

スマートミキサー

— 新しい音信号混合法 —

大脇 渉[†] 高橋 弘太[†]

[†] 電気通信大学 情報理工学研究科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: †{owaki,kota}@ice.uec.ac.jp

あらまし スタジオ技術者として経験のない個人が、プロ並み、さらにはそれをも超えるミキシングを行うための新しいミキシング法として、スマートミキサーを提案する。従来のミキサーは、単なる加算器であった。そこでプロは、豊富な経験に基づき、加算器の前後に様々な処理を施すことにより、目的とする混合信号を得てきた。これに対し、スマートミキサーでは、入力信号を時間周波数平面に展開し、平面どうしの非線形な重ねあわせにより、混合信号を得る。この原理により、従来では不可能であった、きめ細かなミキシングが可能となる。本報告では、従来法として、プロがどのようなテクニックを駆使してミキシングを行っているかを紹介したのち、スマートミキサーにより、それを超えるミキシングが実現できることを示す。さらに、具体的な実装例として、振幅操作によるスマートミキサーの動作例を示す。

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

Smart Mixer

— A Novel Method for Mixing Sound Signals —

Wataru OWAKI[†] and Kota TAKAHASHI[†]

[†] Graduate School of Informatics and Engineering, The University of Electro-Communications,
Chofugaoka 1-5-1, Chofu-shi, Tokyo, 182-8585, Japan

E-mail: †{owaki,kota}@ice.uec.ac.jp

Abstract We propose a novel method for mixing sound signals, which is called “smart mixer”. Smart mixer will provide post professional quality mixing for an amateur. A conventional mixer is just an adder. In the case of professional mixing, a variety of processes have to be utilized in order to get the fine output signals. In contrast, with a smart mixer, user can reduce labor for selecting and adjusting many kind of effectors. Smart mixer makes this possible using the mixing at time-frequency planes. Furthermore, smart mixer evolves fineness of mixing process, with this principle. In this paper, some of conventional mixing techniques and results of simple application of smart mixer are described, as the first report of smart mixer.

Key words mixing, mastering, masking, time-frequency representation

1. はじめに

音信号のミキサーを高度化する試みとして、スマートミキサーを提案する。本稿では、スマートミキサーの着眼点を紹介し、簡易なアルゴリズムによるミキシング効果を報告する。

これまでの音信号のミキサーは、単純な加算器であった。加算器による混合では、入力信号の一部が聴き取れない場合がある。これは、時間周波数平面上での成分の分布どうしがぶつかりあって、信号のスペクトルの特徴が壊れてしまうのが原因で

ある。

そこで、これに対処するため、レコーディングスタジオなどで現在行われている手法は、ミキサーの前後に、コンプレッサ、イコライザなどの処理を追加する方法である。追加した処理によって、時間周波数平面上での成分の分布を操作し、出力信号における複数の入力信号の聞こえ方を調整する。つまり、追加する処理の選択が、ミキシングの成功の鍵を握っていると言える。

さて、この音信号のミキシングには、画一化された方法は存

在しない。現状では、熟練者それぞれが各自で蓄積した経験と勘によって、処理が施されている。このため、ミキシングは、一般ユーザにとって難解なものとなっている。最適なミキシングとは、究極的には個人の好みである [1, 2] ことも、ミキシングの難しさの一因である。

多様なミキシングの処理について、アルゴリズム化する試みも為されてきた [3-6]。しかし、いずれもミキサーの前後に付加する処理についての研究であり、ミキサー自体を高度化する試みはなされてこなかった。

そこで我々は、スマートミキサーを発想した。スマートミキサーは、加算器自体に入力信号を解析し処理する機構を持たせることで、加算器自体の高度化を行う。

加算器自体が高度化されることで、音信号の緻密なミキシングを、より多くの場面で実施できるようになる。特に一般ユーザが、高度なミキシングを手に入れることができる意味は大きい。さらに熟練者にとっては、スマートミキサーを処理のあたりをとる場面で用いることができ、これにより作業効率の向上と、デジタル信号処理ならではの緻密な作業を提供することができる。

2. 音信号のミキシング

本節では、従来の音信号のミキシングについて紹介する。まず、音信号のミキシングの目的を、プロフェッショナルの証言をもとに整理する。次に、音信号のミキシングの処理例を紹介する。最後に、これらを踏まえて、スマートミキサーの発想を導出する。

