側抑制性重み付けを用いた雑音環境下におけるSTRAIGHT分析合成系の品質改善

戸田　智基[†]　坂野　秀樹[††]　梶田　将司[††]　武田　一哉[††]
板倉　文忠[††]　鹿野　清宏[†]

Improvement of STRAIGHT Method under Noisy Conditions Based on Lateral Inhibitive Weighting

Tomoki TODA[†], Hideki BANNO[††], Syoji KAJITA[††],
Kazuya TAKEDA[††], Fumitada ITAKURA[††], and Kiyohiro SHIKANO[†]

あらまし　雑音環境下におけるSTRAIGHT分析合成系の音声の品質を改善する方法として，側抑制性重み付け（LIW: Lateral Inhibitive Weighting）によるスペクトルシェーピングを提案する．LIWのディップの深さと帯域幅を適切に選ぶことによって，SN比が0dBから10dB程度の範囲では，ケプストラムひずみが男性女性ともに0.2dB程度改善されることがわかった．ところが，SN比が高い場合には，LIWを利用したSTRAIGHTで合成した場合，ひずみが大きくなる傾向が見られた．この問題への対策として，瞬時フレームごとにSN比を推定して，SN比が高くなるにつれてLIWの効果を弱める処理を導入した結果，SN比が高いときでも，従来のSTRAIGHTと同程度の品質で音声が合成できることがわかった．

キーワード　雑音抑圧，STRAIGHT，側抑制性重み付け，SN比推定

1. まえがき

音声符号化技術の発展に伴う近年の携帯電話の普及には，目を見張るものがある．しかし，携帯電話の品質は十分なものとはいいがたい．その理由の一つとして，雑音環境下における極端な品質低下が挙げられる．この問題の解決法としては，波形符号化を用いるという方法などがある[1]．しかし，波形符号化を用いた場合には，伝送に必要な情報量が多くなってしまうため，昨今の携帯電話の事情を考えると情報量がより少なくてすむ音声符号化方式が望ましい．そこで，本論文では，情報量が少なくてすむといわれている音声分析合成系[2]においてこの問題に対処することを考える．

ところが，音声分析合成を用いた場合には，雑音環境下での音声に対し処理を行うと，品質が極端に低下するという問題点がある[1]．この原因の一つとして，求められるスペクトルは雑音の影響を大きく受けるということが考えられる．そこで，雑音を取り除く方法として，スペクトルサブトラクション法[3]を用いるという方法がある．しかし，雑音スペクトルを推定するために，非音声区間の検出が必要となる．そこで本論文では，非音声区間の検出が必要でない方法として，スペクトルに対して側抑制性重み付け（Lateral Inhibitive Weighting）[4],[5]を行い，雑音の影響の少ないスペクトルを求めることにより，品質を改善することを検討する．Lateral Inhibitive Weightingの耐雑音性は，小原らによって示されており[6]，近年の研究では，1994年に梶田らによって提案された帯域分割–自己相関（SBCOR）分析法（Subband–Crosscorrelation Analysis）[7]～[9]などで用いられている．

本論文では，音声分析合成方式として，STRAIGHT（Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum）[10],[11]を用いる．STRAIGHTは音声分析合成方式でありながら高品質な音声を合成することが可能である[12]．一方で，高品質な音声を合成するために情報

[†] 奈良先端科学技術大学院大学情報科学研究科，生駒市
Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0101 Japan
[††] 名古屋大学大学院工学研究科，名古屋市
Graduate School of Engineering, Nagoya University, Nagoya-shi, 464-8603 Japan

量を大幅に拡大するという欠点がある．しかし，原理的には VOCODER [13] そのものであるため，品質を保ったまま情報量を削減することができる可能性がある [12].

本論文の構成は，まず **2.** で提案手法の構成について説明する．**3.** では提案手法の性能評価実験と，その問題点について述べる．**4.** ではその問題点を改善するための提案手法と，その性能評価実験について述べる．**5.** では提案手法の主観評価実験について述べ，最後に **6.** において本論文の結論と今後の課題について述べる．

2. 手　　法

2.1 音声分析合成系 STRAIGHT

STRAIGHT は，音声の基本周期に応じたガウス窓で分析し，平滑化操作により時間周波数特性から周期性の影響を取り除き，合成時には，位相制御による音源を構成するといった操作を行う音声分析合成方式である．実時間処理への適用のため，逐次近似のような収束の判定のための繰返し演算は含んでおらず，すべて短時間フーリエ変換に基づいているため，携帯電話などへの応用に適していると考えられる．また，STRAIGHT によって合成された音声の品質は高く，高い品質を保ったまま，基本周波数，スペクトル包絡，発声速度等のパラメータを各々独立に変換，加工することができることが特徴である．しかし，雑音環境下においては STRAIGHT によって合成される音声の品質は極端に低下してしまう．この原因の一つとして，雑音を含んだスペクトルを求めてしまい，それを用いて音声を合成することが挙げられる．本論文では，この問題に対処するため，雑音抑圧効果のある，Lateral Inhibitive Weighting 処理を行う．

2.2 Lateral Inhibitive Weighting

2.2.1 Lateral Inhibitive Weighting の概要

パワースペクトルに対する Lateral Inhibitive Weighting（LIW）は，自己相関領域では，低次成分の抑圧効果をもつ．このため，白色雑音のように自己相関領域の値が 0 次に集中する雑音を抑圧することができる．

本論文で用いる LIW は次式で表される．

$$w(f) = e^{-2\left(\frac{D}{\frac{2}{3}W}\right)^2 (f-f_c)^2} \cos\left\{\frac{2\pi(f-f_c)}{\frac{2}{3}W}\right\} \tag{1}$$

