

Title	位相変化に対する 音色知覚機構に関する研究
Author(s)	西澤, 雅人
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1342
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士



修 士 論 文

**位相変化に対する音色
知覚機構に関する研究**

指導教官 赤木正人 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

西澤雅人

2000 年 2 月 15 日

要 旨

従来より「耳は位相に聾である．」とされ、音声符号化、認識、合成などほぼすべての音声研究において専ら振幅スペクトル形状に焦点があてられてきた．しかし、少数ながら、ある条件下では位相変化による音色変化が知覚されるという報告もあり、昨今、位相制御によってより高品質な音、自然な合成音が実現できることも期待されている．

そこで本稿では、過去に報告されている実験において注目されていなかった実験系による位相歪みを重視し、これを可能な限り除き、位相変化を忠実に再現させることを試みた上で、位相特性の異なる調波複合音の基本周波数を様々に変化させて心理物理実験を行い、特に位相変化による音色知覚と基本周波数の関係について調べた．更に、聴神経の発火パターンを模擬した Auditory Image Modl(AIM) においてシミュレーションを行い、その結果と実験結果と比較し、聴覚機構における位相知覚特性について考察を行った．

目次

1	はじめに	1
1.1	本研究の背景	1
1.1.1	位相に関する研究の歴史的背景	1
1.1.2	実験機器と群遅延	3
1.1.3	位相知覚のモデル	4
1.1.4	安武らの実験	4
1.2	本研究の目的, 特色	4
2	実験系の位相補正について	6
2.1	実験機器について	6
2.2	実験系の特性	10
2.2.1	インパルス応答	10
2.2.2	TSP 法	10
2.2.3	TSP 法による実験系の測定結果	13
2.3	位相補正の手順	16
2.4	位相補正の性能評価	18
3	実験 1	22
3.1	目的	22
3.2	刺激音	23
3.3	実験条件	24
3.4	結果	24
3.5	考察	26
4	実験 2	27
4.1	目的	27

4.2	刺激音	28
4.3	実験条件	28
4.4	結果	29
4.5	考察	30
5	実験 3	31
5.1	目的	31
5.2	刺激音	31
5.3	実験条件	32
5.4	結果	33
5.5	考察	37
5.5.1	実験結果について	37
5.5.2	実験データの分析	38
6	Auditory Image Model(AIM)[10] によるシミュレーション	39
6.1	位相変化による Neural Activity Pattern(NAP) の変化	39
6.2	Hiss 相関	45
7	おわりに	49
7.1	まとめ	49

第 1 章

はじめに

音色に関する研究として位相情報が扱われることは少ない。これは長い間「人間は位相に対して聾である。」という通説が信じられているためである。実際、音声の音韻などの知覚においては、位相はほとんど影響を与えないため、音声の符号化、認識、合成などの研究は主に振幅スペクトルの形状との関連においてなされ、位相については余り考慮されていない。しかし、位相が音色知覚に影響を及ぼすことは幾つか報告されており ([3][4][5][6][8][11][16][17])、位相と音色知覚の関連を明確にすることは聴覚系の時間情報処理、音色の知覚機構を検討する上で重要視されている。

本研究では、実験系による位相歪みの影響を厳密に補正した上で、様々な基本周波数の調波複合音を用いて心理物理実験を行なう。そして、調波複合音の高調波成分の位相を変化させた場合に音色に及ぼす影響について調べる。その上で、聴神経の発火パターンを模擬したモデルであるパルスリボンモデル [8] の妥当性について検証する。更に、パルスリボンモデルを発展させた Auditory Image Model(AIM)[10] によるシミュレーションを行い、聴覚機構における位相知覚特性について考察する。

1.1 本研究の背景

1.1.1 位相に関する研究の歴史的背景

位相変化による音色の変化について最初に研究したのはHelmholtzである [1]。Helmholtzは、独自の方法で位相が可変な基本周波数 120Hz と 240Hz の 8 成分調波複合音を生成し、様々な位相特性の調波複合音について実験を行った。その結果、Helmholtz は次のように結論づけた。(1) 位相変化による複合音の音色の変化は小さく、ある母音が他の母音に聞

こえる程ではない。(2)6番目から8番目の成分では不協和音やうなりが生じた。高次の高調波成分については位相の影響がある。ところが、一般的な見解としては、Helmholtzの結果から、音色は位相に依存しないと単純化され、後に「耳は位相に聾である。」(The ear is deaf to phase)という表現を生み、これが通説となってしまった。

Helmholtzの後Koningはサイレンの小穴の配置を変化させて位相変化による音色の変化を調べた。その結果、明らかに音色は振幅とはかかわりなく、位相の影響を受けるとしている。しかし、この結果はHermannにより批判された。Hermannはサイレンの小穴によって正確に空気の流れが調整されるわけではなく、したがって、サイレンは位相の役割についての研究には適切ではないとしている。今日では、位相の影響は小さいため、サイレンを用いた研究では問題は解決されないとされている。

20世紀に入り、音声の研究には電機音響機器が導入されるようになった。このころFletcher[2]は音色に主に関係するのは倍音構造であり、倍音構造は周波数、強度、位相で特徴付けられるが、強度が高い場合を除いて位相の変化が知覚に与える影響はわずかであるとした。実際、アンプなどの電機音響機器には位相歪みがあったが、再現される音楽や音声の品質にほとんど影響を与えないので、一般的には位相のずれは無視できるとされていた。

しかし、Mathes and Millerは3つの純音からなる複合音を用いた実験[3]を行い、同一スペクトルで成分音の位相関係により包絡線(波形)が違う場合、音色が変わることを見出している。

Lickliderは16成分調波複合音を用いた実験[4]で、(1)高域成分の位相を変化させる方が低域成分の位相を変化させるよりも音色の変化が知覚しやすく(2)基本周波数は低い方が音色の判別をしやすいことを報告している。彼はこの論文でThe auditory system is by no means “phase deaf”と明確に述べている。

また、Shroederは31成分調波複合音を用いた同様な実験[5]を行い、上述のMathes and Millerの結論のとおり、音色は刺激波形の尖鋭度(peak factor)に強く依存し、位相の変化によって波形の尖鋭度が変化することにより音色に変化が現れると結論づけた。

Plompらは、各高調波成分の位相を交互に $\pi/2$ シフトさせた10成分調波複合音を用いて実験[6]を行い、その結果、このような音は全て位相が同じ調波複合音とは音色が異なっていた。さらにPlompらは位相をシフトさせる量を $\pi/16$ にかえて実験を行い次のようにまとめている。位相の変化による調波複合音の音色の違いは基本周波数が低い方が影響は大きく、150Hzを越えると影響が小さくなる。

また、Pattersonは31成分調波複合音の各成分の位相特性が様々である刺激音を用いて音色比較実験を行なった[8]。その結果、各成分の位相が $\cos$ 位相である刺激音とランダ

ムに変化する刺激音の比較の場合、基本周波数が 400Hz 以上であれば聴覚は、位相の変化を聞き取れないが、200Hz 以下であれば最初の 12 成分からなる刺激であっても判別でき、また、奇数成分と偶数成分の位相が異なった刺激音の比較においては判別能力は基本周波数に大きく依存するということが明らかになった。

以上の報告において共通している点は

- 1 基本周波数は低い方が音色の違いを判別しやすい。
- 2 低域成分よりも高域成分において位相の変化による音色の変化が大きい

ということである。

1.1.2 実験機器と群遅延

群遅延特性は位相歪みを測る一つの尺度である。位相に関する実験において実験機器の位相歪みにまで、触れられているものは少ない。しかし、電気音響機器導入以降、スピーカーからの放射音で実験を行った場合、電気音響変換部での歪みや、部屋の残響によって各周波数成分の位相は乱されてしまう。また、ヘッドホンで実験を行った場合でもアンプ等の特性により位相は変化する。

Blauert らは電気音響機器による群遅延がどのように知覚に影響するかを調べた。群遅延を系統的に変化させた all-pass フィルタを音声、音楽、雑音、調波複合音パルス列に作用させ、音色変化の検知限を測定した。[7] その結果 7kHz 付近のみ群遅延特性を変化させた場合、群遅延が 1.1ms で違いが知覚できる。また、ヘッドホン (Sennheiser HD414) の群遅延特性に似せた位相特性を制御した場合群遅延のピークと谷の差が 0.5ms で知覚できる、という結果が得られた。

また、河原らは時間的に非対称な群遅延のばらつきを持つ信号とこれに対して時間軸を反転した信号を組み合わせることで音色の検知限を測定した [15]。河原らは非対称性の大きさの制御は群遅延の標準偏差を制御することにより行っている。これにおいて、群遅延の標準偏差が適切ならば、2 つの音には明確な音色の違いが存在し、群遅延の標準偏差が 1.0ms 以上ではほぼ 100%、0.5ms でも 75% 以上違いが分かるという結果が得られている。

これらの実験結果より、位相に関する実験を行う場合実験機器の位相歪みは無視できないと考えられる。

1.1.3 位相知覚のモデル

Patterson[8] は、位相変化の知覚モデルとして、聴覚神経での神経発火を模擬したパルス列を用いたモデルを提案した。このモデルは基底膜を模擬したフィルタバンクにより周波数分析された波形を基にパルスを生成して生成されたパルス列 (パルスリボン) の様相の違いにより、知覚特性を説明しようとしたものである。例えば、1つのフィルタを複数の周波数成分が通過すると、各周波数成分の位相の違いによりパルス間隔が変化することを根拠に説明を試みている。

また、Patterson はパルスリボンモデルを発展させ、新たに Auditory Image Model (AIM) [10] を用いて、時間反転音の知覚特性の説明を試みている。時間反転音は位相スペクトルの符号を反転したものであり、位相のみが異なっている。

本研究においても聴覚の位相知覚特性を検討する際、AIM を用いたシミュレーションを行う。

1.1.4 安武らの実験

1.1.2 で述べたように、実験機器には位相歪みがある。しかし、位相に関する研究においてこれについて言及しているものは少ない。

これに対し、安武らは実験系の特性を測定し実験機器による位相歪みを可能な限り、補正した上で基本周波数が 120Hz と 250Hz の 31 成分調波複合音を用いた実験を行った [11]。実験系の位相歪みを補正することにより、刺激音の位相特性が正確に再現され、実験の信頼性の向上を図った。この実験の結果について安武らは 1.1.3 に述べたパルスリボンモデルの妥当性について議論した。安武らは実験結果から、聴覚フィルタを複数の周波数成分通過するとすると各周波数成分の位相の違いによりパルス間隔が変化し差が検出できるとするパルスリボンモデルに対し、聴覚フィルタ内に高調波成分が 2-3 本以上含まれていれば音色の違いが知覚できる可能性を示唆している。

1.2 本研究の目的、特色

従来の位相に関する研究における問題点を指摘する。Helmholtz や Fletcher の時代に行われた実験は、電気音響機器導入以前のものであったり、音場への放射音が使われていたりして、反射などにより位相の変化がかき消され、音色の差が知覚されない可能性が高い。ヘッドホンを使用した実験であっても実験系により生じる群遅延歪みなどは無視している場合が多く、使用したヘッドホンの諸特性も明示されていない場合もある。このようにし

て得られた実験結果の頑健性は明らかでない。

さらに、位相差による音色の違いについて、聴覚における時間情報処理や音色の知覚特性についての系統的な検討はほとんどなされていないのが現状である。安武らの実験 [11] ではこれらの問題に対して、実験系の位相歪みの補正を行っているが、その精度に対しては議論の余地がある。また、聴神経の発火パターンを模擬したモデルであるパルスリボンモデルの妥当性の検証を試みているが、十分な説明が出来るものではない。

そこで、本研究では安武らの研究と同様に実験系の位相歪みを厳密に調べ、より精度を求めた位相補正について検討する。その上で、種々の心理物理実験を行うことによって、位相変化による音色知覚特性について調べる。更に、聴神経の発火パターンを模擬したモデルであるパルスリボンモデルを発展させた、AuditoryImage Model(AIM) においてシミュレーションを行い、実験結果と比較し聴覚系における位相知覚機構について考察する。

第 2 章

実験系の位相補正について

2.1 実験機器について

位相歪みの尺度である群遅延特性は極めて不安定であり，常に微細な変動がある．実験系の歪みを可能な限り補正する為には比較的安定な特性であり，実験に最適であると考えられる機器を選択する必要がある．次に本研究で使用する機器について説明する．

ヘッドホン

平原 [12] によると聴覚実験に理想的なヘッドホンとは，実験者の意図するとおりの刺激音を被験者に呈示できるものであり，その基本的な条件として

- 周波数特性が平坦であること
- 歪み率が低いこと
- 群遅延特性が平坦であること
- ダイナミックレンジが広いこと
- クロストークが少ないこと

があげられ，また，実験の再現性を保証する為に必要な条件として

- ヘッドホンの個体差が少ないこと
- 装着状態による特性変化が少ないこと

- 物理的，電氣的に堅牢であること

があげられるとしている．しかし，現実的にはこのような条件を全て満たすヘッドホンは無い．特に可聴帯域 (20Hz～20kHz) 全域にわたって平坦な周波数特性を持つヘッドホンは無い．従ってヘッドホンの特性を承知した上で使用することになる．通常，ヘッドホンの周波数特性はカップラにヘッドホンを装着して測定するが，実耳装着時の特性との差は問題になっている．そこで，本研究では実用帯域の周波数特性が比較的平坦であり，実耳レスポンスとカップラレスポンスの差が少ないとされている，ヘッドホン，Lambda Nova (STAX) を使用した．

- Lambda Nova (STAX), SRM-1/MK-2 P.P(STAX)

Lambda Nova はコンデンサ型¹・耳覆い式・開放型のヘッドホンで，元来は音楽観賞用であるが聴覚実験にも用いられている．成極電圧は 580V/DC，耳当ては実耳装着性を向上させるために平板ではない．ヘッドホンアンプとして，音量調整用ボリュームをステップアッテネータに交換した SRM-1/MK-2 P.P. を使用する．

カップラ

- B&K4153

IEC318 に規定される耳覆い型イヤホンの広い周波数範囲の校正用カップラで，使用周波数範囲は 20Hz～10kHz．

マイク，マイクロホンプリアンプ，マイクロホンパワーサプライ

- B&K4134

1/2 インチの音圧型コンデンサマイクロホン²で，拡散音場あるいは密閉空洞およびカップラ内の測定用として設計されている．周波数範囲 4Hz～20kHz(± 2 dB)，ダイナミックレンジ 21～160dB，成極電圧 200V/DC．

- B&K2669B

マイクロホン前置増幅器．1/2 インチコンデンサマイクロホンに使用．

¹ 薄い振動体 (膜) に直流電圧を加えておき，それを挟んだ 2 枚の電極に音楽信号を供給することで起きる静電気で，振動体 (膜) を駆動する．

² 振動膜とこれに対抗して配置された固定電極で形成され，音圧に応じて振動膜が振動することによる両者の間の静電容量の変化により電気出力信号を取り出す．構造が簡単で，安定性が高いことから音響計測に適している．

- B&K5935

バッテリー動作2チャンネルマイクロホン電源. 0~50dB を 10dB ステップでゲイン選択. Lin./A 荷重.