2.1 ミキシングの目的

ミキシングの目的について、プロのミキシングエンジニアである山内 [7] は、“主役をはっきりさせる”としている。そして、主役をはっきりさせるためには、マスキングを解消する必要があるとしている。

マスキングとは、聴感特性の一種で、ある周波数の成分が、時間方向、周波数方向にある近い成分を聴こえなくしてしまう現象 [8] を指す。マスキングには、周波数方向の「同時マスキング」と、時間方向の「継時マスキング」とがある。つまり音信号は、聴感特性において時間周波数平面上での拡がりを持っている。

マスキングについて、ミキシングのプロフェッショナルの間では、モノラルでのミキシングのとき、マスキングの影響がもっとも顕著となることが知られている [1]。そこで本報告では、モノラル信号について検討を行う。

マスキングを解消する処理は、時間周波数平面上において成分の分布どうしの相互の干渉を解消することと等しい。この処理は、時間周波数平面上で画像処理的にゲインマスクを生成することと一致する。

2.2 ミキシングの処理例 1: 単純加算 (図 2(a))

最も単純なミキシングでの処理は単純加算である。

この方法の利点は、まず処理が単純だということである。しかし、マスキングにより音の聞き取りが妨害されることが、致命的である。

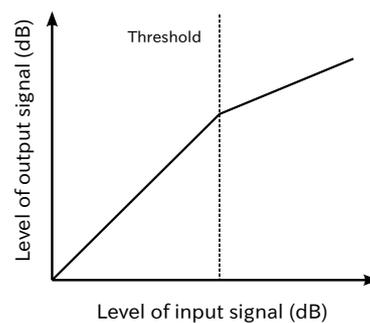


図 1 コンプレッサの基本的な特性

Fig. 1 The basic characteristics of the compressor.

2.3 ミキシングの処理例 2: 信号の切り替え (図 2(b))

単純な足し算からの簡単な工夫として、出力信号とする入力信号の切り替えが挙げられる。例えば、カーナビゲーションシステムにおいて案内音声再生の際に、既に BGM が再生されていたとしよう。このとき、一旦 BGM の音量を切り、できた無音区間で音声再生するというのが、この方法での動作である。割り込む音声は確実に聞こえるが、BGM を楽しんでいたユーザに不快感を与える。

2.4 ミキシングの処理例 3: 音量差の付与 (図 2(c))

例えばラジオ放送で、話者によるナレーションと BGM とをミキシングする際に用いられる手法である。このミキシングでは、実質的に入力信号どうしに優先度の差が存在する。ナレーションが主役であり、BGM は脇役である。そこで、主役を十分に聞かせるために、脇役の音量を主役に対して小さくする。つまり、主役と脇役とに音量差を付与している。

このミキシングを実現するために、加算器への入力の前に乗算器を接続する。主役の入力信号の時間平均エネルギー、もしくはピーク値に応じて、脇役への乗算器の係数を時変とする。このような処理は、一般にダッカー (Ducker) と呼ばれている。また、処理対象の信号とは別の信号の情報によって、処理を制御する方法を、サイドチェイン (side-chain) と呼んでいる。

ナレーションと BGM との最適な音量差について、小森ら [4] は、“ナレーションの重み付ラウドネスレベルが背景音より 9 phon 程度大きければ、最適レベルと判断する”としている。

2.5 ミキシングの処理例 4:

非線形関数を用いた音量制御 (図 2(d))

入力信号どうしの音量差を調整する上で問題になるのが、各々の入力信号が時間によってエネルギーが変動することである。そこで、音量差を与える乗算器の前に、入力信号のダイナミックレンジを調整する非線形関数を追加することが行われる。ここで用いられる非線形関数は、一般にコンプレッサ (Compressor) と呼ばれている。コンプレッサの基本的な特性を、図 1 に示す。コンプレッサは、入力信号の局所的な時間平均エネルギーが、しきい値以上である時区間で駆動する。駆動する時区間において、入力信号を設定した比率でエネルギーを圧縮する。

非線形関数での処理の問題点として、パラメータの設定によっては、音質を著しく損なうことがあげられる。非線形関数では、スペクトルの考慮なしに振幅を時間軸で変化させる。こ

れによって、波形そのものが書き換えられ、同時にスペクトルが変化するためである。

この問題点を解決する方法として、周波数帯域ごとに異なるパラメータを設定するという方法がある。これを一般に、マルチバンドコンプレッサ (Multi Band Compressor) と呼ぶ。

