

修士論文

音源のエネルギー密度向上を目的と
した位相最適化の高速・高品質化



2012 年 2 月 8 日

指導教員 広瀬 啓吉 教授

電子情報学専攻

48-106419 砂田 宜宏

内容概況

近年の音楽編集では、特にポピュラー音楽において、音質の劣化を最小限に留めながら音を大きくするために波形を編集することに多大な労力が払われている。これは音が大きくなることにより音に迫力が出る、環境雑音や再生機器の雑音に埋もれて音が聞き取りにくくなることを防ぐといった効果が期待されてのことである。この目的を達成するため録音後の音源に対し様々な編集が行われるが、現在用いられている手法では音質の劣化は避けることができない。

音質劣化の非常に少ないエネルギー密度向上の手法の検討として、音源の位相を変更する手法が近年提案されている。この手法では、音源波形をフーリエ変換を用いて周波数成分へと分割し、その後それらの振幅には編集を加えず、位相のみを変更してエネルギー密度が最大となる位相の組み合わせを探索する。人間の聴覚特性として、音源の位相が変化しても音質の変化が全く知覚されない場合があることが知られており、位相の変化方法が適切であれば非常に小さい劣化でエネルギー密度を向上させられることが期待できる。しかしながらこの手法では、音質の劣化が発生しないような位相の変化について十分な検討がされておらず、音質の劣化が発生してしまっていた。またエネルギー密度が最大となる位相の組み合わせを見つけるために全探索を行なっているため、計算時間も非常に長く、現実的な時間ではおよそ 1 [dB] から 3 [dB] 程度しかエネルギー密度を向上させることが出来ず、実用レベルに達しているとは言い難かった。

本論文では、音源の位相を変更してエネルギー密度を向上させる手法を改良し、音質の劣化を最小限に留めながら高速に音源のエネルギー密度を向上させる手法を提案する。本論文で提案する手法では、位相の変化量を平滑化することにより音質の劣化を抑制する。また探索に最急降下法を用いることにより、高速な探索を実現する。

本論文で提案した手法を用い、幾つかの楽曲についてエネルギー密度の向上を行ったところ、およそ 5 [dB] から 7 [dB] のエネルギー密度の向上が実現された。また ITU-BS. 1116 で推奨される方法を用いて聴取実験を行い、音質の劣化が非常に小さいことを確認した。

目次

第 1 章	序論	1
1.1	研究の背景	1
1.2	研究の目的	2
第 2 章	Auto Gain Control	3
2.1	はじめに	3
2.2	Automatic Gain Control の原理	3
2.3	Automatic Gain Control の音質への影響	4
2.4	実際の運用	5
2.5	まとめ	5
第 3 章	従来 of 位相最適化手法	7
3.1	はじめに	7
3.2	音の位相	7
3.3	位相変化と音質	8
3.4	位相最適化によるエネルギー密度向上の基本原則	10
3.5	従来手法によるピーク最小となる 位相の組み合わせの探索	13
3.6	従来手法による音質劣化防止のための制約条件	13
3.7	まとめ	13
第 4 章	最急降下法と なめらかな位相変化を用いた 高速・高品質な位相最適化	15
4.1	はじめに	15
4.2	最急降下法	15
4.3	ピーク値のなめらかな関数での代用	16
4.4	ピーク値の微分	19
4.5	微分計算の高速化	22
4.6	制約条件	22
4.7	フレームの切り出し	26
4.8	まとめ	26

第 5 章	実験と考察	30
5.1	はじめに	30
5.2	制約を用いない場合の探索速度の評価	30
5.2.1	実験条件	30
5.2.2	結果と考察	30
5.3	制約条件を用いた場合の探索速度の評価	33
5.3.1	実験条件	33
5.3.2	結果と考察	34
5.4	音質の劣化の評価	34
5.4.1	条件	34
5.4.2	結果と考察	36
5.5	提案手法と Auto Gain Control の比較	37
5.5.1	音質	37
5.5.2	エネルギー密度の時刻変化	39
5.5.3	リアルタイム性	42
5.5.4	両者の使い分け	42
第 6 章	結論	43
6.1	まとめ	43
6.2	今後の展望	43
	謝辞	45
	参考文献	46
	発表文献	48
付録 A	ローパスフィルタを用いた 制約条件の適用例	49

図目次

2.1	Function of Auto Gain Control.	4
2.2	Characteristic of Auto Gain Control.	5
2.3	Original waveform.	6
2.4	Waveform that AGC is applied.	6

3.1	Fourier Transform using constant length window function.	9
3.2	Phase characteristic of all pass filter.	11
3.3	Phase characteristic of random change.	11
3.4	Sine waves added to make a large peak.	12
3.5	Sine waves added to make a moderate peak.	12
3.6	Weighted overlap and add.	14
4.1	Part of waveform.	17
4.2	Exponential of the part of waveform.	17
4.3	Change of peak value for varying phase.	18
4.4	Change of approximation of peak value for varying phase.	18
4.5	Edit windowed waveform to minimize total waveform.	21
4.6	Gain characteristic of an FIR low pass filter.	24
4.7	Phase characteristic of an FIR low pass filter.	24
4.8	Impulse response of an FIR low pass filter.	24
4.9	Change of variable vector θ_{delta} and smoothed phase change ϕ_{delta}	25
4.10	Windowed waveform.	27
4.11	Optimized windowed waveform without the constraint.	27
4.12	Optimized windowed waveform with the constraint.	27
4.13	How the process work if large peak exist on the end of the frame.	28
4.14	Procedure of effective windowing.	29
5.1	A part of the test signal.	31
5.2	RMS gain as a function of computational time by proposed and conventional method.	31
5.3	A part of the test signal optimized by proposed method.	32
5.4	A part of the test signal optimized by conventional method.	32
5.5	RMS change as a function of computational time for varying smoothness of phase change.	33
5.6	Scores of each clips.	38
5.7	Original waveform.	38
5.8	Phase-optimized waveform.	39
5.9	Local RMS of original, phase optimized and AGCed samples.	40
5.10	Local RMS of original sample.	40
5.11	Local RMS of sample optimized by proposed method.	41
5.12	Local RMS of sample compressed by AGC.	41
A.1	Steepest decent using modified axis.	50

表目次

5.1	Musics used for experiment.	34
5.2	Parameter setting of AGC.	35
5.3	Evaluation scale of BS.1116.	36

第1章

序論

1.1 研究の背景

1877年のレコードの発明により、音楽の楽しみ方として、生演奏を聞くだけでなく電氣的に記録された演奏を聞くという楽しみ方が生まれた。それ以来データとしての音楽は発展・普及し、今では生演奏を聞く機会よりも、記録された演奏を聞く機会のほうがはるかに多いであろう。

またその後の電気回路の発達は音楽編集のあり方も変容させた。レコードが開発された当初は、音楽編集といえば録音した音楽に簡単なフィルタをかけた程度であったが、その後音楽に対し編集を加える様々な手法が開発され、今では編集者の仕事は演奏された音楽をそのままの形で記録することではなく、録音された音を加工し楽曲として最適な形へと編集することとなっている。更に近年ではコンピュータの発達により、従来は全てアナログ回路で行われていた音楽編集にもデジタル編集が用いられるようになり、より柔軟な音楽編集が可能となっている。

それらの編集手段としてアナログ編集を行っていた時代から用いられているものの一つに Auto Gain Control (AGC) がある。詳しい原理については後述するが、簡単に言えば AGC は増幅率を各時刻毎に変化させ、音割れを発生させずに音源のエネルギー密度を向上させるための手法である。音源のエネルギー密度が向上することにより、音に迫力が出る、楽音が環境音や再生機器の雑音に埋もれて聞き取りづらくなることを防ぐといった効果があり、現在ではことポピュラー音楽においてはすべての楽曲で AGC が用いられていると言っても過言ではないだろう。

AGC は広くそして古くから用いられている手法であるが、幾つか問題点も存在する。AGC はその原理上音源に非線形な編集を加えるため、適用することにより音質が変化する。エネルギー密度向上の度合いが少ない場合であれば、パラメータ調整が多少いい加減でも発生する音質の変化はそれほど耳障りなものではないものの、楽曲によってはその音質変化を嫌い、AGC がかけられない場合もある。またエネルギー密度向上の度合いを高くしようとするとパラメータ調整は困難になり、適切に設定されていなければ非常に耳障りな劣化が発生する。近年のポピュラー音楽はエネルギー密度を非常に高く編集する傾向があり、耳障りな劣化を抑えながら十分なエネルギー密度を達成するには楽器毎にパラメータを調整しながら個別に AGC をかける、異なるパラメータの AGC を複数回かけるなど、高度な技術が要求される。また時々刻々と増幅率が変化するため、奏者の意図した強弱の変化が反映されなかったり、奏者の意図しない音の強弱の変化が発生したりする。このよ

うな理由から、AGCに代わる、もしくはAGCと併用できる、より効果的なエネルギー密度向上の手段が望まれていた。

そのような手段の一つとして、各周波数成分の位相を最適化する手法が藍原らにより提案されている [1]。人間の聴覚は位相の変化に対して鈍感であり、位相を極端に変化させない限り元の音との聞き分けは非常に困難である。したがって、この手法は原理的には音質の劣化を最小限に抑えながらエネルギー密度を向上させられることを期待できる。しかしながら、[1]の手法は本来補聴器への応用を第一の目的として開発されたものであり、音質の劣化を引き起こすような極端な位相変化を制限するための制約が洗練されておらず、適用すると音質劣化が生じていた。また、位相の最適化の方法も非常に効率の悪いものであり、現実的な計算時間ではエネルギー密度を僅かしか向上させることができなかった。

1.2 研究の目的

本論文では、[1]の手法を改良し、大幅なエネルギー密度の向上を高速に実現する手法と、音質の劣化を抑えるための制約についての提案を行う。

これらの改良により、[1]の手法を音楽用途にも十分適用可能なものとする。本論文で提案する手法を単独でもしくはAGCと組み合わせて用いることにより、よりよい音質で大きなエネルギー密度を達成することがより容易になると考えられる。

第2章

Auto Gain Control

2.1 はじめに

この章では、音源のエネルギー密度を向上させるために古くから用いられている手法である、Automatic Gain Control(AGC)の動作について説明を行う。

AGCは、前後の波形のエネルギー等に依存して増幅率を変動させることにより、種々の効果を与える編集である。音楽編集に用いられる際、AGCはその利用方法によってコンプレッサー、リミッター、ノイズゲート等様々な名前と呼ばれるが、これらのうちエネルギー密度を向上させる用途に用いられるものは、コンプレッサーとリミッターである。またリミッターはコンプレッサーのパラメータを特定の値で固定したものであるため、以下ではコンプレッサーの動作について説明を行う。

2.2 Automatic Gain Controlの原理

コンプレッサーは、入力波形のエネルギーが一定以上となったとき、入力のエネルギーの大きさに応じて増幅率を減少させ、波形を圧縮する [2]。この処理により、エネルギーの小さい箇所は波形に変化が無いが、エネルギーの大きな箇所では波形が小さくなり、ピークが抑えられる。結果として、データの量子化ビット数で表現可能な値を超えるような過大出力を発生させることなく全体の増幅率を上げることができ、全体としてはエネルギー密度が向上する (図 2.1)。図 2.1 の縦軸の Local RMS(Root Mean Square) は、波形に対し横軸の値を開始時刻とした長さ約 46 [ms] のハン窓をかけ、Root Mean Square を取ったものであり、各時刻での音の大きさに対応する。コンプレッサーの入力と出力のエネルギーの関係の例を図 2.2 に示す。圧縮が始まる入力エネルギーをスレッシュホールドと呼び、また入力がスレッシュホールド以上のときの、入力と出力のエネルギーの比をレシオと呼ぶ。入力エネルギーと出力エネルギーの関係はスレッシュホールドを境界とした 2 つの直線で表されるが、この境界をまたぐ時に突然増幅率が変化しはじめるため、その箇所が不自然に感じられる場合がある。そこでこの 2 つの直線を滑らかに繋ぐ機能が存在するが、これをどの程度滑らかに繋ぐかを設定するパラメータがニーである。入力のエネルギーがスレッシュホールドを超え圧縮がかかった後に、入力のエネルギーがスレッシュホールドを下回ると圧縮は解除され増幅率は元に戻る。この圧縮の解除が非常に短い時間で行われると非常に不自然で耳障りな音となる。したがって、圧縮の解除はある程度の時間をかけて行われる。この時間を設定するパラメータがリリースで、圧縮の解除までにかかる時間を示す。入力がスレッシュ

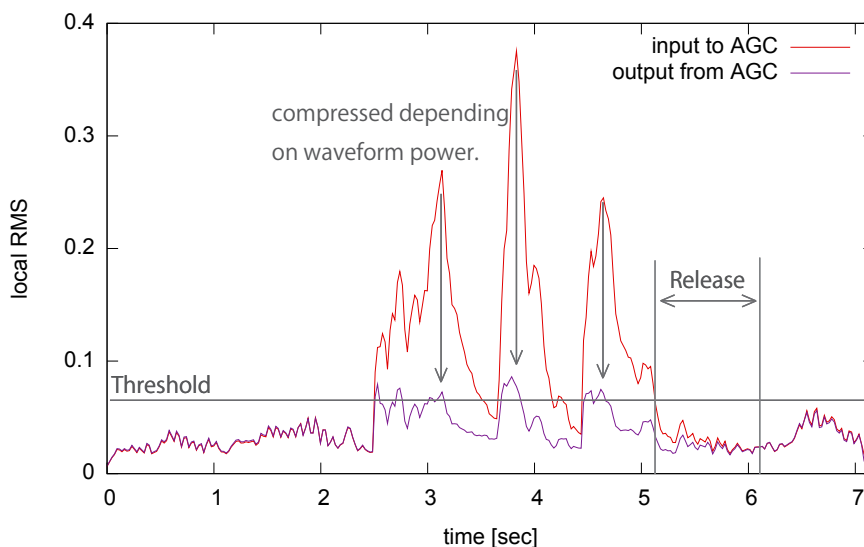


図 2.1: Function of Auto Gain Control.

しおールドを超えてから実際に圧縮がかかるまでの時間を調整するパラメータがアタックである。

ある音源の波形を図 2.3 に、その音源に AGC を適用した例の波形を図 2.4 に示す。前半の部分では両者の波形は殆ど変わらないが、およそ 220 [ms] 付近の波形の振幅が大きくなった箇所から AGC によって増幅率が下げられ、ピークが小さくなっている。波形のピークが小さくなった分増幅率を上げることができるため、波形全体としてはエネルギー密度を高くすることができる。

2.3 Automatic Gain Control の音質への影響

AGC では波形に対し非線形事変な編集を加えるため、編集の加えられた波形のスペクトルは複雑な変化をするが、結果として AGC が動作している箇所の音色は「こもった音」として認識される場合が多いようである。この音色の変化は多くの場合望ましいものではないが、楽器やジャンル、楽曲によってはこの音色の変化が好まれ、積極的に音質の変化が発生するように運用される場合もある。

エネルギー密度を向上させるための用途では、アタックは設定できる中で最短の値、レシオは 8 : 1 程度の値として、所望のエネルギー密度が得られるまでスレしおールドを下げる。このとき、スレしおールドの値をある一定まで下げると、音源全体に圧縮がかかった状態となり、この状態で更にスレしおールドを下げたとしてもそれ以上のエネルギー密度の向上は見込めない。これ以上のエネルギー密度を達成したい場合は、リリースを短くすることによりゲインの変化する速度を上げることになるが、急激なゲインの変化は波形を著しく歪め、音割れと類似した音となる。この音質の変化は前述のものとは異なり、意図的に発生させられることは殆ど無い。

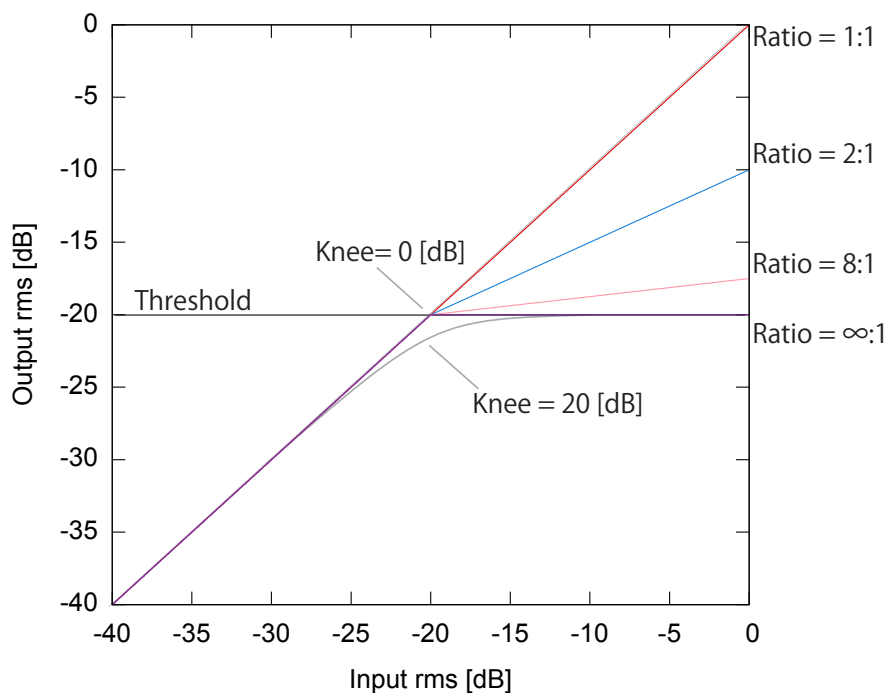


図 2.2: Characteristic of Auto Gain Control.

