

弦楽三重奏に対するステレオ音楽音響信号の周波数特性に着目した自動ミキシング

三輪 明宏 守田 了

山口大学工学部

宇部市常盤台 2557

本稿では音楽音響信号の周波数特性に着目し、自動ミキシングを行う。ミキシング後の理想とするステレオ音楽音響信号を指定すると、各楽器の音楽音響信号の周波数特性が理想とする音楽音響信号の周波数特性に近づくように、ミキシングを行うためのパラメータを決定する。ミキシングを行うためのパラメータとしてヴォリュームとパニングのみを使用する。まず各楽器の音楽音響信号の周波数特性とミキシング後の理想とするステレオ音楽音響信号の周波数特性を求める。周波数領域で各楽器のヴォリュームおよびパニングの値を最急降下法を用いて最適化を行う。

ミキシング前の個々の音響信号が得られる場合について、ミキシングの基本的な性能を評価した後に、ミキシング前の個々の音響信号を得ることが出来ない場合についてミキシングのパラメータを推定する。実際に、市販されているベートーベンの弦楽三重奏のCDに対して、我々が録音した各楽器の音響信号を自動ミキシングした結果、CDで使用されている各楽器の位置関係を再現することが出来た。

自動ミキシング, ステレオ音楽音響信号, 三重奏

Automatic Mixing based on Frequency Characteristics of Stereo Music Signal for a String Trio

Akiniro MIWA and Satoru MORITA

Faculty of Engineering, Yamaguchi University

2557 Tokiwadai, Ube, 755, Japan

In order to make the performance CD, we need to mix each audio signal with taking volume and panning into consideration. It is important to hear sounds of the important instrument that compose music clearly. Concretely, in order to interfere neither one frequency band nor another frequency band, we adjusted the parameter of the volume and the panning. It is difficult to adjust these parameters because we need severe training to hear the frequency band. Generally, the mixing has been done by mixing engineer's experience and intuition.

In this paper, we proposed a method of the giving frequency characteristic from CD to estimate the volume and the panning parameters automatically. local minimum. We shows the effectiveness of our system when we can get the original audio signal before mixing. Furthermore we shows the effectiveness of our system when we can not get the original audio signal before mixing, nor when we do not know these parameters of the volume and the panning. In this case, we substitute the original audio signal by the audio signal we record the performance based on the score. This experiment shows that we can estimate the parameter of the panning in the same order as CD music.

keyword: Automatic Mixing, Frequency Characteristics, Stereo Sound Signal

1 まえがき

近年、コンピュータの高性能化、音響機器の低価格化などの影響で家庭でも手軽にミキシングを行うことが出来るようになってきた。譜面を元に人間が演奏するだけでなく、MIDI 楽器などで演奏を行い、デジタル録音を行うことで高音質の音響信号を手軽に作成することが出来るようになってきた。このようにして得られた音響信号を CD にする場合、各楽器の演奏のヴォリュームやパンニングを考慮してミキシングする必要がある。ミキシングの際には楽曲を構成する重要な各楽器の音がバランス良く分離して聴こえる必要がある。具体的には、ひとつのパートの音を構成する周波数帯域が他のパートの周波数帯域と重ならないようにヴォリュームやイコライザのパラメタを調整し、それでも周波数帯域が重なっている場合はパンニングを用いて周波数帯域の重なりを回避する等の作業を行う。これらの作業を人間が効率的に行うためには、周波数帯域を聞き分けるための訓練が必要であるためミキシング作業は一般的に困難である。そのため、たとえ市販の CD の譜面を得ることが出来て、個々の楽器の音響信号を作成することが出来たとしても、CD のミキシングを再現することは非常に困難である。従来、このミキシング作業はミキシングエンジニアの経験と勘によって行われてきた [1]。

本稿では弦楽三重奏の市販の CD に対して、演奏の譜面しか得られない状況において、音響信号を独自に作成し、ミキシングを再現することを目的とする。本稿ではミキサの構成要素としてヴォリューム、パンニングのみを使用する。