2.6 ミキシングの処理例 5:

線形フィルタを用いた音色調整 (図 2(e))

多くの音信号は、固有のスペクトルの形状を持っており、音色を決定づける特徴的な周波数帯域というものが存在する。

しかし、数多くの信号を加算することで、各入力信号の特徴的な周波数帯域どうしが被り、音色が判別できなくなってしまう。出力信号において、各入力信号の特徴を生かすためには、工夫が必要である。

そこで、各入力信号のスペクトルの特徴が引き立て合うように、線形フィルタ (イコライザ, Equalizer) を用いて音色を調整する。この手法は、2 信号にかかる周波数特性が鏡に写したように逆になることから、“Mirrored equalization [1]” と呼ばれている。

2.7 スマートミキサーへ (図 2(f))

処理例 4 で振幅への処理を、処理例 5 で周波数領域での処理を紹介した。いずれの処理でも、時区間によって処理の効果が一律でないという問題点がある。これは、入力信号非定常であることが原因である。

この問題点を解決するために、我々は振幅への処理では周波数帯域別に、周波数領域での処理ではエネルギー別に、それぞれパラメータを調整することを考えた。

3. スマートミキサー

2 節において、ミキシングで行われる処理が、処理が緻密になればなるほど、時間周波数平面上での分布の操作という、一点に集約されことを示した。そこで、加算器自身にアルゴリズムとして、時間周波数平面上での分布を操作する能力を与えることを発想した。

この新しい発想の加算器が、スマートミキサーである。そしてスマートミキサーを、図 2(f) で用いた記号で表すこととした。

4. 局所抑制法

スマートミキサーの原理については、[9] でも報告を行う予定だが、本稿では、振幅処理によってスマートミキサーの効果の確認を行った結果を報告する。簡単な振幅処理によるスマートミキサーの実装例が、局所抑制法である。

局所抑制法は、モノラル 2 信号を入力とし、モノラル 1 信号を出力とする。そして、2 つの入力信号の間に優先度の差があるとすると、優先度の高い入力信号が“主役”であるとし、優先度の低い入力信号を抑制することで、「“主役”をはっきりさせる」ことを狙う。

従来のミキシング法との関連性を表現すると、“サイドチェインによって駆動するマルチバンド・ダッカーを用いた処理の緻密化と一般化”である。パラメータの設定によって、従来のダッカーと同様の処理も実現できる。

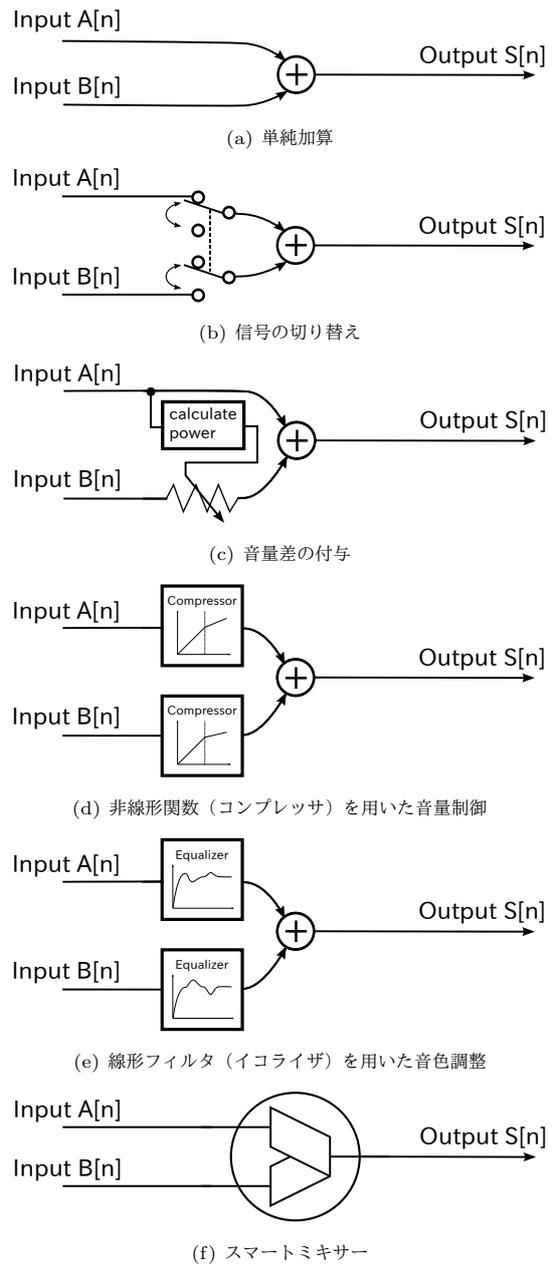


図 2 単純加算からスマートミキサーへ
Fig. 2 Improvement from a simple adder to the smart mixer.

