修 士 論 文

両耳による選択的聴取を
補助する雑音残響環境下音声強調手法の研究

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

佐々木 裕吉

2012 年 3 月

修 士 論 文

両耳による選択的聴取を補助する雑音残響環境下音声強調手法の研究

指導教官　赤木正人 教授

審査委員主査　赤木正人 教授
審査委員　鵜木祐史 准教授
審査委員　党建武 教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0910025 佐々木 裕吉

提出年月: 2012 年 2 月

概 要

音声認識において，雑音や残響の影響を受けることにより，性能の大幅な低下が見られる。また，軽度難聴者は健聴者と比較して，人混みやホールの中など雑音や残響が多い環境下において，聴き取り能力が著しく低下するという報告がある。そのため，雑音や残響を抑圧するために音声強調手法を音声認識や補聴器に導入する試みが盛んに行われている。これまでに様々な音声強調手法が提案されているが，その中に人の両耳聴機能に着目した手法が存在する。

Usagawa らはLindemann の両耳聴モデルに基づき，周波数領域両耳聴モデル (Frequency Domain Binaural Model; FDBM) を提案した。この手法では，左右の受信信号におけるクロススペクトルを算出することにより，両耳間位相差と両耳間レベル差を計算する。そして，音源の方向情報の推定を行い，雑音環境下において目的信号を抽出する。処理信号に各音源の方向情報を保存することで，使用者の両耳による選択的聴取を補助することに成功している。Li らによって提案された Two–Stage Binaural Speech Enhancement with Wiener Filter (TS–BASE/WF) では， 1 つの処理体系で雑音抑圧を行わず雑音推定部と雑音抑圧部から成る 2 段階の処理体系を持つことにより，処理性能を向上させている。前述の FDBM と TS–BASE/WF，どちらも雑音環境下での使用を想定している。しかし，実環境下での使用を考えた場合，屋内での使用も考えられるため，雑音と残響を同時に抑圧する必要がある。

室内インパルス応答において，室の大きさに依存した時刻を境界としたとき，初期反射と後部残響に区別することができる。初期反射は壁からの単一な反射音と考えることができるため，目的信号との相関が高くなる。後部残響は複数の反射音が重なることにより目的信号との相関は低いが，部屋中に拡散していることになる。このような特性を持つ残響と雑音を同時に抑圧できる両耳による選択的聴取を補助する音声強調手法はほとんどないと言える。

本研究では，雑音残響環境下において両耳による選択的聴取を補助する音声強調手法の構築を目的とする。残響環境下での TS–BASE/WF の動作を考えた場合，雑音推定部ではクロススペクトルを用いないため，残響環境下でも目的信号以外の音の推定は可能であると考えられる。一方，雑音抑圧部では目的信号と雑音が無相関であることが前提となる Wiener Filter を用いているため，処理信号に影響を及ぼすことが予測される。この仮説を実験により検証し，得られた結果に基づいて TS–BASE/WF の改良手法を構築した。構築した改良手法には TS–BASE/WF 以上の性能が見られた。これにより，雑音残響環境下での両耳による選択的聴取を補助する音声強調手法の実現の可能性が示唆された。

目 次

図 目 次

第1章　序論

1.1　本研究の背景

1.1.1　音声強調手法

近年，信号処理技術や通信技術の発達に伴い，様々な音声アプリケーションが広く普及している。携帯電話に代表されるような小型携帯端末やインターネット電話などを用いることで，遠距離間での音声通信が手軽に行える。また，自動音声認識器 (Automatic Speech Recognizer; ASR) を導入することにより，機械に入力する際の負担軽減に貢献している。しかし，音声通信は他人の話し声やドアの開閉音などの様々な雑音が存在する環境下での使用が想定される。加えて室内での使用を想定した時，壁からの反射音，すなわち残響の影響も考慮に入れる必要がある。これら雑音や残響の影響が，音声通信の妨げになる可能性は否めない。ASR についても同様の問題が生じている。雑音や残響が存在しない環境ではほぼ完璧な認識が可能であっても，実環境下においては認識率が大幅に低下する [1][2]。このため，雑音や残響を抑圧するために音声強調手法を音声アプリケーションがフロントエンドとして導入されている。

また，高齢者や軽度難聴者は，健聴者と比較して雑音や残響の影響により，聴き取り能力が著しく低下する [3]。健聴者は，人混みやホールの中などの騒音が多い環境でも，聞き取りたい音を抽出することが可能である [4]。しかし，難聴者は内耳による周波数分解能などが低下しているため，任意の音を聞き分けることができない。このような理由から，補聴器により音の振幅を線形増幅しても完全な聴力回復は難しい。そこで，音声強調手法を備えた補聴器を用いて雑音 · 残響を抑制し，難聴者の聴力回復を目指す試みがなされている [5]。音声アプリケーションや補聴器への需要は，今後さらに高まることが予想されるため，音声強調手法は既に欠かすことのできない技術の一つである。

1.1.2　これまでに提案された音声強調手法

1.1.1 節で述べた通り，音声強調手法を音声アプリケーションや補聴器に導入することでそれらの性能を高めてきた。音声強調手法は受信信号に見られる特徴に基づいて，様々な信号処理技術を駆使しながら発展している。その用途は大きく分けて，雑音抑圧と残響抑圧，そして雑音 · 残響抑圧に分類される。

雑音抑圧

雑音抑圧を目的とした音声強調手法の中に Boll のスペクトルサブトラクション (Spectral Subtraction; SS) [6] がある。受信信号の振幅スペクトルから雑音の振幅スペクトルの推定平均値を差し引くことで雑音抑圧を行う。ホワイトノイズのような定常雑音を対象としており，非定常な雑音には対応していない。また，雑音の推定誤差などから生じるミュージカルノイズが問題とされる。しかしながら，処理が非常に簡潔であるため，現在に至るまで様々な改良法が提案されている [7][8][9]。Boll が提案した SS は単点受信であるが，マイク素子を複数にすることで音源の方向情報を利用し雑音スペクトルの推定精度を向上させた手法 [10] も存在する。

また統計的信号処理に基づく手法も存在する。Ephraim & Malah により提案された Minimum Mean Square Error – Short–Time Spectral Ampritude (MMSE–STSA) [11] では，推定短時間振幅スペクトルの平均二乗誤差を最小にする。MMSE–STSA では雑音環境下での音声強調を可能とするが，非定常雑音に対して非常に弱い。Lim & Oppenheim によって提案された手法では，Wiener が提案した Wiener filter を音声に適用している [12]。この手法では，フレーム間で推定した雑音の振幅スペクトルに対して平均二乗誤差を算出し，最適なゲイン関数を合成していく。どちらも，雑音と目的信号が無相関であることを仮定しているため，残響環境下では適用が難しい。

残響除去

Nakatani *et al.*, により提案された Hermonic–based dEReverBeration (HERB) [13] は，残響が付加された音声信号を回復する音声強調手法である。この手法は，残響環境下における音声の調波構造を回復するため，室内インパルス応答 (RIR) の逆フィルタを受信信号に適用する方法である。しかし，その効果は残響時間が短い場合に限られている。複数のマイク素子を用いる手法として，Multiple–input/output Inversee Theorem (MINT) [14] を Miyoshi & Kaneda が提案している。MINT では，室を 1 入力多出力型の線形システムに見立てて逆フィルタを合成する。このため，RIR が非最小位相特性を持っていたとしても逆フィルタを合成することができる。

HERB と MINT の抱える共通の問題点として，事前に RIR を測定しておく必要がある。このため，室の時間による環境変化に追従することが難しい。このような問題を解決するため，semi–blind MINT [15] が提案されている。この手法では音源から一番近いマイク素子が既知であるとし，各マイクロホン間の相関行列から RIR を推定する。しかしながら，音源から一番近いマイクが既知である必要があるため，完全なブラインド処理とは言えない。

雑音 · 残響抑圧

Asano により，独立成分分析 (Independent Component Analysis; ICA) に基づいたブラインド音源分離 [16] が提案されている。ICA とは音源同士が統計的に独立であるという仮定のもと，受信信号を複数の加法的成分に分離する計算手法である。Asano のブラインド音源分離を参考に Takahashi *et al.*, は複数のマイク素子を用いて ICA を行う手法を提案している。この手法では，ICA と SS を用いて，マイク素子数と同等かそれ以下の音源分離することが可能となっている。また，Furuya *et al.*, により 耐残響耐性を加味した手法の検討がなされている [18]。これらの手法は ASA のフロントエンドとして使用することを念頭に置いて構築されているため，音声通信の様な人が聴く音声アプリケーションや補聴器への導入には適しているとは言い難い。また，分離できる雑音源の数がマイク素子数と同等かそれ以下と制限されている。このため，雑音源の数が未知である実環境への適用を考えた場合，システムの規模が大規模になる可能性が否めない。

Kinoshita *et al.*, は複数のマイク素子を利用した雑音残響環境下における音声強調手法 [19] を提案している。SS による雑音抑圧と多段線形予測に基づく残響抑圧の 2 段階処理から成る音声強調手法である。この手法を用いることで ASA の認識率を向上させることが可能であるが，周波数の高域に雑音の取り残しが見られる。このため，人が聴くことを想定するアプリケーションへの適用には疑問が残る。

1.1.3 2 入力 2 出力型音声強調手法

人は両耳で受信した信号から，音源の位置する方向や距離を知覚することができる。音源の方向情報を知覚することにより，複数の音源が存在する環境下でも聴き取りたい音を抽出することができる。したがって，特定の方向に注意を向けて音を聴くことにより，その方向に存在する音が聴こえやすくなる場合がある [20]。このような知見から，音声強調手法の形式を 2 入力 2 出力型にすることにより，出力信号に各音源の方向情報を付加する手法が存在する。これにより，使用者自身の両耳による選択的聴取を補助することが可能となり，今まで述べてきた 1 出力型の音声強調手法よりも聴感上の音質が良くなる。

Wiener post–filter を用いる手法

Zelinski は異なるマイク間での雑音が無相関であるという仮定に基づく雑音環境下音声強調手法 [21] を提案した。Dörbecker & Ernst は Zelinski の手法を 2 入力 2 出力型の音声強調手法 [22] に発展させた。2 点のマイク間で受信された雑音が無相関であるという仮定から，雑音間のクロススペクトルを推定し IIR–filter を構築する。構築した IIR–filter をフレーム間での平均二乗誤差が最少になるよう Wiener post–filter を用いて補正する。最後に補正した IIR–filter を受信信号に適用することで雑音の抑圧を行う。この手法では，定常雑音や非定常雑音だけでなく，拡散雑音や残響の抑圧を目的として構築されている。

しかしながら，IIR–filter の構築に受信信号間でのクロススペクトルを用いているため，目的信号と相関の高い残響の抑圧が完全に行われているか不明である。

周波数領域両耳聴モデル

Usagawa *et al.*, は Lindemann の両耳聴モデル [23] に基づき，周波数領域両耳聴モデル (Frequency Domain Binaural Model; FDBM) [24] を提案した。Lindeman の両耳聴モデルでは時間領域で処理を行うのに対し，FDBM では周波数領域で処理を行っている。FDBM はクロススペクトルを用いて IPD と ILD を算出し，音源の方向推定を行う。次にその推定方向に対し，音源分離を行うことで雑音の抑圧を行う。雑音環境下において，音源の方向推定と音源分離を一括して行う優れた手法と言える。しかし，FDBM はクロススペクトルを用いて音源の方向推定を行うため，残響環境下での性能低下が見られる。

Two–Stage Binaural Speech Enhancement; TS–BASE/WF

Li *et al.*, によって提案された Two–Stage Binaural Speech Enhancement with Wiener Filter (TS–BASE/WF) [25] では， 1 つの処理体系で雑音抑圧を行わず雑音推定部と雑音抑圧部から成る 2 段階の処理体系を持つ手法である。雑音推定部では，適用フィルタを用いて両耳受信信号間で減算することにより，受信信号に含まれる雑音を推定する。推定した雑音に基づき，雑音抑圧部では Wiener filter を用いてゲイン関数を合成し，受信信号に適用することで雑音の抑圧を行う。しかし，TS–BASE/WF の雑音推定には目的信号の到来方向が既知である必要がある。そこで，近年 Duc *et al.*, が TS–BASE/WF のフロントエンドとして，目的信号の方向推定部を加えた手法 [26] を提案している。TS–BASE/WF は耐残響性に関する検討は行われていないものの，目的信号と雑音が無相関であることを仮定する Wiener filter を雑音抑圧部に用いているため，残響環境下での使用に疑問が残る。

1.2 研究の目的

現在までに提案されてきた音声強調手法の問題点を踏まえ，本研究では，音声アプリケーションや補聴器への導入を考慮した雑音残響環境下における音声強調手法の構築を目的とする。前節で述べたとおり，2 入力 2 出力型の音声強調手法は聴感上の音質が良い。そのため，音声アプリケーションや補聴器への導入に適した音声強調手法の形式と言える。このことから，本研究で提案する音声強調手法は 2 入力 2 出力型を想定する。

TS–BASE/WF の雑音推定部では，他の手法と異なり受信信号間のクロススペクトルを用いていない。このため，雑音推定部は残響環境下でも有効であることが考えられる。しかし，前述の通り雑音抑圧部では，目的信号と雑音が無相関であることを仮定する Wiener

filter を用いているため，正常に動作しない可能性がある。そこで本研究では，まず TS–BASE/WF の残響環境下での性能評価を行うことで，その問題点を明らかにする。次に明らかになった問題点から TS–BASE/WF の改良手法を検討·構築し，最後に TS–BASE/WF と改良手法との性能比較を行う。

1.3 本論文の構成

本論文は，全 8 章により構成される。以下に各章の概要を簡潔に記す。

第 1 章

本研究の背景と関係する先行研究を紹介する。そして，現段階では解決されていない問題点を示す。これに基づき，本研究で解決する目的を述べる。

第 2 章

雑音環境下を想定した音声強調手法である TS–BASE/WF に着目し，その処理工程について示す。また，残響環境下での TS–BASE/WF の動作を想定し，その問題点を予測する。

第 3 章

本研究のシミュレーション実験で用いるデータベースと客観評価尺度について述べる。また，本件急で用いる頭部伝達関数を含む室内インパルス応答の作成手順について詳しく記述する。

第 4 章

本章では，シミュレーション実験により TS–BASE/WF の性能評価を行う。まず，雑音環境下における TS–BASE/WF の性能を示すことにより，その有効性を確認する。次に，TS–BASE/WF に対して残響耐性評価実験を行うことにより，第 2 章で予測した問題点を検証する。

第 5 章

第 4 章で明らかにした TS–BASE/WF の問題点を補うための改良案を検討する。また，改良手法の構成とその処理工程について詳しく記述する。

第 6 章

本章では，ケプストラム平均サブトラクション (Cepstral Mean Subtraction; CMS) が TS–BASE/WF のフロントエンドとして有効であることを確認する。まず，CMS のパラメータ設定を行うため，シミュレーション実験を行った。そして，CMS がエコーに対して有用であること，同様に残響に対しても効果が見られることを示す。

第 7 章

第 5 章で提案した TS–BASE/WF の改良手法である CMS + TS–BASE/WF の性能評価を行う。まず，残響耐性評価実験を行い，TS–BASE/WF と CMS + TS–BASE/WF の性能を比較する。次に，雑音環境下と雑音残響環境下における TS–BASE/WF と CMS + TS–BASE/WF の性能を評価する。

