令和 4 年度 卒業論文

バンド編曲に向けたギター音源からベース音源を生成する CNN モデル

指導教員 北原鉄朗教授

日本大学 文理学部 情報科学科

香西 智雄

2023 年 2 月 提出

概　　要

ギターは，軽音楽において中心的な楽器の 1 つである．そのため，多くのアマチュアギタリストが存在し，ギターを弾きながら作曲を楽しむ者も少なくない．しかし，作曲した楽曲をバンドで演奏するには，ベースやドラムスなど各楽器パートの演奏内容を決める作業（編曲）が必要になる．編曲を行うには，各楽器の特性を知る必要があったり，編曲した結果を人に伝えるには楽譜に書くか DTM (desktop music) を用いる必要があるため，簡単にできるものではない．

本研究が目指すのは，ユーザが作曲した楽曲の伴奏がギター 1 本で与えられたときに，他の楽器パートの演奏内容を自動で決めて、バンドで演奏できるようにするシステムの実現である．他の楽器パートとしては，ベース，ドラムス，キーボードなどが考えられるが，本稿ではベースのみを扱うこととする．

実験として，三つの条件と三つの特徴量抽出手法でそれぞれ学習を行い，どの特徴量抽出手法を用いたモデルが適切なベース音源を生成することができるかを比較する．評価基準として，予測音源と元音源との基本周波数の推定の重なり具合をグラフで表示するとともに，その一致度を割合を正解率として数値で表示させる．

どの条件下でもクロマグラムを用いたモデルが最も精度が高かった．だが，正解率の平均値を見てみると，8 割を超える値になったものは一つもなく，適切なベース音源を生成することができているかという部分はまだ達成できていなかった．ただ，クロマグラムを用いたモデルに関しては基本周波数の推定結果から，オクターブのずれも正解だと考慮した場合，モデルの平均正解率はかなり上がった．今後

の課題としては，今回の実験で用いた音源はかなり条件付けているので，多種多様な音源に対応できるようにデータセットの数を増やし，学習精度を上げていく．他にも，特徴量抽出手法の設定条件を変更して，どの条件が最も精度が高くなるかの検証を行う必要がある．

目　　次

図 目 次

表 目 次

第1章　序　　論

1.1　本研究の背景

ギターは，軽音楽において中心的な楽器の 1 つである．そのため，多くのアマチュアギタリストが存在し，ギターを弾きながら作曲を楽しむ者も少なくない．しかし，作曲した楽曲をバンドで演奏するには，ベースやドラムスなど各楽器パートの演奏内容を決める作業（編曲）が必要になる．編曲を行うには，各楽器の特性を知る必要があったり，編曲した結果を人に伝えるには楽譜に書くか DTM (desktop music) を用いる必要があるため，簡単にできるものではない．

そこで知識いらずで自動的にバンド編曲を行うシステムがあれば，各楽器の特性や楽譜に書く，DTM を用いるなどの工程をせずに編曲を楽しむことができ，自分の音楽の幅を広げることが可能になる．

1.2　本研究の目的

本研究が目指すのは，ユーザが作曲した楽曲の伴奏がギター 1 本で与えられたときに，他の楽器パートの演奏内容を自動で決めて、バンドで演奏できるようにするシステムの実現である．他の楽器パートとしては，ベース，ドラムス，キーボードなどが考えられるが，本稿ではベースのみを扱うこととする．

第2章　関連研究

2.1　CNNを扱った音楽研究

（研究の背景，目的，従来研究との違いなどを，過去の論文を引用しながら述べる）

2.1.1　How Low Can You Go?　Reducing Frequency and Time Resolution in Current CNN Architectures for Music Auto-tagging[1]

小さいフィルターを何重にも重ねて畳み込みを行う，VGG-CNNと，音楽的情報をうまく学習するために，サイズの異なるフィルターを多く含んでいる，MUSICNNの二つを用いて，音源ファイルをメルスペクトログラムに変換する際に，メルスペクトログラムの周波数帯域と，フレームレートを下げたものを学習させた結果の精度比較を行っている．サンプリング周波数や，ホップサイズの大きさを変えながらそれぞれの結果をクラス分類評価として，ROC_AUCと，PR_AUCの値から比較している．どちらの結果も大きな差はなく，メルスペクトログラムの周波数帯域と，フレームレートを下げても，精度を保つことができた．

2.1.2 Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification [2]

音源ファイルに対して，メルスペクトログラムを用いて，音楽ジャンルの特定の精度が人間の耳と同等の精度になるような CNN の構築を行っている．人間の耳が約 3 秒間の音声で音楽のジャンルを 70% 当てれることに対して，通常の CNN の精度では，そのレベルまで達成できていないことを，音声をメルスペクトログラムに変換して，人間の耳に近い状態のデータでの学習を行っている．結果として人間と同程度の精度まで引き上げることに成功している．

2.1.3 CMKD: CNN/Transformer-Based Cross-Model Knowledge Distillation for Audio Classification[3]

CNN と AST を用いて，音楽分類に関して，どちらのモデルが精度の高いモデルを作れる教師あり学習を行えるかの比較を行っている．片方の学習済みモデルを教師として，もう片方のモデルの学習を行うことで，学習結果の精度比較を行っている．結果としてお互いに教師モデルの精度よりも上回る学習モデルを作ることができた．

2.1.4 音響信号からのベース奏者の認識に関する研究 [4]

CNN を用いて，あるベース演奏者のもつ音の選び方，音の強弱，リズムといった演奏特徴を模したアドリブ演奏音源を自動生成するシステムの検討について述べている．データセットとして，実際のアーティストの楽曲を音源分離を用いて，ベース音源を抽出させている．データセットの音源を 5 秒ずつメルスペクトログラムに変換していき，RGB 画像として保存させている．今回の論文内では，実際の実験を行っておらず，手法に関してのみ述べていた．

2.2 ピアノを主軸にしている編曲研究

2.2.1 演奏者の好みを反映した自動編曲システム [6]

編曲に関して専門的な知識を持たないユーザーを対象に，自分の好みを反映できるピアノ用の自動編曲システムである．ニューラルネットワークを用いて，楽曲特徴量と感性語対の関係を学習することで，個人の好みを反映できる自動編曲システムの構築を行った．被験者実験による評価より，本システムの有用性は示すことはできたものの，個人の技量を反映した編曲を行うことはできなかった．

2.2.2 ユーザの技術に合わせた自動編曲機能をもつピアノ演奏練習システム [7]

演奏楽曲の楽譜とユーザによる実際の演奏データを入力とした，奏者の技術に合わせたピアノ演奏練習システムである．ユーザの演奏の音響分離と，実演奏との時間的ずれを除去するために楽譜との同期を行い，それぞれのピアノロールの比較することで，ユーザの弾き間違えた箇所を認識させ，弾き間違えた箇所周辺の楽譜から予め決めてあるパターンに沿って，楽譜の簡略化を行っている．

2.2.3 Piano Arrangement System Based on Composers' Arrangement Processes[8]

総譜からソロピアノ譜に変換する際に生じる，演奏不可能な箇所や初心者にとって難しい箇所を警告し，編曲の手助けを行うシステムである．実際に作曲家が編曲している様子を撮り，その様子をシンクアラウンド分析をすることで，どのような条件下でどの編曲を行うべきかを判断させている．

2.2.4 Automatic Piano Reduction from Ensemble Scores Based on Merged-Output Hidden Markov Model[10]

アンサンブル譜からソロピアノ譜に自動編曲を行うシステムである．演奏難易度の制約を考慮した，アンサンブル譜に対する忠実度の最適化問題に基づいて，隠れマルコフ法を用いて，ソロピアノ譜への編曲を行っている．演奏難易度に関しての定量化は隠れマルコフ法を用いて，運指を考慮することで，アンサンブル譜に対する忠実度は，音符の数を考慮することで実現している．システムとして単純な制約のもとでの演奏難易度の考慮はできたが，出来上がった譜の中に演奏不可能な箇所や，音楽的に不十分な声部が存在し，さらなる改良が必要だった．

2.3 ギターを主軸にしている編曲研究

2.3.1 与えられたメロディーとコード進行に基づくギター用編曲システムの構築 [5]

メロディーとそれに対応するコード進行を入力として，ギターで演奏できるように自動的に編曲を行うシステムである．それぞれの弦のフレット番号に弾きやすさの点数を与え，その合計点を評価指標として最適な調を決定させている．初心者，中級者が演奏可能かどうかを判定するためにポジションの移動距離を考慮し，移動距離が少ないポジションになるように編曲を行っている．被験者実験による評価より，市販の楽譜と編曲譜の差には大きな違いはなく，本システムの有効性を示すことができた．

2.3.2 Song2Guitar: A Difficulty-aware Arrangement System for Generating Guitar Solo Covers from Polyphonic Audio of Popular Music[9]

ポピュラーな楽曲に対して，難易度を考慮したソロギター譜に編曲するシステムである．ポリフォニック音源からコード，メロディ，ビートを抽出し，これらのデータを隠れマルコフ法を用いて，easy，normal，hard の三つの難易度ごとにソロギター譜に変換している．被験者実験により，これらの難易度の妥当性と編曲精度の検証を行ったが，難易度に関してうまく分けられているという意見もあったが，easy が最も難しく感じるといった意見があり，またインターフェイス面でも見づらい部分があるなど，改良の余地があった．

2.4 本研究のアプローチ

従来の研究における編曲という分野は，主にピアノを主軸にしている研究が数多く存在し，総譜からソロピアノ譜，アンサンブル譜からソロピアノ譜，楽曲の簡略化や，自分好みの編曲を行う [6][7][8][10] など，幅広く存在している．ギターを主軸にしている研究では，ほとんどがソロギター譜への編曲を行うシステム [5][9] であり，バンド編曲という分野を取り扱っている研究は存在しなかった．

本研究では，CNN を用いて，ギター音響信号からベース音響信号を自動生成するモデルを提案する．ここで，オーディオ音源からオーディオ音源を返すようにしたのは，世の中には midi ギターというものが存在するが，実際に弾いた音を midi に書き起こす精度が不十分であることが多く存在する．また，対象とするユーザーがギターに関する知識しか持たないことを想定しているため，入力を midi などで作成する必要もなく，出力として楽譜などに起こしても，楽譜が読めない可能性があるため，この入出力の形にした．

第3章　提案手法

3.1　システムの流れ

本手法では，畳み込みニューラルネットワーク（CNN）を用いてギター音源からベース音源を生成する．この手法では，ギター音源とベース音源のペアデータが学習用に与えられることを前提とする．ギター音源を 0.5 秒ごとに区切り，そのスペクトログラムを CNN の畳み込み層に入力し，逆畳み込み層に与えることでベース音源のスペクトログラムを得る．ここで，0.5 秒ずつ音源を区切って学習を行う理由は，一つのコードに対する分析をより細かく学習させるために行っている．0.5 秒という短い時間の中に含まれるコードはどれだけ早いテンポの曲に対しても，一つのみであり，複数のコードを含んだ状態で学習させると，一つ一つのコードに対する特徴量がうまく学習できない可能性があるため，このような設定にした．

3.1.1　入力音源のスペクトログラムの計算

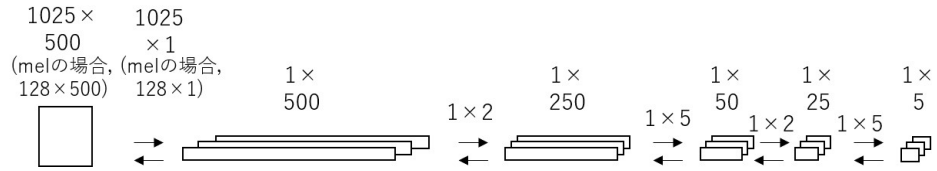
ギター音源（学習時はベース音源も）に対して，短時間フーリエ変換を用いてスペクトログラムを計算する．まず，サンプリング周波数を 22050Hz にダウンサンプリングする．次に，短時間フーリエ変換を行う．窓関数は hann 関数，窓幅は 2048，ホップサイズはサンプリング周波数の 1/1000 とした．その後，各値の絶対値を取ることで振幅スペクトログラムを得る．

3.1.2　入力音源に対する特徴抽出

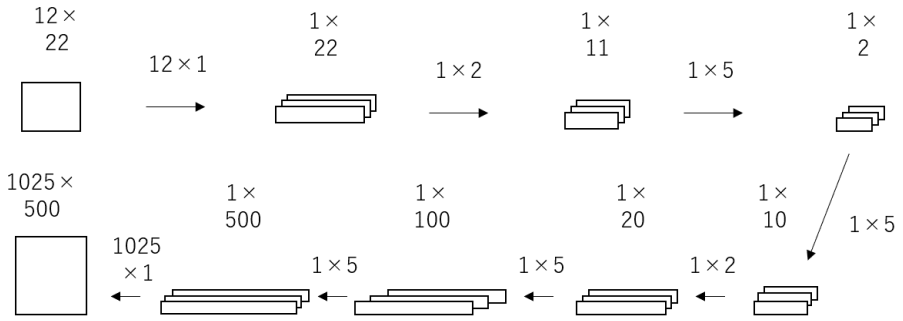
入力音源のスペクトログラムを次項で述べる CNN に入力する他，スペクトログラムから特徴抽出したものを CNN に入力する方法も試行する．抽出する特徴として，メルスペクトログラムおよびクロマグラムを用いる．ただし，クロマグラムに関してはホップサイズを 512 とした．

3.1.3　CNN によるベース音源のスペクトログラムの生成

前項までの方法で得たギター音源のスペクトログラム（またはメルスペクトログラム、クロマグラム）を畳み込み層で圧縮を行い，その後，逆畳み込み層を適用することでベース音源のスペクトログラムに変換する．モデルの概要を図 3.1 に示す．入力されるデータは形状が 1024 × 500（周波数軸：1024 要素，時間軸：500 要素）であり（メルスペクトログラムのの場合は 128 × 500，クロマグラムの場合は 12 × 22），そこから図 3.1 の畳み込み層によって 1 × 5（クロマグラムの場合は 1 × 2）に圧縮したのち，逆畳み込み層によって 1025 × 500 のスペクトログラムに変換する．各層におけるフィルタのチャネル数は 1024 とし，ストライドは 1，パディングはなし，活性化関数は ReLU とした．



(a) スペクトログラム，メルスペクトログラムを用いる場合



(b) クロマグラムを用いる場合

図 3.1: CNN モデルのアーキテクチャー．長方形の上の数値はデータの形状，矢印の上の数値はフィルタの形状を表す．右向きの矢印は畳み込み層，左向きの矢印は逆畳み込み層である．

3.1.4 位相復元による音響信号の生成

出力されたベース音源のスペクトログラムに対して逆フーリエ変換および位相復元を行うことで，ベース音源の音響信号を得る．位相復元には Griffin-Lim アルゴリズム法を用いる．反復回数は 32 回，窓幅は 2048，ホップサイズはサンプリング周波数の 1/1000 とした．得られた音響信号に対して，HPSS（調波打楽器音分離）を用いて，調波楽器音と打楽器音の分離を行い，調波楽器音の抽出を行う．これによって，生成された音響信号に含まれている雑音の分離を行う．margin に関しては最小値を 1.0，最大値を 1.5 とする．抽出された調波楽器音を量子化ビット数 16，サンプリング周波数 44100Hz で wav 音源に変換する．

第4章　評価実験

4.1　データセット

Cakewalk by BandLab を用いてギターおよびベースパートの MIDI シーケンスを入力し，同ソフトウェアに内蔵されているソフトウェア音源を用いて wav 形式に変換した．BPM は 120，小節数は 4 小節（8 秒）とした．ギター音源には外部から持ってきた sforzand，ベース音源には Cakewalk by BandLab に付属している SI-Bass Guitar を用い，どちらにもエフェクターは適用しなかった．コード進行は 1 小節あたり 1 コードとし，メジャースケールに基づいて決定した．ベースは各コードのルート音とした．ギターおよびベースパートのリズムは八分音符とした．この基準に基づいてギター音源およびベース音源のペアを 16 種類作成した．また，同様なコード進行であるが，抑える箇所が違うボイシングを行ったギター音源を 4 種類作成した．作成したもの一例を図 4.1 に示す．このうち，10 個を学習用に，10 個をテスト用に割り当てた．

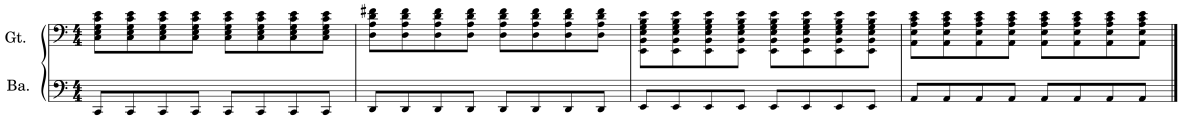


図 4.1: 作成したギター，ベースのスコアの例

表 4.1: 学習データのコード進行リスト

コード進行
CDEmAm
C♯D♯FmA♯m
DEF♯mC♯m
EABmC♯m
FCAmG
F♯G♯C♯D♯m
GABmD
ABC♯mG♯m
A♯FGmCm
A♯CDmEm

4.2　実験条件

本実験において注目すべきは，入力されるギター音源の音響的特徴などが学習時に用いたものからどの程度異なっていても適切にベース音源の生成かかどうかである．そこで，次の 3 つの条件を設定した．

条件 1 学習データとテストデータとでコード進行が異なるものの，用いるソフトウェア音源や音響的条件が同じものと，学習データとテストデータとでコード進行が同じであるが，ボイシングを用いており，用いるソフトウェア音源や音響的条件が同じ。

条件 2 学習データとテストデータとでコード進行が異なるものと，学習データとテストデータとでコード進行が同じであるが，ボイシングを用いているもの．また，用いるソフトウェア音源は学習データとテストデータで同じであるも

のの，学習データにだけローパスフィルタ（設定：1 オクターブ上がるごとに-3db する）をかけた。

条件 3 学習データは前節の方法で作成したものであるが，テストデータは第 1 著者が本物のギターで演奏したもの。演奏の録音には M-Audio 社の M-Track を用いた.