4.1 局所抑制法の一般形

局所抑制法の一般形を、図 3 に示す。入力信号を優先度の高いものから順に $A[n]$, $B[n]$ とし、出力信号を $S[n]$ とする。ここで n は時刻 (サンプル番号) である。 $X_A[i, k]$, $X_B[i, k]$, $X_S[i, k]$ は、各音信号の時間フレーム番号 i , 周波数ビン番号 k の時間周波数平面上の位置 (i, k) での、短時間フーリエ変換値である。

関数 U は、対応する音信号の時間周波数表現の近傍領域の値の集合 $U(X_A, i, k; m_1, m_2, n_1, n_2)$, $U(X_B, i, k; m_1, m_2, n_1, n_2)$ を生成するものとする。時間周波数平面上の点 (i, k) の、近傍領域 $U(X, i, k; m_1, m_2, n_1, n_2)$ の領域の概略を、図 4 に示す。ここで、 m_1 , m_2 , n_1 , n_2 は、近傍領域の範囲を決定するパラメータである。そして、 $U(X_A, i, k; m_1, m_2, n_1, n_2)$,

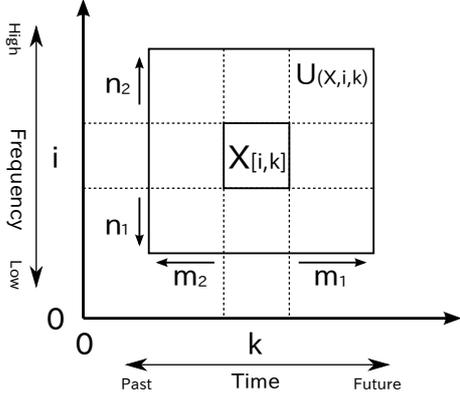


図4 局所抑制法における，時間周波数平面上での近傍領域

Fig.4 The rectangular area of the neighborhood on the time-frequency plane for the local suppression method.

$U(X_B, i, k; m_1, m_2, n_1, n_2)$ の振幅値に， (i, k) からの距離に応じた重み関数をかけた上で最大値をとることで，位置 (i, k) におけるエネルギー値 $L_A[i, k]$ ， $L_B[i, k]$ を算出する．この近傍領域への処理は，聴感特性であるマスキングの考慮を狙ったものである．

$L_A[i, k]$ ， $L_B[i, k]$ より，優先度の低い音信号への抑圧係数を決定する非線形関数を， $F(L_A[i, k], L_B[i, k])$ とする．非線形関数 $F(L_A[i, k], L_B[i, k])$ により算出される，優先度の低い音信号 $X_B[i, k]$ への抑圧係数を $W_B[i, k]$ とする． $W_B[i, k]$ により抑制された優先度の低い音信号の時間周波数表現と，優先度の高い音信号の時間周波数表現とを足しあわせた信号を，出力信号の時間周波数表現 $X_S[i, k]$ とする．

5. 実験と考察

まず，スマートミキサーの研究のはじめとして，図3で示した局所抑制法の簡単な処理を組み，どれほどの効果が得られるかの確認をすることとした．今回は，C++言語を用いて計算機上で実装した．以下，結果を報告する．

5.1 実験諸元

本稿では，時間周波数平面上で抑制すべき成分についての解析を行った結果を報告する．

実験素材として，音声とBGMそれぞれ二つのデータを用意した．音声には，話速バリエーション型音声データベース:SRV-DB [10] より，男性と女性の発話データを一文ずつ用いた．以後，男性のデータをPM00，女性のデータをPF00と表す．BGMには，クラシック(BGM1)とジャズ(BGM2)を選択した．そしてこれら4つの素材について，“PM00とBGM1”，“PM00とBGM2”，“PF00とBGM1”，“PF00とBGM2”の，4つの組み合わせのミキシングを試した．