W は LIW の帯域幅を変化させるパラメータであり，D は LIW のディップの深さを変化させるパラメータである．f は周波数を示し，f_c は LIW の中心周波数を示す．W の値に係数 2/3 を掛けることにより $w(f_c \pm W/2) = 0$ となる．$w(f) = 0$ となる $f = f_c \pm f_1, f_2, \ldots$ において，$2f_2$ を LIW の帯域幅と考えれば W は帯域幅を示すことになる．そのため，本論文では係数 2/3 を用いる．また，本論文では，いかなる中心周波数においても，帯域幅，ディップの深さは固定しており，パワースペクトルに対してすべての中心周波数において LIW を用いる．つまり，パワースペクトルに対して LIW を畳み込む処理を行う．なお，その際には積分値が 1 となるように正規化された LIW を用いる．

中心周波数が 2000 Hz，パラメータ $W = 1500$，$D = 1.5$ のときの LIW の形状を図 1 に示す．また，パラメータ $D = 1.5$ とし，パラメータ W を変化させたときの LIW の帯域幅の変化を図 2 に，パラメー

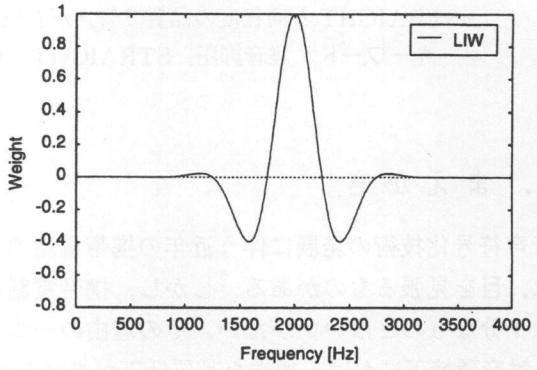


図 1　LIW の形状 ($f_c = 2000$, $W = 1500$, $D = 1.5$)
Fig. 1　Shape of lateral inhibitive weighting.

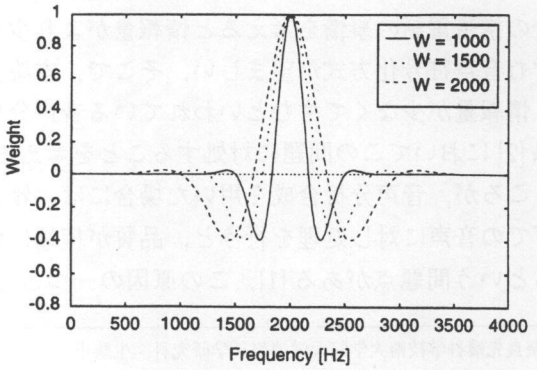


図 2　W の変化に対する LIW の帯域幅の変化 ($f_c = 2000$, $D = 1.5$)
Fig. 2　Band width of LIW controlled by W.

タ $W = 1500$ とし，パラメータ D を変化させたときの LIW のディップの深さの変化を図 3 に示す．どちらも中心周波数は 2000 Hz である．これらの図から，帯域幅を変化させるパラメータ W の値を大きくするにつれて，LIW の帯域幅は広くなっていき，ディップの深さを変化させるパラメータ D の値を大きくするにつれて，振動が少なくなり，ディップの深さは浅くなっていくことがわかる．

2.2.2 Lateral Inhibitive Weighting の耐雑音性の原理

白色雑音のような周期性をもたない雑音の自己相関の値は 0 次に集中する．一方，音声信号の自己相関の値は周期成分をもつ．音声信号に雑音が加法的に付加されるとすると，雑音の影響は，自己相関係数の 0 次，つまり，パワーを検出するよりも，音声信号の周期性が現れる自己相関係数の高次成分を取り出した方が少ない．このような原理で LIW の耐雑音性はもたらされる．パワースペクトルに対して，すべての中心周波数において式 (1) の LIW を用いることは，自己相関領域において図 4 に示すような窓を掛けることに相当する．図 4 より，自己相関係数の高次成分を強調していることがわかる．

2.3 LIW を用いた STRAIGHT

雑音環境下における音声に対し，STRAIGHT を用いると，求められるスペクトルは雑音の影響を大きく受ける．そこで，STRAIGHT により得られたスペクトルに対し，すべての中心周波数において LIW 処理を行うことにより，雑音の影響の少ないスペクトルを求めることを考える．

先に述べたように，LIW は音声の周期性が現れる自己相関係数の高次成分を強調するものである．しかし，従来の STRAIGHT のガウス窓で得られたスペクトルには，LIW の効果が十分得られるだけの周期成分は含まれていない．そのため，ガウス窓の幅を従来の STRAIGHT で用いている幅より広げる必要がある．従来の STRAIGHT で用いられているガウス窓を次式に示す．

$$g(t) = e^{-\pi(t/\eta T_0)^2} \tag{2}$$

T_0 は音声の基本周期を示す．η は窓の時間方向の伸長の程度を示すパラメータであり，本論文では従来の STRAIGHT に対して $\eta = 1.4$ のガウス窓を用いる [14]．また，ガウス窓の幅を広げても，平滑化窓 [15] を自己相関領域で掛けるため，周期成分はほとんどなくなってしまう．そこで，ガウス窓を広げるにつれて，平滑化窓の時間領域における幅も広げることにする．具体的にはガウス窓の幅を従来の n 倍にしたときには，平滑化窓の幅も時間領域において n 倍にする．この処理を行うことにより STRAIGHT の平滑化条件は満たされなくなる．しかし，音声の周期成分が現れる自己相関領域の高次成分が失われないために，白色雑音のような周期性をもたない雑音環境下においてはある程度の品質の改善が期待できる．白色雑音環境下において予備実験を行ったところ，ガウス窓の幅と平滑化窓の時間領域における幅を広げることにより，STRAIGHT の品質が向上することがわかった．そこで，本論文では雑音環境下における STRAIGHT の更なる品質の改善を行うために，幅が拡大された平滑化窓による処理が行われたパワースペクトルに対して最適な帯域幅とディップの深さをもつ LIW を畳み込むことを考える．

