴取するとパートごとの音が空間上に配置されるように体感することが可能である。これら[6-9]の目的は、仮想の音源位置を固定することで臨場感を高めることであつたため、本研究のように特定のパートを強調する目的では使われてこなかった。したがって、たとえば聴きたい楽器のすぐ近くに聴きたくない楽器が定位していた場合、聴きたい楽器のみを選択的に聴けるようにすることは困難であつた。

一方、空間音響システム [10, 11] では、聴取者のアバターや各パートの位置を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、各パートのミキシングを適切に変更するためには、ソロが始まったパートを近くに配置したり、ソロが終わり伴奏に戻ったパートを遠くに配置しなおすなどの煩雑な操作が必要なため、初心者には困難であつた。

我々は、iPhone/iPad の本体の加速度・ジャイロセンサによって得られた端末の角度情報と、端末のフロントカメラで取得できるユーザの顔の角度情報を統合することで、専用のイヤフォン/ヘッドフォンを用いずともインタラクティブな空間音響の体験を可能とする。

3. サウンドスコープパッド

サウンドスコープパッドを iPad にインストールすることでユーザは以下の体験が可能となる(図 5)。

3.1 VR コンサート体験

アプリを起動して再生ボタンを押すと iPad の画面上に 360 度映像から切り出された背景上に 3 名の奏者が現れる(図 5)。iPad を左に向けると 360 度映像上で左側で演奏していた奏者が現れる。同じように iPad を右側に向けるとその方向で演奏していた奏者が現れる。奏者の 360 度映像上での位置はプログラムにより任意に設定可能であるが、現在登録されている曲ではユーザの周囲を取り囲むように配置してある。ユーザから奏者までの距離はアプリ上で変更可能である。

3.2 空間音響体験

ヘッドフォンあるいはイヤフォンから出力される音は、空間音響ライブラリ OpenAL[12]を用いて各パートの音が空間上に配置されるよう処理されている。iPad をユーザの目の前に構えると、アプリケーションがディスプレイ上でユーザの顔の方向を認識し、その状態で顔を左や右に動かすと顔の角度を常に追従して、出力音を調整するため、顔を動かしても空間に配置された音源の位置が固定されているように感じられる。たとえば、ユーザが 180 度振り返ると、後ろで演奏されていると聞こえたパートの音が、振り返った後には正面で聞こえているように聞こえる。