まず，各素材の有音部の二乗平均値について，音声，BGMそれぞれについては等しくなり，BGMが音声より12 dB大きくなるよう，レベルを調整した．このとき単純加算では，すべての組み合わせで音声を聞き取ることができない．ここで4つの実験素材の有音部は，図5のように指定した．(a)，(b)の波形の上部のひらがなは，該当する箇所での発話内容である．

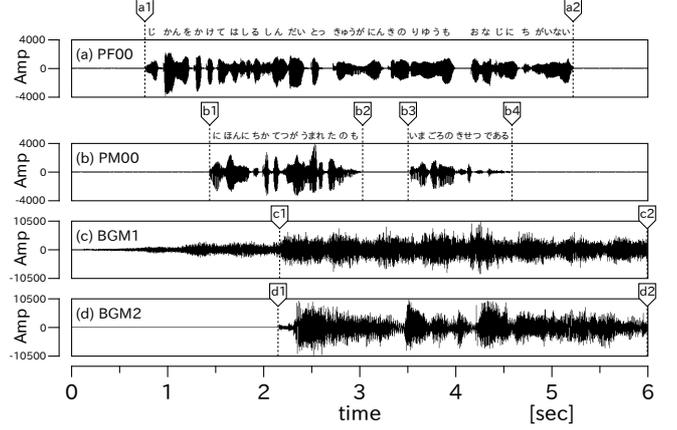


図5 実験素材の波形

Fig.5 The waveforms of source sound signals.

本実験において，局所抑制法の非線形関数 $F(L_A[i, k], L_B[i, k])$ は，最大抑圧量を w_{Max} とし，エネルギー値 $L_A[i, k]$ ， $L_B[i, k]$ ，周波数の3つの関数として， $X_B[i, k]$ への抑圧量 $W_B[i, k]$ を決定することとした．つまり抑圧量 $W_B[i, k]$ は，3次元空間中の位置で決定される． $L_A[i, k]$ ， $L_B[i, k]$ ．周波数それぞれに4つずつのパラメータを与え，図6の特性で，抑圧量 w_A ， w_B ， w_f を算出する． $W_B[i, k]$ は，式(1)~(4)によって算出する．

$$w_f = \begin{cases} \frac{i - l_{f1}}{l_{f2} - l_{f1}} \cdot w_{Max} & l_{f1} \leq i < l_{f2} \\ w_{Max} & l_{f2} \leq i < l_{f3} \\ \frac{l_{f4} - i}{l_{f4} - l_{f3}} \cdot w_{Max} & l_{f3} \leq i < l_{f4} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$w_A = \begin{cases} \frac{L_A[i, k] - l_{A1}}{l_{A2} - l_{A1}} \cdot w_f & l_{A1} \leq L_A[i, k] < l_{A2} \\ w_f & l_{A2} \leq L_A[i, k] < l_{A3} \\ \frac{l_{A4} - L_A[i, k]}{l_{A4} - l_{A3}} \cdot w_f & l_{A3} \leq L_A[i, k] < l_{A4} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

表1 スマートミキサーの実験諸元

Table 1 The parameters of a smart mixer.

サンプリング周波数	F_s	44.1 kHz
入力信号 A	$A[n]$	音声 (PM00:日本に地下鉄が生まれたのもいまごろの季節である PF00:時間をかけて走る寝台特急が人気の理由も同じに違いない)
入力信号 B	$B[n]$	BGM (BGM1:ベートーベン作曲交響曲第9番 BGM2:A 列車で行こう)
FFT 点数	N_{FFT}	4096 点
解析時ウィンドウ関数	$H[n]$	ハンニング窓形 2048 点
合成時ウィンドウ関数	$G[n]$	ハンニング窓形 1024 点
フレームシフト	N_{SFT}	512 点

