楽譜情報を援用した多重奏音楽音響信号の音源分離と調波・非調波統合モデルの制約付きパラメータ推定の同時実現

糸山克寿[†1] 後藤真孝[†2] 駒谷和範[†1]
尾形哲也[†1] 奥乃博[†1]

本論文では，多重奏の音楽音響信号とその楽曲に含まれるすべての単音の音高・音長・音量・発音時刻・楽器の種類の組である楽譜情報を入力として，単音ごとの音響信号を出力する音源分離手法と，そのための制約付きモデルパラメータ推定手法について述べる．本分離手法では，Standard MIDI File（SMF）などから抽出された楽譜情報を用いることで混合音のパワースペクトルを単音ごとに分離し，調波構造と非調波構造のそれぞれを表現する 2 つのモデルを統合した新たな重み付き混合モデルを用いることで，単音に複数の調波構造が含まれることを防ぎ，かつ音高を超えた楽器音の音色類似性を考慮することを実現する．モデルパラメータは，楽譜情報に基づいて MIDI 音源から生成したテンプレート音によって初期化し，EM アルゴリズムを用いた最大事後確率推定により反復推定する．さらに，モデルの過学習を防ぎ，同一楽器の単音のモデルに類似した音色を持たせるための制約条件も同時に用いる．ポピュラー音楽の SMF を用いた評価実験で，本手法により SNR が 0.4 – 0.9 dB 向上することを確認した．

Simultanious Realization of Score-informed Sound Source Separation of Polyphonic Musical Signals and Constrained Parameter Estimation for Integrated Model of Harmonic and Inharmonic Structure

Katsutoshi Itoyama,[†1] Masataka Goto,[†2]
Kazunori Komatani,[†1] Tetsuya Ogata[†1]
and Hiroshi G. Okuno[†1]

This paper describes a sound sourse separation method for polyphonic sound mixtures of musical signals which include both harmonic instrument sounds and inharmonic instrument sounds, and a constrained parameter estimation method by using a score which includes pitch, duration, volume, onset time, and instrument of each note as prior information. We separate a power spectrum of sound mixtures into each musical note by using an integrated weighted-mixture model consisting of both harmonic-structure and inharmonic-structure tone models (generative models for the power spectrogram). The integrated model realize a parameter estimation method under a constraint of parameter similarity in the same musical instruments. We initialize model parameters using template sounds which are recorded from a MIDI tone generator. On the basis of the Maximum *A Posteriori* Probability estimation using the EM algorithm, we estimated all parameters of this integrated model under several original constraints for preventing over-training and maintaining intra-instrument consistency. Using standard MIDI files as prior information of the model parameters, we confirmed that the integrated model increased the SNR by 0.4 – 0.9 dB.

1. はじめに

デジタルオーディオが普及し，価値観が多様化する中で，より能動的に音楽を楽しみたいというユーザの要求が現れてきた．これまでのオーディオ再生技術は，受動的な音楽の楽しみ方をより豊かにする方向に進歩し，ユーザの要求に応えてきた．たとえば，5.1 次元や 7.1 次元などの大がかりなシステムで忠実な音環境の再現を目指すというものや，アクティブノイズキャンセルなどの簡便な装置で静かな音環境を作ることでどこでも手軽に音楽鑑賞を楽しむというものがある．一方，能動的な音楽の楽しみ方には作曲や編曲，演奏などがある．一般的には能動的に音楽を楽しめる

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

のは技術や道具を持っている人に限られており，受動的な楽しみと能動的な楽しみの間には大きなギャップがあった．

能動的な音楽鑑賞[1]という要求に応える研究事例として，Yoshii らはドラムスを対象とした楽器音イコライザ INTER:D[2] および Drumix[3] を実現した．ユーザは Drumix を使って楽曲中のドラムスの音量を操作し，音色を置き換え，また，ドラムパターンを編集でき，その結果能動的な音楽鑑賞がより簡便に可能となった．しかし，これらのシステムはドラムスだけを対象としており，一般の楽器音に対して適用するまでには至っていなかった．

これに対して我々の目的は，CD などによる音楽音響信号（混合音）中のあらゆる楽器パートに対して自由に音量を操作できる楽器音イコライザを実現することである．従来のグラフィックイコライザやパラメトリックイコライザでは，特定の周波数帯域ごとの音量を調整して周波数特性を変化させることはできたが，楽器ごとの音量を調整することはできなかった．

楽器音イコライザを実現するためには，楽曲中に含まれるすべての楽器音を楽器パートごとに，もしくは単音ごとに分離する必要があり，そのためにはどの楽器が，どの時刻に，どの音高で演奏されているのかといった，「楽譜情報」が必要となる．そこで本論文では，音楽音響信号とその楽曲の「楽譜情報」を入力とし，音楽音響信号を楽器パートごとに分離して出力する手法について論ずる．本論文での「楽譜情報」とは，「楽曲に含まれるすべての単音の音高・音長・音量・発音時刻・楽器種類」である．インターネット上での標準 MIDI ファイル（Standard MIDI File; SMF）の販売サービス[4),5)]などによって，最新の楽曲であっても音楽音響信号に対応した SMF を入手することが容易になっており，これらの SMF から楽譜情報を抽出することが可能であるため，音楽音響信号と楽譜情報の組を得ることはさほど困難ではないと考える．ただし，音響信号と SMF とは何らかの従来法[6)–10)]を用いて同期がとられていると仮定し，本論文では同期の問題については扱わない．

