

VR音源を用いたミックスダウンにおける 音像幅調整のための技術研究

Technical research for adjusting the sound image width in mixdown using a VR Audio source

坂本 昭人 SAKAMOTO Akihito

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

VR音響技術を用いた音楽コンテンツは年々増加し、音楽業界においてもインディーズレーベルからメジャーレーベルまで、企業規模に関わらず活用されている。その要因のひとつに、VR音響専用のマイクの普及や従来のステレオ(以下、2mix)音源をVR音響化する編集技術やツールが広まってきたことが挙げられる。VR音響のフォーマットに従い、音を録音したり編集することができるようになった今日、次に求められるのは、それらの音をまとめ上げるVR音響のミックスダウン(以下、VRミックス)である。従来の2mixにおいても同様で、このミックスダウン(以下、ミックス)は制作の最終工程にもなり得、非常に重要である。ミックスでは、定位や音質、音圧等様々な要素をバランスよく調整することが求められる。本研究では、その中において、音源の音像幅調整を図るものである。また、VRミックスの本質追求や教育性、汎用性等を鑑み、特定のツール機能には依存せず、多くのハードウェアまたはソフトウェアに導入されている機能を用いて研究を進める。尚、本来、VR音響という言葉は広義として扱われるが、ここでは、Spatial Audioと呼ばれる空間音響テクノロジーにおけるシーンベースのフォーマットをプラットフォームとし、VR音響と定義していく。

キーワード：VR音響、Spatial Audio、空間音響、音楽制作、音像

1. 背景と目的

ミックスとは、これまで録音や編集を施した複数の音源をバランスよくまとめ上げる工程である。音量バランスを整えることで、楽曲内での音源の主役と脇役を明確化したり、音質を整え聴き心地を良くしたりする、音楽制作の最終工程であり、調整要素は非常に多岐にわたる。よって、ミックスの良し悪しが最終クオリティに影響すると言っても過言ではない^[1]。

その要素の中に音像の調整がある。音像とは、音源が音響空間の中で存在する位置や方向等を示す広さのことを言う。本研究では、VRミックスにおいて、音源が持つ音像の幅を調整することを目的として行うものである。

音楽コンテンツだけを用いた音響空間再現の場合、視覚情報は無く聴覚情報を優先し制作することができるが、VR動画を併用し空間を再現する場合には、視覚と聴覚が相互に補完する中で成立する。そのため、音響を設定する上では動画像との関連性に配慮する必要がある。さらに、特に商業コンテンツの場合、動画像単独、音響単独で再生してもコンテンツとして成立する空間を再現しなければならない。これら全ての目的達成を目指す際、時に両者からは相反する調整を求められることがある。解決手法のひとつとして、動画像から音響に対する配慮、または、音響から動画像に対する配慮がある。そこで、後者における技術的配慮として、音像幅の調整が考えられる。VR音響を制作する過程において、各音源の音像幅を意図的に調整することができれば、音響品質の担保と動画像との関連性のバランスを自由に調整することができる。

具体的な音響幅調整のメリットのひとつとして、2mix、VRミックスに関わらず、音像幅が広がるほど、他の音源と混ざり合うことで生まれる新しい周波数成分が発生し、単一音源では成し得ない特定周波数の音圧増加または減少を実現したり、定位感の強弱を表現すること等、様々なパラメーターを調整することができる。

そこで、2mixとVRミックスにおける音源エネルギーを比較し、

特徴の違いを示しつつ、VRミックスで有効な音像幅調整の技術を模索し提言していく。

2. 先行研究と仮説

音源の定位状況や再生フォーマットにもよるが、従来の2mixの音像最大幅を左チャンネル(Lch)が -90° 、右チャンネル(Rch)が 90° の 180° 度幅と仮定すると、VR音響ではその幅が 360° と倍になり、さらに上下軸も対象となることから、音像幅の調整がしにくいことは容易に想像できる。具体的には、VRミックスでは、音楽表現ができる空間となる音像の範囲は2mixより広いため、ひとつの音源が持つ音響空間全体に占める音像幅は狭いと考えられる。そこで、意図的に音源の音像幅を広げることができれば、VRミックスとして有効な表現手法が増えると考ええる。