本来であれば，ギターやベースパートのリズムなどにも変化を付けるべきであるが，今後の課題とした.

生成されたベース音源の評価は，フレームごとに正解音源との音高の差を求め，差が 50cent 以内のときに正解とみなしたときの，正解率を用いて行う.

4.3　実験結果

表 4.2: モデルごとの正解率

条件	テストデータ	STFT	mel	chroma
条件 1	A♯CDmEm_voicing	0.37	0.32	0.66
	EABmC♯m_voicing	0.39	0.20	0.64
	CDEmAm_voicing	0.39	0.49	0.53
	GABmD_voicing	0.55	0.54	0.62
	GCDEm	0.58	0.56	0.62
	CDmEmDm	0.42	0.21	0.17
	DmEmAmEm	0.57	0.40	0.32
	EmAmFG	0.59	0.39	0.81
	AmFGC	0.70	0.54	0.79
	FAmGDm	0.58	0.35	0.63
	平均	0.51	0.40	0.58
条件 2	A♯CDmEm_voicing	0.29	0.26	0.58
	EABmC♯m_voicing	0.22	0.26	0.62
	CDEmAm_voicing	0.17	0.30	0.51
	GABmD_voicing	0.28	0.49	0.67
	GCDEm	0.24	0.31	0.52
	CDmEmDm	0.15	0.11	0.23
	DmEmAmEm	0.23	0.17	0.22
	EmAmFG	0.35	0.26	0.68
	AmFGC	0.17	0.35	0.76
	FAmGDm	0.41	0.42	0.58
	平均	0.25	0.29	0.54
条件 3	CDEmAm_Audio	0.20	0.09	0.35

テストデータの名称はコード進行を表す．学習に用いたコード進行と同じだがヴォイシングを変更したものは「_voicing」を付与した．

4.4 考　　察

これらの実験条件の結果から考察を行う.

実験条件 1

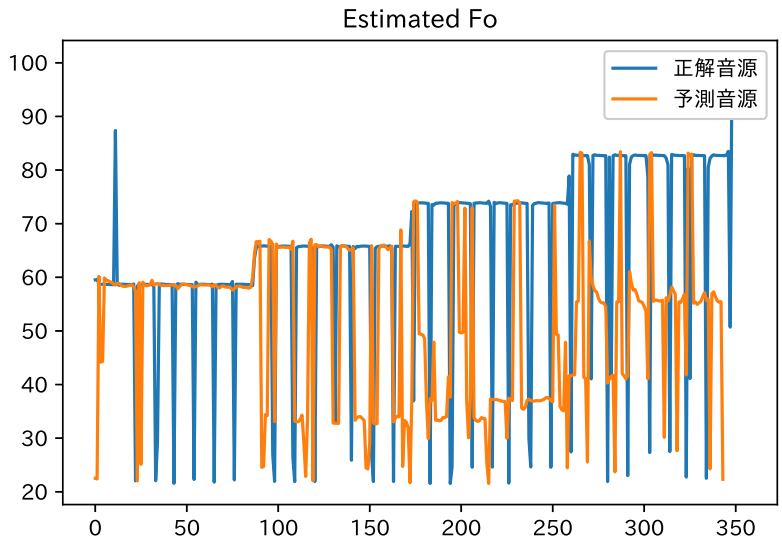


図 4.2: STFT での A♯CDmEm_voicing の基本周波数推定結果

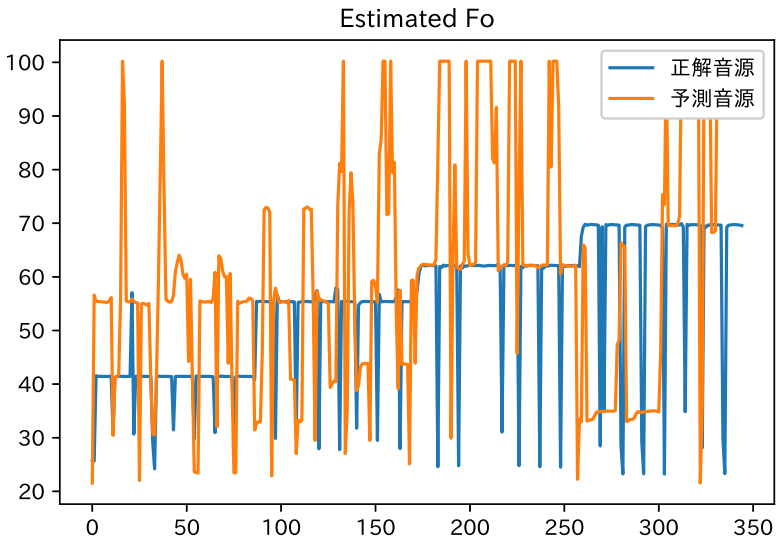


図 4.3: mel での EABmC♯m_voicing の基本周波数推定結果

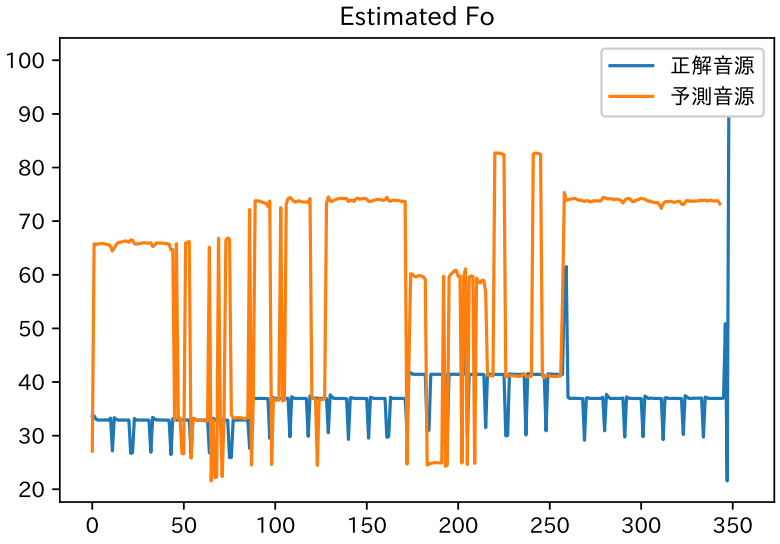


図 4.4: chroma での CDmEmDm の基本周波数推定結果

表 4.2 より，実験条件 1 において，それぞれのモデルでの正解率の最低値であった予測結果に関して見ていくと，STFT モデルでの A♯CDmEm_voicing の結果である図 4.2 では，最初の A♯ のところはあっているが，他のコードの部分では，ほとんど正解音源に対してずれている．だが，ずれている箇所を見ていくと，正解音源の基本周波数に対して，約 1/2 になっている箇所も存在し，実際にオクターブずれを考慮した際の正解率を見ると，0.55 となり，0.18 も上がったため，オクターブずれが生じていた．

mel モデルでの EABmC♯m_voicing の結果である図 4.3 では，予測音源の基本周波数が全体的に正解音源に対してずれていた．オクターブずれの可能性があると言えるような箇所もほとんどなく，基本周波数も一定になることがなく，常に変動していた．オクターブずれを考慮した際の正解率は 0.29 となり，0.09 上がったが，精度としては 3 割を切っているため，適切なベース音源を生成することはできていなかった．

chroma モデルでの CDmEmDm の結果である図 4.4 では，かなり大幅にずれが生じていたが，図 4.2 や図 4.3 と比べ，基本周波数が常に変動しているわけではなく，一定にはなっている．また，推定結果から見て，正解音源の基本周波数に対して，約 1/2 になっている箇所が多く存在し，実際にオクターブずれを考慮した際の正解率を見ると，正解率が 0.77 となり，0.60 も上がったため，オクターブずれがかなり存在していた．

それぞれのモデルの平均を見ていくと，chroma が最高値で 0.58，mel が最低値で 0.40 となっており，mel に関しては，予測音源の半分以上が実際のベース音源とずれていることが分かった．

実験条件 2

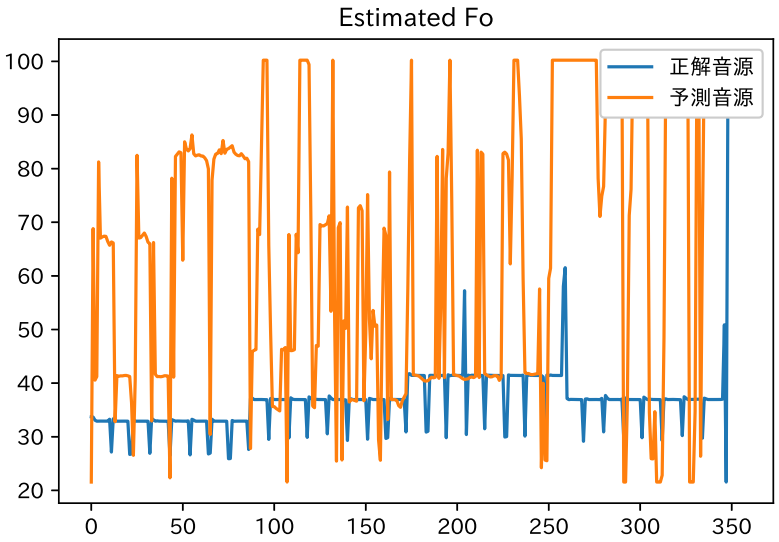


図 4.5: STFT での CDmEmDm の基本周波数推定結果

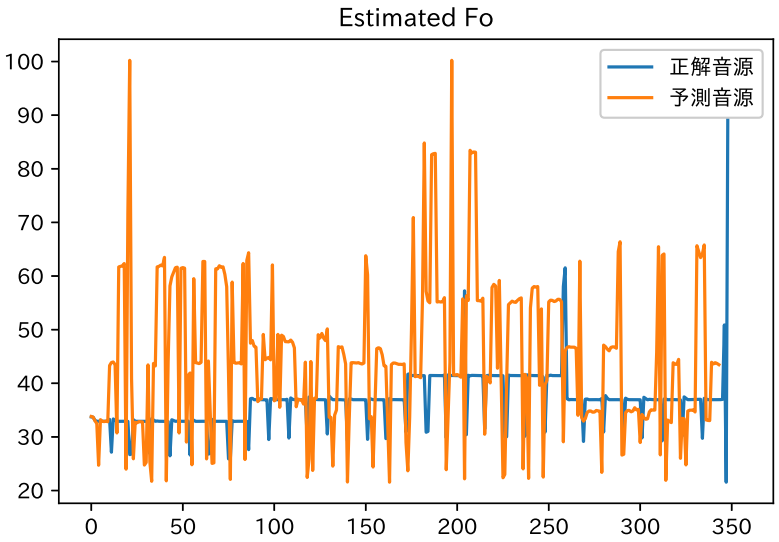


図 4.6: mel での CDmEmDm の基本周波数推定結果

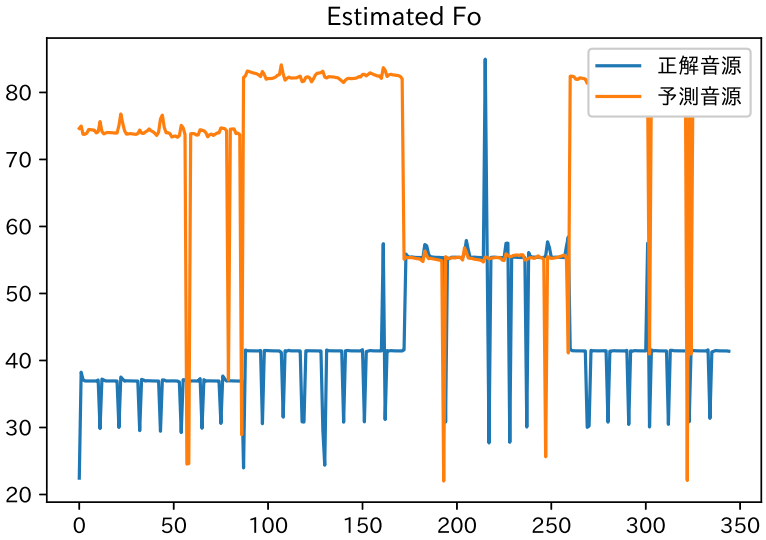


図 4.7: chroma での DmEmAmEm の基本周波数推定結果

表 4.2 より，実験条件 2 において，それぞれのモデルでの正解率の最低値であった予測結果に関して見ていくと，STFT モデルでの CDmEmDm の結果である図 4.5 では，全体的に常に基本周波数が変動しており，正解音源と一致する箇所も一瞬しかなかった．オクターブずれを考慮した際の正解率を見ると，0.20 となり，0.50 上がったが，それでも 2 割しか正解音源と一致せず，適切なベース音源を生成することはできていないと考えられる．

mel モデルでの CDmEmDm の結果である図 4.6 では，図 4.5 よりも基本周波数の振れ幅は小さいが，変わらず常に変動していて，正解音源と一致する箇所も一瞬しかなかった．オクターブずれを考慮した際の正解率を見てみると，変わらず 0.11 となり，オクターブずれは存在せず，予測自体がうまくいっていないと考えられる．

chroma モデルでの DmEmAmEm の結果である図 4.7 では，条件 1 の図 4.4 同様にかなり大幅にずれが生じていたが，基本周波数が常に変動しているわけではなく，一定にはなっている．また，推定結果から見て，正解音源の基本周波数に対

して，約 1/2 になっている箇所が多く存在し，実際にオクターブずれを考慮した際の正解率を見ると，0.85 となり，0.63 も上がったため，オクターブずれがかなり存在していた．

それぞれのモデルの平均値を見ていくと，chroma が最高値で 0.54，STFT が最低値で 0.25 となっており，どのモデルもローパスフィルタをかけ，加工を行うと正答率が低くなった．

