

シーンベースのVR音響における上下定位の再現性向上に関する研究

Research Related to Improving the Reproducibility of Vertical Localization in Scene-based VR Audio

坂本 昭人 SAKAMOTO Akihito

デジタルハリウッド大学 助教
Digital Hollywood University, Assistant Professor

シーンベースのVR音響を用いた音楽制作において、上下定位の再現性を向上させることに関する研究である。360度動画の視点移動に連動し定位変化する音源のVR音響化において、左右の変化は比較的分かりやすく再現できることに對し、上下移動に伴う定位の変化は分かりにくい。そこで、VRコンテンツを再生するプラットフォームの仕組みを分析し、その条件下において、従来よりも上下定位の変化する様が認識しやすくなるような制作技術を構築し、これにより再現性を向上させることを目的とする。

キーワード：VR音響、Spatial Audio、Ambisonics、音楽制作、定位

1. 背景と目的

VRコンテンツが広く普及してきた今日において、その代表的な表現手法のひとつに360度動画がある。

360度動画の特徴のひとつは、視聴者が自由に動画像を視点移動させることに連動し定位等が変化する音である。この再現には、Spatial Audioや空間音響と呼ばれる音専用のVR技術が用いられ、その中でもシーンベースの仕組みを用いた制作が多い。シーンベースとは、視聴者の立ち位置を各音源の中心とし、視点移動に応じ音響が変化する仕組みのことで、360度動画と相性が良い。そこで本研究では、シーンベースの空間音響システムをVR音響と定義していく。

従来においてもVR音響の制作ワークフロー、そしてそこで作られたコンテンツを再生するプラットフォームは存在する。しかしながら、これらの環境下で再現されるコンテンツの特徴として、左右の視点移動に伴う音の定位変化は分かりやすいが、上下視点の移動に伴う音の定位変化が分かりにくいことが挙げられる。そもそも人は左右に比べ上下の変化を正確に認識しにくい特質を持っているが^[1]、別の要因のひとつとして、WebプラットフォームにVRコンテンツをアップロードする際、独自のエンコードがなされることで、そこで指定された仕様に沿った再生表現に変わってしまうことが考えられる。

そこで、再生プラットフォームの上下定位における可変システムを分析し、その仕組みを活用した制作ワークフローを構築することで、上下の視点移動に伴う音の変化という点における再現性を向上していく。

また、本研究における音の変化の測定には、音の大きさを表す音圧の値を用いる。上下軸角度が0度となる正面定位(以下、正面定位)での音圧と、上下軸角度-90度となる底面定位(以下、底面定位)、または、上下軸角度が+90度となる頂面定位(以下、頂面定位)での音圧を比較し、変化の値が増大することで再現性向上としていく。

2. 先行研究と仮説

VR音響の定位を設定する際は、対となる360度動画の描写に合わせる事が多い。ボーカル音源の定位を設定する場合は、ボーカリストの口元に定位を設定するというように、左右、そして上下の定位が視覚的、聴覚的にずれることのないよう留意する。この制作フローで構築されたVRコンテンツを再生し、視聴視点を上下に動かさない上下軸角度0度となる正面定位の音圧を基準値としながら、

上下90度に視点移動させた状態で再生した音の変化量を比較する。この差分が大きいほど聴感上でも違いが分かりやすくなることから、再現性の向上としていく。

コンテンツ制作には、VR音響の音楽制作ワークフロー「シーンベースのVR音響においてスタンダードとなり得る音楽制作ワークフローの構築」^[2]を用いる(図1)。

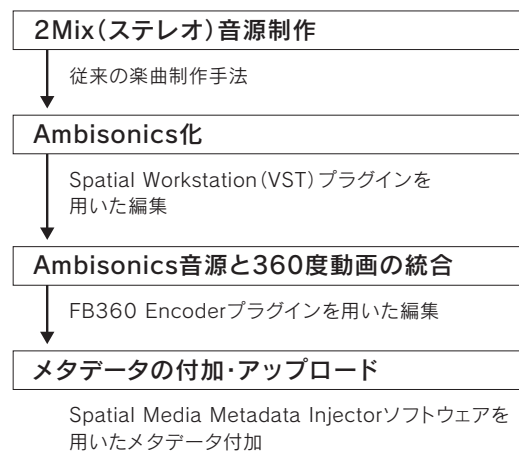


図1：音楽制作ワークフロー概念図

本研究では、コンテンツをVR音響化するために、Ambisonicsの形式を活用する。Ambisonicsとは、音響空間を再現するためのテクノロジーである。空間の中心点で全指向の情報(W)を取得し、そこから前後(X)、左右(Y)、上下(Z)それぞれの位置情報を計測処理することで、360度の音響情報を再現する。