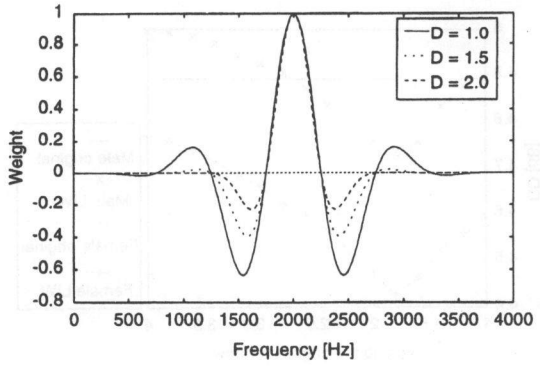


図 3 D の変化に対する LIW のディップの深さの変化 ($f_c = 2000$, $W = 1500$)
Fig. 3 Side-dip depth of LIW controlled by D.

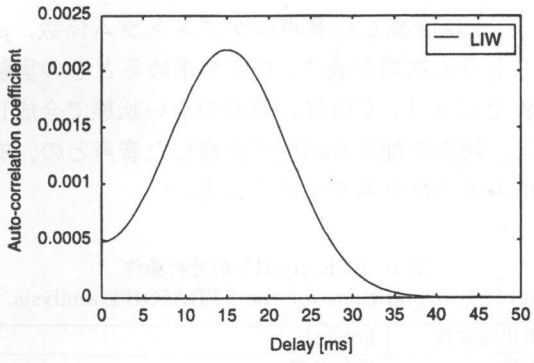


図 4 LIW の自己相関領域表現 ($W = 100$, $D = 1.5$)
Fig. 4 LIW function in auto-correlation domain ($W = 100$, $D = 1.5$).

また，LIW は負の係数を含んでいるので，LIW をすることにより，スペクトルに負の部分が生じることがある．そこで，非負でかつ滑らかなスペクトルを得るために，LIW をした後に半端整流関数 [15] を用いている．半端整流関数は，次式で表されるものを用いる．

$$\eta(x) = \frac{1}{2}\log_e(1 + e^{2x}) \tag{3}$$

ここで，変換前の値を x，変換後の値を $\eta(x)$ としている．なお，STRAIGHT では近似の誤差を各帯域で知覚的に一様とするために，平均値が 1 となる正規化パワースペクトルを求めそれを対象に平滑化を行う [15]．そのため，x は平滑化窓による処理が行われた正規化パワースペクトルに LIW を畳み込むことによって得られたパワースペクトルの値となる．

LIW を STRAIGHT に利用するにあたり，帯域幅，ディップの深さの決定が重要となる．予備実験を行ったところ，LIW の帯域幅を固定幅にしても LIW の効果は得られたが，ピッチに応じて変化させたときの方が，よりいっそうの効果が見られた．そのため，LIW の帯域幅は，ピッチ同期にする．

3. STRAIGHT に LIW を用いることによる雑音環境下における品質改善

3.1 実験条件

STRAIGHT の分析条件を表 1 に示す．

音声のひずみを表す尺度としてケプストラムひずみ（CD: Cepstrum Distortion）を用いている．CD は次式で表される．

$$CD = \frac{10}{\log_e 10}\sqrt{2\sum_{i=1}^{p}(c_i^{(x)} - c_i^{(y)})^2} \tag{4}$$

$c_i^{(x)}$，$c_i^{(y)}$ は合成した音声のケプストラム係数，p はケプストラム次数を表す．CD を求めるときの実験条件を表 2 に示す．CD は，雑音のない状態で合成した音声と，雑音を加えた状態で合成した音声との，有声区間のみでのひずみを求めている．

表 1　STRAIGHT の分析条件
Table 1　Conditions for the STRAIGHT analysis.

標本化周波数	8000 Hz
分析窓	ガウス窓
フレーム長	可変 {式 (2) において $\eta = 1.4$ に設定}
シフト長	1 ms
FFT ポイント数	2048

音声データとしては，日本人男性話者 2 名女性話者 2 名が発声した各 8 文章，計 32 文章を用いている．

3.2 最適なパラメータの決定

まず，最も効果的に雑音抑圧を行うために，分析合成に用いる様々なパラメータについて，最適な値を求める．

3.2.1 最適なガウス窓の幅の決定

グローバル SN 比が 0 dB になるように白色雑音を加えた音声に対して，LIW を利用した STRAIGHT 分析合成を行い，CD を求める．ガウス窓の幅と CD の関係について調査する．LIW は，帯域幅を変化させるパラメータ W が音声の基本周波数，ディップの深さを変化させるパラメータ $D = 1.5$ のものを用いる．

実験結果を図 5 に示す．キャプションに関しては，original が従来の STRAIGHT による結果を示し，LIW が LIW を利用した STRAIGHT による結果を示す．CD の値は各話者 8 文章，男性女性各々計 16 文章による平均値を示す．図 5 から，ガウス窓の幅を従来の 1.8 倍にしたときに，男性女性ともに CD の改善度が大きくなることがわかる．

3.2.2 最適な LIW の帯域幅の決定

次に，先ほどと同じ条件の音声に対して LIW を利用した STRAIGHT 分析合成を行い，CD を求める．LIW の帯域幅を変化させるパラメータ W と CD の関

表 2　CD 計算時の実験条件
Table 2　Conditions for calculating cepstrum distortions.

