

時間周波数平面を用いた音声の埋もれない音信号混合法の評価

大脇 渉^{†a)} 高橋 弘太[†]

Evaluation of the Sound Signal Mixing Method without Hiding the Voice
Using the Time-Frequency Plane

Wataru OWAKI^{†a)} and Kota TAKAHASHI[†]

あらまし 本論文では、音のミキシングに時間周波数平面を用いることの有効性を示す。著者らは、時間周波数平面を用いた新しい音信号混合法を提案してきた。時間周波数平面を用いることで、各入力信号の特徴を捉え、信号間の特徴の相互関係に沿った緻密な処理が実現できる。この考えに立脚したミキシング手法として、音声とBGM(Background Music)とのミキシング時に、音声の埋もれない音信号混合法を提案してきた。本論文では提案法が時間周波数領域での処理であることの有効性を、聴取実験により検証する。具体的には、提案法での変化量を、周波数領域、時間領域での処理で近似し、了解度と主観評価を比較して評価する。これらの検討により、時間周波数平面を用いた音信号混合法の有効性を示す。

キーワード ミキシング、時間周波数平面、短時間フーリエ変換、聴覚心理モデル、音響信号処理

1. ま え が き

本論文では、音のミキシングに時間周波数平面を用いることの有効性を示す。著者らは、時間周波数平面を用いた新しい音信号混合法 [1], [2] を提案してきた。時間周波数平面を用いることで、入力信号の特徴を捉え、信号間の特徴の相互関係に沿った緻密な処理を実現できる。この考えに立脚したミキシング手法として、音声とBGMのミキシングにおいて、音声の埋もれない混合法 [3]~[12] を提案してきた。ここで、音声の埋もれないとは、カーステレオの音楽へのナビ音声の重ね合わせや、館内BGMへのアナウンスの重ね合わせ等の場面において、音声の聴きとりやすさが確保されていることを指す。

従来の音のミキシング手法は、時間領域での処理(ダッカー [13])と周波数領域での処理(イコライザ [14])とそれらの組合せ [14]~[18] であり、時間周波数平面の任意の点を独立に調整することはできなかった。本手法ではそれが可能となるため、スペクトル遷移パターンを捉えた処理も実現できる。

本論文では、[12]の提案法について、時間周波数領域での処理の有効性を検討する。検討に際し、従来法に相応の手法として、提案法が生成する時間周波数領域での変化量を、周波数領域、時間領域での変化量でそれぞれ近似する方法を、イコライザ近似、ダッカー近似として定義する。これらに単純加算と提案法を加えた4種類の混合法について、聴取実験による了解度と主観評価と比較することで、周波数領域単独の処理、時間領域単独の処理では不十分であることを示す。

2. で音声の埋もれない音信号混合法 [12] を紹介する。3. で提案法での変化量の、時間領域、周波数領域での近似処理について述べる。4. で聴取実験について述べる。5. で考察を述べる。そして 6. をむすびとする。

2. 音声の埋もれない音信号混合法

音声とBGMをミキシングしたとき、音声の埋もれない音信号混合法として、著者らは時間周波数領域でゲイン調整を行う手法 [12] を提案した。

処理方法の概要を図1にまとめる。まず、入力音声、入力BGM、出力、それぞれの複素時間周波数表現を $X_A[i, k]$, $X_B[i, k]$, $Y[i, k]$ とする。ここで、 i は時間フレーム番号、 k は周波数ビン番号である。本手法では、ゲインマスク $W_A[i, k]$, $W_B[i, k]$ を用いて出力 Y

[†] 電気通信大学、調布市

The University of Electro-Communications, 1-5-1 Chofugaoka,
Chofu-shi, 182-8585 Japan