2.4 実際の運用

AGC を用いることが一般的であるポピュラー音楽において、AGC は編集の各所で目的毎に複数回に分けて利用されることが多い。AGC がよく用いられる箇所と目的として、

- 録音時の過大入力防止
- 楽器毎の編集において音質を積極的に変えるため
- 楽器毎の編集においてエネルギー密度向上のため
- 楽曲の仕上げにおいてエネルギー密度向上のため

がある。本論文で提案する位相最適化の手法は、楽曲の仕上げの際のエネルギー密度の向上のために用いられる AGC を代替もしくは補助することを目的としている。

2.5 まとめ

エネルギー密度を向上させるための手法として一般的に用いられている AGC の原理と、音質への影響、運用方法について説明を行った。AGC はエネルギー密度向上のために非常に効果的であるが、音質が変化することや、一定以上のエネルギー密度の向上を見込めないといった問題点が存在する。

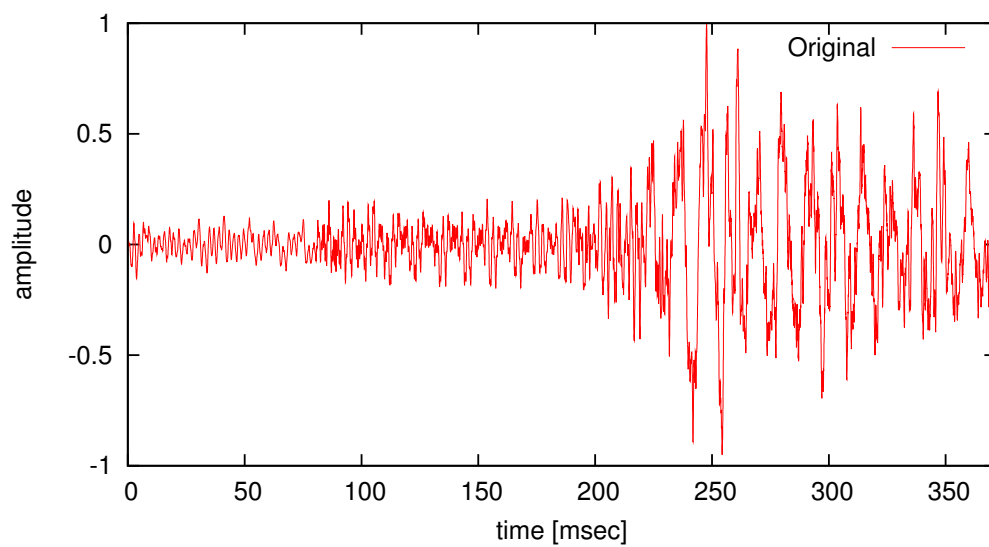


図 2.3: Original waveform.

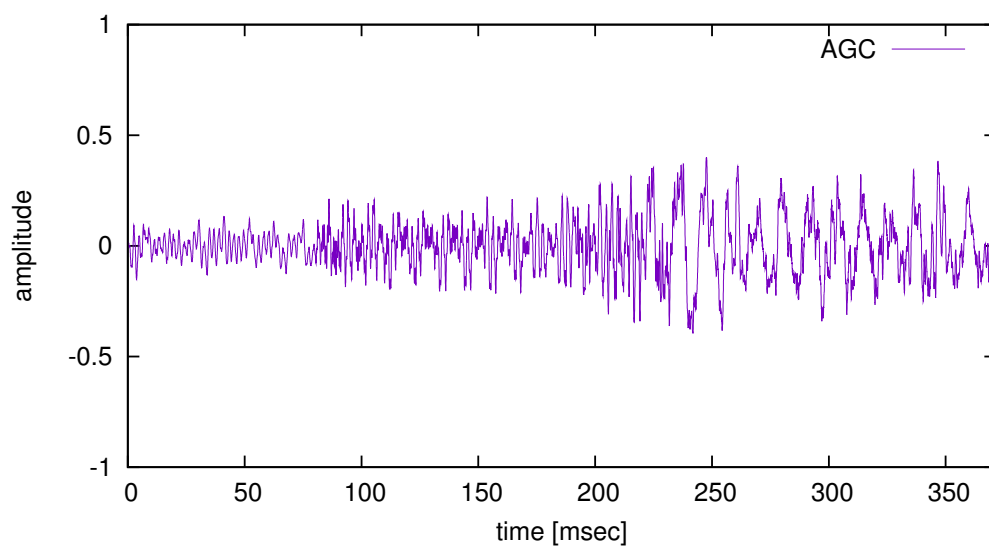


図 2.4: Waveform that AGC is applied.

第3章

従来の位相最適化手法

3.1 はじめに

本章では、葦原らによって提案された、音源のエネルギー密度向上手法について解説する [1]。[1] の手法では、音源波形を短時間フーリエ変換を用いて時刻ごとの振幅と位相を用いた表現へと変換した後、音源の各周波数成分の振幅は変更せず、位相のみを変更してピークを最小化することにより、音源のエネルギー密度の向上を行う。

過去に行われた振幅や位相を改変した音源の聴取実験から、位相の変化が音質の知覚に与える影響は非常に小さく、音源の位相が変化した場合でも音質の変化が知覚できない場合があることが古くから知られている [3]。その一方で、近年の位相と音質知覚に対するより詳細な研究において位相のみを変更した場合であっても音質が劣化する場合があることが示され [4][5]、またフーリエ変換による分析の際に用いられる窓関数の種類や長さといった条件によっては、位相も振幅に劣らない情報を持つことが明らかになっている [7][8]。

したがって、位相を適切に変更した場合であれば非常に小さな音質劣化でエネルギー密度の向上を実現できることが期待できるが、位相の変更方法が適切でなければ音質の劣化が発生することとなる。

3.2 音の位相

時間波形を振幅と位相を用いた表現へと変換する方法は様々なものが提案されているが [9][10]、[1] の手法は計算機上で音の波形を扱う際に最も一般的に用いられている、固定長の窓関数を利用した短時間離散フーリエ変換を用いられている。

分析対象の時間波形をある長さの窓関数を用いて切り出し、短時間離散フーリエ変換することによりその時刻での振幅スペクトルと位相スペクトルを得る。このとき、窓関数によって切り出された一区間をフレームと呼ぶ。あるフレームでの、それぞれの周波数の振幅 A と初期位相 ϕ は以下の式によって求められる。(離散フーリエ変換) フレームの波形を x_t 、フレームの長さを N とすると、

$$A_k \exp(\phi_k) = \sum_{t=0}^{N-1} x_t \exp\left\{-j \frac{2\pi}{N} kt\right\} \quad (3.1)$$

また、以下の式を用いることにより、振幅と位相から波形を得ることができる (離散逆

フーリエ変換)

$$x_t = \frac{1}{N} \sum_{k=0}^{N-1} A_k \exp \left\{ j \left(\frac{2\pi}{N} kt + \phi_k \right) \right\} \quad (3.2)$$

フレームの開始時刻での各周波数成分の位相は初期位相と呼ばれる。各周波数成分は定常であるため、初期位相さえ解ればそのフレーム内の任意の時刻での位相が一意に決定する。よって位相の議論に必要なのは初期位相のみであり、初期位相は単に位相と呼ばれることも多い。本論文でも、特別の断りが無い限り位相と言った場合は初期位相を意味する。

波形の特徴の時間変化を観察するためには波形を切り出す窓の中心の位置をずらしながら分析を行う。その際の窓の中心の間隔をシフト長、もしくはフレームシフトと呼ぶ。窓長を長くすると周波数分解能が向上する代わりに時間分解能が低下し、窓長を短くするとその逆である。窓長によって分析結果が変わってくるため、用途によって適切な窓長を選ぶことが重要である。音の波形は 20 [ms] ~ 40 [ms] 程度の区間であれば定常的であるとみなすことができるため、通常の用途の場合は窓長はその程度の長さに設定されることが多い [11]。シフト長は窓長の半分の長さに設定されることが多い。窓関数は窓の中心近くで大きな値となり、逆にはじめと終わりでは小さな値をとる関数が用いられる。短時間離散フーリエ変換では窓関数をかけた区間が何度も繰り返されているとして計算されるため、窓の境界の部分で波形の不連続が発生することを防ぐためである。具体的な窓関数の種類としてはハン窓やハミング窓、ブラックマン窓がよく用いられる。ハン窓を用い、窓長 32 [ms]、フレームシフトを 16 [ms] とした場合の分析の模式図を図 3.1 に示す。

3.3 位相変化と音質

人間の聴覚は音の位相変化に鈍感であり、音の位相を変化させても音質の変化が知覚できない場合がある事が知られている [3]。その一方で、位相の変化のさせ方によっては音質が大きく変化する場合もあることが知られている。どのような場合であれば音質が変化し、どのような場合では変化しないかについては様々な研究がなされているが [12][13][14][15]、現状では殆どの研究は定常音に関するものであり、本研究で主な対象となる非定常音については殆ど解明されていない。

非定常音を対象とし、位相を変化させても音質が変化しないという特性を活かした先行研究として、音の位相を変化させ知覚困難型の電子透かしを埋め込む事に利用した研究がある [16][17]。その研究では、特性の異なる 2 種類の Infinite Impulse Response(IIR) オールパスフィルタを切り替えて用いる事により音に対し情報を埋め込んでいる。用いられている IIR オールパスフィルタは以下の伝達関数を持つものであり、これは位相を変化させても音質が殆ど変化しない代表例として知られている。

$$H(j\omega) = -\frac{j\omega - a}{j\omega + a} \quad (3.3)$$

a はフィルタの特性を決めるパラメータであり、 $\omega = a$ となる周波数で $H(j\omega) = -j$ となる。計算機上に実装する場合は双一次変換などを用い、 z 領域で実装する。 $a = 1$ [kHz] と

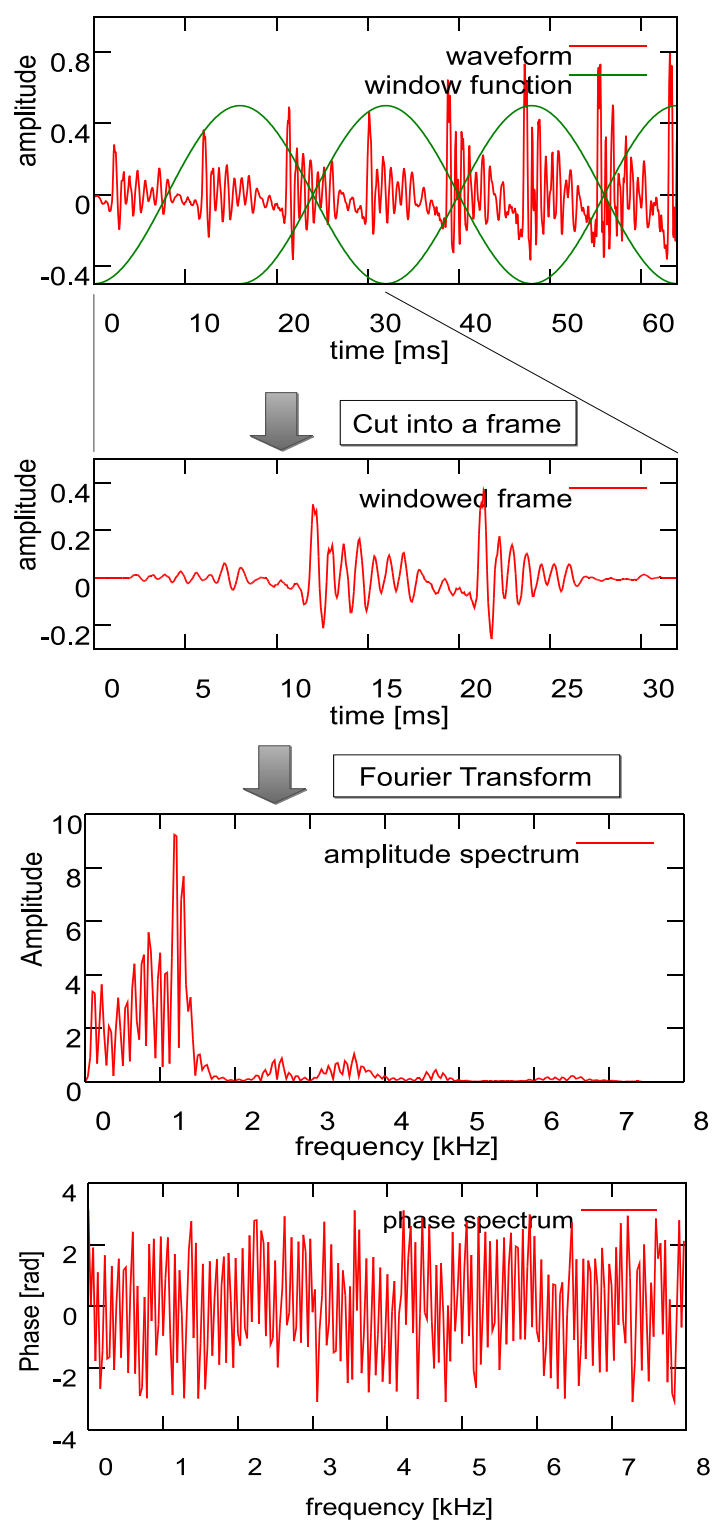


図 3.1: Fourier Transform using constant length window function.

したときの式 3.3 で表されるオールパスフィルタの位相特性を図 3.2 に示す。振幅特性は全ての周波数において 1 である。

また逆に、音質を変化させるオールパスフィルタの位相特性の一例を図 3.3 に示す。これは波形の位相に対し、平均 0 [rad]、標準偏差 0.2 [rad] のランダムな変化を与え、また、振幅特性は全ての周波数において 1 となる特性であり、Finite Impulse Response(FIR) フィルタを用いて実現されている。このフィルタでは、式 3.3 のフィルタを用いた場合と比較して位相の変化量は小さいが、このフィルタで位相を変化させた音源は残響感が発生しており、元音源と比べて明確に音質が変化している。

3.4 位相最適化によるエネルギー密度向上の基本原則

各周波数成分の山と山もしくは谷と谷が重なると、波形に大きなピークが発生する。そこで、各周波数成分の位相をずらし、山と山、谷と谷が重ならないようにすれば、ピークを小さくすることができる。このとき、波形のピークは小さくなっているが、各周波数成分の振幅は変化しておらず、また周波数の異なる正弦波は互いに直交しているためエネルギー密度は小さくならない。よってピークが小さくなった分だけ波形全体を増幅してやることにより、エネルギー密度を向上させることが可能となる。図 3.4、3.5 に例を示す。どちらも振幅 1 の 2 つの正弦波から成り立つ波形であるが、図 3.4 の波形ではピーク値は 2.0 であるのに対し、図 3.5 の波形ではおよそ 1.76 になっている。以上の原理を利用し、音源の位相を最適化しピークを最小化すると、ピークが小さくなった分歪ませることなく増幅率を上げることができ、エネルギー密度を向上させることができる。一般の音源に対して以上の原理を適用するためには、2 つの問題を解決する必要がある。

まず一つはどのようにしてピークが最小となる位相の組み合わせを探索するかということである。ピークの位置と値は各周波数の振幅と位相を変数とした複雑な関数となっており、ピーク値が最小となる位相特性を解析的に得ることは困難である。またフーリエ変換のフレーム長を N とすると、ピーク最小となる位相の組み合わせを探索することは $N/2$ 個の周波数成分の位相の組み合わせ問題となり、全探索を行うことは実質的に不可能である。したがって、何らかの方法を用いて現実的な時間で準最適解を得ることが必要となる。

また、前述したとおり位相の変化のさせ方によっては音質の劣化が発生する場合がある。したがって探索の際に何らかの制約条件をかけ、音質を損なわない範囲でのピーク最小化を行う必要がある。このとき、非定常音についてどのような位相変化が音質の劣化を引き起こし、またどのような位相変化であれば音質が劣化しないのかについては詳細な条件は解明されていない。したがって制約条件として利用しやすく、かつ音質の劣化を抑えることのできる適当な条件を見つける必要がある。

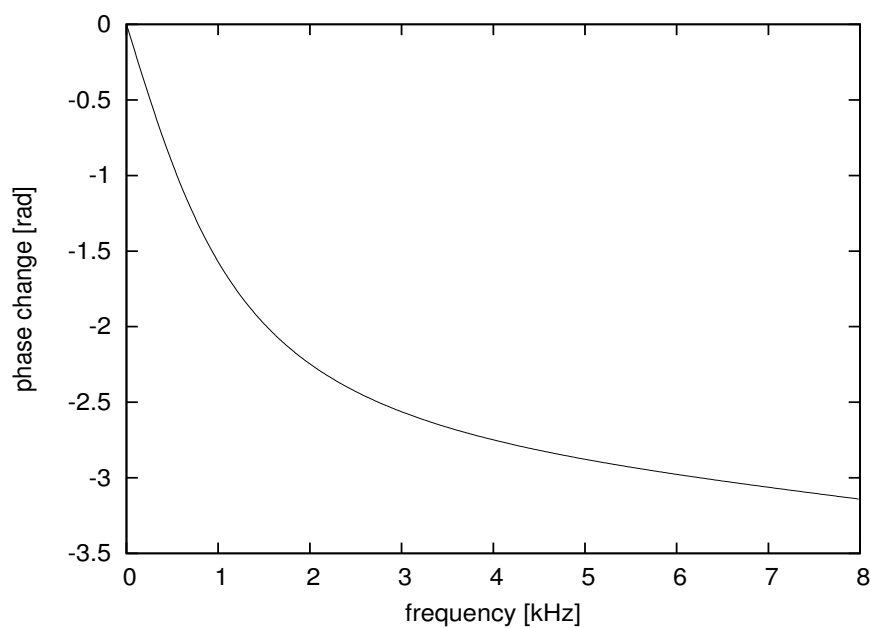


図 3.2: Phase characteristic of all pass filter.