第 8 章

本研究で得られた結果を要約し，本研究で残された課題を記述する。

第2章　TS–BASE/WF の概要

TS–BASE/WF のブロックダイアグラムを 図 2.1 に示す。

2.1　想定する信号モデル

TS–BASE/WF は雑音環境下での使用を想定しているため，受信信号は目的信号と雑音の和で表すことができる。i 番目のフレームにおける周波数 k [Hz] の両耳受信信号スペクトルを $X_L(k,l)$ ，$X_R(k,l)$ とすると，信号モデルは次式のように表すことができる。

$$X_i(k,l) \ = \ S_i(k,l) \ + \ N_i(k,l), \ i = L, R \tag{2.1}$$

式中の $S_i(k,l) = HRTF_i(k,l)S(k,l)$ は目的信号スペクトルを表しており，$HRTF_i(k,l)$ は 頭部伝達関数 (HRTF) のスペクトルを示している。また，式中の $N_i(k,l)$ は雑音を表しており，突発性雑音や複数の雑音源，拡散雑音もこれに含まれる。目的信号の到来方向が既知であり，雑音と目的信号が無相関であるものとする。

2.2　雑音推定部

Equalization–cancellation (EC) モデルは，Durlach により提案され [27]，Culling & Sumerfiled により発展した [28] 人の両耳聴モデルである。TS–BASE/WF の雑音推定部は，EC モデルを参考に構築された。両耳信号では HRTF の影響を受けることにより，左右間の信号において位相や到来時間などが異なってくる。従来の EC モデルではこのような左右間の信号特性を生かして雑音を消約する。一方，TS–BASE/WF の雑音推定部では目的信号を消約することにより，雑音のみを出力する。TS–BASE/WF の雑音推定部は以下の等価処理と消約処理により成り立っている。

2.2.1　等価処理

等価処理では，Normalized Least Mean Square アルゴリズムを用いて，図 2.1 における等価フィルタ $W_L(k,l)$, $W_R(k,l)$ の合成を行う。

$$W_L(l+1) = W_L(l) + \mu \frac{X_L}{\|X_L\|^2}[X_R(l) - W_L^T(l)X_L(l)], \tag{2.2}$$

$$W_R(l+1) = W_R(l) + \mu \frac{X_R}{\|X_R\|^2}[X_L(l) - W_R^T(l)X_R(l)], \tag{2.3}$$

ただし，$W_i(l) = [W_i(1,l), W_i(2,l), ..., W_i(K,l)]^T$，$X_i(l) = [X_i(1,l), X_i(2,l), ..., X_i(K,l)]^T$ $(i = L, R)$, を指し，K は窓長，上付文字 T は転地記号，μ はステップサイズを表す。

等価フィルタは周波数領域において入力信号に乗算することにより，両耳間で目的信号成分が等しくなるよう事前に学習される。ただし，信号モデルを見て明らかなように目的信号の到来方向は既知であるものとする。学習試料には，HRTF が畳み込まれた 10 s の時間長を持つホワイトノイズが用いられる。

2.2.2 消約処理

図 2.1 が示す通り，目的信号の到来方向に対応した適応フィルタを用いて左右の受信信号から目的信号を取り去る。等価処理により，等価フィルタは雑音の存在しない環境下により学習されている。したがって，等価フィルタをぞれぞれ左右の受信信号に適用することにより，目的信号成分が左右の受信信号間でほぼ等しくなる。等価フィルタが適用された受信信号を左右間で減算することにより，目的信号の雑音成分のみが抽出される。

$$\begin{aligned} Z_L(k,l) &= X_L(k,l) - W_R(k,l)X_R(k,l) \\ &\approx N_L(k,l) - W_R(k,l)N_R(k,l), \end{aligned} \tag{2.4}$$

$$\begin{aligned} Z_R(k,l) &= X_R(k,l) - W_L(k,l)X_L(k,l) \\ &\approx N_R(k,l) - W_L(k,l)N_L(k,l), \end{aligned} \tag{2.5}$$

式中の Z_L, Z_R は推定した雑音である。この消約処理により，目的信号の方向外から到来する雑音を推定する。このことから，複数の雑音源（もしくは拡散雑音）や非定常雑音を推定することが可能となる。

2.3 雑音抑圧部

推定した雑音を元にゲイン関数を合成し，雑音の抑圧を行う。前節で述べた雑音推定部により，受信信号における雑音成分は推定されたことになる。しかし，等価フィルタがそれぞれ左右の受信信号に適用されているため正確に推定されているとは言い難い。そこで補償因子 $\hat{C}_i(k,l)$ を受信信号と推定した雑音との平均二乗誤差が最小になるよう合成する。図 2.1 が示す様に補償因子を推定した雑音に適用することで，受信信号中の雑音成

分と推定された雑音が対応付けられる。ただし，目的信号と雑音は無相関であるものとする。

$$\hat{C}_i(k,l) = \arg\min_{C_i} E[X_i(k,l) - Z_i(k,l)C_i(k,l)],\ i = L, R \tag{2.6}$$

式中の E は期待値算出の演算子を表す。式 (2.6) の右辺における期待値が 0 に近づくよう，補償因子は最適化されていく。Wiener の理論により，最適化された補償因子 $C_i^{opt}(k,l)$ は以下のように与えられる。

$$C_i^{opt}(k,l) = \frac{\phi_{X_iZ_i}(k,l)}{\phi_{Z_iZ_i}(k,l)},\ i = L, R \tag{2.7}$$

式中の $\phi_{X_iZ_i}(k,l)$ は $X_i(k,l)$ と $Z_i(k,l)$ とのクロススペクトルを表し，$\phi_{Z_iZ_i}(k,l)$ は $Z_i(k,l)$ の自己相関スペクトルを示す。次に Wiener filter を用いてゲイン関数を合成していく。まず，*priori* SNR を以下のように算出する。

$$\xi = \frac{E[S_LS_L^* + S_RS_R^*]}{E[(C_LZ_L)(C_LZ_L)^* + (C_RZ_R)(C_RZ_R)^*]}\ , \tag{2.8}$$

式中の $\xi(k,l)$ は算出された *priori* SNR を表し，上付き記号 $*$ は複素共役を示している。*priori* SNR はミュージカルノイズが残留しないよう算出されていく。これを用いてゲイン関数 $G_{WF}(k,l)$ は以下のように求めることができる [29]。

$$G_{WF} = \frac{\xi(k,l)}{1+\xi(k,l)}, \tag{2.9}$$

図 2.1 の示す Gain calculation は式 (2.8) と式 (2.9) に相当する演算を行っている。最後にゲイン関数を左右の受信信号に適用することにより，目的信号が推定される。

$$\hat{S}_i(k,l) = G_{WF}(k,l)X_i(k,l).\ i = L, R \tag{2.10}$$

式 (2.10) の様に左右の受信信号間で共通のゲイン関数を用いることにより，各音源の方向情報を保存することが可能となっている。

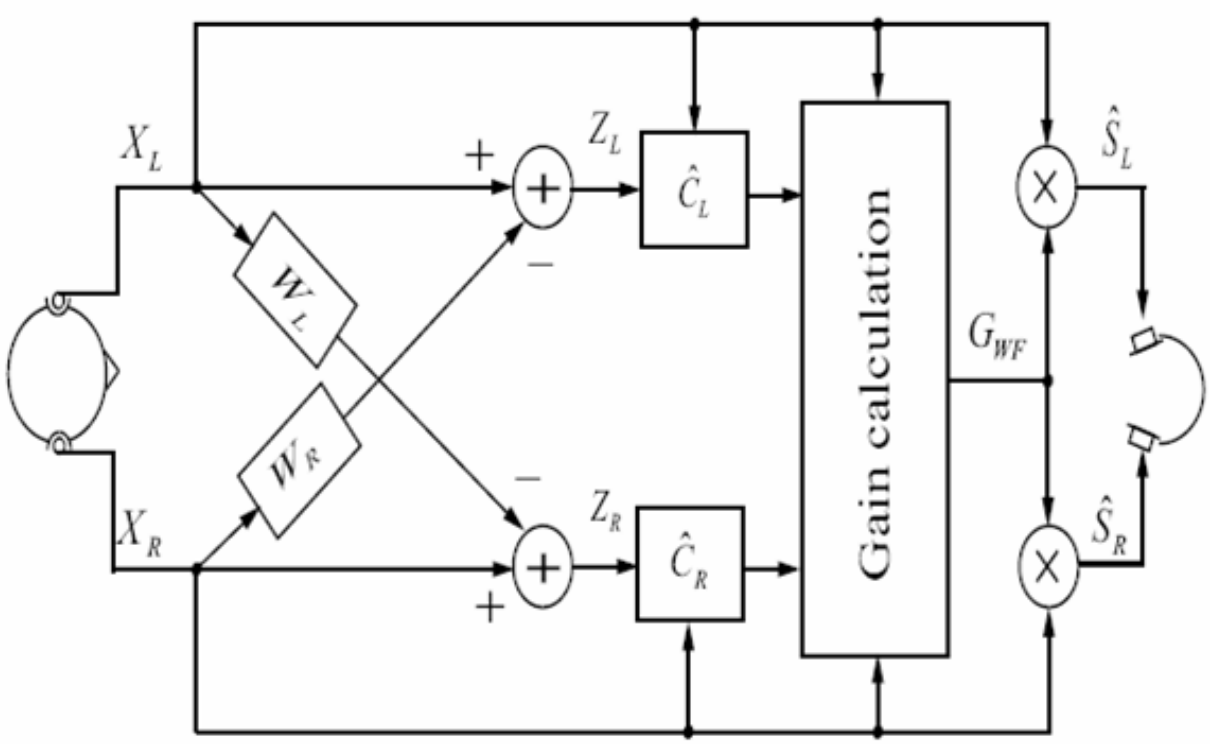


図 2.1: TS–BASE/WF のブロックダイアグラム

2.4 残響環境下における TS–BASE/WF の問題点

目的信号に RIR が畳み込まれると残響が付加される。この信号を受信信号として TS–BASE/WF に処理させることを考える。RIR は室の大きさに依存した時刻を境界として，初期反射と後部残響に区別することができる。(図 2.2) 初期反射は壁によって反射した単一のエコーと考えることができる。このため，初期反射は目的信号との相関が高く，音源の方向情報が保持されている。一方，後部残響は複数の反射音が重なることにより目的信号との相関は低い。しかし，反射音が部屋中に拡散しているため音源の方向情報が希薄になっている。

TS–BASE/WF における雑音推定部はクロススペクトルを用いずに処理が行われる。このため，音源の方向情報が十分に保持されているならば，初期反射と後部残響の推定が可能となる。しかし，先述の通り後部残響には方向情報が希薄であるため，十分に雑音推定できるか不明である。一方，雑音抑圧部では目的信号と雑音が無相関であることが前提である Wiener Filter を用いるため，初期反射の抑圧に影響が生じる可能性がある。以上の点を踏まえて，第 4 章では TS–BASE/WF の性能評価実験を行い，ここで述べた仮説を実証する。

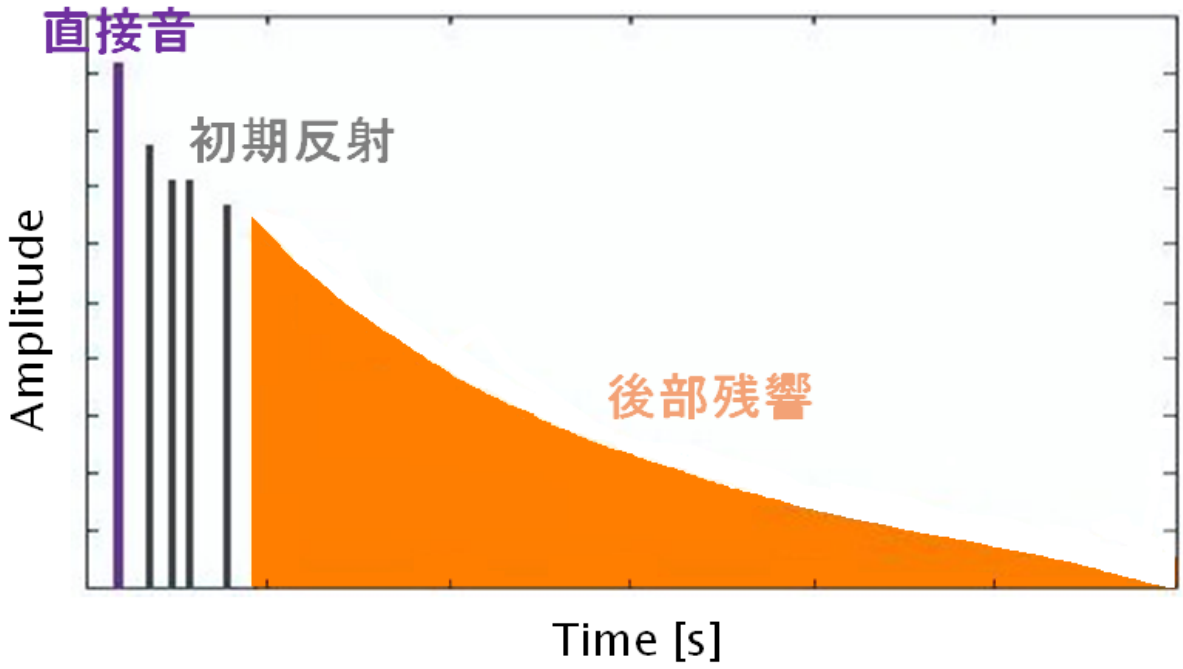


図 2.2: 室内インパルス応答の概略図

第3章　データベースと評価尺度

3.1　データベース

3.1.1　音声試料と頭部伝達関数

本研究における実験の音声試料には NTT–AT の音素バランス 1000 文広帯域音声データベースを使用した。サンプリング周波数は 44.1 kHz，量子化 16 bit となっている。実験で用いる話者はそれぞれ，FIH (女声)，FSS (女声)，MIY (男声)，MYK (男声) となっている。また，各話者ごとに違う文章を発話した音声試料を用いた。また HRTF には，MIT データベース [32] を使用した。サンプリング周波数は 44.1 kHz，量子化 16 bit となっている。HRTF を信号に畳み込むことで，測定した角度と仰角の方向情報を付加することが可能である。本研究において，音源はすべて仰角 0° に設置しており，正面方向を水平角 0°，右側を +，左側を − とした。

3.1.2　頭部伝達関数を含む室内インパルス応答

RIR を擬似的に作成する手法として，鏡像音源法 [33] がある。まず，反射音を壁にかけられた鏡に写る鏡像音源から到来すると仮定する。鏡像音源から到来する反射音に壁の反射率と距離に応じた減衰率を乗算し，時系列に並べていくことで RIR を作成する。しかしながら，両耳間で異なる RIR を作成しても直接音と反射音の方向情報が十分に反映されていると言いがたい。本研究では鏡像音源法を発展させ，直接音と反射音に対応する HRTF を畳み込むことを検討した。

反射音の到来方向推定

音源とマイク中間点との水平角 $\theta(n,m,l)$，仰角 $\phi(n,m,l)$ を求める。n と m，l はそれぞれ，x 軸方向の部屋番号と y 軸方向，z 軸方向の部屋番号を示している。部屋番号がすべて 0 を示すときは，実音源が設置されている実空間を表す。また部屋番号が負の値を示す場合，軸上で負の領域に鏡像空間が設置されていることを表している。鏡像音源が設置されている座標から，逆正接関数を用いて $\theta(n,m,l)$ と $\phi(n,m,l)$ を割り出した。