a) E-mail: wataru@uec.ac.jp

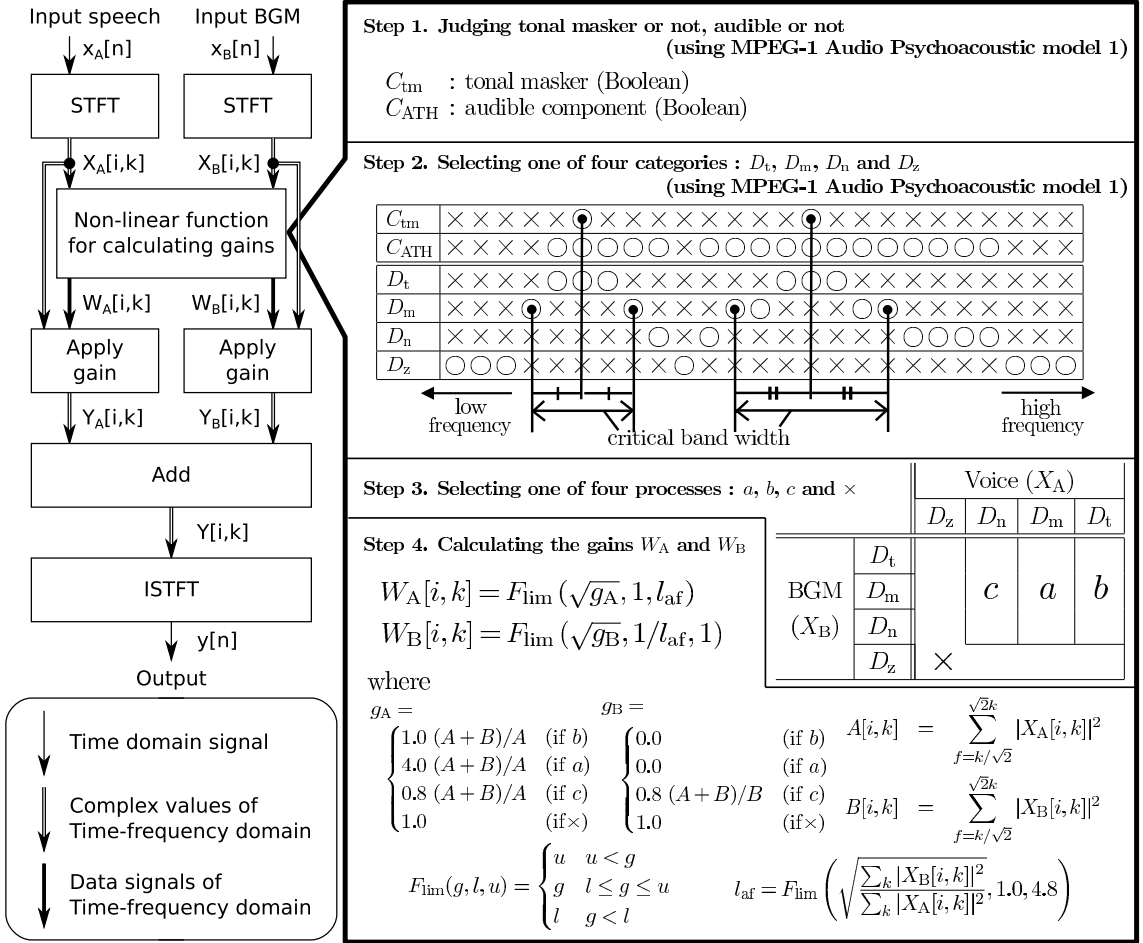


図1 音声 が埋もれない音声信号混合法のブロック図及びゲイン生成部における処理の概要
 Fig.1 Block diagram of the proposed method and overview of the processing for calculating gains.

を $Y = W_A X_A + W_B X_B$ で生成する。

ゲインマスクを決めるにあたって, Step 2 では, 各 $[i, k]$ について $X_A[i, k]$, $X_B[i, k]$ を以下の四つのカテゴリー, 純音 D_t , 純音近傍 D_m , 非純音 D_n , 非可聴 D_z に分類する。その分類のために, Step 1 では, MPEG-1 Audio Psychoacoustic model 1 [19] で採用されている純音判定 C_{tm} と有音判定 C_{ATH} の基準を利用している。

Step 3 では, 音声と BGM の四つのカテゴリーの組合せで決まる $16 (= 4 \times 4)$ 属性のそれぞれに対して対応表 (図 1 の Step 3 欄を参照) を使って, 四つの処理 a, b, c, \times を対応させる。

これら四つの処理について順に説明する。まず, \times

では少なくとも片方の入力が無音であるから単純加算が良い。次に, a, b, c に対するゲインマスク設定法について述べる準備として, X_A, X_B の $[i, k]$ におけるパワーを周波数方向に平滑化したものをそれぞれ $A[i, k]$, $B[i, k]$ とおく。出力パワーは, 互いの位相がランダムであるとすれば $P = W_A^2 A + W_B^2 B$ と見積もることができる。また, 出力パワー中の X_A 成分の割合は $Q = W_A^2 A / P$ で見積もることができる。我々は, P, Q を A, B の関数として適切に決定することで, 音声 が埋もれない混合ができるものと考えた。

P, Q は図 2 に従って決定する。まず, a では $P = A + B, Q = 1.0$ とし, 出力を強制的に音声化する。次に, b では $P = 4.0(A + B), Q = 1.0$ と

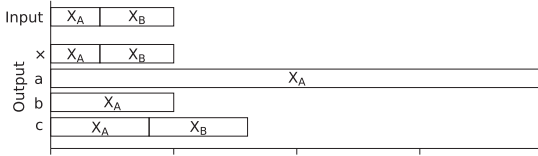


図2 P, Q の決定法. X_A, X_B は, 由来する入力信号を示す

Fig. 2 Determination of P and Q . X_A and X_B are labels showing one of input signals.

表1 実験諸元

Table 1 Parameters used in experiments.

