

# スマートミキサー

## — 基本原理と有効性 —

高橋 弘太<sup>†</sup> 大脇 渉<sup>†</sup>

<sup>†</sup> 電気通信大学情報理工学研究所 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: †{kota,owaki}@ice.uec.ac.jp

あらまし スマートミキサーは、信号を時間周波数平面に展開し、時間周波数平面上の各点において別個の非線形な混合処理を行うことにより、従来のミキサーでは成し得なかつたきめ細かなミキシングを行う方法である。本報告では、スマートミキサーの原理と具体的な構成例を示すとともに、時間周波数平面の各点において振幅を調整するだけでなく、位相も調整することで、さらに効果的なミキシングが行えることを示したい。スマートミキサーは、音楽製作におけるミックスダウン時のみならず、音楽にカーナビ音声を重ねたり、パソコン内部での音信号の混合などの用途にも有効であり、幅広い応用を前提に研究を進めている。

キーワード ミキシング, マスタリング, マスキング, 時間周波数表現, 短時間フーリエ変換, フィルタバンク

## Smart Mixer

### — Basic Concepts and Performance —

Kota TAKAHASHI<sup>†</sup> and Wataru OWAKI<sup>†</sup>

<sup>†</sup> Graduate School of Informatics and Engineering, The University of Electro-Communications,  
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan

E-mail: †{kota,owaki}@ice.uec.ac.jp

**Abstract** The smart mixer is a proposed new type of sound mixer, that expands input signals on the time-frequency planes and mixing is performed by the nonlinear blending on each point of the plane. In this paper, basic concepts of smart mixer and typical examples of configuration for smart mixing are described. Furthermore, efficiency of phase modification on a time-frequency plane is also described. Using the concepts of smart mixer, excellent and fine sound mixing can be realized. Typical applications of the smart mixer are mixing consoles in a recording studio, car navigation systems, a sound mixing part of a personal computer and so on.

**Key words** mixing, mastering, masking, time-frequency representation, short-time Fourier transform, filter bank

## 1. はじめに

現行のサウンドミキサーは、単なる加算器である。単純に加算することによって、ある音が別の音を妨害してしまうという問題が起こりうる。これを緩和するため、ミキサーに入力する前に線形フィルタを通過させ、信号のスペクトル形を変更することで互いに重要な帯域を優先させるという手法が広く用いられてきた。すなわち、ミキサーの外側での工夫である。

これに対し、我々は、図1のように、ミキサーそのものを高度化することを目指している。これを、スマートミキサーと名付け [1]、新しい信号混合法として確立すべく研究を進めている。

スマートミキサーは、信号を時間周波数平面に展開し、時間

周波数平面上の各点において別個の非線形な混合処理を行う。これにより、従来のミキサーでは成し得なかつたきめ細かなミキシングの実現を狙っている。本報告では、スマートミキサーの原理と具体的な構成例を二例示すとともに、位相の調整を行うことの有効性について実証したい。

## 2. 定式化

### 2.1 スマートミキサーの一般形

入力信号数が3以上の場合についても同様に拡張できるので、ここでは2信号の混合で議論する。2信号をAとBとしよう。図2に、スマートミキサーの一般形を示す。以後、離散時間系で処理するものとし、サンプル番号  $n$  における両信号を、 $A[n]$ ,

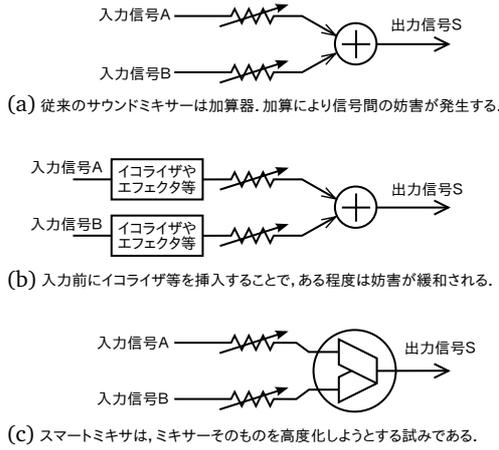


図 1 従来型ミキサーとスマートミキサー

Fig. 1 A conventional mixer and the smart mixer.