まずミキシング前の個々の音響信号が得られる場合について実験を行う。3 楽器までの演奏についてさまざまなパラメタを用いて音響信号をミキシングする。ミキシングした音響信号のパラメタを推定することで自動ミキシングの基本的な性能を評価する。次に、ミキシング前の個々の音響信号を得ることが出来ない場合について実験を行うことで、弦楽三重奏の市販 CD に対する自動ミキシングの性能を評価する。

2. では自動ミキシングの処理の流れについて述べ、3. では弦楽三重奏に対する自動ミキシングを評価する。

2 自動ミキシングの処理の流れ

時刻 t での i 番目の音楽音響信号を $f_i(t)$ と表す。これらに演算を加えずにミキシングした音楽音響信号は $\sum_{i=1}^n f_i(t)$ となる。ここでミキシングの際になんらかの演算 $X(f_i(t))$ を加えた場合を考える。演算 X が線形で

あれば次式で記述出来る。

$$X\left(\sum_{i=1}^n f_i(t)\right) = \sum_{i=1}^n X(f_i(t)) \quad (1)$$

さらに、式 (1) のフーリエ変換はフーリエ変換の線形性を用いて次式で記述出来る。

$$F\left\{X\left(\sum_{i=1}^n f_i(t)\right)\right\} = \sum_{i=1}^n X\{F(f_i(t))\} \quad (2)$$

ただし、

$$F(f_i(t)) = \int_{-\infty}^{\infty} f_i(t)e^{-j\omega t} dt \quad (3)$$

このことから、ミキシングの際に線形演算を行うなら時間領域でミキシングする代わりに周波数領域でミキシングできることがわかる。

各周波数のパワーをヴォリュームとみなす。元のヴォリュームを $(V_L, V_R)^T$ 、演算後のヴォリュームを $(V'_L, V'_R)^T$ とし、演算行列を M とし、以下のように記述する。

$$\begin{pmatrix} V'_L \\ V'_R \end{pmatrix} = M \begin{pmatrix} V_L \\ V_R \end{pmatrix} \quad (4)$$

本稿ではヴォリューム演算とパンニング演算を行う。

ヴォリューム演算は元のヴォリュームを a 倍する演算である。演算行列を次式で定義する。

$$M = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \quad (5)$$

パンニング演算はステレオ音楽音響信号の定位を変更する演算である。パンニング演算には、左右の音量を線形に分配する方法 [2]、sin 曲線を用いて分配する方法 [?] 等が提案されている。本稿では [2] の方法を採用する。定位 P を $-100 \leq P \leq 100$ の範囲で定義し、 $P = -100$ のとき左、 $P = 0$ のとき中央、 $P = 100$ のとき右とする。

演算行列 M を次式で定義する。

$$M = \begin{pmatrix} LL & RL \\ LR & RR \end{pmatrix} \quad (6)$$

ここで、 LL, RL, LR, RR は分配を行うためのパラメタを表し、 P を用いて次式のように決定する。 $-100 \leq P \leq 0$ のとき、

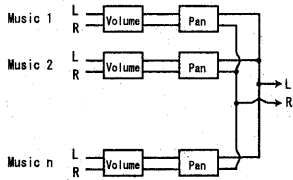


図 1: ミキサの構成

Fig. 1: The composition of the mixer.

$$LR = 0 \quad (7)$$

$$RR = \frac{100 - \|P\|}{100} \quad (8)$$

$0 < P \leq 100$ のとき、

$$RL = 0 \quad (9)$$

$$LL = \frac{100 - \|P\|}{100} \quad (10)$$

演算後の V_L, V_R それぞれの音量の合計は変わらないため次式が成立する。

$$LL + LR = 1 \quad (11)$$

$$RL + RR = 1 \quad (12)$$

音響信号はサンプリング周波数 44.1 kHz、16 bit ステレオで AD 変換を行う。その後、矩形窓を用いた FFT によりパワースペクトラムに変換する。矩形窓を用いたのは、周波数領域でミキシングを行った後にパワースペクトラムを時間領域に復元する必要があるためである。FFT の観測区間は 2048 点とする。この区間内には約 46.4ms の音響信号が含まれる。楽器 i 、周波数 f での左右のパワースペクトラムをそれぞれ $V_{i,L}(f)$ 、 $V_{i,R}(f)$ とする。