Sampling frequency (F_S)	44100 Hz
Samples of FFT (N_{FFT})	1024
Samples of frame shift	384
Type of analyze window	Hanning
Samples of analyze window	1023
Type of synthesise window	Hanning
Samples of synthesise window	767

して, a に比較して P を 4.0 倍することで混合音においてもピークを保存することを狙った. 一方, c では音声は重要ではないため, $P = 1.6(A + B)$, $Q = 0.5$ とし, 音声と BGM を等音量化するに留める. なお, 係数である 4.0 と 1.6 は予備実験により決定した.

Step 4 では, 以上で求めた P, Q から W_A, W_B を逆算する. ただし, 音声の抑制及び BGM の強調が行われないよう, $W_A \geq 1.0, W_B \leq 1.0$ の条件をつける. また, 過度の音声の強調及び BGM の抑制が起こることを防止する幾つかの条件を設けているがここでは説明を省く. 提案法の諸元を表 1 に示す.

3. 時間領域, 周波数領域での近似処理

本章では, 提案法が時間周波数領域でゲイン調整を行っていることの利点を示すために, それを周波数領域に限定して近似したゲイン調整 (以下, イコライザ近似という) と, 時間領域に限定して近似したゲイン調整 (以下, ダッカー近似という) を定義する.

まず, 準備として $Y_A[i, k], Y_B[i, k]$ を, それぞれ ISTFT により時間信号 $y_A[n], y_B[n]$ に変換する. 次に, $y_A[n], y_B[n]$ を, 提案法と同一諸元 (表 1) の STFT で再解析し, その結果をそれぞれ $V_A[i, k], V_B[i, k]$ とする.

イコライザ近似では, まず, 時間周波数平面の観測により音声と BGM がともに有音であると判断したフレームの集合 \mathbb{I}^E について, 周波数ビンごとのパワーの総和が提案法に等しくなるゲイン $G_c^E[k]$ を以下に

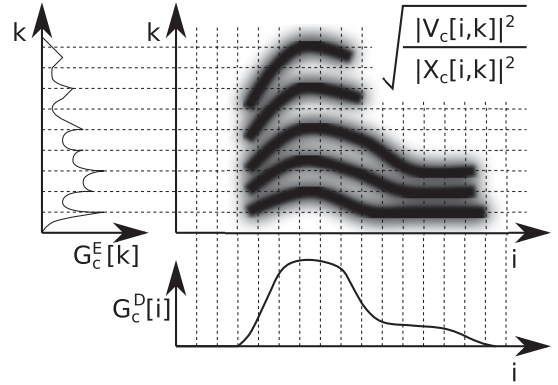


図3 $G_c^E[k], G_c^D[i]$ 算出の概念図

Fig. 3 The concepts of the calculation for $G_c^E[k]$ and $G_c^D[i]$.

よって求めておく.

$$G_c^E[k] = \sqrt{\frac{\frac{1}{I^E} \sum_{i \in \mathbb{I}^E} |V_c[i, k]|^2}{\frac{1}{I^E} \sum_{i \in \mathbb{I}^E} |X_c[i, k]|^2}} \quad (1)$$

ここで I^E は, \mathbb{I}^E の総フレーム数である. また, c は A または B である. G^E を用いて, イコライザ近似での出力信号の時間周波数表現 $Z^E[i, k]$ を,

$$Z^E[i, k] = G_A^E[k]X_A[i, k] + G_B^E[k]X_B[i, k] \quad (2)$$

として算出する.

ダッカー近似では, まず, フレームごとのパワーの総和が提案法に等しくなるゲイン $G_c^D[i]$ を以下によって求めておく.

$$G_c^D[i] = \sqrt{\frac{\frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} |V_c[i, k]|^2}{\frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} |X_c[i, k]|^2}} \quad (3)$$

G^D を用いて, ダッカー近似での出力信号の時間周波数表現 $Z^D[i, k]$ を,

$$Z^D[i, k] = G_A^D[i]X_A[i, k] + G_B^D[i]X_B[i, k] \quad (4)$$

として算出する.

$G_c^E[k], G_c^D[i]$ 算出の概念図を, 図 3 に示す.

以上のようにして求めた $Z^E[i, k], Z^D[i, k]$ を ISTFT することで, イコライザ近似とダッカー近似の出力信号を得る.

4. 聴取実験

単純加算 (A), 提案法 (S), イコライザ近似 (E),

表 2 了解度での、入力信号 3 セットの設定
Table 2 Characteristics of input signals for the hearing experiments of the speech intelligibility.