3.3 能動的音楽鑑賞体験

手を耳にかざし耳を澄ますようなジェスチャーをすると

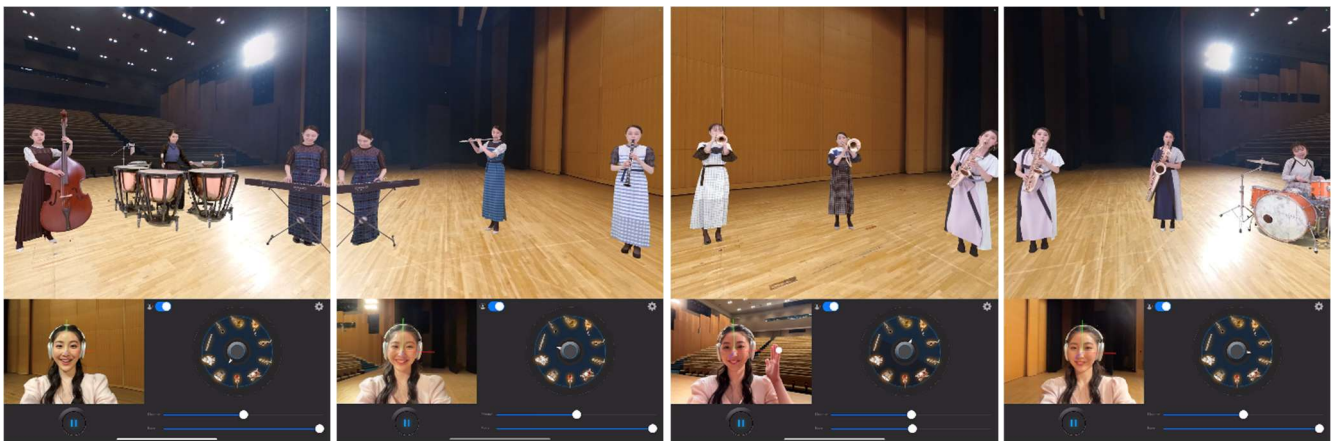


図 5 サウンドスコープパッドのスクリーンナップショット

アプリがそれを検出し、正面で聴こえているパートの音が強調される。手と耳を近づけたり遠ざけたりすると、強調したい奏者の演奏音の強調度合いや、強調するパートの範囲を調整することができる。iPad 画面左下の顔認識や手の認識のプレビュー画面(図 6)の右側にはパート配置画面(図 7)があり、強調度合いや強調範囲を確認可能である。

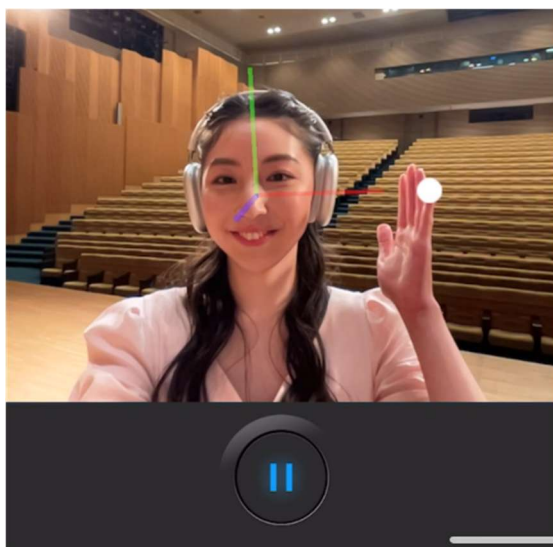


図 6 プレビュー画面

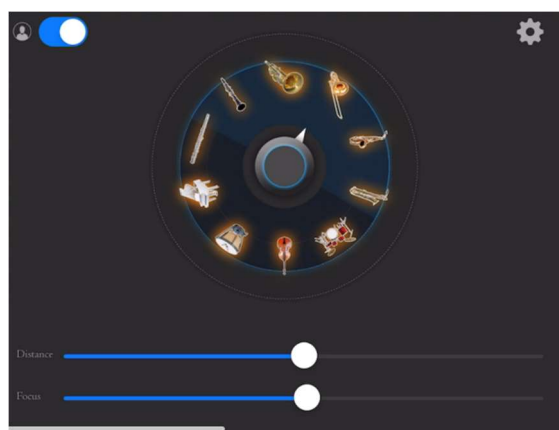


図 7 パート配置画面

4. 実装

360 度映像から切り取られて iPad に表示される複数の奏者を見ながら、頭部方向やハンドジェスチャーで自分の聴きたいパートを強調できる iPad アプリ、サウンドスコープパッドの実装について述べる。その特徴は以下の 3 つである。

4.1 フェース・ハンドトラッキング

頭部方向は、iPad のフロントカメラが取得した画像を解析して検出する。これはアプリ開発時の iOS のバージョン 14 では、Apple 製のイヤフォンやヘッドフォンに搭載されているジャイロセンサから情報を取得するためのアプリケーションプログラミングインターフェース (API) が、公開されていなかったためである。結果的に、特定のハードウェアに依存することなく、センサを搭載していないイヤフォン/ヘッドフォンでもインタラクティブな空間音響体験が可能となった。iPad の位置は絶えず変化しているため、iPad に搭載されているジャイロセンサによる本体の角度情報とフロントカメラが取得した頭部方向の情報を統合することで、実空間上でのユーザの頭部方向を検出することが可能となった。小指の付け根を特徴点とすることで、特定の部分を強調したいときに耳に手をかざすジェスチャーを安定して検出することが可能となった。iOS のバージョンが 15 となり Apple 製のイヤフォンやヘッドフォンに搭載のジャイロセンサの情報を取得する API が公開となり、今後、ジャイロセンサから得られる情報も利用する予定である。