実際に音の定位設定をするのは、音源をある程度作りAmbisonics化する時である。ここで、360度の動画像に合わせ、定位を設定することで、視点移動に連動した音の変化を再現することができる。コンテンツが完成した後はVR音響が再生可能なプラットフォームにアップロードすることで自動的にエンコードされ、VR音響をステレオ環境で聴くことができる仕組みになる。このコンテンツを再生し、上下軸において複数の定位で変化する音圧を調べることで、全体あるいは特定の周波数において値の変化が生じると考

える。特に特定の周波数による音圧変化に期待する。周波数ごとに音圧が変われば、任意の周波数において音の大きさを変化させるエフェクト機能であるイコライザー等を使うことができ、制作時には、変化が著しい周波数を事前に目立たせるよう調整することができ、上下の定位が変化した際の再現性を向上させることができる。

向上を定義するための具体的な数値検証については、音源から発生する音のパワーを表す音響パワーレベルという観点で見れば、2倍の変化を示す3dBを超す値を基準としていく。

また、ここで分析した結果をワークフローに反映させ、そのプラットフォームのための新しい制作手法を構築していく。

3. 検証

3.1 概要

左右の定位は一定とし、上下の定位変化を比較していく。具体的には正面定位を基準とし、そこから-90度下に視点移動させた底面定位と、基準から+90度上に視点移動させた頂点定位における再生音の変化を比較する。こうすることで、正面定位からの変化量を比較し、変化量が多いほど再現性の向上が確認できる。計測には、各周波数における音圧の変化を測定するスペクトラムアナライザーを用いる。尚、検証結果が分かりやすいよう、今回は全てLch(左チャンネル)の信号で検証する。

3.2 制作環境

本研究におけるVRコンテンツの再生プラットフォームはYouTubeを使用する。ここでは、次のような最小要件が指定されている^[3]。

- (1) メタデータがファイルに追加されている。
- (2) 使用している音声トラックが1つだけである。
- (3) 空間音声にアンビソニクス(AmbiX)形式が使われている。
- (4) 1次アンビソニクス(FOA)がサポートされている。
- (5) ヘッドロックステレオを組み合わせた1次アンビソニクス(FOA)がサポートされている。

これらの条件を満たすVRコンテンツ制作のための環境で実施する。

(1) コンピュータ

特別な設定あるいは高機能なマシンスペックは必要とせず、またWindowsでもMacでも可能とする。今回使用したコンピュータは下記の通りである(制作、再生、検証で共通)。

本体 : MacBook Air

OS : macOS High Sierra

プロセッサ : 1.7GHz Intel Core i5

メモリ : 4GB 1600MHz DDR3

グラフィックス : Intel HD Graphics 400

(2) DAW (Digital Audio Workstation)

今回、VR音響を制作するために、1次Ambisonicsのフォーマットで制作をしていく。1次Ambisonicsでは1つのトラックにおいて、4チャンネルの録音や再生ができる機能を必要とする。また、Ambisonics化するためにはAAXやVST形式に対応したプラグインを用いる。そこで、Cockosによって開発された「REAPER Version 6.12c」を使用する。

(3) プラグイン

2Mix音源をAmbisonics化し、動画像を見ながら定位を設定するため、Facebookが提供しているFacebook 360 Spatial

Workstation (VST) - macOS - v3.3.3の「FB360 Spatialiser (ambiX) (Two Big Ears) (16ch)」(以下、FB360 Spatialiser)と、「FB360 Control (ambiX) (Two Big Ears) (16ch)」(以下、FB360 Control)、「FB360 Encoder」を使用する。