DATリンク

- DAT-Link+ (T.C.T.)

計算機と各種音響機器とを接続するデジタルオーディオインタフェース. デジタルデータの転送のほかに, DATデッキをフロントエンドとして用いることで, アナログデータの計算機への入力も可能である.

DAコンバーター

- DAC-TALENT BD (STAX)

2組のバッテリーを持ち, 一方がDAプロセッサに電源を供給している間, もう一方のバッテリーが充電を行なうため, AC と DC を完全に切り離すことができる.

DAT

- TCD-D10PRO2 (SONY)

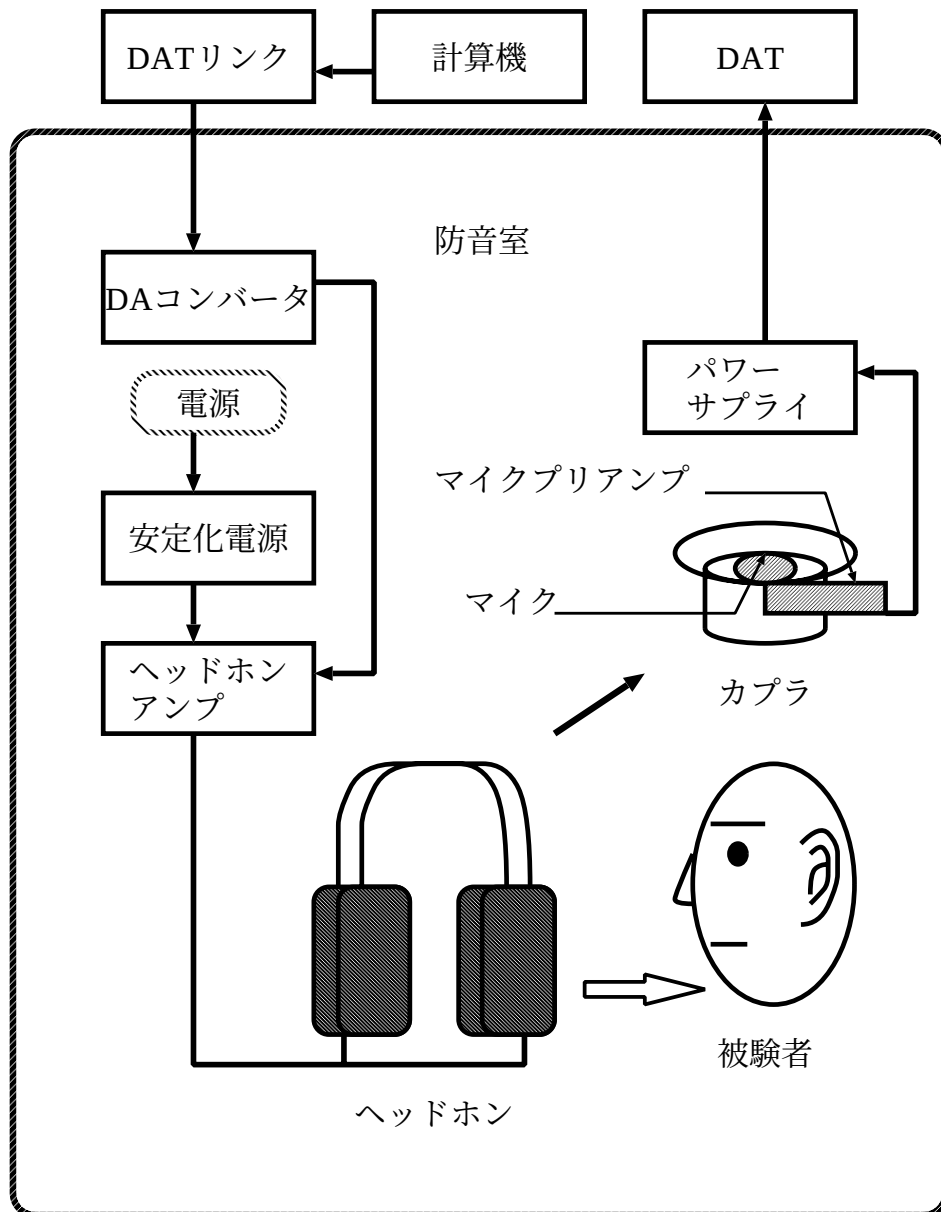
ポータブルタイプのデジタルオーディオレコーダー. DC 駆動のため, 他の機器からノイズの影響を受けない.

安定化電源

- HYPERSINE REGULATOR HSR-510(SINANO)

音響映像機器専用電源で, 外来より混入するノイズや波形歪み, 負荷機器から発生する高周波ノイズ等も瞬時波形制御方式の高速波形コントロールにより, 常に最適なサイン波を負荷機器に供給する. また, 出力高安定化回路により, 電源入力に不安定な電圧や周波数が入力されても, 出力には安定した電力を供給する.

図 2.1: 実験系



防音室

本研究では，物理特性の測定および心理物理実験に，暗騒音約 23dB(A) の防音室を使用した．

図 2.1 に，以上の機器により構成される本研究の実験系の概略を示す．

2.2 実験系の特性

2.2.1 インパルス応答

実験系を一つのシステムとして考え伝達関数を求め，種々の特性を測定する．そのためインパルス応答の測定が必要になる．インパルス応答の導出には直接法や Time stretched pulse(TSP) 法，M 系列法等が知られている．直接法は持続時間の短い矩形波などの疑似パルスを入力信号とする方法で測定結果がそのままインパルス応答となる．しかし，音響系の伝達関数は一般にダイナミックレンジが広く，インパルス応答が長いという特徴を持ち，また，音響系の伝達関数を測定する場にはかなりの暗騒音が存在することが多い．そのため，パルスのようなエネルギーの小さい信号では測定誤差が大きくなり，十分な SN 比をとるためには，同期加算の回数が非現実的な大きさになってしまう．

これに対して M 系列法や，TSP 法ではエネルギーの大きい M 系列や TSP 信号が入力信号として利用される．この場合，測定結果と，入力信号を時間軸上で反転したものを巡回的にたたみこむことでインパルス応答が得られる．

本研究では TSP 法を改良した OATSP 法(以下 TSP 法)[13] を用いて実験系のインパルス応答を求める．

2.2.2 TSP 法

OATSP(以下 TSP) は周波数領域で設計され，それを IDFT することで時間関数としての信号となる．TSP は式 (2.1) のように定義される．

$$H(n) = \begin{cases} a_0 \exp(j4m\pi n^2/N^2) & (0 \leq n \leq N/2) \\ H^*(N - n) & (N/2 + 1 \leq n \leq N) \end{cases} \quad (2.1)$$

* : 複素共役

a_0 : 振幅

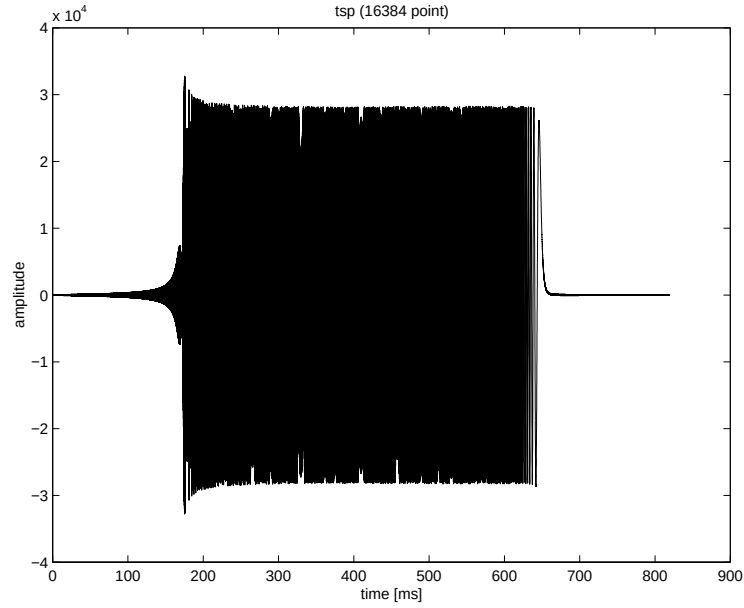


図 2.2: TSP 信号:16384 点

N : データ長 ($N = 2^i$)

m : 時間引き伸ばしパラメータ ($m \leq N/2$)

このとき, ITSP は式 (2.2) のようになる.

$$H^{-1}(n) = \begin{cases} a_0 \exp(-j4m\pi n^2/N^2) & (0 \leq n \leq N/2) \\ H^*(N-n) & (N/2 + 1 \leq n \leq N) \end{cases} \quad (2.2)$$

TSP の応答と ITSP を畳み込むことにより, 時間引き延ばしが逆に圧縮されて, 被測定系のインパルス応答が求められる. また, 周波数応答はインパルス応答を DFT することで求まる. サンプルング周波数 20kHz とし, TSP の各パラメータを $N=16384, m=4800, a_0 = 32760$ として作成した, TSP 信号と ITSP 信号を図 2.2, 2.3 に示す. これらをたたみこむと, ほぼきれいなインパルス応答になることがわかる. (図 2.4)

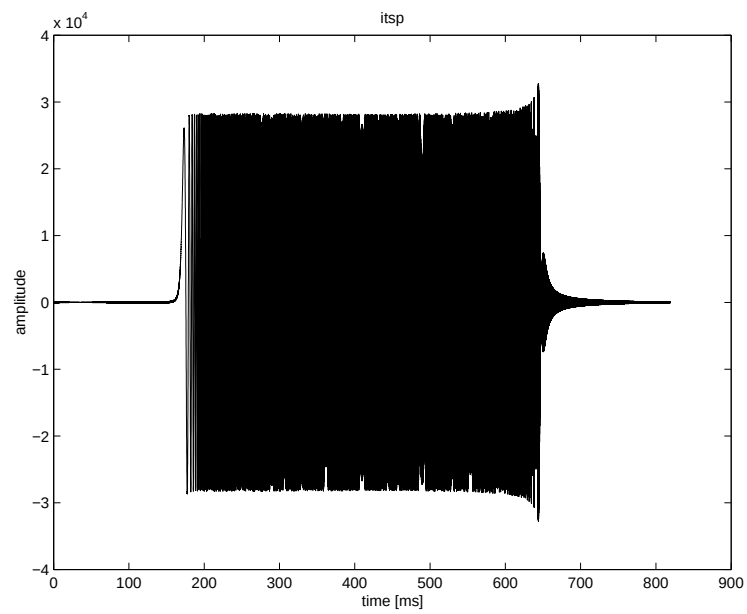


図 2.3: ITSP 信号:16384 点

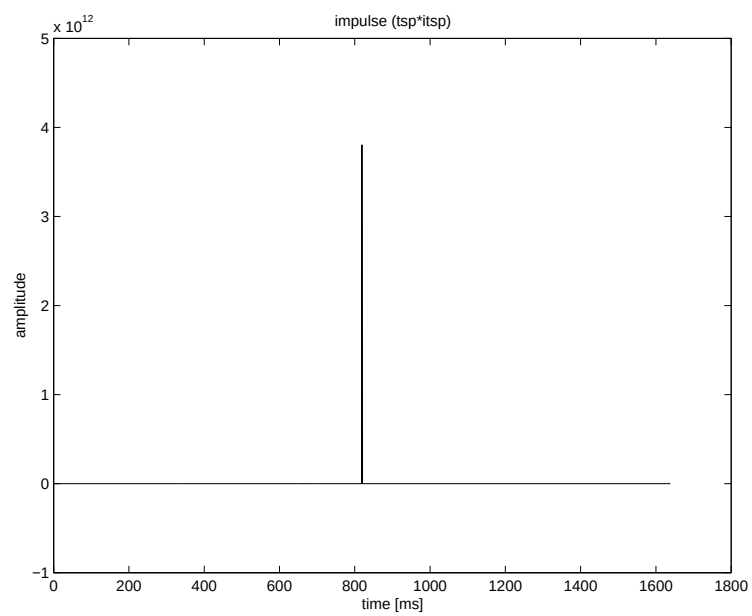


図 2.4: TSP と ITSP を畳み込んだ図

2.2.3 TSP 法による実験系の測定結果

実験系の測定を行うにあたり、考慮すべき点として温度、湿度などの環境条件の影響、AC 電源の影響などが挙げられる。このなかでも、特に AC 電源は群遅延特性に与える影響が大きい。そのためアナログ部においては、ヘッドホンアンプ以外の機器は全てバッテリー駆動で使用し、ヘッドホンアンプの電源は、安定化電源を使用した。

実験系の特性の測定は具体的には、次のように行う。まず、実験系に TSP 信号を入力し最終的にヘッドホンからでる音がカップラ内に設置されたマイクロホンにより収音される。その音を DAT に記録し、DAT から計算機にデータとして取り込まれる。これが TSP の時間応答波形として取り扱われる。

図 2.5 に実験系に入力した TSP の時間応答波形を示す。位相歪みなどにより、図 2.2 に示される元の波形とはかなり異なっていることがわかる。この、TSP 時間応答波形と前述の ITSP とをたたみこみインパルス応答を求め、これを DFT して、振幅特性と群遅延特性を求める。これらを図 2.8、図 2.9 に示す。これらの図から、振幅特性は低域で安定せず 4.5kHz 付近と 7~8kHz 付近において大きなディップが現れていることがわかる。また、群遅延特性をみると、やはり低域での変動が大きく、同じように 4.5kHz 付近、7~8kHz 付近においてディップが現れる。更に、全体的に数百 μsec 程の微細な変動があることがわかる。この群遅延の変動量は 1.1.2 でも述べられているように、音色の差として現れてくる可能性がある。