4.2 360 度映像

アプリケーションの開発当初は、リアカメラで撮影した背景画像に奏者の映像を拡張現実 (AR) 合成することを検討していたが、フロントカメラで頭部方向検出を行いながらリアカメラを起動することができないことがわかったため、背景は事前に撮影した 360 度映像をもいいることにした。奏者の映像は、グリーンバックの前での演奏を撮影し、

事前にクロマキー処理したものをを用いている。iPad のジャイロセンサが取得した本体の角度情報に応じて、360 度映像から切り出して iPad に表示する範囲が変更される。したがって、ユーザは見たい奏者を探しながら能動的な音楽鑑賞をすることが可能である。

4.3 空間音響

従来のサウンドスコープヘッドフォンでは、検出した動作とオーディオミキサーの操作とを関連づけることで、聴きたいパートの音が正面で強調されて聞こえるようにしていた。一方、サウンドスコープフォンでは、3D 音響ライブラリ OpenAL[12]で定義される 3D 空間上に n 個のパートを配置することで、頭部の方向に配置されたパートの音を正面で聴くことができる。このとき、聴者のアバターから n 個のパートそれぞれとの作る角度を θ_n とする。

頭部の方位に応じて次式で増幅率 h_n^θ ($0 \leq h_n^\theta \leq 1$) を変化させることにより、正面に近い角度にあるパートの音をより強調して聴くことができる。以下、頭部の正面方向とパートの配置されている位置とのつくる角度を θ ($-\pi \leq \theta < \pi$)、手と耳との距離を δ ($0 \leq \delta \leq 1$) とする。なお、実装上での手と耳の距離は、手の小指の先端がインカメラの撮影範囲の 10% 内側の位置するときを 1、インカメラの撮影範囲の中心に位置するときを 0 とし、その間を正規化した値である。インカメラの撮影範囲の外側から 10% は、認識対象外とすることで動作を安定させている。

$$h_n^\theta = \begin{cases} 0 & \widetilde{h}_n^\theta < 0 \\ \widetilde{h}_n^\theta & 0 \leq \widetilde{h}_n^\theta \end{cases} \quad (1)$$

ただし、

$$\widetilde{h}_n^\theta = \begin{cases} 0 & \delta = 0 \\ 1 - (\alpha \cdot |\theta_n|) / (\pi \cdot \delta) & \delta > 0 \end{cases} \quad (2)$$

α ($0 \leq \alpha \leq 1$) は調節可能なパラメータで、 $\delta < 1$ の場合の増幅率の変化を設定する。 $\alpha = 0$ の場合には、耳に手を近づけても各パートの増幅率に変化はないが、 $\alpha > 0$ の場合には、耳に手が近づくとつれて増幅率が減少する。このとき聴者が向いている方向よりも、向いていない方向の増幅率の減少のほうが大きいので、正面の音が相対的に大きな音で聞こえるようになる。

4.4 ブーストモード・ミキシングコンソールの実装

4.3 節の実装のみでは、耳に手を近づけると聴者の向いていない方向の増幅率が減少する。これにより正面のパートの音は聴きやすくなるが、すべてのパートの増幅率の合計の値は小さくなるため、曲としての迫力に欠けた状態になりやすい。そこで、増幅率の合計が一定となるよう正規化したブーストモードを実装した。ブーストモードでは一つのパートの音量が大きくなりすぎる場合があるため、増幅率に閾値を設け、それ以上大きくならないようにするブーストリミッターを設けた (図 8 下端)。さらに、各楽器の音を個別に調整可能とするため、ミキシングコンソールを実装した。

