

# アンビソニックスマイクによるバンドライブ演奏の多目的リミックス生成

## Multi-purpose remix generation of live band performances using an Ambisonics microphone

三代川凌

Ryo Miyokawa

法政大学情報科学部コンピュータ科学科

ryo.miyokawa.5h@stu.hosei.ac.jp

### Abstract

A method is proposed to generate part-wise listen-back signals and a two-channel remix from a full-ensemble band recording captured by a single Ambisonics microphone (FOA, 4 channels). Part directions are determined from short prior single-part recordings, and each part is emphasized using spherical-harmonic (SH) beamforming. Inter-source leakage and reverberation are then suppressed using a time-frequency-domain multichannel Wiener filter (MWF). The resulting part-wise outputs are adjusted with faders and panning to enable post-recording remixing with a preferred balance. Objective evaluation using SI-SDR [1] on Ambisonics simulations derived from multitrack sources showed improvements of 17.63 dB for Drums and 14.96 dB for Guitar. An improvement of 9.96 dB was obtained for Vocals, while the additional improvement of MWF over SH beamforming was negative for Vocals in this experiment. Piano and Bass showed improvements of 8.14 dB and 2.46 dB, respectively. These results suggested the effectiveness of training-free listen-back support based on on-site direction setting. Future work will investigate evaluation criteria for multi-purpose remixing, including subjective listening tests.

### 1 はじめに

近年、バンド演奏の録音は、自己評価や振り返り（聞き返し）のために広く行われている。スマートフォンやポータブルレコーダーの普及により、小規模ライブや日常練習でも手軽に録音できる一方、マイク 1 本または 2 本による一斉録音を中心であるため、各楽器音が混ざり合い、残響や音量バランスの偏りが生じやすい。その結果、録音を聞き返す際に「自分のパートが埋もれる」「他パートが過度に支配的になる」といった問題が起き、演奏内容の確認や改善点の把握が難しくなる。

この問題に対し、各楽器を個別マイクで収録してマルチトラック編集する方法は有効である。しかし、機材準備や設営に要する時間・コストが大きく、録音後のミキシング環境と技能も必要となるため、日常的な練習や小規模ライブの現場に常に適用することは現実的でない。したがって、少ない機材で一斉録音しつつ、録音後に聞き返しに適した音へ整える手段が求められる。

本研究では、単一のアンビソニックスマイクによる一次アンビソニックス (FOA) 録音を入力として、パート別の聞き返し信号と 2ch リミックスを生成する枠組みを提案する。アンビソニックスは空間情報 (方向) を含む信号として扱えるため [2]、方向手がかりを利用して特定方向の音を相対的に強調し、各パートの明瞭性を高めることが期待できる。本研究の狙いは、厳密な音源分離そのものではなく、(1) 練習を想定して、聞き返したいパートを相対的に前に出す「聞き返し支援」を成立させること、(2) 録音後に音量バランスや左右配置を調整できるように、2ch 出力としての操作可能性を与えること、の 2 点である。

提案手法では、事前に収録した短い単音 (単独パート) 録音を用いて各パート方向を決定し、SH ビームフォーミングにより方向強調を行う。さらに、実録音で問題となりやすい回り込みや残響の影響を抑えるため、時間周波数領域の MWF を適用し、パート別信号を得た後にフェーダーと PAN で 2ch リミックスへ接続する。また、距離・角度条件の観察、実録音スペクトルの比較、シミュレーション音源を用いた SI-SDR 評価により [3]、提案枠組みが聞き返し支援に対してどの程度有効であるかを検討する。

以降、2 章で従来研究を整理し、3 章で提案手法を述べる。4 章で実験条件・評価方法・実験結果・考察を示し、5 章で結論と今後の課題をまとめる。

### 2 従来研究

本章では、一斉録音の聞き返しにおける困難さを整理した上で、アンビソニックス録音が提供する「方向」情報と、方向強調・回り込み抑制の代表的枠組みを概観する。これにより、本研究が「分離」ではなく「聞き返し支援」と「録音後の操作可能性」に焦点を置く理由を明確化する。

#### 2.1 一斉録音における聞き返しの困難さ

バンド練習や小規模ライブの録音では、ステレオ/モノラルの一斉録音が広く用いられている。これらは手軽である一方、マイク設置位置や部屋の音響特性に強く依存し、音量バランスが一定しない。特にドラムやベースなどの低域成分が支配的な場合、他パートが時間周波数領域で埋もれるマスキングが生じやすく、ボーカルやギターの明瞭性が低下する。さらに実環境では壁面反射や残響により音の輪郭や定位が不明瞭になりやすく、一斉録音の単一混合信号から録音後に特定パートを独立に操作することは一般に難しい。

