

優先度付き非線形演算による新しいサウンドミキサー

— スマートミキサーの提案 — *

○高橋弘太, 有北知弘, 大脇渉, 手塚歩, 宮地紘司 (電通大)

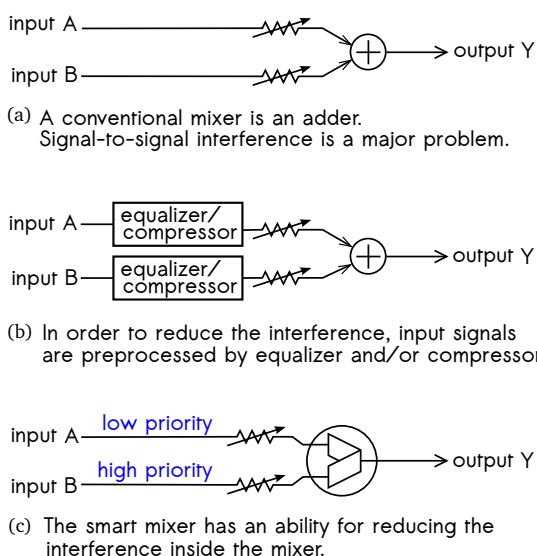


Fig. 1 A conventional mixer and the smart mixer.

1 スマートミキサーとは

サウンドミキサーは、音響信号処理の中核を成す要素であり、音楽製作におけるミックスダウン時のみならず、パソコン内部での信号混合、カーオーディオシステムにおけるカーステレオとナビ音声の混合など、様々な場面で使用されている。しかし、どのような場面であっても、混合自体は、Fig. 1(a)に示すとおり、単純な加算によるものが圧倒的であった。

加算を行うことの弊害は明らかである。たとえば、音楽製作において、混合によりマスキングなどの現象が生じて、ボーカルや、楽器音の聴取が妨害される。

これを緩和するために、Fig. 1(b)に示すように、イコライザやコンプレッサなどの信号処理機構をミキサーの前後に付加する方策が知られている。すなわち、ミキサー自身の改良ではなく、ミキサーの外側での対処である。

我々は、Fig. 1(c)に示すように、ミキサー自身を高度化することで、信号混合における問題点を根本的に解消できないかと考えた。

入力信号に優先度の概念を導入し、単純加算では実現できないきめ細かなミキシングを行うことをめざし、これを、スマートミキサー [1, 2] と呼び、研究を進めている。Table 1 に想定している応用を示す。

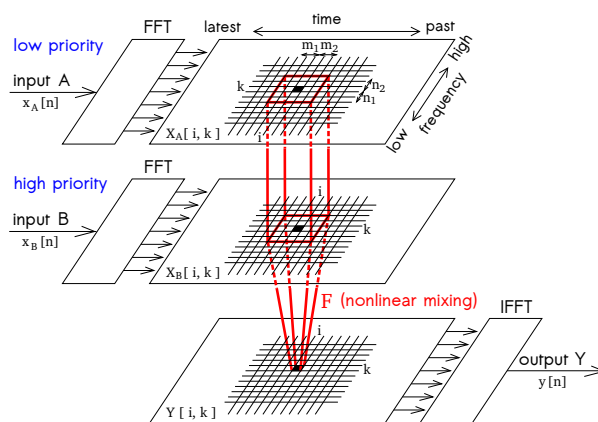


Fig. 2 The general structure of the smart mixer. Input signals are mixed on the time-frequency plane.

Fig. 2 に、スマートミキサーの基本的な構造を示す。入力信号を A, B としたとき、それらの波形 $x_A[n]$, $x_B[n]$ は、窓関数をかけて短時間FFTされた後、時間周波数平面上(以下、tf平面と書く)での信号 $X_A[i, k]$, $X_B[i, k]$ として展開される。ここで、 i は時間フレーム番号、 k は周波数 bin 番号である。混合は、tf平面上で行われ、混合処理後に得られた信号 $Y[i, k]$ を逆FFT処理することで、出力信号 $y[n]$ を得る。

スマートミキサーにおいて、tf平面上でのある点 (i, k) に対する混合結果 $Y[i, k]$ は、その点の近傍における $X_A[i, k]$ と $X_B[i, k]$ の値から算出する。このとき、「優先度」の考え方を導入するのがスマートミキサーの特徴である。信号 B を優先信号としたとき、 $Y[i, k]$ を算出するにあたって、できるだけ信号 A が信号 B を妨害しないように信号 A に対して非線形な処理が加えられる。これによって、従来の単純加算での混合では成し得なかった高度な重ねあわせを行うことをめざしている。

Table 1 Typical applications of the smart mixer.