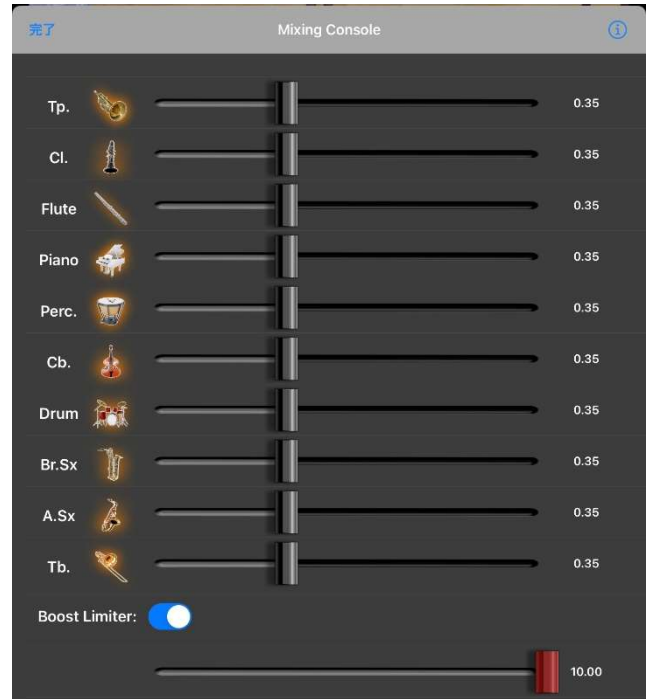


図 8 ブーストモードとミキシングコンソールの実装

5. おわりに

本稿では、360 度映像から切り出した奏者を表示しながら頭部方向やハンドジェスチャーで聴きたいパートを強調しながら聴くことのできるアプリケーション、サウンドスコープパッドについて述べた。サウンドスコープパッドは、アプリ名「SoundScopePad」として、Apple App Store 上で公開している[13]。また、利用方法を理解するための紹介動画を用意している[14](図 9)。アプリを 30 回利用するとアンケートを依頼するページの案内が表示される。ご協力いただければ幸いです。

iOS15 より Apple 製のイヤフォン/ヘッドフォンに搭載されたジャイロセンサの情報を取得する API が公開となった。今後、それらの情報もアプリで利用していく。



図 9 サウンドスコープフォンの紹介動画

参考文献

- [1] Masatoshi Hamanaka, SeungHee Lee. Sound Scope Headphones. ACM Siggraph2009 Talks TK-201 /Emerging Technologies ET-201, August 2009.
- [2] 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, 2007, vol. 12, no. 3, p. 295-304.
- [3] 浜中雅俊. サウンドスコープフォン. 情報処理学会 音楽情報科学研究会 研究報告 2021-MUS-131, Vol.2022, No.61, pp.1-3, June 2022.
- [4] 浜中雅俊. SoundScopePhone. <https://gttm.jp/hamanaka/soundscopephone/>, (参照 2022-05-16)
- [5] バードデータチャレンジ. 国立研究開発法人 国立環境研究所 生物多様性領域. <https://www.nies.go.jp/kikitori/contents/bdc.html>, (参照 2022-5-16)
- [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, 2004, Vol. 14, no. 2, p. 66-73.
- [7] 久木元 伸如, Ewe C. Huat, 竹田 仰. プロジェクション 型没入ディスプレイにおける 3次元音場生成のための実用的検討. 電子情報通信学会 技術研究報告, 1999, EA99-32, p. 29-36.
- [8] Dolby. ドルビーアトモス. <https://www.dolby.com/jp/ja/technologies/dolby-atmos.html>, (参照 2022-05-16)
- [9] Sony. 360 Realty Audio. https://www.sony.jp/headphone/special/360_Realty_Audio/, (参照 2022-05-16)
- [10] 鈴木陽一, 西村 竜一. 超臨場感音響の展開. 電子情報通信学会誌, 2010, vol. 93, no. 5, p. 392-396.
- [11] 安藤彰男. 高臨場感音響技術とその理論. 電子情報通信学会 基礎・境界ソサイエティ, 2009, vol. 3, no. 4, p. 33-46.
- [12] OpenAL. <https://www.openal.org/>, (参照 2022-05-16).
- [13] SoundScopePad. <https://apps.apple.com/jp/app/soundscopepad/id1600250919>, (参照 2022-05-16).
- [14] SoundScopePad. <https://gttm.jp/hamanaka/soundscopepad/>, (参照 2022-05-16).