## 2.2 アンビソニックス録音と方向手がかり

アンビソニックス録音は空間全体の音響情報を多チャンネルで記録し、方向情報を含む信号として扱える方式である。市販の一次アンビソニックスマイクでは、一般に4つの単指向性マイク出力(A-format)を取得し、線形結合によりB-formatへ変換して利用する。一次アンビソニックス(FOA)のB-formatは無指向性成分Wと直交3軸方向の一次成分X、Y、Zから構成され、任意方向への指向性合成や再生方式変換が可能となる。実演環境では各パートの音源(アンプやドラムなど)の配置が固定されることが多い。このとき、録音信号内に音源方向の手がかりを保持できることは、特定パートの聞き返し支援やリミックス生成に有効である。

## 2.3 SH ビームフォーミングに基づく強調と限界

SH ビームフォーミングはFOA 信号の線形結合により任意方向を強調でき、深層学習型分離のような非線形推定を伴わないため音質を保ちやすい利点がある。一方で実録音では、反射・残響や他パートの回り込みにより、単純なビーム出力だけでは目的パートが十分に前に出ない場合がある。またFOA は次数が低いため空間分解能に制約があり、音源間角度差が小さい条件では強調が困難になりやすい。

## 2.4 MWF による回り込み抑制(時間周波数後処理)

SH ビーム出力に残る他パート混入や残響成分を抑制するため、時間周波数領域の多チャンネルWiener フィルタ(MWF) [4]を適用する。観測信号の統計量(PSD/共分散)に基づいて周波数ごとの線形フィルタを設計し、目的成分を推定した後、逆STFTにより時間波形へ戻して各パート出力とする。

## 3 方位情報に基づくパート別強調・リミックス生成

本章では、単一のFOA 録音からパート別の聞き返し信号を生成し、最終的に2ch リミックスとして出力するまでの処理フローを述べる。本研究の要点は、方向推定を含む運用を明確にし、線形な方向強調に統計的抑制を組み合わせることで、実環境の回り込み・残響に対しても「聞き返しやすさ」を狙う点にある。

### 3.1 録音環境と処理フロー

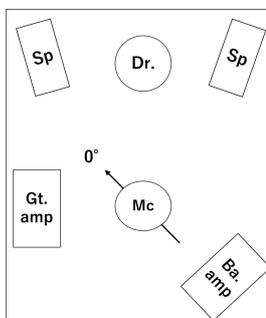


Fig.1: 録音環境(例): Zoom H3-VR を中心に配置したバンド練習の収録

本研究では、Zoom H3-VR を用いてアンビソニックス形式の4チャンネル収録を行う。録音は主として屋内のバンド練習スタジオで実施し、アンビソニックスマイクは図1のように演奏

空間の中央付近に設置する。各パート(Vo、Gt、Ba、Dr、Keyなど)がマイク周囲に分布する状況を想定し、単一のアンビソニックス収録からパート別の聞き返しが可能な信号生成を目指す。

収録信号はFOA として扱い、A-format(単指向性マイク4ch)からB-format(W、Y、Z、X)へ変換した後、指向性合成・回り込み抑制・リミックスに用いる。図2に、FOA 入力(4ch)から2ch リミックスへ至る処理系列を示す。ここでKは対象パート数(同時に生成するパート別出力の本数)であり、各パートについて方向( $\theta, \phi$ )を与えてSH ビームを1本生成する。続いてMWFでは、各パートのSH ビーム出力を参照(目的成分の手がかり)として共分散を推定し、時間周波数領域で回り込み・残響を抑制したパート別出力を得る。MWF後はパートごとにモノラル信号として取り出し、フェーダとPANで2chへ合成して書き出す。



Fig.2: 提案手法の全体フロー

### 3.2 パート方向の推定・設定

本研究では、各パート方向(方位角 $\theta$ 、仰角 $\phi$ )を事前に与えてビームを形成する。方向は、(i)収録時の配置情報に基づく手動設定、または(ii)FOA 信号に基づく推定で決定する。