反射音の作成

マイク中間点と音源との距離から，音波が到来する時間 $T(n,m,l)$ を求める。そして，反射回数に対応する壁の反射率 $r^{|n|+|m|+|l|}$ と空気を伝番することによる減衰率 $b(n,m,l)$ との乗算を反射音 $\delta(t)$ に代入していく。

$$\delta(t) \ = \ \begin{cases} r^{|n|+|m|+|l|}b(n,m,l), \ if \ t \ = \ T(n,m,l) \\ 0, \ otherwise \end{cases} \tag{3.1}$$

反射音と頭部伝達関数との畳み込み

求めた角度$\theta(n,m,l)$と仰角$\phi(n,m,l)$に対応する頭部伝達関数$hrtf_i(\theta(n,m,l),\phi(n,m,l))$を反射音 $\delta_i(t)$ に畳み込む。その後，$\max(|n|+|m|+|l|)$ が反射回数の最大値 N 以下となるよう，反射音成分の総和を求めたものを室内インパルス応答 $h_i(t)$ とする。

$$h_i(t) \ = \sum_{\max(|n|+|m|+|l|)\leq N} \delta(t) * hrtf_i(\theta(n,m,l),\phi(n,m,l)), \ i = L,R \tag{3.2}$$

室の環境設定

本研究で再現する室の大きさは (5 m × 5 m × 3 m) となっており，2 つのマイクにおける中間点の位置は (1 m × 1 m × 1.5 m) とした。音源とマイク中間点 (もしくは雑音源) との距離は 1.4 m に設定し, マイク間の距離は 0.17 m となっている。

3.2 評価尺度

本研究の実験において，Segmental Signal–to–Noise Ratio (SEGSNR) [30] と Log–Spectral Distortion (LSD) [31] の 2 つの客観評価尺度を用いた。SEGSNR はその値が大きければ大きいほど SN 比が改善されたことになり，LSD はその値が小さいほど信号に歪みが小さいことになる。SEGSNR と LSD の計算方法を以下に示す。

$$SEGSNR = \frac{10}{L}\sum_{l=0}^{L-1}\log_{10}\left(\frac{\sum_{k=0}^{k-1}[s(lK+k)]^2}{\sum_{k=0}^{K-1}[s(lK+k)-\hat{s}(lK+k)]^2}\right), \tag{3.3}$$

$$LSD = \frac{10}{L}\sum_{l=0}^{L-1}\left(\frac{1}{K}\sum_{k=0}^{K-1}[\log_{10}AS_d(k,l)-\log_{10}A\hat{S}(k,l)]^2\right). \tag{3.4}$$

式 (3.3) と式 (3.4) の s は目的信号を表し，$\hat{s}$ は実験音もしくは処理信号である。いずれも，両耳信号である場合は左右間で平均をとる。L はフレームの総数，K は 1 フレームのサンプル数を示す。式 (3.4) において $AS(k,l) \equiv \max\{|S(k,l)|^2,\delta\}$ と定義する。$\delta = 10^{-50/10}$ であり，log–spectrum 上でダイナミックレンジを制限するために用いる。

第4章　TS–BASE/WF の性能評価実験

4.1　雑音環境下での性能評価実験

4.1.1　目的

雑音環境下における TS–BASE/WF の性能評価を行い，その有効性を確認する。今回は雑音に音声を用いて実験を行った。

4.1.2　実験条件

実験音は 16 kHz にダウンサンプリングを行った。フレームの切り出しにはハニング窓を使用し，TS–BASE/WF における FFT の分析フレーム長は 512 (32 ms)，オーバーラップは 1/2 となっている。そして，TS–BASE/WF のターゲット方向を正面方向として等価フィルタの学習を行った。

4.1.3　実験音の作成手順

目的信号と雑音，共に各話者の音声を用い，4 話者の組み合わせ全 9 通りすべてを作成した。目的信号，あるいは雑音に HRTF を畳み込んだものを使用する。雑音源は常に 1 つであり，その到来方向を 45 度から 45 度刻みで 315 度まで変化させた。SN 比は次式を用いて算出し，値は 0 dB から 2 dB 刻みで 10 dB まで変化させた。

$$SNR = 10 \log_{10} \sum_{t} \frac{|s(t)|^2}{|n(t)|^2}. \tag{4.1}$$

ここで，$s(t)$ は目的信号，$n(t)$ は雑音を表している。どちらも HRTF を畳み込んだ後に算出した。

4.1.4　実験結果と考察

実験結果を図 4.1，図 4.2 に示す。SEGSNR と LSD の改善量は共に受信信号の全区間 SN 比が増加していくに従い減少していくことが分かる。これは全区間 SN 比が増加していくに従い受信信号中の雑音成分が減少していくためだと考えられる。SEGSNR の改善

量は 2 ∼ 4 dB であり，LSD の改善量は 1.5 ∼ 2.5 dB だった。雑音が音声の様に非定常な信号でも，TS–BASE/WF により抑圧できることが分かる。

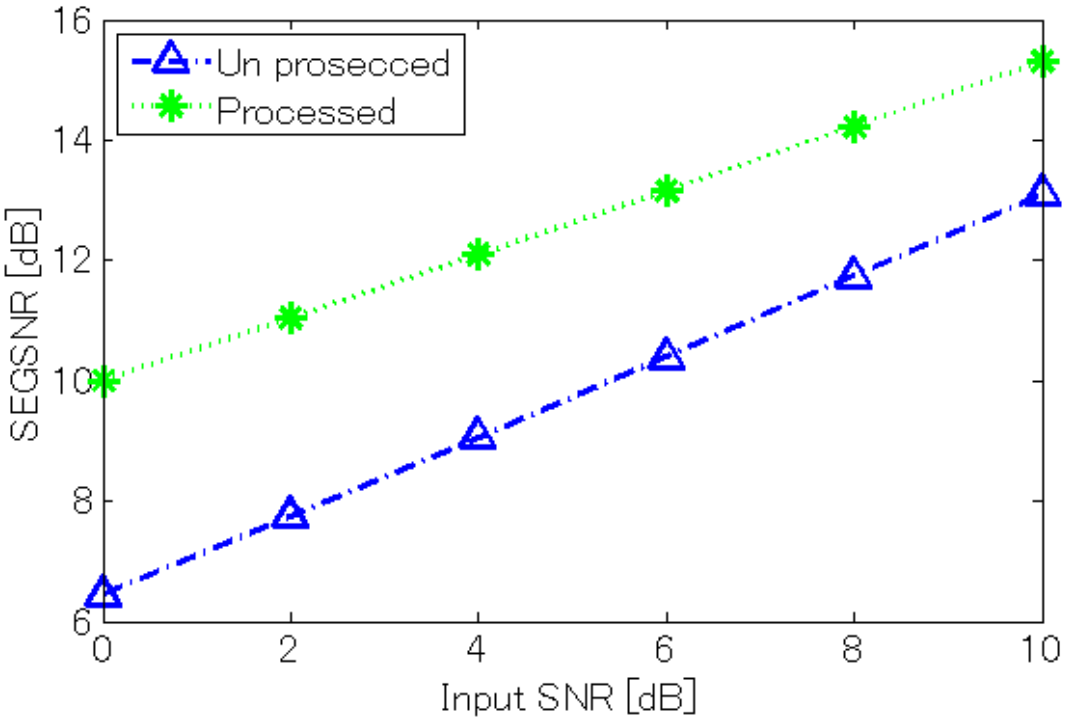


図 4.1: TS–BASE/WF の雑音耐性評価実験結果：縦軸は SEGSNR，横軸は受信信号の全区間 SN 比を示す

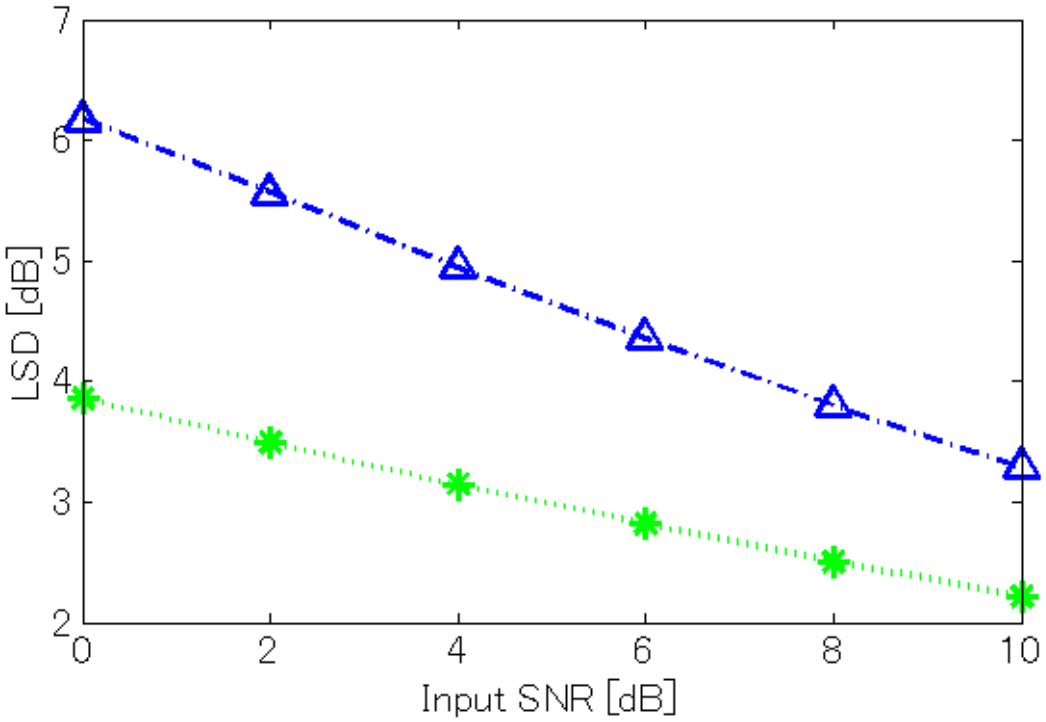


図 4.2: TS–BASE/WF の雑音耐性評価実験結果：縦軸は LSD，横軸は受信信号の全区間 SN 比を示す

4.2 残響環境下での性能評価実験

4.2.1 目的

TS–BASE/WF の残響耐性評価実験を行うことにより，その問題点を明らかにする。雑音推定部 における Wiener filter は目的信号と雑音が無相関であることを仮定しているため，処理信号に何らかの影響があると考えられる。

4.2.2 実験条件

4.1.2 節の実験条件と同様である。

4.2.3 実験音の作成手順

目的信号には各話者の音声を用い，RIR は第 3 章で述べた作成手順を用いて合成した。RIR は目的信号の到来方向を正中面とし，反射音の総数を 40 に設定した。また RIR の残響時間は 0.25 s から 0.25 s 刻みで 2 s まで変化させている。最後に作成した RIR と音声を畳み込むことで実験音とした。

4.2.4 実験結果と考察

実験結果を図 4.3，図 4.4 に示す。SEGSNR と LSD の改善量は共に，残響時間が長くなるのに従い増加する。これは，残響時間が長くなるほど，受信信号に含まれる残響成分が大きくなるためである。しかしながら残響時間が 0.25 s の時，LSD の値が受信信号に比べて処理信号が上回っている。このことから目的信号と相関が高い初期反射により，雑音推定部 の Winer Filter が影響を受けていることが予測される。

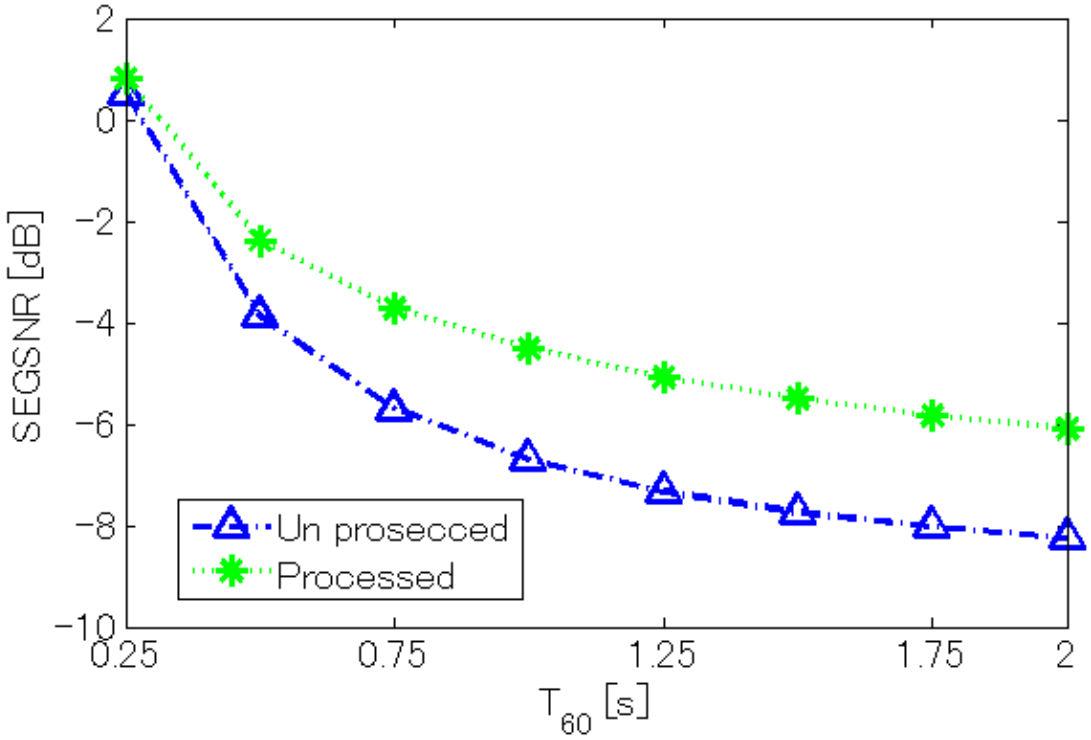


図 4.3: TS–BASE/WF の残響耐性評価実験結果：縦軸は SEGSNR，横軸は残響時間を示す

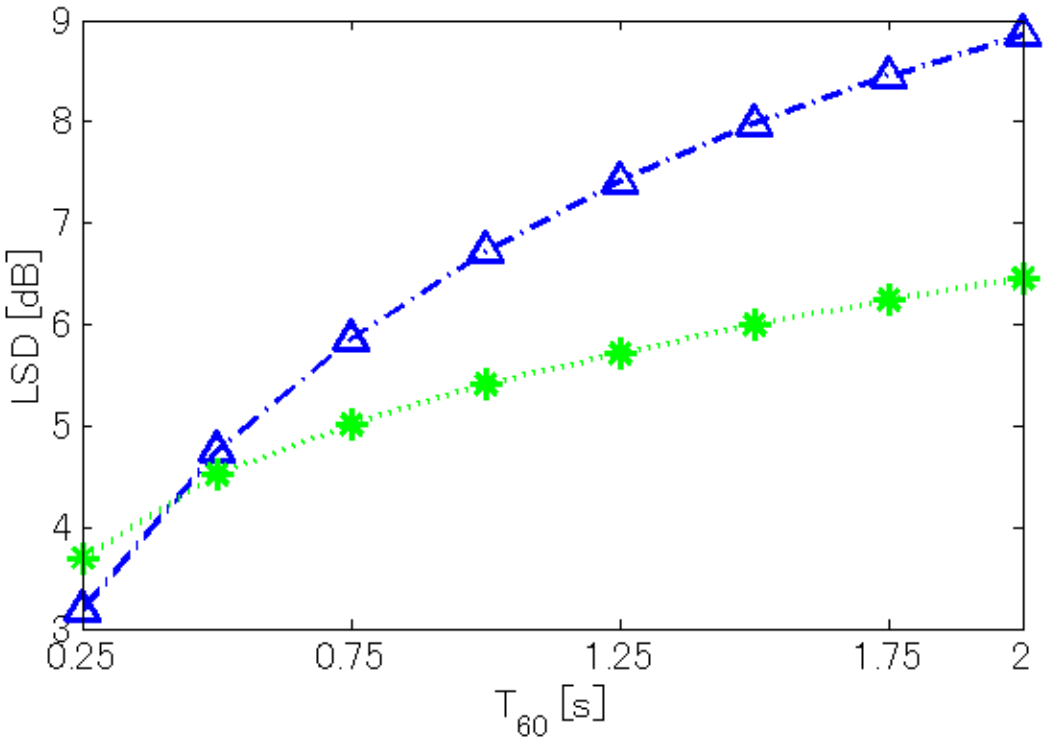


図 4.4: TS–BASE/WF の残響耐性評価実験結果：縦軸は LSD，横軸は残響時間を示す

4.3 初期反射と後部残響に対する TS–BASE/WF の性能評価実験

4.3.1 目的

前節より，雑音推定部における Wiener filter により初期反射を十分に抑圧できない可能性が示唆された。本実験では，初期反射と後部残響を擬似的に切り分け，これらが TS–BASE/WF に与える影響を観察する。