実験条件 3

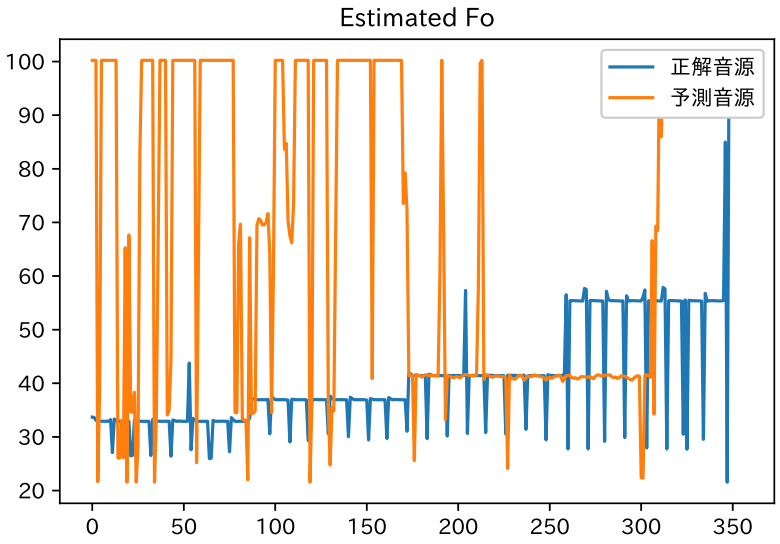


図 4.8: STFT での CDEmAm の基本周波数推定結果

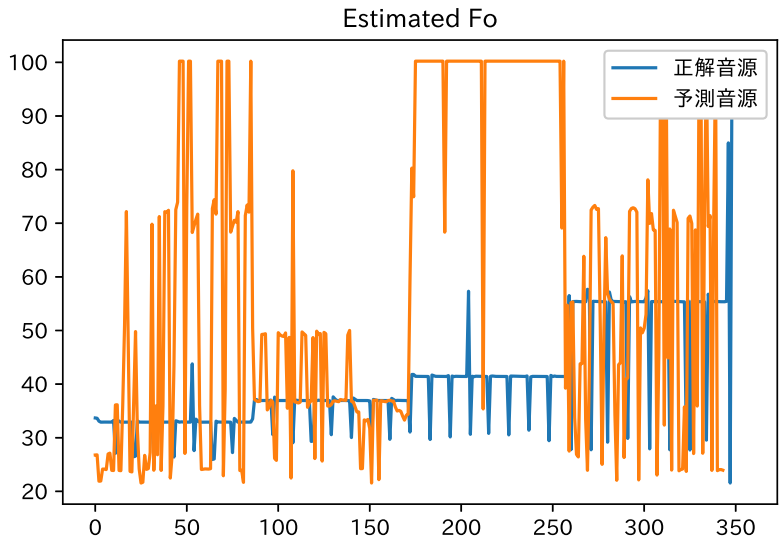


図 4.9: mel での CDEmAm の基本周波数推定結果

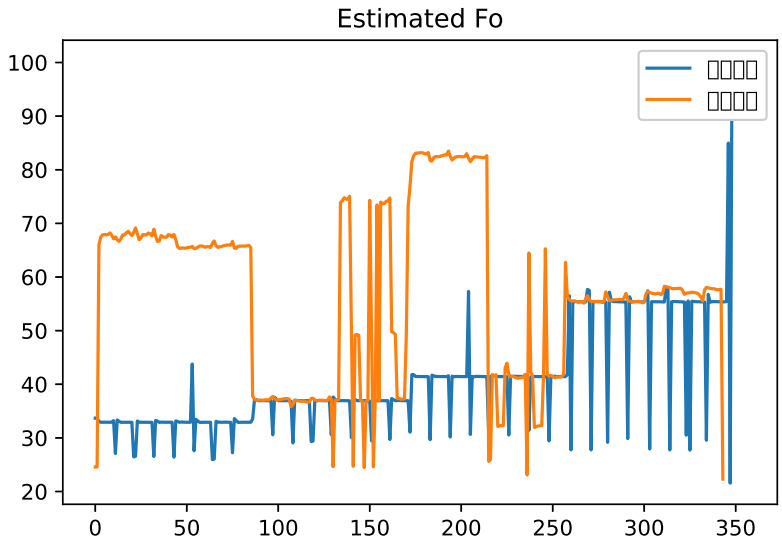


図 4.10: chroma での CDEmAm の基本周波数推定結果

表 4.2 より，実験条件 3 において，それぞれのモデルの予測結果を見ていくと，STFT モデルでの CDEmAm の結果である図 4.8 では，基本周波数が 100Hz になっている箇所が多々あるが，これは基本周波数推定の最大値を 100Hz と定めたために，100Hz 以上の基本周波数だった場合に上に張り付く形になってしまった．Em のところで一致するようになっていたが，それ以外が正解音源に対してかなりずれていて，オクターブずれの可能性もなかった．実際にオクターブずれを考慮した際の正解率を見ると，0.21 となり，0.01 上がったが，オクターブずれを考慮しても大した差は生じなかった．

mel モデルでの結果である図 4.9 では，全体的に基本周波数が常に変動していて，正解音源と一致する箇所も瞬間的にしか存在しなかった．オクターブずれを考慮した際の正解率を見ると，0.11 となり，0.02 上がったが，それでも 2 割を切っているため，適切なベース音源を生成することはできていないことが分かる．

chroma モデルでの結果である図 4.10 では，他の結果とは異なり，基本周波数が常に変動していることはなく，安定した音を生成することができている．また，C や Em の序盤半分のところは，正解音源の基本周波数に対して，約 2 倍の位置に予測結果の基本周波数が位置しているため，オクターブずれの可能性があるように考えられる．実際にオクターブずれを考慮した際の正解率を見ると，0.69 となり，正解率が 0.34 も上がり，オクターブずれがかなり存在していた．

それぞれのモデルの正答率を見ていくと，最も正解率の高かったモデルは，chroma で，最も正解率が低かったモデルは，mel だった．特に mel のモデルの予測結果でのセントの値は低く，正解率が一割を切るほどの結果となった．だが，どのモデルの正解率を見ても，半分は切っており，適切なベースが生成されたとは言い難い結果になった．

総合考察

それぞれのモデルの平均正解率を見ていくと，すべての条件下で chroma が最も正解率が高くなった．また，chroma は条件 2 のような音源に対して，加工をかけてもそこまで正解率が下がることはなかった．他のモデルでは，条件 2 のような加工をかけると精度が落ちていき，特に STFT に関しては条件 1 に対して，条件 2 は正解率が約 1/2 程下がってしまい，加工された音源に対しての予測がうまくいかないことが分かった．

また，オクターブがずれている部分も正解とみなした際の正解率では，chroma が最も多く，オクターブずれが存在し，条件 1 の結果が 0.82，条件 2 の結果が 0.81 となった．同様に STFT では条件 1 の結果が 0.65，条件 2 の結果が 0.33，mel は条件 1 の結果が 0.47，条件 2 の結果が 0.37 となり，オクターブずれを考慮しなかった平均正解率よりも上がった．だが，chroma のような大きな正解率の変化は生じず，オクターブずれを考慮しても正解率は 7 割を超えることはなく，適切なベース音源を生成するための特徴量抽出手法として適しているとは言えない結果となった．chroma がうまくいった理由には，クロマグラムは和音分析に特化していて，コード推定で用いられる特徴量抽出手法であるため，今回のギターのコード進行からベース音源を予測するという学習の形に適していたと考えられる．

第5章　結　　論

5.1　結論

本稿では，畳み込みニューラルネットワーク（CNN）モデルを用いて，ギター音源からベース音源を生成する手法を提案した．その中でどのように音源をデータに変換するかによってのモデル精度の差について調べた．今回の実験結果では，どの条件結果でもクロマグラムを用いたモデルの精度が最も良く，メルスペクトログラム，STFT を用いたモデルの精度はどちらもローパス音源，オーディオ音源に対して，かなり下がることが分かった．ただ，どのモデルも 8 割近い正答率をもったモデルは存在しないため，適切なベース音源を生成することができているとは言いきれない結果になった．結果から，クロマグラムのような和音分析に特化している特徴量抽出手法の方が，オーディオ音源の分析に関して適しているということが分かった．

5.2　今後の課題

今後の課題としては，今回の実験のデータセットは，エフェクターなどの加工を施していないもののみを取り扱うようにしていたため，今後，加工された音源に対して，クロマグラムでの特徴量抽出がうまくいくのかどうかが課題となると考えられる．また，基本周波数の推定結果から，オクターブずれの可能性がある予測音源が多々存在していたので，オクターブずれの可能性を考慮した場合は，どのモデルも今回の結果より多少，正答率は上がると考えられる．今後の課題として

は，BPM，ビートなどのデータセットの設定を変更していき，データセットを増やした上で学習を行うとともに，それぞれの特徴量抽出手法の設定条件を変更し，どの条件下が最も精度が高くなるのかを検証していく．また，音源を0.5秒ずつ区切っていることの妥当性を検証するために，区切る時間を1秒，2秒など伸ばした結果の比較も行っていきたい．予測音源に対して，聞いてみてどうだったかの主観評価などを今回は行っていないため，今後は主観評価でも検証していきたい．

参考文献

[1] Andres Ferraro, Dmitry Bogdanov, Xavier Serra, Jay Ho Jeon, Jason Yoon:“How Low Can You Go? Reducing Frequency and Time Resolution in Current CNN Architectures for Music Auto-tagging”, EUSIPCO2020, 2020.

[2] Mingwen Dong:“Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification”, 2018.

[3] Yuan Gong, Sameer Khurana, Andrew Rouditchenko, James Glass:“CMKD: CNN/Transformer-Based Cross-Model Knowledge Distillation for Audio Classification”, IEEE, 2022.

[4] 西村 奈那子:“音響信号からのベース奏者の認識に関する研究”, 情報処理学会第 83 回全国大会, pp.2-243–244, 2021.

[5] 丸山剛志, 三浦雅展, 柳田益造:“与えられたメロディーとコード進行に基づくギター用編曲システムの構築”, 第 3 回情報科学技術フォーラム, pp.399–400, 2004.

[6] 島田 貴宏, 鬼沢 武久:“演奏者の好みを反映した自動編曲システム”, 第 7 回情報科学技術フォーラム, pp.203–204, 2008.

[7] 福田翼, 池宮由楽, 糸山克寿, 吉井和佳:“ユーザの技術に合わせた自動編曲機能をもつピアノ演奏練習システム, 情報処理学会第 77 回全国大会, pp.2-402–403, 2015.

[8] Onuma Sho, Hamanaka Masatoshi:“Piano Arrangement System Based on Composers ’ Arrangement Processes”, ICMC, 2010.

[9] Shunya Ariga, Satoru Fukayama, Masataka Goto:“Song2Guitar: A Difficulty-aware Arrangement System for Generating Guitar Solo Covers from Polyphonic Audio of Popular Music”, ISMIR, pp568–574, 2017.

[10] Eita Nakamura, Sigeki Sagayama：“Automatic Piano Reduction from Ensemble Scores Based on Merged-Output Hidden Markov Model”, ICMC, pp.298–305, 2015.

謝　　辞

北原鉄朗教授には，本研究を進めるにあたり，手厚いご指導をいただきました．深く感謝いたします．

付 録A　基本周波数推定結果

本章では，実験条件 1，2 のそれぞれの基本周波数の推定結果を示す.