(ii)では、本番前に各パートが単音(または短いフレーズ)を鳴らす事前録音を用いる。GUIで当該パートが鳴っている区間を選択し、粗探索→細探索の2段階で方向探索を行う。各候補方向は、選択区間のビーム出力を短時間フレームに分割してRMSを算出し、無音の影響を避けるためRMS上位20%の代表値で評価する。

推定方向に加えて周辺角度のビームも試聴でき、妥当性確認と微調整を可能にする。最終的に、パート名と角度の対応を設定として書き出し、録音ごとの差を反映しつつ再現可能な条件を構成する。図3に探索結果(Stage1 best、FINAL best)と試聴方向(PLAY)の可視化例を示す。

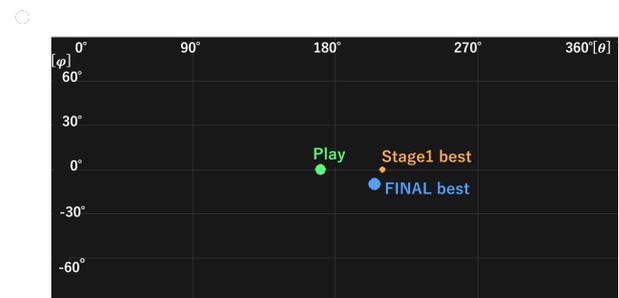


Fig.3: 方向推定の可視化例: Stage1 best(粗探索)、FINAL best(細探索)、PLAY(試聴方向)

### 3.3 SH ビームフォーミングによるパート別強調

方向が与えられた各パートに対して、FOAの球面調和表現に基づくビームフォーミングを適用し、目的方向の音を強調し

た信号（パート別ビーム）を得る。FOA では B-format (W、Y、Z、X) の線形結合により任意方向の指向性を合成できる。本研究では、指向性の鋭さと音質の自然さのバランスを考慮し、MaxRe [5, 6] や MaxDI [7] の指向性設計を用いてビームを形成する。

一方で実録音では反射音・残響・他パートの回り込みにより、ビーム出力だけでは目的パートが十分に前に出ない場合がある。そこで次節で述べる MWF により回り込み抑制を行い、目的方向の強調結果を保ちながら不要成分を追加で抑えることで、聞き返し用途に適したパート別信号を生成する。

### 3.4 MWF による回り込み抑制（時間周波数後処理）

本研究では、SH ビーム出力に残存する他パート混入や残響成分を抑制するため、時間周波数領域の多チャンネル Wiener フィルタ (MWF) [4] を適用する。FOA 観測の STFT ベクトルを  $\mathbf{x}(f, t) \in \mathbb{C}^4$  とし、各パート  $k$  の参照ビームを  $y_k(f, t) = \mathbf{a}_k^H \mathbf{x}(f, t)$  で得る。参照ビーム群からソフトマスク

$$m_k(f, t) = \frac{|y_k(f, t)|^p}{\sum_{j=1}^K |y_j(f, t)|^p} \quad (1)$$

を構成し、 $m_k$  で重み付けした外積  $\mathbf{x}\mathbf{x}^H$  の指数平均（係数  $\alpha$ ）により、目的成分共分散  $\mathbf{R}_{s,k}(f)$  と観測共分散  $\mathbf{R}_{x,k}(f)$  を推定する。数値安定化のため対角負荷  $\lambda \mathbf{I}$  を加え、W 成分参照  $\mathbf{e}_{\text{ref}} = [1, 0, 0, 0]^T$  として

$$\mathbf{w}_k(f) = \mathbf{R}_{x,k}(f)^{-1} \mathbf{R}_{s,k}(f) \mathbf{e}_{\text{ref}} \quad (2)$$

を求める。出力は  $\hat{s}_k(f, t) = \mathbf{w}_k^H(f) \mathbf{x}(f, t)$  とし、逆 STFT で時間波形へ戻して各パートの聞き返し信号とする。

### 3.5 リミックス生成 (MIX UI) と出力

得られたパート別出力 (SH ビーム + MWF 後) を入力として、各パートのフェーダー（音量）と PAN を独立に操作し、録音後に 2ch リミックスを生成する。図 4 に、パート方向マップと連動した方向操作、パート別フェーダー / PAN (Mute / Solo) を含む)、再生操作、および現在設定でのステレオ書き出しを統合した UI 例を示す。本 UI により、「自分のパートを前に出す」「支配的なパートを少し下げる」といった聞き返しに直結するバランス調整を録音後に行える。最終出力はステレオミックスとして書き出し、練習共有や振り返りに利用する。また、事前単音区間の選択から方向決定、本番録音への適用、フェーダー / PAN 調整、2ch 書き出しまでの一連の運用を同一環境で実行できる。