FB360 Spatialiserは、実際に定位を調整したり、音源をAmbisonics化するために使用する。また、Ambisonics化された音源をそのままステレオ環境のDAWで再生しても正しい定位が再現されないため、FB360 Controlを用いて、ステレオ環境における正しい定位のモニタリングをする。そして、ここで制作したVR音響コンテンツと360度動画を統合するために、FB360 Encoderを使用する。

(4) ソフトウェア

VR音響のファイルにYouTubeの仕様に合わせたメタデータを付加するためにGoogleの提供している「Spatial Media Metadata Injector」を使用する。

3.3 再生環境

(1) ブラウザ

こちらもYouTubeの仕様に則り、Googleの提供する「Google Chrome (バージョン84.0.4147.89)」を使用する。

(2) 録音ツール

YouTube上の音をWeb上で再生し、それをリアルタイムにDAW上で測定することはできるが、マシン負荷または回線の安定具合による測定精度の低下、また、比較のための細かい再生設定等の利便性向上のため、視点移動させた状態でコンテンツを再生し、これをDAWに取り込みオフラインで測定する。取り込みにはオープンソースである「Soundflower (バージョン2.0.b2)」を使用する。YouTubeで再生される音をSoundflowerを用いて出力し、DAW側では入力設定をSoundflowerにすることで、信号の伝達経路を確保する。この状態で録音することで、外部機器等を用いたアナログ録音よりも劣化の少ないデジタル録音を可能とする。本研究におけるYouTubeからの録音は全てこの形で行うことで録音状態の均一化を図る。

3.4 検証環境

スペクトラムアナライザー(以下、アナライザー)

20Hzから20kHzの周波数帯において測定が可能で、かつ、2つの音源を同時に比較でき、さらに、AAXやVST等のプラグイン形式に対応し、今回のDAWでも使用することができることから、BLUE CAT AUDIOが提供する「Blue Cat's FreqAnalyst」を使用する。

3.5 比較検証

(1) 一般的な定位設定で構成された楽曲での比較

ここでは6個のトラック(音源)で構成された、坂本音響塾で制作した楽曲「K(仮)」^[4]を使う。各トラックの定位は動画像に従い次のように設定している(図2)。尚、上下させる定位の角度はVRカメラのレンズ位置となる上下軸角度0度、つまり正面定位とし、それより下に移動すれば-に、上に移動すれば+として表される。

ドラムの上下軸の設定角度 : -21.3度

ベースの上下軸の設定角度 : -13.7度

ギター1の上下軸の設定角度 : -10.3度

ギター2の上下軸の設定角度 : -32.3度

シンセサイザーの上下軸の設定角度 : -40.5度

ボーカルの上下軸の設定角度 : -4.8度



図2：楽曲「K (仮)」音源における定位の設定画面

動画に合わせていることから比較的上下軸の角度を0度付近に設定しているものが多い。これをVRコンテンツとして再生した時の、正面定位および底面定位、頂点定位で違いを測る。また、計測する範囲は全ての楽器（音源）が鳴るサビの部分、1分12秒から1分19秒の約4小節分とし、その間のアナライザー変化値の最大値で計測する。計測値に違いが出た主な周波数における値は下記の通りである。