A.1　実験条件 1

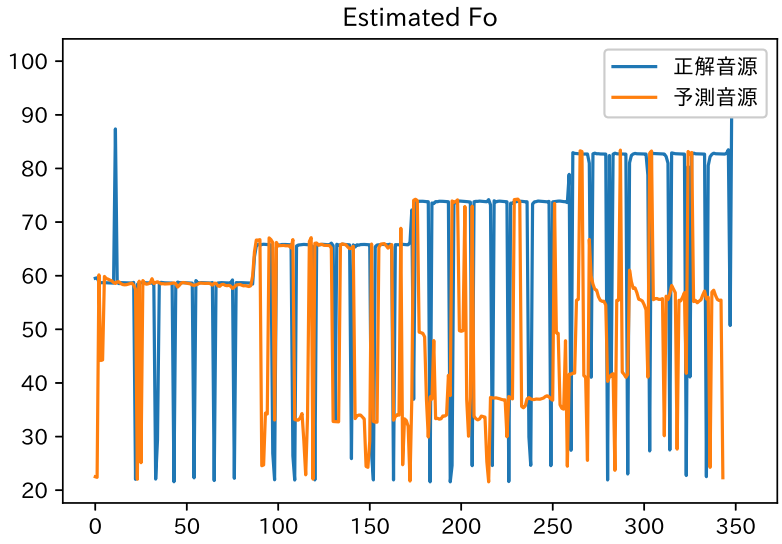


図 A.1: STFT での A♯CDmEm_voicing の基本周波数推定結果

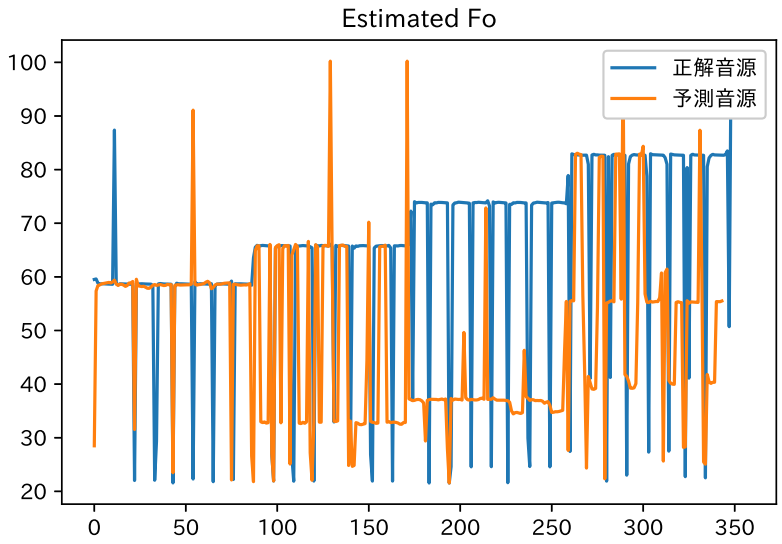


図 A.2: mel での A♯CDmEm_voicing の基本周波数推定結果

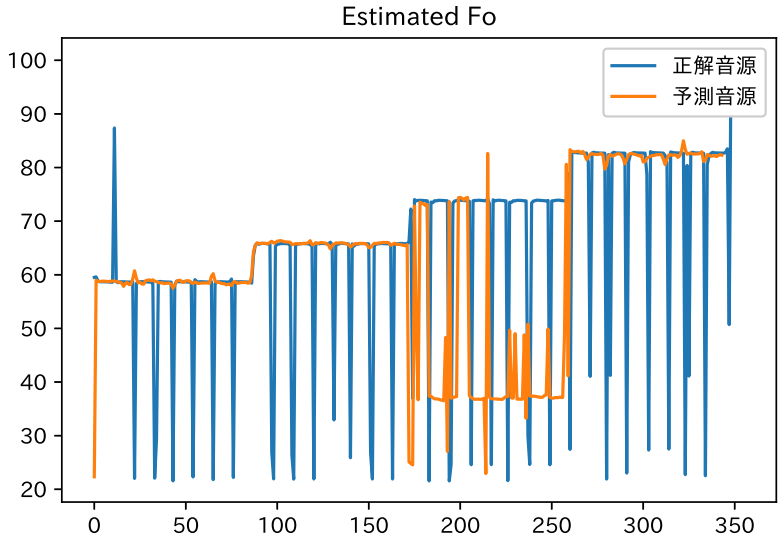


図 A.3: chroma での A♯CDmEm_voicing の基本周波数推定結果

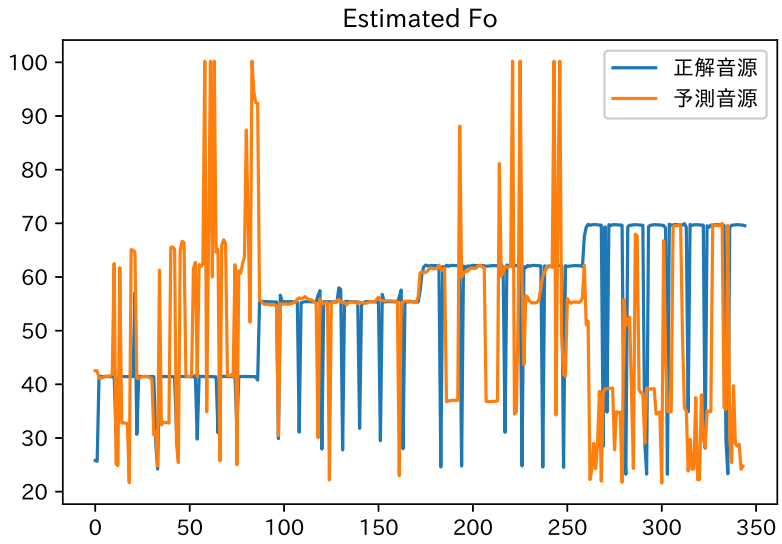


図 A.4: STFT での EABmC♯m_voicing の基本周波数推定結果

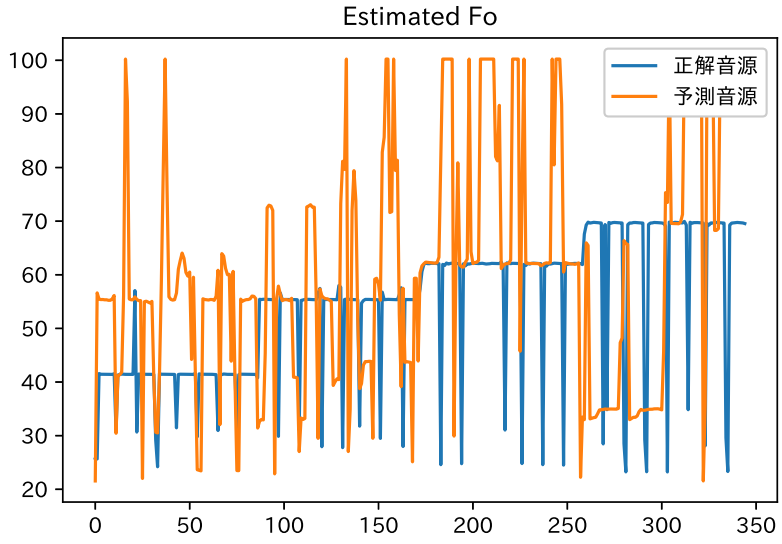


図 A.5: mel での EABmC♯m_voicing の基本周波数推定結果

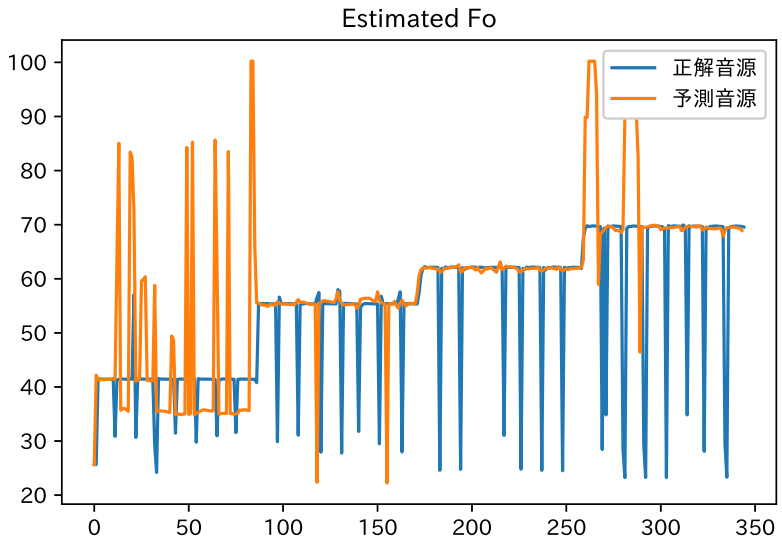


図 A.6: chroma での EABmC♯m_voicing の基本周波数推定結果

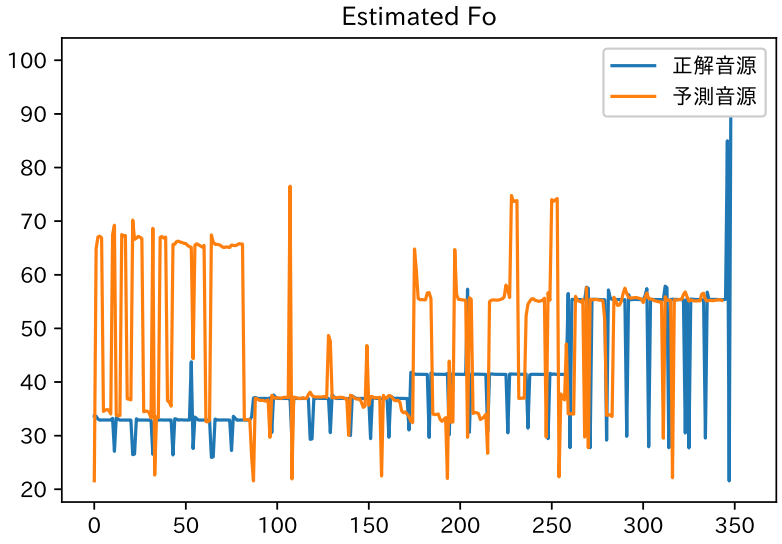


図 A.7: STFT での CDEmAm_voicing の基本周波数推定結果

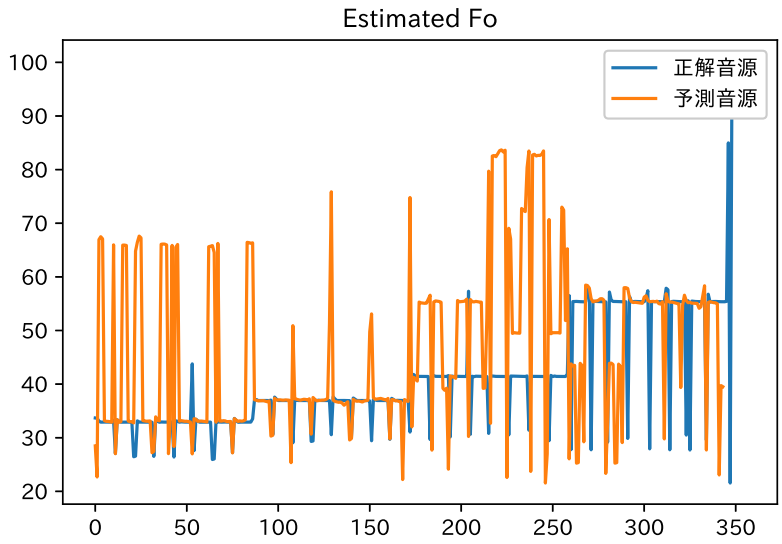


図 A.8: mel での CDEmAm_voicing の基本周波数推定結果

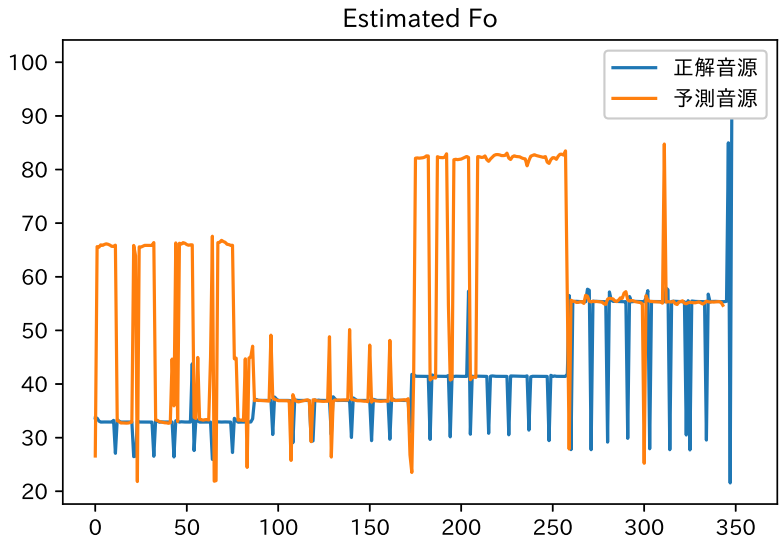


図 A.9: chroma での CDEmAm_voicing の基本周波数推定結果

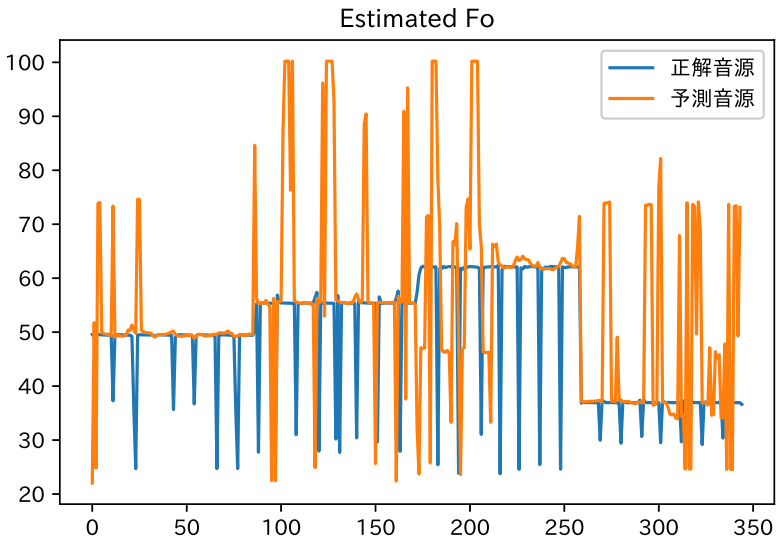


図 A.10: STFT での GABmD_voicing の基本周波数推定結果

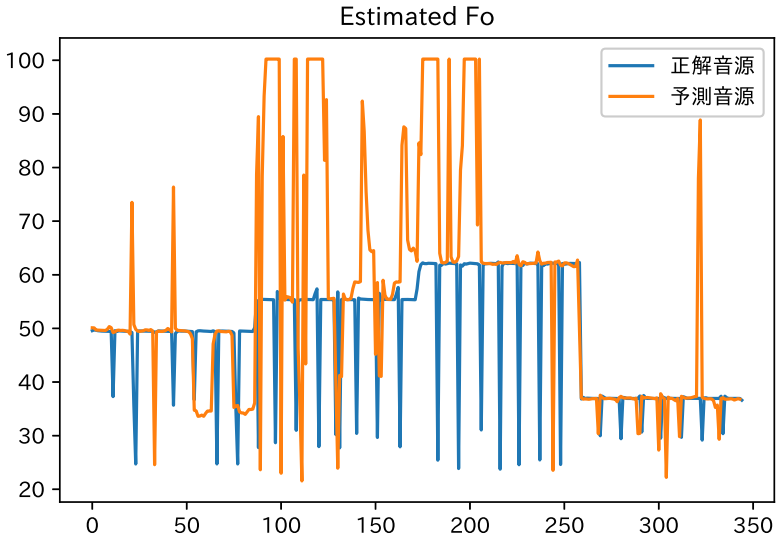


図 A.11: mel での GABmD_voicing の基本周波数推定結果

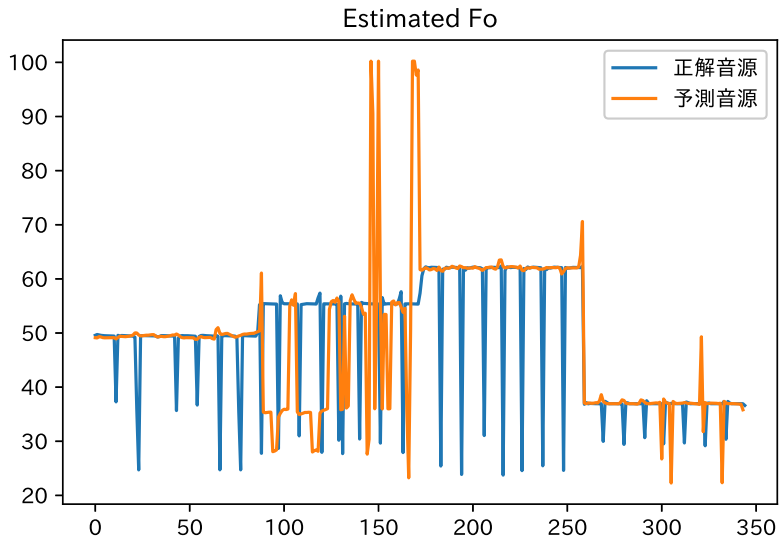