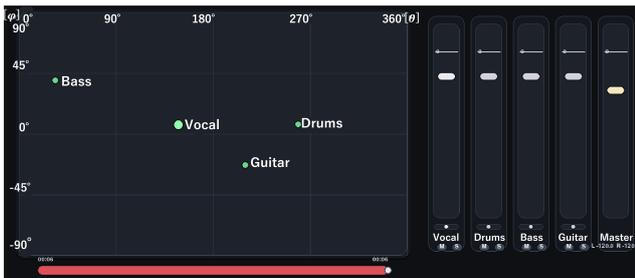


Fig.4: リミックス UI 例：パート方向マップと連動した方向操作、パート別フェーダー / PAN (Mute / Solo)、再生操作、およびステレオ書き出し

## 4 評価

本章では、提案法の有効性を (i) 幾何条件の成立性、(ii) 周波数領域での傾向確認、(iii) 客観指標による比較の 3 観点から評価する。まず実環境における距離・角度条件の変化に対し、SH ビームフォーミングによる方向強調が得られやすい条件の傾向を確認する。次に実録音のスペクトル比較により、強調および抑制の出現帯域を観察する。最後にマルチトラック音源から作成したアンビソニックス・シミュレーション音源を用い、SI-SDR により Baseline (Mix)、SHbeam、MWF の差を比較する。

### 4.1 実環境計測に基づくシミュレーション配置条件

シミュレーション評価に用いる空間および配置条件は、実環境で実施した距離・角度条件の検証結果を参考に設定する。部屋サイズは  $(x, y, z) = (6, 4, 3)$  m とし、マイク（受音点）は中心  $(3, 2, 1.5)$  m に配置する。音源間角度差については、実環境の検証において方向強調が安定して観察されやすかった条件を踏まえ、下限を  $50^\circ$  として統一する。また、各音源の距離は一律に固定せず、想定配置に応じて設定する。以上により、提案法の性能を比較するための評価条件を構成する。

### 4.2 評価項目

評価は次の 3 点から構成する。(1) 幾何条件の成立性として、音源・マイク間距離および音源間角度差に対し、方向強調がどの程度得られるかの傾向を確認する。(2) 周波数領域の分析として、パート強調により目的成分（基音・倍音）が相対的に目立つ帯域、ならびに MWF 適用により抑制が出る帯域をスペクトルで観察する。(3) 客観評価として、シミュレーション音源から切り出した区間を用い、Mix、SHbeam、MWF の性能差を SI-SDR で評価する。

### 4.3 スペクトル解析条件

スペクトル比較では対象区間を揃え、参照信号と処理後信号を同一条件で比較する。本稿で用いた条件の一例は、窓長 0.1 s、窓関数はハンギング窓、FFT 点数は 8,192 点である。表示周波数上限は図に応じて設定し（例：2,048 Hz または 4,096 Hz）、縦軸は dB スケールとし表示範囲は図ごとに固定する。

### 4.4 評価設計と新規性の位置づけ

本稿では 5 音源 ( $K = 5$ ) を対象として評価する。提案法は「方向ごとにビームを構成し、パート数に応じてビーム本数を増やす」枠組みであるため、音源間角度差が確保される条件では  $K > 5$  への拡張も想定される。ただし FOA の空間分解能や残響の影響により限界があるため、 $K > 5$  の成立条件は今後の評価課題として整理する。

### 4.5 SI-SDR の評価データと条件

正解信号は各パートのマルチトラック音源とする。対象パートは Bass、Guitar、Drums、Vocals、Piano の 5 種類である。比較対象は Baseline (Mix)、SHbeam、MWF の 3 種類とした。全パートが同時に発音している区間を選定し、その範囲内から長さ 2 秒の評価区間を一様乱数で 10 回抽出し、各試行で

SI-SDR を算出して平均と標準偏差を求めた ( $N = 10$ 、窓長 2 秒)。

#### 4.6 時間軸整列

推定信号と正解信号の間には処理に起因する遅延が生じうる。そこで相互相関により遅延(ラグ)を推定し、探索範囲を  $\pm 100$  ms に制限した。これは過大な探索範囲が誤整列を招きやすいためである。そして、時間軸を整列してから SI-SDR を計算した。