音楽音響信号の音源分離に関する従来研究は，以下の 2 つに大別できる．

（1） 音高を明示的に扱うことで，調波構造を持ち，音高に依存する楽器音を対象にするもの．人間の知覚において，音高が変化しても変化を感じない成分（高調波成分の強度比など）を適切に扱うことができる．調波構造を表現する混合正弦波モデルを用いるもの[11),12)]，SMF を基に調波構造にフィルタをかけるもの[13)]，時間周波数平面上での調波構造を表現するモデルのフィッティングによるもの[14)]，高調波成分のパワーエンベロープの類似性を用いるもの[15)]，ステレオ信号のパワーと位相の共通性を用いるもの[16)]などがある．これらは主に調波構造を含む楽器音のみを対象としており，一般の楽器音を分離することは困難であった．

（2） 音高を明示的に扱わず，一般の楽器音を対象にするもの．原理的には任意の楽器音を扱うことができるが，各々の従来研究ではピアノやバイオリンなどの楽音とドラムスなどの噪音のいずれか一方を主として扱ってきた．楽音を扱うものでは，Non-negative Matrix Factorization（NMF）やその拡張である Non-negative Tensor Factorization（NTF）を用いるもの[17),18)]などが，噪音を扱うものでは，Independent Component Analysis（ICA）を用いるもの[19)]，NMF を用いるもの[20)]，ドラム音検出の後にスペクトル変調を行うもの[21)]などがある．また，ICA などの統計的手法で楽音と噪音を同時に分離するもの[22)–26)]があるが，発音区間や周波数成分のスパースネスが保証されない複雑な音響信号を扱うには至っていない．このような音響信号を分離するためには楽器音認識が不可欠で，楽器音認識と音源分離を併用したもの[27)]などが研究事例としてあげられるが，対象がドラム音のみであり，調波構造を持つ音には適用されていなかったりした．

これらのアプローチは従来排他的で，双方の長所をあわせ持つ手法はこれまで存在しなかった．

本論文で分離の対象とする音は，「調波的な音」および「非調波的な音」，およびそれらを加算して得られる音である．以下の性質を満たす音を調波的な音と呼ぶ．

- 調波構造を持つ．
- 各高調波成分の相対強度が時間の経過によって変化しない．
- 急激な F0 の変化を含まない．

具体的には，弦（ピアノ弦やギター弦など）や管内の空気（フルートなど）の定常的な振動によって得られる音が相当する．歌声は各高調波成分の相対強度が母音の遷移によって連続的に変化するため，本論文では扱わない．また，パワースペクトルに調波構造を含めた周波数方向への鋭いピークが存在しない音を非調波的な音と呼ぶ．具体的には，ドラム音を想定している[⋆1]．ピアノが発音時にハンマで弦を叩く音のような，調波構造を含む音から調波構造を取り除くことによっ

⋆1 理想的な膜振動から得られる信号には「整数倍でない倍音構造」が含まれるため，ドラム音の中には周波数方向への鋭いピークが存在するものがある．しかし本論文では「ドラム音には周波数方向への鋭いピークは存在しない」と仮定し，ドラム音を扱う．

て得られる音も，周波数方向への鋭いピークがほとんど存在しないと見なし非調波的な音に含める．

従来の音源分離に関する研究の多くは，前述のとおりこれらの 2 種類の音の一方のみに着目していた．Goto[28] は，混合音中の最も優勢な調波構造を抽出する手法について述べており，さらに調波構造モデルを他の任意の関数の重み付き混合モデルに置き換えても調波構造の場合と同様にモデルパラメータ推定を行うことで，パワースペクトル上の任意の構造を扱うことが理論的には可能であるとも述べているが，調波構造以外の構造をどのようなモデルで扱えばよいか，そのようなモデルを実現するうえでどのような問題点があるか，といった具体的な手法については述べていなかった．

我々は Goto[28] の示唆を受け，調波構造モデルと非調波構造モデルを統合した混合モデルを用いた音源分離手法を設計し，実現した．調波構造モデルは，音高を持つ楽器の単音の調波構造を表現するパラメトリックモデルに基づいており，発音時刻，音長，音量，音高（F0）の時間変化，パワーエンベロープの時間変化，各高調波成分の相対強度といったパラメータで表現される．非調波構造モデルは，ノンパラメトリックモデルに基づいており，調波構造では表現が難しいドラム音などのパワースペクトルをそのまま表現する．また前述のように，ピアノやギターなどの調波構造を持つ楽器音であっても，発音時には弦をハンマで叩くことや弦を弾くことに由来する非調波成分を含んでいるので，それらのパワースペクトルも非調波構造モデルで表現する．

SMF などから抽出した各単音の音高，音長，音量，発音時刻，楽器によってこれらのモデルのパラメータを初期化し，モデルパラメータの最大事後確率推定を EM アルゴリズムを用いて実現する．このパラメータ推定における問題点は，非調波構造モデルは大きな自由度を持っており，あらゆるパワースペクトルを表現できるため，パラメータ推定の結果，非調波構造モデルが調波成分も含めてすべての混合音を表現してしまうことである．この問題を解決するため，非調波構造モデルの形状に関する制約や同一楽器のモデルパラメータ類似性に関する制約を導入する．このようにして得られた調波・非調波統合モデルを用いることで，混合音のパワースペクトル上での分離が可能となる．

本論文の構成は以下のとおりである．まず，2 章で音源分離における 2 つの問題点を述べる．続いて，3 章では調波・非調波統合モデルの定式化，4 章では統合モデルに基づく音源分離処理，5 章ではモデルパラメータの推定処理について述べる．6 章で評価実験について述べ，7 章で本論文のまとめを行う．