応用の場面	効能
音楽製作	トラック間での妨害を抑制する
カーナビ	カーステレオを邪魔せず音声混合
ラジオ放送	音楽の音量を絞らずに DJ 音声を混合
カラオケ (1)	伴奏に邪魔されずボーカルを目立たせる
カラオケ (2)	デュエットを美しく聴かせる
パソコン	別アプリケーションの音を妨害なく混合
ゲーム機	音楽や効果音をプレイしやすいよう混合
遠隔会議	司会者や重要人物の音声を際立たせる

* A new type of sound mixer using nonlinear processing based on a priority concept –smart mixer–. by TAKAHASHI, Kota, ARIKITA, Tomohiro, OWAKI, Wataru, TEZUKA, Ayumu, MIYACHI, Hiroshi (The University of Electro-Communications)

Table 2 Various types of smart mixer.

非優先音		優先音		名称
振幅調整	位相調整	振幅調整	位相調整	
○	—	—	—	局所抑制法
—	○	—	—	局所同期法
○	○	—	—	今後, 研究を予定
○	—	○	—	
—	○	—	○	
○	○	○	○	

2 スマートミキサーのバリエーション

以下では、信号 A を非優先信号、信号 B を優先信号として話を進めよう。前章で述べたように、ある点 (i, k) における混合信号 $Y[i, k]$ の値を、その点の近傍 $U(i, k)$ における入力信号の集合 $\{X_A[i', k'], X_B[i', k'] \mid (i', k') \in U(i, k)\}$ の非線形関数として求めるのがスマートミキサーの一般型であり、この非線形関数の選択がスマートミキサーの性能決定に大きくかかわってくる。

我々は、様々な非線形関数を試し、また、関数内のパラメータも調整して、どのような非線形関数を用いるのが好ましいかについての研究に取り組んでいる。

スマートミキサーが調整の対象とする音として、非優先音だけに絞る方法だけでなく、優先音についても含める方法も考える。さらに、その調整を、振幅について行う方法、位相について行う方法、振幅と位相の両方について行う方法の3つの方法が考えられる。これらの観点から、スマートミキサーで我々が研究すべきことを整理すると、Table 2 のようになる。

振幅の調整の考え方の基本は、優先音にとって重要な成分が存在している tf 平面上のある点における非優先音の成分の抑制である。これを、我々は局所抑制法 [1] と呼んでいる。

一方、位相の調整の考え方の基本は、優先音にとって重要な成分が存在している tf 平面上のある点を中心として、非優先音の位相を優先音に揃えることである。これによって、非優先音が優先音に対して逆相になって優先音を弱めてしまうことを防ぐことができる。これを、我々は局所同期法 [2] と呼んでいる。