図 A.12: chroma での GABmD_voicing の基本周波数推定結果

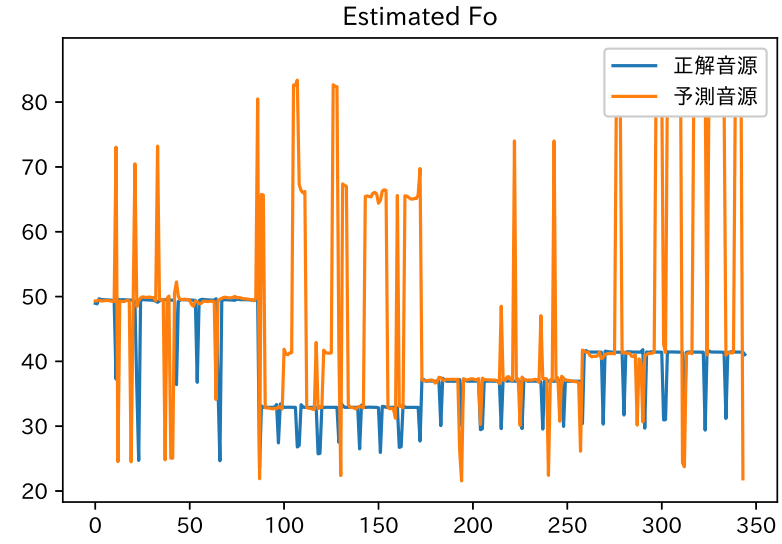


図 A.13: STFT での GCDEm の基本周波数推定結果

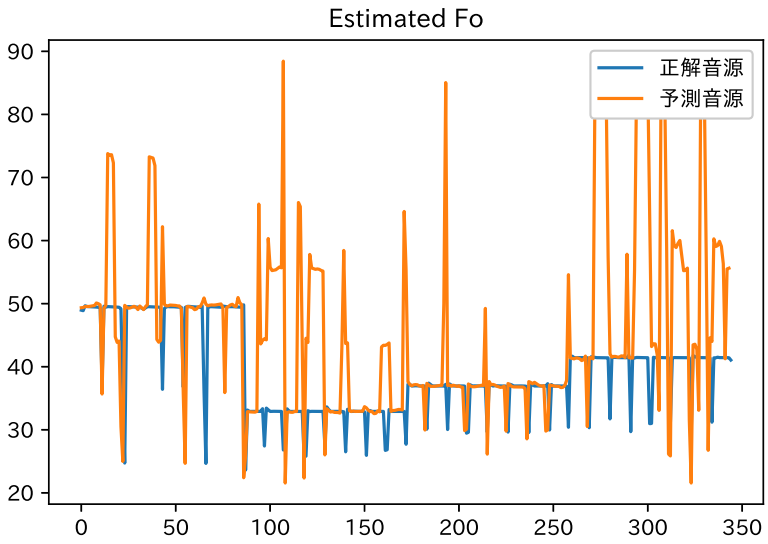


図 A.14: mel での GCDEm の基本周波数推定結果

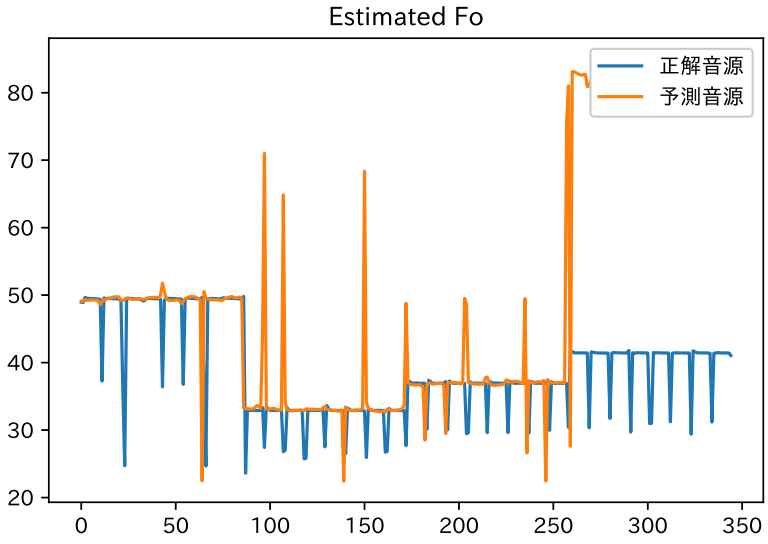


図 A.15: chroma での GCDEm の基本周波数推定結果

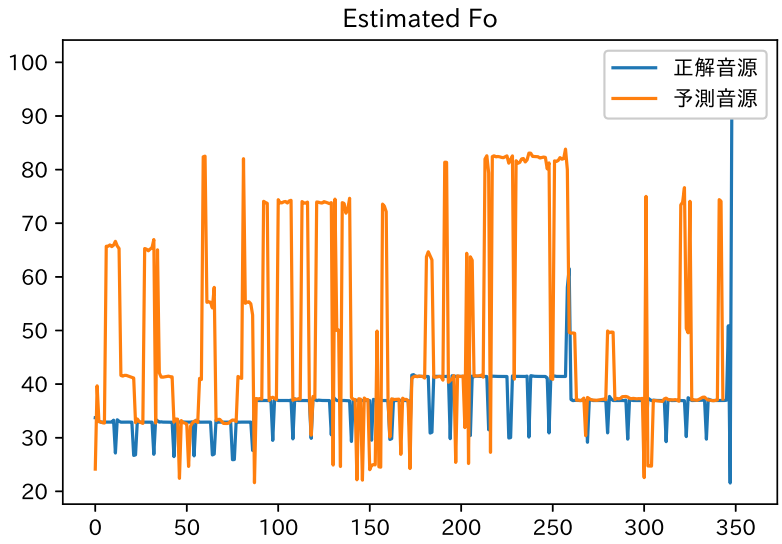


図 A.16: STFT での CDmEmDm の基本周波数推定結果

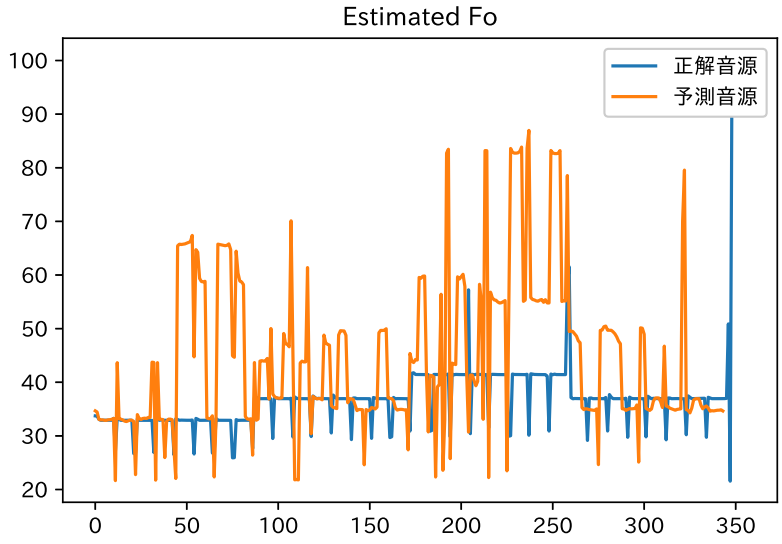


図 A.17: mel での CDmEmDm の基本周波数推定結果

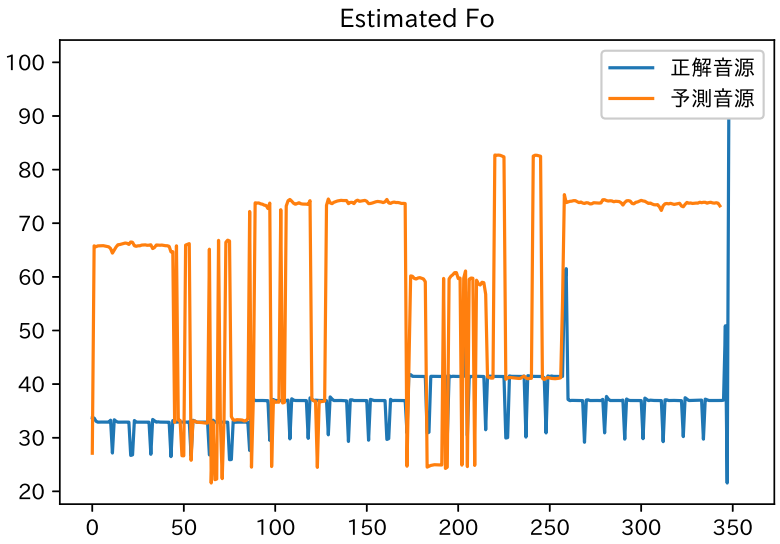


図 A.18: chroma での CDmEmDm の基本周波数推定結果

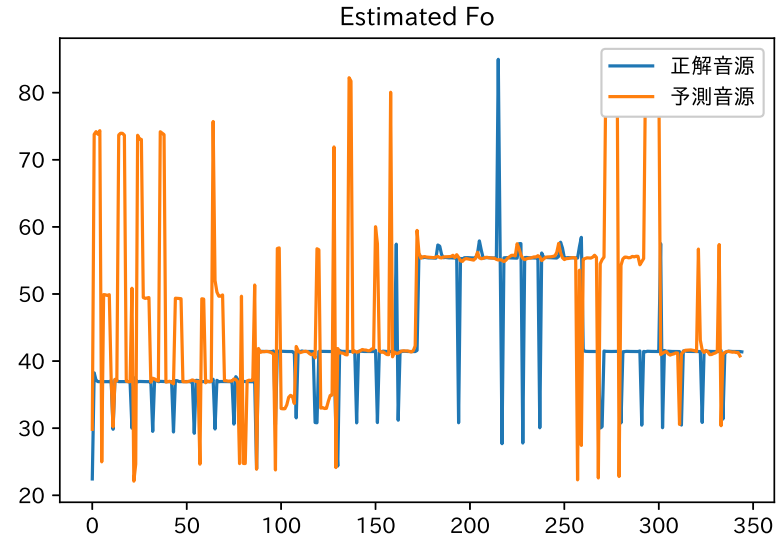


図 A.19: STFT での DmEmAmEm の基本周波数推定結果

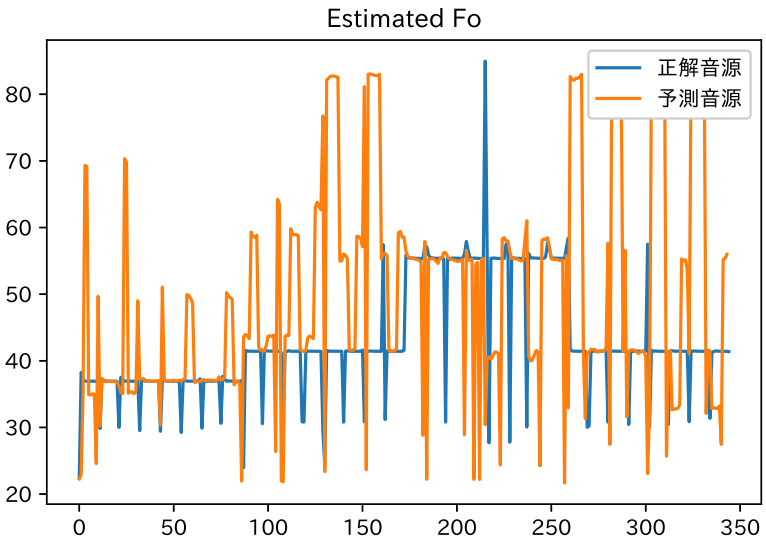


図 A.20: mel での DmEmAmEm の基本周波数推定結果

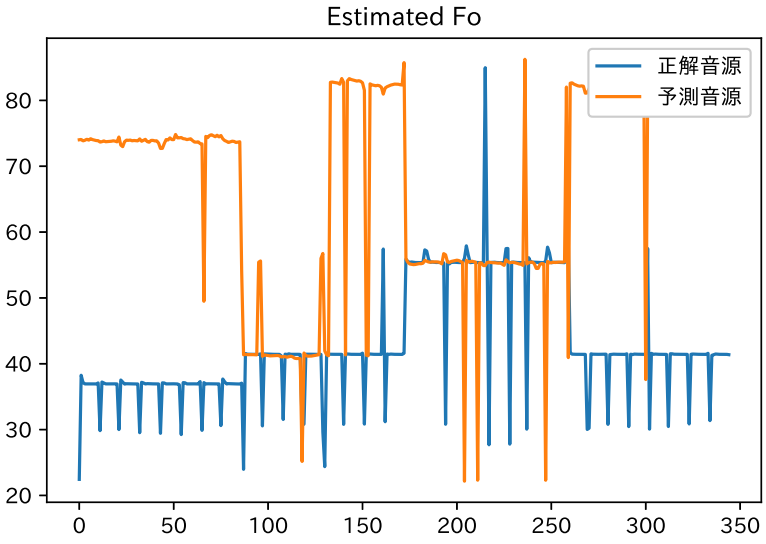


図 A.21: chroma での DmEmAmEm の基本周波数推定結果

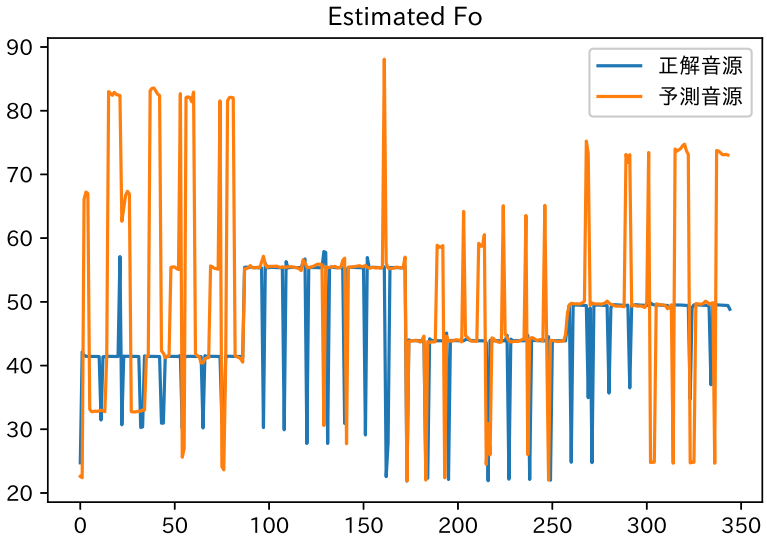


図 A.22: STFT での EmAmFG の基本周波数推定結果

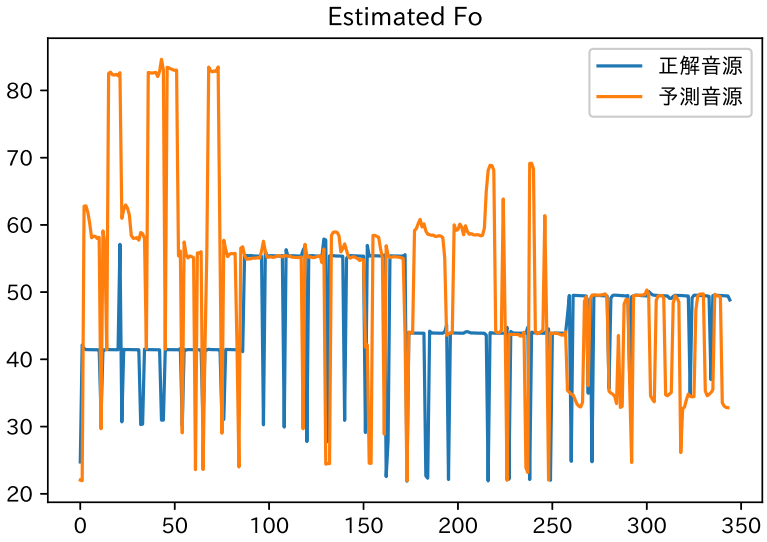


図 A.23: mel での EmAmFG の基本周波数推定結果

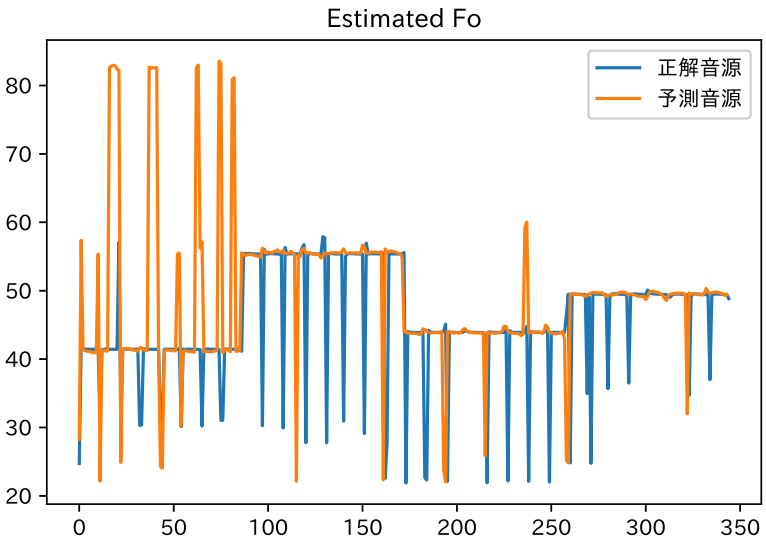