#### 4.7 実験結果

##### 4.7.1 最小距離の測定

2 台のスピーカを  $90^\circ$  に配置し、マイクとの距離を 2.0 m から 0.5 m まで段階的に変化させて録音した。実験配置を Fig. 5 に示す。取得信号に対して SH ビームフォーミングを適用し、目的音源方向へビームを指向させて方向強調の傾向を確認した。

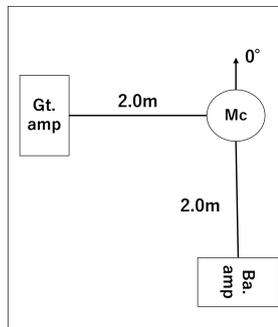


Fig.5: 最小距離を測定するための録音配置図

距離が近い条件では直接音が相対的に支配的になり、方向強調が得られやすい傾向が観察された。一方で近接配置では入力レベルや配置誤差の影響を受けやすく、条件によっては強調が不安定となる場合も見られた。

##### 4.7.2 最小角度の測定

半径 2.0 m の円弧上でスピーカ同士の角度を  $90^\circ$  から  $20^\circ$  まで変化させて録音した。実験配置を Fig. 6 に示す。各角度条件において SH ビームフォーミングにより音源方向へビームを形成し、方向強調の傾向を確認した。

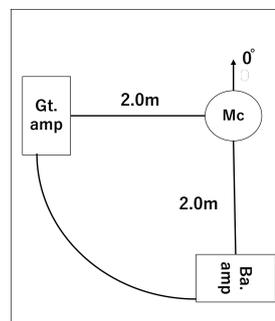


Fig.6: 最小角度を測定するための録音配置図

角度差が大きい条件ほど目的音が相対的に強調され、角度差が小さい条件では強調が困難となる傾向が観察された。ただし角度差が小さい場合でも、ビーム方向を音源方向から意図的にずらすことで目的外音が減衰し、相対的に狙いたい音が目立つ場合があった。これは FOA のビーム幅やサイドローブ特性の

影響を受けるため、厳密な分離というより「聞き返ししやすい方向付け」の調整として位置づけられる。

##### 4.7.3 距離・角度条件のまとめ

屋内環境でも、距離・角度条件によって方向強調の得られやすさが変化する傾向が確認された。角度差が小さい条件では強調が困難になりやすく、手法比較が不安定になり得るため、本稿の以降の評価では角度条件を統一し、シミュレーション配置の音源間角度差の下限を  $50^\circ$  とする。一方で距離は方向強調の成立性に影響する要因であるため、固定せずに変化させて傾向を確認する。

##### 4.7.4 実環境収録データを用いたスペクトル解析

本節では、実環境における SH ビームフォーミングおよび MWF の挙動を確認するため、実際に収録したバンド演奏データを用いて周波数領域で比較を行う。ここでは低域成分が支配的であり、他パートとの干渉が課題となるベースパートを対象として解析した。

##### 4.7.5 SH ビームフォーミングによる強調効果

図 7 に、FOA の W 成分(全指向性、参照)と、ベース方向へ指向させた SH ビーム出力(処理後)の振幅スペクトル(横軸: 周波数、縦軸: dB)を同一条件で比較した例を示す。

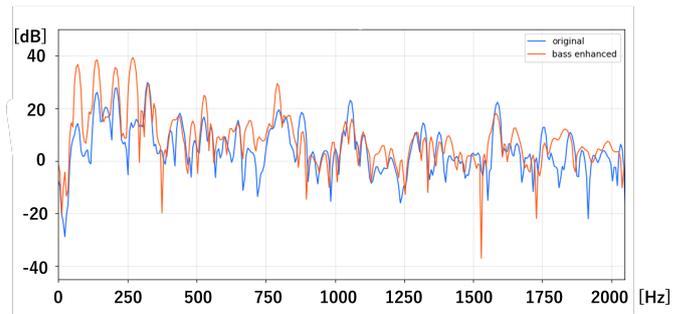


Fig.7: SH ビームフォーミングによるベース強調信号と参照信号のスペクトル比較例

参照信号では他パートの干渉により判別が難しかったベースの基音および低次倍音(100 ~ 400 Hz 付近)が、強調信号では相対的に目立つ傾向が観察された。これは方向情報に基づく空間フィルタリングにより、被り音が低減した可能性を示唆する。

##### 4.7.6 MWF による残留干渉成分の抑制

図 8 に、同一時間区間における SH ビーム出力(処理前)と MWF 適用後出力(処理後)の振幅スペクトル(横軸: 周波数、縦軸: dB)を比較した例を示す。これにより、MWF が追加で抑制した帯域傾向を観察する。