以下、本稿では、局所同期法に絞って考えていく。

3 局所同期法によるスマートミキサー

入力 tf 分布 $X[i, k]$ は、複素数値をとるので、これを、振幅 $R[i, k]$ と位相 $\phi[i, k]$ で表現しておく。

$$X[i, k] = R[i, k] \exp(j\phi[i, k]) \quad (1)$$

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。

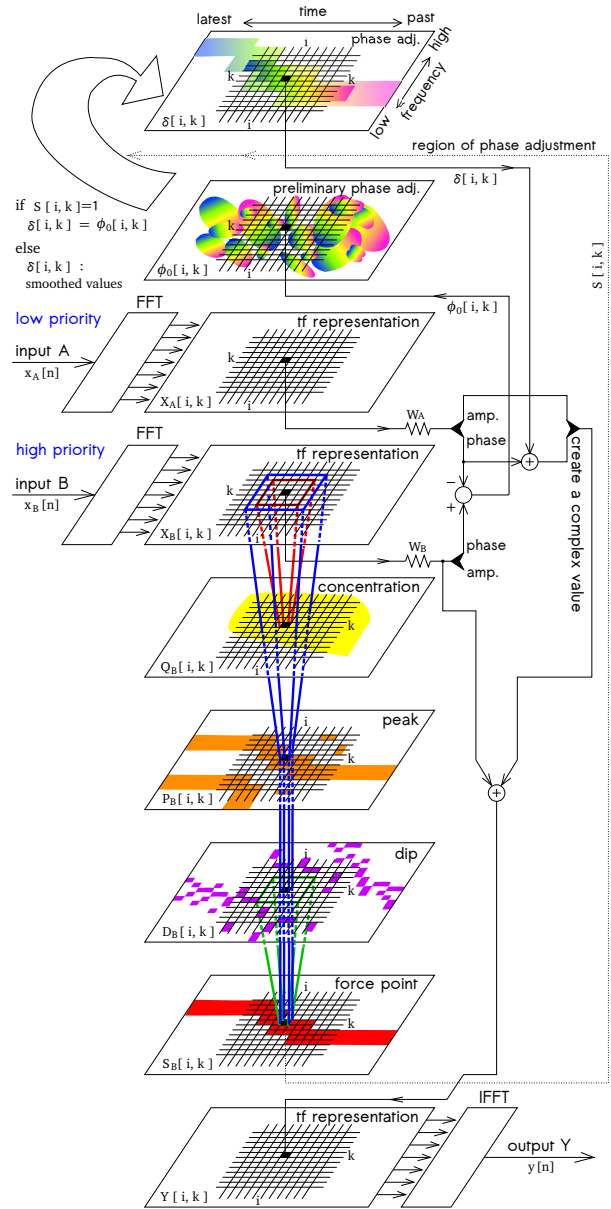


Fig. 3 A smart mixer based on the phase adjustment method and the priority concept.

局所同期法を、Fig. 3 に示す。本手法では、

$$Y[i, k] = W_A R_A[i, k] \exp j(\phi_A[i, k] + \delta[i, k]) + W_B X_B[i, k] \quad (2)$$

に示すように、信号 A の位相に $\delta[i, k]$ の調整を加える。信号 A と信号 B の位相差を ϕ_0

$$\phi_0[i, k] = \phi_B[i, k] - \phi_A[i, k] \quad (3)$$

としたとき、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う点では好ましい。しかし、tf 平面上では矛盾した信号となるため、時間信号に戻す際にフレーム間で干渉が生じ、信号 A の劣化が甚だしくなる。