正面定位を基準とした底面定位での主な変化値

- 22Hzで4dB減少
- 100Hzで2dB減少
- 450Hzで2dB増加
- 1.4kHzで3dB減少
- 11kHzで5dB増加

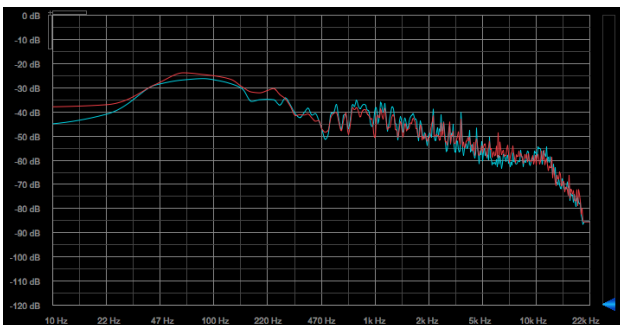


図3：正面（青）と底面（赤）定位における周波数特性

正面定位を基準とした頂点定位での主な変化値

- 22Hzで2dB減少
- 100Hzで3dB減少
- 450Hzで2dB減少
- 1.4kHzで3dB増加
- 11kHzで8dB減少

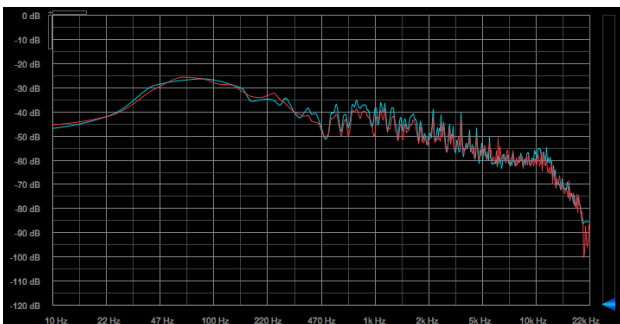


図4：正面（青）と頂点（赤）定位における周波数特性

このことから、上下の定位変化に伴う周波数ごとの変化が確認できる(図3)(図4)。

(2) 正面定位に設定したホワイトノイズ音源を用いた比較

次に、周波数ごとの変化の値をより詳しく把握するため、全ての周波数でおおよそ同じ大きさで再生されるホワイトノイズを用い、これを正面定位(0度)に設定することで比較をしてみた^[5]。左右の定位は0度としている。尚、比較対象とする周波数帯はこれまでと同様とする。

正面定位を基準とした底面定位での主な変化値

- 22Hzで4dB減少
- 100Hzで4dB減少
- 450Hzで4dB減少
- 1.4kHzで4dB減少
- 11kHzで4dB減少

正面定位を基準とした頂点定位での主な変化値

- 22Hzで3dB減少
- 100Hzで3dB減少
- 450Hzで3dB減少
- 1.4kHzで3dB減少
- 11kHzで3dB減少

このように、全周波数帯で変化量が同じとなる結果であった。この現象は、左右に視点を変え、その上で、正面定位、底面定位、頂点定位を比較しても同様である。

さらに正面定位の音源が複数になっても同じ現象となった^[6]。

しかし、楽曲を用いた検証では周波数ごとの変化が見られたことからこの矛盾を考察すると、両者の違う点として、元々設定した上下軸の設定角度の差が要因の一つとして考えられる。楽曲では動画像に合わせた定位設定のため正面定位(0度)のものではなく上下軸の角度が付いている状態である。一方でホワイトノイズを用いた検証では、全て上下軸の設定角度が0度である。このことから、YouTubeの仕組みは、上下軸における定位の移動では特定の周波数に関わらず、全周波数において一律に音圧が変化し、元々音源が定位する場所によってこの増減の値が変化するという新しい仮説ができる。

(3) 上下軸の角度を付けた設定のホワイトノイズ音源を用いての比較

新しい仮説を受け、再びホワイトノイズを使用し、制作時点で正面定位でない、上下軸の角度を付けた状態で検証した^[7]。

ホワイトノイズの上下軸の設定角度：-45度

正面定位を基準とした底面定位での主な変化値

- 22Hzで変化なし
- 100Hzで変化なし
- 450Hzで変化なし
- 1.4kHzで変化なし
- 11kHzで変化なし

正面定位を基準とした頂点定位での主な変化値

- 22Hzで5dB減少
- 100Hzで5dB減少
- 450Hzで5dB減少
- 1.4kHzで5dB減少
- 11kHzで5dB減少

結果、ホワイトノイズを正面定位に設定し比較した検証と同様に、

周波数ごとによる変化は見られないものの、元々設定した角度からの距離が遠くなる頂点では音圧が低下し、元々の定位から距離の絶対値が変わらない底面では音圧に変化が見られず、新しい仮説を証明することができた。