$B[n]$  と書くことにする．これらの入力信号が，短時間フーリエ変換，ウェーブレット変換，フィルタバンク処理，一般化調和解析，正弦波分解（例えば [2] のようなもの），スパース分解（例えば [3] のようなもの）などによって周波数分析され，時間周波数分布（以下  $tf$  分布と言う）として， $X_A[i, k]$ ,  $X_B[i, k]$  と表現されているとしよう．ここで， $i$  は時間軸方向の座標値である．また， $k$  は分析された成分の番号を表し，短時間フーリエ変換においては周波数方向の座標値である．以下，時間周波数平面（以下  $tf$  平面と言う）上の点を  $(i, k)$  と書くことにする．

次に，近傍の定義をする．点  $(i, k)$  を中心として，未来方向に  $m_1$  の広がりを持ち，過去方向に  $m_2$  の広がりを持ち，低周波数方向に  $n_1$  の広がりを持ち，高周波数方向に  $n_2$  の広がりを持つ近傍を， $U(i, k; m_1, m_2, n_1, n_2)$  と書き，点の集合として，

$$U(i, k; m_1, m_2, n_1, n_2) = \{(g, h) \mid g - i \leq m_1, i - g \geq m_2, h - k \leq n_1, k - h \geq n_2\} \quad (1)$$

と定義する．ここで， $m_1, m_2, n_1, n_2$  を固定して考えるときは，これらを省略し，単に  $U(i, k)$  と書くことにする．なお，上式

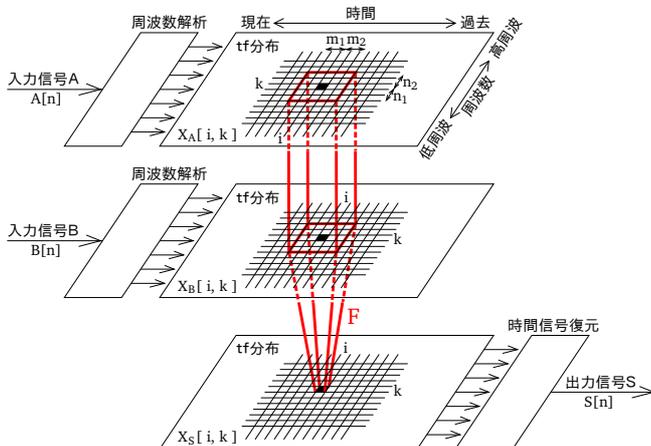


図 2 スマートミキサーの一般的な形

Fig. 2 A general structure of the smart mixer.

では，長方形領域として近傍を定義したが，長方形に限らず，円形，楕円形，台形，三角形領域などとして定義する方法もある．以上の定義を用いて，入力  $tf$  分布  $X$  の点  $(i, k)$  を中心とした近傍  $U(i, k)$  での値の集合を， $U(X, i, k)$  と書くこととし，以下のように定義する．

$$U(X, i, k) = \{X[g, h] \mid (g, h) \in U(i, k)\} \quad (2)$$

以上の準備のもと，我々が提案するスマートミキサーの一般的な形を記述すると以下ようになる．スマートミキサーは，入力  $A, B$  に対し，それぞれの入力  $tf$  分布の近傍での値  $U(X_A, i, k)$ ,  $U(X_B, i, k)$  の関数  $F$  として，出力  $tf$  分布  $X_S[i, k]$  を

$$X_S[i, k] = F(U(X_A, i, k), U(X_B, i, k)) \quad (3)$$

のように生成し， $X_S[i, k]$  を時間領域信号に逆変換することでミキシング出力  $S[n]$  を得る方法である．

ちなみに，信号  $A$  と信号  $B$  を  $W_A : W_B$  で加算 ( $S[n] = W_A A[n] + W_B B[n]$ ) する従来型のミキサーは，式 (3) の形では，

$$X_S[i, k] = W_A X_A[i, k] + W_B X_B[i, k] \quad (4)$$

と書けるので，スマートミキサーの特別な場合と言える．

## 2.2 入力 $tf$ 分布 $X[i, k]$ , $R[i, k]$ , $\phi[i, k]$

短時間フーリエ変換などを用いた場合，入力  $tf$  分布  $X[i, k]$  は，複素数値をとる．これを以下のように，振幅  $R[i, k]$  と位相  $\phi[i, k]$  に分解して表現しておく．

$$X[i, k] = R[i, k] \exp(j\phi[i, k]) \quad (5)$$

特に， $X_A, X_B$  の振幅を，それぞれ， $R_A[i, k]$ ,  $R_B[i, k]$  と書き，位相を，それぞれ， $\phi_A[i, k]$ ,  $\phi_B[i, k]$  と書くことにしよう．

なお，本文書では，位相の加減算において，断りがなくても，演算後，結果を  $-\pi$  から  $\pi$  の範囲にとどめるよう，必要に応じて  $2\pi$  の整数倍を加算する処理が挿入されるものとする．

## 2.3 有音判定分布 $Q[i, k]$

楽音や音声など，多くの信号は， $tf$  平面上でエネルギーが局在化している．スマートミキシングを行う際，ある音  $A$  の  $tf$  平面上でエネルギーが十分小さな成分は， $A$  を聴取するという観点からは存在の値がなく，また  $B$  の聴取を妨害するという観点からも考慮する必要がないので，これらの成分を処理から除外するために，有音である部分を判定しておく都合が良い．