図 A.24: chroma での EmAmFG の基本周波数推定結果

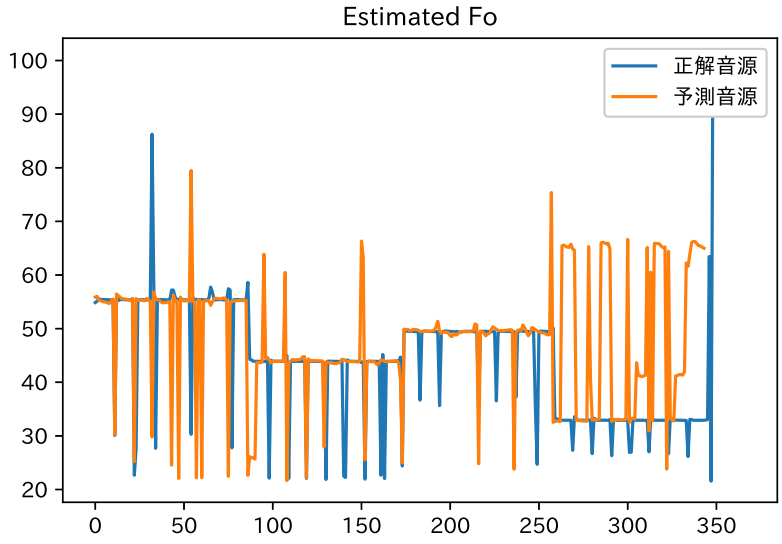


図 A.25: STFT での AmFGC の基本周波数推定結果

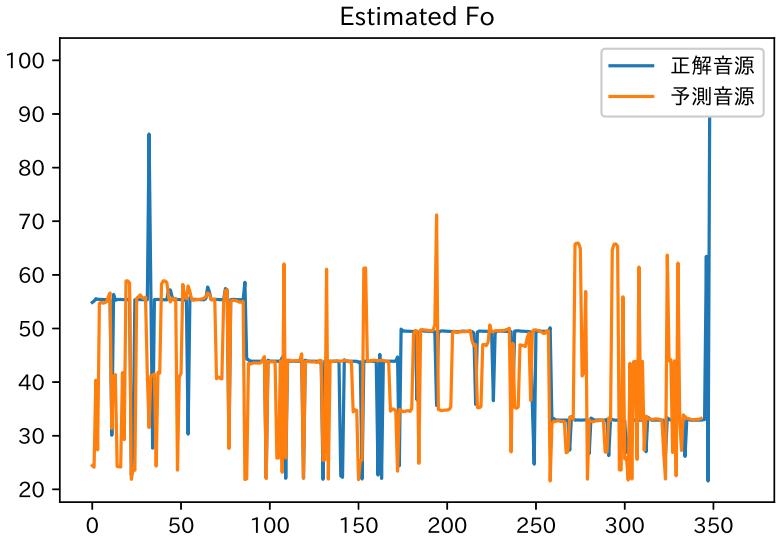


図 A.26: mel での AmFGC の基本周波数推定結果

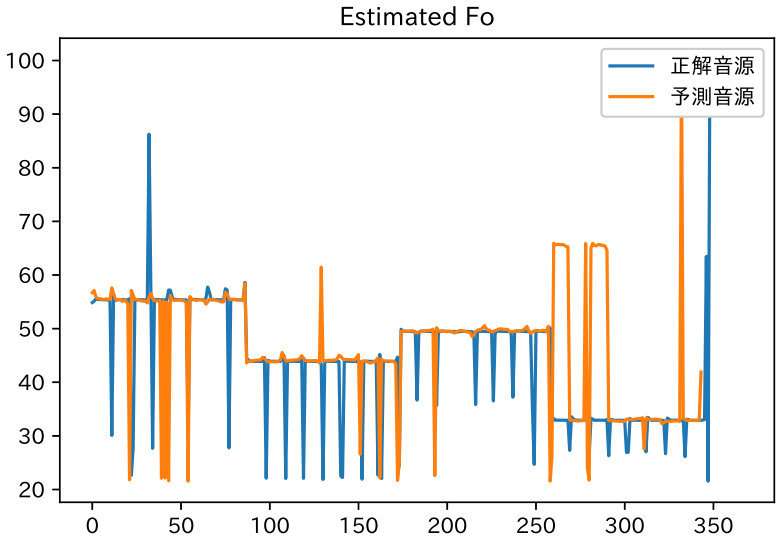


図 A.27: chroma での AmFGC の基本周波数推定結果

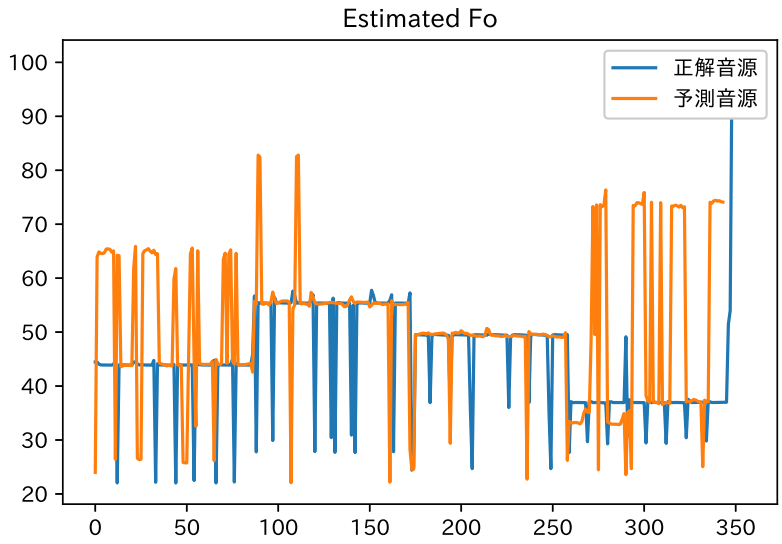


図 A.28: STFT での FAmGDm の基本周波数推定結果

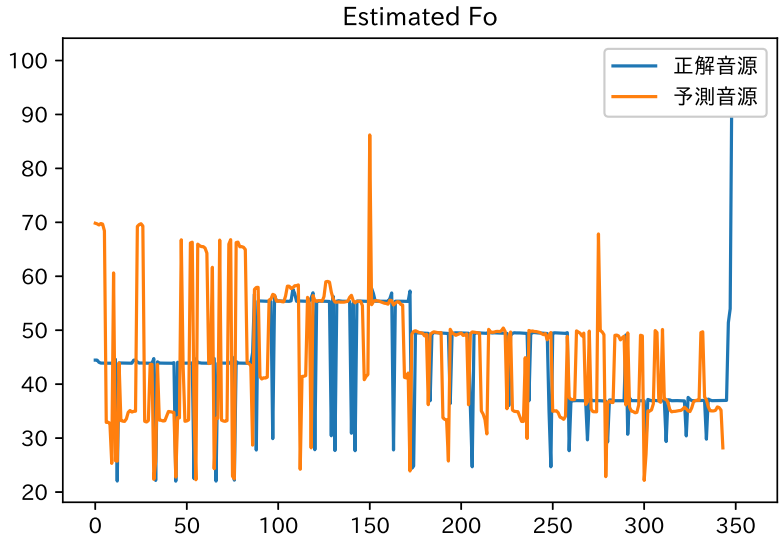


図 A.29: mel での FAmGDm の基本周波数推定結果

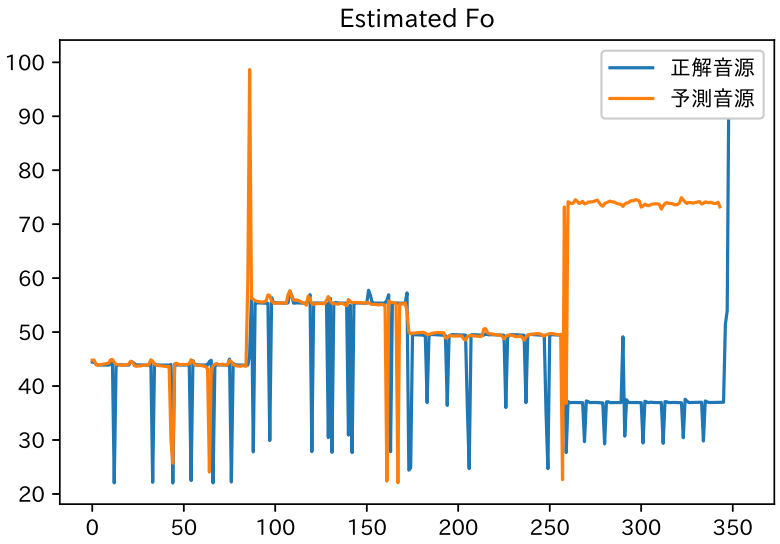


図 A.30: chroma での FAmGDm の基本周波数推定結果

A.2 実験条件 2

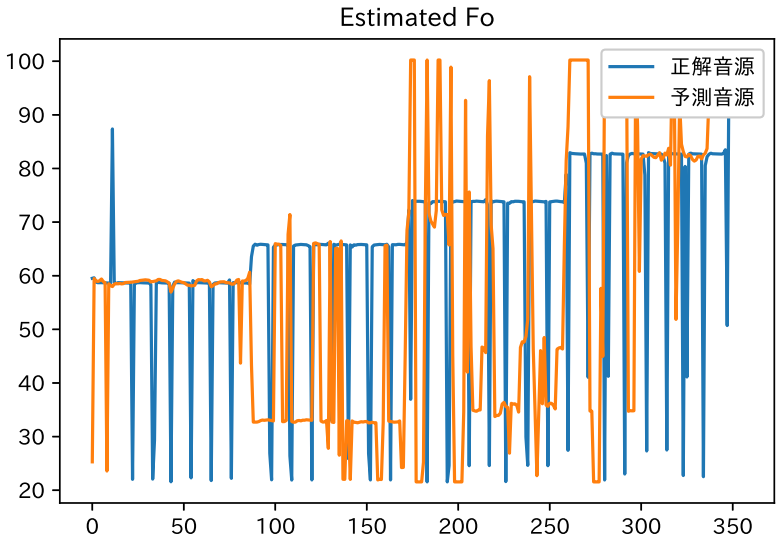


図 A.31: STFT での A♯CDmEm_voicing の基本周波数推定結果

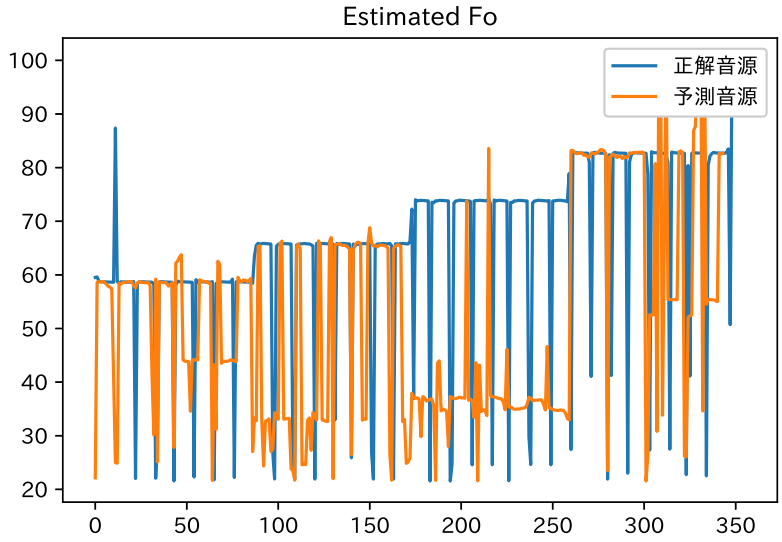


図 A.32: mel での A♯CDmEm_voicing の基本周波数推定結果

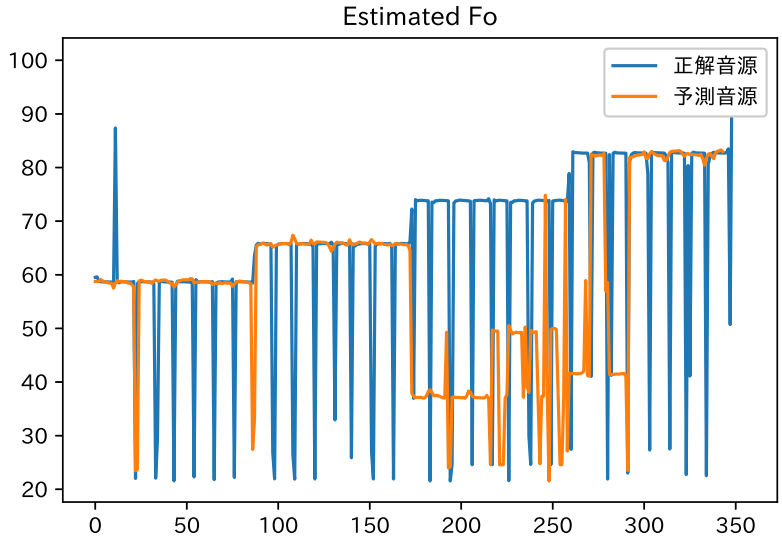


図 A.33: chroma での A♯CDmEm_voicing の基本周波数推定結果

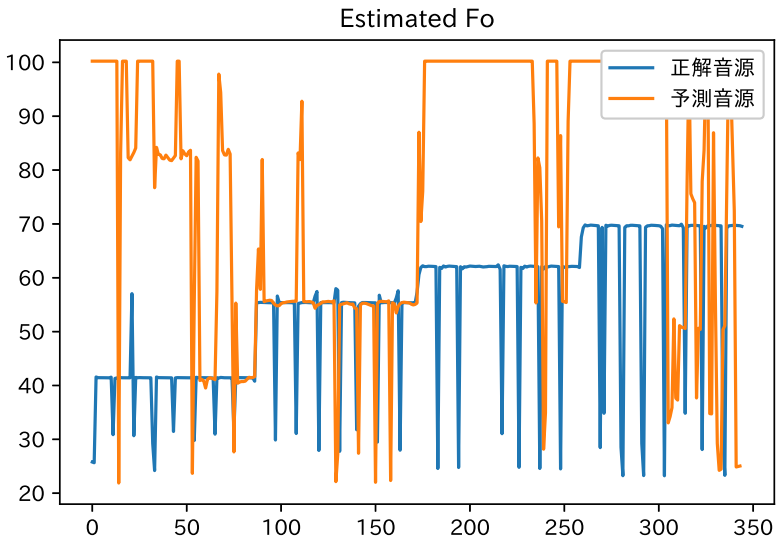


図 A.34: STFT での EABmC♯m_voicing の基本周波数推定結果

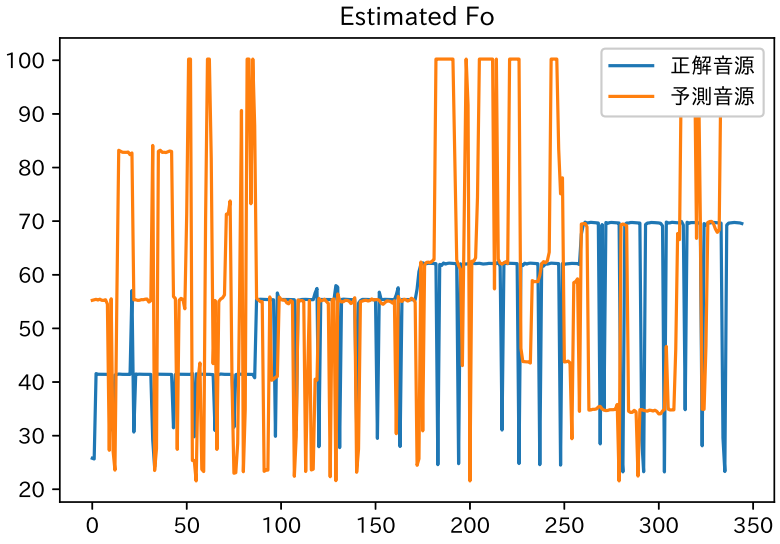


図 A.35: mel での EABmC♯m_voicing の基本周波数推定結果

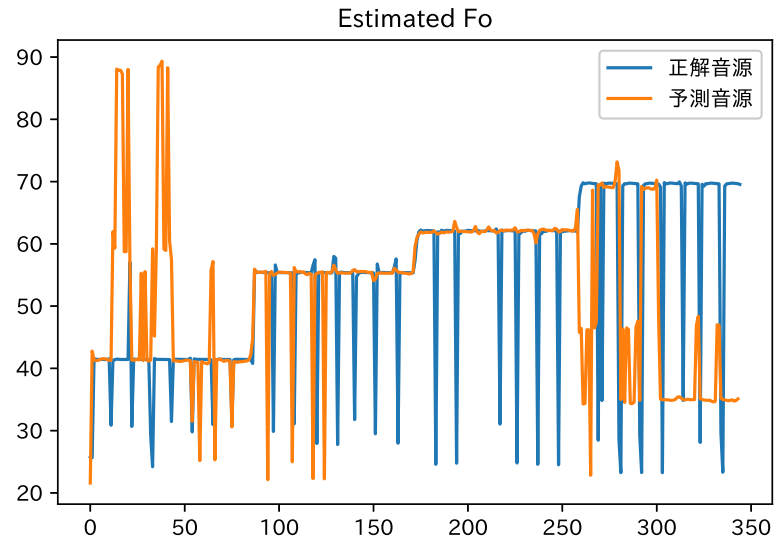