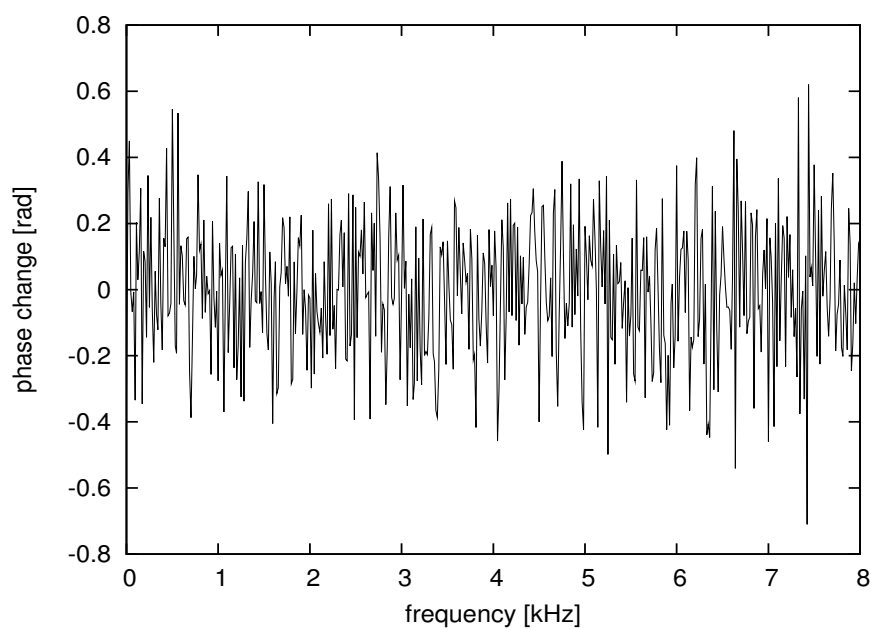


図 3.3: Phase characteristic of random change.

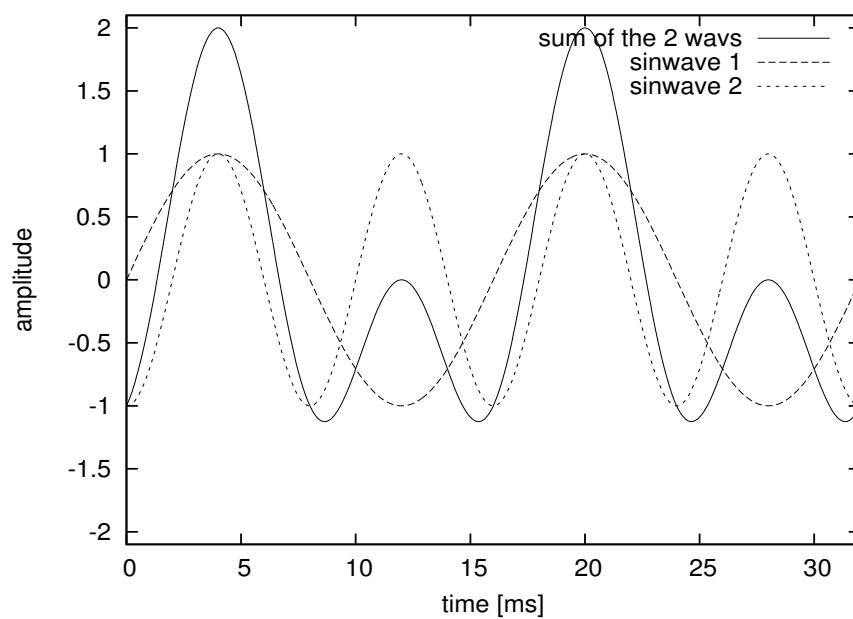


図 3.4: Sine waves added to make a large peak.

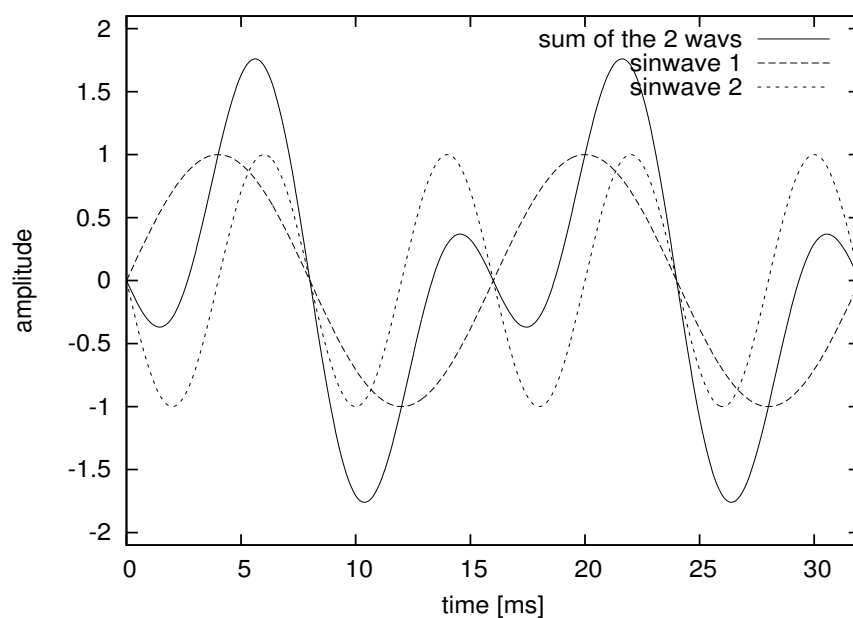


図 3.5: Sine waves added to make a moderate peak.

3.5 従来手法によるピーク最小となる 位相の組み合わせの探索

[1] による従来手法では、各周波数成分の位相を一つずつ決定していくことにより、位相の最適化を行なっている。このとき、高速化のために計算量を削減するためにピークの大きさに対する寄与の大きい、振幅の大きい周波数成分から順に位相を決定していき、振幅の小さい周波数成分については位相を変更しない。位相の変化量については、 $-\pi/2$ から $\pi/2$ まで、 $10 [^\circ]$ 刻みで 18 回変化させる。以上の手順により、得られる結果は最適解ではないものの全探索を行うよりはるかに高速に、ある程度のエネルギー密度向上が実現される。エネルギー密度の向上率は、対象となる音源に依存するが、およそ $1 [dB]$ から $3 [dB]$ 程度である。

この手法では、全ての周波数成分に対して全探索を行うよりはるかに現実的ではあるものの、本質的には全探索と大きく変わらず、現実的な時間で得られる結果は最適解からはるかに遠く、大きなエネルギー密度の向上を実現することはできない。

3.6 従来手法による音質劣化防止のための制約条件

短時間離散フーリエ変換による分析を行う際、フレーム境界での不連続を防止するために、窓関数のはじめと終わりで小さな値を持つような窓関数を用いて波形を切り出すことを前に述べた。従来手法ではこの切りだされたフレームの位相を変更するが、位相変更後の波形ではフレームの初めと終わりで小さな値を持つという条件は保存されていない。そのため、位相変更後の波形をそのまま再合成すると、フレームの境界で波形の不連続が生じ、音質劣化の原因となる。従来手法ではこの不連続が音質劣化の大きな原因であると考え、図 3.6 のように、位相変更後の波形に再び窓関数をかけた後に前後のフレームと足しあわせていた。しかし、この方法では位相改変後に再び窓関数をかけるため、各周波数の位相だけでなく振幅も変化してしまう。振幅が変化すると音質の劣化に繋がるだけでなく、「ピークが小さくなくてもエネルギー密度は変化しない」という条件が崩れるためエネルギー密度の向上にも悪影響を及ぼす。また、§3.3 で述べたとおり、FIR フィルタを用い、音源をフレームに切り分けることなく位相を変化させた場合であっても音質の劣化は生じるが、位相変化後に窓関数をかけてもそのような音質の劣化は取り除くことはできない。

3.7 まとめ

この章では、音の位相が音質に与える影響を説明し、またそのことを利用してエネルギー密度を向上させる手法の基本原則を説明した。また、位相最適化によるエネルギー密度向上のための問題点と、それらの問題を解決するための従来手法を説明した。従来手法では $1 \sim 3 [dB]$ 程度のエネルギー密度の向上を期待できるが、これは最適解からは程遠く、また音質の劣化発生する。

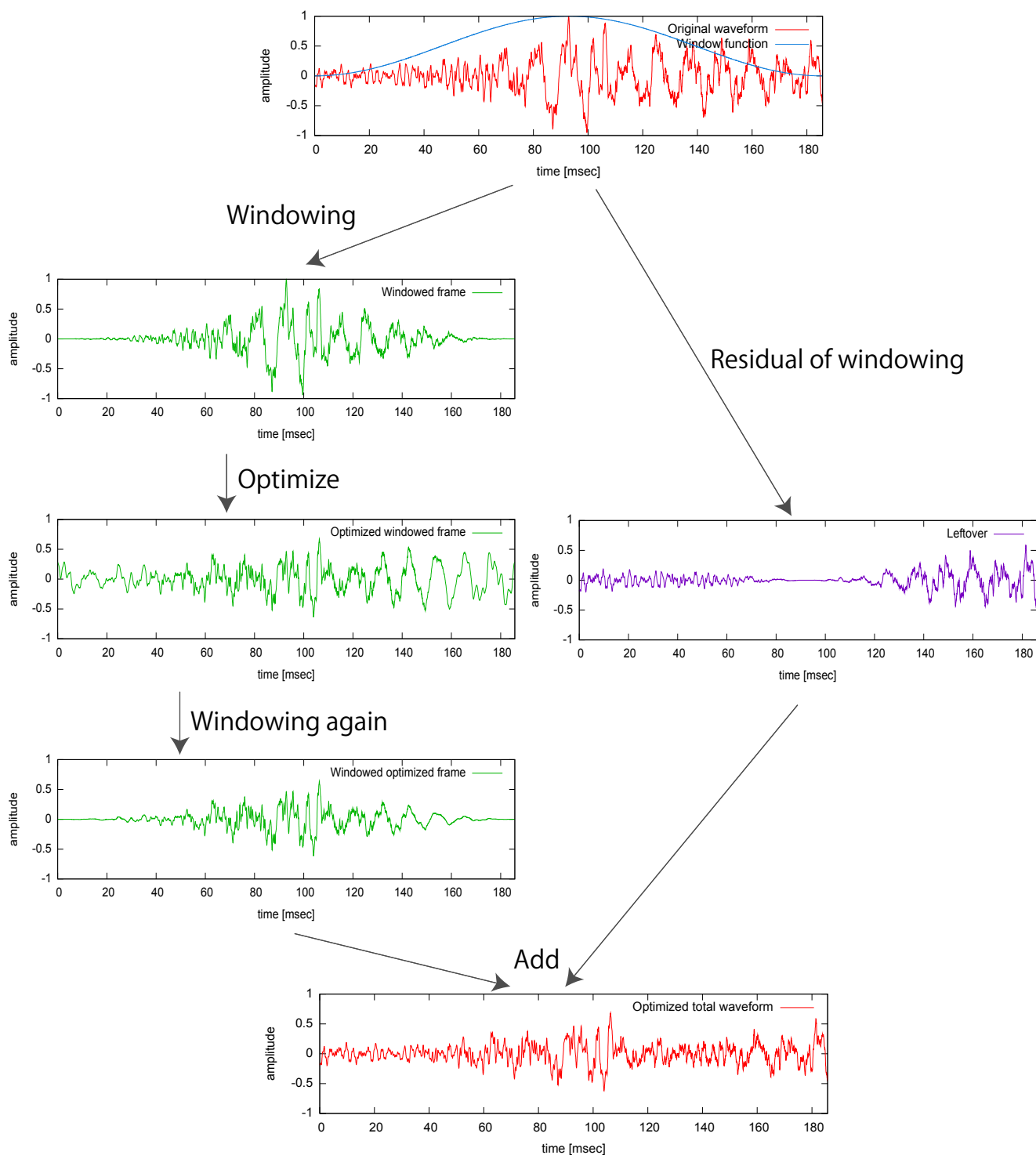


図 3.6: Weighted overlap and add.

第4章

最急降下法と なめらかな位相変化を用いた 高速・高品質な位相最適化

4.1 はじめに

本研究では、波形のエネルギー密度を向上させるために、波形のピークが最小となる位相の組み合わせの探索を行う。ピーク値は波形の各周波数の振幅と位相を変数とした複雑な関数となっており、ピーク値が最小となる位相特性を解析的に得る事は困難である。したがって、そのような場合における常套手段として最急降下法による探索を行い、準最適解を得る。

また、位相変化を平滑化することによって位相変化に伴う音質の劣化を低減する手法の提案を行う。

4.2 最急降下法

最急降下法は勾配法の一つであり、関数の微分値を用いて関数の最小値を探索するアルゴリズムである。局所最適に陥りやすいという欠点が存在するものの、動作が単純で高速であることから、数値解析ではよく用いられる。

一般に、 $f(\mathbf{a})$ で与えられる値を変数ベクトル \mathbf{a} を更新することにより最小化する場合、最急降下法による \mathbf{a} の更新式は以下ようになる。 k 回の更新後に得られた変数ベクトルを $\mathbf{a}^{(k)}$ とすると、

$$\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} + \mathbf{a}_{\text{delta}}^{(k)} \quad (4.1)$$

である。ここで、 $\mathbf{a}_{\text{delta}}^{(k)}$ は一度の更新における \mathbf{a} の変化量であり、

$$\mathbf{a}_{\text{delta}}^{(k)} = -\alpha \begin{pmatrix} \partial f(\mathbf{a}^{(k)})/\partial a_0 \\ \partial f(\mathbf{a}^{(k)})/\partial a_1 \\ \vdots \\ \partial f(\mathbf{a}^{(k)})/\partial a_{N-1} \end{pmatrix} \quad (4.2)$$

である。ここで、 α は更新の大きさを決めるパラメータである。

この式を用いて変数ベクトルの値の更新を繰り返すことにより、 $f(\mathbf{a})$ を最小化することができる。

4.3 ピーク値のなめらかな関数での代用

最急降下法を用いるためには、最小化したい値を変数の各要素で偏微分した値は予め与えられている必要がある。本研究では各周波数成分の位相を変数とし、ピーク値を最小化する。したがってピーク値の各周波数成分の位相での偏微分を求める必要があるが、ピーク値は位相の変化に対し傾きが不連続な関数となるため微分不可能であり、そのまま最急降下法で用いることはできない。したがって、最急降下法を用いるためにはピーク値を位相の変化に対し滑らかな関数で代用する必要がある。

ある区間内のピーク値 x_{peak} は以下の式で得られる。

$$x_{peak} = \max(|x_t|) \quad (t = 0, \dots, N-1) \quad (4.3)$$

x_t は長さ N の区間の波形、 $\max()$ は最大値を与える関数である。また区間の始まりの時刻を $t = 0$ としている。

x_{peak} は指数関数を用いることにより、なめらかな関数で代用することができる。指数関数を用いた x_{peak} の代用値 m は、以下の式で表される。

$$m = \sum_t^{N-1} \{\exp(x_t) + \exp(-x_t)\} \quad (4.4)$$

第一項は正のピーク、第二項は負のピークを検出するための項である。指数関数を用いることにより、ピークの箇所の値は非常に大きくなる一方で、それ以外の箇所の値はピークの箇所の値と比べて比較的小さな値にしかならない。それらの値を全て足したものの値への寄与はピークの値が支配的であり、以下の近似が成り立つと考えられる。

$$m \approx \exp(\max |x_t|) \quad [t = 0, 1, \dots, N-1] \quad (4.5)$$

ある音源波形を図 4.1 に、そしてその波形を x_t としたときの $\exp(x_t) + \exp(-x_t)$ の値を図 4.2 に示す。ピークの位置での値と比較すると、他の箇所の値は殆ど 0 に等しい。

m と x_{peak} の対応関係の例として、ある波形のある周波数の位相変化させた時の m と、 x_{peak} の変化をそれぞれ図 4.3、4.4 に示す。見易さのため、 m の変化は縦軸を対数スケールとしている。図 4.3、4.4 を見ると、 x_{peak} の大小と m の大小には綺麗な対応関係が出来ていることがわかる。また、 x_{peak} では傾きが不連続となっていた部分でも、 m の値は滑らかに変化している。したがって、 m を最小化すれば、 x_{peak} も近似的に最小化される。

尚、実装上の問題として、元の波形の振幅が大きいと、 m の値は非常に大きくなるため 64 [bit] 浮動小数点を用いて計算を行なっても値のオーバーフローが発生する場合がある。一方元の波形の振幅が小さすぎるとピークの箇所の値が十分に大きくなり、 m と x_{peak} の乖離が大きくなる。本研究では、それらの問題を防ぐため予め波形のピークの値が 100 となるよう波形全体をノーマライズした後に位相の最適化を行い、その後に再び元の振幅に戻した。

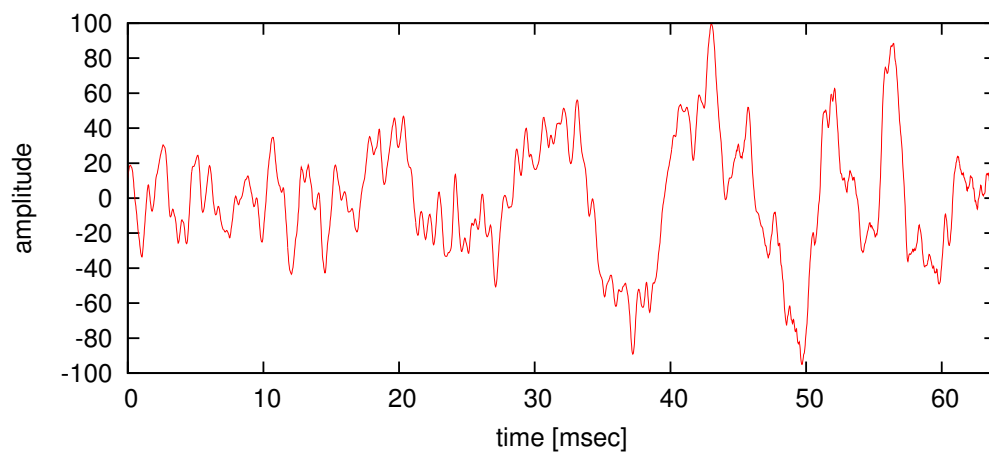


図 4.1: Part of waveform.