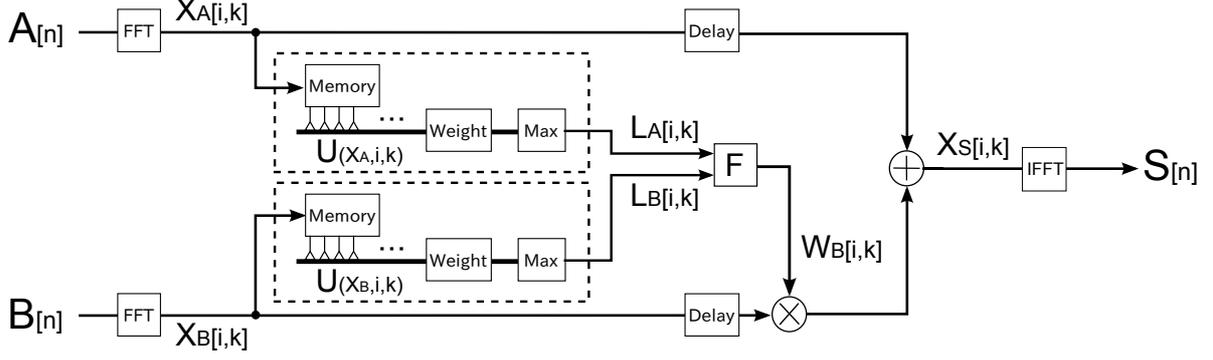


図3 簡単なスマートミキサーである、局所抑制法のブロック図

Fig.3 A block diagram of a simple smart mixer. (local suppression method)

$$w_B = \begin{cases} \frac{L_B[i, k] - l_{B1}}{l_{B2} - l_{B1}} \cdot w_f & l_{B1} \leq L_B[i, k] < l_{B2} \\ w_f & l_{B2} \leq L_B[i, k] < l_{B3} \\ \frac{l_{B4} - L_B[i, k]}{l_{B4} - l_{B3}} \cdot w_f & l_{B3} \leq L_B[i, k] < l_{B4} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$W_B[i, k] = \max(w_A, w_B) \quad (4)$$

なお本稿では、近傍領域の考慮は行っていない。

5.2 スマートミキサー実施例

先述の素材の組に対して処理を行ったなかから、比較的良好な結果が得られたな処理例の一つを紹介する。

まず、予備実験の結果、次の3つの知見を得ている。第1に、優先度の高い信号と低い信号とのエネルギー差に着目すべきである。第2に、周波数毎のエネルギーの分布に即して、抑圧量を調整すべきである。第3に、約50 Hzより低い帯域では、処理を行わないことで、良好な聴感が得られる。

そこで、表2の9つの周波数帯域に分けて、パラメータを調整することとした。このうち第3帯域を例にとって紹介する。第3帯域以外の帯域についても、同様の処理を行っている。

まず図7に、第3帯域において $L_A[i, k]$ を横軸に、 $L_B[i, k]$ を縦軸にとった散布図を示す。図7中の斜めの破線は、 $L_A[i, k]$ と $L_B[i, k]$ のエネルギー差が0 dBのラインである。このラインより上部の時間周波数分布では、優先度の高い信号のエネルギー $L_A[i, k]$ が、優先度の低い信号のエネルギー $L_B[i, k]$ よりも小さい。

図7中の四角枠内の時間周波数分布に対して、非線形関数 $F(L_A[i, k], L_B[i, k])$ を、優先度の低い信号への最大抑圧量

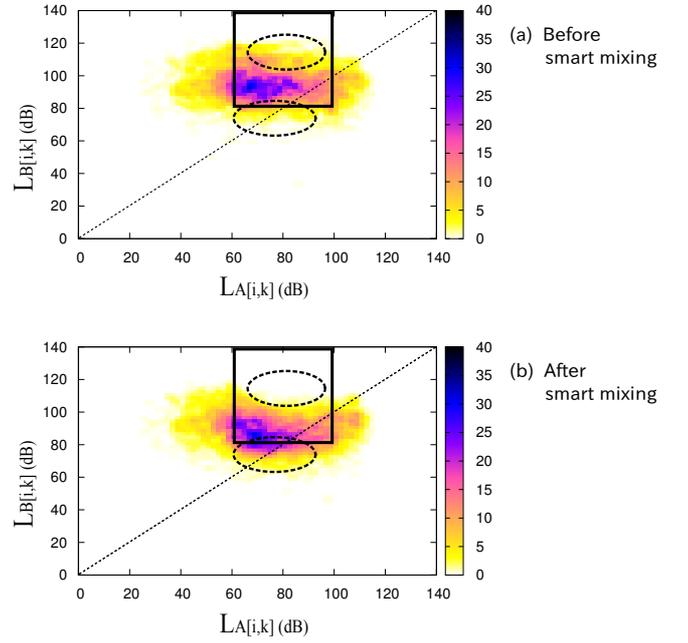


図7 第3帯域(250 - 500 [Hz])での、 $L_A[i, k]$ と $L_B[i, k]$ の散布図
Fig.7 The scattering plots of $L_A[i, k]$ and $L_B[i, k]$ of the 3rd frequency band (250 - 500 [Hz]).