楽器数を n とするとき各楽器の音響信号の接続を図 1 に示す。L には左チャンネルのパワー $V_{i,L}(f)$ および位相を入力し、R には右チャンネルのパワー $V_{i,R}(f)$ と位相を入力する。各入力のパワーに対してヴォリューム演算を行った後、パニング演算を行う。全ての楽器について別々に演算を行い、各楽器に対する演算結果の和をとり、ミキシング波形のパワースペクトラムをミキサの出力とする。左チャンネルのミキサの出力を $\tilde{V}_L(f)$ 、右チャンネルのミキサの出力を $\tilde{V}_R(f)$ とする。 $\tilde{V}_L(f)$ および

$\tilde{V}_R(f)$ がミキシング後の周波数特性を表す。目的とする楽曲の左右の周波数特性を $V_L(f)$ 、 $V_R(f)$ とする。本稿での自動ミキシングとは次式で示される評価関数 z を最小化する、各音響信号に対するヴォリューム演算とパニング演算のパラメタを探ることである。探索するパラメタは楽器数が n のときに $2n$ 個ある。

$$z = \sum_{i=F_{min}}^{F_{max}} \left\{ \|V_L(i) - \tilde{V}_L(i)\| + \|V_R(i) - \tilde{V}_R(i)\| \right\} \quad (13)$$

ただし、 F_{min} および F_{max} は自動ミキシングを行う際に考慮する周波数の下限と上限を示す。本稿では $F_{min} = 10\text{Hz}$ 、 $F_{max} = 11.05\text{kHz}$ とした。式 (13) において加算の対象となるデータ数は、FFT の観測区間を $FFTsize$ として表すと次式で記述できる。

$$DataNumber = \frac{2 \times FFTsize \times F_{max}}{44.1} \quad (14)$$

本稿では最適化の手法として最急降下法 [3] を用いる。解の初期値をランダムに 4 点選び、評価関数を最小とする解を最適解とする。

3 自動ミキシング実験

本稿で提案する手法を用いて自動ミキシングを試みる。

3.1 パラメタが既知の場合

バイオリン、ヴィオラおよびチェロの 3 楽器をミキサの入力とする。ミキサの入力と出力から各演算のパラメタを推定する。

実験結果を図 2(a) ~ (f) に示す。横軸はパニング、縦軸はヴォリュームを表す。図 2(a) はヴァイオリンの音響信号を中央に定位させ、ヴォリュームを $-20 \sim 10[\text{dB}]$ まで $0.1[\text{dB}]$ 刻みで変化させたときのヴォリュームおよびパニングの推定結果を表している。図 2(b) はヴァイオリンの音響信号のヴォリュームを $-3[\text{dB}]$ に固定し、定位を $-100 \sim +100$ まで 1 刻みで変化させたときの推定結果を表している。推定する楽器が 1 つの場合はヴォリューム、パニング共に正確に推定することが出来た。

推定する楽器数を 2 および 3 に増やして同様の実験を行う。図 2(c) はヴィオラを $(-3, 0)$ 、チェロを $(-3, 50)$ に、ヴァイオリンのパニングを -50 に固定し、ヴァイオリンのヴォリュームを $-20 \sim 10[\text{dB}]$ まで $0.1[\text{dB}]$ 刻みで変化させる。このときの推定結果を図 2(e) に示す。