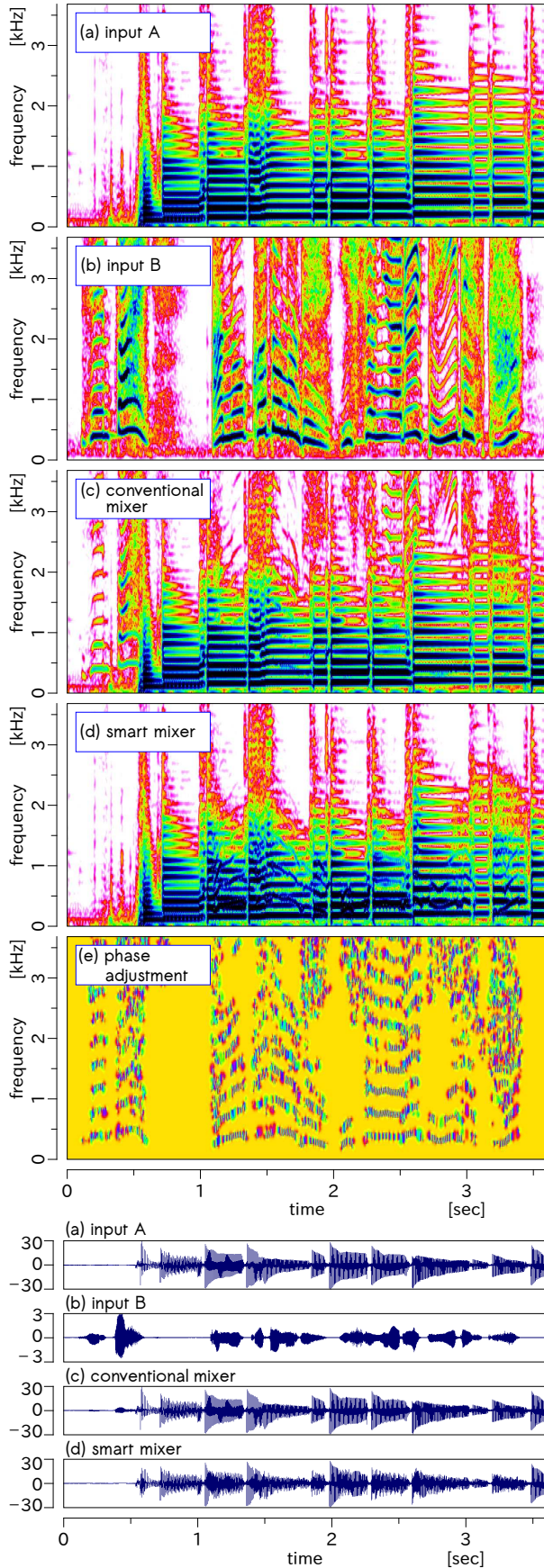


Fig. 4 A smart mixer based on the amplitude and phase adjustment theory and the priority concept.

これを防ぐため、位相強制は、信号 A の聞き取りに最も重要な tf 平面上の部分だけに限られるべきであるし、tf 平面上でなめらかにする必要がある。

この要求をかなえるため、我々は、tf 平面上で定義される 4 つの分布を準備する．有音分布 $Q[i, k]$ ，ピーク分布 $P[i, k]$ ，ディップ分布 $D[i, k]$ ，強制位置分布 $S[i, k]$ である．

有音分布 $Q[i, k]$ は、以下のように定義した．

$$Q[i, k] = \begin{cases} 1 & \left(\max_{(g, h) \in U(i, k)} |X[g, h]|^2 \geq T_Q[k] \right) \\ 0 & \text{(otherwise)} \end{cases} \quad (4)$$

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エネルギーの集中する領域を漏らさず検出しようというのが上式の狙いである．

次に、ピーク分布 $P[i, k]$ を定義する．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_{SFT} は、フレームシフト長であり、 N_{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 $\phi_1[i, k]$ と書き表すこととしよう．

$$\phi_1[i, k] = \phi[i, k] - \phi[i-1, k] - 2\pi k N_{\text{SFT}}/N_{\text{FFT}} \quad (5)$$

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、 $\phi_1[i, k] < 0$ 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、 $\phi_1[i, k] > 0$ となる．この性質をとらえて、我々はピーク判定分布 $P[i, k]$ を以下のように定義した．

$$P[i, k] = \begin{cases} 1 & (\phi_1[i, k] < 0 \text{ かつ } \phi_1[i, k-1] \geq 0) \\ 0 & \text{(otherwise)} \end{cases} \quad (6)$$

上式によって、 $(i, k-1)$ 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．

同じ考えに基づき、エネルギーの谷をあらわすディップ判定分布 $D[i, k]$ は、以下のように定義できる．

$$D[i, k] = \begin{cases} 1 & (\phi_1[i, k] \geq 0 \text{ かつ } \phi_1[i, k-1] < 0) \\ 0 & \text{(otherwise)} \end{cases} \quad (7)$$

以上の準備のもと、tf 平面上で信号 A の位相を強制する領域を示す分布 $S[i, k]$ を作成する．今回は、

$$S[i, k] = \begin{cases} 1 & (P[i, k] = 1 \text{ かつ } Q[i, k] = 1 \text{ かつ } \\ & (i, k) \text{ の近傍領域内で } D(i, k) = 1 \\ & \text{の数が一定閾値以下}) \\ 0 & \text{(otherwise)} \end{cases} \quad (8)$$