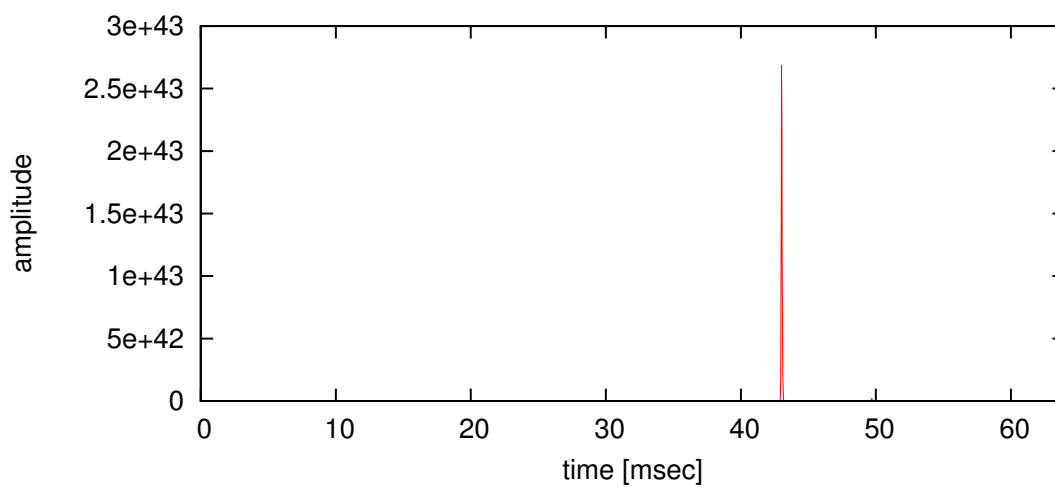


図 4.2: Exponential of the part of waveform.

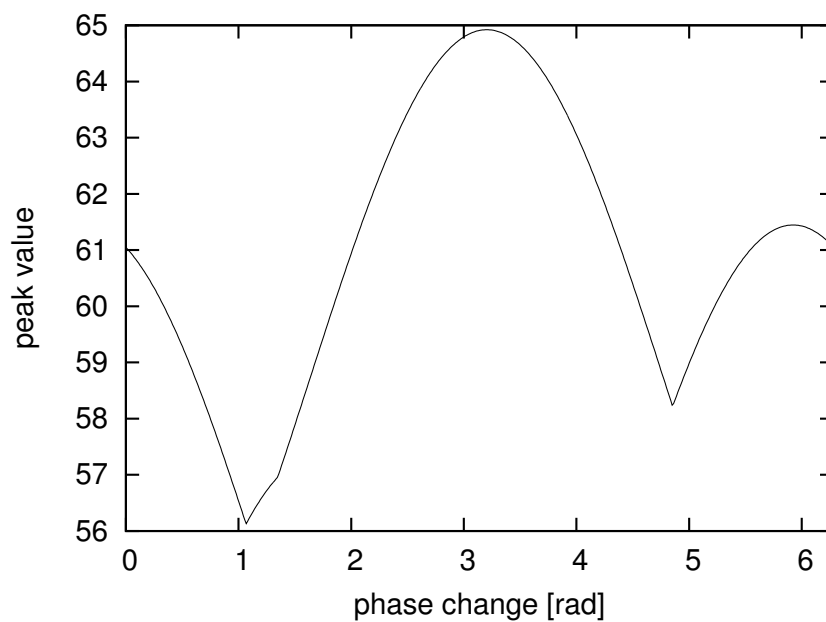


図 4.3: Change of peak value for varying phase.

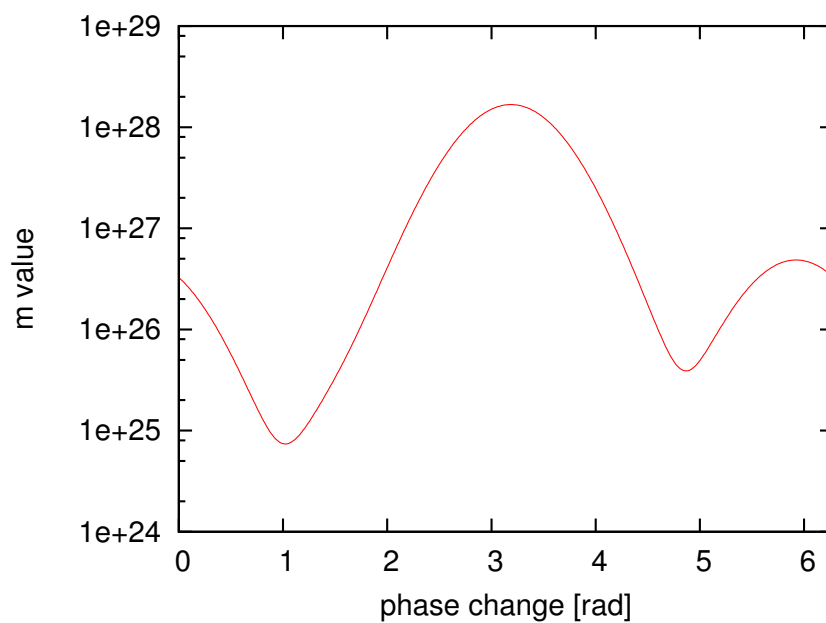


図 4.4: Change of approximation of peak value for varying phase.

4.4 ピーク値の微分

前の節で定義した m を用いることにより、ピーク値を微分可能な値で代替可能となった。最急降下法を用いるためには、 m を各周波数の位相で微分した値を求める必要がある。

最急降下法で用いる勾配を得るため、 m の式の x_t に離散逆フーリエ変換の式を代入し、 l 番目の周波数成分の位相 ϕ_l で編微分する。 l は $0 \leq l < N$ の整数である。

$$\begin{aligned} \frac{\partial x_{smax}}{\partial \phi_l} &= \frac{\partial}{\partial \phi_l} \sum_{t=0}^{N-1} \left[\exp \left\{ \frac{1}{N} \sum_{k=0}^{N-1} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{N} \sum_{k=0}^{N-1} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \right] \end{aligned} \quad (4.6)$$

$$\begin{aligned} &= \frac{\partial}{\partial \phi_l} \sum_{t=0}^{N-1} \left[\prod_k^{k \neq l} \exp \left\{ \frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \cdot \exp \left\{ \frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \right. \\ &\quad \left. + \prod_k^{k \neq l} \exp \left\{ -\frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \cdot \exp \left\{ -\frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \right] \end{aligned} \quad (4.7)$$

$$\begin{aligned} &= \sum_{t=0}^{N-1} \left[\prod_k^{k \neq l} \exp \left\{ \frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \right. \\ &\quad \cdot \frac{\partial}{\partial \phi_l} \exp \left\{ \frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \\ &\quad + \prod_k^{k \neq l} \exp \left\{ -\frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \\ &\quad \left. \cdot \frac{\partial}{\partial \phi_l} \exp \left\{ -\frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \right] \end{aligned} \quad (4.8)$$

$$\begin{aligned} &= \sum_{t=0}^{N-1} \left[\prod_k^{k \neq l} \exp \left\{ \frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \right. \\ &\quad \cdot \exp \left\{ \frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \cdot \frac{1}{N} j A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \\ &\quad + \prod_k^{k \neq l} \exp \left\{ -\frac{1}{N} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \\ &\quad \left. \cdot \exp \left\{ -\frac{1}{N} A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right\} \cdot -\frac{1}{N} j A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right] \end{aligned} \quad (4.9)$$

$$\begin{aligned}
 &= \sum_{t=0}^{N-1} \left[\exp \left\{ \frac{1}{N} \sum_{k=0}^{N-1} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \cdot \frac{1}{N} j A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right. \\
 &\quad \left. + \exp \left\{ -\frac{1}{N} \sum_{k=0}^{N-1} A_k \exp \left(\frac{2\pi j}{N} kt + j\phi_k \right) \right\} \cdot -\frac{1}{N} j A_l \exp \left(\frac{2\pi j}{N} lt + j\phi_l \right) \right] \quad (4.10)
 \end{aligned}$$

再び離散逆フーリエ変換を用いて、

$$\begin{aligned}
 &= \sum_{t=0}^{N-1} \left[\exp(x_t) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi lt}{N} + \phi_l \right) \right\} \right. \\
 &\quad \left. - \exp(-x_t) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi lt}{N} + \phi_l \right) \right\} \right] \quad (4.11)
 \end{aligned}$$

となる。

この式は、 $j A_l \exp\{j(2\pi lt/N + \phi_l)\}$ の部分で表される周波数 l の成分の微分値に、 $\exp\{x_t\}/N$ の項で表されるピークの大きさに応じた重み付けの係数を掛けていると考えることができる。負のピークに対応する部分も同様である。

式 4.11 ではフレームのオーバーラップの影響を考慮していない。実際の最適化では、模式図を図 4.5 に示す通り、窓関数を用いてフレームを切り出し、切り出したフレームの各周波数成分の位相を変更し、その後再び、切りだされた残りの部分と足し合わせる。ピークを最小化すべきは、窓関数によって切り出されたフレーム単体ではなく、窓関数によって切り出されたフレームと切りだされた残りを足した波形である。位相を変更するフレームの波形を x_t 、切りだされた残りを $x_{leftover,t}$ 、それらを足しあわせたものを $x_{total,t}$ とし、以下の式で表されるとする。

$$x_{total,t} = x_t + x_{leftover,t} \quad (4.12)$$

このとき、実際に最小化すべき値 m_{total} は次の式で表される。

$$m_{total} = \sum_t \left\{ \exp(x_t + x_{leftover,t}) + \exp(-x_t - x_{leftover,t}) \right\} \quad (4.13)$$

これを各周波数の初期位相 ϕ_l 微分した値は、 $x_{leftover,t}$ の値は ϕ_l に依存しないことを考えると、式 4.11 と同様に計算できて

$$\begin{aligned}
 \frac{\partial m_{total}}{\partial \phi_l} &= \sum_{t=0}^{N-1} \left[\exp(x_{total}) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi lt}{N} + \phi_l \right) \right\} \right. \\
 &\quad \left. - \exp(-x_{total}) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi lt}{N} + \phi_l \right) \right\} \right] \quad (4.14)
 \end{aligned}$$

となる。 A_l 、 ϕ_l は x_{total} のフーリエ変換により得られる振幅と位相ではなく、 x_t のフーリエ変換により得られる振幅と位相であることに注意が必要である。この値を用いることにより最急降下法による探索を行うことができる。

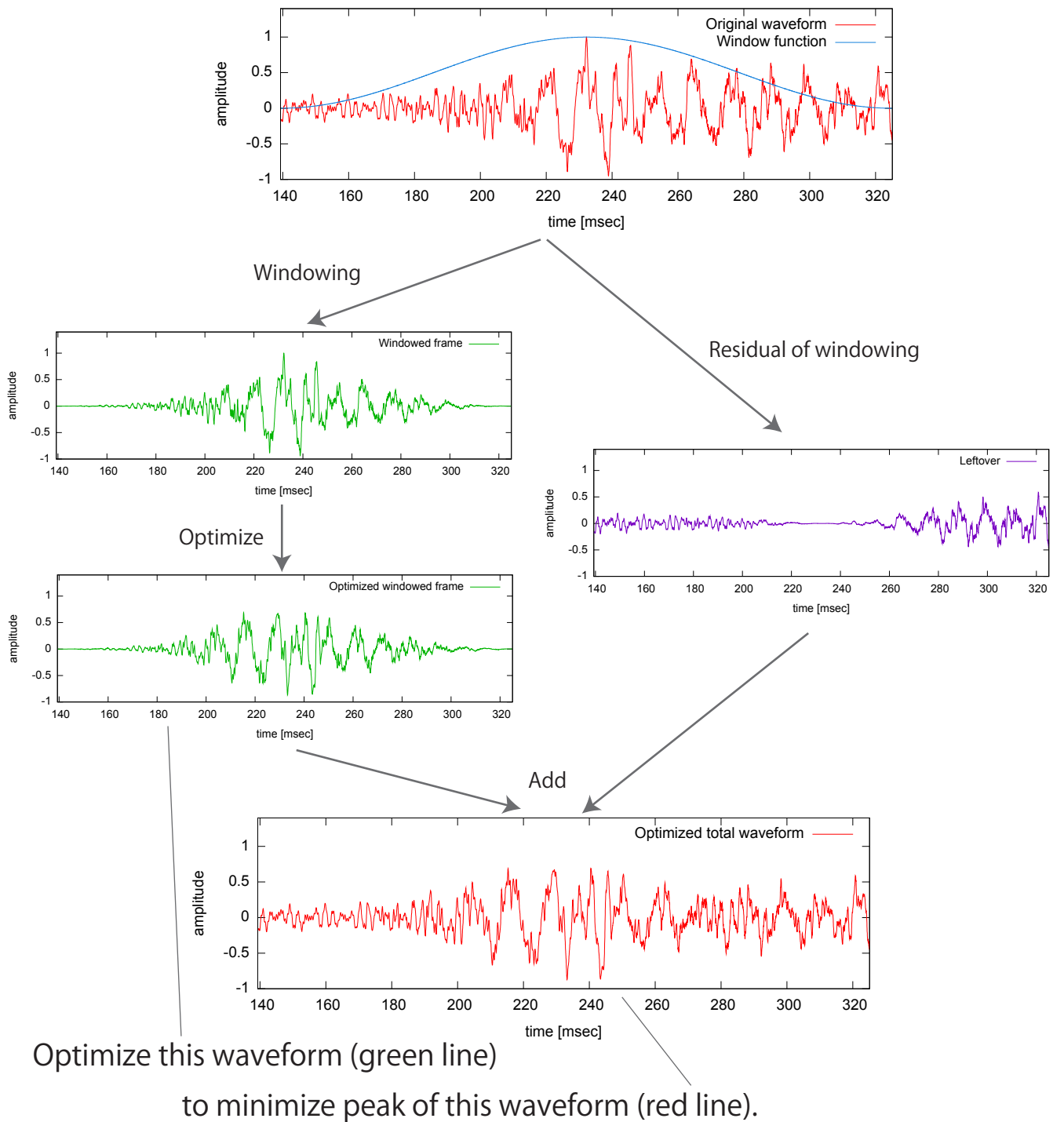


図 4.5: Edit windowed waveform to minimize total waveform.

4.5 微分計算の高速化

m_{total} を各周波数成分の位相で微分した値は、以下の式で表されることを先に述べた。

$$\begin{aligned} \frac{\partial m_{total,t}}{\partial \phi_l} = & \sum_{t=0}^{N-1} \left[\exp(x_{total,t}) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi l t}{N} + \phi_l \right) \right\} \right. \\ & \left. - \exp(-x_{total,t}) \frac{1}{N} j A_l \exp \left\{ j \left(\frac{2\pi l t}{N} + \phi_l \right) \right\} \right] \end{aligned} \quad (4.15)$$

フレームの長さを N とすると、1つの周波数について計算量が N 必要となり、それを $N/2$ 個の周波数について行うため、計算量は $O(N^2)$ となる。この微分の計算は最急降下法による探索の中で何度も行われるため、このまま計算すると非常に時間がかかってしまう。そこでこの式を以下のように変形する。

$$= \sum_{t=0}^{N-1} \left[\exp(x_{total,t}) \exp \left(j \frac{2\pi l t}{N} \right) \frac{1}{N} j A_l \exp(j\phi_l) \right] \quad (4.16)$$

$$- \exp(-x_{total,t}) \exp \left(j \frac{2\pi l t}{N} \right) \frac{1}{N} j A_l \exp(j\phi_l) \right] \quad (4.17)$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} \left[\exp(x_{total,t}) - \exp(-x_{total,t}) \right] \cdot \exp \left(j \frac{2\pi l t}{N} \right) \cdot j A_l \exp(j\phi_l) \quad (4.18)$$

$$= \text{IDFT}_{t=0}^{N-1} \left[\exp(x_{total,t}) - \exp(-x_{total,t}) \right]_l \cdot j A_l \exp(j\phi_l) \quad (4.19)$$

IDFT(\cdot) は離散逆フーリエ変換を表す。

離散逆フーリエ変換の部分は逆高速フーリエ変換を利用する事により、全ての l について計算量 $O(N \log N)$ で同時に計算できる。これにより、微分の式全体での計算量も $O(N \log N)$ となり、プログラムの高速化が可能になる。

4.6 制約条件

§3.3において、人間の聴覚は位相の変化に比較的鈍感であり位相を変化させても知覚できないことがあること、逆に位相を変化させることにより音質の劣化が生じる場合もあること、更にそれらの詳細な条件については未だ解明されていないことを述べた。

音質の劣化を抑えながらエネルギー密度を向上させるためには、探索の結果が音質の劣化を生じさせるような位相変化とならないようなんらかの制約条件を加える必要がある。

本研究では、§3.3で例示した位相変化によって音質の劣化が生じる場合と生じない例からの類推から、周波数方向になめらかな位相変化であれば音質の劣化が生じにくいという仮定をおき、FIR線形位相ローパスフィルタを位相変化量に対してかけることにより平滑化し、音質劣化防止のための制約とする。このとき、ローパスフィルタは時間領域の信号に対してではなく、周波数領域の位相変化量に対して適用されることに注意が必要である。

この制約条件を用いて位相の最適化を行った場合音色の劣化がどの程度抑えられるかについては、実験の章において検証を行う。

使用するFIRフィルタは、フーリエ級数と窓関数を用いる方法によって設計した[18]。FIRフィルタの周波数特性、位相特性、インパルス応答の一例を図4.6、4.7、4.8に示す。このフィルタのカットオフ周波数が低くなるほど、得られる位相変化はなめらかになる。

当然ながら、制約をかけない状態で探索を行った後に平滑化を行ったのでは、得られる結果はピーク値最小となる位相と異なったものとなる。したがって、探索によって得られる結果そのものが、ある位相特性に対して上記のフィルタをかけたものとなっている必要がある。

ここで、 k 回目の位相の更新における位相の変化量を $\phi_{\text{delta}}^{(k)}$ 、平滑化に利用するFIRフィルタのインパルスレスポンスを h_{lpf} とし、

$$\phi_{\text{delta}}^{(k)} = h_{\text{lpf}} * \theta_{\text{delta}}^{(k)} \quad (4.20)$$

となる θ を導入し、これを変数として最急降下法による探索を行う。 $\theta_{\text{delta}}^{(k)}$ は k 回目の更新における θ の変化量である。そうすると、得られる位相変化は $\theta_{\text{delta}}^{(k)}$ にローパスフィルタを適用したのものとなるため、 $\theta_{\text{delta}}^{(k)}$ がどのような値を取った場合であっても滑らかな位相変化となることが期待できる。図4.9に模式図を示す。 m_{total} を、 θ の各要素で偏微分した結果、すなわち θ の微小変化によって m_{total} の値がどれだけ変化するかの一近似的は、重ね合わせの定理より、畳み込みを用いて

$$\begin{pmatrix} \frac{\partial m_{\text{total}}}{\partial \theta_0} \\ \frac{\partial m_{\text{total}}}{\partial \theta_1} \\ \vdots \\ \frac{\partial m_{\text{total}}}{\partial \theta_{N-1}} \end{pmatrix} = h_{\text{lpf}} * \begin{pmatrix} \frac{\partial m_{\text{total}}}{\partial \phi_0} \\ \frac{\partial m_{\text{total}}}{\partial \phi_1} \\ \vdots \\ \frac{\partial m_{\text{total}}}{\partial \phi_{N-1}} \end{pmatrix} \quad (4.21)$$

で与えられる。また最急降下法の更新式より、 α を小さな定数として、

$$\theta_{\text{delta}}^{(k)} = -\alpha \begin{pmatrix} \frac{\partial m_{\text{total}}}{\partial \theta_0} \\ \frac{\partial m_{\text{total}}}{\partial \theta_1} \\ \vdots \\ \frac{\partial m_{\text{total}}}{\partial \theta_{N-1}} \end{pmatrix} \quad (4.22)$$

である。これらの式を用い、 θ を変数として最急降下法を行うことにより、なめらかな位相変化という条件を保った範囲で m_t を最小化することができる。

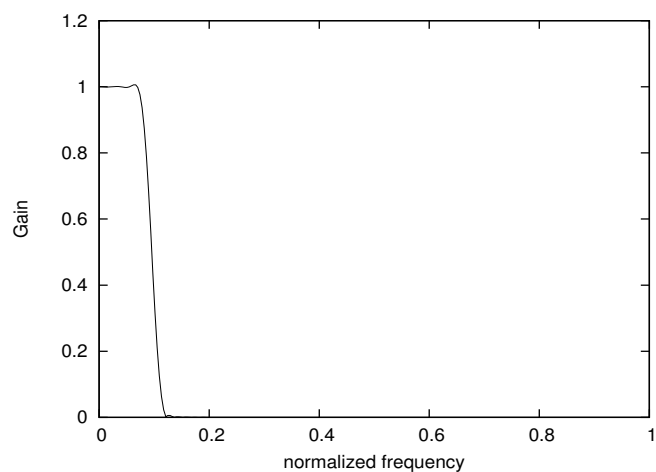


図 4.6: Gain characteristic of an FIR low pass filter.