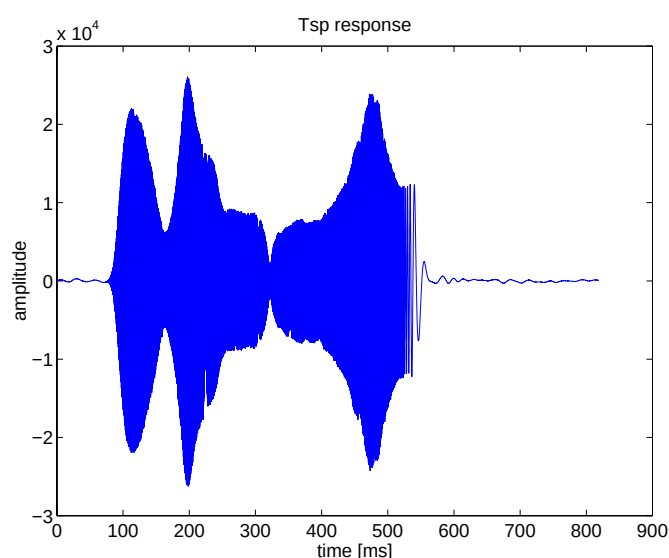


図 2.5: TSP 時間応答波形

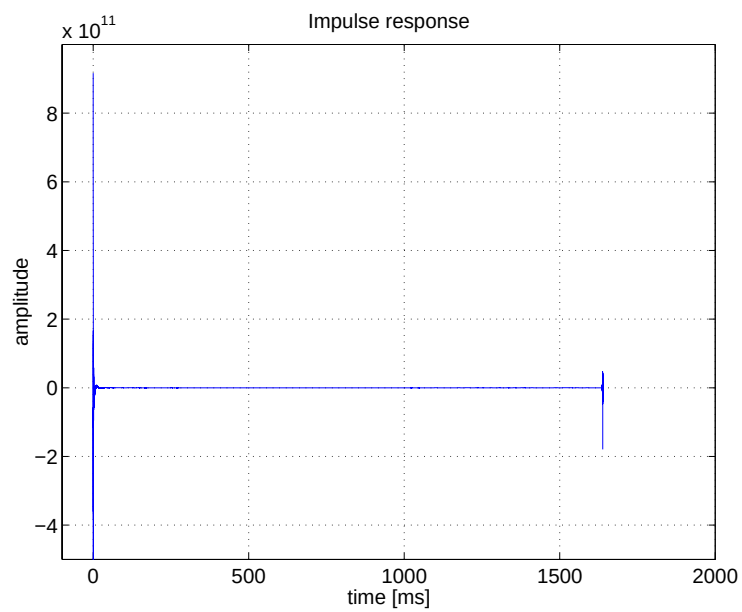


図 2.6: インパルス応答:8 回同期加算平均して最大値を 0 にシフト

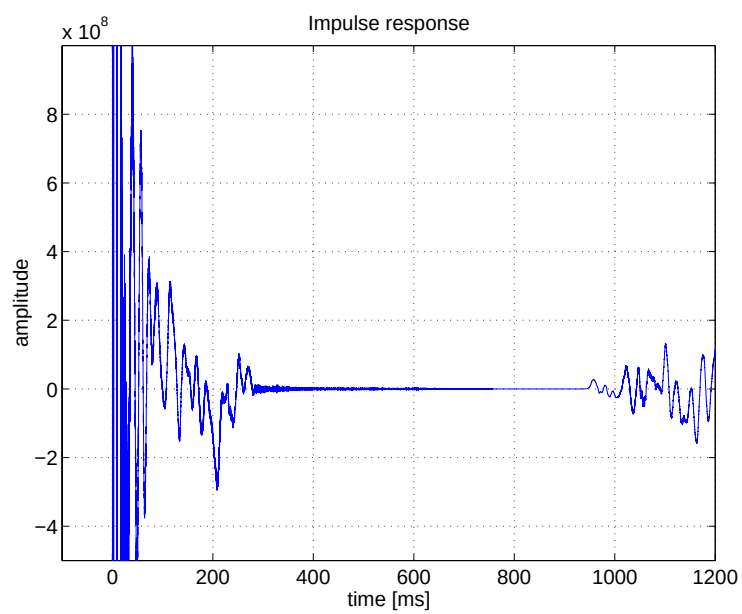


図 2.7: インパルス応答:拡大図

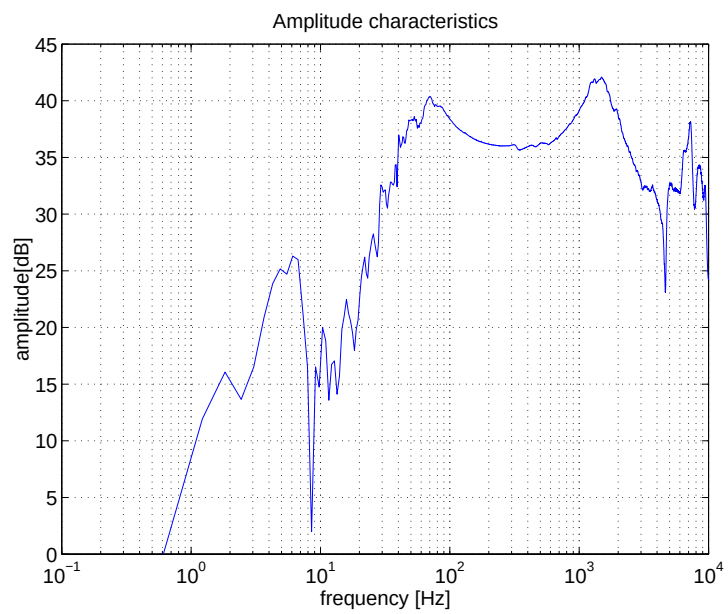


图 2.8: 振幅特性

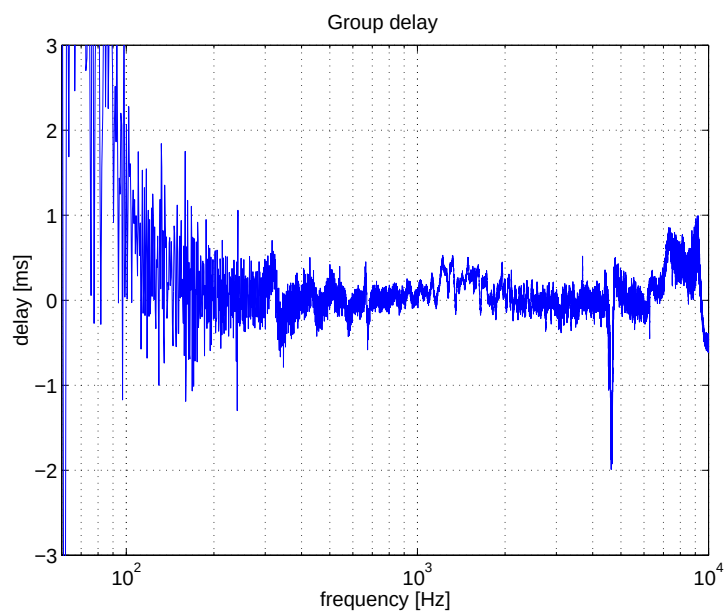


图 2.9: 群延迟特性

2.3 位相補正の手順

TSP 法により実験系の TSP 時間応答波形、振幅特性、群遅延特性などを調べた結果、精度の高い実験を行うためには位相補正が必要であることがわかった。位相補正は実験系の逆特性をもつ逆フィルタを求め、これを予め刺激音にかけておき、刺激音が実験系に入力された場合に実験系の特性と相殺することによって行う。

具体的な位相補正の手順を図 2.10 に示す。まず、最初に TSP 信号を実験系に入力し、TSP 時間応答を得てこれと、ITSP を畳み込むことによって、インパルス応答を計算する。求めたインパルス応答を DFT して周波数特性、群遅延特性を求める。周波数特性の逆数を計算し、これが逆フィルタになる。この逆フィルタと波形データを周波数領域上で乗算し、逆 DFT したものが刺激音になる。

ただし、群遅延特性に関しては微細な変動があるため、そのままこの方法を適用しても完全な補正は困難である。しかし、群遅延の微細構造は波形の形状との相関が低いことから、聴覚上それほど重要でなく、群遅延の概形が特に重要であるという報告もある(坂野ら [14]) ことから、本研究においても、まずは、群遅延特性の概形を平坦に補正することにした。従って、これからは群遅延特性を平滑化した概形から群遅延特性の安定性を判断する。図 2.9 に示される群遅延特性を平滑化し、概形をとりだしたものを図 2.11 に示す。

逆フィルタ

逆フィルタは周波数特性の逆数を計算することにより求められる。しかし、上述のようにそのまま周波数特性の逆数を逆フィルタとして用いても、微細な群遅延特性の変動は測定毎に変化するので、完全に補正することは出来ない。よって群遅延特性を平滑化してその概形が補正されるように逆フィルタを作成する。群遅延特性は位相を周波数領域で微分したものであるので、群遅延を積分することで位相は求められる。

従って、逆フィルタの位相特性は測定された周波数特性の逆数そのものの位相特性ではなく、平滑化した群遅延特性の逆数を取り、積分して求めた位相特性を持たせる。ただし、離散時間関係で処理するため実際には積分を累積和で近似して位相を導出する。また、低域では群遅延の変動が大きすぎて補正は困難である。従って、補正帯域を制限し、補正帯域外では定常になるように逆フィルタを補間した。図 2.12、図 2.13 に逆フィルタの振幅特性、群遅延特性を示す。

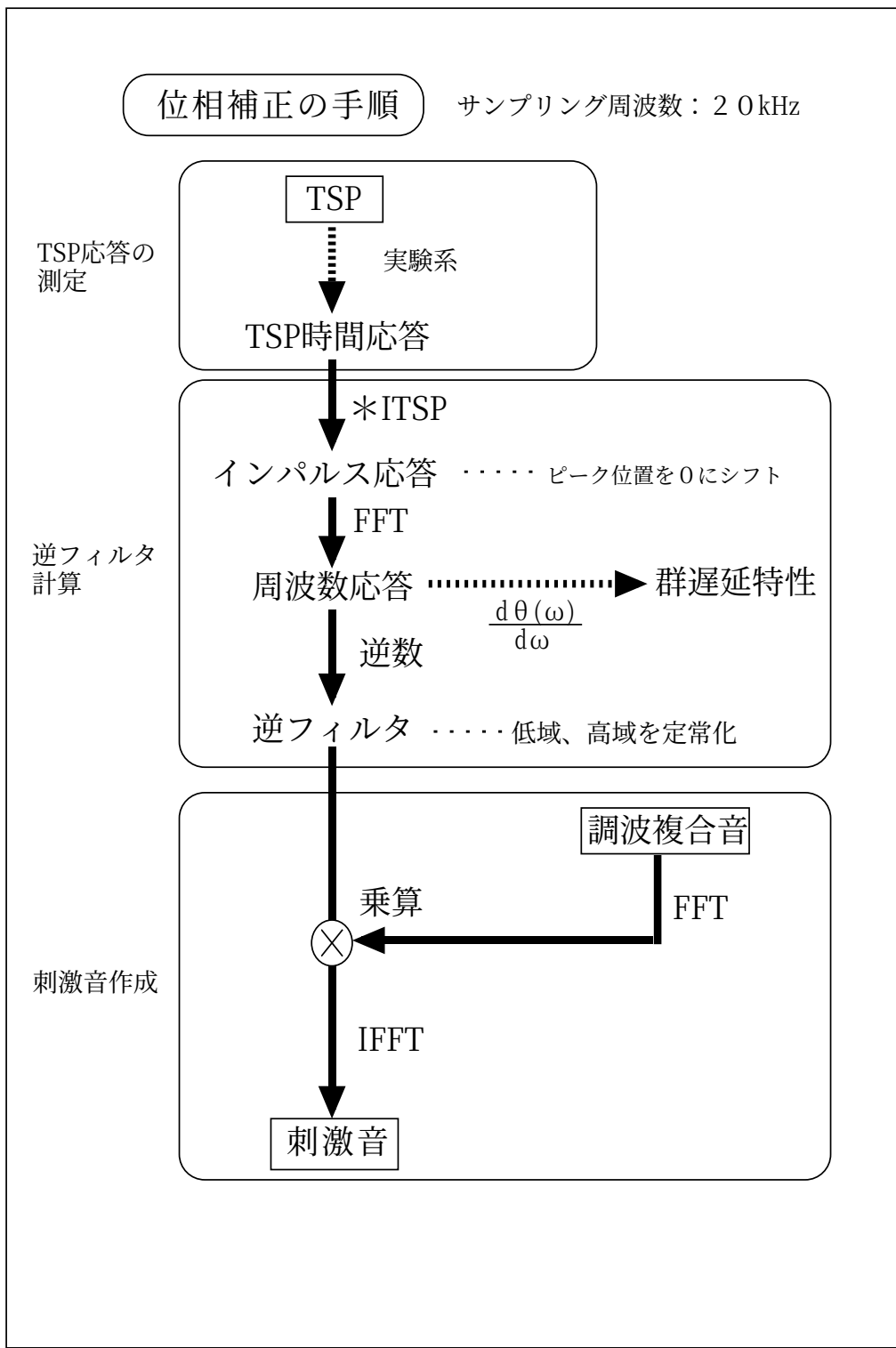


図 2.10: 補正手順

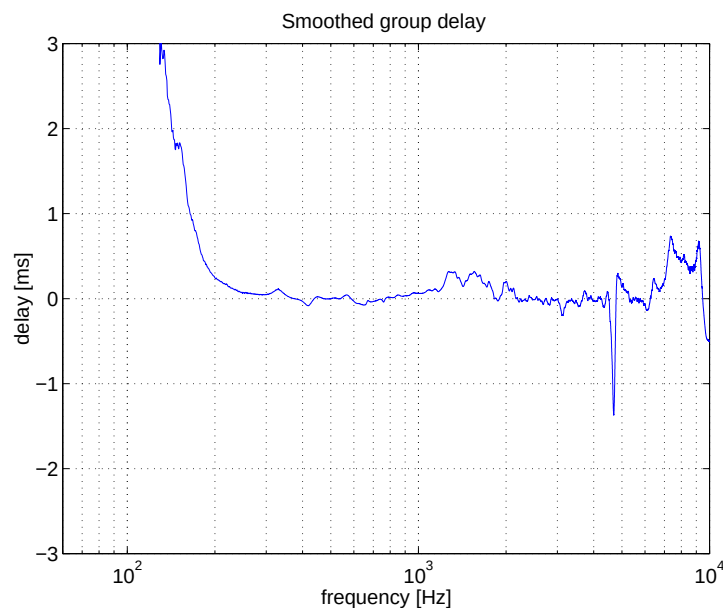


図 2.11: 群遅延特性の概形

2.4 位相補正の性能評価

以上の方法による位相補正の性能の評価を行った．逆フィルタの補正帯域は 300Hz～9300Hz としている．

求めた逆フィルタと TSP 信号を周波数軸上で乗算し，逆 DFT することによって得られた波形 (図 2.14) を実験系に入力して，時間応答波形，周波数特性，群遅延特性を測定した．その図を図 2.15，図 2.16，図 2.17 にそれぞれ示す．最初に時間応答波形をみると，逆フィルタをかけていない TSP 時間応答波形 (図 2.5) と比べると，図 2.15 は原波形 (図 2.2) に近い波形になっており補正が有効であることが伺える．