(4) 視点移動に伴う音圧変化を優先し設定した楽曲での比較

この原理を利用し、楽曲において、上下の視点移動に伴う音圧の変化が現れるよう制作に反映する^[8]。ここでは結果を分かりやすくするため、低音域(90Hz付近)を基音とするベースと中音域(550Hz付近)を基音とするギター1の2つの楽器のみを使用する。両音源とも正面定位を避け、動画像に準じつつも上下どちらかにやや寄った設定で検証している。2つの上下軸の定位角度は下記のような設定である。尚、ここでいう基音は両楽器ともアナライザー測定において最も音圧値の高い値から定義している。

ベースの上下軸の設定角度：-12.4度

ギター1の上下軸の設定角度：+17.9度

正面定位を基準とした底面定位での主な変化値

22Hzで2dB増加
100Hzで1dB減少
450Hzで2dB減少
1.4kHzで1dB増加
11kHzで変化なし

正面定位を基準とした頂点定位での主な変化値

22Hzで2dB減少
100Hzで3dB減少
450Hzで1dB減少
1.4kHzで2dB増加
11kHzで変化なし

このように、元々底面寄りに定位設定したベースの低音域成分が底面定位では増加したり減衰量が控えめとなり、元々頂点寄りに定位設定したギターの中音域成分が頂点定位では増加したり減衰量が控えめとなる結果となった。ここでは、元々の上下軸の定位角度を20度以内と控えめにしていることから、アナライザーによる変化量も小さいものとなっているが、ほとんど変化の見られなかった正面定位の検証とは、明らかに違いが出ていることが確認できる。このことから、YouTubeにおけるVR音響処理の仕組みは、設定する上下軸の設定角度による音圧の変化は周波数に関係なく一定であるが、元々設定した上下軸の角度がこの変化に影響を与えることが分かり、そして、この仕組みを利用することで、意図的に変化量を調整することができるようになった。

3.6 新しい制作手法

プラットフォームの仕様が確認できたことから、この仕様を用い、より再現性が向上する制作手法を構築する。

これまでの検証により、上下軸の定位変化による左右の定位または音質の変化は見られない。そして、上下軸の定位変化は、元々の上下軸の設定角度からの距離により増減されることが確認できている。そこで、音源を複製し同じ音源を2つ使用することで、正面定位における本来の音質や音量等の楽曲バランスを担保しつつ、上下の視点移動による音圧の変化を大きくすることで再現性の精度向上を目指す(図5)。具体的には、複製元となる音源は従来通りの制作手法で活用するものとし、複製後の音源は、視点が上下移動した際に変化する音専用のものとして活用する。各楽器の定位は次の通りである。尚、両者の左右の角度は同様の角度とする^[9]。

ベースの上下軸の設定角度：-12.4度

複製した同じベースの上下軸の設定角度：-90度

ギター1の上下軸の設定角度：+17.9度

複製した同じギター1の上下軸の設定角度：+90度

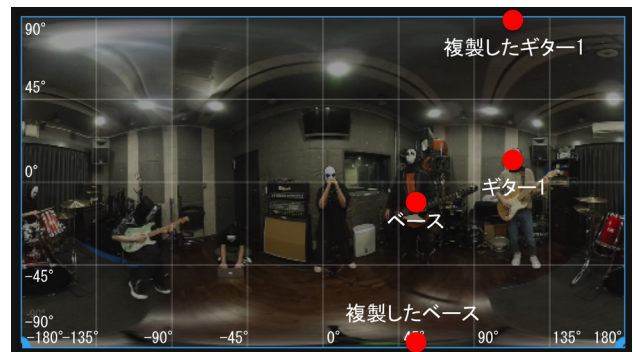


図5：同じ音源を2つ使用した上下定位の設定画面

このように、元々の楽器は動画像に準じた定位設定とし、複製した同じ音源を、ベースは底面定位となる-90度に、ギターは頂点定位となる+90度に行っている。尚、音源が倍増したことで全体の音圧も上がることから、本検証ではピークオーバー防止のため全ての音源を-6dB程減少させているが、本研究では正面定位からの変化量を比較するため、全ての音源が一律で変化していれば問題ない。その他、音質やタイムライン、左右等の設定はこれまでと同じである。