4.3.2 実験条件

4.1.2 節の実験条件と同様である。

4.3.3 実験音の作成手順

反射音の最大次数と最小次数を変化させた RIR の合成を行う。反射音の最大次数を 1 から 40 まで変化させた RIR をまず合成する。次に反射音の最大次数を 40 に固定しながら最小次数を 1 から 39 まで変化させた RIR をそれぞれ合成した。すべての RIR は残響時間を 2 s で固定し，目的信号は正中面に設置する。目的信号には 4 人の話者音声を用い，作成した RIR をそれぞれ畳み込むことで実験音とした。

4.3.4 実験結果と考察

作成した実験音を TS–BASE/WF を用いて処理を施した結果を図 4.5 と図 4.6 に示す。Additive Reflection が反射音の最大次数を増加させていった値であり，Eliminated Reflection は反射音の最大次数を固定しながら最小次数を増加させていった値を示す。All Reflection は反射音の最大次数が 40 の値を表している。図 4.5 と図 4.6 において，Additive Reflection は横軸の値が増えるに従い最大反射回数が増加し，Eliminated Reflection は横軸の値が増えるに従い最小反射回数が増加する。反射次数が 7 のとき Eliminated Refrection の SEGSNR 改善量が最大値をとる。(図 4.5) また反射次数が 9 を示しているとき， LSD 改善量について同様のことが言える。(図 4.6) このとき，低次数の反射音つまり，初期反射の影響を受けていない実験音を TS–BASE/WF が効果的に処理していることが分かる。すなわち，雑音推定部における Wiener Filter が目的信号と相関の高い初期反射を十分に抑圧できていない。

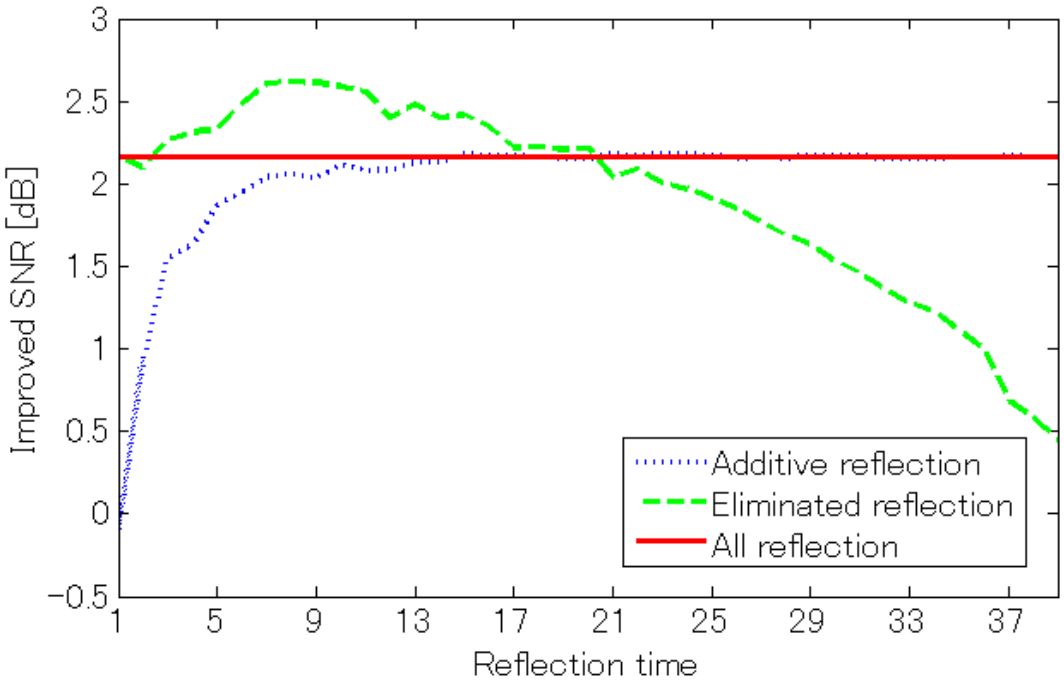


図 4.5: 初期反射と後部残響に対する TS–BASE/WF の性能評価実験結果 : 縦軸は SEGSNR の改善量，横軸は反射音の反射次数を示す

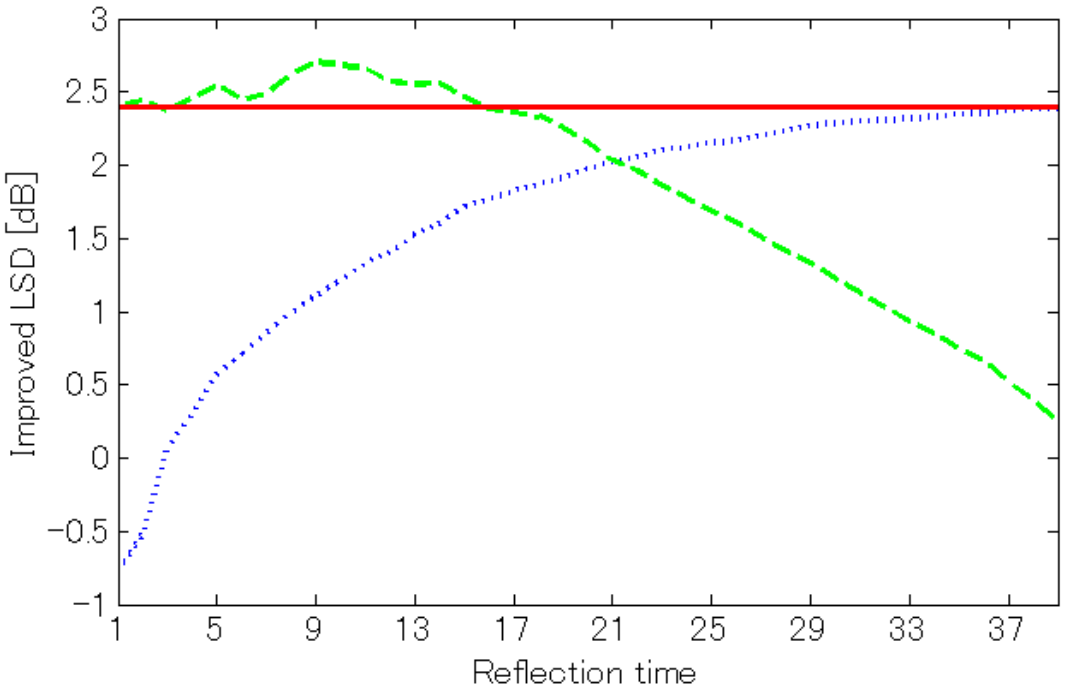


図 4.6: 初期反射と後部残響における TS–BASE/WF の性能評価実験結果：縦軸は LSD の改善量，横軸は反射音の反射次数を示す

第5章　TS–BASE/WF の改良手法に関する検討

5.1　改良案の検討

4.4.3節から，初期反射の影響によりTS–BASE/WF の雑音抑圧部における Winer Filter が十分に動作しないことが分かった。このような問題を解決するため，初期反射を抑圧するフロントエンドを TS–BASE/WF に設置する。設置する手法に関しては以下の3つの中から汎用性に優れ，且つ性能の高い手法を選択する。

最小位相差フィルタ

この手法 [34] ではくし形の RIR を想定する。RIR が最小位相を指し示すとき，逆フィルタが存在するとして残響の抑圧を行う。しかし，実環境下において，RIR は非最少特性を示すことがほとんどない。また，逆フィルタの合成には RIR が必要となるため，使用者の移動を考慮すると汎用性に欠る。

ケフレンシー領域でのリフタリング

受信信号の振幅ケプストラム上において，エコー時間に相当するケフレンシーにおいて特有のピークが見られる。このピークが0に収束するようリフタリングを行うことでエコーを抑圧することが可能となる [35]。しかし，RIR におけるエコーの遅延時間を正確に推定する必要がある。このため，最小位相フィルタと同じく使用者の移動する状況下では使いにくい。

ケプストラム平均サブトラクション

この手法 [36] では，まず受信信号の振幅ケプストラムをフレーム間で平均正規化することにより RIR の振幅ケプストラムを推定する。次に受信信号の振幅ケプストラムから推定した RIR の振幅ケプストラムを差し引くことで，エコーの抑圧を行う手法である。他の2つの手法と違い，比較的簡単にエコーの抑圧が可能な手法である。しかし，RIR における初期反射のみを推定できるか不明である。

5.2 改良手法の概要

本研究では，TS–BASE/WF のフロントエンドとしてケプストラム平均サブトラクション (Cepstral Mean Subtraction; CMS) を採用する。理由としては，前節で述べた通り使用者の移動に合せて RIR やエコー時間の推定を繰り返し測定する必要がないためである。CMS が初期反射の抑圧に有効であるならば，フロントエンドとして用いることで TS–BASE/WF の性能向上に繋がる。

5.2.1 改良手法の構成検討

左右のチャンネルでことなった RIR の振幅ケプストラムを推定し，受信信号に対して減算を行うと各音源の方向情報が保持されない可能性がある。したがって，改良手法では左右のチャンネルに入力された信号を元を平均正規化することで，チャンネル間共通の RIR の振幅ケプストラムを推定する。これをケプストラム領域において，左右の入力信号から差し引くことで初期反射を抑圧する。その後，TS–BASE/WF を用いて後処理を行うことにより，残った後部残響の抑圧を行う。この手法をこれ以後， CMS + TS–BASE/WF と呼ぶ。

ケプストラム平均正規化

まず，入力信号 $x(t)$ が以下のように目的信号 $s(t)$ と RIR $h(t)$ との畳み込みで示されることを仮定する。

$$x(t) = s(t) * h(t), \tag{5.1}$$

次に入力信号にフーリエ変換を施し，振幅スペクトルに対して対数をとる。そして，算出された対数振幅スペクトルに対して，逆フーリエ変換を行うと次式のように示すことができる。

$$c_x(k,l) = c_s(k,l) + c_h(k,l), \tag{5.2}$$

式 (5.2) における l はフレームの番号， k は対応するケフレンシーを示す。c は下付文字に対応する振幅ケプストラムである。次に算出した入力信号を N 個のフレームに区切り，重み付き平均正規化を行う。

$$\begin{aligned} c_{ave}(k,l) &= \frac{1}{\sum_{l=0}^{L-1}\exp(-\alpha\cdot l)}\sum_{l=0}^{L-1} c_x(k,l)\cdot\exp(-\alpha\cdot l) \\ &= c_{ave:s}(k,l) + c_{ave:h}(k,l) \\ &\approx \hat{c}_h(k,l), \end{aligned} \tag{5.3}$$

ただし，下付文字 ave は信号に対して，平均正規化を行ったことを意味する。$\exp(-\alpha \cdot l)$ は忘却係数であり，過去フレームの影響を少なくするために用いる。RIR が時不変である場合，式 (5.3) における $c_{ave:s}(k,l)$ は相殺される。これにより，推定される RIR の振幅ケプストラム $\hat{c}_h(k,l)$ を求めることができる。

5.2.2 ケプストラム領域におけるサブトラクション

フレームごとに入力信号の振幅ケプストラムから，推定した RIR の振幅ケプストラムを差し引く。

$$\hat{c}_s(k,l) = c_x(k,l) - \beta \cdot \hat{c}_h(k,l),\ 1 > \beta > 0. \tag{5.4}$$

式 (5.4) の処理を行った後，求められた目的信号の振幅ケプストラムと対応する入力信号の位相スペクトルを用いて再合成を行う。

第6章　CMS の性能評価実験

6.1　CMS のパラメータ設定に関する検討

6.1.1　目的

CMS におけるサブトラクション係数を適切な値に定めるため，実験を行った。サブトラクション係数を残響環境下において逐次，変化させてゆくことで適切な値を探る。

6.1.2　実験条件

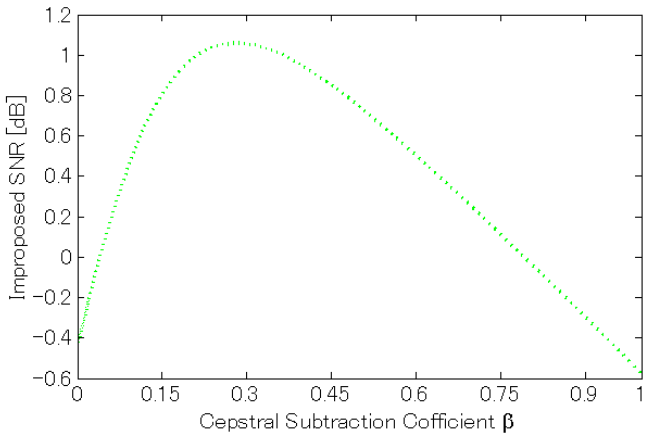
実験音は 16 kHz にダウンサンプリングを行った。フレームの切り出しにはハニング窓を使用し，CMS における FFT の分析フレーム長は 512 (32 ms)，オーバーラップは 1/4 となっている。平均正規化に用いる過去フレームは計 60 個とし，忘却係数のパラメータ α を 0.008 に設定した。また，サブトラクション係数 β は 0 から 0.02 刻みで 1 まで変化させた。

6.1.3　実験音の作成手順

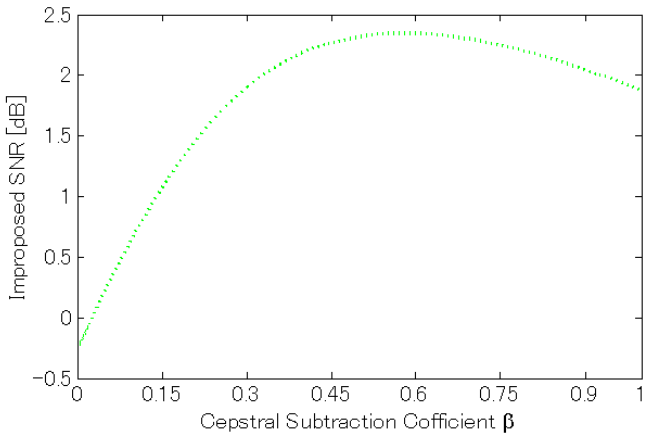
4 人の話者音声を目的信号に用い，それぞれに残響時間を 0.25 s，0.5 s，1.0 s と変化させた RIR を畳み込むことで実験音とした。RIR 作成時における目的信号の到来方向は正中面となっている。