例えば、2mixの場合、やや左に定位させた音源と、やや右に定位させた音源が存在するとき、その中間地点となる中央定位でも両音源は少ない音量で鳴っている。そのため、重なる周波数成分が加減され、別の鳴り方をすることがある。楽曲のミックスの場合、いわゆるこの混ざりが音楽表現を豊かにする、または、リアリティを向上させる等の目的で意図的に用いられることが多い。

しかし、VRミックスの場合、左右の音像の幅が 360° と広く各音源の独立性が高い分、調整が難しい。

言い換えると、この独立性がVRミックスの特徴であるとも言えるが、楽曲制作において音像幅の調整ができれば、より表現の幅が広がることや、従来の2mixの考え方を応用利用できるといったメリットがある。

そこで、2mixとVRミックスの特徴の違いや相関関係等について、音源エネルギーを用い検証しながら、VRミックスにおける音像幅の調整化実現を目指していく。

尚、上下の定位変化による音像変化については、先行研究「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」^[2]

により、全周波数帯一律で最大4dBほどの変化が生じることが分かっているため、今回は上下の垂直軸を0°と固定し、左右の音像幅に特化し行う。

また、本研究で行う制作のワークフローは、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」^[3]を用いる(図1)。

VR音響専用のマイクで録音した音の場合、音源個別の調整が難しいため、音源録音は個別に行い、それらをVR音響化した後ミックスする工程である。

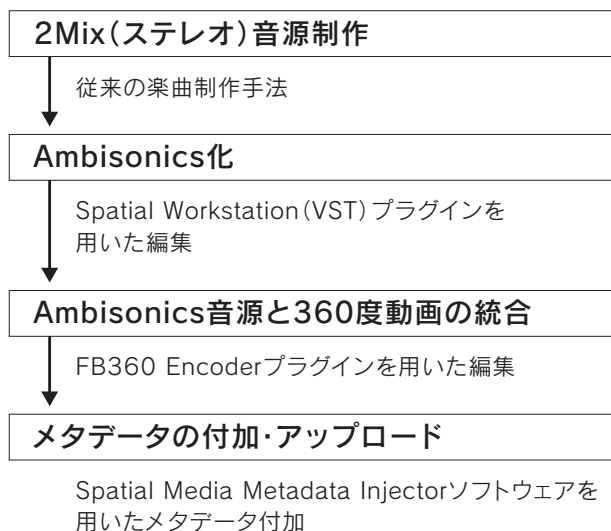


図1：音楽制作ワークフロー概念図

3. 検証

3.1 概要

まず、同じ条件のもと、2mixフォーマットとVRミックスフォーマットにおいて、単一音源の音像幅の標準値を測る。比較対象として、中央に定位した場合と、左チャンネル側-45°に定位した場合の状況を使用する。両フォーマットを比較し違いが生じた場合は、差分の特徴に鑑み、VRミックスフォーマットに適した新しい音像幅調整の技術を模索する。

3.2 制作環境

本研究における音源の再生プラットフォームはYouTubeを使用する。2mixのコンテンツでは比較的多くのフォーマットに対応しているが、VRミックスの場合、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Ambisonics (AmbiX)形式が使われている。
- (4) 1次Ambisonics (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Ambisonics (FOA) がサポートされている。

これらの条件を満たすVRミックスコンテンツで実施する。

(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：1.4GHz Intel Core i5

メモリ：8GB 1600MHz DDR3

グラフィックス：Intel HD Graphics 5000

(2) DAW (Digital Audio Workstation)

VRミックスを行うために、1次Ambisonicsのフォーマットで制作をしていく。Ambisonicsとは、音響空間を再現するためのテクノロジーで、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^[2]。1次Ambisonicsとは、その中で最も扱いやすくコンピュータ負荷等も低い、低次のものであり、First Order Ambisonics (以下、FOA)とも呼ばれる。FOA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が必要とする。また、音源をAmbisonics化するために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 6.16」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Spatial Workstation - OSX - v2.2.1の「FB360 Spatialiser (ambiX) (Two Big Ears) (7->8ch)」(以下、FB360 Spatialiser)と、「FB360 Control (ambiX) (Two Big Ears) (8ch)」(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。音源の定位設定では、視聴者(視聴位置)から音源までの距離が設定できるが、今回はデフォルト設定となる1m距離のまま進める。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ミックスコンテンツと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ミックスしたコンテンツファイルに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。