2. 問題の所在と解決へのアプローチ

本研究の目標は，多重奏の音楽音響信号とその楽曲の楽譜情報（各単音の音高・音長・音量・発音時刻・楽器種類）が与えられたとき，音響信号のパワースペクトルを単音ごとに分離することである．いい換えれば，我々の目標は各楽器パートのすべての単音に対して，単音に対応する調波構造モデルと非調波構造モデルの全パラメータを推定することである．

与えられた楽譜情報の各単音を個別に MIDI 音源で演奏することで，音響信号中の各単音にある程度近い，「音のサンプル」を作成できる．この音のサンプルをテンプレート音と呼ぶ．分離対象の音楽音響信号とその楽譜情報，MIDI 音源を用いて生成したテンプレート音が与えられたとき，我々が解くべき課題は以下の 2 点である．

（1） **テンプレート音と入力音響信号とのずれの吸収．**テンプレート音と入力信号との間には必ず音響的な違いがあるので，テンプレート音をそのまま用いたのでは完全に 1 つ 1 つの音を分離することはできない．そこで，テンプレート音と入力信号との音響的差異を吸収する手法が必要となる．

（2） **奏法に独立な楽器音の音色類似性の達成．**ある楽器を用いて，同一の音高，音長，音量を持つ単音を複数回にわたって演奏したとしても，奏法（ビブラートのかけ方など）が異なれば異なる音色を持つ音響信号が生成される．単音ごとの音色の違いを表現するためには，音色を表現するモデルを単音ごとに作成する必要がある．しかし，単音ごとに独立したモデルを作成すると，パラメータの自由度が大きくなりすぎてしまうため，混合音への適応によってモデルが過学習を起こし，結果として分離性能が低下してしまう可能性がある．これを防ぐためには，同一楽器の単音の間に存在する音色の類似性を満たすような，音色の表現方法を実現する必要がある．

これらの課題を，以下のアプローチで解決する．

（1） **モデルパラメータ適応．**テンプレート音で初期化した音モデルのパラメータを，モデルと入力音響信号とのパワースペクトル上での音響的差異を最小化するように更新する．すなわち，音モデルのパラメータを入力音響信号に適応させることによってテンプレート音と実演奏とのずれを吸収する．

（2） **同一楽器内パラメータ類似性に対する制約．**同一楽器の個々の単音モデル間のパラメータの類似性を保ちつつも各単音の微小な違いを許容するような制

約の下でモデルパラメータの更新を行う．これは，同一楽器に属する各単音のモデルパラメータの平均値と現在着目している単音のモデルパラメータとの間のKullback-Leibler ダイバージェンス（以下，KLD と略す）を最小化することによる，モデルパラメータに対する制約を与えることで達成できる．

2.1 問題の定義

本論文で扱う分離問題とは，入力混合音のパワースペクトル $X(c,t,f)$ を，k 番目の楽器，l 番目の単音（以下，(k,l) 番目と記す）のパワースペクトルに分解することである．ここで，$c \in \{1,\ldots,\mathrm{C}\}$ は左右などのチャネルの番号，$t \in [\mathrm{T}_0,\mathrm{T}_1]$ は時刻，$f \in [\mathrm{F}_0,\mathrm{F}_1]$ は周波数を表す．入力された楽譜情報から，楽曲中では K 種類の楽器が演奏されており，各々の楽器は L_k 個の単音を持つものとする．すなわち，$k \in \{1,\ldots,\mathrm{K}\}$ であり，各々の k に対して $l \in \{1,\ldots,\mathrm{L}_k\}$ である．また，(k,l) 番目の単音を表すモデルを $J_{kl}(c,t,f)$，(k,l) 番目のテンプレート音のパワースペクトルを $Y_{kl}(t,f)$ とする．本論文で用いる楽譜情報にはチャネル間音圧比は含まれていないため，テンプレート音にはチャネル間音圧比を設定せず，モノラル音響信号として生成する．そのため，$Y_{kl}(t,f)$ には c がなく 1 チャネルとなっている．また，

$$X_0 = \frac{1}{\mathrm{C}}\sum_c \iint X(c,t,f)\,dt\,df = \sum_{k,l}\iint Y_{kl}(t,f)\,dt\,df \tag{1}$$

となるように，$Y_{kl}(t,f)$ のパワーを正規化してあるものとする．

3. 調波・非調波統合モデル

調波・非調波統合モデル $J_{kl}(c,t,f)$ は，(k,l) 番目の単音のパワースペクトルを表現するモデルである．調波的な音のパワースペクトルを表現する調波構造モデル $H_{kl}(t,f)$ と非調波的な音のパワースペクトルを表現する非調波構造モデル $I_{kl}(t,f)$ との和に統合モデル全体の重み w_{kl} を乗じた $J'_{kl}(t,f)$ に，さらにチャネルごとの重み $r_{kl}(c)$ を乗じたもので，以下の式で定義する．

$$J_{kl}(c,t,f) = r_{kl}(c)J'_{kl}(t,f) \tag{2}$$

$$J'_{kl}(t,f) = w_{kl}\big(H_{kl}(t,f) + I_{kl}(t,f)\big) \tag{3}$$

w_{kl} および $r_{kl}(c)$ は以下の各条件を満たす．

$$\forall k,l,\ \sum_{k,l} w_{kl} = X_0 \tag{4}$$

$$\forall k,l,\ \sum_c r_{kl}(c) = \mathrm{C} \tag{5}$$

3.1 調波構造モデル

調波構造モデル $H_{kl}(t,f)$ は，パラメトリックな基底関数であるガウス分布関数の重み付き線形和として，パワーエンベロープを表現する関数 $E_{klm}(t)$ と各時刻の調波構造を表現する関数 $F_{kln}(t,f)$ を用いて以下の式で定義する．ただし，M，N は定数で，それぞれパワーエンベロープを表現するガウス分布関数の数と高調波成分を表現するガウス分布関数の数を表す．