$tf$  分布は一般に複素数値をとり， $tf$  平面上でエネルギーが集中している領域内においても，特定の点においては偶発的に値が 0 になることがありうる．そこで， $tf$  平面上の近傍領域でのエネルギーの最大値をとることによって有音部を判定し，その結果を  $Q[i, k]$  とおくことにする．入力  $tf$  分布  $X$  の有音判定分布  $Q[i, k]$  を以下のように定義する．

$$Q[i, k] = \begin{cases} 1 & \left( \max_{(g, h) \in U(i, k)} |X[g, h]|^2 \geq T_Q[k] \right) \\ 0 & (\text{otherwise}) \end{cases} \quad (6)$$

ここで， $T_Q$  は，有音と判定する閾値であり，固定値を用いる方法の他，入力信号強度または出力信号強度に適応させて可

変とする方法や、ミキサー操作者が手動で調整する方法がある。

また、 $Q[i, k]$  の定義のしかたとしては、式(6)の他、近傍内でのエネルギーや振幅の平均値として定義することができる他、分布の値を上式のように2値化せずに連続量として定義しても良いだろう。その場合、エネルギーや振幅と線形関係にある量を用いるだけでなく、非線形関係にある量を用いることも可能であると考えている。さらに、最大値や平均で近傍内を評価するとき、tf平面上での距離に応じて重みをつける方法も有効であろうし、人間の聴覚特性におけるマスキング効果のデータに基づいてtf平面での近傍の距離や閾値を決める方法も考えられる。いずれにしても、以上のようにすれば、tf平面上である音が妨害する領域と、その妨害量を書き表すことができる。なお、単一の入力信号に対して、近傍のパラメータ  $m_1, m_2, n_1, n_2$  を変えることで複数の  $Q[i, k]$  を用意する方法も有望である。

以下、 $X_A, X_B$  の有音判定分布を、それぞれ、 $Q_A[i, k], Q_B[i, k]$  と書くことにしよう。

#### 2.4 ピーク判定分布 $P[i, k]$

前節で定義した有音判定分布  $Q[i, k]$  を用いることで、tf平面上での有音部と無音部を判定することができる。

しかし、 $Q[i, k]$  が有音と判定する領域は、tf平面上で大きな面積を占めるかたまりとなってしまうことがしばしばあり、このかたまり全体で位相調整が行われてしまうと、調整されたほうの信号は、大きく劣化してしまうことが起こり得る。

そこで、tf平面上で、両隣の周波数のエネルギーに比較してエネルギーが集中している領域を判定してやることで、有音領域に比較して限られた領域を選び、この領域を中心として位相調整を行うことを考えた。これによって、音質の劣化のより少ないスマートミキサーを実現することができる。

この目的で作られるピーク判定分布が  $P[i, k]$  である。ピーク判定分布  $P[i, k]$  は、以下の2つの方法のどちらかで生成することができる。

第一の方法は、単純にtf平面上での振幅のピークとして、判定する方法である。例えば、分布  $X$  のピーク判定分布  $P[i, k]$  を以下のように定義する。

$$P[i, k] = \begin{cases} 1 & (|X[i, k-1]| < |X[i, k]| \text{ かつ} \\ & |X[i, k+1]| < |X[i, k]|) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

第二の方法は、位相変化を用いる方法である。今、tf分布  $X$  を持つ信号が、周波数 bin 番号  $k$  にエネルギー中心を持つ成分を含むとしよう。このとき、この成分に対し、時間  $i-1$  と時間  $i$  のあいだでは、 $X$  には  $2\pi k N_{\text{SFT}}/N_{\text{FFT}}$  だけの位相変化が起こる。ここで、 $N_{\text{SFT}}$  は、フレームシフト長であり、 $N_{\text{FFT}}$  は、FFT 点数である。実際の位相変化と、以上の仮定のもとでの位相変化の差を  $\phi_1[i, k]$  と書き表すことにしよう。

$$\phi_1[i, k] = \phi[i, k] - \phi[i-1, k] - 2\pi k N_{\text{SFT}}/N_{\text{FFT}} \quad (8)$$

もし、周波数 bin 番号  $k$  にエネルギー中心を持つ成分が存在するならば、 $\phi_1[i, k] = 0$  となるはずである。それに対し、エネルギー中心が  $k$  より低いところであれば、実際の位相変化は小さ

めになるので、 $\phi_1[i, k] < 0$  となり、逆に、エネルギー中心が  $k$  より高いところであれば、実際の位相変化は小さめになるので、 $\phi_1[i, k] > 0$  となるであろう。そこで、第二の方法では、以下のようにピーク判定分布  $P[i, k]$  を定義する。

$$P[i, k] = \begin{cases} 1 & (\phi_1[i, k] < 0 \text{ かつ } \phi_1[i, k-1] \geq 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