(5) 音源

特定の周波数に限定せず検証を進めるため、全周波数帯でおおよそ同じ強度を再生するホワイトノイズをモノラル形式で使用する。

3.3 再生環境

(1) ブラウザ

YouTubeの仕様に則り、Googleの提供する「Google Chrome (バージョン 92.0.4515.131)」を使用する。

(2) 録音ツール

YouTube上の音をそのまま再生し、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、一度DAWに取り込みオフラインで測定する。取り込みにはSolid State Logic社が提供するオーディオインターフェース「SSL2+」を使用する。同機材のアウトプットから同じく同機材のインプットに直接信号を入力することで、再生環境と録音環境の違いを無くし、ハードウェアによる音質等の変化を極力少なくする。本研究におけるYouTubeの再生および録音は全てこの形で行う。

3.4 検証環境

音源信号の計測には、Waves Audio Ltd.が提供する、「PAZ Analyzer」を使用する。ウェーブレットリアルタイムアナライザによって、ステレオ再生時における音源エネルギーを音像とし計測することで、音像幅を視覚的に知ることができる。本プラグインでは、左

右の定位幅が左チャンネル-45°、右チャンネル45°の90°幅での表記となる。本研究では、VRコンテンツとの比較を前提としているため、ステレオ環境下の条件として、分かりやすく左右±90°の180°幅で想定しているが、今回使用する定位角度は、左チャンネル最大-90°の内、半分となる-45°としているため、プラグイン内では-22.5°と置き換えることができ、大きな問題はない。

3.5 検証

(1) ホワイトノイズを用いた2mix音源^[4]で中央定位と左チャンネル-90°定位(以下、L ch -90°)での音像を測定したところ、どちらも、定位した角度のみに音像が現れ、その幅は限定されたものとなった。

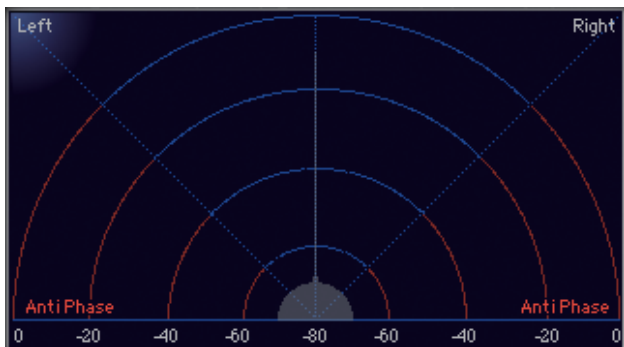


図2：2mixにおける中央定位の音源エネルギー

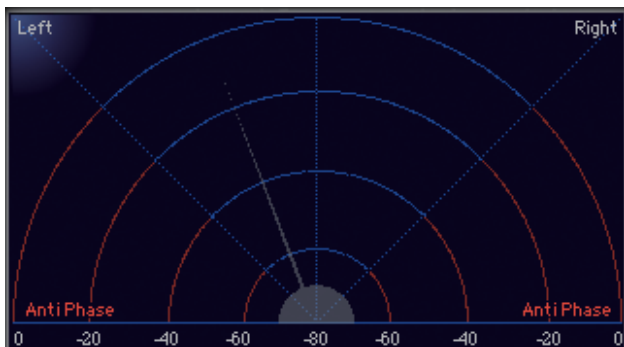


図3：2mixにおけるLch -90° 定位の音源エネルギー

(2) VRミックス音源におけるホワイトノイズの中央定位とL ch -90°での音像^[5]を測定したところ、中央定位の音源では、2mixと同様の結果となったが、L ch -90°に定位した音像には特徴的な形や幅が現れるものとなった。