w_{Max} を18 dBとして駆動させた。この処理による四角枠内の時間周波数分布の図7中での変化は、下方向に最大18 dB下方への移動である。例えば、図7中の2つの破線楕円枠で変化がみとれる。上方にある枠内での密度は、処理前と処理後と比較して、減少している。対して、下方にある枠内では、処理前と処理後で、密度の上昇がみられる。このとおり、データの集中する位置が、図中の下方に移動している。これは、時間周波数平面上の位置 (i, k) で、優先度の高い信号のエネルギーが優先度の低い信号のエネルギーに対して、小さい値となっている箇所が減っていることを意味する。

この設定で処理をした結果、他の設定に対して比較的良好な結果が得られた。単純加算では聞き取りができなかった音声も、聞き取れるようになったとともに、BGMの劣化が比較的少量となった。ただし、BGMの音色の印象は、元信号に対して変化が感じられた。

表2 9つの周波数帯域の設定

Table 2 A table of parameters with 9 frequency bands.

帯域番号	周波数帯域 [Hz]	帯域番号	周波数帯域 [Hz]
1	0 - 125	6	2500 - 5000
2	125 - 250	7	5000 - 10000
3	250 - 500	8	10000 - 20000
4	500 - 1250	9	20000 - 22050
5	1250 - 2500		

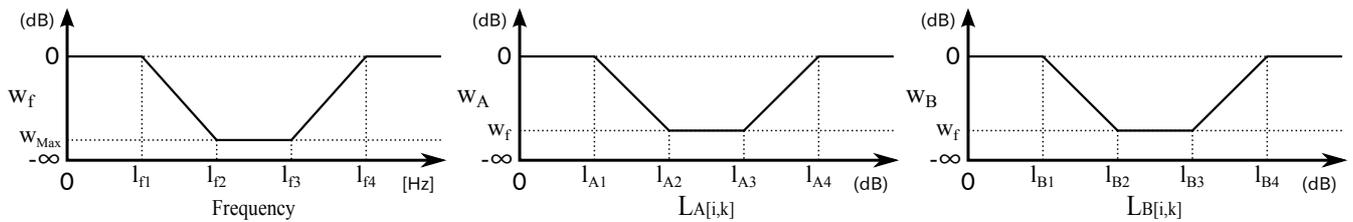


図 6 非線形関数 $F(L_A[i, k], L_B[i, k])$ がもつ 13 のパラメータ間の関係
 Fig. 6 A relationship among 13 parameters for the nonlinear function $F(L_A[i, k], L_B[i, k])$.