振幅特性においても逆フィルタをかけていない TSP 信号の振幅特性 (図 2.8) と比較すると補正帯域では平坦な特性が得られている．群遅延特性においても同様に，図 2.11 と比較すると群遅延特性の概形を，比較的平坦にすることに成功しており，逆フィルタによる補正が有効であることが，周波数領域，時間領域の両方で確認できた．

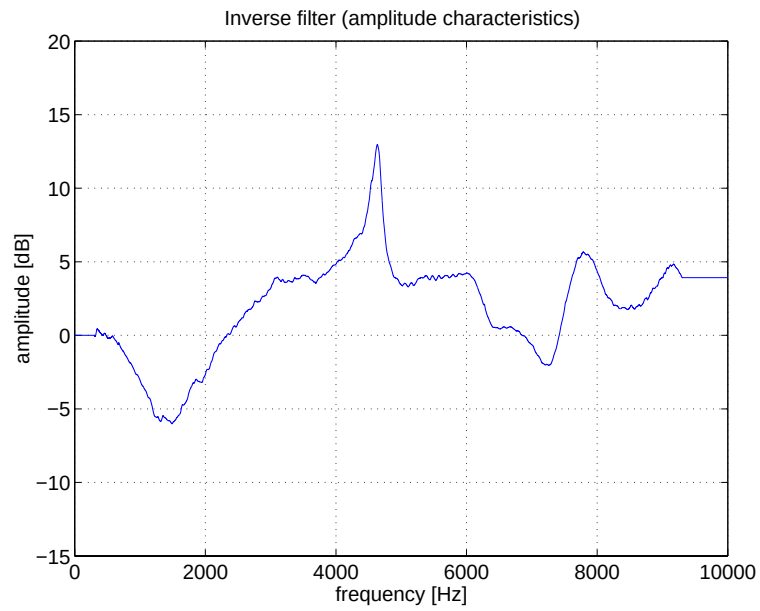


図 2.12: 逆フィルタ:振幅特性

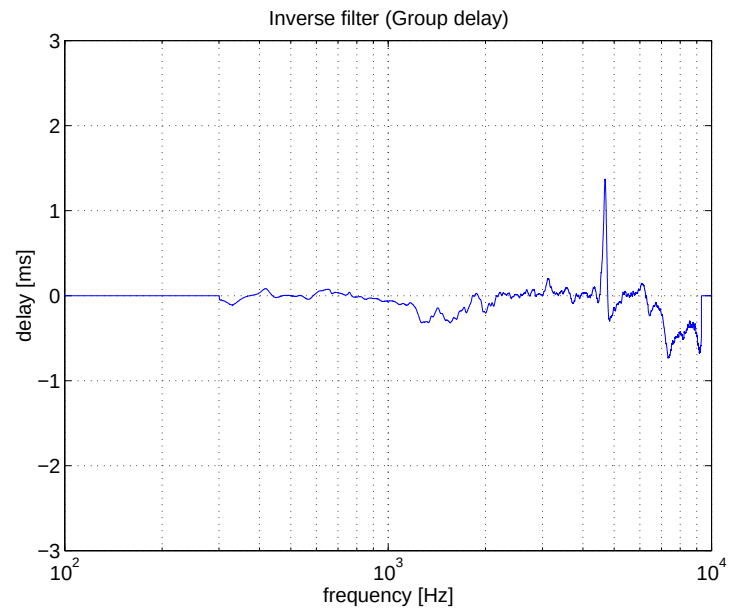


図 2.13: 逆フィルタ:群遅延特性

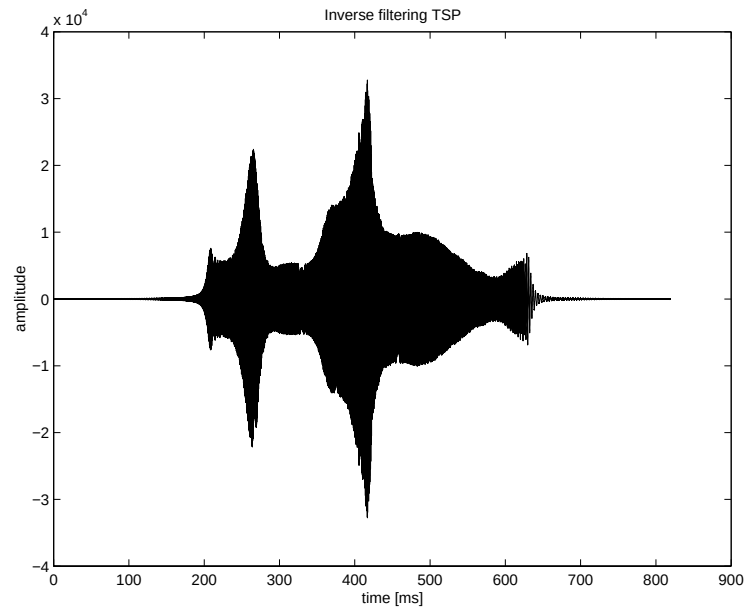


図 2.14: 逆フィルタをかけた TSP 信号

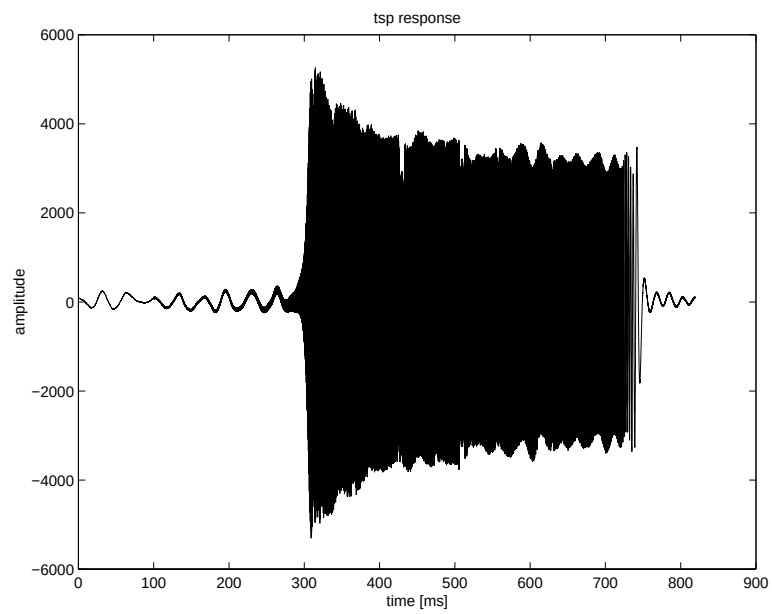


図 2.15: 逆フィルタをかけた TSP 信号の時間応答波形

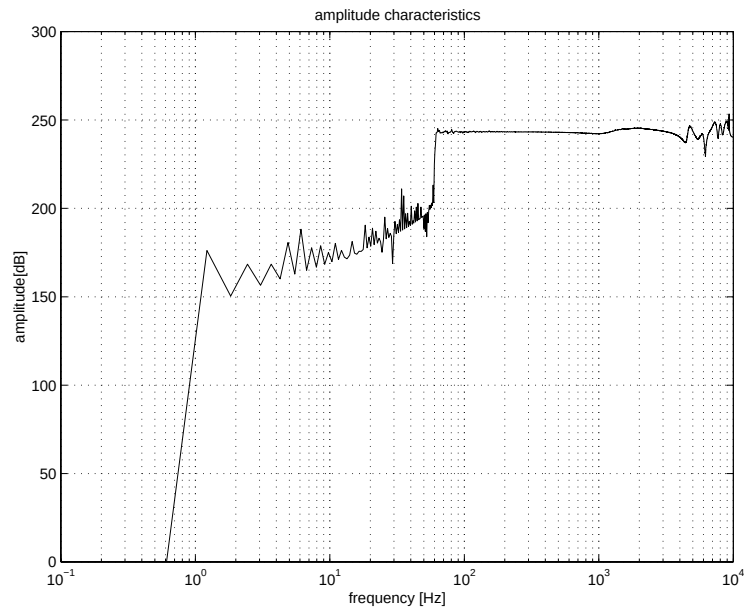


図 2.16: 逆フィルタをかけた TSP 信号の振幅特性

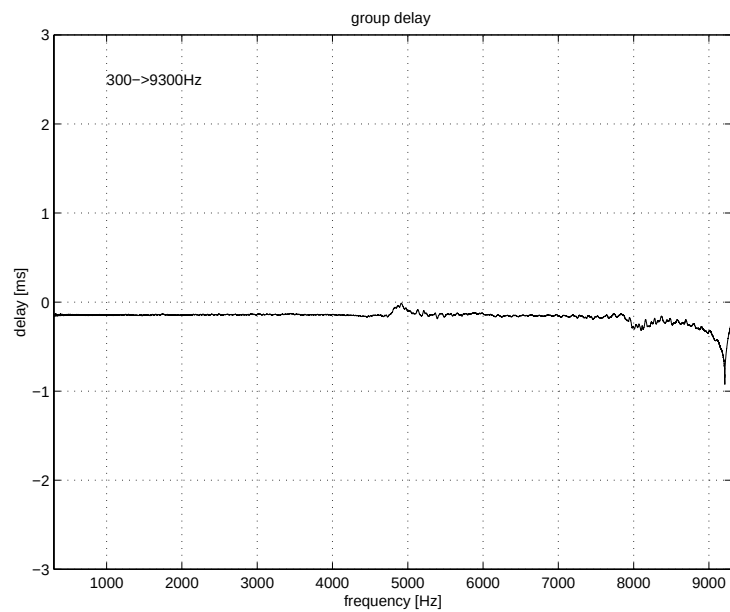


図 2.17: 逆フィルタをかけた TSP 信号の群遅延特性

第 3 章

実験 1

安武らの実験の追試

3.1 目的

2 章で、実験系の群遅延特性の歪みは逆フィルタにより、ほぼ補正できることが確認できた。本研究で行った実験に用いる刺激音は全て上述の手法で位相補正される。

最初の実験として安武らの実験 [11] の追試を行った。安武らの実験は本研究と同様に逆フィルタによる位相補正を行い、基本周波数 125Hz, 250Hz の 31 成分調波複合音を用いて実験を行った。その結果、

- 1 基本周波数が 125Hz のときの方が音色の違いを判別しやすい。
- 2 聴覚フィルタに 2-3 本の高調波成分が通過する場合に音色の違いが現れてくる可能性がある。

という結論を出している。特に 2 の結論は聴神経の発火パターンのモデルであるパルスリボンモデルの妥当性を支持する結果であり、聴覚機構の位相知覚特性を議論する上で重要な示唆を与えるものである。従って、この結論の妥当性を確認する為に、実験の追試をする。

3.2 刺激音

安武らの実験に用いられたのは式 3.1, 式 3.2で表される, 基本周波数 125Hz, 250Hz の 31 成分調波複合音である. 31 個の高調波成分の内, 1 成分のみ位相が π ずれている. 式 3.1, 式 3.2の h は位相を π ずらす成分の次数を表し, その成分が含まれる聴覚フィルタ内の高調波成分が 1~4 本になるように決められる.

なお, 聴覚フィルタのバンド幅は ERB(式 3.3) により計算し, 基本周波数から式 3.4により聴覚フィルタ内に含まれる高調波の数を求めた. 表 3.1に位相変化させる高調波の次数と周波数との対応を示す. また, 逆フィルタの補正帯域は $f_0 = 125Hz$ のときは 300Hz ~3875(= 125 × 31)Hz, $f_0 = 250Hz$ のときは 300Hz~7750(= 250 × 31) Hz に設定し, 刺激音が補正帯域外の成分を持たないようにしている.

$$f_0 = 125Hz$$

$$\sin 2\pi 3f_0t + \cdots + \sin(2\pi hf_0t + \pi) + \cdots + \sin 2\pi 31f_0t \quad (3.1)$$

$$h : \text{位相を}\pi\text{ずらす高調波の次数 } (h = 4, 5, 7, 10, 13, 17, 23, 30)$$

$$f_0 = 250Hz$$

$$\sin 2\pi 2f_0t + \cdots + \sin 2\pi(hf_0t + \pi) \cdots + \sin 2\pi 31f_0t \quad (3.2)$$

$$h : \text{位相を}\pi\text{ずらす高調波の次数 } (h = 3, 4, 5, 8, 11, 16, 22, 30)$$

$$ERB[Hz] = 24.7(4.37f[kHz] + 1) \quad (3.3)$$

$$n = [ERB]/f_0 + 1 \quad n : \text{聴覚フィルタ内の高調波の数}, [] : \text{整数部分} \quad (3.4)$$

表 3.1: 位相を変化させる高調波の次数とその周波数：() 内の数字はその高調波が含まれる聴覚フィルタ内の高調波の数

125Hz	調波の次数	4(1)	5(1)	7(1)	10(2)	13(2)	17(3)	23(3)	30(4)
	周波数 [Hz]	500	625	875	1250	1625	2125	2875	3750
250Hz	調波の次数	3(1)	4(1)	6(1)	8(1)	11(2)	16(2)	22(3)	30 (4)
	周波数 [Hz]	750	1000	1500	2000	2750	4000	5500	7500