$$H_{kl}(t,f) = \sum_{m,n} E_{klm}(t)F_{kln}(t,f) \tag{6}$$

$$E_{klm}(t) = \frac{u_{klm}}{\sqrt{2\pi}\phi_{kl}} e^{-\frac{(t-\tau_{kl}-m\phi_{kl})^2}{2{\phi_{kl}}^2}} \tag{7}$$

$$F_{kln}(t,f) = \frac{v_{kln}}{\sqrt{2\pi}\sigma_{kl}} e^{-\frac{(f-n\omega_{kl}(t))^2}{2{\sigma_{kl}}^2}} \tag{8}$$

このモデルは，Kameoka らの調波時間構造化クラスタリング（Harmonic Temporal Clustering; HTC）[14] で用いられる音源モデルを参考に設計した．Kameoka らの HTC 音源モデルでは，$\omega_{kl}(t)$ は時間 t に関する多項式として定義されていたが，多項式で表現可能な F0 時系列の集合は任意の F0 時系列の集合よりも小さい．そこで本研究では，任意の F0 時系列を表現するために，ノンパラメトリックな関数として $\omega_{kl}(t)$ を定義した．u_{klm}，v_{kln} は以下の条件を満たす．

$$\forall k,l,n,\ \sum_m u_{klm} = 1 \tag{9}$$

$$\forall k,l,m,\ \sum_n u_{kln} = 1 \tag{10}$$

3.2 非調波構造モデル

非調波構造モデル $I_{kl}(t,f)$ は，ノンパラメトリックな関数として，パワースペクトルの各時刻および周波数における周波数成分の強度を直接表現するように以下の式で定義する．

$$I_{kl}(t,f) = w^{(I)}_{kl} I'_{kl}(t,f) \tag{11}$$

ただし，$I'_{kl}(t,f)$，$w^{(H)}_{kl}$ および $w^{(I)}_{kl}$ は以下の各条件を満たす．

$$\forall k,l,\ \iint I'_{kl}(t,f)\,dt\,df = 1 \tag{12}$$

$$\forall k,l,\ w^{(H)}_{kl} + w^{(I)}_{kl} = 1 \tag{13}$$

4. 音 源 分 離

入力パワースペクトル $X(c,t,f)$ を (k,l) 番目の単音へと分離するためのパワースペクトル分配関数 $S_{kl}(c,t,f)$ を導入する．この関数は以下の条件を満

たす．

$$\forall c,t,f,\ \sum_{k,l} S_{kl}(c,t,f)=1 \tag{14}$$

分離された (k,l) 番目の単音のパワースペクトルは

$$X^{(S)}_{kl}(c,t,f)=S_{kl}(c,t,f)X(c,t,f) \tag{15}$$

で表される．

ここで，どのように $S_{kl}(c,t,f)$ を定めると最も良い分離が行えるかを考える．それには分離の良し悪しを計る尺度が必要なので，$X^{(S)}_{kl}(c,t,f)$ と $J_{kl}(c,t,f)$ との Kullback-Leibler Divergence (KLD) Q'_{kl}：

$$Q'_{kl}=\sum_c \iint X^{(S)}_{kl}(c,t,f)\log\frac{X^{(S)}_{kl}(c,t,f)}{J_{kl}(c,t,f)}\,dt\,df \tag{16}$$

でこの尺度を定義する．KLD は距離の公理を満たさないが，あらゆる k，l，c，t，f に対して

$$X^{(S)}_{kl}(c,t,f)=J_{kl}(c,t,f) \tag{17}$$

となるときに限り最小値 0 をとるので，パワースペクトル間の類似度として用いることができる．

このとき，Q'_{kl} を最小化するような $S_{kl}(c,t,f)$ を求めることができれば，それを用いた $X^{(S)}_{kl}(c,t,f)$ が Q'_{kl} に基づく最適な分離結果となる．ただし，$S_{kl}(c,t,f)$ は式 (14) を満たさなければならないので，あらゆる k，l に関して同時に Q'_{kl} を最小化する必要がある．そこで，Q'_{kl} をあらゆる k，l に関して足し合わせ，式 (14) の条件に対する未定乗数 $\lambda^{(S)}(c,t,f)$ による Lagrange の未定乗数項を加えた Q'：

$$\begin{aligned}Q'=&\sum_{k,l}Q'_{kl}\\&-\sum_c\iint\lambda^{(S)}(c,t,f)\\&\cdot\left(\sum_{k,l}S_{kl}(c,t,f)-1\right)dt\,df\end{aligned} \tag{18}$$

を最小化する．Q' は $S_{kl}(c,t,f)$ に関する制約条件を満たす空間において凸関数であるので，連立方程式

$$\frac{\partial Q'}{\partial S_{kl}(c,t,f)}=0,\quad \frac{\partial Q'}{\partial\lambda^{(S)}(c,t,f)}=0 \tag{19}$$