標本化周波数	8000 Hz
分析窓	ハミング窓
フレーム長	32 ms
シフト長	8 ms
FFT ポイント数	1024
ケプストラム次数	16

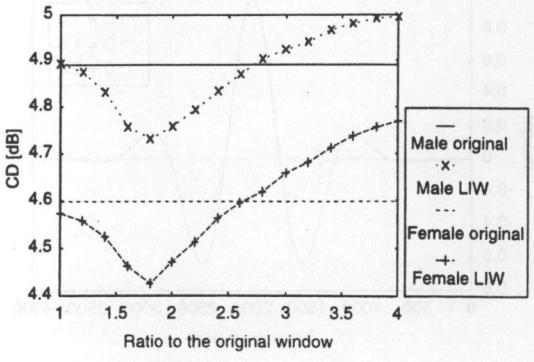


図 5　ガウス窓の幅の変化に対する CD の変化
Fig. 5　CD as a function of the band width of the Gauss window.

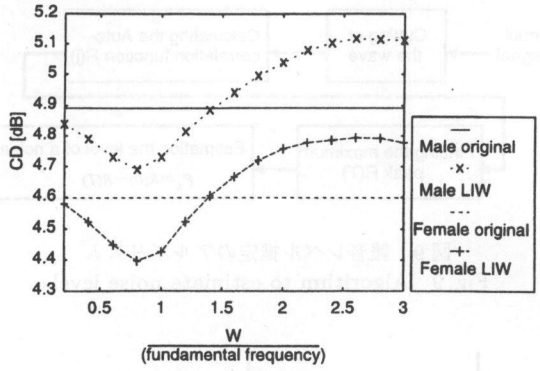


図6 LIWの帯域幅を決定するパラメータ W の変化に対するCDの変化
Fig. 6 CD as a function of the controlling parameter of the band width of LIW.

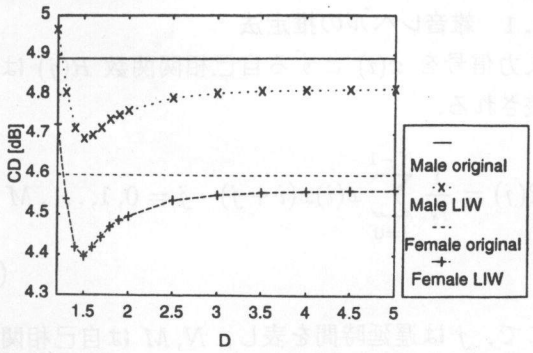


図7 LIWのディップの深さを決定するパラメータ D の変化に対するCDの変化
Fig. 7 CD as a function of the controlling parameter of the dip-depth of the inhibition.

係について調査する．ガウス窓の幅は従来の1.8倍にし，ディップの深さを変化させるパラメータ $D = 1.5$ のLIWを用いる．

実験結果を図6に示す．キャプション等は先ほどと同様である．W を基本周波数の0.8倍にしたときに，男性女性ともにCDの改善度が大きくなることがわかる．

3.2.3 最適なLIWのディップの深さの決定

次に，先ほどと同じ条件の音声に対してLIWを利用したSTRAIGHT分析合成を行い，CDを求める．LIWのディップの深さを変化させるパラメータ D とCDの関係について調査する．ガウス窓の幅は従来の1.8倍にし，LIWの帯域幅を変化させるパラメータ W は基本周波数の0.8倍にする．

実験結果を図7に示す．キャプション等は先ほどと同様である．$D = 1.5$ にしたときに，男性女性ともにCDの改善度が大きくなることがわかる．

3.3 LIWを利用したSTRAIGHTの雑音環境下における品質

グローバルSN比の異なった白色雑音を加えた音声に対して，LIWを利用したSTRAIGHT分析合成を行い，CDを求める．この結果と，従来のSTRAIGHT分析合成を行ったときの結果とを比較する．グローバルSN比を0dBから，40dBまで5dBごとに変化させて実験を行う．ガウス窓の幅は従来の幅の1.8倍にし，W は基本周波数の0.8倍，$D = 1.5$ のLIWを用いる．

実験結果を図8に示す．キャプション等は先ほどと同様である．SN比が0dBから10dB付近までに，LIWを利用することによるCDの改善が見られる．最

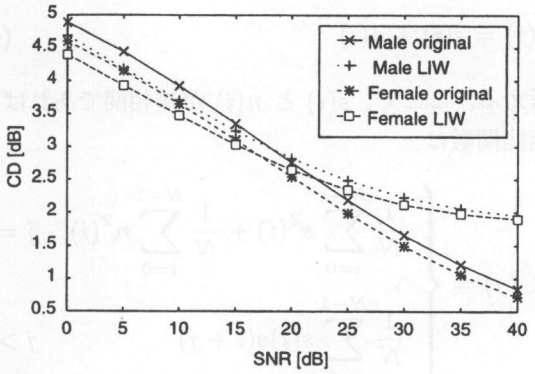


図8 白色雑音のSN比に対するCDの変化
Fig. 8 CD of the analysis-synthesis speech with white noise at various SNR conditions.