3.3 実験条件

- 実験毎に実験系の特性を測定し直し、逆フィルタを更新した。
- 刺激音の立ち上がり、立ち下がり部には 5ms のリニアなテーパーをかけた。
- 基本周波数 125Hz, 250Hz のそれぞれについて 1 本の高調波成分の位相を変化させたものと位相変化を行っていないものの組合せ 8 対とそれを逆順にしたものの 8 対、同じ音同士の組合せ 9 対の計 25 対をランダムに片耳呈示し対比較実験を 2 回行った。
- 被験者は正常聴力を有し、聴取実験の経験がある大学院生 8 人である。
- ひとつの刺激音の信号長は 1 秒、刺激対の二つの刺激音の間隔は 0.5 秒、刺激対と刺激対の間隔は 4 秒である。

3.4 結果

結果を図 3.1 に示す。図の横軸は刺激の対を表し、数字は刺激音を表す。0 は位相を変化させていない刺激音であり、その他の数字は表 3.1 の刺激の実験番号に従う。また、縦軸は刺激音のペアの類似度を表し、位相を変化させた音を聞き分けられなかった被験者の割合を示す。類似度が高いペアであれば位相変化による音色の違いを判別できなかったということを表す。ただし、図 3.1 では左から 9 番目までの棒グラフは同じ音同士のペアの類似度を表す。

つまり、図 3.1 では左から 9 番目までの棒グラフは同じ音同士の対を被験者が正しく同じ音である判断できているかを表し、10 番目から 17 番目の棒グラフは違う音同士 (位相変化のない音と位相変化のある音のペア) が違う音であると判断できているかを表している。

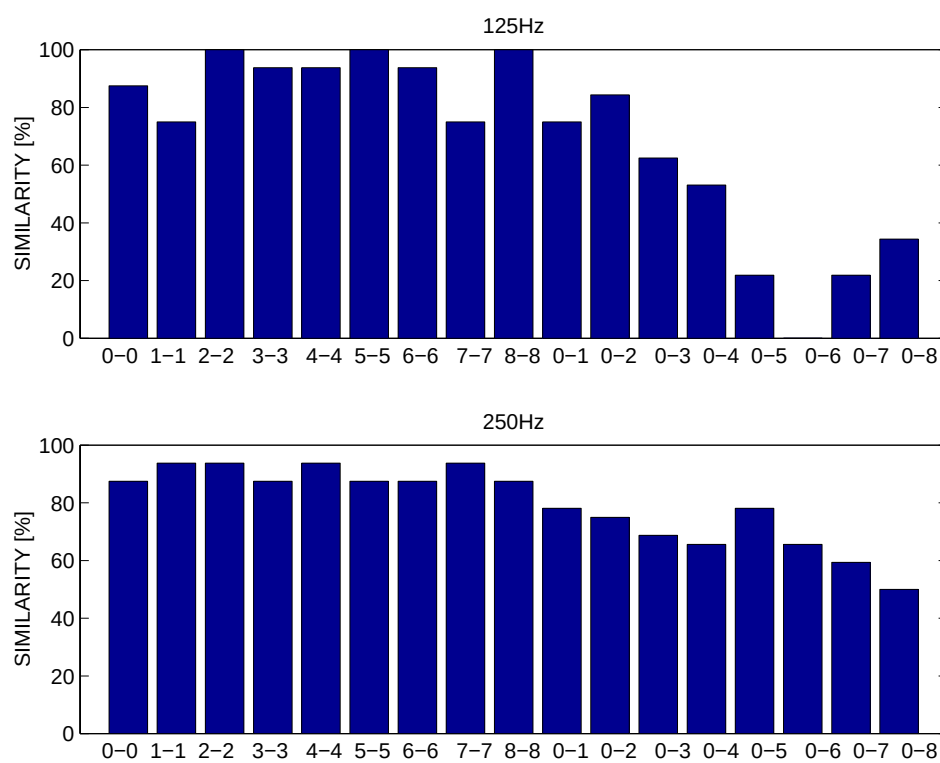


図 3.1: 位相変化のない原波形から作成した刺激音との類似度

3.5 考察

まず、基本周波数 125Hz の 31 成分調波複合音の実験結果の図 (図 3.1 上段) をみると、同じ音同士はほぼ、正しく判断できている。違う音同士の判別になると、0-1,0-2,0-3 のペアは違う刺激音のペアであるのにも関わらず、類似度が高く音色の違いを判断できていない。0-4 のペアで、やや類似度は下がるものの、判断できると断言できない。しかし、0-5 のペアで類似度は急激に下がり、更に、0-6 のペアでは類似度は 0% になり、全ての被験者が違う音であると判断している。音色の違いを判別できていない実験番号 1,2,3 の刺激音の聴覚フィルタ内の高調波成分をみるとすべて一本であることに気づく (表 3.1)。一方、音色の類似度が急激に下がり、音色の違いを判断できるとう結果がでた実験番号 5 の刺激音の聴覚フィルタ内の高調波成分数は、2 本であり、更に 0% の類似度を示した実験番号 6 の刺激音の聴覚フィルタ内の高調波成分数は、3 本である。この結果は安武の実験結果と同様なものであり、安武らの実験の妥当性を確認できた。

基本周波数 250Hz の場合では音色の判別は曖昧になり、全ての刺激音のペアにおいて類似度が高く、音色の違いの判断は著しく難しくなることがわかる。この結果は安武の報告においても同様であり、基本周波数が低いほうが音色の違いは大きいという過去の実験結果と矛盾するものではない。

第 4 章

実験 2

4.1 目的

実験 1 においては 31 成分調波複合音を用いて実験を行った。31 成分調波複合音には、基本周波数が 125Hz の場合、 $125 \times 31 = 3875\text{Hz}$ の高調波成分までしか含まれていない。一方、基本周波数が 250Hz の 31 成分調波複合音には $250 \times 31 = 7750\text{Hz}$ の高調波成分まで含まれている。従って、実験 1 では基本周波数が 125Hz の場合、高々 3175Hz の成分までしか位相変化による音色変化について調べられていない。また、それに伴い、実験 1 では聴覚フィルタ内の高調波数が 1~4 本までのものしか調べられていない。聴覚フィルタ内の高調波成分がそれ以上になる場合においても位相変化による音色の影響が現れるかについて調べる必要がある。

そのためには調波複合音の基本周波数が低い場合では高調波成分数を多くし、より高域の成分を含むようにすれば良い。しかし、基本周波数が同じで、高調波成分数のみが異なる調波複合音の音色が位相変化に対して同じように変化するのか、という疑問がある。

そこで、実験 2 では基本周波数が同じで高調波成分数の異なる調波複合音を用いて実験 1 と同様な位相知覚実験を行い、同様な結果が出るかどうかを調べる。もし、高調波成分数の影響がないならば、基本周波数によって高調波成分数を変化させた調波複合音を用いて実験が行えるようになる。

表 4.1: 位相を変化させる高調波の次数とその周波数：() 内の数字はその高調波が含まれる聴覚フィルタ内の高調波の数 $f_0 = 125\text{Hz}$

31 成分	調波の次数	3(1)	4(1)	8(1)	12(2)	16(2)	18(3)	20(3)	23(4)	28(4)	30(4)
	周波数 [Hz]	375	500	1000	1500	2000	2250	2500	2875	3500	3750
60 成分	調波の次数	4(1)	13(2)	17(3)	23(3)	30(4)	34(4)	38(5)	45(6)	50(6)	60(7)
	周波数 [Hz]	500	1625	2125	2875	3750	4250	4750	5625	6250	7500

4.2 刺激音

刺激音は基本周波数 125Hz の 31 成分調波複合音と 60 成分調波複合音である。実験 1 と同様に調波複合音の位相が 1 成分のみ、 π ずれている。刺激音を表 4.1に位相変化させる高調波の次数と周波数との対応を示す。31 成分調波複合音の場合、聴覚フィルタ内の高調波成分数は 1～4 本であり、60 成分調波複合音の場合は聴覚フィルタ内の高調波成分数は 1～7 本である。逆フィルタの補正帯域は $f_0 = 125\text{Hz}$ のときは $300\text{Hz} \sim 3875 (= 125 \times 31)\text{Hz}$, $f_0 = 250\text{Hz}$ のときは $300\text{Hz} \sim 7500 (= 125 \times 60)\text{Hz}$ に設定し、刺激音が補正帯域外の成分を持たないようにしている。

4.3 実験条件

- 実験毎に実験系の特性を測定し直し、逆フィルタを更新した。
- 刺激音の立ち上がり、立ち下がり部には 5ms のリニアなテーパをかけた。
- 基本周波数 125Hz の 31 成分調波複合音と 60 成分調波複合音それぞれについて 1 本の高調波成分の位相を変化させたものと位相変化を行っていないものの組合せ 10 対とそれを逆順にしたもの 10 対、同じ音同士の組合せ 11 対の計 31 対をランダムに片耳呈示し対比較実験を 1 回行った。
- 被験者は正常聴力を有し、聴取実験の経験がある大学院生 8 人である。
- ひとつの刺激音の信号長は 1 秒、刺激対の二つの刺激音の間隔は 0.5 秒、刺激対と刺激対の間隔は 4 秒である。

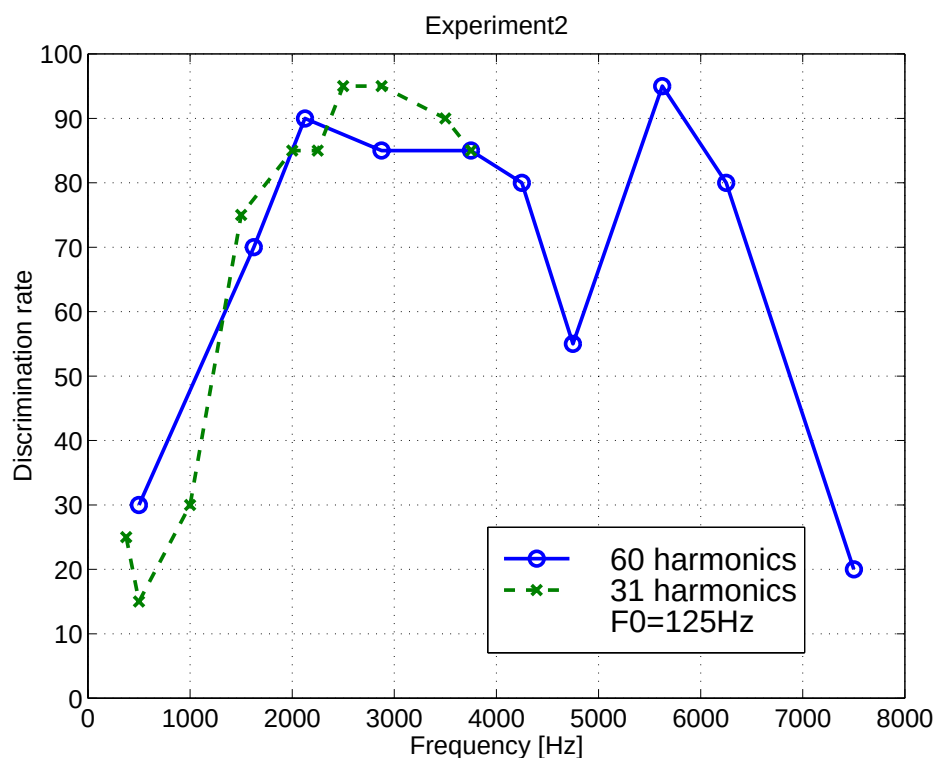


図 4.1: 位相変化のない原波形から作成した刺激音との判別率

4.4 結果

実験結果を図 4.1に示す. 図 4.1 は位相変化のない原波形から作成した刺激音と表 4.1 に示された, 刺激音との判別率を表している. 図 4.1の横軸は位相を π ずらした高調波成分の周波数を表しており, 表 4.1の周波数と対応している. 判別率が高い程, 被験者が位相変化のない原波形から作成した刺激音と一つの調波成分の位相を π ずらした原波形から作成した刺激音とを聞き分けられたことを意味する. 図中, 実線が 60 成分調波複合音の実験結果を表し, 破線が 31 成分調波複合音の実験結果を表す.

4.5 考察

図 4.1によると、60 成分調波複合音の実験結果と 31 成分調波複合音の実験結果にはそれほど差がない。4000Hz くらいまでの 60 成分調波複合音の実験結果と、31 成分調波複合音の実験結果は、ほぼ一致している。31 成分調波複合音は 3875Hz までの成分しか含んでいないので必然的に、図はそこまでの結果しかない。

60 成分調波複合音を用いた実験結果において、判別率が急激に上がるのは 1625Hz-2125Hz の高調波成分の位相を変化させた原波形から作成した刺激音であった。これらの高調波成分について、聴覚フィルタに含まれる調波成分数をみると、実験 1 の結果と同様に 2-3 本である。

従って、位相変化による音色の変化は高調波成分数による影響は小さいということがわかった。また、聴覚フィルタ内の高調波成分数が 2-3 本であるときに音色の変化が知覚しやすいという傾向も高調波成分数によらず、共通しているということがわかった。

実験において被験者の多くは、音色の違いは調波複合音とは別に分離して、純音のような音が聞こえてくるという内観報告をした。この現象は Duifhuis[16][17]によっても報告されている。

第 5 章

実験 3

5.1 目的

実験 1, 2 では基本周波数が 125Hz, あるいは 250Hz の調波複合音を用いて実験を行った。しかし、聴覚フィルタに含まれる高調波成分数が 2-3 本である場合に位相変化による音色の変化が知覚されやすいという傾向が、他の基本周波数の調波複合音においてもみられるかという疑問がのこる。また、位相を変化させる高調波成分の周波数と位相変化知覚との関係について、より詳細に調べる必要がある。

従って、実験 3 では様々な基本周波数の調波複合音を用いて位相変化による音色変化知覚実験を行う。

実験 2 により、調波複合音の位相変化による音色の変化は、高調波成分数の影響をほとんど受けないことがわかった。従って、基本周波数の低い場合は高調波成分数を多くとり、最低でも 5kHz 以上の成分を含むような調波複合音を用いて実験を行い、出来るだけ広帯域にわたって調べる。また、実験 2 において、音色の違いは、調波複合音から純音のような音が分離して聞こえてくるという内観報告が多かったため、この現象がどのような場合に起こるのかについても調べる。

5.2 刺激音

刺激音は基本周波数が 62.5Hz から 250Hz まで 2 オクターブを 1/4 オクターブ毎に設定した調波複合音である。調波複合音の成分数は基本周波数によって最低 5kHz 以上の成分を含むように設定した。逆フィルタの補正帯域は 300Hz から各基本周波数が含む最も高い成分の周波数までとし、刺激音が補正帯域外の成分を持たないようにしている。表 5.1