の解

$$S_{kl}(c,t,f)=\frac{J_{kl}(c,t,f)}{\sum_{k,l}J_{kl}(c,t,f)} \tag{20}$$

が，Q' を最小化する $S_{kl}(c,t,f)$ であり，これにより統合モデルに基づく分離が行われる．

図 1 に全体の処理の流れを示す．最初に楽譜情報と

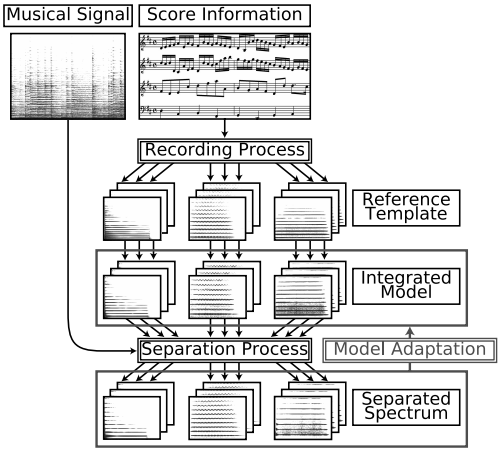


図 1 分離とモデル適応の処理の流れ
Fig. 1 Overview of separation and model adaptation.

MIDI 音源を用いて，テンプレート音の録音を行い，テンプレートを基にモデルパラメータを初期化する．その後，Q' を $S_{kl}(c,t,f)$ に関して最小化することによるパワースペクトルの分離（Separation Process）と，分離パワースペクトルを用いたモデルパラメータ推定（Model Adaptation）を交互に繰り返すことで分離処理が進められる．モデルパラメータ推定については次章で詳細を述べる．

5. パラメータ推定

分離の場合と同様に，推定されたパラメータの良し悪しを計る尺度を $S_{kl}(c,t,f)X(c,t,f)$ と $J_{kl}(c,t,f)$ の間の KLD である Q'_{kl} および制約条件として作用するいくつかの追加コストの重み付き和で定め，この尺度を最小化するようにパラメータ推定を行う．本章では，パラメータ推定における各々の制約条件について述べ，その後パラメータ推定方法および推定手法と EM アルゴリズムとの関連について述べる．

5.1 テンプレート音との類似性

(k,l) 番目の単音のテンプレート音は，混合音中の (k,l) 番目の単音とは楽器個体などが異なっているものの，SMF から抽出した音高，音長，音量，楽器によって MIDI 音源から生成した音響信号であるので，混合音中の (k,l) 番目の単音と「かなり」近いパワースペクトルを持つと考えられる．つまり，モデルとテンプレートスペクトル $Y_{kl}(t,f)$ との差を同時に最小化することで，混合音中の (k,l) 番目の単音から推定される理想的なパラメータに「かなり」近いパラメータを推定することが可能になる．

分離スペクトルとモデルとの差を KLD で定義したことと同様に，テンプレートスペクトルとモデルとの

差 $Q_{kl}^{(Y)}$ を KLD で定義する．ただし，テンプレートはモノラル音響信号であるので，モデルはチャネル間音圧比を含まないものを用いる．

$$Q_{kl}^{(Y)} = \iint Y_{kl}(t,f) \log \frac{Y_{kl}(t,f)}{J'_{kl}(t,f)}\, dt\, df \quad (21)$$

5.2 調波構造モデルへの制約

調波構造モデルのパラメータ $\omega_{kl}(t)$ は大きい自由度を持つため，たとえば，分離音に他の楽器音が混入しており，本来モデルが表現すべき楽器音の音高ではなくそちらの楽器音の音高が推定されてしまった場合，$\omega_{kl}(t)$ に不連続な区間が含まれてしまう．しかし，通常はこのような F0 時系列は楽器音の単音として望ましくなく，単音の F0 時系列は連続的に変化するべきと考える．そこで，以下の式で表現される新たなコストを導入する．

$$Q_{kl}^{(\omega)} = \int \tilde{\omega}_{kl}(t) \log \frac{\tilde{\omega}_{kl}(t)}{\omega_{kl}(t)}\, dt \quad (22)$$

$\tilde{\omega}_{kl}(t)$ は，$\omega_{kl}(t)$ にガウシアンフィルタを畳み込むことで時間方向に平滑化したもので，このコストは $\tilde{\omega}_{kl}(t)$ と $\omega_{kl}(t)$ の KLD で定義されている．$\tilde{\omega}_{kl}(t)$ は $\omega_{kl}(t)$ よりも「滑らか」であるため，Q'_{kl} を最小化すると同時に $Q_{kl}^{(\omega)}$ を最小化することで，推定された $\omega_{kl}(t)$ が「滑らか」であることを強制することができる．

5.3 非調波構造モデルへの制約

1 章で述べたように，非調波構造モデル $I_{kl}(t,f)$ は非常に大きい自由度を持つ．そのため，このモデルは任意のパワースペクトルを表現することが可能で，入力パワースペクトル，および分離パワースペクトルが調波構造モデルを用いることなくこのモデルだけで表現されてしまう可能性がある．しかし，調波構造モデルが表現すべき調波構造までも非調波構造モデルが表現してしまうことは望ましくない．

我々は，パワースペクトル上の調波構造に関する最大の特徴は周波数方向に複数のピークを持つこと，と考える．逆に，周波数方向に強いピークを持たないパワースペクトルが調波構造を含むことはない．したがって，非調波構造モデルが周波数方向に強いピークを持ちにくくさせる制約を与えることができれば，前述の問題を解決できると考える．

この制約を実現するため，以下の式で表現される新たなコストを導入する．

$$Q_{kl}^{(\tilde{I})} = \iint \tilde{I}_{kl}(t,f) \log \frac{\tilde{I}_{kl}(t,f)}{I'_{kl}(t,f)}\, dt\, df \quad (23)$$