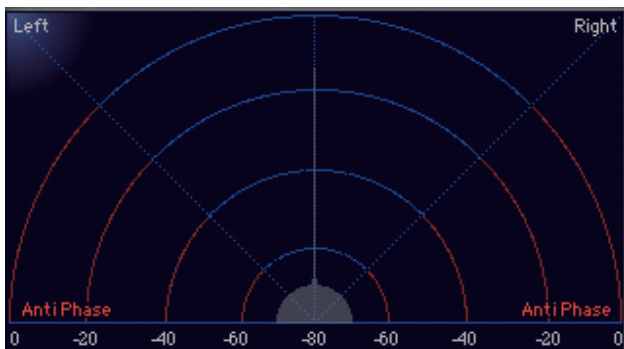


図4：VRミックスにおける中央定位の音源エネルギー

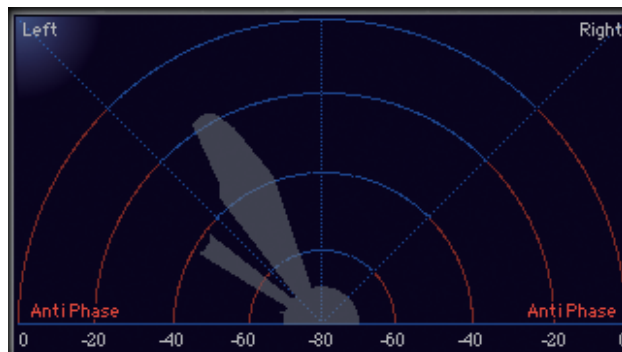


図5：VRミックスにおけるL ch -90° 定位の音源エネルギー

いずれも、おおよそ-15°程度のエネルギーを頂点としている点は変わらないが、VRミックスにおいては、中央定位を外れると2mixよりも音像幅が広く再生されることが分かった。VR音響フォーマットという音像全体の広さを鑑みれば、2mixと比べその幅が広がることは想定できるが、音源からやや離れた位置が最も広い音像幅となっていることが分かる。要因としてVRミックスでは奥行きが存在することから、今回設定した1mの距離が反映されていると考えられる。

(3) VRミックスに適した新しい音像調整手法の検証

音像幅の具体的な調整手法の検証のひとつとして、音像幅を拡大することを目指す。拡大するためにはいくつかの方法が考えられる。例えば、似た音源を複製し擬似ステレオ化したり、イコライザーを用いてバンド幅を調整したりすることである。しかし、VRミックスの場合、いずれもエネルギー量の形、すなわち音質自体を大きく変えてしまう可能性が考えられる。

そこで、極力このエネルギー量の形を変えずに音像幅だけを拡大するための別の手法として、音源の倍音成分に着目した。

倍音とは音高となる基音の周波数を基準とし、その整数倍にあたる周波数を持つ音の成分のことである。例えば、基音が1kHzの場合、2kHzや4kHzが倍音に当たる。また、音楽では周波数の2倍となる倍音がちょうどオクターブ上の音階となることも注目した理由である。音楽においては、音階の変更がなければ、コードやキーの遵守、不協和音防止等の観点からもメリットが多い。

ここでは、1kHzを基音とし、その2倍および1/2倍を含めた2倍倍音の増加調整^[6]と、オクターブ関係にはならない3倍および1/3倍を含めた3倍倍音の増加調整^[7]をし、L ch -90°の定位に限定し比較しながら、音像幅の拡大を図った。

尚、倍音調整には、特定の周波数だけを増減する機能を持つグラフィックイコライザーを使用する。今回は、Apple社が提供する「AUGraphicEQ」を使用した。また、各倍音における増減の値は音圧の強さで2倍となる6dBの増加で行った。