Set	Contents	Standard deviation	Volume difference
I1	Voice	PF02 (Female)	500
	BGM	No.5	4000
I2	Voice	PF02 (Female)	375
	BGM	No.15	4000
I3	Voice	PF02 (Female)	375
	BGM	No.94	6000

ダッカー近似 (D) の 4 手法について、聴取実験による了解度と主観評価で比較する。被験者は成人男性 12 人であり、ヘッドホン (DENON AH-D2000) によって音を提示した。

4.1 了解度の実験概要

入力信号は、表 2 に示す 3 セットを用いた。音声は「話速バリエーション型音声データベース SRV-DB [20]」の「発話のプロフェッショナルによるオリジナル原稿 (カーナビ文章) の読み上げ」から、09 ~ 32 文の 24 データを用いた。BGM は「RWC 研究用音楽データベース: ポピュラー音楽 [21]」に収録の音源を用いた。各セットでの音声と BGM との音量差は、ある一人の被験者に対して単純加算したときに音声が聴きとれないようにあらかじめ調整した。また、音声に SNR が -10 dB の白色雑音を付加した実験も行った。白色雑音はピンクノイズほど騒音を代表する信号ではないが、一部の騒音 (例えばフライパンによる調理時の音) に対しては類似性が高い。音声の 24 データのうち、09 ~ 20 文の 12 データに白色雑音を付加し、21 ~ 32 文の 12 データには白色雑音を付加せずを用いた。

サンプルは実験計画法により、各被験者が聴取する文の回数が 1 回ずつ重複せず、BGM 3 種類とミキシング手法 4 種類との組合せ数が等しくなるように振り分けた。被験者には、サンプルを 1 度だけ聴かせ、聴きとれた内容をひらがなで記入させた。了解度の算出は、サンプルごとに 3 段階 (2: 全音節正解, 1: 一部音節正解, 0: 全音節不正解) で得点づけし、全サンプルで全音節正解の場合に、100% となるよう正規化した。

4.2 主観評価の実験概要

主観評価は、Q1: “入力音声と比較して、内容の聴きとりやすさ”, Q2: “入力音声と比較して、音質の自然さ”, Q3: “入力 BGM と比較して、音質の自然さ” の三つの評価項目とした。評価値は、表 4 に示す 5 段階の DMOS (Degradation Mean Opinion Score)

表 3 主観評価での、入力信号 3 セットの設定
Table 3 Characteristics of input signals for the hearing experiments of the subjective evaluation.

Set	Contents	Standard deviation	Volume difference
E1	Voice	PF00 (Female)	500
	BGM	No.5	2000
E2	Voice	PM00 (Male)	1000
	BGM	No.15	4000
E3	Voice	PF01 (Female)	375
	BGM	No.94	6000

表 4 主観評価での、DMOS 値の設定
Table 4 Descriptions of degradation mean opinion scores (DMOSs) used in the hearing experiment of the subjective evaluation.

Point	Impairment
5	Degradation is inaudible
4	Degradation is audible but not annoying
3	Degradation is slightly annoying
2	Degradation is annoying
1	Degradation is very annoying

値 [22] とした。

入力信号は、表 3 に示す 3 セットを用いた。音声には「話速バリエーション型音声データベース SRV-DB [20]」の「発話のプロフェッショナルによる編集手帳 (読売新聞) の読み上げ」を用いた。4.1 の実験と異なる音声を使用したのは、カーナビ音声は文が短く抑揚が単調で、音質の自然さの評価が困難であるためである。一方、BGM には 4.1 の実験と同じものを用いた。各セットでの音声と BGM との音量差は、ある一人の被験者に対して単純加算したときに音声が聴きとれないようにあらかじめ調整した。更に入力信号の音声について、了解度と同じく SNR が -10 dB の白色雑音を付加した音源もあわせて用意した。全サンプルを全被験者に聴取させた。

4.3 了解度の実験結果

了解度の結果を図 4 に示す。

まずイコライザ近似では、単純加算からの改善が見られない。提案法とダッカー近似で、白色雑音の付加の有無にかかわらず 100% を達成した。提案法が了解度を向上させている要因として、時変なゲイン操作が大きな役割を果たしていることが示された。

4.4 主観評価の実験結果

主観評価の結果を図 5 に示す。

“音声の内容の聴きとりやすさ (Q1)” については、(a) より、提案法とダッカー近似が秀でている。一方、

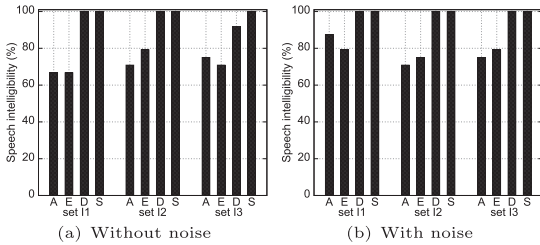


図4 了解度の聴取実験結果. 左 (a): 白色雑音付加なし, 右 (b): 白色雑音付加あり. A: 単純加算, E: イコライザ近似, D: ダッカー近似, S: 提案法

Fig. 4 Results of the hearing experiment of the speech intelligibility. A : Simple addition, E : Approximated equalizer, D : Approximated ducker, S : Proposed method.