正面定位を基準とした底面定位での主な変化値

22Hzで4dB増加
100Hzで2dB増加
450Hzで3dB増加
1.4kHzで1dB減少
11kHzで4dB減少

正面定位を基準とした頂点定位での主な変化値

22Hzで4dB減少
100Hzで3dB減少
450Hzで4dB増加
1.4kHzで4dB増加
11kHzで2dB増加

このように、底面に行くほどベースの基音となる低音域の周波数帯が強調され、中高音域は目立たなくなり、頂点に行くほどギターの基音となる中音域の周波数帯が強調され、低音域は目立たなくなる結果となった。変化量においても、上下の定位を20度以内として検証した数値よりも大きくなっており、再現性の向上を確認することができる。

4. 結果と考察

YouTube上におけるVR音響の再生仕様は、周波数に関わらず一定に増減するが、元々の音源定位によりその度合いは変化する。そこで、制作時において上下の定位に差をつけておくことで、再現したい上下の定位の変化における音の変化を事前に調整することができる。

さらに、同じ音源を複製し、1つは本来の動画像に準じた定位、もう1つは上下の定位変化時に作用する定位と分けることで、制作時における細かい調整が可能となる。

今回の研究では上下の定位を設定することのできる最大角度となる90度で試み、一番大きく変化したところで4dBの差が出たことで、再現性の向上を確認することができた。制作意図にもよるが、上下の定位の音圧変化を十分に果たしたい場合、複製する音源の定位設定

は+90度または-90度というように、変化をもたらしたい上下軸の方向において、できるだけ最大の角度となるように調整する方が良いと考えられる。

5. 今後の課題

再現性向上が見られた半面、その変化量は4dB程度であることから、より大きい変化を付けることができる手法の研究余地はある。

また、今回はトラック数を増やすことで目的を達成したが、処理するトラックが増えるほど、マシン負荷は増え、制作時におけるトラック操作や管理の負荷もかかる。

YouTubeをプラットフォームとし、その仕様に沿った上下軸定位の再現性向上はできたが、VR音響を再生するプラットフォームは他にも存在し、今後も増えることが考えられることから、複数のプラットフォームに共通した制作手法であると利便性が高い。

参考文献

- [1] 倉本 敏行, 佐多 正至, 高橋 誠:『前方平面に配置したスピーカによる音像の定位』電子情報通信学会技術研究報告(2002年), 103-107頁.
- [2] 坂本 昭人:『シーンベースのVR音響のための標準的音楽制作ワークフローの提案と構築』DHU JOURNAL vol.06 2019(2019年), 56-61頁.
- [3] YouTubeヘルプ:“360°動画やVR動画で空間音声を使用する”
<https://support.google.com/youtube/answer/6395969>
(参照2020年7月3日).
- [4] akihito sakamoto:“【VR音響研究】坂本音響塾6期生楽曲「K(仮)」プリプロ VR音響実験”
<https://www.youtube.com/watch?v=dF3fhTGjzE0> (参照2020年7月9日).
- [5] akihito sakamoto:“WhiteNoise30s A VR Audio Movie injected”
<https://www.youtube.com/watch?v=XfvVuue4Gq8>
(参照2020年7月9日).
- [6] akihito sakamoto:“WhiteNoise30s A B VR Audio Movie injected”
<https://www.youtube.com/watch?v=apl7f2ZPW0>
(参照2020年7月10日).
- [7] akihito sakamoto:“WhiteNoise30s C VR Audio Movie injected”
<https://www.youtube.com/watch?v=6eCVJ-5AmUA>
(参照2020年7月13日).
- [8] akihito sakamoto:“【VR音響研究】坂本音響塾6期生楽曲「K(仮)」プリプロ VR音響実験 Gtr1&Ba 定位設定小”
<https://www.youtube.com/watch?v=RCk1C1FDbvl> (参照2020年7月20日).
- [9] akihito sakamoto:“【VR音響研究】坂本音響塾6期生楽曲「K(仮)」プリプロ VR音響実験 Ba+Gtr1 定位設定大(90°)”
<https://www.youtube.com/watch?v=rs2vZkyUIGM> (参照2020年9月11日).