表 5.1: 位相を変化させる高調波の次数とその周波数：() 内の数字はその高調波が含まれる聴覚フィルタ内の高調波の数

62.5Hz	調波の次数	5(1)	13(2)	19(3)	28(4)	34(5)	42(5)	72(9)	84(10)
	周波数 [Hz]	312	812	1187	1750	2125	2625	4500	5250
74Hz	調波の次数	5(1)	13(2)	21(3)	24(3)	27(4)	41(5)	63(8)	73(9)
	周波数 [Hz]	370	963	1556	1778	2000	3038	4668	5409
89Hz	調波の次数	4(1)	11(2)	20(3)	22(3)	27(4)	50(6)	60(7)	65(8)
	周波数 [Hz]	355	978	1778	1956	2400	4445	5334	5778
105Hz	調波の次数	3(1)	11(2)	19(3)	21(3)	25(3)	33(4)	40(5)	50(6)
	周波数 [Hz]	316	1158	2000	2211	2632	3475	4212	5265
125Hz	調波の次数	4(1)	10(2)	13(3)	17(3)	25(4)	30(4)	40(6)	56(7)
	周波数 [Hz]	500	1250	1625	2125	3125	3750	5000	7000
148Hz	調波の次数	3(1)	9(2)	19(3)	24(3)	29(4)	37(5)	42(5)	48(6)
	周波数 [Hz]	444	1333	2814	3554	4295	5480	6220	7108
174Hz	調波の次数	3(1)	11(2)	16(2)	21(3)	27(4)	33(4)	42(5)	44(5)
	周波数 [Hz]	522	1914	2784	3564	4698	5742	7308	7656
210Hz	調波の次数	4(1)	12(2)	17(2)	21(3)	26(3)	31(4)	34(4)	37(5)
	周波数 [Hz]	840	2520	3570	4410	5460	6510	7140	7770
250Hz	調波の次数	2(1)	6(1)	11(2)	14(2)	17(2)	20(3)	28(4)	30(4)
	周波数 [Hz]	500	1500	2750	3500	4250	5000	7000	7500

に刺激音を示す。位相変化幅は π である。

5.3 実験条件

- 実験毎に実験系の特性を測定し直し、逆フィルタを更新した。
- 刺激音の立ち上がり、立ち下がり部には 5ms のリニアなテーパーをかけた。
- 各基本周波数の調波複合音について 1 本の高調波成分の位相を変化させたものと位相変化を行っていないものの組合せ 8 対とそれを逆順にしたもの 8 対、同じ音同

士の組合せ 9 対の計 25 対をランダムに片耳呈示し対比較実験を 1 回行った。

- 被験者は音色の違いを判別し、音色が違ふと判断した場合には音色の違いの性質として、「純音が分離したように聞こえる」か否かを判断してもらい、2 段階で音色を判断してもらった。(付録:回答用紙)
- 被験者は正常聴力を有し聴取実験の経験がある大学院生 8 人である。
- ひとつの刺激音の信号長は 1 秒、刺激対内の二つの刺激音の間隔は 0.5 秒、刺激音と刺激音の間隔は 4 秒である。

5.4 結果

実験結果を図 5.1～図 5.9に示す。各基本周波数の調波複合音について、位相変化のない原波形から作成した刺激音と表 5.1に示された、刺激音との判別率を表している。図 5.1～図 5.9の横軸は位相を π ずらした高調波成分の周波数を表しており、表 5.1の周波数と対応している。判別率が高い程、被験者が位相変化のない原波形から作成した刺激音と一つの調波成分の位相を π ずらした原波形から作成した刺激音とを聞き分けられたことを意味する。

また、×は、75%以上の判別率のときに、調波複合音から純音のような音が分離したように聞こえると感じた被験者の割合を示している。

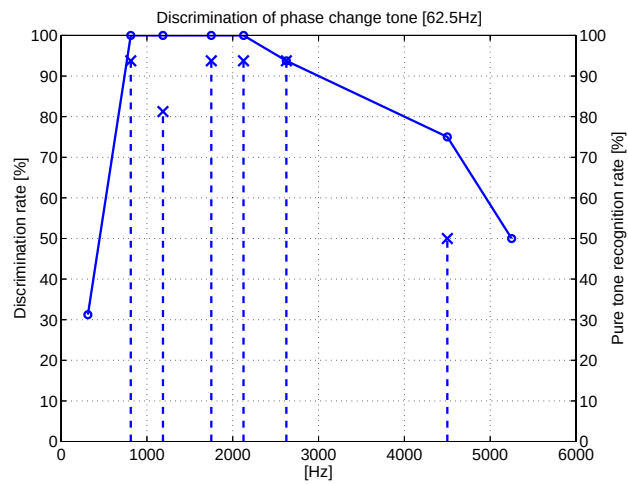


図 5.1: 実験結果: $f_0 = 62.5\text{Hz}$

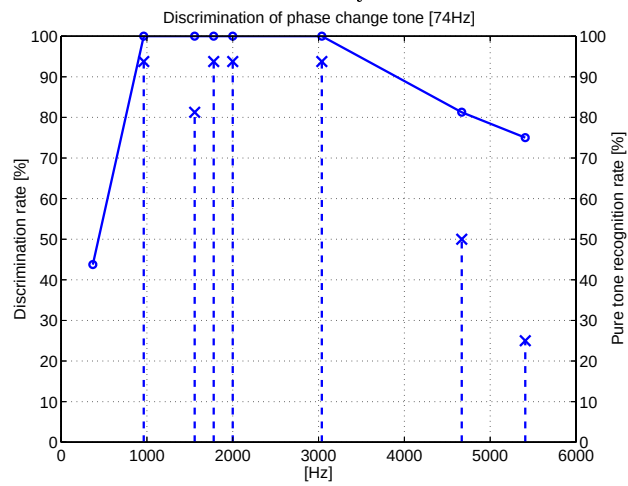


図 5.2: 実験結果: $f_0 = 74\text{Hz}$

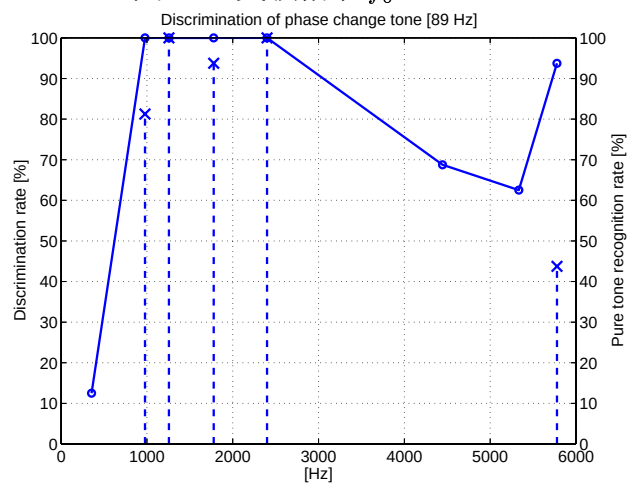


図 5.3: 実験結果: $f_0 = 89\text{Hz}$

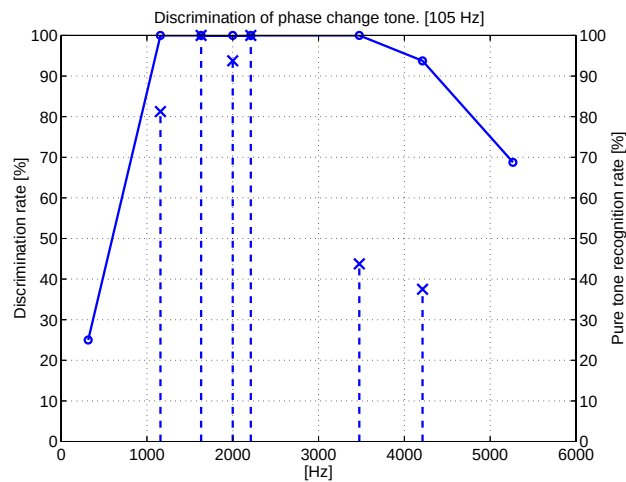


図 5.4: 実験結果: $f_0 = 105\text{Hz}$

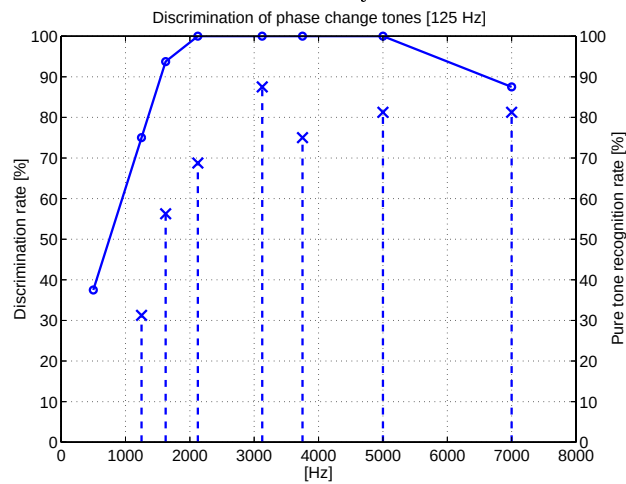


図 5.5: 実験結果: $f_0 = 125\text{Hz}$

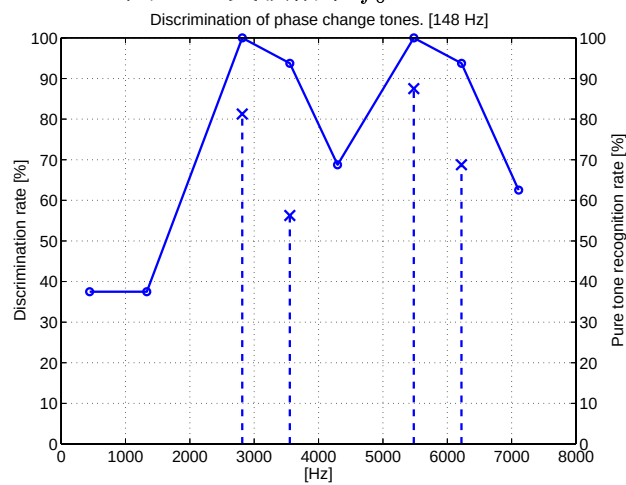


図 5.6: 実験結果: $f_0 = 148\text{Hz}$

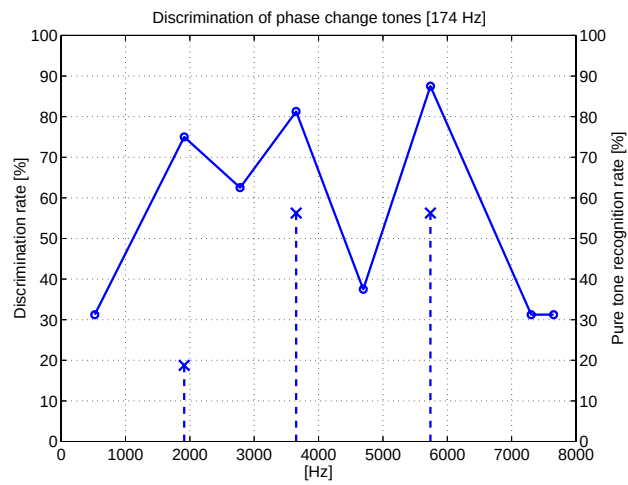


図 5.7: 実験結果: $f_0 = 174\text{Hz}$

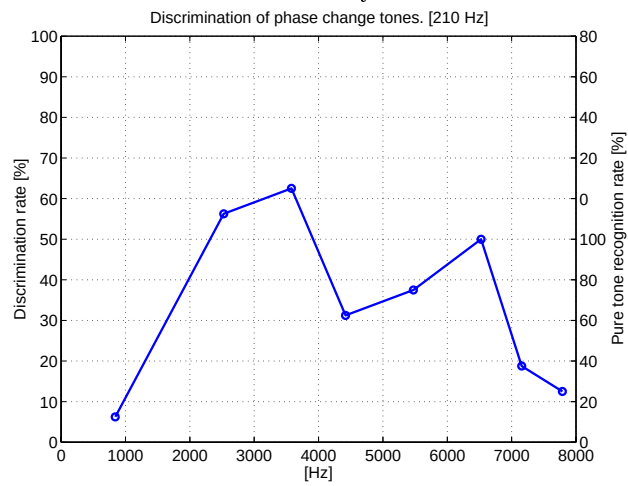


図 5.8: 実験結果: $f_0 = 210\text{Hz}$

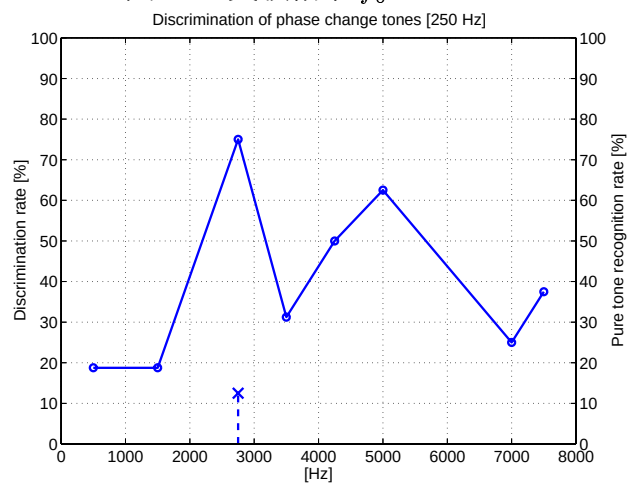


図 5.9: 実験結果: $f_0 = 250\text{Hz}$

5.5 考察

5.5.1 実験結果について

基本周波数が 210Hz より低い調波複合音の場合、実験結果は幾つか共通した特徴を表している。

第一に、実験 1, 2 と同様に、聴覚フィルタ内の高調波成分の数が 2-3 本であるときに判別率が急激に上がるということである。これに関しては、聴覚フィルタ内の成分の位相差からパルスの間隔が変化し、差が検出できるとするパルスリボンモデルの妥当性がより確実なものとなったといえるだろう。

第二に、位相変化による音色の違いが判別できるようになる高調波成分より高い周波数では、かなり広帯域にわたり、高い判別率が持続されるようになる。210Hz より低い基本周波数の調波複合音の実験では、6kHz を越えるような高調波成分においても判別率は 75% を越えるようなものがあった。