も改善が見られたところでは，男性女性ともに0.2dB程度改善される．しかし，SN比が高いときには，LIWを利用することにより，CDに劣化が見られる．この原因として，LIWをすることにより白色雑音を抑圧できるが，音韻性は崩れてしまうといったことが考えられる．

そこで，SN比が高くなるにつれて，LIWの効果を弱くし，従来のSTRAIGHTに戻していくことを考える．

4. パラメータを変化させることによる雑音環境下における品質改善

先に述べた問題点を改善するために，瞬時フレームごとにSN比を推定し，そのSN比が高くなるにつれてLIWの効果を弱め，従来のSTRAIGHTに戻していくことを考える．SN比を推定する方法として，國枝らによって提案された自己相関関数を利用して雑音レベルを逐次推定する方法[16]を用いる．

4.1 雑音レベルの推定法

入力信号を $x(i)$ とする自己相関関数 $R(j)$ は次式で表される．

$$R(j) = \frac{1}{N}\sum_{i=0}^{N-1} x(i)x(i+j) \quad j = 0, 1, \ldots, M \tag{5}$$

ここで，j は遅延時間を表し，N, M は自己相関関数の計算範囲を決定する定数である．また，音声の切出し窓は，時間長 $(N+M)$ の方形窓とする．

周期信号 $s(i)$ に雑音 $n(i)$ が重畳した入力信号として，

$$x(i) = s(i) + n(i) \tag{6}$$

を考える．ここで，$s(i)$ と $n(i)$ が無相関であれば，自己相関関数は，

$$R_x(j) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{i=0}^{N-1} s^2(i) + \dfrac{1}{N}\sum_{i=0}^{N-1} n^2(i) & j = 0 \\ \dfrac{1}{N}\displaystyle\sum_{i=0}^{N-1} s(i)s(i+j) & j > 0 \end{cases} \tag{7}$$

で近似できる．

周期信号の自己相関関数は，その周期を T として

$$R(0) = R(nT) \qquad (n \text{ は整数}) \tag{8}$$

という関係を満たすから，式 (7) より信号レベルを

$$P_S = R(T) \tag{9}$$

によって近似できる．同様にして，雑音レベルは，

$$P_N = R(0) - R(T) \tag{10}$$

として近似できる．雑音レベル推定法の原理を図 9 に，自己相関関数の様子を図 10 に示す．

本論文では，式 (9)，(10) で求められる P_S, P_N を用いて，次式から SN 比 (SNR) を推定する．

$$SNR = 10\log_{10}\left(\frac{P_S}{P_N}\right) \tag{11}$$

実際に，音声データにグローバル SN 比の異なった白色雑音を加えて，SN 比の推定精度について調査する．SN 比は前 10 フレームを含めた平均値で求める．時間長 40 ms の方形窓 $(N = 160,\ M = 160)$ を 20 ms でシフトさせる．グローバル SN 比を 0 dB から 40 dB まで 5 dB ごとに変化させて実験を行う．音声データは **3.** で用いたものと同じであり，標本化周波数は 8000 Hz である．

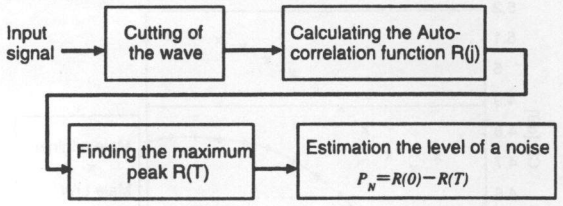


図 9 雑音レベル推定のアルゴリズム
Fig. 9 Algorithm to estimate noise level.

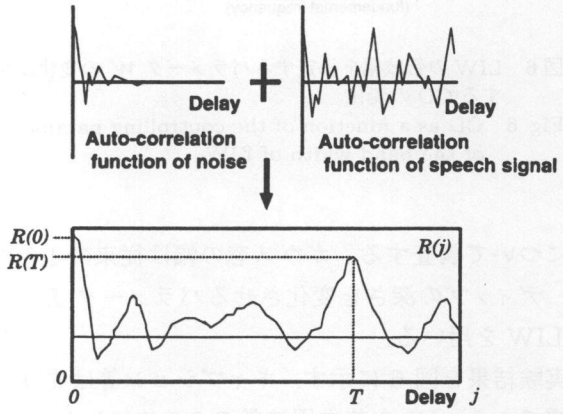


図 10 自己相関関数の様子
Fig. 10 An auto-correlation function of speech signal under noisy conditions.

各々のフレームで推定された SN 比を全フレームで平均した結果を図 11 に示す．推定された SN 比の値は各話者 8 文章，男性女性各々計 16 文章の平均値を示す．

4.2 SN 比に対する最適なパラメータの決定

実験条件は **3.** と同じである．グローバル SN 比が 10 dB，15 dB，20 dB，40 dB になるように白色雑音を加えた音声各々に対して，LIW を利用した STRAIGHT 分析合成を行い，CD を求める．各々の SN 比において，最適なガウス窓の幅，LIW の帯域幅，ディップの深さを求める．

SN 比が 10 dB のときは，ガウス窓の幅を従来の 1.8 倍，W は基本周波数の 0.8 倍，$D = 1.6$ にしたときに，男性女性ともに CD の改善度が大きくなる．15 dB のときは，ガウス窓の幅を従来の 1.8 倍，W は基本周波数の 0.8 倍，$D = 1.7$ にしたときに，男性女性とも

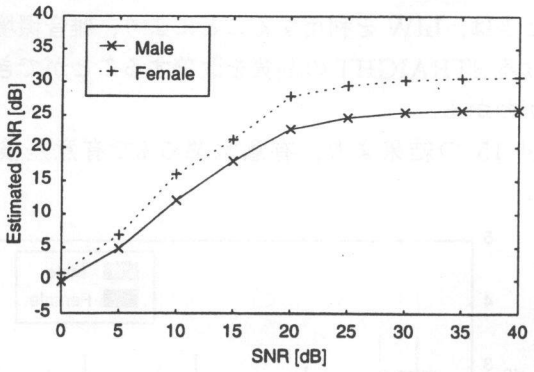


図 11 SN 比の推定精度
Fig. 11 Accuracy in estimating the SNR.