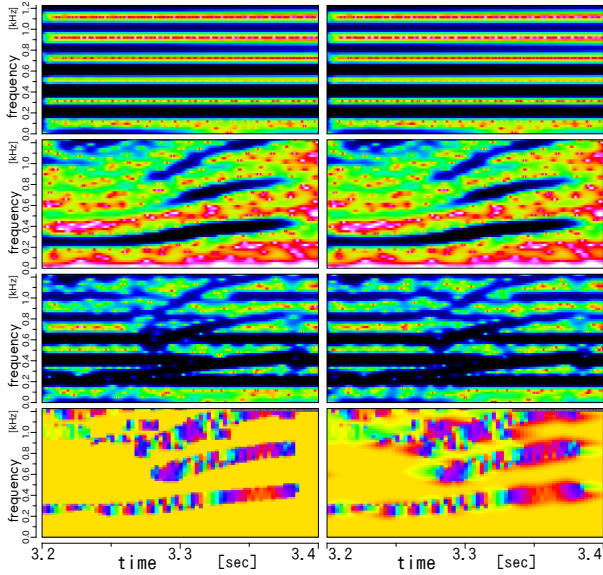


Fig. 5 Effects of smoothing on the phase adjustment.

によって $S[i, k]$ を決定した.

次に, tf 平面上で $S[i, k] = 1$ の部分で, できるだけ位相を強制しつつも, なめらかな位相調整になるような位相調整量 $\delta[i, k]$ 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, 強制位置分布が $S_B[i, k] = 1$ 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.

$$\begin{aligned} \delta[i, k]^{(n+1)} = & \delta[i, k]^{(n)} \\ & + \epsilon \lambda S[i, k] f(\phi_0[i, k] - \delta[i, k]^{(n)}, \pi C_P, E_P) \\ & + \epsilon f(\delta[i, k-1]^{(n)} - \delta[i, k]^{(n)}, \pi C_F, E_F) \\ & + \epsilon f(\delta[i, k+1]^{(n)} - \delta[i, k]^{(n)}, \pi C_F, E_F) \\ & + \epsilon f(\delta[i-1, k]^{(n)} - \delta[i, k]^{(n)}, \pi C_T, E_T) \\ & + \epsilon f(\delta[i+1, k]^{(n)} - \delta[i, k]^{(n)}, \pi C_T, E_T) \end{aligned} \quad (9)$$

ここで, ϵ 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, λ 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. 今回は, $\epsilon=0.2, \lambda=10.0$ とした. その他の定数は, $C_P=0.25, C_F=0.10, C_T=0.10,$

$E_P=1.0, E_F=3.0, E_T=3.0$ とした. 非線形関数 f は,

$$f(\Delta, C, E) = \begin{cases} -\pi & (\Delta < -C) \\ -\pi(\frac{-\Delta}{C})^E & (-C \leq \Delta < 0) \\ \pi(\frac{\Delta}{C})^E & (0 \leq \Delta < C) \\ \pi & (C \leq \Delta) \end{cases} \quad (10)$$

である. 以上をソフトウェアで実装して実験した結果を, Fig. 4 に示す. (a) の信号 A はギター演奏, (b) の信号 B は音声である. (c) の従来型の単純加算のミキサー出力に比較して, (d) のスマートミキサー出力は, 音声のスペクトル形が残存しているという点で秀でている. (e) は位相調整量 $\delta[i, k]$ であり, 黄色が 0 を示し, 緑→青→赤→黄の順で位相が 1 回転する.

式 (9) のイテレーション演算で位相をなめらかにすることの有効性を, Fig. 5 に示す. Fig. 5 は, 左側の図が位相調整として $\phi_0[i, k]$ を直接用いたもの, 右の図がなめらかな $\delta[i, k]$ を用いたものである. 表示区間は, Fig. 4 の 3.4 s ~ 3.6s, 0~1.2 kHz の範囲である. 位相をなめらかにすることで, スペクトル構造の崩れが若干少なくなっていることがわかる.

4 おわりに

スマートミキサーを紹介し, そのタイプを整理した. 位相調整を行う局所同期法について, 処理方法を述べた. 音楽と音声の混合に関して実験を行い, スマートミキサーの有効性を示した. また, 位相をなめらかに調整することの必要性を示した.

スマートミキサーは, スタンドアロン型のミキサーとしての実現の他, Fig. 6 に示すように, ProTools など, 音楽製作などで使われている DAW に外付けされるプラグインとしての実装も予定している.

参考文献

- [1] 大脇, 高橋, 信学技報, EA2011-102, 37-42, 2011.
- [2] 高橋, 大脇, 信学技報, EA2011-103, 43-48, 2011.



Fig. 6 We are planning to make a smart mixer implemented as a plug-in for DAW systems.