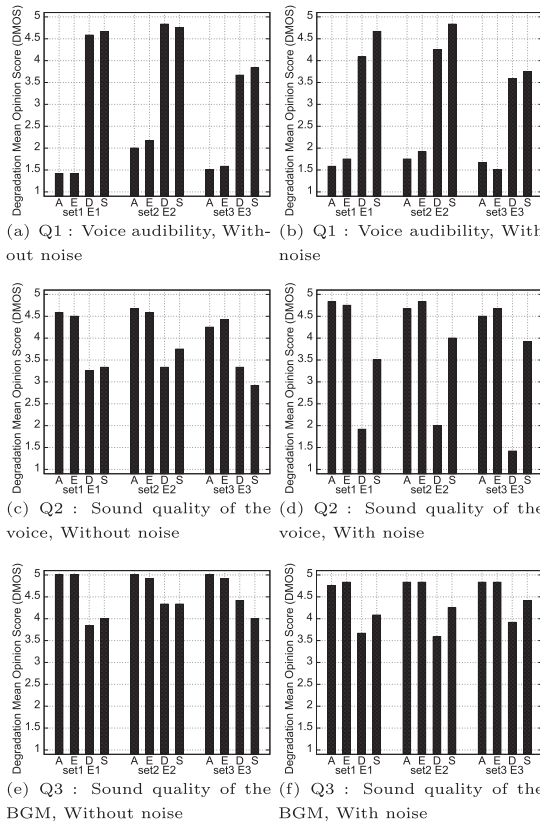


図5 主観評価の聴取実験結果. 上 (a,b): Q1 音声の内容の聴きとりやすさ, 中 (c,d): Q2 音声の音質の自然さ, 下 (e,f): Q3 BGM の音質の自然さ. 左 (a,c,e): 白色雑音付加なし, 右 (b,d,f): 白色雑音付加あり. A: 単純加算, E: イコライザ近似, D: ダッカー近似, S: 提案法

Fig. 5 Results of the hearing experiment of the subjective evaluation. A : Simple addition, E : Approximated equalizer, D : Approximated ducker, S : Proposed method.

白色雑音付加ありでの評価 (b) を (a) と比較すると, ダッカー近似でのみ全セットで評価が下がっている.

“音声の音質の自然さ (Q2)” については, (c) より, 単純加算とイコライザ近似では評価が高く, ダッカー近似と提案法では低い. 一方, 白色雑音付加ありでの評価 (d) を (c) と比較すると, ダッカー近似のみ評価が下がり, 提案法では逆に評価が上がっている.

“BGM の音質の自然さ (Q3)” については, (e) より, 単純加算とイコライザ近似では評価が高く, ダッカー近似と提案法では低い. 一方, 白色雑音付加ありでの評価 (f) を (e) と比較すると, 提案法のみ評価が上がり, ダッカー近似では逆に評価が下がっている.

Q2 と Q3 の比較から, 音声への白色雑音の付加による音質の自然さの劣化は, BGM の劣化としても評価されたことが分かる. 提案法では, 音声, BGM ともに評価が向上している. 逆にダッカー近似では, 音声, BGM ともに評価が低下している.

以上から, 時間領域でのゲイン操作により, 音声の了解度と内容の聴きとりやすさを得られることが示された. 周波数帯域ごとに処理を異にすることで, 音質劣化を低減できることが示された. そして, 両者を組み合わせた時間周波数平面での処理により, 聴きとりやすさと音質の両立が実現できることが示された.

5. 考 察

聴取実験の結果から得た知見を, 時間周波数平面での観察と照らし合わせる. 入力信号として表3のE3を用いたときの入出力信号の時間周波数平面を図6に示す (パワーが大きい部分を黒で表している). 図6の12枚の時間周波数平面のうち, 左の6枚は白色雑音を加えない場合の入出力信号であり, 上から入力音声 (CV), 入力 BGM (CB), 単純加算 (CA), イコライザ近似 (CE), ダッカー近似 (CD), 提案法 (CS) である. 右の6枚は白色雑音を加えた場合の入出力信号であり, 上から入力音声 (WV), 入力 BGM (WB, ただし CB と同一である), 単純加算 (WA), イコライザ近似 (WE), ダッカー近似 (WD), 提案法 (WS) である. これら12枚の図に対して考察の対象とする4領域 $\alpha, \beta, \gamma, \delta$ を破線で囲った. α には音声の母音認識で重要なフォルマントと周期成分, β には音声の話者性に関連するフォルマントと周期成分, γ には音声の非周期成分, δ には BGM のアタック成分が存在する. 各領域内での濃淡の模様が注目し, 入力信号の濃淡の模様が, 出力信号の濃淡の模様として明瞭に