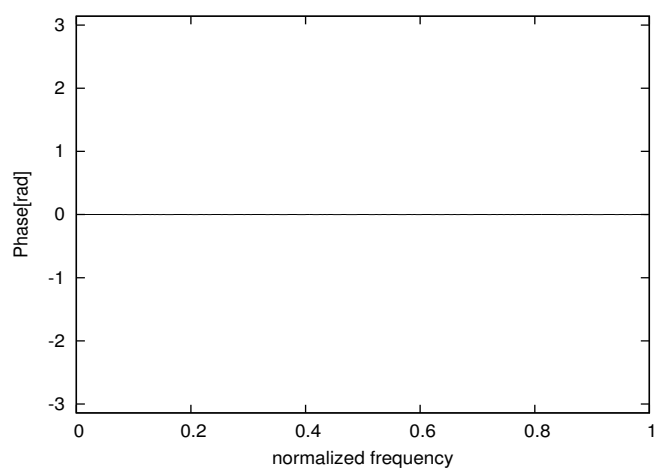


図 4.7: Phase characteristic of an FIR low pass filter.

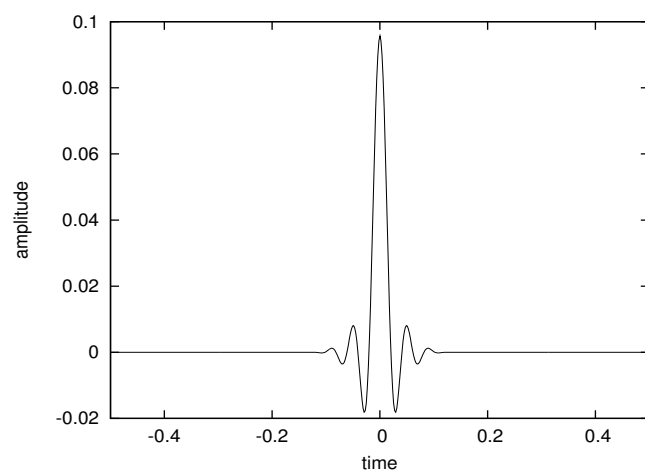


図 4.8: Impulse response of an FIR low pass filter.

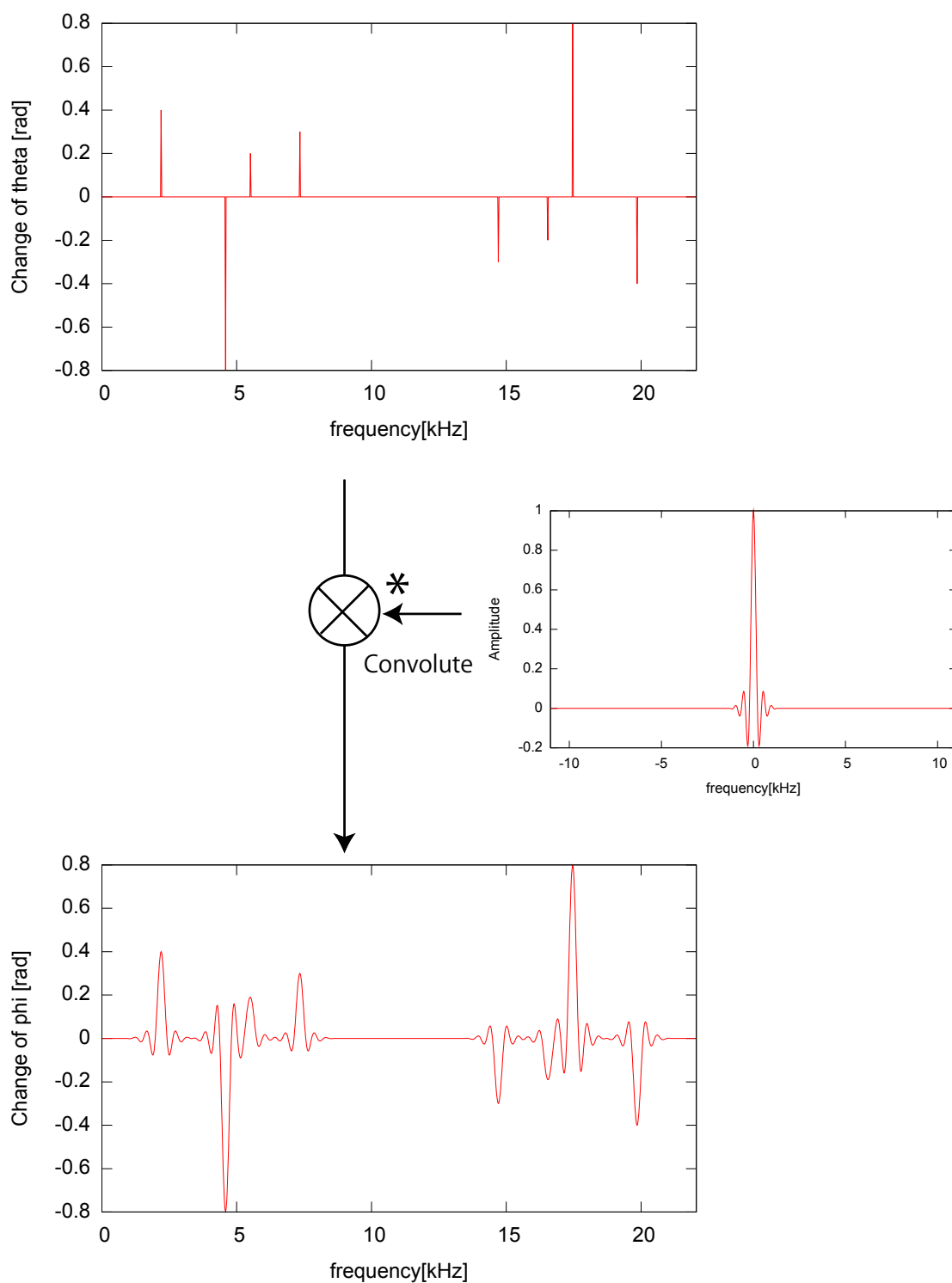


図 4.9: Change of variable vector θ_{delta} and smoothed phase change ϕ_{delta} .

尚、フレーム内の各周波数成分の位相を変更した場合、フレーム境界での波形の不連続が発生し、音質が劣化する場合がある。従来手法では各周波数成分の位相を変更した後にもう一度窓掛けを行うことで波形の不連続を解消していたが、これにより波形の振幅スペクトルが変化し、それによって音質が劣化するといった問題が生じていた。

位相変化量の平滑化による制約条件を用いた場合であれば、位相変化量が平滑化されることにより群遅延量も小さくなり、フレーム境界での不連続も抑えられることが期待できる。図 4.10、4.11、4.12 にそれぞれ、窓掛けを行ったフレームの位相最適化前の波形、窓掛けを行ったフレームに対し制約条件を用いずに 10 ステップの位相最適化を行った波形、窓掛けを行ったフレームに対し位相を平滑化する制約条件を用いて 10 ステップの位相最適化を行った波形を示す。制約条件を用いずに位相の最適化を行った場合窓関数の包絡が大きく崩れ、フレーム境界で不連続性が発生するような波形となっているが、位相を平滑化する制約条件を用いた場合であれば窓関数の包絡はよく保存されており、特別な後処理を行わなくてもフレーム境界での不連続性が発生しにくいことがわかる。

4.7 フレームの切り出し

音の信号処理を行う場合、フレームの切り出しは一定の時間間隔毎に切り出しを行い、前のフレームから順に処理をしていくことが一般的である。しかし今回の用途では、位相を変更するフレームの終端付近に大きなピークが存在する場合(このピークを仮にピーク A とする)、窓関数の包絡を保ったままではピークの最小化は有効に行われぬ 4.13。これは最大のピークを優先的に最小化しようとするためである。フレームの中央付近にピーク A より少し小さいピーク(ピーク B とする)が存在した場合、仮に次のフレームにおける処理でピーク A が小さくなったとしても、ピーク B はそのまま残ってしまい、波形全体でのピーク最小化の効率は非常に悪くなる。仮に窓関数の包絡を保つという条件をなくした場合でも、そのピークを最小化するためにはフレームの終端付近で大きな振幅を持つ必要があり、フレーム境界での波形の連続性に悪影響を及ぼし、音質の劣化につながる。したがって効率的にピーク値を最小化するためには、位相を変更するフレームの中央付近にピークが存在することが望ましい。そこで図 4.14 に示すように、最急降下法の 1 ステップ毎に波形全体で最大のピークを探し、そこを中央とした窓関数を用いてフレームを切り出し、切り出したフレームと切り出された残りの和のピークを最小化するように切り出されたフレームの位相を変更するようにした。

4.8 まとめ

本章では、指数関数を用いたピーク値のなめらかな関数での代用法とそれを各周波数成分の位相で微分した値を求めた。また、FIR ローパスフィルタを用いた音質を劣化させないための制約条件を導入し、それを最急降下法へ組み込む方法を提案した。更に、処理を行う際には窓関数を用いてフレームを切り出す必要があるが、効果的に位相の最適化が行えるフレームの切り出し方を提案した。

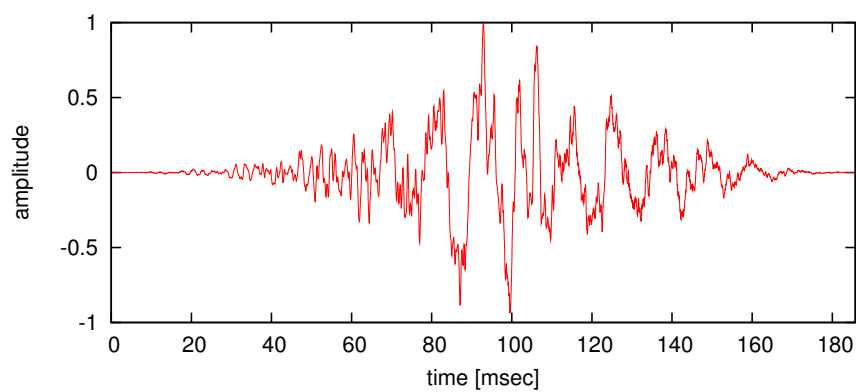


図 4.10: Windowed waveform.

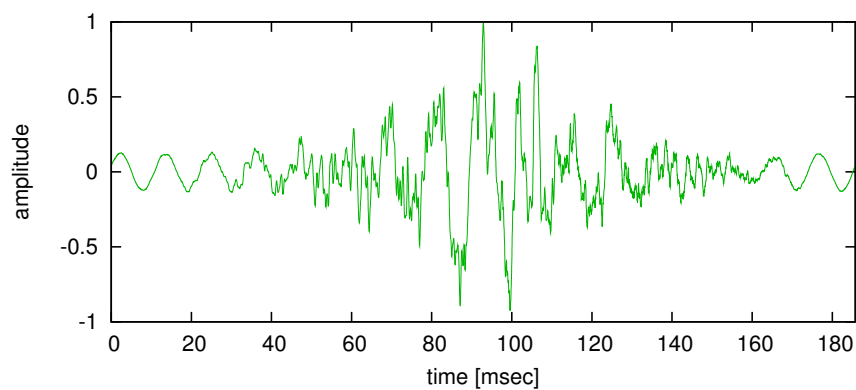


図 4.11: Optimized windowed waveform without the constraint.

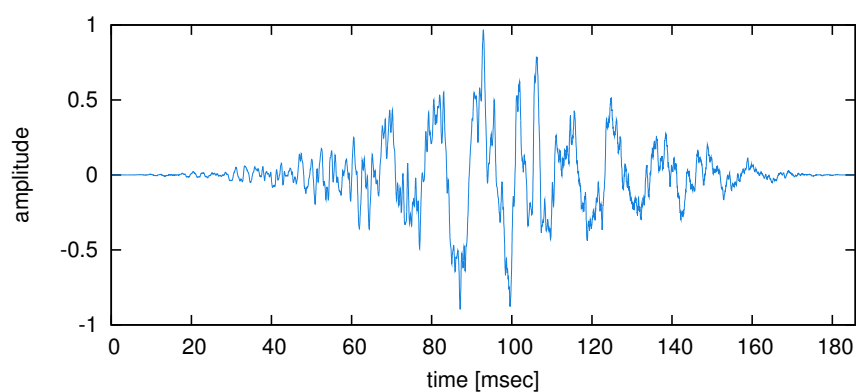
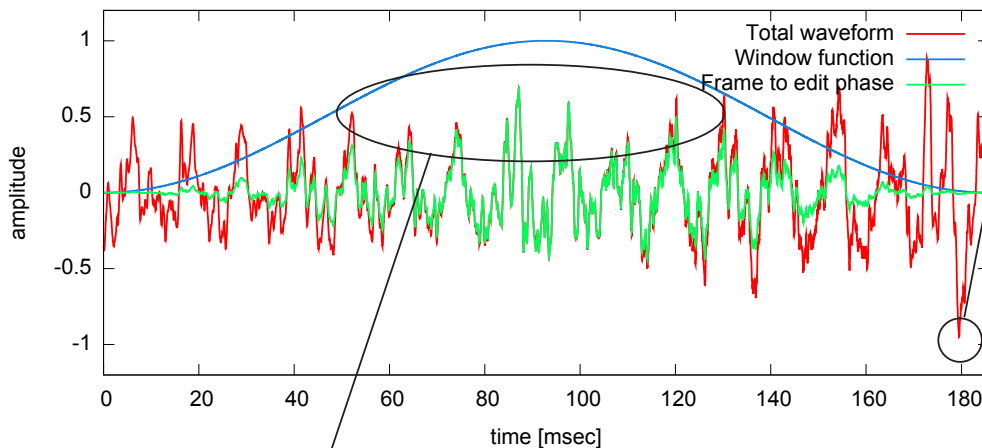


図 4.12: Optimized windowed waveform with the constraint.

Phases of windowed waveform (green line) are optimized to minimize peak of total waveform (red line).

The process tries to minimize this peak (Peak A), but can not.



These peaks is (Peak B) left almost unchanged.

図 4.13: How the process work if large peak exist on the end of the frame.