これによって、 $(i, k-1)$  と  $(i, k)$  との間にエネルギー中心(ピーク)があることが、間接的に判定できる。スマートミキサーは、位相を主要な指標とする方法であるから、式(7)よりも、この式(9)のほうが同一の指標を用いているという点で親和性が高い。ただし、式(7)と式(9)の判定結果は類似しており、ピーク判定分布としては、どちらの定義を使っても構わない。

なお、tf平面上で信号が調波構造を有しない部分や白色雑音に類似したスペクトル構造を持つ部分において、式(7)と式(9)のどちらを用いても異常に多くのピークが検出されてしまい、相対する信号の位相を強制するための領域としては好ましくない場合がある。そこで、以下のようにして  $P[i, k]$  を改良するのがよい。

まず、ピーク判定分布  $P[i, k]$  を定義した発想にならって、ドイツ判定分布  $D[i, k]$  を以下のどちらかで定義する。

$$D[i, k] = \begin{cases} 1 & (|X[i, k-1]| > |X[i, k]| \\ & \text{かつ } |X[i, k+1]| > |X[i, k]|) \\ 0 & (\text{otherwise}) \end{cases} \quad (10)$$

$$D[i, k] = \begin{cases} 1 & (\phi_1[i, k] \geq 0 \text{ かつ } \phi_1[i, k-1] < 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (11)$$

式(10)が振幅による定義、式(11)が位相による定義である。

以上の準備のもと、 $P[i, k]$  を選別する。選別の基準としては、たとえば、「 $P[i, k]$  が選別の結果生き残るのは、 $(i, k)$  を中心とした領域  $U(i, k; V_D, V_D, W_D, W_D)$  において、 $D[i, k] = 1$  が1個も存在せず、かつ、領域  $U(i, k; V_P, V_P, W_P, W_P)$  において、 $P[i, k] = 1$  が  $2V_P + 1$  の0.8倍以上存在するときのみ」とするなどが考えられる。ここで、 $V_D, V_P, W_D, W_P$  は、選別のためのtf平面上の局所領域の大きさを決めるためのパラメータである。以上の操作によって、時間軸方向で連結した成分のみを選抜することができる。

なお、このような選別の実施の有無にかかわらず、式(7)と式(9)のどちらも、tf平面でピークと判定する点が非常に疎である。そこで、最後にピーク判定分布  $P[i, k]$  を、いわゆるモルフォロジー的操作によって「太らせて」おいたほうがよい場合が多い。この処理は、たとえば、

$$P[i, k] \leftarrow \begin{cases} 1 & (\text{近傍 } U(i, k; V, V, W, W) \text{ の中に} \\ & P(i, k) = 1 \text{ が存在する}) \\ 0 & (\text{otherwise}) \end{cases} \quad (12)$$

の処理による  $P[i, k]$  の修正で、実現することができる。

### 3. スマートミキサー・第1の方法

今、音楽 A に対して、音声 B をミキシングすることを考え

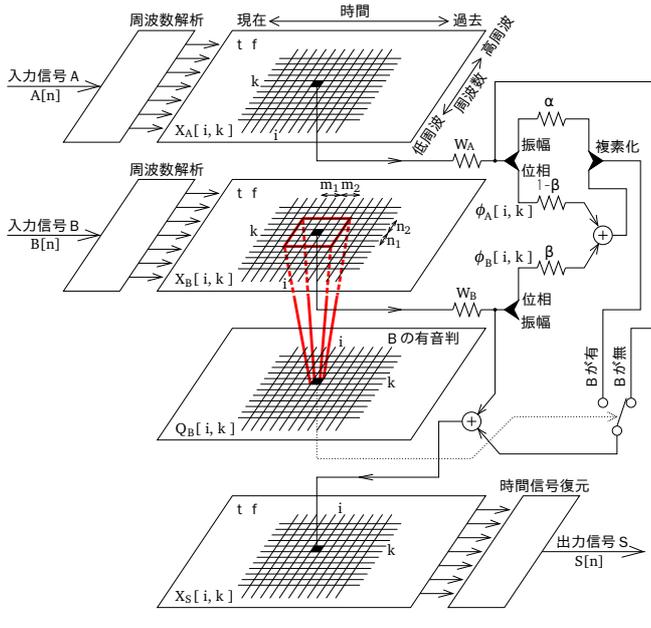


図3 スマートミキサー・第1の方法

Fig. 3 A structure of a simple smart mixer. (method 1)

る。スマートミキサー・第1の方法では、音声Bの聞き落としが生じないように、以下のようにミキシングする。

$$X_S[i, k] = \begin{cases} \alpha W_A R_A \exp j(\beta \phi_B + (1-\beta) \phi_A) \\ \quad + W_B X_B[i, k] & (Q_B[i, k] = 1) \\ W_A X_A[i, k] + W_B X_B[i, k] & (Q_B[i, k] = 0) \end{cases} \quad (13)$$