ここで，$\bar{I}_{kl}(t,f)$ は $I'_{kl}(t,f)$ にガウシアンフィルタを畳み込むことで，周波数方向に平滑化したものであり，$I'_{kl}(t,f)$ よりも周波数方向に滑らかであるため周波数方向のピークを持たない，もしくは持っていたとしてもピークにおける周波数成分のパワーは $I'_{kl}(t,f)$ のそれより小さい．したがって，パラメータ推定時にこのコスト $Q_{kl}^{(\tilde{I})}$ も最小化することで，非調波構造モデルが調波構造を表現せず，それ以外の「非調波的な音」のパワースペクトルだけを表現することを強制できる．

5.4 同一楽器内のパラメータ類似性

2 章で述べたように，調波・非調波統合モデル $J_{kl}(c,t,f)$ のパラメータは，単音ごとの微小な誤差を許容しつつ，かつ同一楽器ごとの類似性を満たす必要がある．この性質を満たすようなパラメータを推定するために，新たに 2 つの制約を導入する．

第 1 の制約は調波構造モデルの v_{kln} に対する制約で，以下の式で表されるコスト関数として与えられる．

$$Q_{kl}^{(v)} = \sum_n \bar{v}_{kn} \log \frac{\bar{v}_{kn}}{v_{kln}} \quad (24)$$

$\bar{v}_{kn}$ は楽器ごとに v_{kln} を平均したものである．このコストは $\bar{v}_{kn}$ と v_{kln} の KLD で定義されており，これを最小化することで v_{kln} を $\bar{v}_{kn}$ へと近づけることが可能になる．すなわち，v_{kln} の楽器ごとの類似性を強制できる．

第 2 の制約は非調波構造モデル $I'_{kl}(t,f)$ に対するもので，以下の式で表されるコスト関数として与えられる．

$$Q_{kl}^{(\bar{I})} = \iint \bar{I}_k(t,f) \log \frac{\bar{I}_k(t,f)}{I'_{kl}(t,f)}\, dt\, df \quad (25)$$

$\bar{I}_k(t,f)$ は楽器ごとに $I'_{kl}(t,f)$ を平均したものである．このコストは $\bar{I}_k(t,f)$ と $I'_{kl}(t,f)$ の KLD で定義されており，これを最小化することで $I'_{kl}(t,f)$ を $\bar{I}_k(t,f)$ へと近づけることができる．すなわち，$I'_{kl}(t,f)$ の楽器ごとの類似性を強制できる．

5.5 基底関数への分配関数

$J_{kl}(c,t,f)$ は $H_{kl}(t,f)$ と $I_{kl}(t,f)$ の線形結合で定義されており，さらに $H_{kl}(t,f)$ は基底関数であるガウス分布関数の線形結合で定義されている．そこで，パラメータ推定のために，$X_{kl}^{(S)}(c,t,f)$ および $Y_{kl}(t,f)$ を調波構造モデルの m 番目のパワーエンベロープのガウス分布，n 番目の調波構造のガウス分布（以下 (m,n) 番目と記す）および非調波構造モデルへと分配する関数 $S_{klmn}^{(H)}(t,f)$，$S_{kl}^{(I)}(t,f)$ を導入する．この場合，Q'_{kl} は以下の Q''_{kl} に置き換えられることになる．

$$Q''_{kl} = Q''^{(H)}_{kl} + Q''^{(I)}_{kl} \quad (26)$$

ただし，

$$X^{(H)}_{klmn}(c,t,f) = S^{(H)}_{klmn}(t,f)S_{kl}(c,t,f)X(c,t,f) \quad (27)$$

$$X^{(I)}_{kl}(c,t,f) = S^{(I)}_{kl}(t,f)S_{kl}(c,t,f)X(c,t,f) \quad (28)$$

$$J^{(H)}_{klmn}(c,t,f) = r_{kl}(c)w_{kl}w^{(H)}_{kl}E_{klm}(t)F_{kln}(t,f) \quad (29)$$

$$J^{(I)}_{kl}(c,t,f) = r_{kl}(c)w_{kl}w^{(I)}_{kl}I_{kl}(t,f) \quad (30)$$

$$Q''^{(H)}_{kl} = \sum_{c,m,n}\iint X^{(H)}_{klmn}(c,t,f) \cdot \log\frac{X^{(H)}_{klmn}(c,t,f)}{J^{(H)}_{klmn}(c,t,f)}\,dt\,df \quad (31)$$

$$Q''^{(I)}_{kl} = \sum_{c}\iint X^{(I)}_{kl}(c,t,f) \cdot \log\frac{X^{(I)}_{kl}(c,t,f)}{J^{(I)}_{kl}(c,t,f)}\,dt\,df \quad (32)$$

である．これらの分配関数の最適な値は，Lagrange の未定乗数項を加えることで $S_{kl}(c,t,f)$ と同様に導出することが可能で，それぞれ

$$S^{(H)}_{klmn}(t,f) = \frac{w_{kl}w^{(H)}_{kl}E_{klm}(t)F_{kln}(t,f)}{J_{kl}(c,t,f)} \quad (33)$$

$$S^{(I)}_{kl}(t,f) = \frac{w_{kl}w^{(I)}_{kl}I_{kl}(t,f)}{J_{kl}(c,t,f)} \quad (34)$$