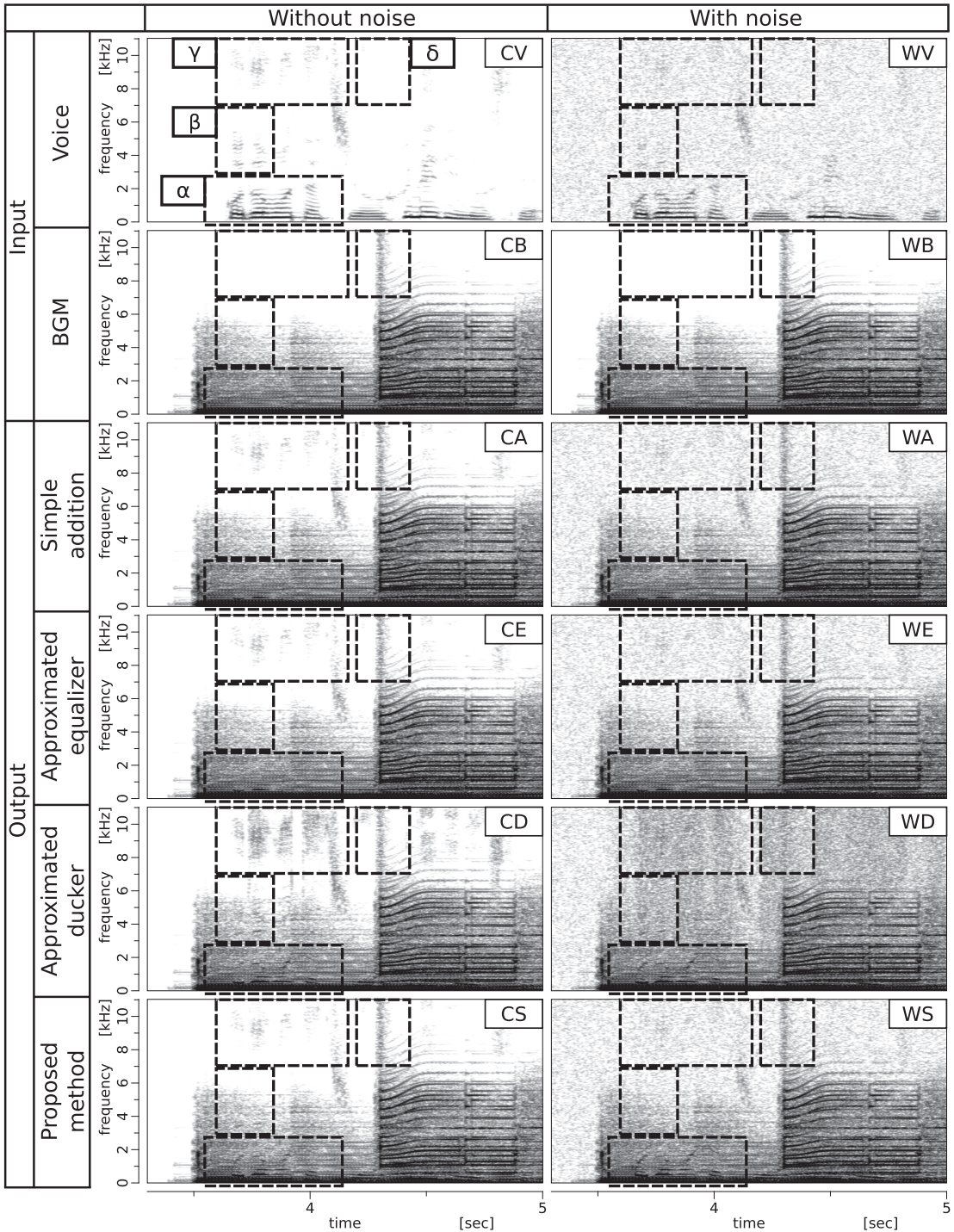


図 6 入出力信号の時間周波数平面. 右: 白色雑音付加なし, 左: 白色雑音付加あり. 上から, 入力音声, 入力 BGM, 単純加算, イコライザ近似, ダッカー近似, 提案法

Fig. 6 Time-frequency planes of input and output signals. Left : Without noise, Right : With noise. From top, input signals of the voice, input signals of the BGM, output signals of the simple addition, output signals of the approximated equalizer method, output signals of the approximated ducker method, and output signals of the proposed method.

表 5 図 6 の枠内における入出力信号間についての相関係数

Table 5 The correlation coefficients between input and output signals in the frames of Fig.6.

(a) α , Without noise			(b) α , With noise		
	CV	CB		WV	WB
CA	0.226	0.965	WA	0.237	0.952
CE	0.227	0.964	WE	0.239	0.951
CD	0.347	0.881	WD	0.415	0.830
CS	0.325	0.855	WS	0.439	0.787

(c) β , Without noise			(d) β , With noise		
	CV	CB		WV	WB
CA	0.467	0.948	WA	0.420	0.648
CE	0.467	0.948	WE	0.420	0.648
CD	0.612	0.822	WD	0.733	0.236
CS	0.491	0.929	WS	0.473	0.631

(e) γ , Without noise			(f) γ , With noise		
	CV	CB		WV	WB
CA	0.955	-0.150	WA	0.997	-0.038
CE	0.955	-0.150	WE	0.997	-0.039
CD	0.912	-0.193	WD	0.860	-0.097
CS	0.955	-0.151	WS	0.997	-0.039

(g) δ , Without noise			(h) δ , With noise		
	CV	CB		WV	WB
CA	0.049	0.929	WA	0.403	0.581
CE	0.048	0.929	WE	0.403	0.581
CD	0.241	0.781	WD	0.749	0.106
CS	0.049	0.928	WS	0.413	0.583