第三に、実験 2 の「音色の変化は純音のような音が、調波複合音とは分離したように聞こえる。」という被験者の内観報告をうけて、実験 3 においては、音色が違うと判断した場合にはこの“純音分離現象”によるものか否かを判断してもらったが、実験の結果、判別率が高いときはこの“純音”の認識率も同時に高いものが多かった。Duifhuis[16][17] においては、調波複合音の高調波成分の位相を反転させるとその成分が分離して知覚されると報告されている。実験では特に、基本周波数が低い場合にこの傾向が顕著にみられた。

基本周波数が高くなると、やはり、判別は曖昧になる。基本周波数が 174Hz ではそれより低い基本周波数の調波複合音のときよりも判別率が 75% を越えるような刺激音は少なくなり、判別率のグラフも変動が大きくなっている。また、判別率が 75% を越える場合の純音の認識率も低くなっている。

基本周波数が 210Hz, 250Hz になると、もはや、判別率が 75% を越えることはほとんどなかった。また、多くの被験者が「純音が分離して聞こえてくるという現象はほとんどなかった」という内観報告をしている。この実験の結果、基本周波数が高くなると位相変化による音色の違いはわかりにくくなるという過去の報告に対して、174Hz～210Hz の間に調波複合音の基本周波数の検知限があるという結果を得た。また、Patterson[8] では基本周波数が 200Hz 以下であれば最初の 12 成分からなる調波複合音でも位相変化は知覚できると報告されていることに対し、表 5.1 をみると、基本周波数 200Hz より低い場合で聴覚フィルタ内の高調波成分数が 2 本以上になるのは、最初の 12～13 本以下の成分のときであり、この報告と今回の実験結果は、ほぼ一致していた。

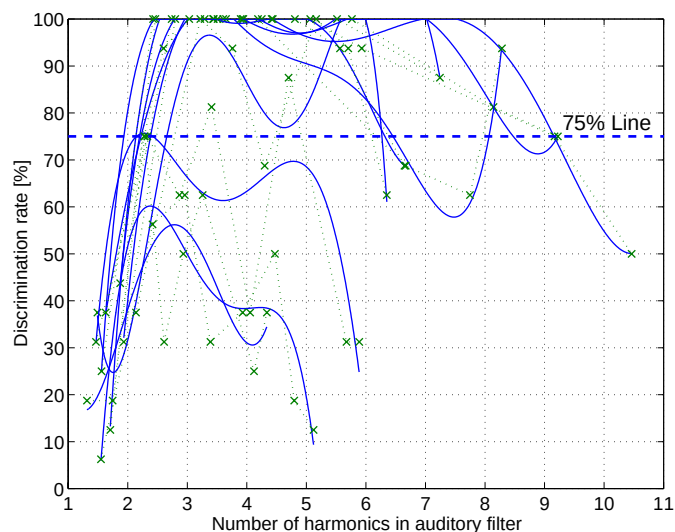


図 5.10: 実験結果を多項式近似した曲線，横軸:聴覚フィルタ内の成分数，縦軸: 判別率

5.5.2 実験データの分析

実験を通して，わかったことは

- 基本周波数が約 200Hz より低い 調波複合音において位相変化させる高調波成分が含まれる聴覚フィルタ内に含まれる高調波成分数が 2-3 本であるとき音色の違いを判断できる。

ということである．位相変化による音色変化と聴覚フィルタ内の高調波成分数の関係を明確にするために，実験 3 の結果をもとに位相変化を知覚できる聴覚フィルタ内の高調波数の検知限を予測した．実験 3 の判別率のデータを多項式の最小 2 乗法により近似して描いた曲線 (図 5.10) において，初めて 75% の直線と交差する点を求めた．その結果を表 5.2 に示す．表 5.2 の値の平均値を計算し，位相変化を知覚できる聴覚フィルタ内の高調波数の検知限は 2.24 であった．

f0[Hz]	62.5	74	89	105	125	148	174	210	250
成分数	2.38	2.20	2.14	1.94	2.23	2.66	2.16	-	-

表 5.2: 判別率が 75% を初めて越えるときの聴覚 フィルタ内の成分数

第 6 章

Auditory Image Model(AIM)[10] によるシミュレーション

6.1 位相変化による Neural Activity Pattern(NAP) の変化

以上より，基本周波数が 200Hz より小さい場合，聴覚フィルタ内の高調波成分数が 2-3 本以上であるときは位相変化による音色の変化が知覚できるということがわかった．

本節では聴神経の発火パターンのモデルであるパルスリボンモデルをさらに発展させた Auditory Image Model[10] によってシミュレーションを行い，実験結果と比較することにより聴覚機構における位相知覚特性について考察する．

AIM は Gammatone フィルタを用いて聴覚系の処理機構を模擬したモデルである (詳細は Patterson[10] など)．Gammatone フィルタの出力をその中心周波数に対応した基底膜上の位置での振動パターン (Basilar Membrane Movement:BMM) として捉え，これに内有毛細胞のモデルを付け加えることにより聴神経の発火パターン (Neural Activity Pattern:NAP) を計算する．ここでは，位相変化による NAP の変化を見る．

AIM によりシミュレーションを行った波形データは

- 1 位相を全く変化させていない調波複合音
- 2 聴覚フィルタ内の成分が 3 本以下の高調波成分位相反転音
- 3 聴覚フィルタ内の成分が 3 本以上 5 本以下の高調波成分位相反転音
- 4 聴覚フィルタ内の成分が 5 本以上の高調波成分位相反転音

である．その結果の一部を図に示す(図 6.1 ～ 図 6.9)．図は聴神経の発火パターン(NAP)である．

ここで用いた Gammatone フィルタバンクは中心周波数 300～8000Hz を ERB-rate 上で 32 に等分割した中心周波数をもつバンド幅 1ERB のフィルタバンクである．図は $F0=62.5, 125, 250\text{Hz}$ の場合を示す．それぞれの周波数において聴覚フィルタ内の成分数が 1 本, 3 本, 5 本以上のときの成分位相反転音における, その成分の周波数に一番近いチャンネルの NAP を, 原音のときのそのチャンネルの NAP とともに示した．

$F0=62.5\text{Hz}$ において, 聴覚フィルタ内の成分数が 1 のとき(図 6.1)の成分を位相反転させた調波複合音の NAP は, 原音の NAP とあまり変化はない．この場合, 実験結果も判別率が低い．一方フィルタ内の成分数が 3 のとき(図 6.2)の成分を位相反転させた調波複合音の NAP は原音の NAP と比べて発火パターンが異なっている．原音の場合では発火していない点においても, パルスが生じている．この場合, 実験結果も判別率は高くなっている．更にフィルタ内の成分数が増えて 9 (図 6.3) になると, 大きなパルスの間に振幅変調された波形が現れる．この場合においても実験結果では, 判別率は高い．

いずれにせよ, 判別率が高い刺激音に対しては NAP の様相は原音と比較して異なっていることがわかる．

この, NAP の変化の傾向は $F0=125\text{Hz}, 250\text{Hz}$ においても同様である．しかし, $F0=250\text{Hz}$ の場合ではどの場合でも, 実験結果は曖昧であり判別率が 75%を越えることはなかった．

基本周波数が 250Hz では判別が曖昧になるのに対して, NAP では聴覚フィルタ内の高調波成分数が 3 本以上の場合, はっきりと様相の違いが表れてしまう．これは過去の実験結果とも対応していない．基本周波数が 250Hz の場合, 聴覚フィルタ内の高調波成分が 2-3 本以上になる周波数は 4～5kHz 以上であり, AIM によるシミュレーションにおいてもこれに対応するチャンネルの NAP を観測した．しかし, AIM によってシミュレートされる NAP は, 内有毛細胞において位相同期によって発火する神経パルスを表現することを主な目的としているので, 位相同期が起らないと考えられている 5kHz 以上の周波数成分に対しては, あまり考慮されていない可能性がある．そのために, 実験結果と AIM のシミュレーション結果が対応していないのかもしれない．

あるいは, 今回, 可能な限り位相補正を行なったが, 補正できるのは比較的大きな群遅延の変動であり, 実際はわずかな変動が全周波数帯域にある．このわずかな変動は基本周期が長い場合はあまり影響無いが, 基本周期が短くなると刺激音の生成時に, ある成分を π , つまり半周期ずらしていたとしても, 実際には, 群遅延のわずかな変動の影響で位相の差がかき消され, 検出できなくなっている可能性がある．従って, 更に位相補正の精度を向上させて実験を行った場合, 基本周波数が 200Hz 以上の場合でも位相変化による音

色変化が知覚できる可能性もある。

また、今回の実験では位相を π 変化させる成分の周波数が 5kHz を越えても 8kHz 位までは音色の判別ができるという結果を得た。一般に、位相差の検出は原波形との位相同期した神経発火によるものであると考えられているが、5kHz 以上の成分においては位相同期は起こらない。にもかかわらず、実験結果は 5kHz 以上の成分においても音色の違いが判別できるという結果が出た。NAP においても原音との違いがはっきりとでている。前述のように、5kHz 以上の成分においての AIM のシミュレーションの妥当性は定かではないが、5kHz 以上の成分においては原波形の位相そのものに同期して神経発火が起こっているのではなく、波形の包絡を聞いている可能性もあり、更に議論が必要である。

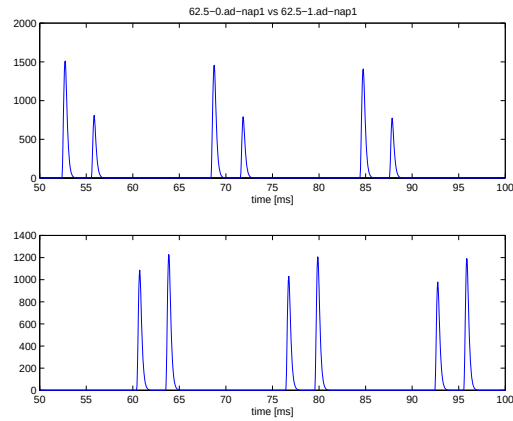


図 6.1: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 1 のチャンネル $f_0 = 62.5\text{Hz}$

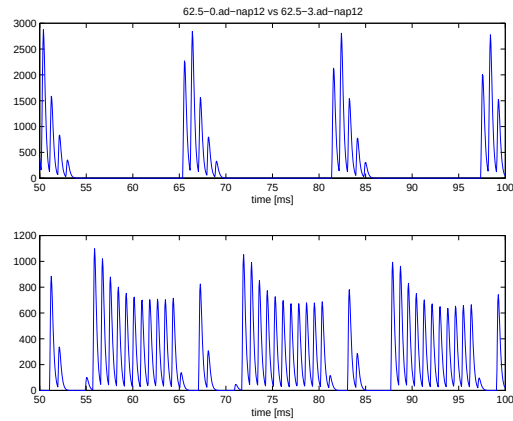


図 6.2: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 3 のチャンネル $f_0 = 62.5\text{Hz}$

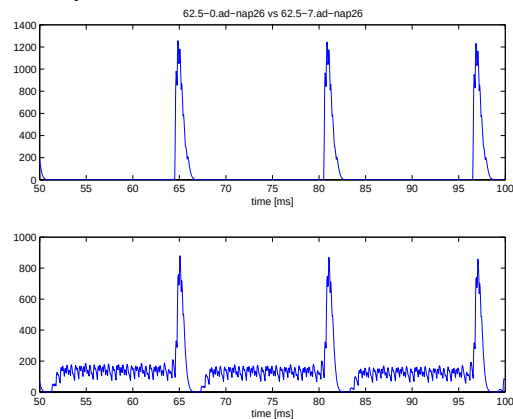


図 6.3: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 9 のチャンネル $f_0 = 62.5\text{Hz}$

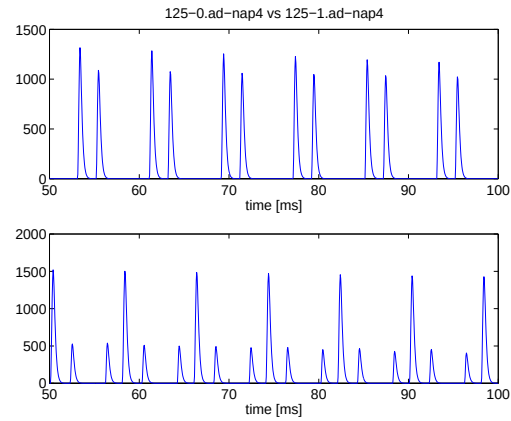


図 6.4: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 1 のチャンネル $f_0 = 125\text{Hz}$

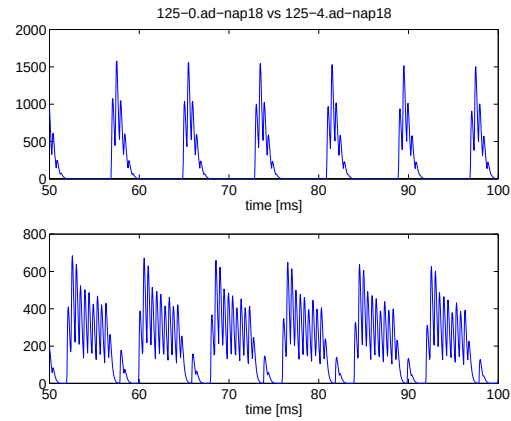


図 6.5: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 3 のチャンネル $f_0 = 125\text{Hz}$

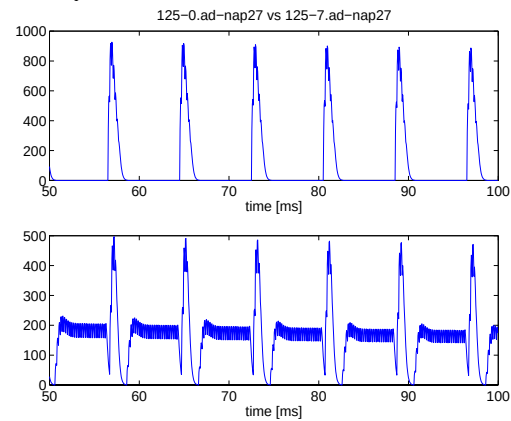


図 6.6: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 6 のチャンネル $f_0 = 125\text{Hz}$

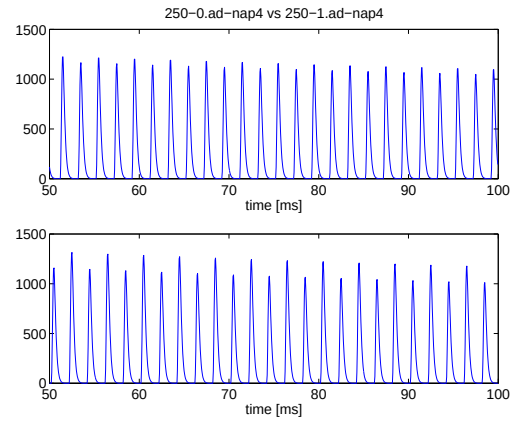


図 6.7: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 1 のチャンネル $f_0 = 250\text{Hz}$