同様に図 2(d) はヴィオラを (-6, 0)、チェロを (-6, 50)、ヴァイオリンのヴォリュームを $-3[\text{dB}]$ に固定し、ヴァイオリンのパニングを $-100 \sim 100$ まで 1 刻みで変化させる。このときの推定結果を図 2(f) に示す。

実験はヴォリュームを動かす場合についてはそれぞれ上述の通り 300 回、パニングを動かす場合については 200 回行っている。正解値を (P_i, V_i) 、推定値を $(\tilde{P}_i, \tilde{V}_i)$ としたときに両者の間のパニングの距離の平均 $d_{p\bar{a}n}$ およびヴォリュームの距離の平均 $d_{v\bar{o}l}$ を次のように求める。

$$d_{p\bar{a}n} = \frac{1}{200N} \sum_{i=1}^N \sum_{j=1}^{200} \sqrt{(P_{i,j} - \tilde{P}_{i,j})^2} \quad (15)$$

$$d_{v\bar{o}l} = \frac{1}{300N} \sum_{i=1}^N \sum_{j=1}^{300} \sqrt{(V_{i,j} - \tilde{V}_{i,j})^2} \quad (16)$$

ただし、 N は楽器の個数とする。 $d_{p\bar{a}n}$ 、 $d_{v\bar{o}l}$ およびそれぞれの分散 $d_{p\bar{a}n}^2$ 、 $d_{v\bar{o}l}^2$ および z の平均と分散を表 1 に示す。

表 1 では、評価関数に正解を与えた際 z が最小になっている。評価関数 z は目標とする周波数成分と推定した周波数成分のパワーの差の総和である。推定するパラメタはヴォリュームとパニングである。評価関数が小さくなればヴォリュームとパニングの値が正解値に近づくことを示す。パニングを横軸、ヴォリュームを縦軸にとり、正解値を (P_i, V_i) 、推定値を $(\tilde{P}_i, \tilde{V}_i)$ としたときに両者の間の距離の平均 d_e を次のように求める。

$$d_e = \frac{1}{N} \sum_{i=1}^N \sqrt{(P_i - \tilde{P}_i)^2 + (V_i - \tilde{V}_i)^2} \quad (17)$$

ただし、 N は楽器の個数を表す。

表 1 では推定結果の z の値は 1 楽器の場合は真の解と等しいが楽器数が増えるに従い増加している。正解を与えた場合の評価関数の値は小さいままである。このため、楽器数が増えるに従いローカルミニマムに陥っている。

N は楽器の個数を表す。

以上より、1 楽器の場合について評価関数 z を用いてミキシングのパラメタを 100% 推定できた。楽器数を 3 に増やした場合、パニングは約 10、ヴォリュームは約 0.8 の誤差で推定することができた。

3.2 パラメタが未知の場合

実験にはベートーヴェン弦楽三重奏ニ長調作品 9 の 2 第 1 楽章を使用する。理想とする音楽音響信号とし

表 2: 各楽器の使用音色名

Table 2: The patch name of used instrument

instrument	No.	patch name
Violin	50	Solo Vln 1
Viola	55	Solo Vla 1
Cello	60	Solo Vc 1

て、同曲の CD [4] を使用する。自動ミキシングする音響信号は同曲の譜面を MIDI で演奏することによって作成する。各楽器の音源は Roland JV-2080 SR-JV80-2 ORCHESTRAL EXPANSION 上の各音色を使用する。使用音色は表 2 に示す通りである。重要な点は、本稿で使用する音源は CD で使用されている音源とは異なることである。A/D 変換には SEK'D ProdiPLUS を使用する。

各楽器が同時に鳴っている箇所を選んで周波数特性を求めた。

自動ミキシングの結果求められた各楽器に対する各演算パラメタを表 3 に示す。このとき得られた評価値 z は 1528×10^{-4} であった。初期値が異なれば評価値も異なったため、評価値が最小となるものを解として選ぶ。