視認できれば、その成分は出力信号でも聴きとりやすくと考える。

あわせて、信号間での模様の比較を、数値指標でも確認する。第一の数値指標として、領域内の各点でのパワー (dB 表現) の相関係数を用いた (表 5)。第二の数値指標として、領域内での平均パワーを用いた (表 6)。

まず、 α と β に着目する。表 6 (a), (c) より、入力 CV に対する出力 {CA, CE, CD, CS} の相関は、CA の場合に比較して、CE では改善が見られず CD と CS では改善が見て取れる。これは、音声の内容の聴きとりやすさの結果 (図 5 (a)) と一致している。すなわち、イコライザ近似に対してのダッカー近似と提案法の優位性が数値指標によって裏付けられた。

次に、 γ に着目する。図 6 を見れば明らかに WD でのみ γ でのパワーが突出している。実際、平均パワー (表 6(b)) は WA に比べて約 10 dB ($\approx 48.4 - 38.1$) 増えている。一方、表 5(f) より WV との相関が WD でのみ下がっていることから、その音声は劣化してい

表 6 図 6 の枠内における入出力信号の平均パワー
Table 6 The average power of the input and output signals in the frames of Fig.6.

(a) Without noise (dB)					(b) With noise (dB)				
	α	β	γ	δ		α	β	γ	δ
CV	58.4	33.2	31.8	17.5	WV	58.4	38.6	38.1	37.4
CB	82.7	43.9	5.9	42.7	WB	82.7	43.9	5.9	42.7
CA	82.8	44.3	31.8	42.7	WA	82.8	45.1	38.1	43.8
CE	82.4	44.3	31.8	42.7	WE	82.3	45.1	38.1	43.8
CD	82.4	46.2	42.4	41.7	WD	82.1	50.2	48.4	48.9
CS	82.3	44.6	31.8	42.7	WS	82.0	46.7	38.1	43.8

る。これらの結果は、音声の音質の自然さ (図 4(d)) での結果と一致し、ダッカー近似に対しての提案法の優位性が数値指標によって裏付けられたことになる。

続いて、 δ に着目する。図 6 より BGM のアタック成分は WD でのみ消失している。表 5(h) より WB と WD の相関が 0.106 になっていることから、BGM が劣化していると確かめられた。これは、BGM の音質の自然さ (図 4(f)) での結果と一致し、ここでもダッカー近似に対する提案法の優位性が示されたことになる。

時間周波数平面での観察により、以下のことが分かった。まず、音声の聴きとりやすさの確保では、出力信号での音声の時間周波数平面での濃淡の模様の維持が重要であると考え、それは 4.4 の結果と一致した。一方、音質の評価の維持では、背景雑音を処理の適用外とすることが重要である。提案法では両者を達成できており、提案法のイコライザ近似、ダッカー近似では達成できない。よって、提案法での時間周波数領域での処理の有効性が示された。

なお、イコライザ近似、ダッカー近似での変化量は、提案法での時間周波数領域での処理に基づいて算出している。したがって、ダッカー近似が提案法と同等の評価になる場合があったとしても、計算量の観点でダッカー近似法は実用的ではない。

6. む す び

本論文で、音声と BGM のミキシングで音声埋もれない音信号混合法を実現する方法として、時間周波数領域でのゲイン調整が有効であることを示した。

音声と BGM のミキシングを、単純加算、提案法、イコライザ近似、ダッカー近似の 4 手法で実施し、聴取実験により了解度と主観評価を比較した。音声は、白色雑音を付加した音源とそうでない音源との 2 種類

で検討した。

白色雑音付加なしでは、提案法とダッカー近似が同等の良い評価を得た。一方、白色雑音付加ありでは、ダッカー近似では評価が低下した。時間周波数領域でのゲイン調整により、音声に混入する背景雑音レベルに依らず、良好な結果が得られることが示された。

また、時間周波数平面での観察と客観的な数値指標によって、聴取実験での評価を裏付けることができた。以上の実験により、音声の聴きとりやすさと音質の両方が、提案法によってのみ達成できることを明らかにした。すなわち、時間周波数領域での処理の有効性が示された。

謝辞 本研究は、JSPS 科研費 20500111, 23500147, 26330188 の助成を受けたものです。本研究では、RWC 研究用音楽データベース (ポピュラー音楽) より、RWC-MDB-P-2001-M01 No.5, RWC-MDB-P-2001-M01 No.15, RWC-MDB-P-2001-M07 No.94 を利用しました。ここに感謝いたします。