低域(約 500 Hz 以下)では振幅が概ね維持され、高域(1 kHz 以上)ではレベル低下が見られた。この傾向は、不要成分の抑制と同時に目的音の高域成分まで抑制された可能性も含むため、聴感上の自然さとのトレードオフを考慮して解釈する必要がある。

##### 4.7.7 シミュレーション音源に対する SI-SDR

本節ではマルチトラック音源から作成したシミュレーション音源を用いて、提案処理(SH ビームフォーミング、MWF)の効果を SI-SDR により評価する。Baseline はシミュレーションで生成した FOA 混合音源とし、同一区間に対して SHbeam および MWF を適用した音源を比較する。

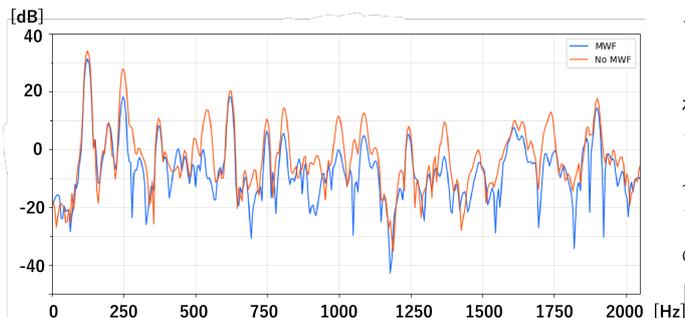


Fig.8: ベース強調信号に対する MWF 適用有無のスペクトル比較例

Table 1: SI-SDR 結果。ランダム 2 秒区間 ( $N = 10$ ) の平均 ± 標準偏差 [dB] と改善量 ( $\Delta$ )。太字は、各音源について  $\Delta(\text{MWF} - \text{SH})$  が正の値であることを示す。

Source	Baseline(Mix)	$\Delta\text{SHbeam}$	$\Delta\text{MWF}$	$\Delta(\text{MWF} - \text{SH})$
Vocals	$-11.92 \pm 3.58$	$10.52 \pm 2.90$	$9.96 \pm 3.77$	$-0.56 \pm 1.03$
Drums	$-13.55 \pm 1.84$	$15.53 \pm 1.08$	$17.63 \pm 2.10$	<b><math>2.11 \pm 1.11</math></b>
Bass	$-2.17 \pm 1.88$	$-0.44 \pm 0.45$	$2.46 \pm 1.51$	<b><math>2.90 \pm 1.60</math></b>
Piano	$-19.44 \pm 1.58$	$5.16 \pm 1.08$	$8.14 \pm 1.63$	<b><math>2.98 \pm 0.86</math></b>
Guitar	$-18.92 \pm 0.96$	$10.08 \pm 1.10$	$14.96 \pm 1.80$	<b><math>4.88 \pm 1.03</math></b>

Table 2: 各楽器音源の入力 RMS(Root Mean Square) とマイクからの距離。RMS は距離減衰前の値、距離はマイク位置からの直線距離を示す。

Source	Input RMS	Distance [m]
Vocals	0.102	3.33
Drums	0.148	1.96
Bass	0.0896	1.25
Piano	0.0393	1.14
Guitar	0.0587	1.04

#### 4.7.8 SI-SDR による客観評価結果

表 1 に時間軸整列後の SI-SDR 評価結果を示す。 $\Delta\text{SHbeam}$  は SH ビームフォーミング単体による改善量、 $\Delta\text{MWF}$  は MWF 適用後の改善量、 $\Delta(\text{MWF} - \text{SH})$  は MWF が SH ビームフォーミングに対して追加で与えた改善量を表す。

Vocals, Drums, Guitar では Baseline に対して 10 dB 以上の改善が確認された。Bass は SHbeam 適用時に  $-0.44$  dB とわずかに悪化した一方、MWF 適用により  $+2.90$  dB の追加改善が得られ、最終的に正の改善幅となった。Piano は SHbeam で  $+5.16$  dB の改善であったが、MWF により追加で  $+2.98$  dB の改善が得られ、合計で  $+8.14$  dB となった。なお Vocals では、MWF が SHbeam に対して追加改善が負 ( $-0.56$  dB) となった。