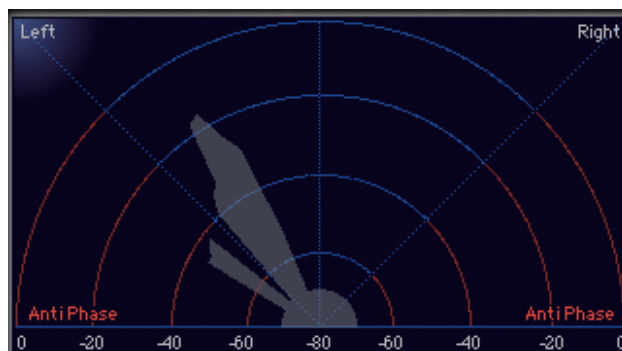


図6：2倍倍音において6dB増加の後の音源エネルギー

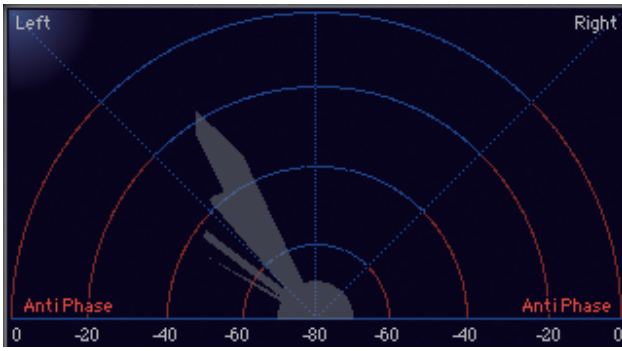


図7：3倍倍音において6dB増加の後の音源エネルギー

検証の結果、オクターブ間隔となる2倍倍音は比較的元音のエネルギーの形を保持しつつ、音像幅の拡大が図れた(図6)。一方オクターブ関係にならない3倍倍音では、音像幅こそ拡大してはいるが、エネルギーの形が大きく変化し、特に距離1mの地点では大きく異なる結果となった(図7)。本研究では、音像幅調整を目的としてはいるが、音源エネルギーの形が変わるということは、音質自体が変わるということであり、音楽制作において有効とは言えない。

(4) 倍音成分増加量の調整

2倍倍音の調整により目的を達成した面もあるが、細部を見ると、元音と全く同一の音源エネルギーの形とは言えない結果ともなっている。特に音楽制作では、1Hz単位または0.1dB単位での微妙な調整が必要になることから、さらに元音を担保した音像幅の拡大を図りたい。これまでの検証では、結果を分かりやすくするため6dBでの調整を図ったが、その分が音質の変化に影響したと考えられるため、2倍倍音における増加量を2dBに抑え、再度検証^[6]した。結果、音像幅は広くなり、さらに6dB増加させたときより元音に近い形となり、増加量による元音への影響を確認することができた(図8)。

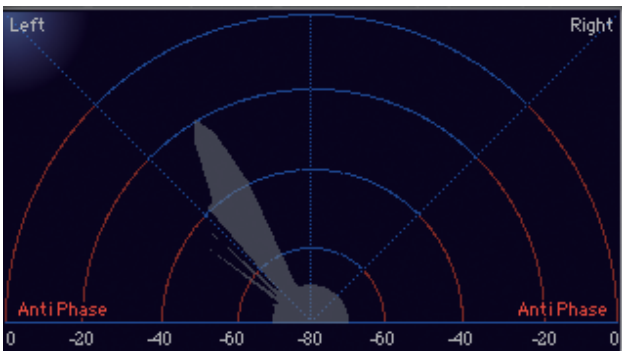


図8：2倍倍音において2dB増加の後の音源エネルギー

(5) 周波数による音源エネルギーへの影響の検証

これまでは倍音を一律に増加させることで音像幅の拡大をしてきたが、より実務的に有効な知見を得るため、周波数帯域による音源エネルギーへの影響を検証した。ここでは、Avid社が提供するパラメトリックイコライザー「EQ III 7-Band」を使用し、低音域と高音域の調整に分けて行った。まず、低音域が音像変化に与える影響を検証するため、基音となる1kHzには影響しない範囲で高音域のアッテネートを施したため、シェルピングタイプで1.8kHzを基準に、同プラグインで最大の調整値となる-12dBのアッテネートを施した。結果、視聴位置に近い音像幅のみが強調される結果となった(図9)。

次に、高音域が与える影響を検証するため、シェルピングタイプで500Hzを基準に-12dBの減少処理を施したところ、距離の長さや定位感の強さを示しやすい結果となった(図10)。