上式は、 $tf$ 平面上、音声Bが存在する部分では、音楽Aに細工を加えるようになっている。なお、 $Q[i, k]$ が、2値化関数でなく連続関数のときは、場合分けせず、滑らかに切り替わるようにすればよい。Aの振幅については、Bを妨害ないように係数 $\alpha$ を乗じて減衰させている。また、AとBの位相が異なっているとミキシングすることでBの音が打ち消されてしまうので、Bの位相をAの位相へ近づける。その係数が $\beta$ である。上式による実現を図3に図示する。

図3をソフトウェアとして実装し、表1に示す諸元で実験した場合の結果を、図4と図5に示す。

図4は、時間信号波形で、上から、(a)入力信号Aのエレキギター、(b)入力信号Bの女性音声、(c)従来法によるミキシング出力、(d)スマートミキサー・第1の方法によるミキシング出力、となっている。エレキギターの振幅が女性音声の振幅に比較して一桁上の大きさを持っていることに注意されたい。

図5は、各信号の $tf$ 分布を示したものである。 $tf$ 分布の各点において、振幅を濃淡に置き換えて表示している。白色が最も振幅が小さく、濃くなるにしたがって振幅が大きい。2.6秒～3.0秒の時間区間に注目して欲しい。この時間区間において、エレキギターは、横縞状の $tf$ 分布形状を呈している。一方、音声は、発音のイントネーション変化により、曲線状の $tf$ 分布形状となっている。従来法によるミキシング出力(c)は、横縞状の $tf$ 分布形状であり、エレキギターのスペクトル形が全面を覆っており、ほとんどエレキギターのスペクトルと言って過言ではないほどであることがわかる。聴取実験を行っても、この時間

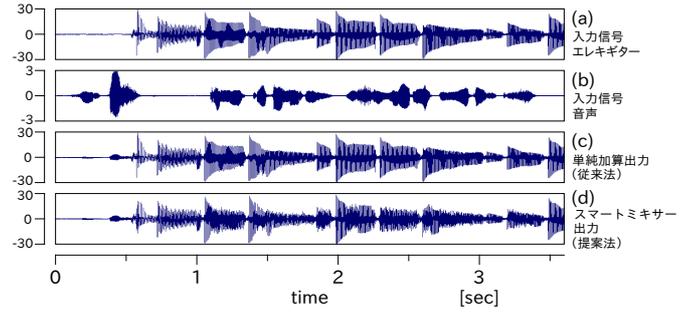


図4 従来法とスマートミキサーの波形による比較(第1の方法)

Fig. 4 A waveform comparison of conventional mixing and smart mixing. (method 1)

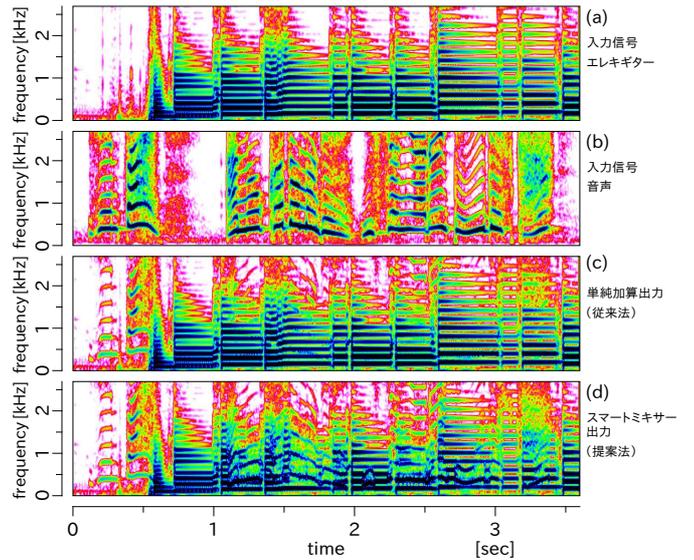


図5 従来法とスマートミキサーの $tf$ 分布による比較(第1の方法)

Fig. 5 A  $tf$ -map comparison of conventional mixing and smart mixing. (method 1)

区間においては著しく音声が聞き取りにくい。これに対し、スマートミキサーによるミキシング出力では、横縞状と曲線状のスペクトルが共存しており、音声とエレキギターの双方のスペクトルがうまく勘案されていることがわかる。聴取実験を行った

表1 スマートミキサー・第1の方法の実験諸元

Table 1 Experimental parameters of a smart mixer. (method 1)

サンプリング周波数	$F_s$	44.1 kHz
入力信号 A	$A[n]$	エレキギター (茶摘みのメロディー)
入力信号 B	$B[n]$	女性音声 (「先輩、あの角まで一緒に歩いていいですか」と発言)
FFT 点数	$N_{FFT}$	4096 点
解析時ウィンドウ関数	$H[n]$	余弦関数形 128 点+平坦部 255 点+余弦関数形 128 点
合成時ウィンドウ関数	$G[n]$	ハンニング窓形 255 点
フレームシフト	$N_{SFT}$	64 点
ミキシングパラメータ	$\alpha$	0.95
ミキシングパラメータ	$\beta$	0.85