ここで、FFT の観測区間を 16384 点として同様の実験を行う。この区間内には 371.5ms の音響信号が含まれる。観測区間を増すことでより推定の安定性に変化が起こる。自動ミキシングの結果求められた各楽器に対する各演算パラメタを表 4 に示す。このとき得られた評価値 z は 4 回とも 7995×10^{-4} であった。ランダムに与えた解の初期値を図 5 に示す。初期値が異なっても 4 回とも推定結果が同じになったため、解としての信頼性が増す。

自動ミキシングの結果得られた定位に着目する。CD で演奏されている配置は左から、ヴァイオリン、ヴィオ

表 3: 2048 点の FFT において自動ミキシングの結果求められた各演算のパラメタ

Instrument	Volume[dB]	Panning
Violin	-27.4	-23.0
Viola	-23.3	-12.0
Cello	-40.6	88.0

Table 4: Our result parameter

表 1: 評価値 z の平均と分散

Table 1. Average and variance of the evaluation value z .

楽器数	種別	平均 (10^{-4}) (正解時)	分散	平均 (10^{-4}) (推定時)	分散	d_{pan}	d_{viol}	d_{pan}	d_{viol}
1	Pan	3.68	2.68×10^{-11}	3.68	2.68×10^{-11}	0	0	0	0
1	Vol	3.68	5.85×10^{-11}	3.68	5.85×10^{-11}	0	0	0	0
2	Pan	4.86	1.17×10^{-10}	703	5.37×10^{-3}	15.24	0.898	249.70	1.116
2	Vol	4.71	5.06×10^{-10}	631	5.43×10^{-3}	13.27	0.816	209.16	1.101
3	Pan	4.43	5.25×10^{-11}	616	3.44×10^{-3}	11.74	0.710	169.74	0.707
3	Vol	4.39	3.47×10^{-11}	936	9.61×10^{-3}	12.23	0.780	168.43	0.764

ラ、チェロの順になっている。表 4 から、我々のシステムも同様の順にミキシングしたことがわかる。

ミキシング前の個々の音響信号を得ることが出来、さらにミキシングのパラメタがわかっている場合、1 楽器についてさまざまなボリュームおよびパニングについて推定を行う実験を行った。その結果、100% の精度で推定を行うことができた。楽器数が増えるに従い、ローカルミニマムに陥るため推定に誤差が生じた。本稿で用いた自動ミキシングの基本的な性能が評価できる。次に、ミキシング前の個々の音響信号を得ることが出来ず、ミキシングのパラメタもわからない場合についてボリュームおよびパニングを推定する実験を行った。我々が独自に演奏したヴァイオリン、ヴィオラおよびチェロの音響信号と CD の音響信号の周波数特性を比較することにより、CD で演奏されている定位と同様の推定結果を得た。得られた評価値 z は 1528×10^{-4} であり、ミキシング前の個々の音響信号を得ることが出来る場合の推定結果と比べて 2 倍程度増えるのみである。またこの評価値は式 (14) より 1024 個のデータの総和である。各周波数あたり 1.49×10^{-4} の誤差が生じていることになる。この誤差は dB 表記すると $-56.52[dB]$ にあたる。推定のための FFT 区間を 8 倍増すことで評価値は 7995×10^{-4} に増えたが、評価に使用されるデータも 8 倍増えたためである。式 (14) を元に各周波数あたりの誤差を求めると 0.976×10^{-4} となり、dB 表記では $-60.2[dB]$ となる。推定のための FFT 区間を増す

表 4: 16384 点の FFT において自動ミキシングの結果求められた各演算のパラメータ

Table. 4: Our result parameter

Instrument	Volume[dB]	Panning
Violin	-33.4	-23.0
Viola	-16.9	-17.0
Cello	-25.9	33.0

ことで、推定誤差を小さくすることができる。

4 むすび

本稿では音楽音響信号の周波数特性に着目し、自動ミキシングを行った。ミキシング前の個々の音響信号を得ることが出来る場合、1 楽器の推定では 100% の精度で推定を行うことができた。楽器数を 3 に増やした場合、パニングは約 10、ボリュームは約 0.8 の誤差で推定することが出来た。