これらをまとめると、実際の位相最適化の手順は以下のようになる。

1. 波形全体のピーク値が 100 となるよう波形全体を定数倍する。
2. 波形の中で最大のピークを探し、その時刻を中心とした窓関数を用いてフレームを切り出す。
3. 式 4.19 を用い、 m_{total} を ϕ の各要素で微分した値を得る。
4. 式 4.21 を用い、 m_{total} を θ の各要素で微分した値を得る。
5. 式 4.2 の $\mathbf{a} = \theta$ 、 $f(\mathbf{a}) = m_{total}$ とし、今回の更新における θ の変化量 θ_{delta} を得る。
6. 式 4.20 を用い、なめらかな位相変化量 ϕ_{delta} を得る。
7. 切り出したフレームの各周波数の位相を ϕ_{delta} に応じて変化させる。
8. 手順 2 から手順 7 までを任意の回数繰り返す。
9. 波形全体のピーク値が適当な値 (16 [bit] の wav であれば 2^8 等) となるよう波形全体を定数倍する。

以上の手順により、なめらかな位相変化という条件を満たしつつ位相を最適化し、エネルギー密度を向上させることができる。

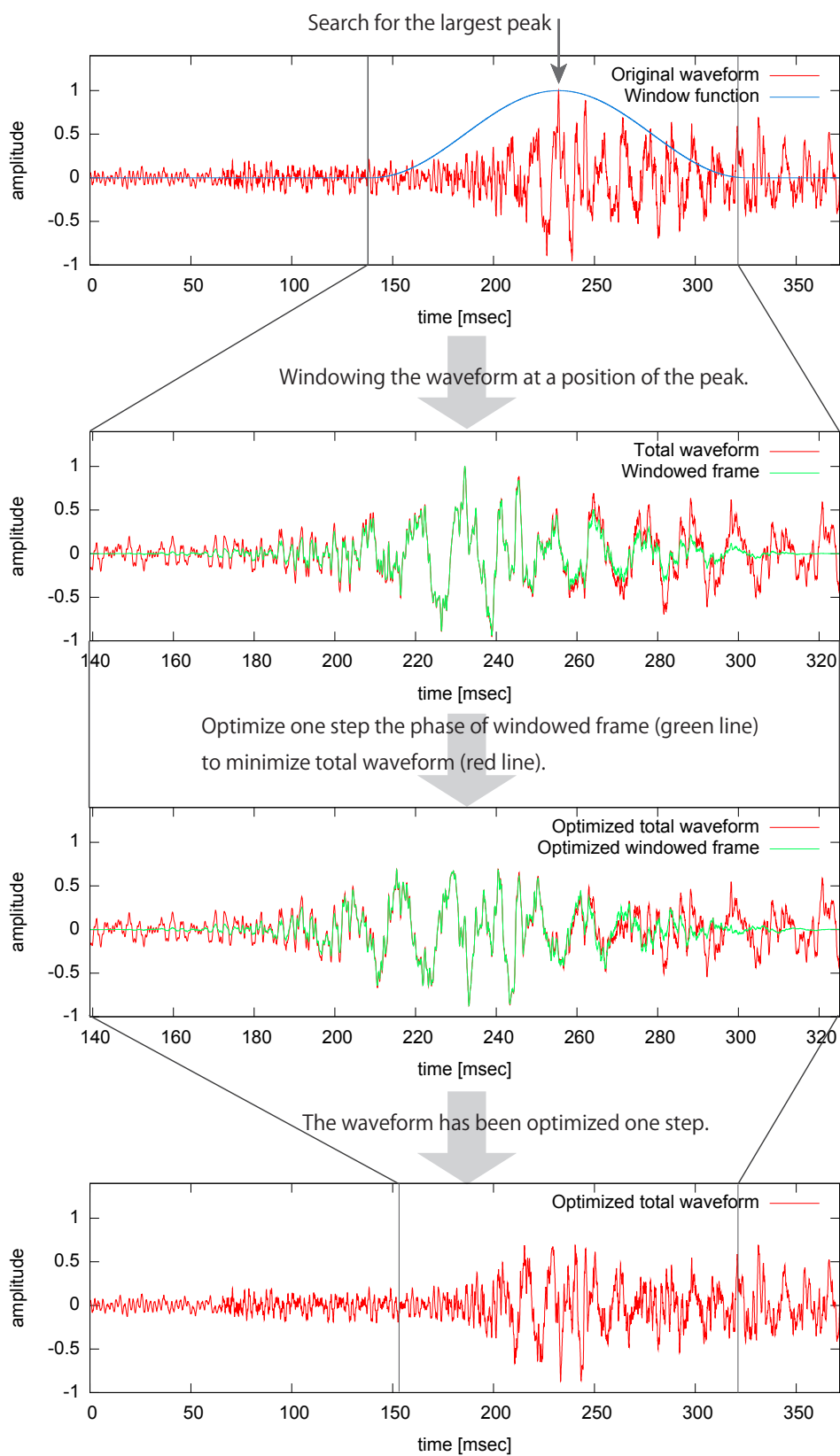


図 4.14: Procedure of effective windowing.

第5章

実験と考察

5.1 はじめに

本章では、本論文で提案した手法と、[1]の手法について探索にかかる時間とエネルギー密度の向上について比較を行う。また本論文で提案した制約条件が探索の速度に与える影響について評価を行う。本論文提案した手法、[1]の手法、AGCの3つの手法を用いてエネルギー密度の向上を行った場合の音質の劣化について評価を行う。更にAGCとの特性の比較を行い、それらの効果的な使い分けについての検討を行う。

5.2 制約を用いない場合の探索速度の評価

5.2.1 実験条件

従来手法と提案手法による最適化の速度を比較するため、位相の最適化を行い一定のRoot Mean Square(RMS)の利得が得られるまでの計算時間を測定する。ここでRMSの利得とは、位相を最適化することによりピークがどの程度減少するかを表す。減少した分だけ増幅率を上げる事ができることに対応する指標である。

実験に用いる信号として、ここでは長さが8192のガウシアンノイズを生成して用いた。図5.1に実験に用いた信号のピーク付近の256サンプルを示す。離散フーリエ変換のフレーム長 N は、提案手法・従来手法の双方で8192とし、1フレームの信号の最適化を行った。ここでは純粋に探索の部分の性能のみを比較するために、従来手法・提案手法共に制約条件は用いず、また窓関数は矩形窓を用いた。従来手法・提案手法共にJavaで実装され、CPUはIntel®Core 2 Duo P8700 (2.53 [GHz])、メモリは4 [GB] DDR2 RAMを搭載した計算機により実行した。また計算時間はシステムタイマーを用いて計測した。

5.2.2 結果と考察

図5.2に、計算時間とRMSの利得の関係を示す。従来手法と提案手法どちらの場合でも、計算時間が長くなるに従ってRMSの利得は上昇しているが、途中から徐々に飽和が生じている。飽和が生じるRMSの利得は従来手法ではおよそ3 [dB]であるが、提案手法であれば飽和が生じるまでにおよそ11 [dB]向上させることができている。

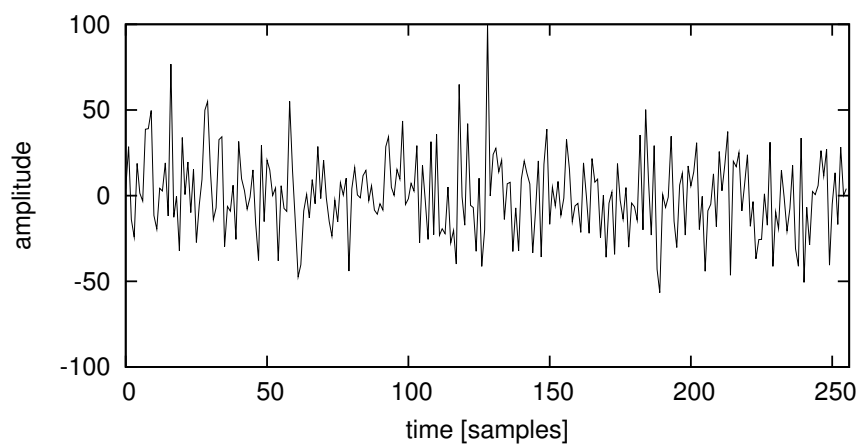


図 5.1: A part of the test signal.

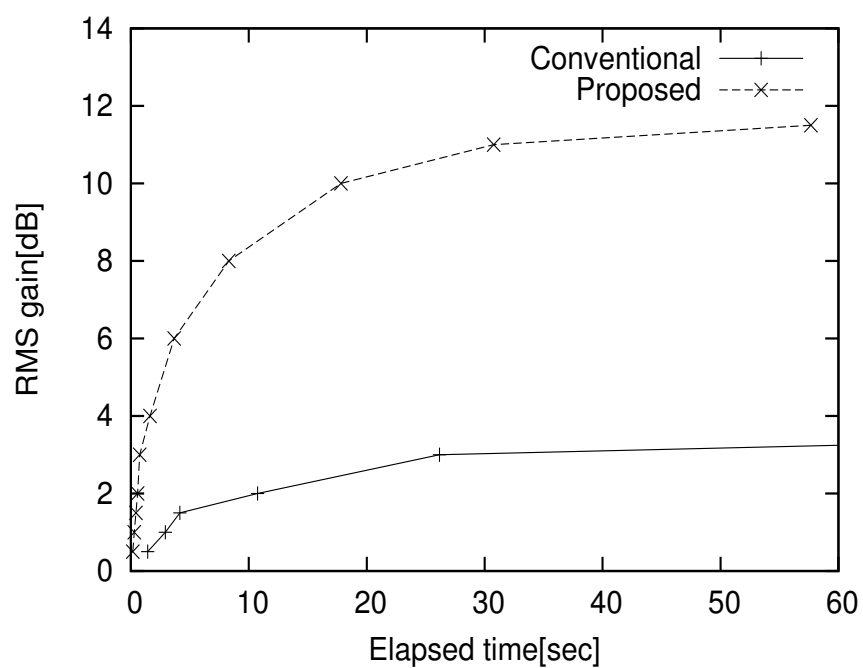


図 5.2: RMS gain as a function of computational time by proposed and conventional method.

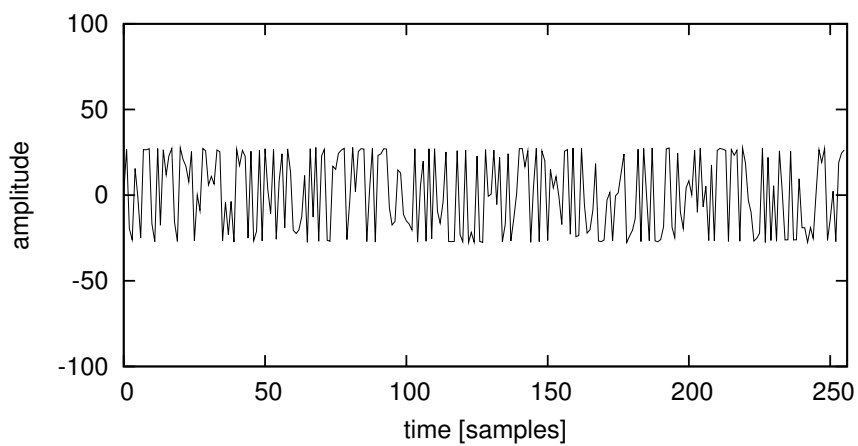


図 5.3: A part of the test signal optimized by proposed method.

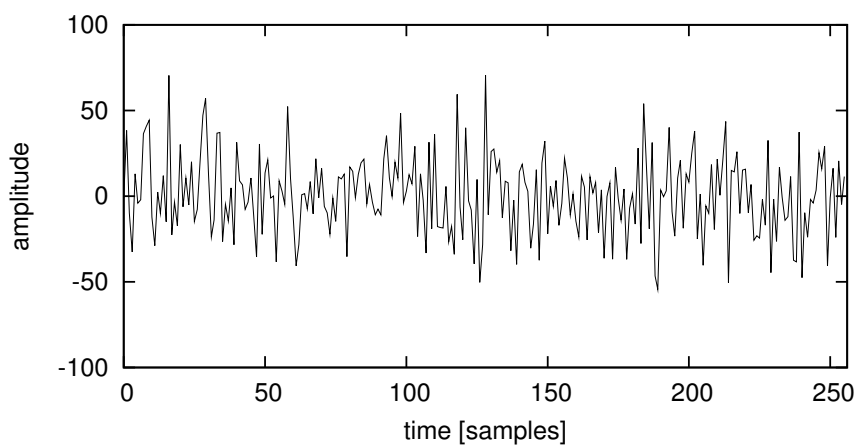


図 5.4: A part of the test signal optimized by conventional method.

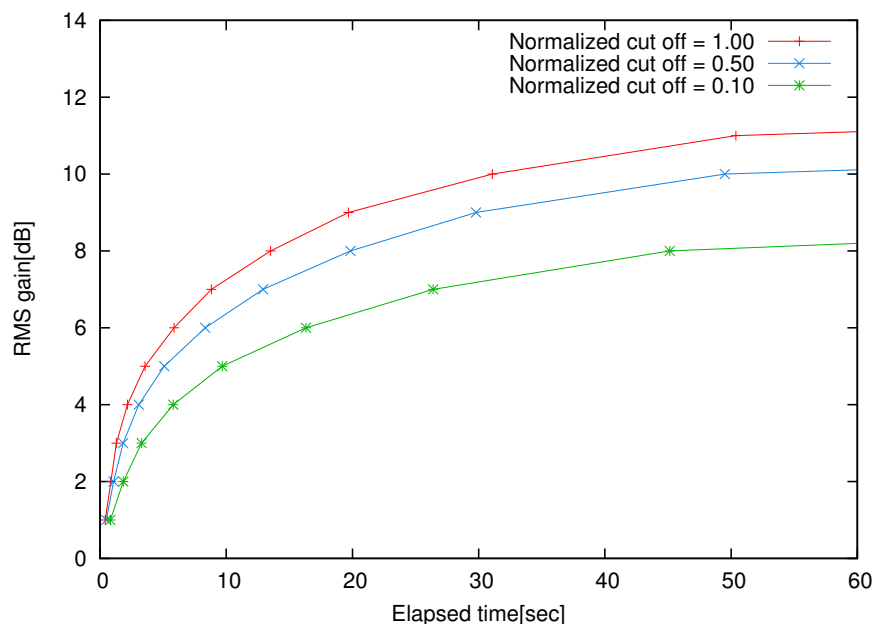


図 5.5: RMS change as a function of computational time for varying smoothness of phase change.

図 5.3 と図 5.4. に、それぞれお従来手法と提案手法でおよそ 30 [sec] の最適化を行った場合の波形を示す。ピーク値の変化の比較が容易になるよう、従来手法による最適化では、波形のピークがある程度減少してはいるが大きくは変化していないのに対し、提案手法による最適化では波形が大きく変化し、ピーク値が非常に小さくなっていることが見て取れる。

5.3 制約条件を用いた場合の探索速度の評価

5.3.1 実験条件

位相を平滑化する制約条件が探索の速度にどの程度の影響を与えるかを検証するため、位相を平滑化するローパスフィルタのカットオフを変化させて位相の最適化を行い、一定の Root Mean Square(RMS) の利得が得られるまでの計算時間を測定する。

実験に用いる信号は、前節で用いたのと同じ、長さが 8192 のガウシアンノイズを生成して用いた。離散フーリエ変換のフレーム長 N は 8192 とし、窓関数には矩形窓を用いて 1 フレームの信号の最適化を行った。§5.2 での実験と同様、位相最適化のプログラムは全て Java で実装され、CPU は Intel®Core 2 Duo P8700 (2.53 [GHz])、メモリは 4 [GB] DDR2 RAM を搭載した計算機により実行した。また計算時間はシステムタイマーを用いて計測した。

表 5.1: Musics used for experiment.

Music No.	Position[mm:ss]
RWC-MDB-P-2001 No.100	04:13 - 04:22
RWC-MDB-C-2001 No.3	00:11 - 00:18
RWC-MDB-C-2001 No.36	00:00 - 00:15
RWC-MDB-J-2001 No.31	00:42 - 00:52

5.3.2 結果と考察

図 5.5 に、位相変化量を平滑化するフィルタのカットオフ周波数を変化させたときの計算時間と RMS の利得の関係を示す。カットオフ周波数は、ナイキスト周波数を 1 とする正規化周波数で示している。カットオフ周波数が 1.00 での探索結果は、制約をかけない場合の探索結果と理論上同一である。一定の RMS の利得を達成するためにかかる時間が前節での示した提案手法を用い制約をかけず探索を行った場合のおよそ 2 倍となっているが、これは位相の平滑化の計算をしている分計算時間が増加し、最急降下法 1 ステップあたりの計算時間がおよそ 2 倍になっているためであると考えられる。

位相変化量を平滑化するフィルタのカットオフ周波数が小さくなると一定時間あたりの RMS の利得は小さくなる。また RMS の利得の飽和も RMS の利得が比較的小さい段階で起こるようになる。これは制約条件が厳しくなるに従って、取れる位相の組み合わせが少なくなっていると考えると妥当である。したがって位相変化量を平滑化するフィルタのカットオフ周波数は、音質の劣化が抑えられる範囲でなるべく大きな値とすることが望ましい。

5.4 音質の劣化の評価

5.4.1 条件

i) 実験用音源

§4.6 において提案した音質劣化防止のための制約条件により、音質の劣化がどの程度抑えられるかを検証するための聴取実験を行う。実験用のデータとして、RWC 音楽データベース [19][20][21][22] より、ポピュラー音楽 1 曲、クラシック音楽 2 曲、ジャズ音楽 1 曲の、計 4 つの楽曲を選択し、楽曲中で最大のピークの周辺の 10 [sec] 程度の部分を切り出しこれを用いた。表 5.1 に、用いた楽曲番号と、サンプルを切り出した時刻を示す。

今後、これらの楽曲番号のうち、全ての楽曲番号に共通している部分を省略し、"RWC-MDB-P-2001 No.100" は "P-100"、"RWC-MDB-C-2001 No.3" は "C-3" のように表記することとする。

これらのサンプルを、提案手法による位相の最適化、従来手法による位相の最適化、AGC によるエネルギー密度向上、mp3 への変換の 4 通りの方法により編集し、編集方法による音質の変化の比較を行う。

表 5.2: Parameter setting of AGC.

Parameter	Value
Ratio	8:1
Knee	6.0 [dB]
Threshold	Depending on clips
Attack	0.0 [ms]
Release	40,200, or 1000 [ms]

提案手法を用いた編集では、制約条件として用いるフィルタとしてフィルタの長さ 16 の線形位相ローパスフィルタを用いた。フィルタのカットオフ周波数は、予備実験においてカットオフ周波数が正規化周波数で 0.1 以下としたとき音質劣化が十分に抑えられたため、0.1 とした。またそのとき、“P-100”では約 5 [dB]、それ以外の楽曲においてはおよそ 7 [dB] エネルギー密度が向上した時点で、探索は飽和した。また、飽和した後に長時間探索を続けると、顕著な音質の劣化が発生する場合が確認された。そのため、聴取実験に用いるデータとして、提案手法を用いて“P-100”に対しては -5.0 [dB]、それ以外の楽曲に対しては -6.0 [dB] ピーク値の削減を行ったものを用いた。またフレーム長 N も音質の劣化に影響し、予備実験において 8192 以上のとき音質の劣化が十分に抑えられたため、フレーム長 $N = 8192$ とした。

従来手法を用いた編集では、各フレームで振幅の大きい周波数から順に 24 の周波数について最適化を行った。このとき、ピークの削減量はおよそ 0.0 [dB] から -3.8 [dB] であった。

AGC を用いた編集では、cakewalk®”sonitus:fx compressor”を用いて編集を行った。各パラメータは表 5.2 に示す通りに設定し、threshold の値を適宜変更することにより、提案手法で行ったのと同様の RMS が得られるまでピークを削減した。ただし、“P-100”に対しては threshold の調整だけでは必要なピーク削減量が得られなかったため、release を 40 [ms] まで減少させた。また“C-3”に対しては、release の値が 200 [ms] の場合 AGC による圧縮が解除される際の音量の変化が顕著であったため、release は 1000 [ms] とした。その他の 2 つのサンプルに対しては、release は典型的な値である 200 [ms] とした。

これら 3 つのエネルギー密度向上の手法の他に、ビットレート 128 [kbps] の mp3 へ変換したサンプルを用意した。これは聴取実験のスコアから実際の音質劣化の度合いを理解しやすくするための参考である。

これら計 16 の編集されたサンプルと 4 つの編集の行われていないのサンプルは、音の大きさを区別がつけられないことがないように、同じ RMS になるように振幅が調整された。

ii) 聴取実験の方法

聴取実験は、ITU-R BS.1116[23] で推奨される方法によって行われた。

ITU-R BS.1116 では、リファレンス音 1 つと評価対象音 2 つの合計 3 つのサンプルが 1 セットとして提示され、2 つの評価対象音に対し、リファレンスからどの程度劣化しているか、表 5.3 に従い、1.0 ~ 5.0 まで 0.1 刻みの点数が付けられる。

表 5.3: Evaluation scale of BS.1116.