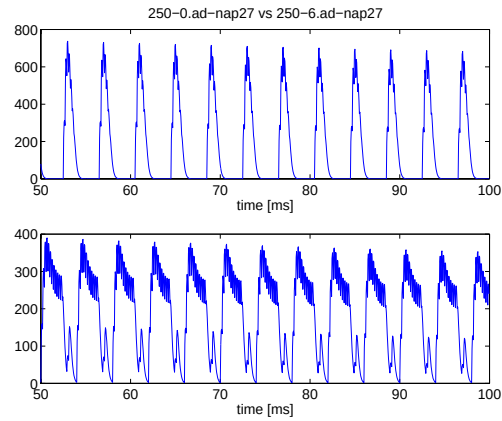


図 6.8: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 3 のチャンネル $f_0 = 250\text{Hz}$

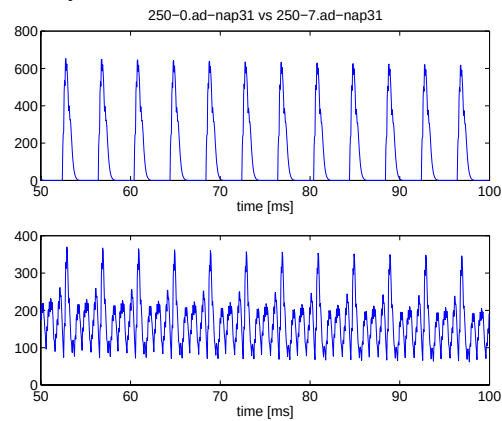


図 6.9: 位相変化のない波形の NAP(上段) と位相変化のある波形の NAP(下段) :聴覚フィルタ内の成分数 4 のチャンネル $f_0 = 250\text{Hz}$

6.2 Hiss 相関

図 6.1～図 6.6 に表した NAP の自己相関を図 6.10～6.15 に示す。図は基本周波数 62.5Hz と 125Hz の場合である。自己相関をとると、NAP の違いはよりはっきりと表れる。特に図 6.11 の下段の図では約 1200Hz の周波数で発火が起こっている。これはこのときの位相を π ずらした成分の周波数 (1187Hz) とほぼ一致している。また、図 6.14 の下段の図においても NAP の自己相関の周波数と位相を π ずらした成分の周波数がほぼ一致している。つまり、これらの場合においては位相を変化させた成分とほぼ同期した発火が起こっていると考えられる。これが事実ならば、位相変化させた成分に位相同期した神経発火が起こることによってこの成分が知覚され、このために実験において純音が分離したように知覚されるという現象が起こっていた可能性がある。実際、基本周波数 62.5Hz のとき 1187Hz の成分の位相を変化させた刺激音についての実験結果 (図 5.1) では判別率、純音の認識率がともに高い。

また、音色を表す指標の一つに、Hiss 相関がある [9]。Hiss 相関は NAP の自己相関の平均値をとり、更にその平方根を計算したものである。基本周波数が 62.5Hz, 125Hz, 250Hz において、位相変化のある調波複合音と位相変化のない調波複合音の Hiss 相関の差を求めた。その結果を図 6.16 に示す。横軸は聴覚フィルタ内に含まれる高調波成分の本数である。その結果、基本周波数が 62.5Hz, 125Hz において、Hiss 相関の差が大きく表れるのは聴覚フィルタ内の高調波成分数が 2-3 本以上になるときであり、この点に関しては実験結果と対応しており Hiss 相関は音色に関して妥当な指標であることがわかる。しかし、基本周波数が 250Hz の場合においては実験結果と対応しておらず、妥当性を欠く。

以上より、基本周波数が低い場合には AIM と実験結果はよい対応を示しており、位相変化は聴神経の発火パターンに影響していることがわかった。しかし基本周波数が高い場合においては更なる議論が必要である。

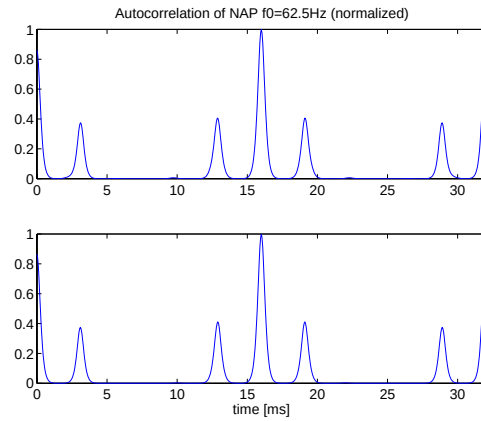


図 6.10: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 1 のチャンネル $f_0 = 62.5\text{Hz}$

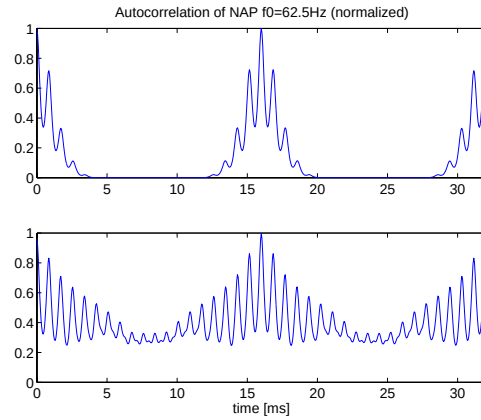


図 6.11: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 3 のチャンネル $f_0 = 62.5\text{Hz}$

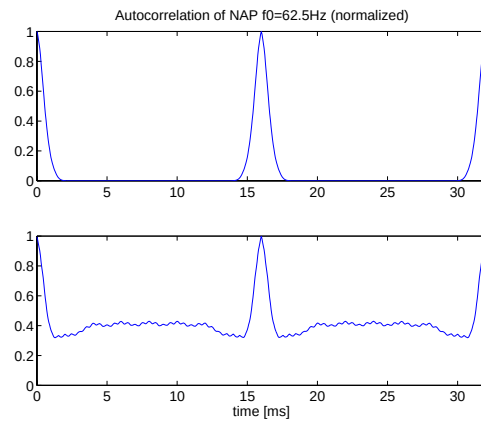


図 6.12: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 9 のチャンネル $f_0 = 62.5\text{Hz}$

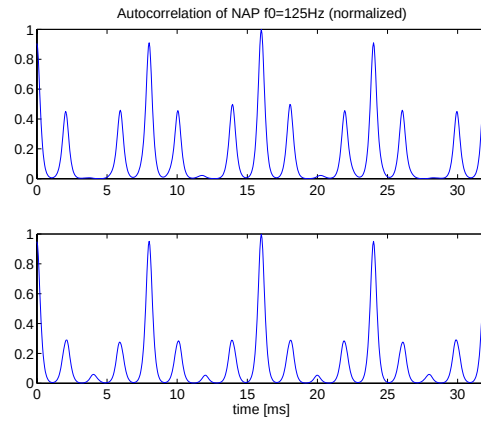


図 6.13: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 1 のチャンネル $f_0 = 125\text{Hz}$

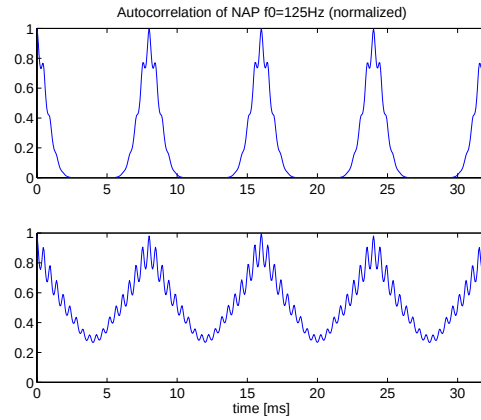


図 6.14: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 3 のチャンネル $f_0 = 125\text{Hz}$

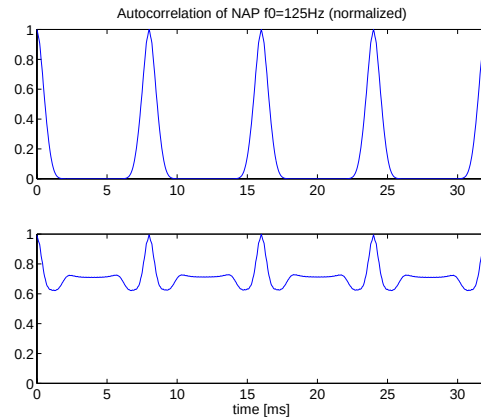


図 6.15: 位相変化のない波形の NAP の自己相関 (上段) と位相変化のある波形の NAP の自己相関 (下段) : 聴覚フィルタ内の成分数 6 のチャンネル $f_0 = 125\text{Hz}$

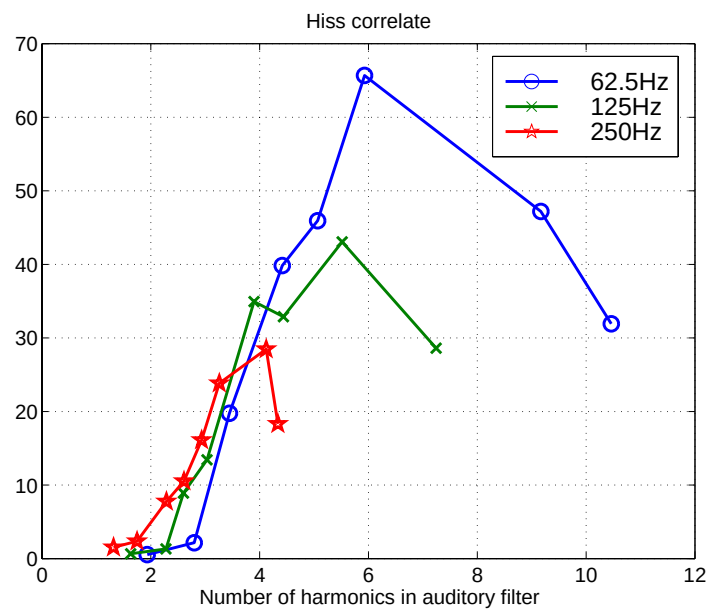


図 6.16: Hiss 相関

第 7 章

おわりに

7.1 まとめ

- 位相に関する過去の実験における実験精度に関する疑問から，実験系の位相歪みを測定し，これを可能な限り補正した．
- 様々な基本周波数の調波複合音を用いて，位相変化に対する音色変化の知覚実験を行い，その結果，およそ 200Hz より低い基本周波数の調波複合音においては，聴覚フィルタ内の高調波成分数が 2-3 本以上である場合に位相変化による音色変化が知覚できることがわかった．
- AIM によるシミュレーションを行い，基本周波数が 200Hz より低い場合，AIM におけるシミュレーションと実験結果は良く対応していることを示した．

今後の課題として，基本周波数が 200Hz を越える場合において位相変化が知覚できなくなるという実験結果に対して，更なる考察が必要である．また，これに関して群遅延特性の微細な変動の影響についての調査が求められる．

謝辞

本論文の作成にあたり，熱心に御指導して頂いた赤木 正人 教授に厚く御礼申し上げます。

また，本研究を進める過程において，多大なアドバイスを下さった水町 光徳 氏，伊藤 一仁氏をはじめ，実験に御協力いただいた赤木研究室の皆様に感謝致します。最後に，本研究に対する議論とアドバイスを頂いた，すべての皆様に謝意を表します。

参考文献

- [1] von Helmholtz, H.L.F.: “Ueber die Klangfarbe der Vocale,” Ann.Physik.Chem., 18, 280-290, (1859).
- [2] Fletcher, H.: “Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency and the Over Structure,” J.Acoust.Soc.Am., 6, 2, 59-69, (1934).
- [3] Mathes, R.D., and Miller, R.L. : “Phase effects in monaural perception,” J.Acoust Soc.Am., 19, 780-797, (1947).
- [4] Licklider, J.C.R.: “Effects of Changes in the Phase Pattern upon the Sound of a 16 Harmonic Tone,” J.Acoust.Soc.Am., 29, Suppl.O3, 780, 53rd meeting, (1957).
- [5] Shroeder, M.R.: “New Results Concerning Monaural Phase Sensitivity” J.Acoust.Soc.Am., 31, Suppl.J5, 1579, 58th meeting, (1959).
- [6] Plomp, R. and Steeneken, H.J.M.: “Effect of Phase on the Timbre of Complex Tones,” J.Acoust.Soc.Am., 46, 2, 409-421, (1969).
- [7] Blauert, J. and Laws, P.: “Group delay distortions in electroacoustical systems,” J.Acoust.Soc.Am., 63, 5, 1478-1483, (1978).
- [8] Patterson, R.D.: “A pulse ribbon model of monaural phase perception,” J.Acoust Soc.Am., 82, 5, 1560-1568, (1987).
- [9] Akeroyd, M.A. and Patterson, R.D.: “Discrimination of wideband noises modulated by a temporally asymmetric function” J.Acoust Soc.Am., 98, 5, 2466-2474. (1995)

- [10] Patterson,R.D. and Allerhand,M.H: “Time-domain modeling of peripheral auditory processing:A modular architecture and a software platform” J.Acoust.Soc.Am.,98,4,1890-1894.(1995)
- [11] 赤木, 安武: “時間方向情報の知覚の検討一位相変化の音色知覚に及ぼす影響について” 信学技報,EA98-19 (1998-06).
- [12] 平原: “聴覚実験用ヘッドホン再考” 聴覚研究会資料,H-95-55 (1995).
- [13] 鈴木, 浅野, 金, 曾根: “時間引き延ばしパルスの設計法に関する考察” 信学技報,EA92-86 (1992-12)
- [14] 坂野, 陸, 中村, 河原 “短時間位相の効率的表現方法” 聴覚研究会資料,H-97-46 (1997)
- [15] 河原, 津崎, パターソン: “オールパスフィルタの位相操作による時間構造制御とその知覚への影響について” 聴覚研究会資料,H-96-79 (1996)
- [16] Duifhuis,H : “ Auditory of high harmonic in a periodic pulse” J.Acoust Soc.Am.,48.888-893,(1970).
- [17] Duifhuis,H : “ Auditory of high harmonic in a periodic pulse II” J.Acoust Soc.Am.,49.1155-1162,(1971).