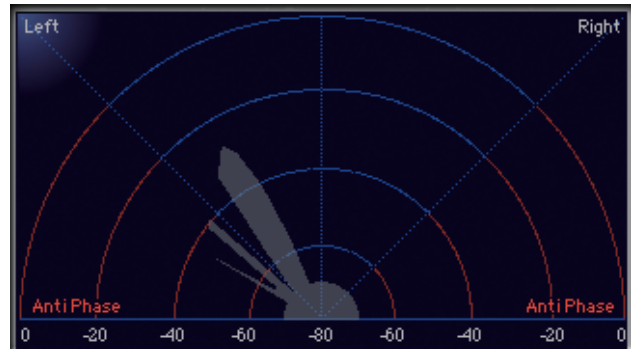


図9：ハイカットで低音域の影響に特化した音源エネルギー

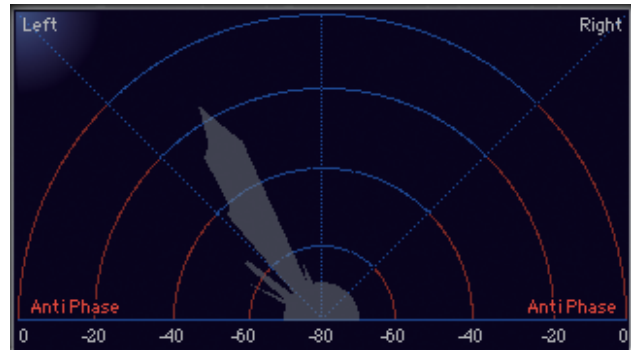


図10：ローカットで高音域の影響に特化した音源エネルギー

4. 結果と考察

VRミックスにおいて、元音にできるだけ影響を与えず音像幅を拡大するためには、2倍倍音の調整が有効だということが分かった。

また、倍音調整の度合いが音質変化の度合いと相関関係にあるため、少ない増加量の調整が望ましいことも分かった。

一方で、2倍倍音以外の周波数に影響を与えると、音質や音階自体が変化してしまうため、元音を担保したVRミックスには有効となりにくいことも分かった。

さらに、音像幅のみに注目してみると、高音域よりも低音域が強く影響を与える結果となったため、より音質等を担保しながら音像幅のみを調整したい場合は、低音域の音源または低音域の倍音を細かく調整することで、精度の高い調整が可能となることが分かった。

5. 課題

今回は、基音に対するオクターブ音となる2倍倍音を活用することで音像幅の調整を行ったが、何倍であれ倍音成分を変化させるということは、音質や音色そのものを変えることにも繋がる。そのため、元音の担保と音質変化に対する補正等の調整を同時に持ってミックスを進めることが現実的であるとも言えるが、特に商業音楽の場合、さらに細かい音質等の調整が求められることから、より高い実用性の向上を図るために、音質変化が限りなく少ない音像幅の調整が求められる。

また、音像幅の調整と定位の分かりやすさは、相関関係になるため、VRミックスは音像幅だけでなく、定位や音質や音量、音圧等の総合的なバランスを調整する必要がある。そのために必要となる調整項目はまだ多数存在することを今後の課題とする。

参考文献

- [1] 坂本昭人：『映像・動画制作者のためのサウンドデザイン入門』株式会社ビー・エヌ・エヌ(2020年), 112-113頁。
- [2] 坂本昭人：『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』DHU JOURNAL 2020(2020年), 40-44頁。
- [3] 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 2019(2019年),
56-61頁

[4]sakamoto akihito: "[VR音響研究]ホワイトノイズを使った2mix
(2D)音源での定位による音の変化"

<https://youtu.be/-XfFz1yJ2z0> (参照2021年8月12日)

[5]sakamoto akihito: "[VR音響研究]ホワイトノイズを使ったVR
ミックス(3D)音源での定位による音の変化"

<https://youtu.be/SBoZcYUatvg> (参照2021年8月12日)

[6]sakamoto akihito: "[VR音響研究]2倍倍音において6dB増加
の後の音の変化"

<https://youtu.be/r7mDu7jkd74> (参照2021年8月12日)

[7]sakamoto akihito: "[VR音響研究]3倍倍音において6dB増加
の後の音の変化"

<https://youtu.be/OQexX6qNgzM> (参照2021年8月12日)

[8]sakamoto akihito: "[VR音響研究]2倍倍音において2dB増加
の後の音の変化"

<https://youtu.be/ROOMFqEmk9Q> (参照2021年8月12日)