Impairment	Grade
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

このとき、2つの評価音のうち片方はリファレンス音と全く同一な音となっている。従って、2つの評価対象音のうちどちらか片方は必ず5.0の評価を与えることとなる。2つの評価対象音のうちどちらがリファレンスと同一の音であるかは毎回ランダムに決められる。3つのサンプルは任意の順番で聞くことができ、また何度でも聞き直すことができる。

編集の方法や評価者の好みによっては、リファレンス音より編集されたサンプルの方が「いい音」と判断される場合が考えられが、評価者の好みによって結果が変動することを防ぐため、リファレンス音からの変化は全て劣化として点数はつけられる。評価者がリファレンスと同一なサンプルと編集されたサンプルを正しく判別し、編集されたサンプルに4.9以下の点数を付けた場合であれば付けられた点数がそのままそのサンプルのスコアに、逆にリファレンスに4.9以下の点数を付けた場合であれば、5.0からリファレンスに付けられた点数を引いた値がサンプルのスコアとなる。編集されたサンプルに全く音質の劣化がない場合であれば、約半数の人が編集されたサンプルに4.9以下の点数を、残り半数の人がリファレンスと同一のサンプルに4.9以下の点数を付けることが期待されるため、評価者が十分に多い場合、全ての評価者について平均すればスコアは5.0に十分に近い値になる。評価者の数がそれほど多くない場合、統計誤差により、編集されたサンプルのスコアが5.0より大きい値になる場合も存在する。

この手法は毎回3つの音を比較するため評価に時間がかかるが、僅かな劣化でも検出することができるため劣化の少ないシステムの評価に向いているとされている。

聴取実験には、14名の正常な聴力を持つ大学院生、大学生が参加した。全てのサンプルはSONY®”MDR-CD900ST.”を用いて提示された。

5.4.2 結果と考察

図5.6に編集されたサンプルのスコアを示す。また、図5.7、5.8に”J-31”のリファレンス音のピーク付近の波形と、制約条件付きの提案手法によって位相最適化された、同じ箇所の波形を示す。

従来手法による位相最適化では、スコアは全体的に非常に低い。”P-100”のスコアは比較的高い値となっているが、これは”P-100”は他の音源と比較して音数の多い音源であり、音質の劣化が比較的目立ちにくかったためであると考えられる。これらのサンプルはRMSの利得も他の編集方法と比較して低いことから、音質・実行速度の両面で性能が低いといえる。

AGCによるスコアは”C-36”と”J-31”については高いスコアとなっているが、”P-100”のスコアは非常に低く、また”C-36”のスコアもそれほど高くない。”P-100”の音質の劣化が激しいのは、”P-100”が予めAGCを適用されていたサンプルであったことが原因であると考えられる。§2で述べたとおり、AGCを用いるとある程度のRMSまでは比較的小さな音質劣化で上昇させることができるが、それ以上のRMS値を達成しようとする、顕著な音質の劣化が発生するようになる。”P-100”ではreleaseの値を小さくしなければ5.0 [dB]のRMSの上昇を達成することができず、急峻な増幅率の変化が発生し、それが音質劣化の原因になったと考えられる。また”C-3”はサンプルの途中に急激に音量が大きくなる箇所が存在する。その箇所において、AGCは増幅率を急激に下げるため、増幅率の変化を容易に知覚することができ、その箇所が不自然な音として知覚されたと考えられる。またAGCは前述の箇所ですで一度下げた増幅率をその後徐々に上げていくが、その箇所はリファレンス音とは異なる音量の変化をしていることが知覚しやすく、ここでもリファレンス音と編集されたサンプルとの判別が容易になったと考えられる。

提案手法による位相の最適化では、4種全てのサンプルについて非常に高いスコアとなっており、特に”C-3”、”C-36”ではリファレンス音から全く劣化していないと判断できる。”C-3”においてスコアが5以上となっているが、これは統計誤差によるものだと考えられる。これら4つのサンプルに対し、提案した制約条件は有効に働いたと考えられる。ここで興味深い点は、図5.8に示した提案手法によって位相を最適化した波形は波形だけ見れば出力の飽和した信号の波形とよく似ているが、音質は出力が飽和した場合と異なり、殆ど変化していないことである。尚、スコアには殆ど影響が出ていないが、”P-100”ではベースの音に僅かな劣化が認められた。これは他のサンプルではRMSの利得の上昇が飽和する前に位相の最適化を終了しているのに対し、”P-100”ではRMSの利得の上昇が飽和するまで位相の最適化を行ったことが一因であると考えられる。

Mp3によるスコアについては、本論文の趣旨からはずれるため、説明を割愛する。

5.5 提案手法と Auto Gain Controlの比較

本論文で提案した手法は、探索効率・音質の両面で従来の位相最適化手法より優れていることが確認された。これら2つの位相最適化手法は基本方針は同じであり、音質・実効速度の両面で勝る本論文で提案した手法を用いることが適当であると考えられる。一方で位相最適化手法とAGCではエネルギー密度を向上させるという点では同じだが、動作が異なるため目的によって使い分ける必要があると考えられる。そこで本節では位相最適化手法とAGCの幾つか特徴についての考察を行う。

5.5.1 音質

先の節で述べたとおり、位相の最適化によってエネルギー密度の向上が行われた場合、判別不可能な程度しか音質は変化しない。それに対してAGCではある程度の音質の変化は避けられない。楽器ごとに個別に編集する際にはこの音質の変化が積極的に利用される

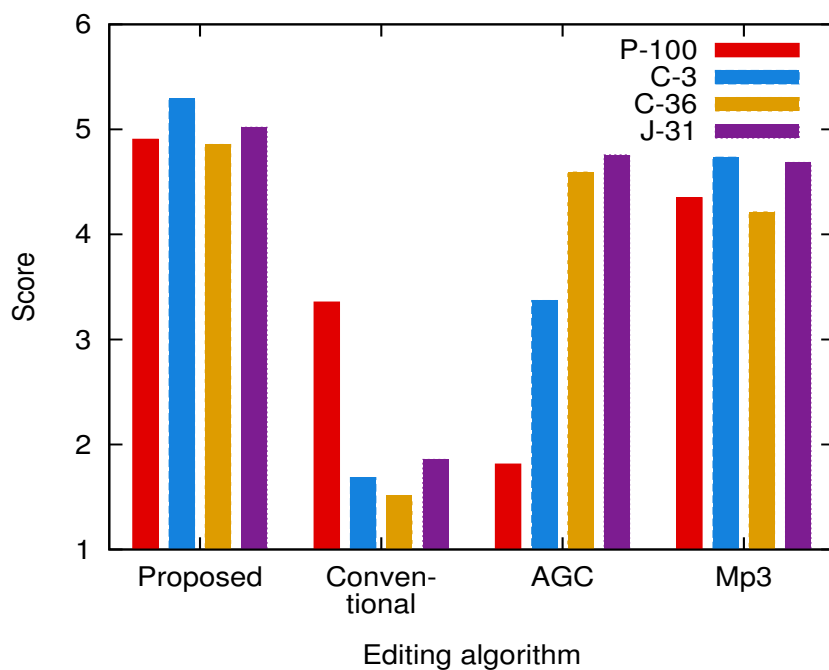


図 5.6: Scores of each clips.

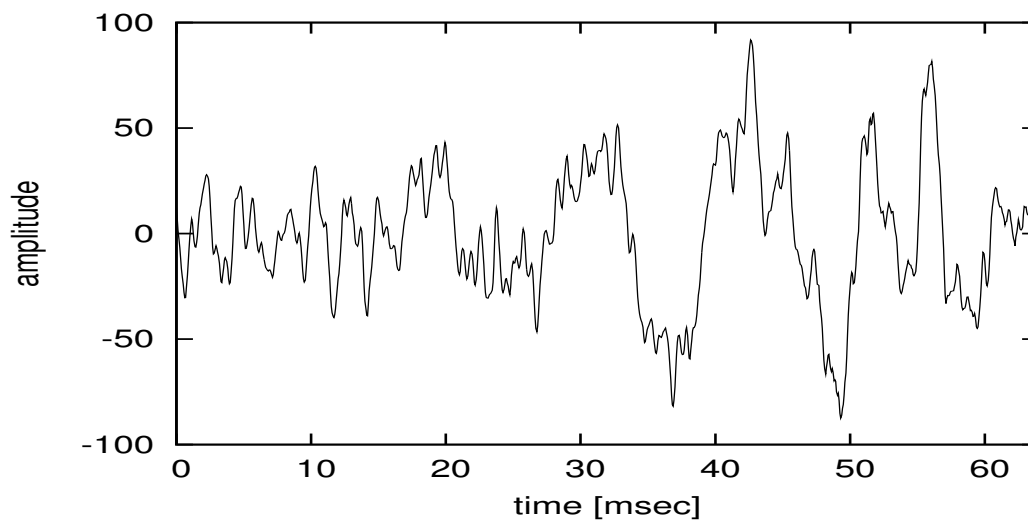


図 5.7: Original waveform.

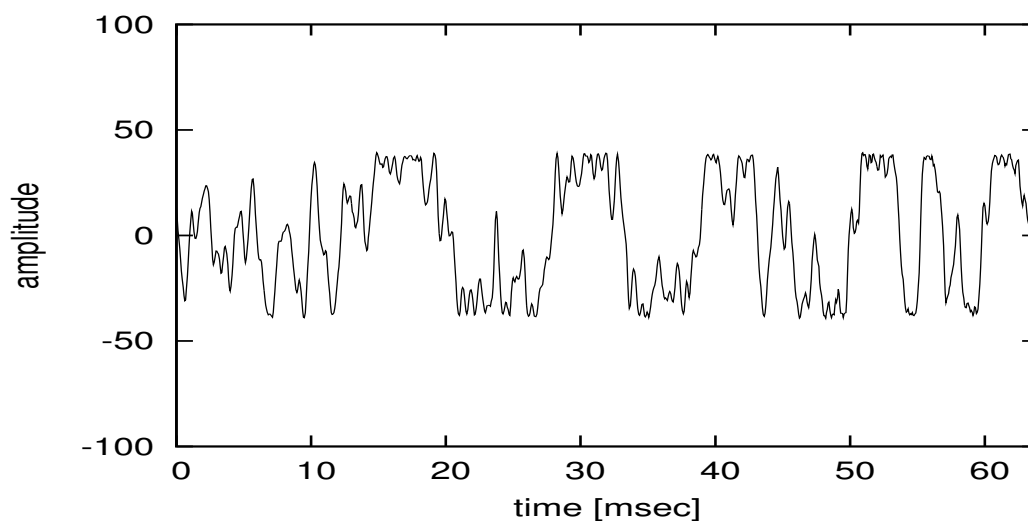


図 5.8: Phase-optimized waveform.

場合もあるが、楽曲の仕上げの際に用いられるときは音質の変化が小さい方が望ましい。

5.5.2 エネルギー密度の時刻変化

図 5.9 に”C-3”の、編集の行われていないサンプル、提案手法による位相最適化を行ったサンプル、AGCによる編集を行ったサンプルのエネルギー密度の時間変化を示す。また図 5.10、5.11、5.12 に、上記の3つのサンプルに対し編集後の波形の最大値が1.0となるようノーマライズを行った場合のエネルギー密度の時刻変化を示す。縦軸の Local RMS (Root Mean Square) は、波形に対し横軸の値を開始時刻とした長さ約 46 [ms] のハン窓を掛け、Root Mean Square を取ったものであり、各時刻でのエネルギー密度に対応する。

図 5.9 を見ると、2.5 [sec] より以前や 6.0 [sec] よりも後のようなエネルギー密度の小さい箇所では本論文で提案した手法による位相最適化、AGC 共にオリジナルのエネルギー密度の変化とよく一致しているが、エネルギー密度の比較的大きな箇所では AGC はオリジナルのエネルギー密度から大きく離れている。特に 2.5 [sec] から 3.2 [sec] までの AGC の動作が始まったと思われる区間では、オリジナルの波形のエネルギー密度は時間と共に大きくなっていったのに対し、AGC をかけた波形のエネルギー密度は逆に時間と共に小さくなってしまっている。一方本論文で提案した手法による位相最適化では、時刻 3.9 [sec] の前後でオリジナルよりエネルギー密度が小さくなっている他はオリジナルのエネルギー密度の変化とよく一致している。位相のみを変えているのにエネルギー密度が変化してしまっているのは、位相を変更するフレームのみに注目すればエネルギー密度は変化しないものの、位相を変更した後に前後のフレームと足し合わせる際、それらのフレームとの位相の関係性が変化し振幅が小さくなるよう働いたためであると考えられる。

また図 5.10、5.11、5.12 を見ると、時刻 3.9 [sec] 付近にあるエネルギー密度が最大となる点を見ると、AGC で編集された波形では local RMS が 0.4 程度であり、オリジナルのものと殆ど変わっていない。一方本論文で提案した手法による位相最適化では、local RMS

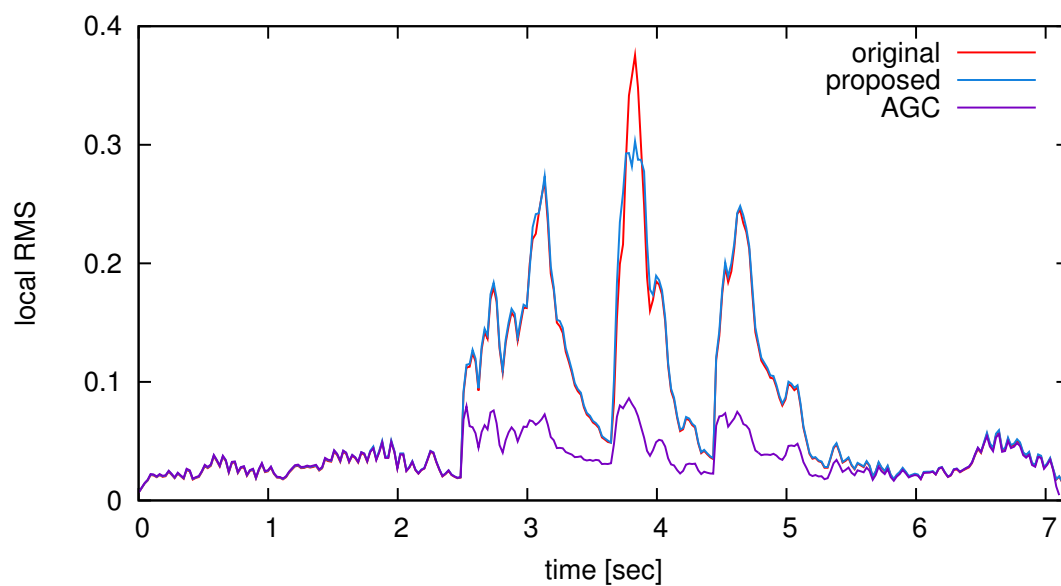


図 5.9: Local RMS of original, phase optimized and AGCed samples.

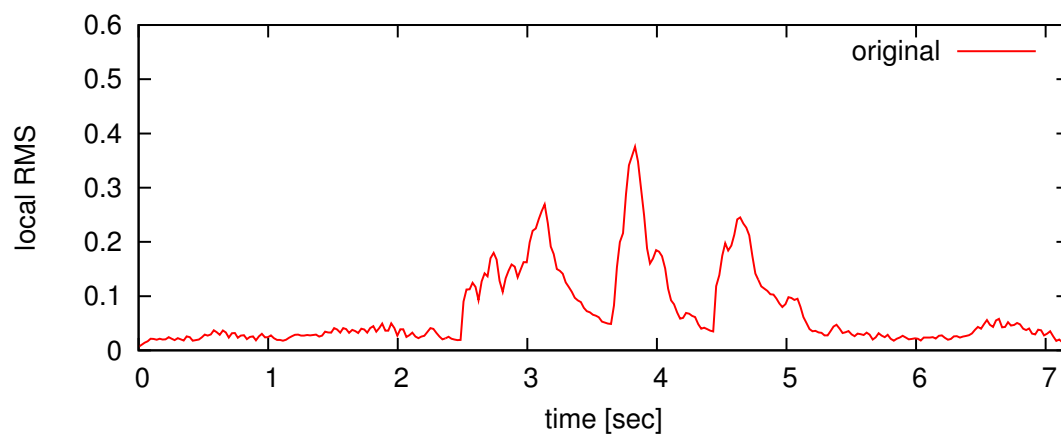


図 5.10: Local RMS of original sample.