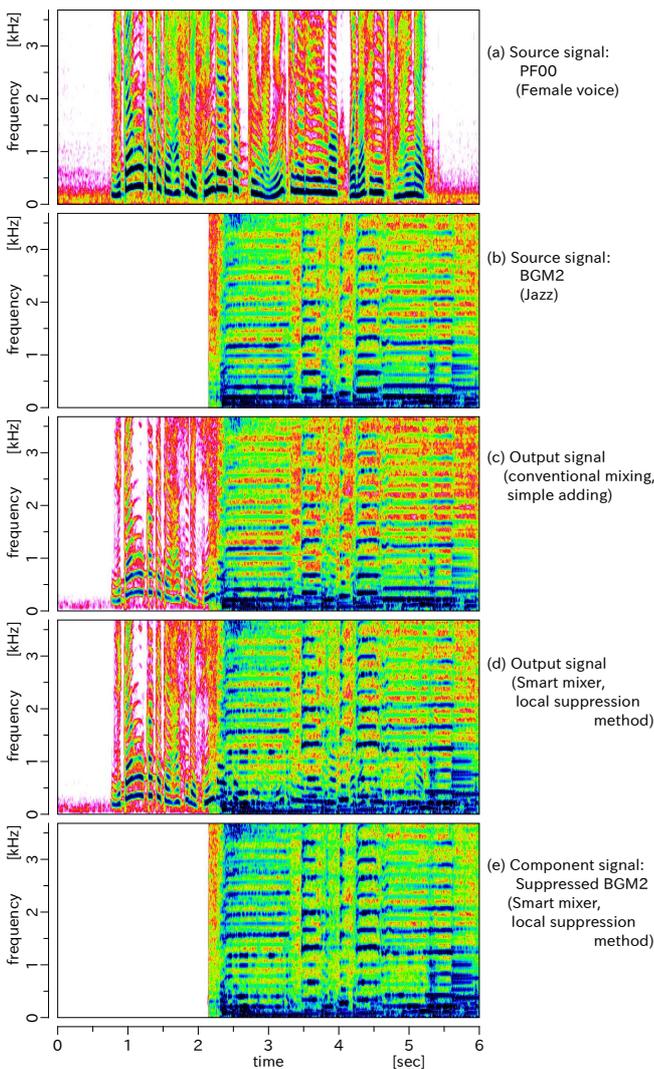


図 8 処理前後の信号の時間周波数平面
 Fig. 8 The time-frequency maps of conventional mixing and smart mixing.

図 8 に、二つの入力信号、単純加算、処理結果、抑制後の優先度の低い信号の時間周波数平面のうち、0 から 3 kHz までの範囲を示す。単純加算では、音声の線スペクトル構造の特徴が際立っておらず、ほとんど音楽に埋もれている。対して、処理後の信号では、音声の線スペクトル構造がみられるとともに、BGM の線スペクトル構造が削られている様子がみられる。処理によって、BGM の時間周波数分布を、音声のスペクトル構造を崩さないように、変更することができたといえる。

処理結果の向上のために、パラメータ設定の細分化、近傍領域の考慮、ピーク検出等の信号の意味的情報による処理の細分化といった改良を、今後検討していく。

6. まとめ

本稿では、新しい音信号混合法であるスマートミキサーを提案した。スマートミキサー研究の第一報として、スマートミキサーの着眼点を、従来のミキシング法を交えて示した。

スマートミキサーの実施例としては、簡易な振幅処理である局所抑制法を示し、実際の音信号を用いて効果を検証した。実験は、音声と BGM の組み合わせについて行った。用いた音信号の調整として、音声と BGM の音量差を、有音部分どうしの時区間で、平均して BGM が 12 dB 大きい値とした。この実験素材について、単純加算と比較して聴感が良好な出力結果が得られた。

今後の課題として、振幅のみならず位相操作も含んだアルゴリズムの検討、パラメータ設定の自動化、あらゆる種類の入力信号への対応、アルゴリズムの多チャンネル対応、リアルタイム処理の実現、聴感特性との対応関係の解析、といったものが挙げられる。

文 献

- [1] R. Izhaki, "Mixing Audio Concepts, Practices and Tools," Focal Press, 2008.
- [2] 社団法人日本音楽スタジオ協会, "サウンドレコーディング技術概論改訂版," 兼六館出版, 2008.
- [3] 谷井章夫, 後藤真孝, 片寄晴弘, "ミックスダウンデザインの抽出と適用," FIT2003(情報科学技術フォーラム) 情報技術レターズ Vol.2 2003, pp.109-110, 2003.
- [4] 小森智康, 小宮山撰, 壇寛弥, 都木徹, 庄田清武, 黒住幸一, 星英明, 村川一広, "ラウドネスを指標とした音声ミキシングバランスに関する調査," 信学論 (A), vol.J92-A No.5 pp344-352, 2009.
- [5] E. Perez-Gonzalez, J. Reiss, "Automatic Equalization of Multichannel Audio Using Cross-Adaptive Methods," Audio Engineering Society Convention 127, 10, 2009.
- [6] A. Tsilfidis, C. Papadakos, J. Mourjopoulos, "Hierarchical Perceptual Mixing," Audio Engineering Society Convention 126, 5, 2009.
- [7] 山内"Dr." 隆義, "ミックス・ダウン研究所," Sound & Recording Magazine, Vol.29(1), pp.154-155, 2010.
- [8] 岩宮眞一郎, "図解入門よくわかる最新音響の基本と仕組み," 秀和システム, 2007.
- [9] 高橋弘太, 大脇渉, "スマートミキサー ~基本原理と有効性~, " 信学技法, 本件の次の発表, 2011.
- [10] "話速バリエーション型音声データベース:SRV-DB," <http://www.it.uec.ac.jp/SRV-DB/>, (参照 2011-10-27).