6.1.4　実験結果と考察

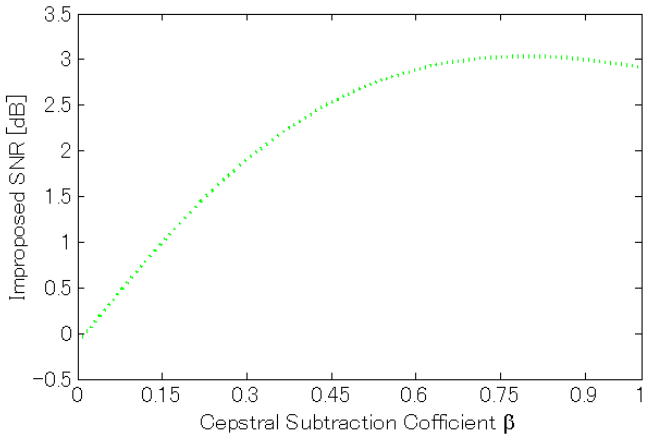
実験結果を図 6.1，図 6.2 に示す。各残響時間において，LSD と SEGSNR の改善量が最大値を取るサブトラクション係数 β の値が異なる。このことから，残響時間により最適なサブトラクション係数 β の値が一意に定まることが分かる。したがって，室の残響時間を計測後，サブトラクション係数 β の値を設定することが望ましい。しかし，本研究で提案する CMS + TS–BASE/WF では，残響時間を測定する処理が含まれていない。そのため，残響時間が短い環境下において処理信号の LSD 値を悪化させない様，サブトラクション係数を設定する必要がある。本研究では，残響時間 0.25 s の時，LSD と SEGSNR の改善量が共に大きい値を示す $\beta = 0.14$ に設定する。



(a) $T_{60} = 0.25\,s$

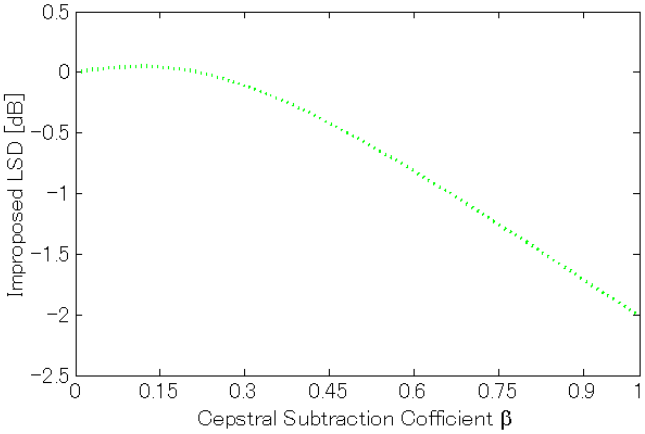


(b) $T_{60} = 0.5\,s$

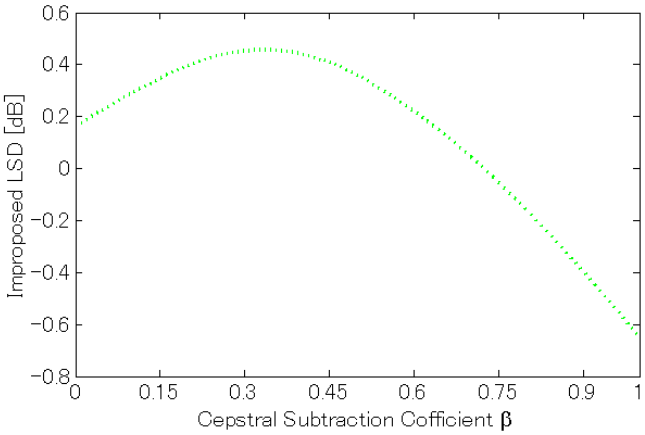


(c) $T_{60} = 1.0\,s$

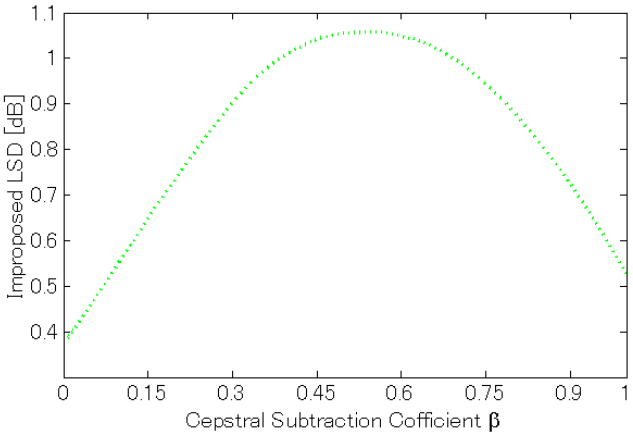
図 6.1: CMS のサブトラクション係数 β を変化させた時の実験結果：縦軸は SEGSNR の改善量，横軸はサブトラクション係数 β を示す



(a) $T_{60} = 0.25\,s$



(b) $T_{60} = 0.5\,s$



(c) $T_{60} = 1.0\,s$

図 6.2: CMS のサブトラクション係数 β を変化させた時の実験結果：縦軸は LSD の改善量，横軸はサブトラクション係数 β を示す

6.2 CMS のエコーに対する耐性評価実験

6.2.1 目的

CMS の性能を評価するため，まずエコーに対しての有効性を検証する。RIR における初期反射は複数のエコーが重なりあわず，時系列上に並んでいる状態と見なせる。このため，まず単一なエコーに対する CMS の有効性を検証する必要がある。

6.2.2 実験条件

実験音は 16 kHz にダウンサンプリングを行った。フレームの切り出しにはハニング窓を使用し，CMS における FFT の分析フレーム長は 512 (32 ms)，オーバーラップは 1/4 となっている。平均正規化に用いる過去フレームは計 60 個とし，忘却係数のパラメータ α を 0.008，サブトラクション係数 β は 0.14 に設定した。

6.2.3 実験音の作成手順

目的信号に遅延時間 δ を加えたエコーを加えることで受信信号を作成する。

$$X_i(k,l) = S_i(k,l) + \gamma S_i(k-\delta,l). \tag{6.1}$$

ただし，係数 γ はエコーと目的信号との全区間 SN 比を調整するための係数である。この係数の調整には，式 (4.1) を用いて行った。また，$S_i(k,l) = HRTF_i(k,l)S(k,l)$ は目的信号スペクトルを表しており，$HRTF_i(k,l)$ は HRTF のスペクトルを示している。本実験では，正中面に対応する HRTF を用いた。エコーの遅延時間を 0.05 s から 0.05 s 刻みで 1 s まで，全区間 SN 比を 0 dB から 2 dB 刻みで 20 dB まで変化させた。

6.2.4 実験結果と考察

実験結果を図 6.3 ~ 7 に示す。図 6.4 と図 6.5 は全区間 SN 比を 0 dB から 2 dB 刻みで 10 dB まで，エコーの遅延時間を 0.05 s から 0.05 s 刻みで 1 s まで変化させた実験音を用いた結果である。また，図 6.7 と図 6.8 は全区間 SN 比を 0 dB から 2 dB 刻みで 20 dB まで，エコーの遅延時間を 0.05 s から 0.05 s 刻みで 0.4 s まで変化させた実験音を用いた結果である。CMS は 14 dB までの全区間 SN 比を持つエコーに効果が見られる。また，エコーに付加される遅延が 0.5 s を超えると CMS では効果が見られず，それ以上の遅延が付加されると処理信号を改悪してしまうことが分かった。これは，CMS において平均正規化を行う計 60 フレームをオーバーラップを考慮してその長さを合計すると 480 ms となることに起因する。480 ms 以上の遅延時間を加えると平均正規化により RIR の振幅ケプストラムが正確に推定できないことになるからである。以上のことから，14 dB