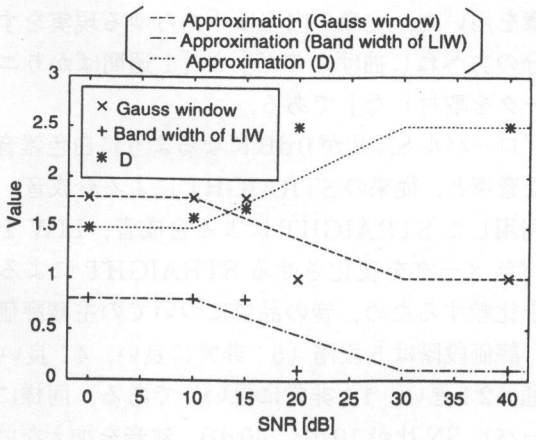


図 12 SN 比に対する各々のパラメータの変化
Fig. 12 Optimal values for the controlling parameters of LIW at various SNR conditions.

に CD の改善度が大きくなる．

SN 比が 20 dB，40 dB のときは，LIW を用いることによる CD の改善は見られない．そこで，SN 比が 20 dB 以上では LIW をパルス化するために，帯域幅は狭く，ディップの深さは浅くする．また，急激にガウス窓の幅，LIW の帯域幅，ディップの深さを決定するパラメータを変化させると，合成された音声の時間変動が大きくなってしまう．そこで，滑らかに変化する直線で SN 比に対する各々のパラメータの変化を近似する．

SN 比に対するパラメータの変化を図 12 に示す．キャプションに関しては，各々Gauss window がガウス窓の幅，Band width of LIW が LIW の帯域幅，D が LIW のディップの深さを決定するパラメータを示す．SN 比が 0 dB，10 dB，15 dB，20 dB，40 dB における最適なパラメータの値を点で示す．SN 比が 20 dB，40 dB のときは，W は基本周波数の 0.1 倍，$D = 2.5$ を最適値としている．また，パラメータの急激な変動を防ぐために，滑らかに変化するように近似した直線も図 12 に示す．以下の実験では，パラメータは図 12 で示されている直線で変化させている．

4.3 パラメータを変化させることによる雑音環境下における品質改善

グローバル SN 比の異なった白色雑音を加えた音声に対して，LIW を利用し，瞬時フレームごとに推定した SN 比に応じて各々のパラメータを変化させる STRAIGHT 分析合成を行い，CD を求める．この結果と，従来の STRAIGHT 分析合成を行ったときの結果と，LIW を利用し，グローバル SN 比が 0 dB で最適化したパラメータを用いる STRAIGHT 分析合成を行ったときの結果を比較する．グローバル SN 比を

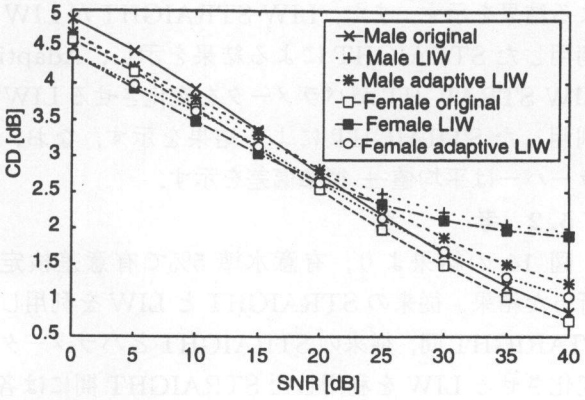


図 13 白色雑音の SN 比に対する CD の変化
Fig. 13 CD of the analysis-synthesis speech with white noise at various SNR conditions.

0 dB から，40 dB まで 5 dB ごとに変化させて実験を行う．

実験結果を図 13 に示す．キャプションに関しては，adaptive LIW が推定した SN 比に応じてパラメータを変化させる STRAIGHT による結果を示す．他のキャプションは図 8 と同様である．瞬時フレームごとに推定した SN 比に応じてパラメータを変化させることにより，グローバル SN 比が 20 dB より大きいときにおける，LIW を用いることによる CD の劣化を抑えることができる．

5. 主観評価

5.1 主観評価実験

主観評価における実験条件を表 3 に示す．実験に用いた音声データは **3.** の実験で用いた 32 文章に含まれている男性話者 2 名女性話者 2 名各々2 文章，計 8

文章を用いた．文章の内容は「あらゆる現実をすべて自分の方へねじ曲げたのだ」と「1 週間ばかりニューヨークを取材した」である．

グローバル SN 比が 0 dB になるように白色雑音を加えた音声と，従来の STRAIGHT による合成音，LIW を利用した STRAIGHT による合成音，LIW を利用しパラメータを変化させる STRAIGHT による合成音を比較するため，音の品質についての主観評価を行う．評価段階は 5 段階（5：非常に良い，4：良い，3：普通，2：悪い，1：非常に悪い）である．同様に，グローバル SN 比が 10 dB，40 dB，雑音を加えないときにおいても主観評価を行う．結果を図 14～図 17 に示す．横軸に関しては，Original Speech が原音声による結果を示し，STRAIGHT が従来の STRAIGHT による結果を示す．また，LIW STRAIGHT が LIW を利用した STRAIGHT による結果を示し，Adaptive LIW STRAIGHT がパラメータを変化させる LIW を利用した STRAIGHT による結果を示す．なお，エラーバーは平均値 ± 標準偏差を示す．

5.2 考　　察

図 14 の結果より，有意水準 5%で有意差検定を行った結果，従来の STRAIGHT と LIW を利用した STARIGHT 間，従来の STRAIGHT とパラメータを変化させる LIW を利用した STRAIGHT 間には各々有意差があることがわかった．よって，SN 比が 0 dB のときは，LIW を利用することにより，雑音環境下における STRAIGHT の品質を改善することができることがわかる．

図 15 の結果より，有意水準 5%で有意差検定を

表 3　主観評価における実験条件
Table 3　Experimental conditions for subjective tests.