#### 4.7.9 多目的リミックスの操作性と特徴

提案法では、ビーム + MWF 後のパート別出力に対してフェーダーと PAN を適用し、録音後に 2ch リミックスを生成する。ただし、FOA の空間分解能の制約や残響・回り込みにより、各パート出力は完全に独立ではなく、他パート成分が残留しうる。そのため、特定パート以外を完全に無音にするような「完全分離に近い操作」や、ごく細かな音量調整の自由度は限定

される。

一方で、聞き返し支援として必要とされる範囲の操作、すなわち「Bass と Drums を相対的に前へ出す」「Vocals を最も大きくする」「支配的なパートを少し下げる」といった粗いバランス調整は、パート別信号の生成とフェーダー操作により実現できる。このように、本稿の枠組みは完全な自由度を目標とするのではなく、演奏の確認や振り返りに必要な可聴性の確保を主目的として有効であると位置づけられる。

## 4.8 考察

### 4.8.1 目標の再確認

本研究の目標は、単一の FOA 録音からパート別の聞き返し信号を生成し、(1) 練習を想定した聞き返し支援として、各パートを明瞭に聞ける状態を得ること、(2) 2ch 出力として、各パートの音量バランスを調整可能なリミックスを生成すること、の 2 点である。以下では、それぞれの到達度と限界を整理する。

### 4.8.2 聞き返し支援としての到達度と配置条件

表 1 より、Vocals, Drums, Guitar では Baseline に対して 10 dB 以上の改善が確認され、目的パートを明瞭に確認しやすい状態が得られている。Piano は SHbeam 単体では約 5 dB ( $\Delta\text{SHbeam} = +5.16$  dB) の改善にとどまったが、MWF 適用により合計で約 8 dB ( $\Delta\text{MWF} = +8.14$  dB) まで改善した。Bass は SHbeam 適用時にわずかに悪化した一方、MWF 適用により正の改善 ( $\Delta\text{MWF} = +2.46$  dB) が得られ、低域が支配的な条件でも一定の聞き返し支援が成立することが示された。また、入力 RMS が最小である条件 (表 2) でも正の改善が得られており、音量が出にくいパートに対しても一定の聞き返し支援が成立することが示された。

この達成に寄与した要因として、音量が出にくいパートがマイク近傍に配置されている点が挙げられる。表 2 では、Piano は入力 RMS が小さい一方で距離が 1.14 m と近く、距離減衰の影響を相対的に受けにくい条件であった。一方、音量が出やすい Drums は距離が 1.96 m であり、結果として目的パートの相対的な音量が確保され、方向強調が得られやすかった可能性がある。また、距離・角度に関する実験から、角度差が小さい条件では強調が困難となる傾向が確認されたため、実運用では音源間角度差を確保することを前提とし、音量が出にくい楽器を近づけ、音量が出やすい楽器 (例: Drums) を相対的に離す配置が有効と考えられる。

### 4.8.3 2ch リミックス生成としての到達度と課題

2ch リミックス生成の理想は、不自然な音を生じさせずに各パートの音量を自由に変更できることである。しかし本稿の条件では、各パート出力に回り込み成分が残留しうるため、特定パート以外を完全に無音にするような「完全に独立な操作」や、ごく細かなバランス調整の自由度には限界がある。

一方で、聞き返し用途で想定する粗い操作、すなわち「Bass と Drums を相対的に前へ出す」「Vocals を最も大きくする」「支配的なパートを少し下げる」といったバランス調整は、パート別信号とフェーダー / PAN 操作により実現できる。したがって、本研究の枠組みは、完全な分離を前提とした制作用途のミキシングというより、演奏の確認や振り返りに必要な可聴性を確保するためのリミックスとして有効であると解釈できる。

音量操作の自由度をさらに高めるためには回り込み（他パート混入）をより強く抑制する必要がある一方、抑制を強めるほど目的音の帯域まで削られ、アーティファクトが生じやすくなる。実環境スペクトル（図 8）では、MWF 適用により高域（1 kHz 以上）でレベル低下が観察され、不要成分の抑制と同時に目的音の高域成分まで抑制された可能性がある。また Vocals では、MWF が SHbeam に対して追加改善が負となる条件が見られ、抑制のかけ方がパートにより有利不利を持つ可能性が示唆される。したがって今後は、単に抑制を強めるのではなく、高域の過抑制を避ける周波数依存の制御などにより、回り込み抑制と聴感上の自然さの両立を図る必要がある。また、用途に応じて SH 強調と MWF 抑制の比率を調整できる運用とすることで、聞き返し支援（明瞭さ優先）とリミックス生成（自然さ優先）の要求差に対応しやすくなると考えられる。