以下の全区間 SN 比を持ち，なお且つ 0.5 s 以下の遅延時間が付加されたエコーに対して CMS は有効であることが示された。

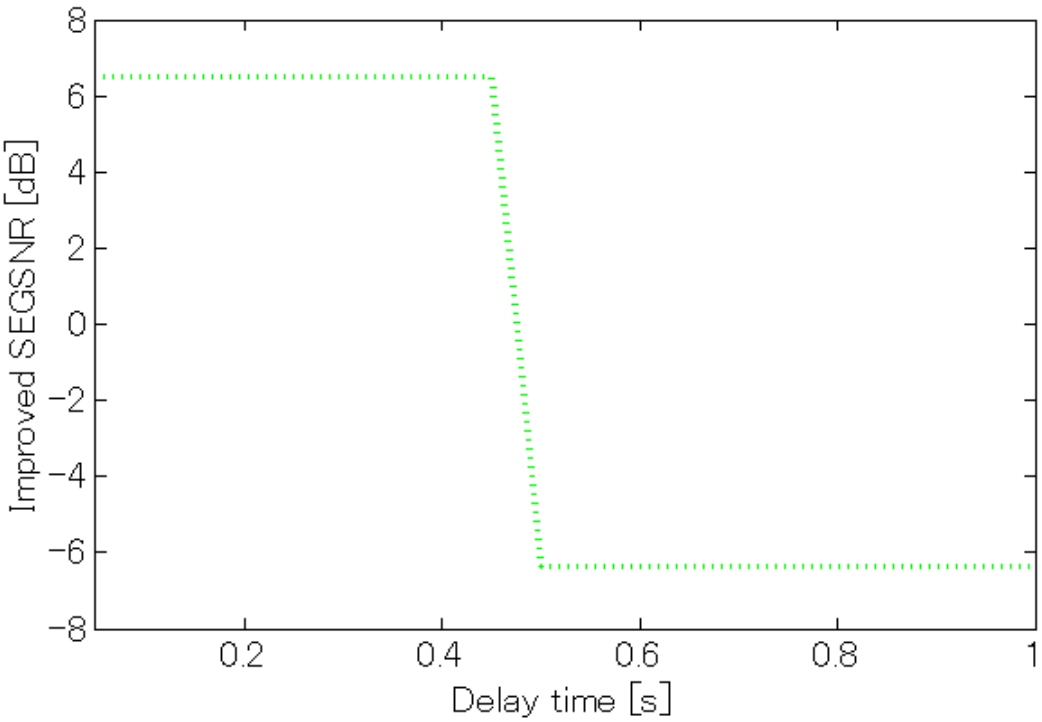


図 6.3: CMS のエコーに対する耐性評価実験結果：縦軸は SEGSNR の改善量，横軸はエコーに付加した遅延時間を示す

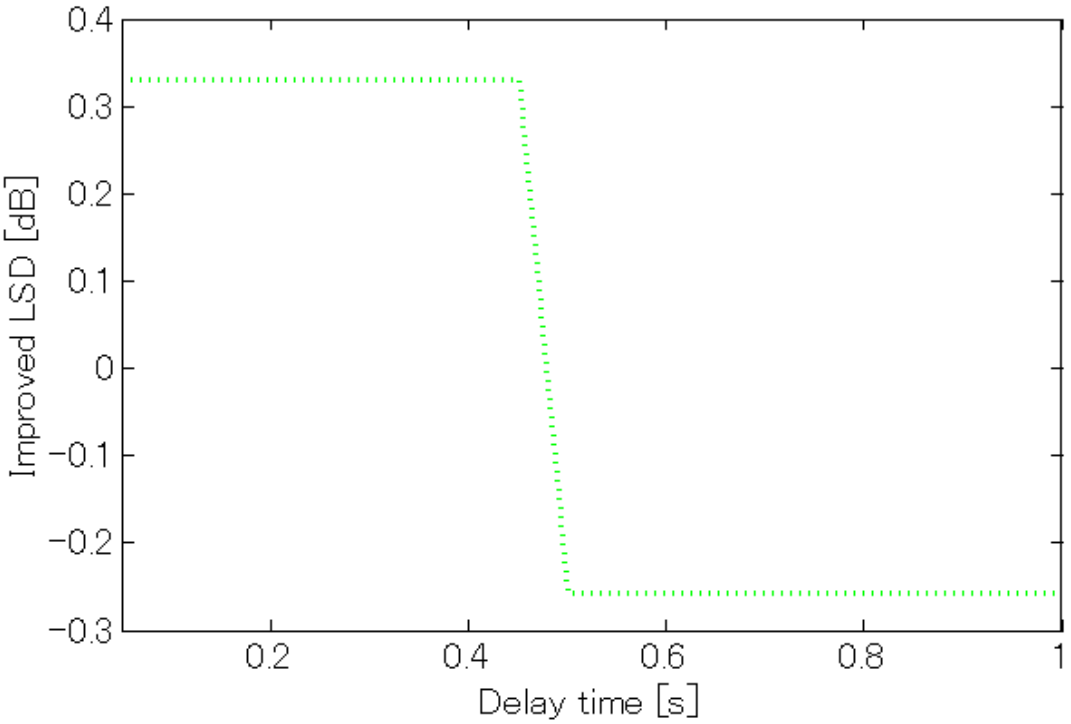


図 6.4: CMS のエコーに対する耐性評価実験結果：縦軸は LSD の改善量，横軸はエコーに付加した遅延時間を示す

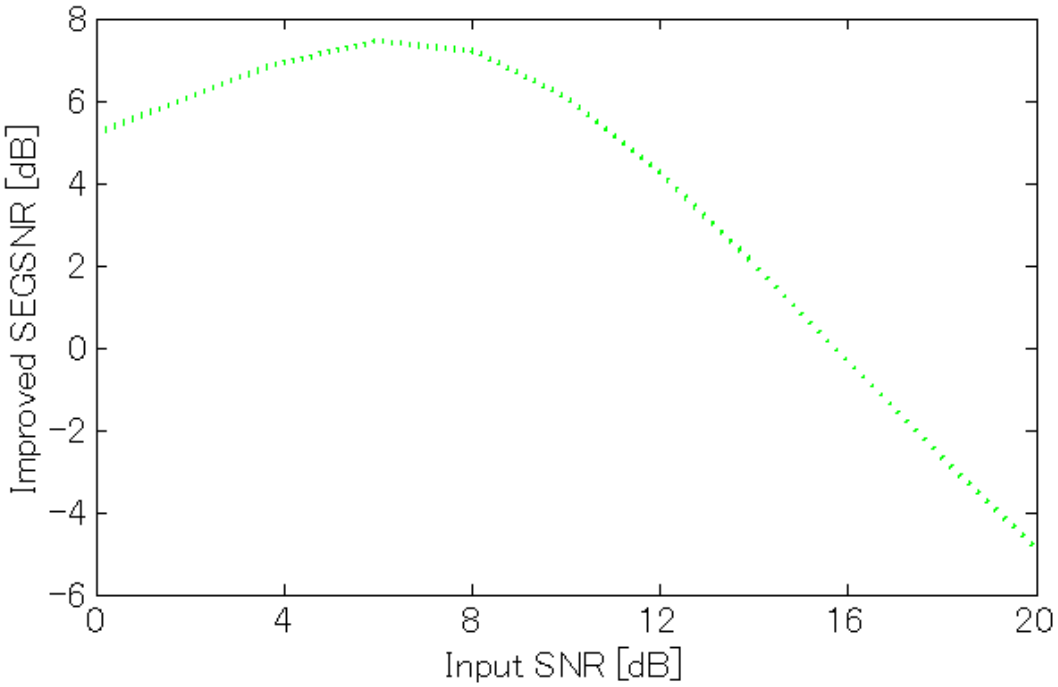


図 6.5: CMS のエコーに対する耐性評価実験結果：縦軸は SEGSNR の改善量，横軸は受信信号の全区間 SN 比を示す

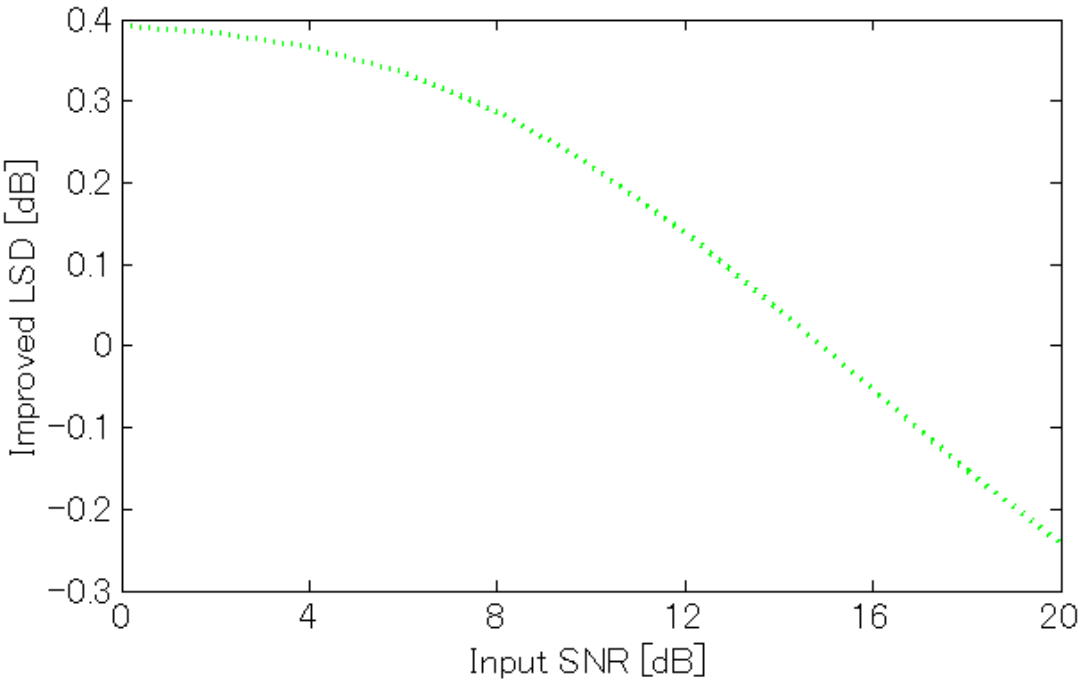


図 6.6: CMS のエコーに対する耐性評価実験結果：縦軸は LSD の改善量，横軸は受信信号の全区間 SN 比を示す

6.3 CMS の残響耐性評価実験

6.3.1 目的

CMS が単一なエコーだけでなく，残響抑圧にも効果があることを実験により証明する。実験音には残響を付加した音声信号を用いる。

6.3.2 実験条件

6.2.2 節の実験条件と同様である。

6.3.3 実験音の作成手順

4.2.3 節と同様の実験音を用いる。

6.3.4 実験結果と考察

実験結果を図 6.7，図 6.8 に示す。残響時間が 0.75 s のとき，SEGSNR 改善量が最大値を取り，それ以降は SEGSNR の改善量がなだらかに減少していく。このことから，CMS は残響時間が長くなるにつれて増大して行く後部残響成分の抑圧には効果を示さず，一定量の初期反射成分を抑圧していることが分かる。また LSD の改善量は残響時間が長くなるのに従い，僅かながら増加していく傾向にある。以上のことから，CMS は残響抑圧に一定の効果を示していることが分かる。

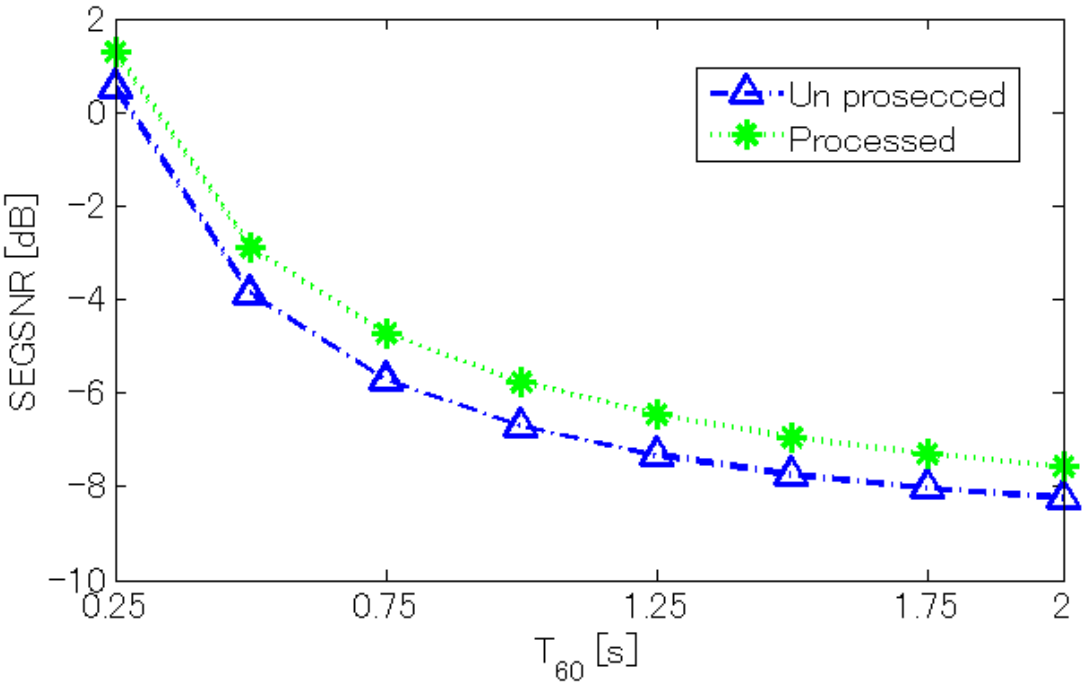


図 6.7: CMS の残響耐性評価実験結果：縦軸は SEGSNR，横軸は残響時間を示す

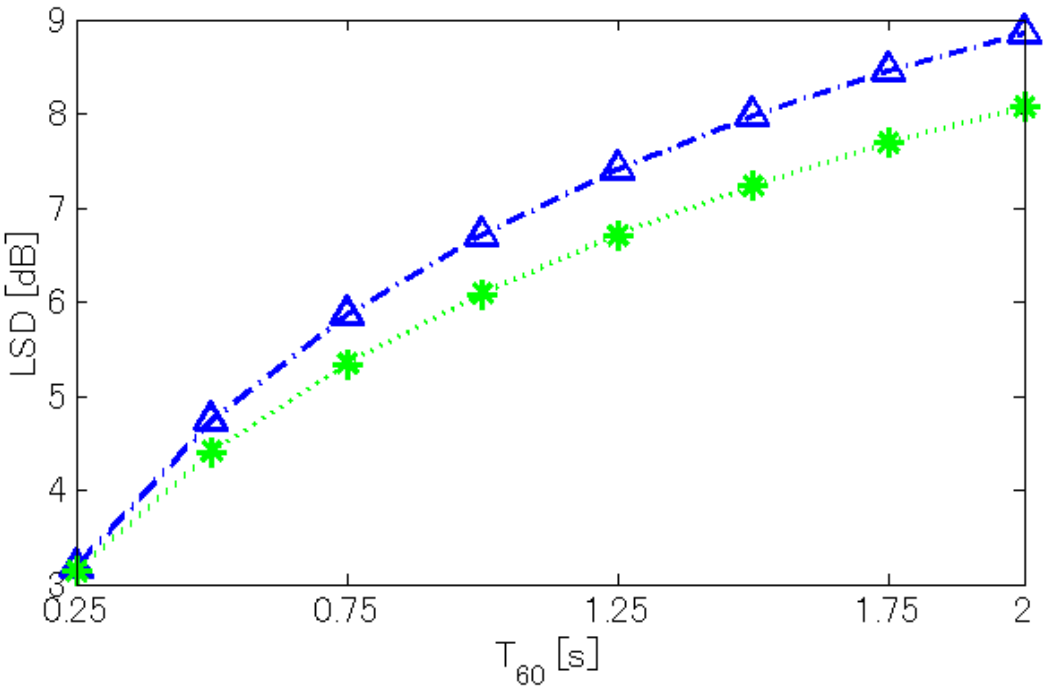


図 6.8: CMS の残響耐性評価実験結果：縦軸は LSD，横軸は残響時間を示す

第7章　CMS + TS–BASE/WF の性能評価実験

7.1　CMS + TS–BASE/WF の残響耐性評価実験

7.1.1　目的

残響環境において，CMS + TS–BASE/WF の性能が従来法である TS–BASE/WF の性能を上回っているか検証する。これにより，TS–BASE/WF のフロントエンドとして CMS が有効に働き，性能の向上に繋がっていることを確かめる。

7.1.2　実験条件

実験音は 16 kHz にダウンサンプリングを行った。フレームの切り出しにはハニング窓を使用し，TS–BASE/WF における FFT の分析フレーム長は 512 (32 ms)，オーバーラップは 1/2 となっている。TS–BASE/WF のターゲット方向を正面方向として等価フィルタの学習を行う。CMS における FFT の分析フレーム長は 512 (32 ms)，オーバーラップは 1/4 となっている。平均正規化に用いる過去フレームは計 60 個とし，忘却係数のパラメータ α を 0.008，サブトラクション係数 β を 0.14 に設定した。

7.1.3　実験音の作成手順

4.2.3 節と同様の実験音を用いた。

7.1.4　実験結果と考察

実験結果を図 7.1，図 7.2 に示す。CMS + TS–BASE/WF における SEGSNR と LSD の改善量は共に，従来法のそれを上回っている。しかしながら，CMS + TS–BASE/WF と従来法どちらも，残響時間が 0.25 s の時，処理信号が悪化している。これは残響時間が短い場合，CMS による LSD 改善量が小さくなることが原因と考えられる。6.2.4 節の実験結果を見て分かるとおり，CMS は直接音に比べてエコーの振幅が一定以上小さくなるとその効果を発揮できない。以上のことから，CMS が受信信号における初期反射成分

を抑圧することにより，TS–BASE/WF の性能向上が見られるものの，残響時間が 0.25 s 以下の場合，処理信号の LSD が改善されない可能性が示唆されている。

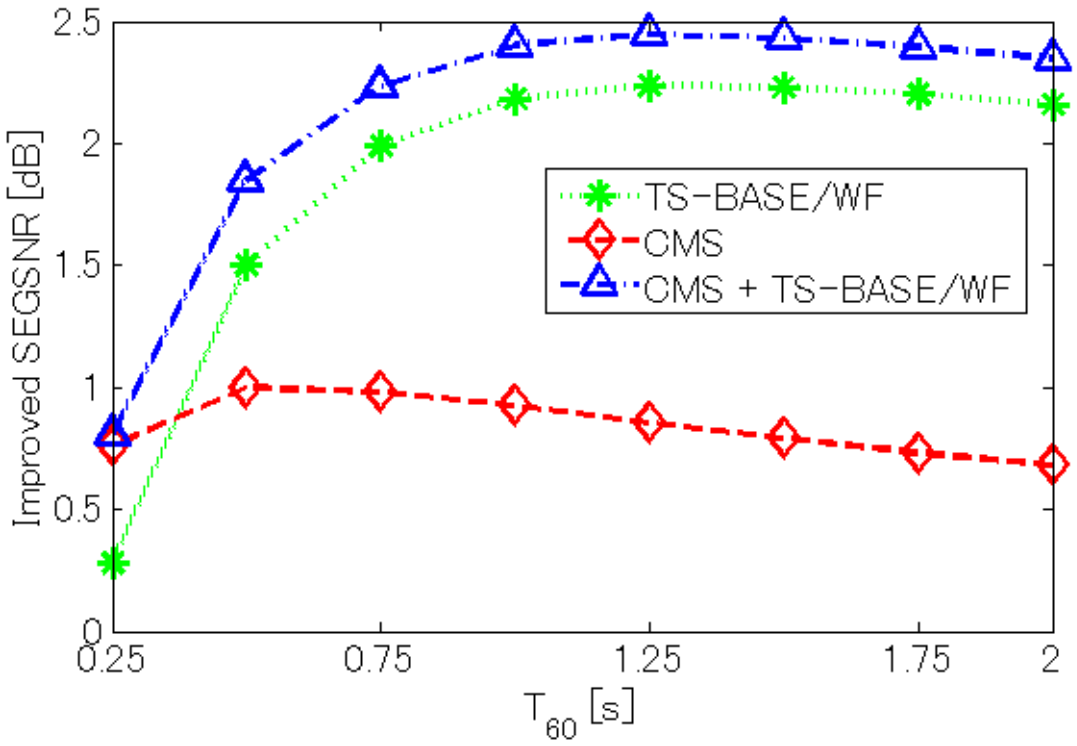


図 7.1: 各手法の残響耐性評価実験結果：縦軸は SEGSNR 改善量，横軸は残響時間を示す

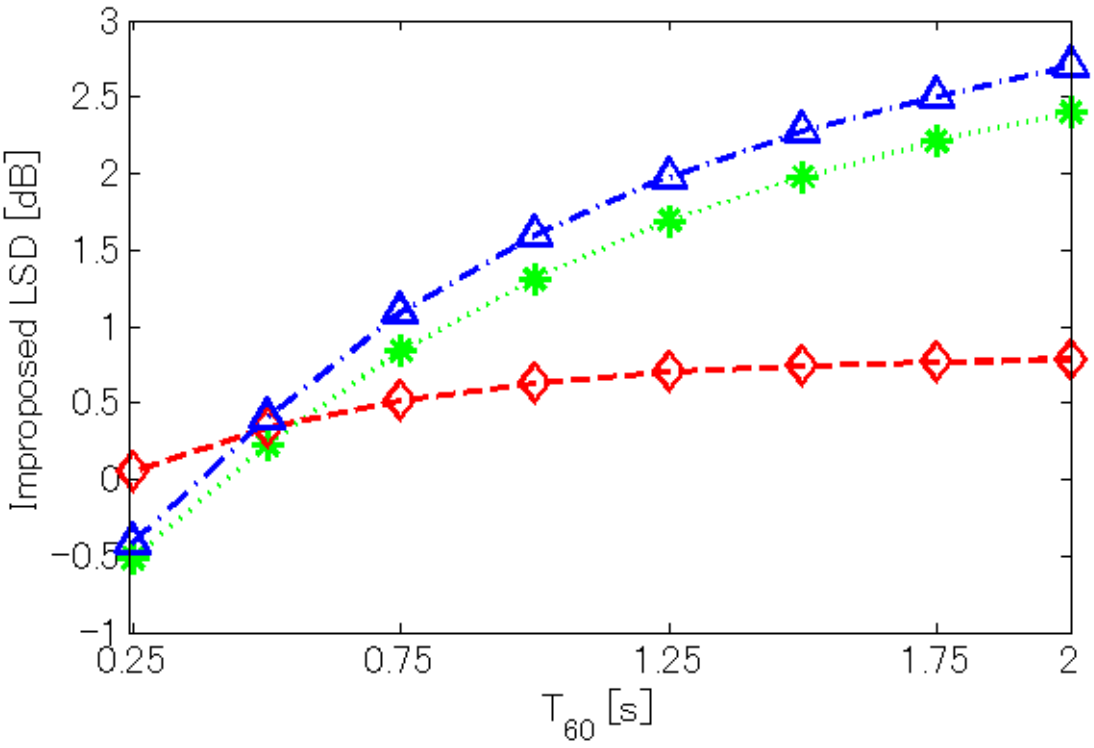


図 7.2: 各手法の残響耐性評価実験結果：縦軸は LSD 改善量，横軸は残響時間を示す

7.2 CMS + TS–BASE/WF の初期反射と後部残響に対する耐性評価実験

7.2.1 目的

前節では残響時間が 0.75 s 以上の場合，CMS + TS–BASE/WF の性能が従来法を上回っていることが確認できた。しかし，前節の結果だけでは CMS が効果的に機能しているとは言い切れない。そこで CMS が受信信号の初期反射成分を十分に抑圧し，当初の目論見通り TS–BASE/WF が残された後部残響成分の抑圧に機能していること実験で確かめる。