となる．

5.6 コスト関数

推定されたパラメータの分離音に対する良し悪しを計る尺度 Q''_{kl}，ここまでに述べた追加コスト $Q^{(Y)}_{kl}$，$Q^{(\omega)}_{kl}$，$Q^{(\tilde{I})}_{kl}$，$Q^{(v)}_{kl}$，$Q^{(\bar{I})}_{kl}$ の重み付き和，さらにモデルパラメータと分配関数の Lagrange の未定乗数項の総和で定義した Q_{kl} を，調波・非調波統合モデルの各々のパラメータに関して最小化することで，モデルパラメータの推定を行う．各々のコストに対する重みは，α，$(1-\alpha)$，β_ω，$\beta_{\tilde{I}}$，β_v，$\beta_{\bar{I}}$ とする．α を最初は $\alpha = 0$ に設定し，徐々に 1 に近づけることで，分離とパラメータ推定の繰返しにおいてモデルの過学習を防ぐことができると考えられる．また，$r_{kl}(c)$，w_{kl}，$(w^{(H)}_{kl}, w^{(I)}_{kl})$，$u_{klm}$，$v_{kln}$，$\omega_{kl}(t)$，$I'_{kl}(t,f)$，$(S^{(H)}_{klmn}(t,f),\ S^{(I)}_{kl}(t,f))$ の各々に関する Lagrange の未定乗数を $\lambda^{(r)}_{kl}$，$\lambda^{(w)}$，$\lambda^{(wHI)}_{kl}$，$\lambda^{(u)}_{kl}$，$\lambda^{(v)}_{kl}$，$\lambda^{(\omega)}_{kl}$，$\lambda^{(I)}_{kl}$，$\lambda^{(SHI)}_{kl}(t,f)$ とする．

さらに，Q_{kl} をすべての (k,l) に対して合計したコスト関数 Q を定義すると，分離とパラメータ推定の両方をコスト関数 Q の最小化としてとらえることができる．分離に関係がある項は Q''_{kl} だけであるので，重み α がかけられていたり，$Q^{(Y)}_{kl}$ などのその他のコストが加えられていても Q を最小化して得られる分配関数には影響しない．また，パラメータ推定に関しては，(k,l) 番目の単音のモデルパラメータ推定に関係がある項は Q_{kl} だけであるので，Q_{kl} 以外の Q の項は $J_{kl}(c,t,f)$ のパラメータ推定には影響しない．したがって，分配関数およびモデルパラメータに関する Q の最小化を交互に繰り返すことで，分離とパラメータ推定を行うことができる．

Q は，以下の式で定義される．

$$\begin{aligned} Q = {} & \sum_{k,l} Q_{kl} \\ & - \lambda^{(w)}\left(\sum_{k,l} w_{kl} - X_0\right) \\ & - \sum_{c}\iint \lambda^{(S)}(c,t,f) \\ & \cdot \left(\sum_{k,l} S_{kl}(c,t,f) - 1\right) dt\,df \end{aligned} \quad (35)$$

$$\begin{aligned} Q_{kl} = {} & \alpha Q''_{kl} + (1-\alpha)Q^{(Y)}_{kl} + \beta_\omega Q^{(\omega)}_{kl} \\ & + \beta_{\tilde{I}} Q^{(\tilde{I})}_{kl} + \beta_v Q^{(v)}_{kl} + \beta_{\bar{I}} Q^{(\bar{I})}_{kl} \\ & - \lambda^{(r)}_{kl}\left(\sum_{c} r_{kl}(c) - \mathrm{C}\right) \\ & - \lambda^{(wHI)}_{kl}\left(w^{(H)}_{kl} + w^{(I)}_{kl} - 1\right) \\ & - \lambda^{(u)}_{kl}\left(\sum_{m} u_{klm} - 1\right) \\ & - \lambda^{(v)}_{kl}\left(\sum_{n} v_{kln} - 1\right) \\ & - \lambda^{(\omega)}_{kl}\int \left(\omega_{kl}(t) - \tilde{\omega}_{kl}(t)\right) dt \\ & - \lambda^{(I)}_{kl}\left(\iint I'_{kl}(t,f)\,dt\,df - 1\right) \\ & - \iint \lambda^{(SHI)}_{kl}(t,f) \\ & \cdot \left(\sum_{m,n} S^{(H)}_{klmn}(t,f) + S^{(I)}_{kl}(t,f) - 1\right) dt\,df \end{aligned} \quad (36)$$

ただし，コスト $Q^{(\omega)}$，$Q^{(\tilde{I})}$ に関しては，これらを追加すると分離とパラメータ推定の反復を行う過程で Q のパラメータに関する極小点がパラメータの値によって変化してしまうため，パラメータの局所的な収束性が保証されなくなる．しかしながら，本論文で実施した実験においては非調波構造モデルが調波構造を表現したり F0 時系列が不連続になったりすることはなく，かつパラメータが発散している様子は確認できなかっ

表 1 本論文で用いる記号の一覧
Table 1 List of symbols.