#### 4.8.4 音源数の拡張性

本実験では、4ch の FOA を用いて 5 音源の混在下で実用的なパート別強調を確認した。各パートが破綻なく強調されたことから、空間的な角度差が確保できれば、6 音源以上の構成でも適用余地があると考えられる。一方で、FOA は空間分解能に制約があるため、音源間角度差が小さい条件では回り込みが増え、パート別出力の独立性が低下しやすい。したがって、音源数を増やす場合は、配置（角度差の確保）や目的に応じた抑音量（MWF 設定）を含む運用条件の設計が重要となる。

## 5 結論

本研究では、単一のアンビソニックスマイクによる一次アンビソニクス（FOA）一斉録音から、パート別の聞き返し信号と 2ch リミックスを生成する枠組みを提案した。提案法は、(i) 事前の短い単音録音に基づく方向設定、(ii) SH ビームフォーミングによる方向強調、(iii) MWF による回り込み・残響の追加抑制、(iv) フェーダーと PAN による 2ch リミックス生成、を直列に接続し、多マイク収録や学習ベース分離に依存せずに録音後の聞き返し調整を可能とする点に特徴がある。

実験では、距離・角度条件により方向強調の得られやすさが変化する傾向が確認され、角度差が小さい条件では強調が困難になりやすいことを観察した。また、実録音スペクトルの比較では、SH 処理により目的成分が相対的に目立つ帯域が見られ、MWF 適用時には高域でレベル低下が観察されるなど、抑制と音質のトレードオフが示唆された。さらに、マルチトラックから作成したアンビソニクス・シミュレーション音源に対する SI-SDR 評価では、Drums で 17.63 dB、Guitar で 14.96 dB、Vocals で 9.96 dB の改善が確認された。Piano では 8.14 dB、Bass では 2.46 dB の改善が得られた。ただし Vocals では、MWF が SHbeam に対して追加改善が負となる条件が見られた。

以上より、聞き返し支援という目的に対しては、複数パートで明瞭性の改善が確認され、方向設定に基づく学習不要な処理として有効である可能性が示された。また、2ch リミックス生成については、特定パート以外を完全に無音にするような自由度は難しい一方で、聞き返し用途で想定する粗いバランス調整（例：Bass / Drums を前に出す、Vocals を最も大きくする等）は実現できることを示した。一方、回り込み抑制を強めるほど

アーティファクトが生じやすいことから、より細かな操作自由度の達成には、抑制と音質の両立が課題である。したがって、実運用では音源間角度差を確保しつつ、音量が出にくい楽器を相対的にマイクへ近づけ、音量が出やすい楽器を相対的に離すなど、録音条件の工夫と併用することが有効と考えられる。

今後の課題として、(1) 多目的リミックス生成に対する評価軸の整理（主観聴取試験を含む）、(2) 回り込み抑制と聴感上の自然さを両立するための MWF 制御（高域の過抑制回避など）、(3) 音源数の拡張（6 音源以上を含む）における成立条件の検証（角度差の確保や配置条件の整理）、を挙げる。

## 参考文献

- [1] Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R. Hershey. Sdr – half-baked or well done? pp. 626–630, May 2019.
- [2] Jérôme Daniel, Rozenn Nicol, and Sébastien Moreau. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. No. 5788, March 2003.
- [3] Francesc Lluís, Nils Meyer-Kahlen, Vasileios Chatziioannou, and Alex Hofmann. Direction specific ambisonics source separation with end-to-end deep learning. *Acta Acustica*, Vol. 7, p. 29, 2023.
- [4] Mehrez Souden, Jacob Benesty, and Sofiène Affes. On optimal frequency-domain multichannel linear filtering for noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 2, pp. 260–276, 2010.
- [5] Daniel P. Jarrett, Emanuël A. P. Habets, and Patrick A. Naylor. *Theory and Applications of Spherical Microphone Array Processing*, Vol. 9 of *Springer Topics in Signal Processing*. Springer Cham, Cham, 1 edition, 2017.
- [6] Boaz Rafaely. *Fundamentals of Spherical Array Processing*, Vol. 8 of *Springer Topics in Signal Processing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1 edition, 2015.
- [7] Franz Zotter and Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Vol. 19 of *Springer Topics in Signal Processing*. Springer Cham, Cham, 1 edition, 2019.