評価者	成人男女 10 名
受聴条件	両耳ヘッドホン 最適受聴レベル

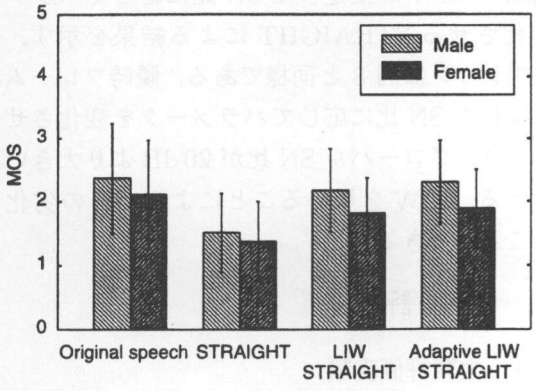


図 14　白色雑音（SN 比 0 dB）を加えたときの MOS 値
Fig. 14　Results of the subjective tests under noisy conditions (SNR 0 dB).

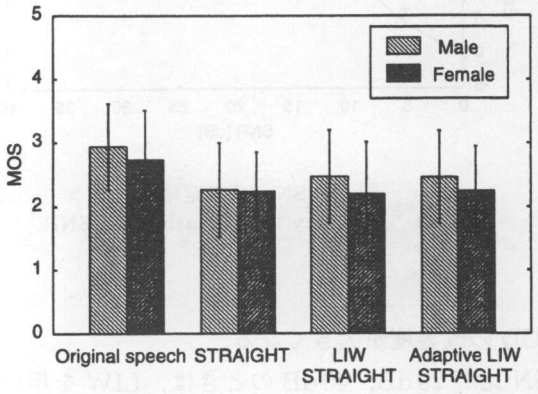


図 15　白色雑音（SN 比 10 dB）を加えたときの MOS 値
Fig. 15　Results of the subjective tests under noisy conditions (SNR 10 dB).

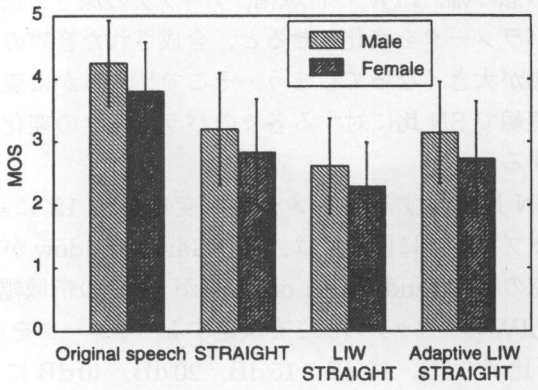


図 16　白色雑音（SN 比 40 dB）を加えたときの MOS 値
Fig. 16　Results of the subjective tests under noisy conditions (SNR 40 dB).

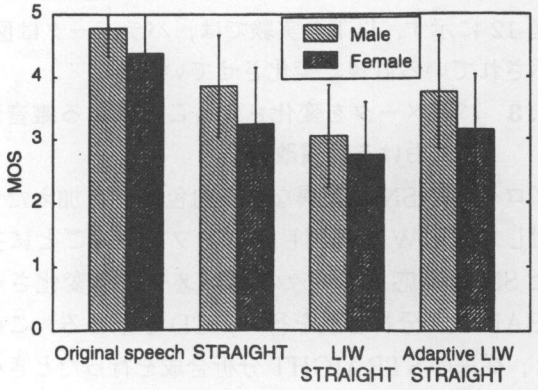


図 17　白色雑音を加えないときの MOS 値
Fig. 17　Results of the subjective test under clean conditions.

行った結果，従来のSTRAIGHT，LIWを利用したSTRAIGHT，パラメータを変化させるLIWを利用したSTRAIGHT間には有意差がないことがわかった．よって，SN比が10 dBのときは，LIWを利用しても従来のSTRAIGHTと同程度の品質であることがわかる．

図16，図17の結果より，有意水準5%で有意差検定を行った結果，従来のSTRAIGHTとLIWを利用したSTRAIGHT間，パラメータを変化させるLIWを利用したSTRAIGHTとLIWを利用したSTRAIGHT間には有意差があることがわかった．また，従来のSTAIGHTとパラメータを変化させるLIWを利用したSTRAIGHT間には有意差がないこともわかった．よって，SN比が40 dB，雑音を加えないときは，LIWを用いることにより品質が低下することがわかる．これは，音韻性が崩れてしまうからであると考えられる．しかし，パラメータを瞬時フレームごとに推定したSN比に応じて変化させてLIWの効果を弱めることにより，品質の低下を抑えることができる．その結果，従来のSTRAIGHTと同程度の品質が得られることがわかる．

6. むすび

本論文では，まず，グローバルSN比が0 dBの白色雑音を加えたときに，最も品質が改善されるように，ガウス窓の幅，Lateral Inhibitive Weighting（LIW）の帯域幅，ディップの深さといったパラメータの値を求め，その値を用いて，LIWを利用したSTRAIGHTの雑音環境下における品質についての実験を行った．その結果から，SN比が低いときにはLIWを用いることによりケプストラムひずみ（CD）を改善することができた．しかし，SN比が高くなるにつれて，CDが若干劣化することがわかった．そこで，瞬時フレームごとにSN比を推定し，それに応じて各々のパラメータを変化させることにより，SN比が高いときに見られたCDの劣化を抑えることができた．

また，品質についての主観評価を行った結果，推定したSN比に応じて各々のパラメータを変化させることにより，SN比が低いときには品質を改善させることができ，SN比が高いときには従来のSTRAIGHTと同程度の品質を得られることがわかった．