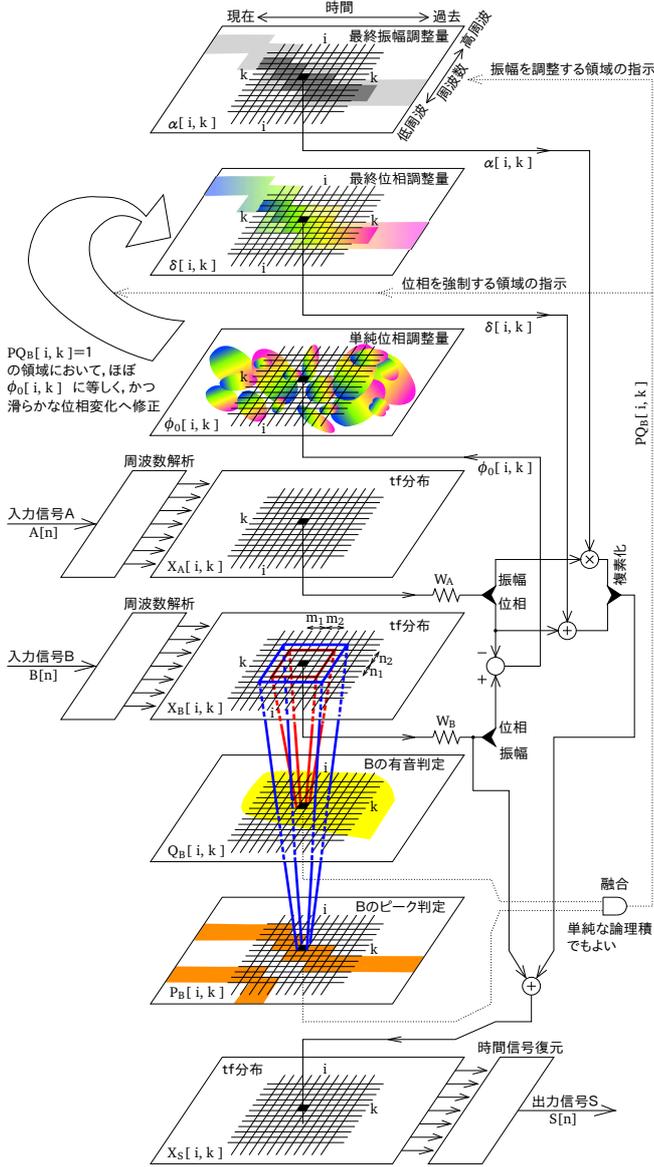


図6 スマートミキサー・第2の方法

Fig. 6 A structure of sophisticated smart mixer. (method 2)

でも、この時間区間において音声は明らかに聞き取り易くなっている。

音声を聞き取り易くするだけであれば、音声の音量を上げれば良いと思うかもしれない。しかし、それでは、音楽が存在しない時間区間の音量も上がってしまう。スマートミキサーがそのようなミキシングを行っていないことは、図5の0.0秒~0.5秒の時間区間を観察することで確かめることができる。音楽が存在しない時間区間においては、スマートミキサーの音量は単純加算(従来法)と同じである。tf平面上で、音楽が妨害を与えている部分だけに必要な手を打つというのがスマートミキサーの考え方である。

#### 4. スマートミキサー・第2の方法

第1の方法では、信号Aの音質の劣化が無視できない場合がある。より高音質を目指す場合には、tf平面での位相修正量を滑らかにするとよい。この考えに基づき、修正量を平滑化する

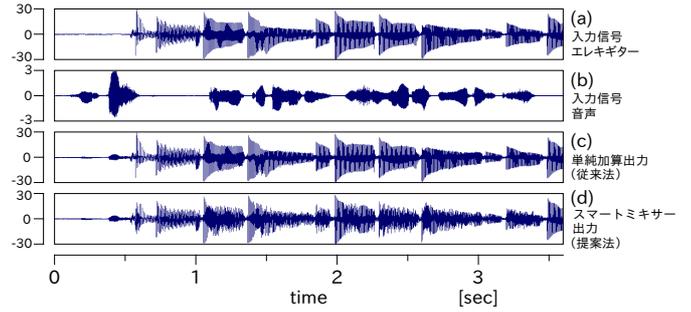


図7 従来法とスマートミキサーの波形による比較(第2の方法)

Fig. 7 A waveform comparison of conventional mixing and smart mixing. (method 2)