さらに、ミキシング前の個々の音響信号を得ることが出来ない場合について実験を行った結果、ミキシング前の個々の音響信号を得ることが出来る場合に比べて 2 倍程度誤差で推定を行うことが出来た。このとき、市販の CD で演奏されている弦楽三重奏の各楽器の位置関係を再現することが出来た。

参考文献

- [1] C. Roads, "The Computer Music Tutorial," The MIT Press, 1996.
- [2] C. Roads, "The Computer Music Tutorial," The MIT Press, pp. 458-460, 1996.
- [3] Polak, "Computational Methods in Optimization," New York Academic Press, 1971.
- [4] BEETHOVEN, "STRING TRIOS Op. 9 Nos 1/3," NIPPON COLUMBIA CO., 33CO-2303, 1988

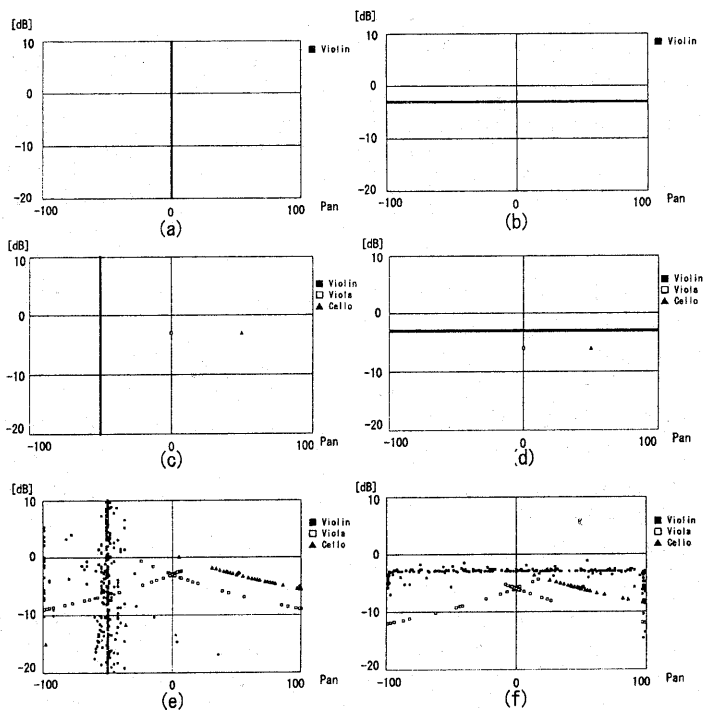


図 2: パラメタ推定結果の変化 (a)1 楽器でヴォリュームを変化させた場合の推定結果

(b)1 楽器でパニングを変化させた場合の推定結果

3 楽器を用い 1 楽器の、

(c) ヴォリュームを変化させた場合の正解値

(d) パニングを変化させた場合の正解値

(e) ヴォリュームを変化させた場合の推定結果

(f) パニングを変化させた場合の推定結果

Fig. 2 Our estimations (a)changing the volume(1 instrument) (b) changing the panning(1 instrument)(c>true values changing the volume(3 instruments)(d>true values changing the panning(3 instruments)(e) our estimations changing the volume(3 instruments)(f)our estimations changing the panning(3 instruments).

表 5: ランダムに与えられた解の初期値 (Pan, Vol[dB])

Table. 5: Initial value of solution given at random. (Pan, Vol[dB])

Instrument	1st	2nd	3rd	4th
Violin	(-63.0, -34.0)	(58.0, 0.0)	(-69.0, -1.0)	(37.0, -4.0)
Viola	(-92.0, -49.0)	(5.0, -2.0)	(-5.0, -15.0)	(74.0, -11.0)
Cello	(84.0, 9.0)	(-70, -42)	(16.0, 5.0)	(-90.0, -4.0)