図 A.36: chroma での EABmC♯m_voicing の基本周波数推定結果

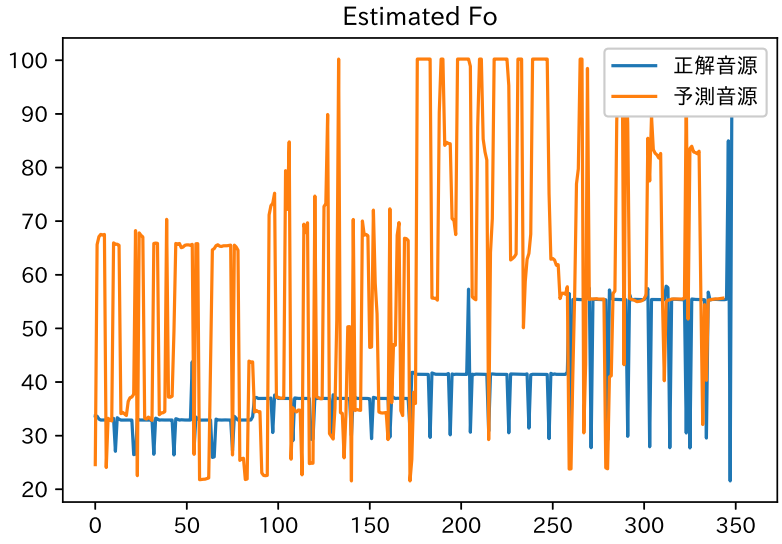


図 A.37: STFT での CDEmAm_voicing の基本周波数推定結果

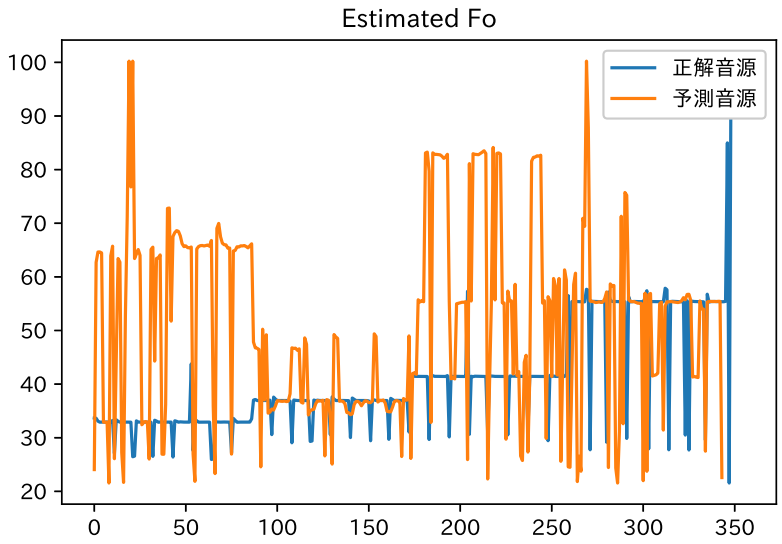


図 A.38: mel での CDEmAm_voicing の基本周波数推定結果

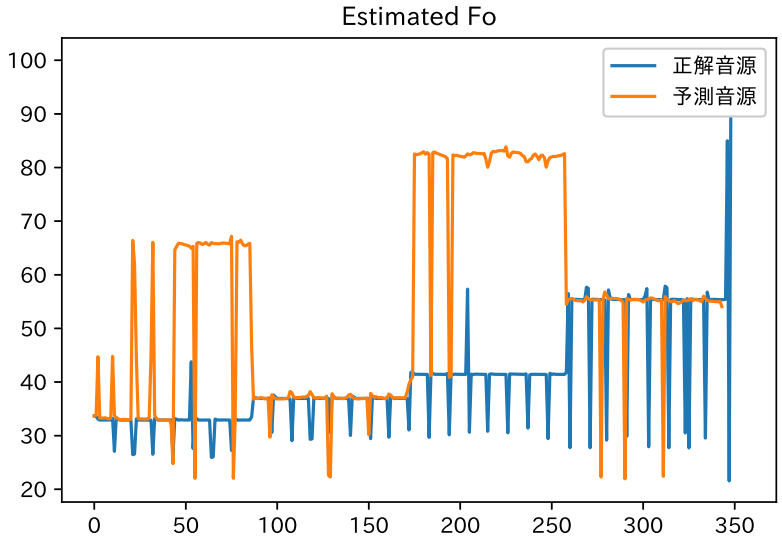


図 A.39: chroma での CDEmAm_voicing の基本周波数推定結果

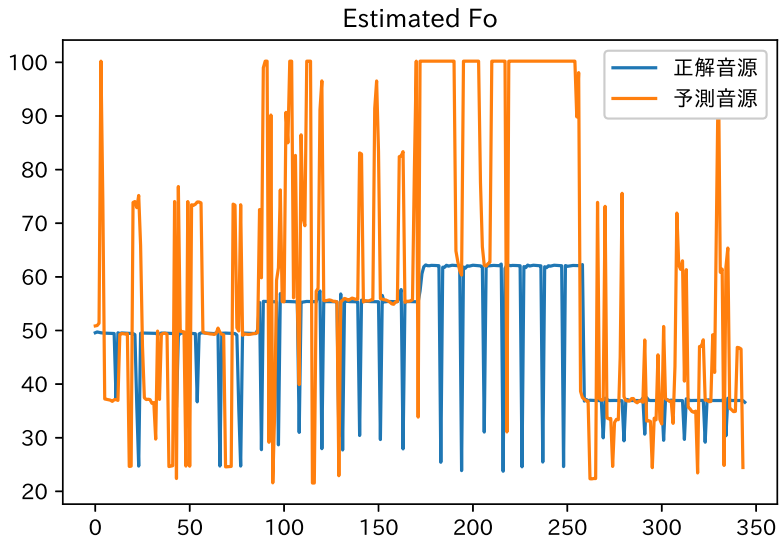


図 A.40: STFT での GABmD_voicing の基本周波数推定結果

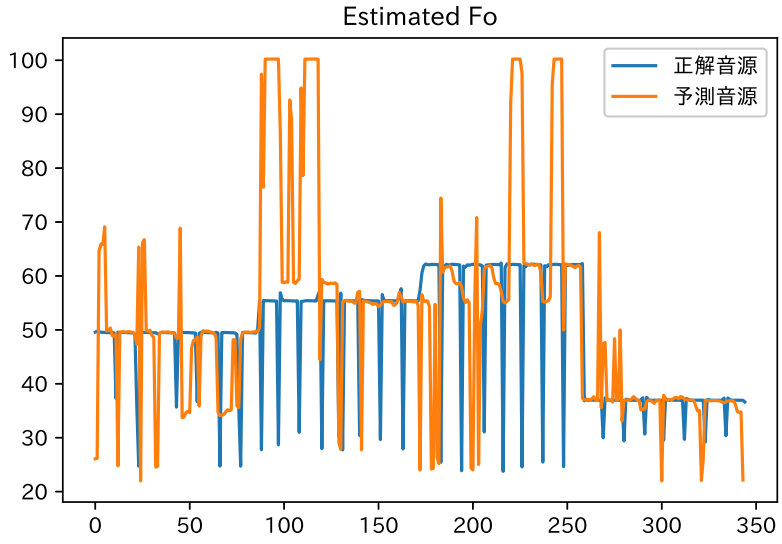


図 A.41: mel での GABmD_voicing の基本周波数推定結果

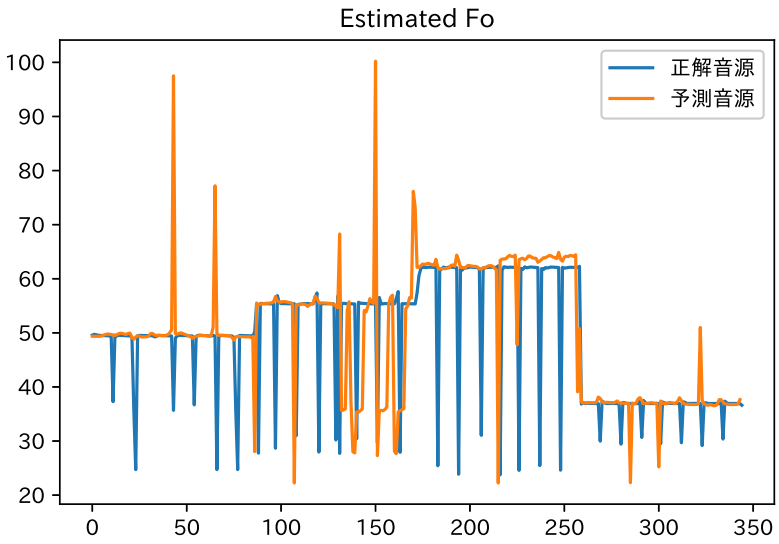


図 A.42: chroma での GABmD_voicing の基本周波数推定結果

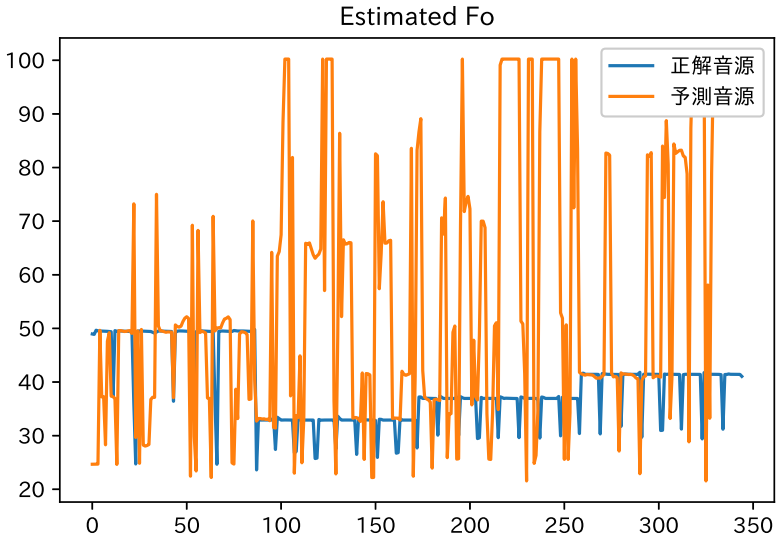


図 A.43: STFT での GCDEm の基本周波数推定結果

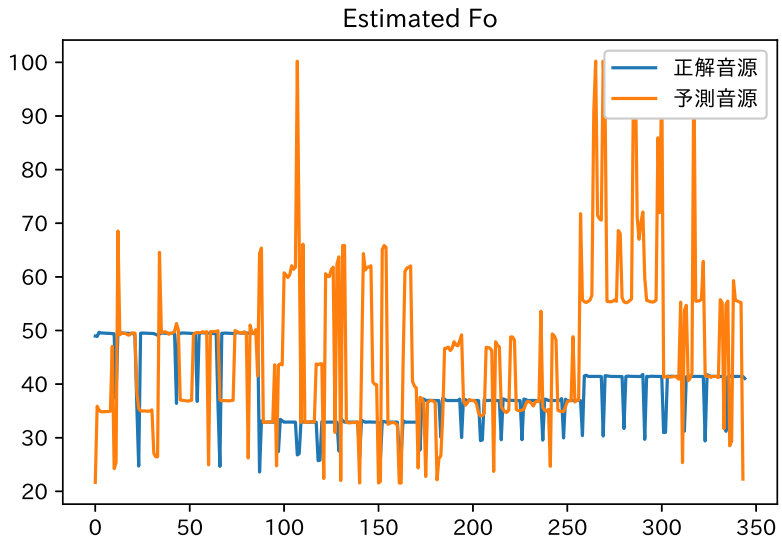


図 A.44: mel での GCDEm の基本周波数推定結果

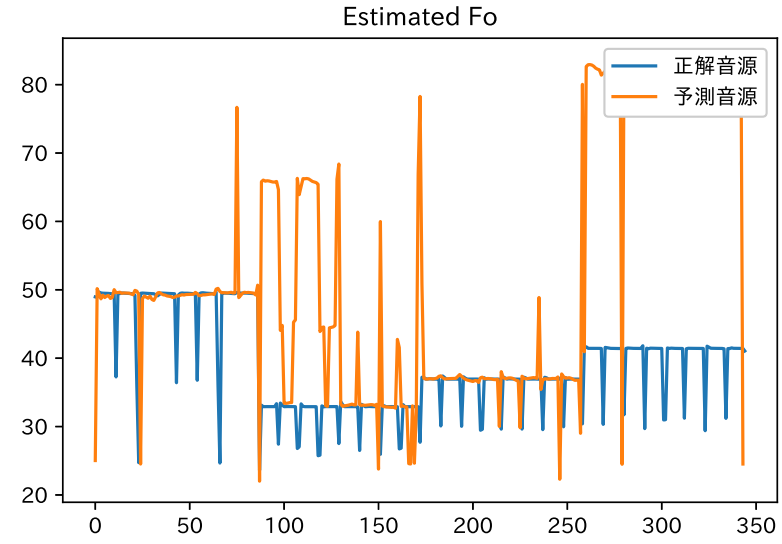


図 A.45: chroma での GCDEm の基本周波数推定結果

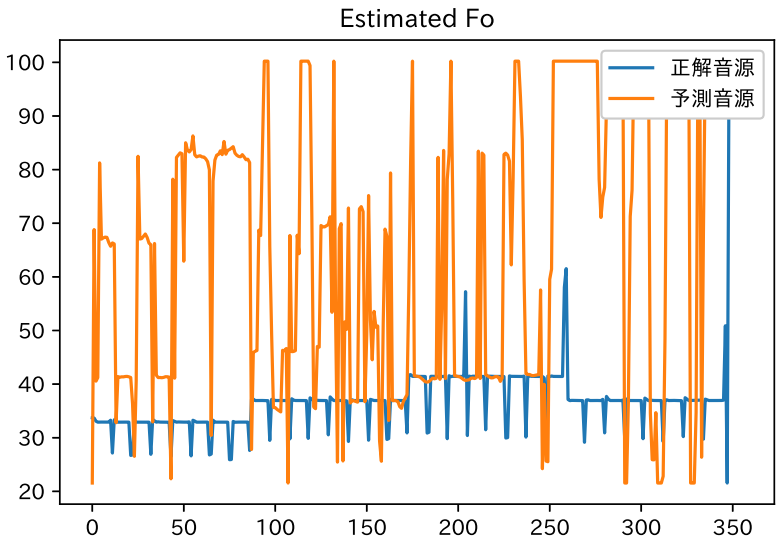


図 A.46: STFT での CDmEmDm の基本周波数推定結果

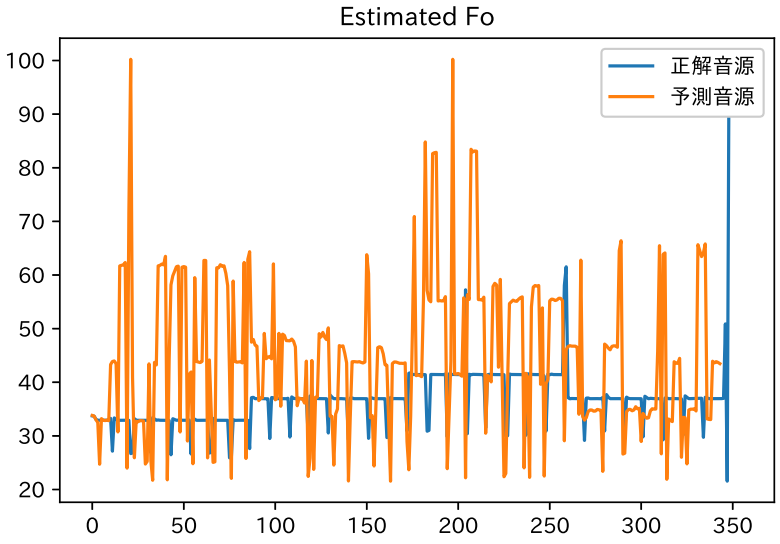


図 A.47: mel での CDmEmDm の基本周波数推定結果

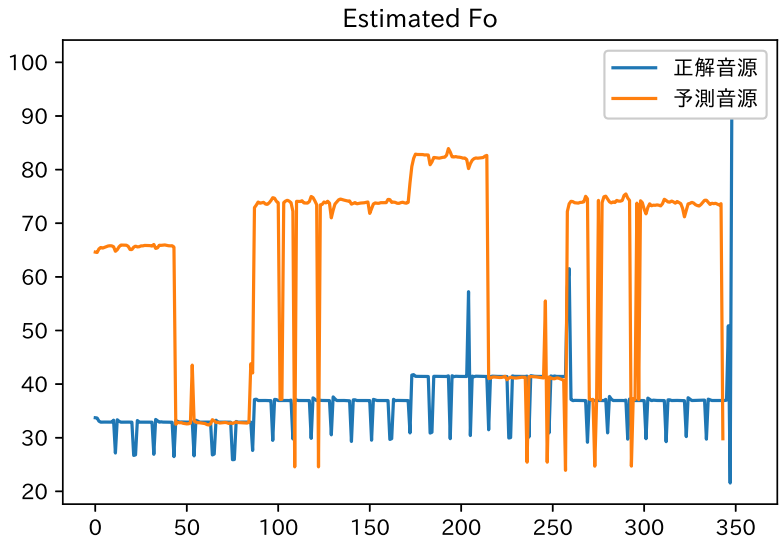