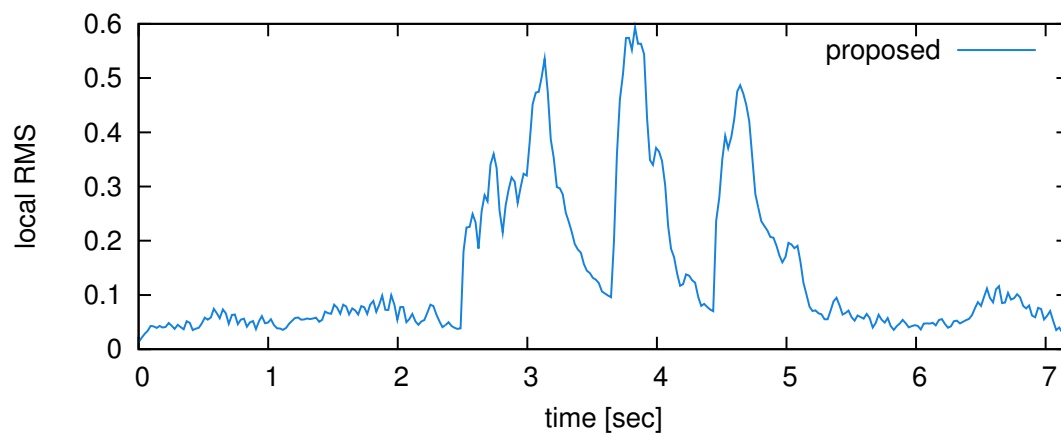


図 5.11: Local RMS of sample optimized by proposed method.

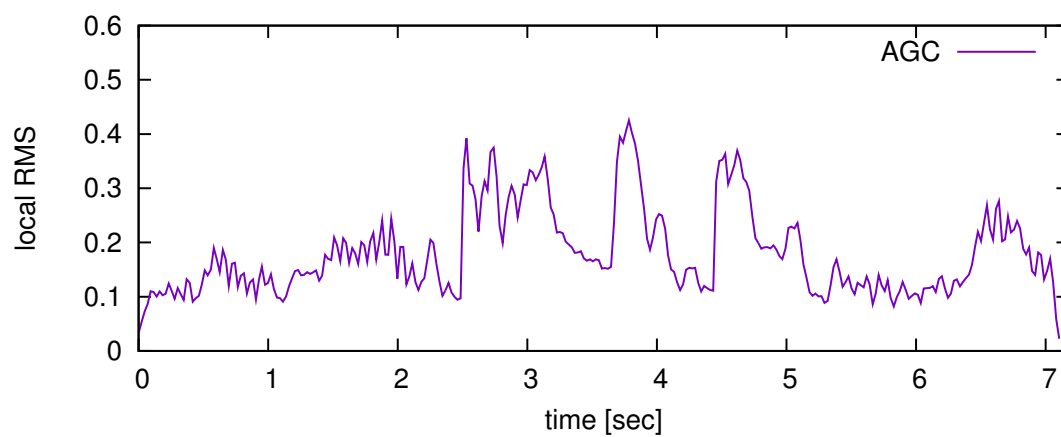


図 5.12: Local RMS of sample compressed by AGC.

は0.6程度まで上昇している。

従って、本論文で提案した手法はAGCと比較して、

- 元の音源の強弱の変化に忠実である
- ピーク付近のエネルギー密度を更に大きくできる

という点において優れていると言える。

5.5.3 リアルタイム性

AGCはその原理の単純さから処理が非常に軽く、現在の計算機の処理能力であれば多数の音源に対し、リアルタイムで適用することも十分に可能である。またハードウェアでの実装もアナログ回路・デジタル回路のどちらをもちいた場合でも容易に行うことができる。

一方で位相最適化の手法では、従来の位相最適化の手法と比較して格段に高速化したものの、何度もFFTを繰り返す必要があり、計算時間はどれだけエネルギー密度を向上させるかにもよるが、音源の長さの100倍程度の時間が必要である。そして繰り返し処理を含むため、アナログ回路での実現にも向かない。

したがって、本論文で提案した位相最適化の手法は、録音済の音源に対しオフライン処理を行うという運用に限定される。また、処理に時間がかかることから、処理された音を聞きながらパラメータを調整するということも難しい。

5.5.4 両者の使い分け

以上の特性を考慮すると、AGCと提案した位相最適化手法の効果的な使い分けは以下のようになる。

楽曲の仕上げより前の段階では従来通りAGCを用いて編集した後、楽曲の仕上げの段階において、目標とする値より数[dB]小さい値までAGCを用いてエネルギー密度を上げる。その後、提案した位相最適化手法を用いて目標とする値までエネルギー密度を向上させる。

以上のような運用により、AGCによって発生していた音質の変化を抑えながら、従来と同等、もしくは従来以上のエネルギー密度を実現することが可能となる。

第6章

結論

6.1 まとめ

本論文では、最急降下法を用いた高速な位相最適化手法と、位相の変化量を平滑化することにより、音質の劣化を最小限に留める手法を提案した。従来手法ではエネルギー密度の改善は最大でも3 [dB] 程度、音質の劣化も顕著であったが、これらの手法により、元音源と比べておよそ6 [dB] のエネルギー密度の改善を、知覚困難なレベルの音質劣化で実現することが可能となった。

これらの改良により、エネルギー密度の最大化を目的として位相を最適化する手法は、大きく実用レベルへと近づいたと考えられる。

6.2 今後の展望

本論文で提案した手法により、位相最適化の速度は格段に高速化された。しかしながら、AGCと比較すると処理は非常に重く、リアルタイムでの運用や、「少しずつパラメータを変化させながら音質と音量の最適点を探る」というような運用は行いづらい。本論文で提案した手法では、探索の部分は単純な最急降下法を用いるに留まっているが、一度の位相変化量を動的に変更したり、局所最適に陥りにくくする等の改良によってより高速化することが可能であると考えられる。また先日、高速フーリエ変換のさらなる高速化を実現するアルゴリズムが提案された [24]。そのアルゴリズムでは、長さ N の信号に対し、それが高々 k 個の成分しか持たない (k -sparse) 場合 $O(k \log N)$ でフーリエ変換を行うことができる。本論文の手法では非常に大きな N を用いているほか、微分の計算の際には、フーリエ変換される信号の特性上、非常に少ない k で k -sparse 性がよく成り立つことが期待できる。したがって [24] の手法を用いることにより、計算量を大きく削減することが期待できる。

また本論文で提案した音質劣化を最小限に留める制約条件は、今回聴取実験に利用した4つの音楽に関してはかなり有効に働いたが、予備実験の段階では、同一の制約条件をかけていた場合でもフレームの長さや、どれだけエネルギー密度を向上させるかといった条件によって音質劣化が発生したりしなかったりすることが確認された。本論文で提案した制約条件は、まだ位相と音色の関係を完全に捉えているとは言い難い。位相変化によって音質劣化の発生するより詳細な条件が解明されればよりよい制約条件を利用することができるようになると思われる。

本論文では4種類の音楽について音質の劣化の検証を行ったが、より多くの種類の音源に対し適用し効果を検証することが必要である。例えば、フルート独奏のような各時刻での周波数成分がほぼ単一の正弦波で構成されるような音源に対しては、位相最適化の手法は相性が悪いと考えられる。更に、編集された音楽データはその後様々な形式で圧縮されて利用されることが多く、それらの圧縮手法との相性の検証も必要である。

本論文ではモノラルの音源を対象としてプログラムを実装し実験を行ったが、実際の音源は殆どが2ch以上である。したがって、多チャンネル音源の位相を変更した場合の音質への影響を調査するとともに、そのような音源に対して位相の最適化を行えるよう拡張を行う必要がある。

謝辞

本研究を進めるにあたり、丁寧なご指導をいただいた、指導教官である広瀬啓吉教授ならびに峯松信明准教授に深く感謝いたします。

また、様々な面で研究活動を支えていただいた高橋登技官、秘書の池上恵氏、楠本由香里氏にも深く感謝いたします。

広瀬・峯松研究室の皆様には日頃より様々なご協力をいただきました。お忙しい中聴取実験に参加頂いた皆様のご協力により、無事に結果の評価を行うことができました。残念ながら全ての方を挙げることはできませんが、研究室の先輩、同期、後輩、OBの皆様のお陰で有意義な研究生生活を送ることができました。

ここに心よりお礼申し上げます。

2012年2月8日

砂田 宜宏

参考文献

- [1] 藍原郁 *et al.* 位相変化を用いた非線形歪低減法. 電子情報通信学会技術研究報告. EA, 応用音響, Vol. 103, No. 252, pp. 15–19, 2003.
- [2] 杉山勇司. レコーディング/ミキシングの全知識. リットーミュージック, 2004.
- [3] J. Lim, D. Wang. The unimportance of phase in speech enhancement. *IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-30*, pp. 679–681, 1982.
- [4] A. Oppenheim, J. Lim. The importance of phase in signals. *Proceedings of the IEEE*, Vol. 69, No. 5, pp. 529–541, 2005.
- [5] 赤木正人, 安武浩二郎. 時間方向情報の知覚の検討: 位相変化の音色知覚に及ぼす影響について. 電子情報通信学会技術研究報告. EA, 応用音響, Vol. 98, No. 157, pp. 15–22, 1998-06-29.
- [6] M. Schroeder, H. Strube. Flat-spectrum speech. *J. Acoust. Soc. Am*, Vol. 79, No. 5, pp. 1580–1583, 1986.
- [7] 大塚真司, 王龍標, 中川聖一. 位相情報と MFCC の併用による話者認識の高精度化. *spectrum*, Vol. 60, No. 700Hz, pp. 75–1, 2008.
- [8] 濱上知樹, 知樹. 音源波形形状を高調波位相により制御する音声合成方式. 日本音響学会誌, Vol. 54, No. 9, pp. 623–631, 1998.
- [9] M. Mathews, J. Miller, E. David Jr. Pitch synchronous analysis of voiced sounds. *The Journal of the Acoustical Society of America*, Vol. 33, p. 179, 1961.
- [10] N. Huang *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings: Mathematical, Physical and Engineering Sciences*, pp. 903–995, 1998.
- [11] 古井貞熙. デジタル音声処理. デジタル音声処理. 東海大学出版会, 1985.
- [12] R. Patterson. A pulse ribbon model of monaural phase perception. *Journal of the Acoustical Society of America*, 1987.
- [13] 赤木正人. 聴覚フィルタとそのモデル. 電子情報通信学会誌, Vol. 77, No. 9, pp. 948–956, 1994.

- [14] K. Paliwal, L. Alsteris. On the usefulness of STFT phase spectrum in human listening tests. *Speech Communication*, Vol. 45, No. 2, pp. 153–170, 2005.
- [15] E. Aiba, M. Tsuzaki. Perceptual judgment in synchronization of two complex tones: Relation to the cochlear delays. *Acoustical Science and Technology*, Vol. 28, No. 5, pp. 357–359, 2007.
- [16] 浜田大樹, 鷓木祐史. Ds-4-2 蝸牛遅延に基づいた電子音響透かし法の検討. 電子情報通信学会総合大会講演論文集, Vol. 2008, No. 2, 2008-03-05.
- [17] 宮内良太 *et al.* 蝸牛遅延特性に基づいた位相偏移変調が知覚不可能性に与える影響. 日本音響学会 2010 年秋季研究発表会公演論文集, pp. 511–514, 2010.
- [18] 中村尚五. ビギナーズ デジタルフィルタ. 東京電機大学出版局, 1989.
- [19] 後藤真孝 *et al.* 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. 情報処理学会 音楽情報科学研究会 研究報告, Vol. 2001, No. 103, pp. 35–42, 2001.
- [20] 後藤真孝 *et al.* Rwc 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース. 情報処理学会 音楽情報科学研究会 研究報告, Vol. 2002, No. 104, pp. 25–32, 2002.
- [21] M. Goto *et al.* Rwc music database: Popular, classical, and jazz music databases. In *Proc. ISMIR*, Vol. 2, pp. 287–288, 2002.
- [22] M. Goto *et al.* Development of the rwc music database. In *Proceedings of the 18th International Congress on Acoustics (ICA 2004)*, Vol. 1, pp. 553–556, 2004.
- [23] B. ITU. 1116, methods for the subjective assessment of small impairments in audio systems, including multi-channel sound systems. *International Telecommunication Union, Geneva*, 1998.
- [24] H. Hassanieh *et al.* Nearly optimal sparse fourier transform. *Arxiv preprint arXiv:1201.2501*, 2012.

発表文献

- [1] 砂田宜宏, 松田徹也, 広瀬啓吉, 峯松信明, “アンサンブル経験的モード分解における効果的なノイズ付加の検討”, 日本音響学会春季講演論文集, 2-Q-33, 2010-3
- [2] Nobuhiro Sunada, Tetsuya Matsuda, Kikichi Hirose, Nobuaki Minematsu, “Use of paired white noises with inverse polarity in ensemble empirical mode decomposition”, Proc. Int. Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP’2010), 2010-3 (Student Paper Award)
- [3] Nobuhiro Sunada, Tetsuya Matsuda, Kikichi Hirose, Nobuaki Minematsu, “Use of Paired White Noises with Inverse Phases in Ensemble Empirical Mode Decomposition”, Journal of Signal Processing, Vol.14, No.4, pp. 277-280, 2010-6.
- [4] 砂田宜宏, 清水信哉, 広瀬啓吉, 峯松信明, “音源のエネルギー密度向上を目的とした位相最適化”, 日本音響学会春季講演論文集, 2-Q-20, 2012-3 (発表予定)
- [5] Nobuhiro Sunada, Shimizu Shinya, Kikichi Hirose, Nobuaki Minematsu, “EFFICIENT PHASE OPTIMIZATION FOR SOUNDS WITH CONDENSED ENERGY DENSITY”, Proc. Int. Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP’2012), 2012-3 (発表予定)

付録 A

ローパスフィルタを用いた 制約条件の適用例

§4.6 で導入したローパスフィルタの畳み込みを用いた制約条件は、軸変換であると解釈できる。 x 、 y の 2 次元の場合で例を示す。制約として用いるローパスフィルタを

$$h_{lpf,0} = \alpha \quad (\text{A.1})$$

$$h_{lpf,1} = 1 - \alpha \quad (\text{A.2})$$

$$0.5 \leq \alpha \leq 1.0 \quad (\text{A.3})$$

とした場合、導入する変数を x' 、 y' として、

$$x = \alpha x' + (1 - \alpha)y' \quad (\text{A.4})$$

$$y = (1 - \alpha)x' + \alpha y' \quad (\text{A.5})$$

となる。 x 、 y の値が近い、つまり $y = x$ の線上近いほど値がなめらかであると言え、反対に $y = -x$ の線上に近ければなめらかでないと言える。したがって α が 0.5 に近い程カットオフ周波数が低く、1.0 に近い場合カットオフ周波数が高くなる。これを变形すると、

$$x' = \frac{-\alpha x + (1 - \alpha)y}{1 - 2\alpha} \quad (\text{A.6})$$

$$y' = \frac{-\alpha y + (1 - \alpha)x}{1 - 2\alpha} \quad (\text{A.7})$$

となる。最急降下法では、これら軸の方向に微小に値が変化した時に、最小化するべき関数の値がどれだけ変化するかを計算し、その値に比例するように各軸毎に値を変更する。したがって、 x' 軸、 y' 軸が $y = x$ に近ければ、探索結果も $y = x$ に近い値になりやすい。 $\alpha = 0.8$ 、 $\alpha = 0.6$ としてある関数に対して最急降下法を行った例を A.1 に示す。制約をかけずに最急降下法を行った場合を赤の線で、 $\alpha = 0.8$ として探索を行った場合の結果を緑の線で、 $\alpha = 0.6$ として探索を行った場合の結果を青の線で示している。 $\alpha = 0.5$ に近くなるにしたがって探索結果も $y = x$ に近い値となっていることがわかる。

ただし、最急降下法の繰り返し回数が非常に大きくなると $y = x$ の線から離れた値となってしまうことが予想されるため、適当な回数で繰り返しを打ち切る必要がある。

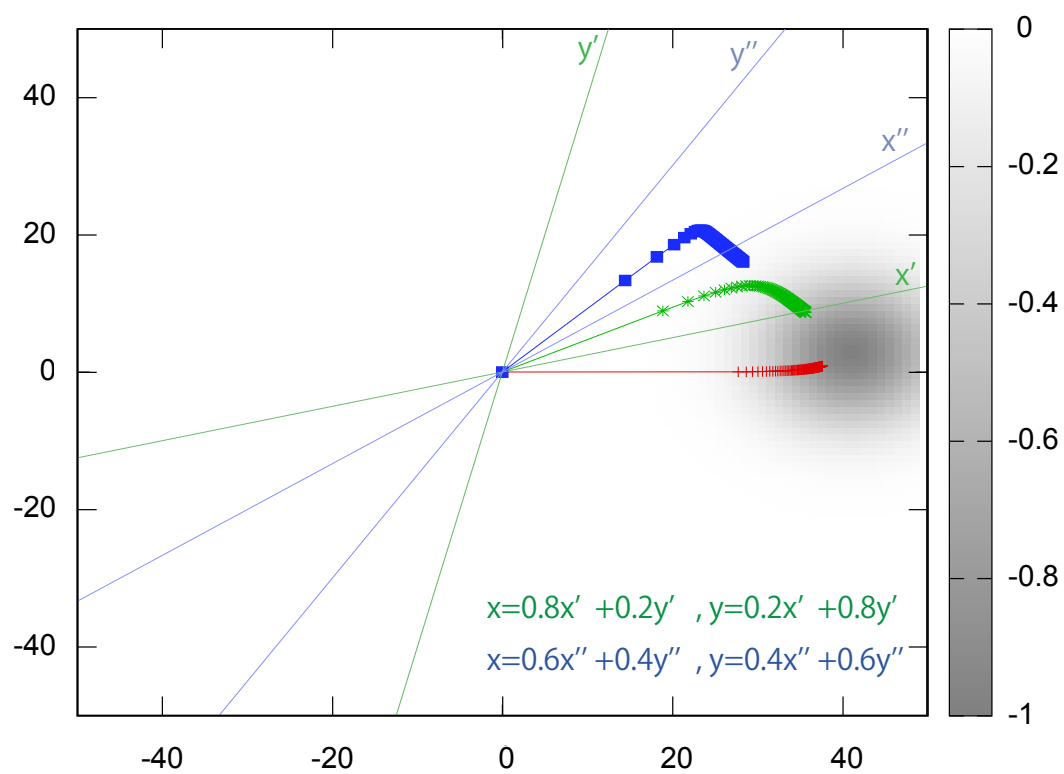


図 A.1: Steepest descent using modified axis.