7.2.2 実験条件

7.1.2 節の実験条件と同様である。

7.2.3 実験音の作成手順

4.3.3 節と同様の実験音を用いた。

7.2.4 実験結果と考察

実験結果を図 7.3，図 7.4 に示す。ここで，4.3.4 節実験結果である図 4.5 と図 4.6 との比較する。図 4.5 では反射回数が 20 回までの反射音が TS–BASE/WF により十分に抑圧できていないことが分かる。また，図 4.6 においては反射回数が 16 回までの反射音を抑圧できていない。一方，図 7.3 と図 7.4 より，同様の反射回数の反射音を CMS により抑圧できている。しかしながら，図 7.3 では反射回数が 10 まで，図 7.4 では反射回数が 2 までの反射音を完全に抑圧できない結果となった。

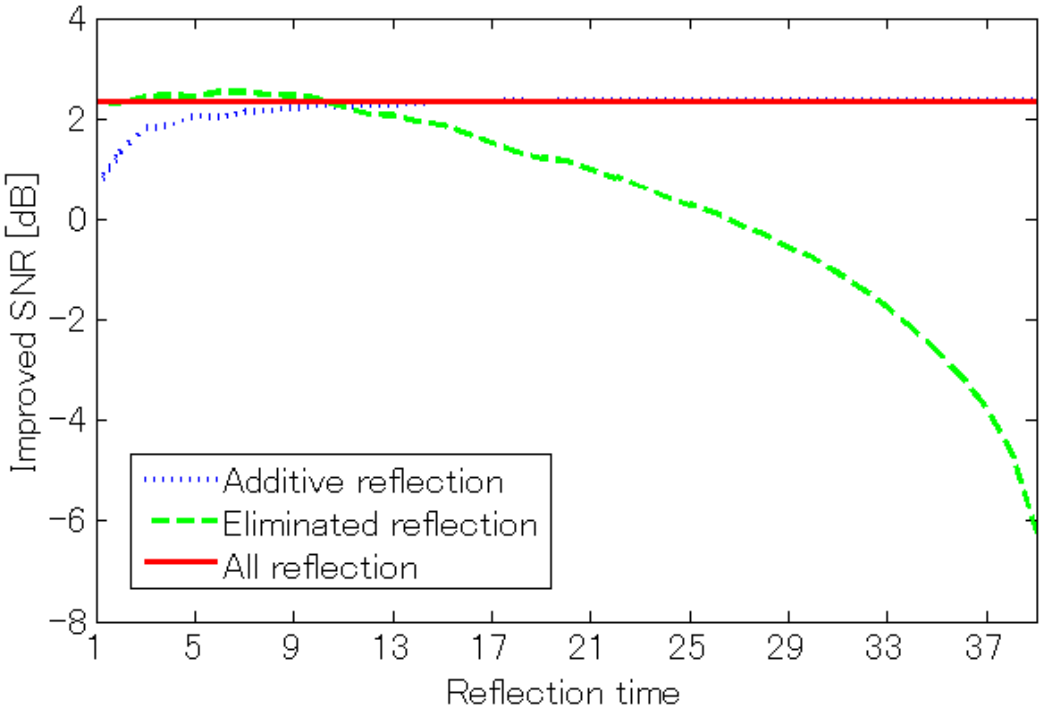


図 7.3: 初期反射と後部残響に対する CMS + TS–BASE/WF の性能評価実験：縦軸は SEGSNR の改善量，横軸は反射音の反射次数を示す

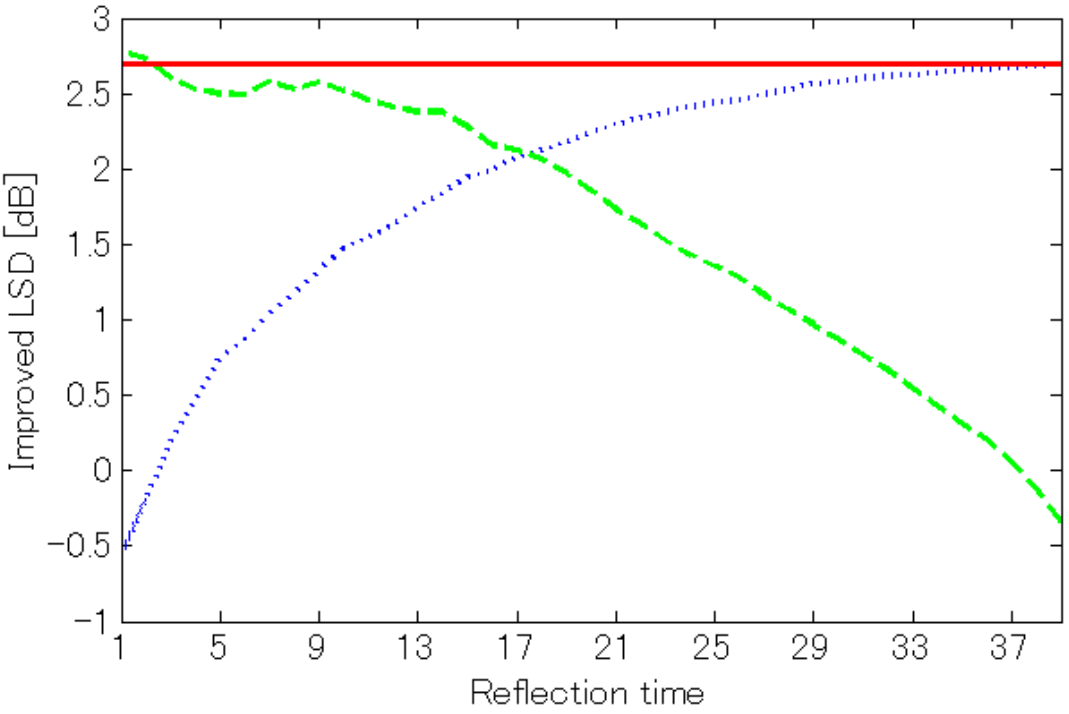


図 7.4: 初期反射と後部残響に対する CMS + TS–BASE/WF の性能評価実験： 縦軸は LSD の改善量，横軸は反射音の反射次数を示す

7.3 CMS + TS–BASE/WF の雑音耐性評価実験

7.3.1 目的

TS–BASE/WF は雑音環境下での使用を想定した音声強調手法である。しかし，CMS + TS–BASE/WF は TS–BASE/WF の残響環境下における問題点に基づいて構築された。そこで，雑音環境下における CMS + TS–BASE/WF の動作を明らかにするため実験を行った。

7.3.2 実験条件

7.1.2 節の実験条件と同様である。

7.3.3 実験音の作成手順

4.1.3 節と同様の実験音を用いた。

7.3.4 実験結果と考察

実験結果を図 7.5，図 7.6 に示す。CMS + TS–BASE/WF の SEGSNR の改善量は負の値を示しており，処理信号を改悪している。LSD の改善量は改悪していないものの，従来法に大きく劣る結果となった。これは CMS が実験音を処理する際，受信信号のフレーム間で平均正規化することにより，雑音成分を十分に推定できないことに起因する。以上のことから，CMS + TS–BASE/WF の対雑音性能は従来法と比べて劣る結果となった。

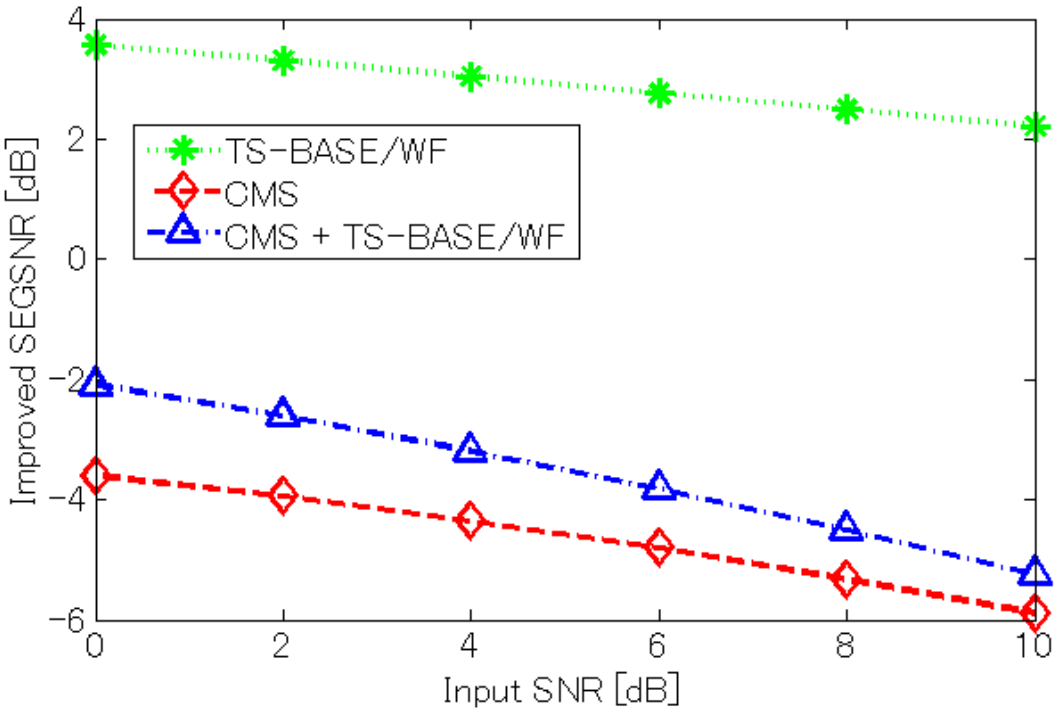


図 7.5: 各手法 の雑音耐性評価実験結果：縦軸は SEGSNR 改善量，横軸は受信信号の全区間 SN 比を示す

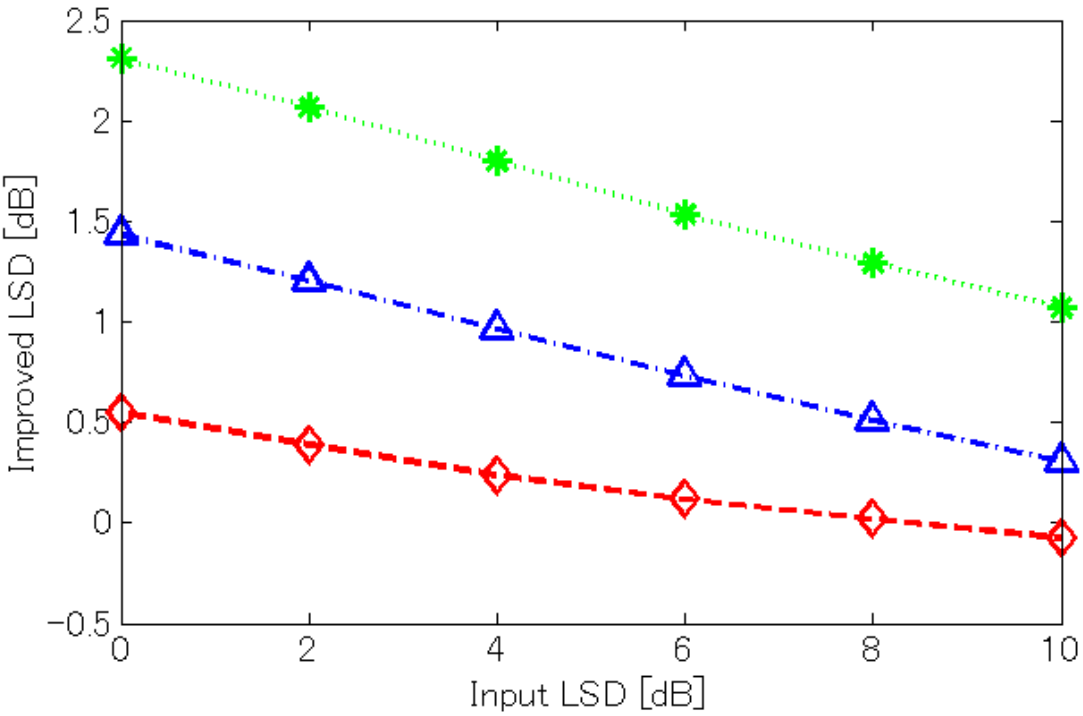


図 7.6: 各手法 の雑音耐性評価実験結果 : 縦軸は LSD 改善量，横軸は受信信号の全区間 SN 比を示す

7.4 CMS + TS–BASE/WF の雑音残響耐性評価実験

7.4.1 目的

7.3.4 節の結果，雑音環境下における CMS + TS–BASE/WF の性能が明らかになった。従来法を大きく下回る結果となったが，残響が存在する環境下であれば CMS が有効に働く可能性がある。そこで，CMS + TS–BASE/WF の雑音残響耐性評価実験を行うことにより，従来法の性能を上回ることか検証を行う。

7.4.2 実験条件

7.1.2 節の実験条件と同様である。

7.4.3 実験音の作成手順

目的信号と雑音，共にデータベース上の音声を用い，4 話者の組み合わせ全 9 通りすべてを作成した。RIR は第 3 章で述べた作成手順を用いて合成し，残響時間は 2.0 s，反射音の総数を 40 に設定した。目的信号に対応する RIR は正中面に音源を配置し，雑音に対応する RIR は 45 度の位置に音源を配置した。雑音源は室の座標 (2.19 m $\times$ 1.74 m $\times$ 1.5 m) に設置する。また式 (4.1) を用いて，全区間 SN 比を 0 dB から 2 dB 刻みで 10 dB まで変化させた。

7.4.4 実験結果と考察

実験結果を図 7.7，図 7.8 に示す。CMS + TS–BASE/WF の SEGSNR と LSD の改善量が共に TS–BASE/WF の値よりも大きいことが分かる。一方，CMS の改善量はほぼ一定の値である。したがって，雑音残響環境下において CMS が雑音の影響を受けずに初期反射成分を抑圧することが可能であることが分かった。このことから，CMS + TS–BASE/WF は雑音と残響共に効果が見られ従来法であるの性能を上回る結果を得ることができた。