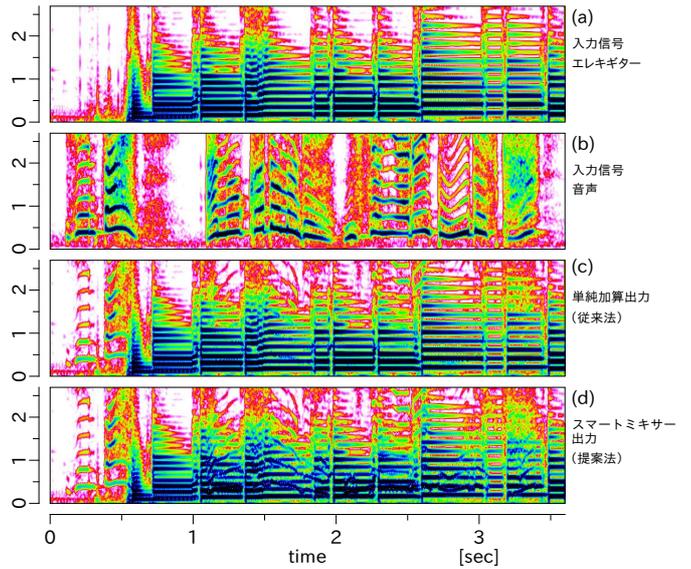


図8 従来法とスマートミキサーのtf分布による比較(第2の方法)

Fig. 8 A tf-map comparison of conventional mixing and smart mixing. (method 2)

仕組みを持つスマートミキサーを、図6に示す。これを第2の方法と呼ぼう。第2の方法では、混合は次式で行われる。

$$X_S[i, k] = \alpha[i, k] W_A R_A[i, k] \exp j(\phi_A[i, k] + \delta[i, k]) + W_B X_B[i, k] \quad (14)$$

すなわち、信号Aに対して、位相に $\delta[i, k]$ の調整を加え、振幅に $\alpha[i, k]$ の調整を加えている。もちろん、信号Bの振幅や位相にも調整を加えてもよいが、複雑になるので、ここでは簡単のため、信号Aに対する調整のみの場合について解説する。

まず、信号Aと信号Bの位相差を $\phi_0$ としよう。

$$\phi_0[i, k] = \phi_B[i, k] - \phi_A[i, k] \quad (15)$$

もし、tf平面全体について、 $\delta[i, k] = \phi_0[i, k]$ とするならば、信号Aの位相は、信号Bの位相そのものになり、信号Bを聞き取らせるといふ点では好ましい。しかし、信号Aの音が不自然なものになってしまうため、工夫しなければならない。

そこで、図6に示すように、信号Bの有音判定分布 $Q_B[i, k]$ を式(6)で生成する。さらに、ピーク判定分布 $P_B[i, k]$ を式(12)で生成する。両者を用いて、位相を強制すべき領域で1となる分布 $PQ_B[i, k]$ を、以下のように生成する。

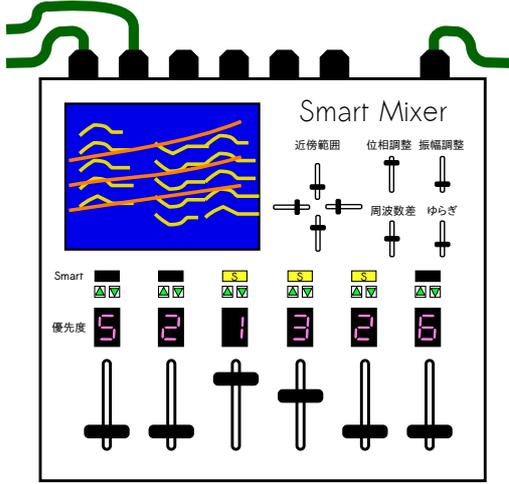


図 9 装置としてのスマートミキサーのイメージ図

Fig. 9 A image figure of a hardware implemented smart mixer.

$$PQ_B[i, k] = \begin{cases} 1 & (P_B[i, k] = 1 \text{ かつ } Q_B[i, k] = 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (16)$$

以上の準備のもと、位相調整量  $\delta[i, k]$  の生成アルゴリズムを以下に述べる。このアルゴリズムは、 $tf$  平面上の点  $(i, k)$  における調整量  $\delta[i, k]$  は、 $(i, k)$  のまわりで近接する  $\delta[i, k]$  の値とスムーズにつなげようとする力で引っ張られ、かつ、 $PQ_B[i, k] = 1$  の  $(i, k)$  においては、 $\phi_0[i, k]$  に引き寄せる力でも引っ張られるというものである。アルゴリズムは、 $tf$  平面全体に対する数回～数百回のイテレーション演算として実現できる。 $n$  回目のイテレーション演算での  $\delta[i, k]$  を、 $\delta[i, k]^{(n)}$  と書くとき、初期状態として  $\delta[i, k]^{(0)} = 0$  と置き、 $\delta[i, k]^{(n+1)}$  は、 $\delta[i, k]^{(n)}$  から以下の更新則で求める。これによって、最終的な  $\delta[i, k]$  は、全ての力が釣り合う値に収束する。