なお，本論文では，ピッチ情報は雑音のない状態で抽出したものを用いた．しかし，雑音環境下では，ピッチ抽出の精度が劣化するといった問題がある[11]．今後検討する課題としては，雑音環境下における精度の良いピッチ抽出法を研究する必要があると考えている．

謝辞 本研究の一部は，科学技術振興事業団の戦略的基礎研究推進事業CRESTの援助を受けて行われた．

文献

[1] 大室 仲，間野一則，“雑音環境下での音声符号化—実用における課題，” 音講論，2-2-3, pp.237–240, Sept. 1998.
[2] 古井貞熙，音響，音声工学，pp.122–124, 近代科学社，東京，1992.
[3] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-33, no.27, pp.113–120, 1979.
[4] H. Hamada, T. Hirahara, A. Imamura, T. Matsuoka, and R. Nakatsu, “Auditory-based filter-bank analysis as a front-end processor for speech recognition,” Proc. Eurospeech, pp.396–399, Paris, France, Sept. 1989.
[5] T. Hirahara and H. Iwamida, “Auditory spectrograms in HMM phoneme recognition,” Proc. ICSLP, pp.381–384, Kobe, Japan, Nov. 1990.
[6] 小原和明，平原達也，“DTW単語認識システムにおける聴覚フィルタフロントエンドの評価，” 音響誌，vol.50, no.6, pp.452–464, 1994.
[7] S. Kajita and F. Itakura, “Speech analysis and speech recognition using subband-autocorrelation analysis,” J. Acoust. Soc. Jpn. (E), vol.15, no.5, pp.329–338, 1994.
[8] 梶田将司，武田一哉，板倉文忠，“中心周波数の逆数の整数倍の相関係数を用いた帯域分割—自己相関分析，” 音響誌，vol.54, no.2, pp.111–118, 1998.
[9] S. Kajita, K. Takeda, and F. Itakura, “Noise robust speech recognition using subband-crosscorrelation analysis,” IEICE Trans. Inf. & Syst., vol.E81-D, no.10, pp.1079–1086, Oct. 1998.
[10] H. Kawahara, “Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited,” Proc. ICASSP, pp.1303–1306, Munich, Germany, April 1997.
[11] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” Speech Communication, vol.27, no.3–4, pp.187–207, 1999.
[12] 東山恵祐，陸 金林，中村 哲，鹿野清宏，河原英紀，“音声分析・変換・合成方法STRAIGHTの音声符号化への適用について，” 信学技報，SP98-10, April 1998.
[13] H. Dudley, “Remaking speech,” J. Acoust. Soc. Am., vol.11, no.2, pp.169–177, 1939.
[14] 河原英紀，勝瀬郁代，東山恵祐，“音声分析・変換・合成方法STRAIGHT-TEMPOにおける相補的な時間窓の利用について，” 信学技報，H-97-47, July 1997.

[15] 河原英紀，増田郁代，“音声分析･変換･合成法 STRAIGHT のスペクトル近似特性の評価と改良について，” 信学技報，SP96-97, Jan. 1997.
[16] 國枝伸行，“雑音レベルの変動を考慮したスペクトルサブトラクション法，” 音講論，2-2-6, pp.245–246, Sept. 1998.

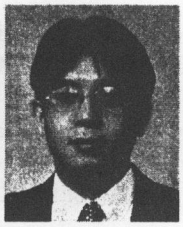
（平成 12 年 2 月 25 日受付，6 月 27 日再受付）

戸田　智基

1999 名大・工・電子情報卒．同年，奈良先端大・情報科学研究科・博士前期課程入学．音声分析合成に関する研究に従事．日本音響学会会員．

坂野　秀樹　（学生員）

1996 名大・工・電子情報卒．1998 奈良先端大・情報科学研究科・博士前期課程了．同年，名大院・工・博士後期課程入学．音声分析合成に関する研究に従事．日本音響学会会員．



梶田　将司　（正員）

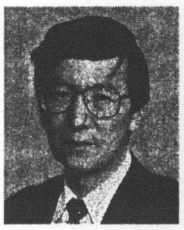
1990 名大・工・情報卒．1995 同大大学院博士課程了．同年同大・工・助手．音声認識のための音響分析の研究に従事．1998 同大・情報メディア教育センター・助手．工博．IEEE，日本音響学会，米音響学会，情報処理学会各会員．

武田　一哉　（正員）

1983 名大・工・電気卒．1985 同大大学院修士課程了．同年国際電信電話（KDD）株式会社入社，ATR 自動翻訳電話研究所，KDD 研究所において音声合成・認識システムの研究を行う．1994 名大・工・助教授．工博．IEEE，日本音響学会，情報処理学会，映像情報メディア学会各会員．

板倉　文忠　（正員）

1963 名大・工・電子卒．1968 同大大学院博士課程了．同年電電公社（現 NTT）武蔵野通研入所．音声処理の研究に従事．工博．1973～1975 ベル研究所にて音声認識・音声分析の研究を行う．1984 名大・工・教授．1998 同大・情報メディア教育センター・教授．1970，1978，1981 年度論文賞，1972，1981 年度業績賞，1996 IEEE Signal Processing Society Award 各受賞．IEEE，日本音響学会各会員．



鹿野　清宏　（正員）

1970 名大・工・電気卒．1972 同大大学院修士課程了．同年電電公社（現 NTT）武蔵野通研入所．1984～1986 カーネギーメロン大客員研究員．1986～1990ATR 自動翻訳電話研究所音声情報処理研究室長．1992NTT ヒューマンインタフェース研究所主席研究員．1994 奈良先端大・情報科学研究科・教授．工博．音声・音情報処理の研究及び研究指導に従事．1975 本会米沢賞，1991 IEEE SP 1990 Senior Award，1994 日本音響学会技術開発賞各受賞．IEEE，日本音響学会，情報処理学会各会員．