図 A.48: chroma での CDmEmDm の基本周波数推定結果

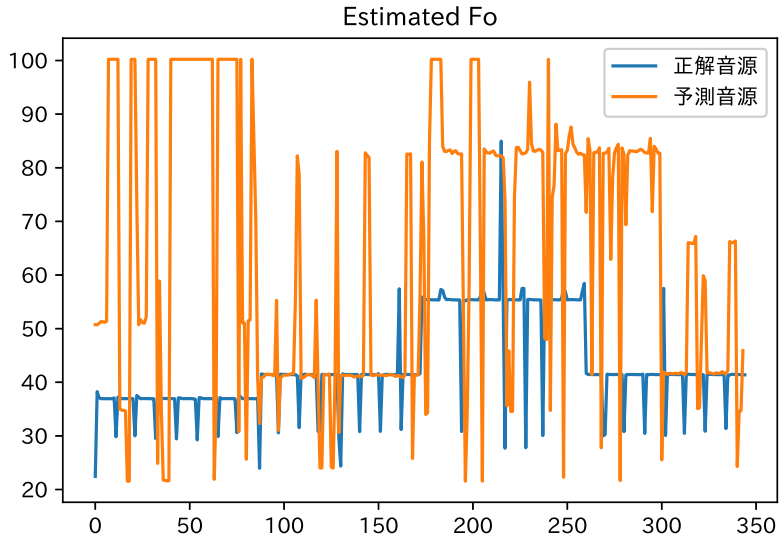


図 A.49: STFT での DmEmAmEm の基本周波数推定結果

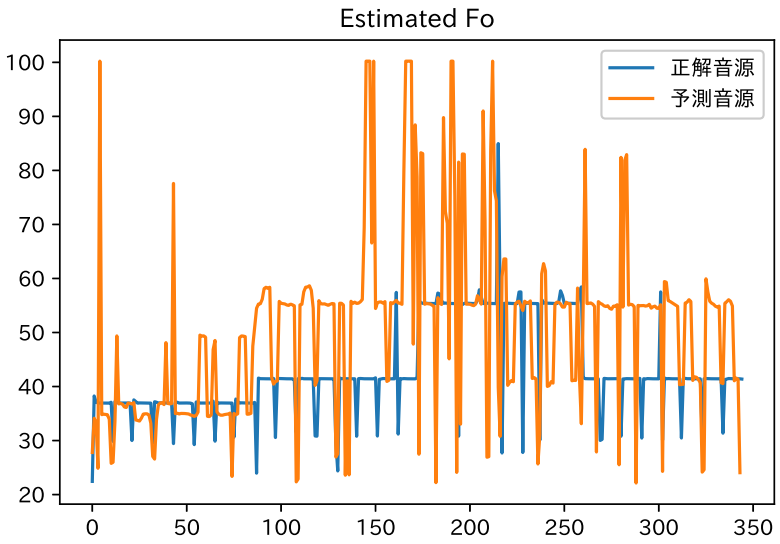


図 A.50: mel での DmEmAmEm の基本周波数推定結果

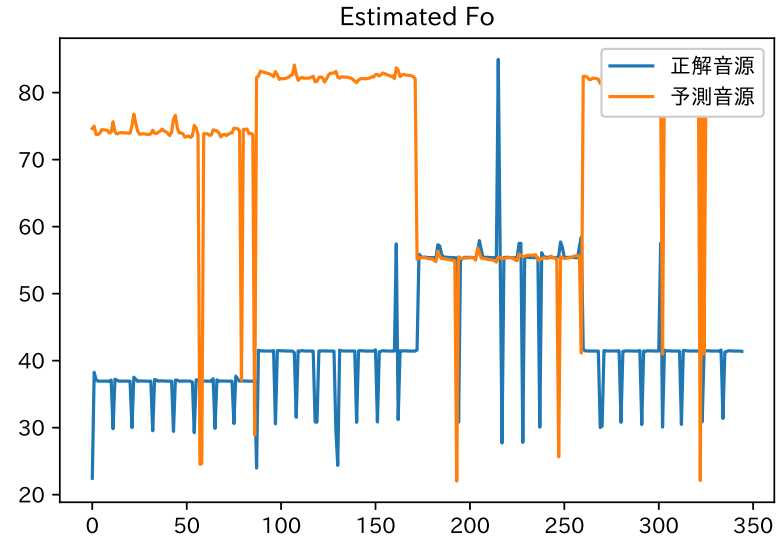


図 A.51: chroma での DmEmAmEm の基本周波数推定結果

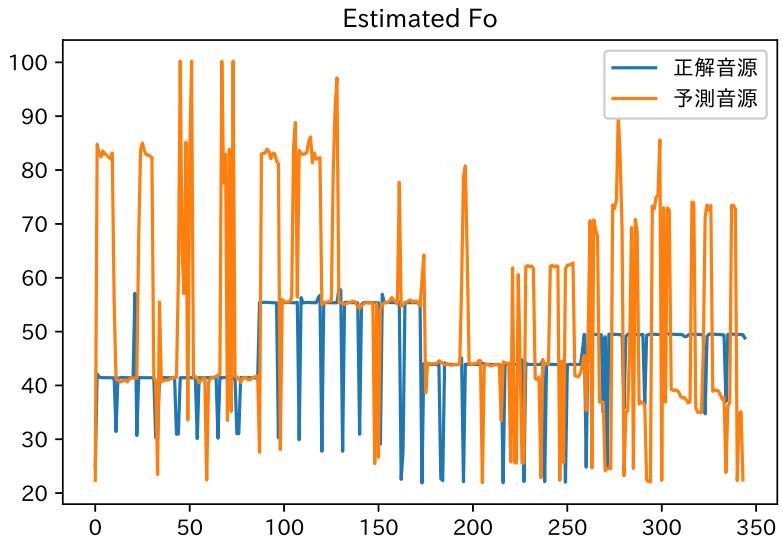


図 A.52: STFT での EmAmFG の基本周波数推定結果

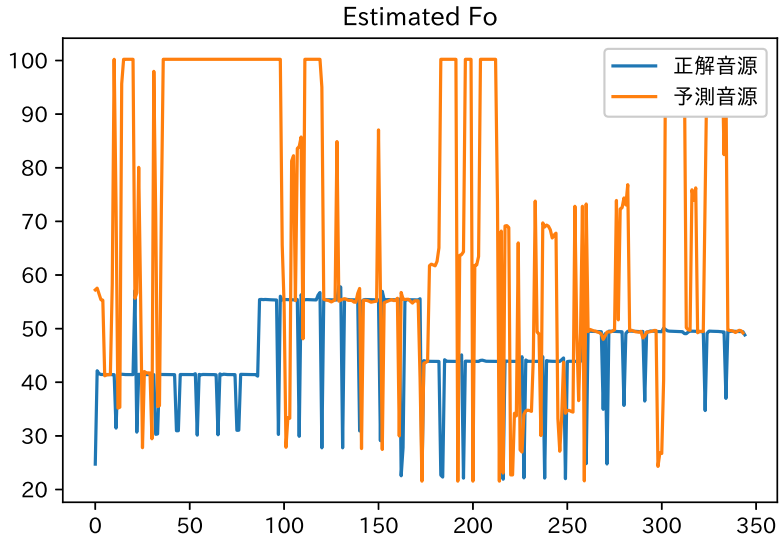


図 A.53: mel での EmAmFG の基本周波数推定結果

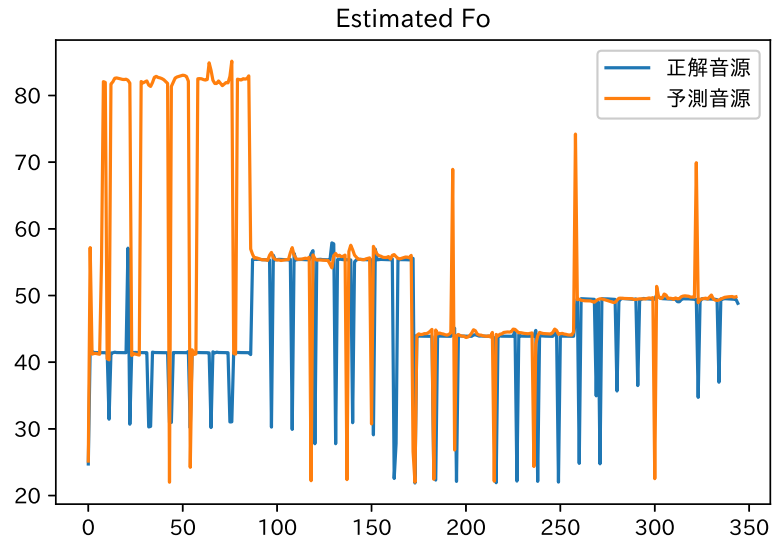


図 A.54: chroma での EmAmFG の基本周波数推定結果

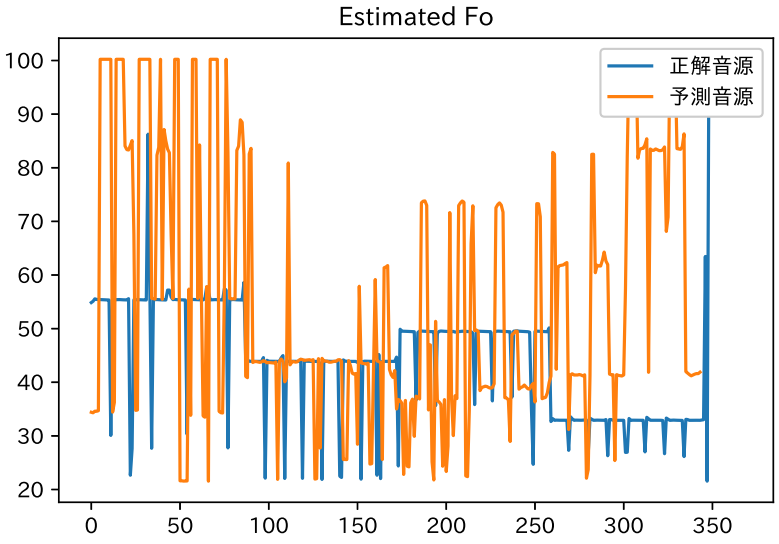


図 A.55: STFT での AmFGC の基本周波数推定結果

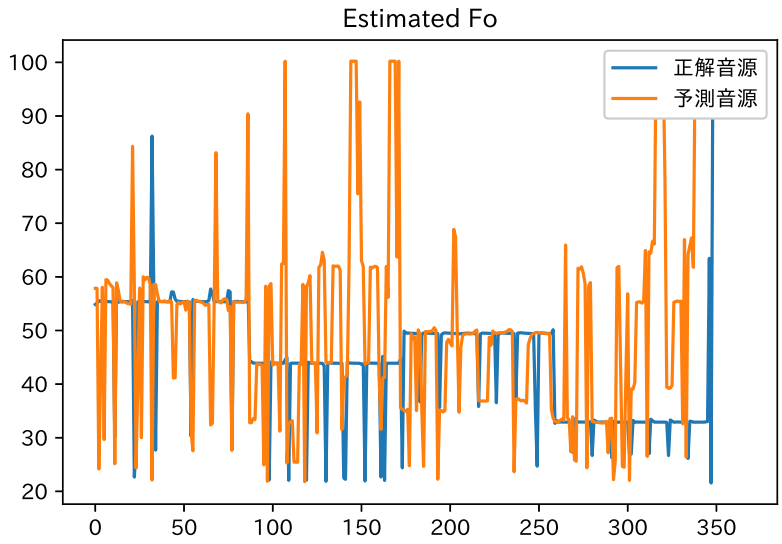


図 A.56: mel での AmFGC の基本周波数推定結果

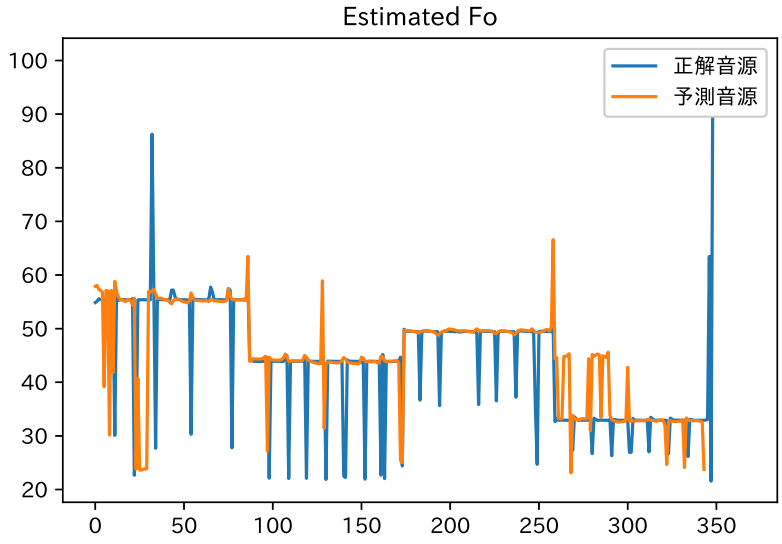


図 A.57: chroma での AmFGC の基本周波数推定結果

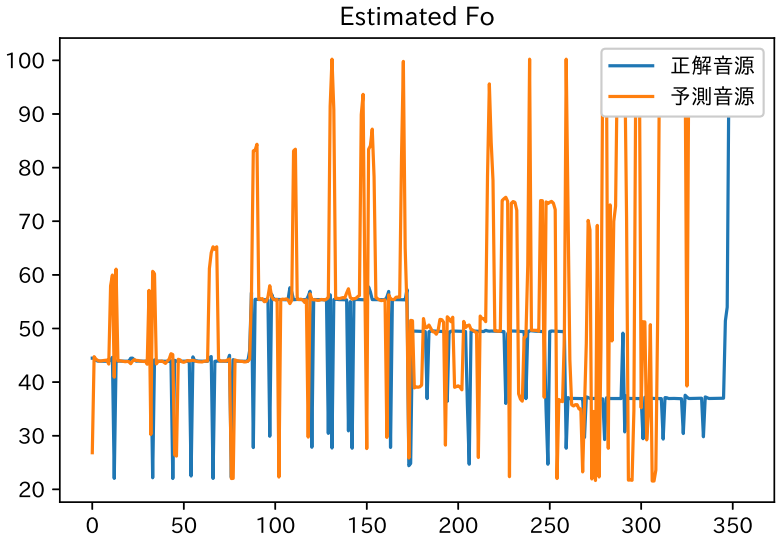


図 A.58: STFT での FAmGDm の基本周波数推定結果

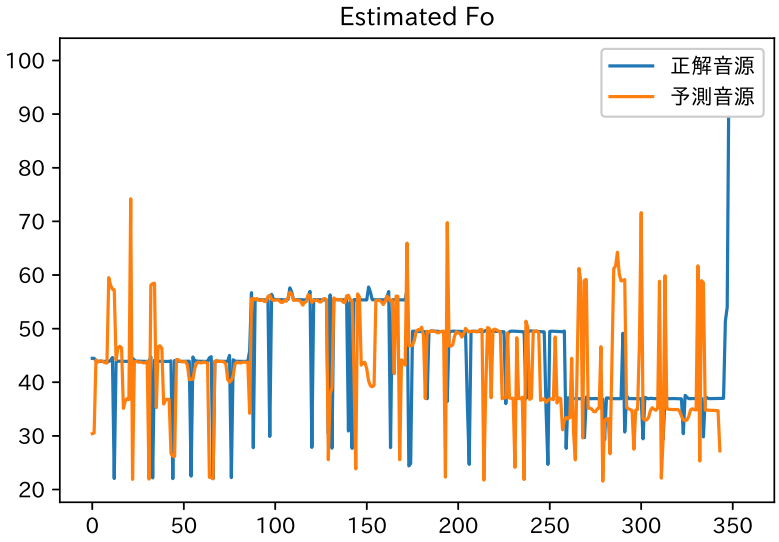


図 A.59: mel での FAmGDm の基本周波数推定結果

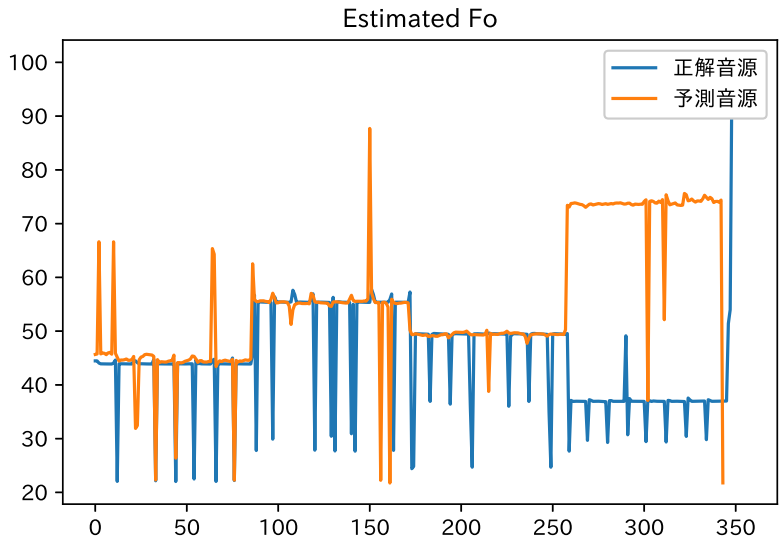


図 A.60: chroma での FAmGDm の基本周波数推定結果