$$\begin{aligned} \delta[i, k]^{(n+1)} = & \delta[i, k]^{(n)} \\ & + \epsilon \lambda PQ_B[i, k] f(\phi_0[i, k] - \delta[i, k]^{(n)}, \pi C_P, E_P) \\ & + \epsilon f(\delta[i, k-1]^{(n)} - \delta[i, k]^{(n)}, \pi C_F, E_F) \\ & + \epsilon f(\delta[i, k+1]^{(n)} - \delta[i, k]^{(n)}, \pi C_F, E_F) \\ & + \epsilon f(\delta[i-1, k]^{(n)} - \delta[i, k]^{(n)}, \pi C_T, E_T) \\ & + \epsilon f(\delta[i+1, k]^{(n)} - \delta[i, k]^{(n)}, \pi C_T, E_T) \end{aligned} \quad (17)$$

ここで、 $\epsilon$  は、学習速度であり、イテレーションの回数に応じて調整する。また、 $\lambda$  は、 $\phi_0[i, k]$  による引き寄せ力と、 $\delta[i, k] = 0$  を  $tf$  平面上で滑らかにする力のバランスを決める定数である。 $C_P, C_F, C_T, E_P, E_F, E_T$ , も定数で、非線形関数  $f$  は、

$$f(\Delta, C, E) = \begin{cases} -\pi & (\Delta < -C) \\ -\pi \left(\frac{-\Delta}{C}\right)^E & (-C \leq \Delta < 0) \\ \pi \left(\frac{\Delta}{C}\right)^E & (0 \leq \Delta < C) \\ \pi & (C \leq \Delta) \end{cases} \quad (18)$$

である。図 6 をソフトウェアで実装し、表 1 に示す諸元で実験した結果を、図 7 と図 8 に示す。図 5 と比較すると、ギターのスペクトルの崩れが少なく、より良好であることがわかる。

表 2 スマートミキサー・第 2 の方法の実験諸元

Table 2 Experimental parameters of a smart mixer. (method 2)

サンプリング周波数	$F_s$	44.1 kHz
入力信号 A	$A[n]$	エレキギター (茶摘みのメロディー)
入力信号 B	$B[n]$	女性音声 (「先輩、あの角まで一緒に歩いていいですか」と発言)
FFT 点数	$N_{FFT}$	4096 点
解析時ウインドウ関数	$H[n]$	余弦関数形 128 点+ 平坦部 255 点+ 余弦関数形 128 点
合成時ウインドウ関数	$G[n]$	ハンニグ窓形 255 点
フレームシフト	$N_{SFT}$	128 点
振幅の調整	$\alpha[i, l]$	簡単のため恒等的に 1 とした。
有音判定(時間軸)	$m_1, m_2$	$m_1 = m_2 = 3$
有音判定(周波数軸)	$n_1, n_2$	$n_1 = n_2 = 4$
有音判定時周波数補正	$G_{adj}$	12 dB/oct. (本文説明は略)
ピーク判定(時間軸)	$V_P$	2 点
ピーク判定(周波数軸)	$W_P$	4 点
ディップ判定(時間軸)	$V_D$	2 点
ディップ判定(周波数軸)	$W_D$	2 点
ピーク肥大(時間軸)	$V$	4 点
ピーク肥大(周波数軸)	$W$	4 点
非線形関数パラメータ	$C_T$	0.10
非線形関数パラメータ	$C_K$	0.10
非線形関数パラメータ	$C_P$	0.25
非線形関数パラメータ	$E_T$	3.0
非線形関数パラメータ	$E_K$	3.0
非線形関数パラメータ	$E_P$	1.0
イテレーション回数	$N_{it}$	100
イテレーションゲイン	$\epsilon$	0.2
フォースバランス	$\lambda$	10.0

## 5. おわりに

ミキサーの前後に高度な信号処理系を配置するのではなく、ミキサーそのものを高度化する試みとして、時間周波数平面上の各点において別個の非線形な混合処理を行う方法を提案し、これをスマートミキサーと名付けた。

簡単な方法と、やや複雑な方法の 2 種類のスマートミキサーをソフトウェアで実装し、その有効性を実証し、比較を行った。

今後は、非線形な混合処理の部分をさらに練り上げるとともに、FPGA によるリアルタイム処理をめざし、図 9 のような、スタンドアロン型スマートミキサーを実現したいと考えている。

## 文献

- [1] 大脇渉, 高橋弘太, “スマートミキサー～新しい音信号混合～,” 信学技報, 本件の前の発表, 2011
- [2] R. J. McAulay, T. F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” IEEE Trans. ASSP-34(4), 744-754, 1986.
- [3] D. L. Donoho, X. Huo, “Uncertainty principles and ideal atomic decomposition,” IEEE Trans. IT-47(7), 2845-2862, 2001.