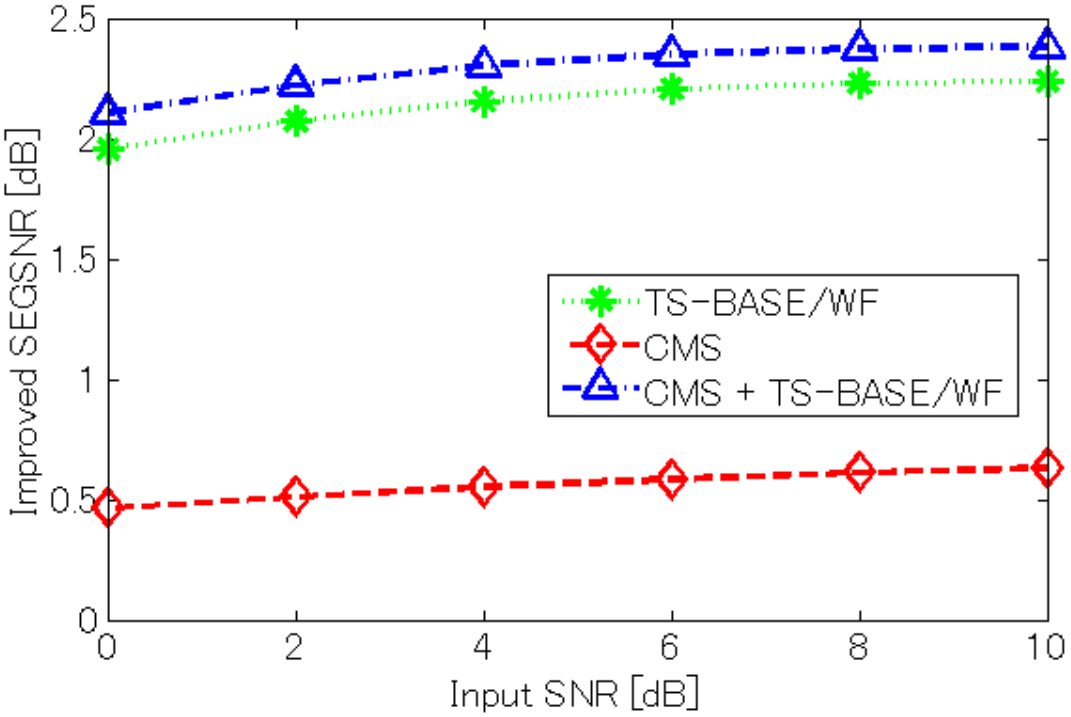


図 7.7: 各手法の雑音残響耐性評価実験結果：縦軸は SEGSNR 改善量，横軸は受信信号の全区間 SN 比を示す

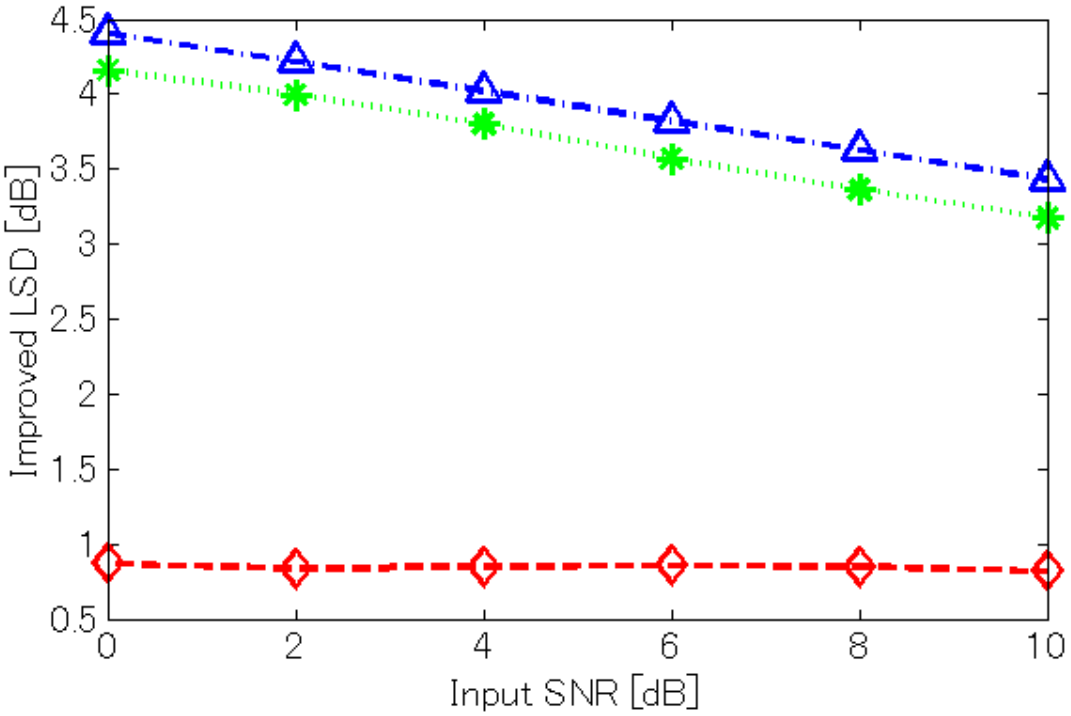


図 7.8: 各手法の雑音残響耐性評価実験結果：縦軸は LSD 改善量，横軸は受信信号の全区間 SN 比を示す

第8章　結論

8.1　本研究で明らかになったこと

本研究では，音声アプリケーションや補聴器への導入を想定した雑音残響環境下における音声強調手法を提案した。以下に第 4 章 ～ 第 7 章までの要約を記載する。

第 4 章

TS–BASE/WF の雑音耐性と残響耐性を測定した。特に残響耐性を測定することにより，雑音抑圧部における Wiener filter が初期反射の影響を受け，処理信号を悪化させることが分かった。

第 5 章

第 3 章で明らかになった TS–BASE/WF の問題点を参考に TS–BASE/WF の改良案を検討した。その結果，初期反射の抑圧に効果が期待できる CMS を TS–BASE/WF のフロントエンドとして採用した。

第 6 章

改良手法のフロントエンドとして採用した CMS の性能評価を行った。直接音に対して 14 dB 以下の全区間 SN 比を持ち，0.5 s 以下の遅延時間が付加されいる単一のエコーに対しては，CMS によって抑圧可能であることを示した。また，CMS の残響耐性に関しては初期反射抑圧に効果があることを示唆する結果が得られた。

第 7 章

CMS + TS–BASE/WF の性能評価実験を行った。残響環境下においては CMS + TS–BASE/WF の性能が従来法と比べて向上していることが分かった。これは CMS がフロントエンドとして初期反射を抑圧しているためである。また，CMS + TS–BASE/WF の雑音耐性は従来法に及ばないものの，雑音残響環境下においては，CMS + TS–BASE/WF

の性能が従来法を上回る結果となった。このことから，CMS をフロントエンドとして用いる場合，残響が存在しない環境下において有効ではないことが示された。

以上のことから，本研究で提案する CMS + TS–BASE/WF は雑音残響環境下において有効な手法であることが言える。また，音声アプリケーションや補聴器に CMS + TS–BASE/WF が導入されることで性能の向上が期待できる。

8.2 今後の展望

本研究で提案した音声強調手法の性能向上を図るため，今後の展望を以下に示す。

- 本研究では，従来法と提案した CMS + TS–BASE/WF を性能評価実験において，SEGSNR と LSD の客観評価尺度を用いている。しかしながら，この 2 つの客観評価尺度だけでは処理信号の音質が十分に改善されているとは言い切れない。そのため，SEGSNR と LSD 以外の客観評価尺度を用いた性能評価実験を行うことが望ましい。

- 客観評価だけでなく主観評価を行っていないため，CMS + TS–BASE/WF により聴感上の音質が改善されているか不明である。

- 6.1.4 節から，CMS のサブトラクション係数 β が室の残響時間によって最適な値が異なることが分かっている。したがって，CMS + TS–BASE/WF を使用する室の残響時間計測を事前に行い，サブトラクション係数 β をその残響時間に適した値に逐次設定するが望ましい。このことから，正確な残響時間推定を行う手法と組み合わせることにより，CMS + TS–BASE/WF の性能向上が見込まれる。

- 7.1.4 節の実験結果では，残響時間が 0.25 s の時，CMS + TS–BASE/WF を用いたとしても処理信号が歪んでしまった。そのため，残響時間が短い場合においては，まだ改良の余地があると言える。

謝辞

本研究を行うに当たり，終始ご指導賜りました北陸先端科学技術大学院大学 情報科学研究科 赤木 正人 教授に深謝致します。また，折に触れてご指導いただきました北陸先端科学技術大学院大学 情報科学研究科 鵜木 祐史 准教授，宮内 良太 助教に心より感謝致します。加えて，本研究を始めるにあたり貴重な助言を賜りました中国科学院声学研究所 李 軍鋒 教授に深謝致します。さらに本研究を遂行していく上で，熱心な議論と多面にわたる協力を賜った北陸先端科学技術大学院大学 情報科学研究科 党 建武 教授，末光 厚夫 助教，川本 真一 助教に厚く御礼申し上げます。

本研究を行うにあたり，多面に渡りご協力いただいた赤木研究室ならびに，鵜木研究室の諸先輩方および皆様に感謝致します。

最後に，大学院在学中に自由な研究の場を与えていただき，暖かく見守ってくれた両親，両祖父母，妹に心から感謝致します。

参考文献

[1] J. C. Junqua and J. P. Haton, *Robustness in automatic speech recognition*, Kluwer Academic Publishers, Boston, 1996.

[2] 飛田 端広, 菅村 昇, “音声認識における周囲環境の影響,” 音響誌, Vol. 51, No. 4, pp. 331–335, 1995.

[3] A. J. Duquesnoy and R. Plomp, “Effect of reverbration and noise on the intelligibility of sentences in case of presbycusis,” J. Acoust. Soc. Am., Vol. 68, pp. 537-544, 1980.

[4] 境久雄著, 中山剛共著, “聴覚と心理,” コロナ社,1978.

[5] 水島 昌英, 伊藤 憲三, “自動利得制御と雑音抑圧処理が難聴者の音声知覚に及ぼす影響,” 信学技報, SP 96–35, pp.17–24, 1996.

[6] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. ASSP, Vol. 27, No. 2, pp. 113–120, 1979.

[7] 谷口 賢一, 津村 尚志, 福留 公利, “スペクトルサブトラクション法における雑音推定方式,” 音講論 (秋), Vol. I, pp. 175–176, 1994.

[8] 園枝伸行, “雑音レベルの変動を考慮したスペクトラルサブトラクション法,” 音講論 (秋), Vol. I, pp. 245–246. 1994.

[9] H. Gustafsson, S. Nordholm, I. Claesson, “Spectral subtraction with adaptive averaging of the gain function,” EUROSPEECH’99, Vol. 6, pp. 2599–1602, 1999.

[10] 金 学胤, 浅野 太, 鈴木 陽一, 曽根 敏男, “短時間振幅スペクトル推定を用いた 2 チャンネル音声強調法における振幅スペクトル推定について,” 音講論 (秋), Vol. I, pp. 533–534. 1994.

[11] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean–square error ahort–time spectral amplitude estimator,” IEEE Trans. ASSP, Vol. ASSP–32, No. 6, pp. 1109–1121, 1984.

[12] J. S. Lim and A. V. Oppenheim, “All–pole modeling of degraded speech,” IEEE Trans. ASSP, Vol. 26, No. 3, pp. 197–210, 1978.

[13] T. Nakatani, K. Kinoshita, and M. Miyoshi, “Harmonicity–based blind dereverberation for single–channel speech signals,” IEEE Trans. Audio. Speech, and Lancuage Processing, Vol. 15, No. 1, pp. 80–95, 2007.

[14] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” IEEE Trans. ASSP, Vol. 36, pp. 145–152, 1988.

[15] 古家 賢一, 片岡 章俊, “チャンネル間相関行列と音声の白色フィルタを用いた Semi–blind 残響抑圧,” 電子情報通信学会論文誌 A, Vol. J88, No. 10, pp. 1089–1099, Oct. 2005.

[16] 浅野太, “ICA による音響信号の分離,” 電子情報通信学会誌, Vol. 87, No. 3, pp. 175–181, 2004.

[17] 高橋 祐, 高谷 智哉, 猿渡 洋, 鹿野 清宏, “独立成分分析に基づく空間的サブトラクションアレーによる雑音抑圧,” 電子情報通信学会 技術研究報告 EA, Vol. 106. No. 125, pp. 13–18, 2006.

[18] 古屋 武志, 金田 圭一, 五反田 博, “独立成分分析に基づく耐残響音源分離に関する研究,” 電子情報通信学会 技術研究報告 NC, Vol. 105, No. 131, pp. 7–12, 2005.

[19] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, “Multi–step linear prediction based speech enhancement in noisy reverberant enbironment,” Proc. Interspeech 2007, pp. 854–857, 2007.

[20] M. Ebata, T. sone, and, “Improvement of hearing ability bydirectional information,” J. Acoust. Soc. AM, Vol. 43, pp.289–297, 1968.

[21] R. Zelinski, “A microphone array with adaptive post–filtering for noise reduction in reverbrant rooms.” Proc. ICASSP, pp. 2578–2581, 1988.

[22] M. Dörbecker & S. Ernst, “Combination of two–channel spectral subtraction and adaptive Wiener post–filtering for noise reduction and dereverberation.” Proceedings EUSIPCO, pp. 995–998, 1996.

[23] W. Lindemann, “Extension of a binaural cross-correlation model by contralateralinhibition.I. Simulation of lateralization for stationary signals,” J.Acoust. Soc. AM., 80, 1608–1622, 1986.

[24] T. Usagawa, K. Sakai and M. Ebata, “Frequency domain binaural model as the front end of speech recongnition system,” Proc. ICSL98, 1998.

[25] J. Li, S. Sakamoto, M. Akagi, and Y. Suzuki, "A two–stage binaural speech enhancement with wiener filter (TS–BASE/WF) for high–quality speech comunication," Proc. IEEE WSPAA, New Paltz, New York, 2009.

[26] C. T. Duc, J. Li, M. Akagi, "A DOA estimation algorism based on equalization–cancellation theory." Proceeding of INTERSPEECH 2010, pp.2770–2773, 2010.

[27] N. I. Durlach, "Equalization and cancellation theory of binaural masking level differences," JASA, Vol. 35, no. 8, pp. 1206–1218, 1979.

[28] J. F. Culling, M. L. Hawley and R. Y. Litovsky, "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources" Jounal of Acoustic Society of America, p1057–1065, 2004.

[29] P. Scalart, J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in Proc. ICASSP, vol. 2, pp. 629–632, 1996.

[30] J. Li and M. Akagi, "Noise reduction method based on generalized subtractive beamformer," Acoust. sci. and Tech., Vol. 27, No. 4, pp. 206–215, 2006.

[31] I. Cohen, "Multichannel post–filtering in nonstationary noise environments," IEEE trans. Signal Processing, Vol. 52, No. 5, pp. 1149–1160, 2004.

[32] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy–head microphone," URL:http://sound.media.mit.edu/KEMAR.html, 1994.

[33] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," Journal of Acoustic Society of America, p912–915, 1979.

[34] S. T. Neely and J. B. Allen, "Invertibility of a room impulse responce," Journal of Acoustical Society of America 66, 165-169, 1979.

[35] R. W. Schafer, "Echo Removal by Distance Generalized Linear Filtering," Tech. Rept. 466, MIT Reserch Laboratory of Electronics, MIT, Cambridge, Mass., Feb 1969. Also Ph. D. Thesis, Department of Elec. Engineering, MIT, Feb. 1968.

[36] T. G. Stockham, Jr., "Restoration of old acoustic recordings by means of digital signal processing," Preprint, 41st Convention, Audio Engineering Society, New York, Oct. 1971.

本研究に関する業績

学会発表リスト

- 佐々木 裕吉, 赤木 正人, “残響環境下における TS–BASE/WF の性能評価,” 日本音響学会聴覚研究会資料, Vol. 41, No. 7, pp. 567–570, 2011.
- Yuuki Sasaki and Masato Akagi, “Speech enhancement technique in noisy reverberant environment using two microphone arrays,” Proc. NCSP12, 2012. (to appear)
- 佐々木 裕吉, 赤木 正人, “二チャンネルマイクロホンアレイによる雑音残響環境下音声強調手法,” 日本音響学会 2012 年 春季研究発表会, 講演論文集, 1–Q–7, 2012. (to appear)