文 献

- [1] 大脇 渉, 高橋弘太, “スマートミキサー — 新しい音信号混合法,” 信学技報, EA2011-102, Dec. 2011.
- [2] 高橋弘太, 大脇 渉, “スマートミキサー — 基本原理と有効性,” 信学技報, EA2011-103, Dec. 2011.
- [3] 大脇 渉, 有北知弘, 手塚 歩, 宮地絃司, 高橋弘太, “音声を埋もれさせない音信号混合法 — スマートミキサーの提案,” 音響春季講義集, 1-1-15, pp.699-702, March 2012.
- [4] 高橋弘太, 有北知弘, 大脇 渉, 手塚 歩, 宮地絃司, “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,” 音響春季講義集, 3-6-8, pp.1035-1038, March 2012.
- [5] 大脇 渉, 有北知弘, 宮地絃司, 高橋弘太, “スマートミキサー — 聴覚心理モデルに基づいたゲインマスク生成,” 信学技報, EA2012-117, Dec. 2012.
- [6] 大脇 渉, 有北知弘, 宮地絃司, 高橋弘太, “スマートミキサー — 時間周波数平面での位相操作,” 信学技報, EA2012-143, March 2013.
- [7] 宮地絃司, 大脇 渉, 高橋弘太, “音響圧縮形式の時間周波数解析を利用した音信号混合法,” 信学技報, EA2013-128, March 2014.
- [8] 有北知弘, 大脇 渉, 宮地絃司, 高橋弘太, “急峻な周波数変化に対応した信号解析法とスマートミキサーへの応用,” 信学技報, EA2013-129, March 2014.
- [9] 田本祐樹, 大脇 渉, 高橋弘太, “画像を媒介とした直観的操作が可能なスマートミキサー,” 信学技報, EA2014-26, Oct. 2014.
- [10] 旭岡舜介, 大脇 渉, 高橋弘太, “フォルマント構造維持を規範とした音声信号混合法,” 信学技報, EA2014-27, Oct. 2014.
- [11] 大脇 渉, 旭岡舜介, 高橋弘太, “フォルマント構造維持を規範とした音声信号混合法,” 情処研報音楽情報科学, 2015-MUS-107, no.66, pp.1-4, May 2015.
- [12] W. Owaki and K. Takahashi, “Novel sound mixing method for voice and background music,” Proc. IEEE International Conference on Acoustics, Speech Signal Process. (ICASSP 2015), pp.290-294, April 2015.
- [13] 小森智康, 壇 寛弥, 都木 徹, 庄田清武, 黒住幸一, 小宮山拱, 星 英明, 村川一広, “ラウドネスレベルを指標とした音声ミクシングバランスに関する研究,” 信学論 (A), vol.J92-A, no.5, pp.344-352, May 2009.
- [14] 葛巻善郎, エンジニアが教えるミックス・テクニック 99, リットーミュージック, 2009.
- [15] R. Izhaki, Mixing Audio: Concepts, Practices and Tools, Focal Press, 2008.
- [16] E. Perez-Gonzalez and J.D. Reiss, “Automatic equalization of multichannel audio using cross-adaptive methods,” Audio Engineering Society Convention 127, 2009.
- [17] J.D. Reiss, “Intelligent systems for mixing multichannel audio,” 2011 17th International Conference on Digital Signal Processing (DSP), pp.1-6, July 2011.
- [18] U. Zölzer, DAFX: Digital Audio Effects, Wiley, 2011.
- [19] ISO/IEC, JTC1/SC29/WG11 MPEG, “Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-Part 3: Audio,” IS11172-3, 1992 (“MPEG-1”).
- [20] 高橋弘太, 葛木圭悟, 吉原 亨, “話速管理機能を持った原稿提示収録システム (ReCoK5) と話速バリエーション型音声データベース (SRV-DB) の公開について,” 信学技報, SP2008-117, Dec. 2008.
- [21] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一, “RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース,” 情処研報音楽情報科学, 2001-MUS-42-6, 35-42, Oct. 2001.
- [22] ITU-T Recommendation P.800. Methods for subjective determination of transmission quality, 1996 Telecommunication standardization sector of ITU (ITU-T).

(平成 27 年 6 月 23 日受付, 11 月 23 日再受付)



大脇 渉 (正員)

2011 電気通信大学電気通信学部卒業。2013 電気通信大学大学院情報理工学研究科博士前期課程修了。修士 (工学)。現在、同大学院情報理工学研究科博士後期課程在学中。デジタル信号処理, 音響信号処理, スマートミキサーの研究に従事。



高橋 弘太 (正員)

1984 東京大学工学部計数工学科卒. 1986 同大学院修士課程修了. 同年, 東京大学工学部助手. 1994 同大講師. 同年, 電気通信大学講師. 1996 同大助教授. 2007 同大准教授. 博士 (工学). 統計的信号処理, センサ信号のデジタル信号処理, マイクロ

ホンアレイや独立成分分析などに特化した実時間センシングシステムの DSP の並列処理による実現, 映像と音響情報によるマンマシンインターフェース, 効率的視聴法と話速推定法, スマートミキサーの研究に従事. また, 話速推定法の研究者用に, 話速を厳密に制御して原稿を読み上げ収録した独自の音声データベース SRV-DB を製作し提供している. 日本音響学会, IEEE 会員.