記号	意味	備考
c	チャネル番号	$c \in \{1, \ldots, \mathrm{C}\}$
t	時刻	$t \in [\mathrm{T}_0, \mathrm{T}_1]$
f	周波数	$f \in [\mathrm{F}_0, \mathrm{F}_1]$
k	楽器番号	$k \in \{1, \ldots, \mathrm{K}\}$
l	単音番号	$\forall k, \exists \mathrm{L}_k,\ l \in \{1, \ldots, \mathrm{L}_k\}$
m	パワーエンベロープを表現するガウス分布関数の番号	$m \in \{1, \ldots, \mathrm{M}\}$
n	調波構造を表現するガウス分布関数の番号（第 n 次倍音に相当）	$n \in \{1, \ldots, \mathrm{N}\}$
$J_{kl}(c,t,f)$	(k,l) 番目の単音の調波・非調波統合モデル	
$J'_{kl}(t,f)$	チャネル間音圧比 $r_{kl}(c)$ を取り除いた $J_{kl}(c,t,f)$	
$H_{kl}(t,f)$	調波構造モデル	
$E_{klm}(t)$	調波構造モデルのパワーエンベロープ関数	
$F_{kln}(t,f)$	調波構造モデルの調波構造関数	
$I_{kl}(t,f)$	非調波構造モデル	
$X(c,t,f)$	入力パワースペクトル	
$Y_{kl}(t,f)$	テンプレートスペクトル	
$S_{kl}(c,t,f)$	(k,l) 番目の単音へのパワースペクトル分配関数	$\forall c,t,f,\ \sum_{k,l} S_{kl}(c,t,f) = 1$
$S^{(H)}_{klmn}(t,f)$	(k,l) 番目の単音，(m,n) 番目の調波構造モデルのガウス分布へのパワースペクトル分配関数	$\forall k,l,t,f,\ \sum_{m,n} S^{(H)}_{klmn}(t,f) + S^{(I)}(t,f) = 1$
$S^{(I)}_{kl}(t,f)$	(k,l) 番目の単音の非調波構造モデルへのパワースペクトル分配関数	
$r_{kl}(c)$	チャネル間音圧比	$\forall k,l,\ \sum_c r_{kl}(c) = \mathrm{C}$
w_{kl}	調波・非調波統合モデル全体の重み	$\sum_{k,l} w_{kl} = 1$
$w^{(H)}_{kl}$	調波構造モデルの重み	$\forall k,l,\ w^{(H)}_{kl} + w^{(I)}_{kl} = 1$
u_{klm}	パワーエンベロープを表現する m 番目のガウス分布関数の重み係数	$\forall k,l,\ \sum_m u_{klm} = 1$
v_{kln}	n 次高調波成分の相対強度	$\forall k,l,\ \sum_n v_{kln} = 1$
τ_{kl}	発音時刻	
ϕ_{kl}	パワーエンベロープを表現するガウス分布関数の広がりを表すパラメータ	
$\omega_{kl}(t)$	F0 時系列	
σ_{kl}	高調波成分を表現するガウス分布関数の広がりを表すパラメータ	
$w^{(I)}_{kl}$	非調波構造モデルの重み	
$I'_{kl}(t,f)$	パワーを正規化した非調波構造モデル	$\forall k,l,\ \iint I'_{k,l}(t,f)\,dt\,df = 1$

たため，実験的には大きな問題にはならないと考える．

本論文で用いる記号の一覧を，**表 1** に示す．

5.7 EM アルゴリズムとしての解釈

ここまでに述べた分離とパラメータ推定の繰返しは，Expectation-Maximization（EM）アルゴリズムを用いた最大事後確率推定として解釈することもできる．

観測確率密度関数 $p(c,t,f)$ が与えられたとき，$p(c,t,f)$ を近似する確率密度分布 $p(k,l,c,t,f|\theta)$ のパラメータ θ を推定する問題を考える．k，l に関する隠れ変数の分布 $p(k,l|c,t,f)$ および θ の事前分布 $p(\theta)$ を導入すると，θ を $\tilde{\theta}$ へと変化させたときの対数事後確率の期待値の増加量 $Q(\theta,\tilde{\theta})$ は，

$$\begin{aligned} Q(\theta,\tilde{\theta}) = \sum_{k,l,c} \iint & p(k,l|c,t,f,\theta) p(c,t,f) \\ & \cdot \log p(k,l,c,t,f|\tilde{\theta})\,dt\,df \\ & + \log p(\tilde{\theta}) \end{aligned} \tag{37}$$

と表される．

ここで，入力パワースペクトル $X(c,t,f)$，パワースペクトル分配関数 $S_{kl}(c,t,f)$，調波・非調波統合モデル $J_{kl}(c,t,f)$ をそれぞれ $p(c,t,f)$，$p(k,l|c,t,f,\theta)$，$p(k,l,c,t,f|\theta)$ に対応させ[⋆1]，さらにモデルに関する制約を表現するコスト関数 $Q^{(Y)}_{kl}$，$Q^{(\omega)}_{kl}$，$Q^{(\tilde{I})}_{kl}$，$Q^{(v)}_{kl}$，$Q^{(\bar{I})}_{kl}$ の総和を $-\log p(\theta)$ に対応させると，EM アルゴリズムにおける E ステップ，すなわち $Q(\theta,\tilde{\theta})$ を最大化する $p(k,l|c,t,f,\theta)$ の推定はコスト関数 Q を最小化する $S_{kl}(c,t,f)$ の導出に対応する．同様に，EM アルゴリズムにおける M ステップ，すなわち $Q(\theta,\tilde{\theta})$ を最大化する $\tilde{\theta}$ の推定はコスト関数を最小化するモデルの各パラメータの導出に対応する．

6. 評価実験

本手法の性能を確認するため，評価実験を行った．

6.1 実験の目的

本実験の目的は，本論文で構築した調波・非調波統

⋆1 ただし，入力パワースペクトルとモデルはあらゆる変数に関して積分した値が 1 になるように正規化したものを対応させる．

合モデルの有効性を確認することである．具体的には，以下の 3 つの条件：

(1) 調波・非調波統合モデルを用いた場合（本手法）
(2) 調波構造モデルのみを用いた場合
(3) 非調波構造モデルのみを用いた場合

において楽曲の音響信号を楽器パートごとおよび単音ごとに分離し，単音ごとにに分離したパワースペクトルとミックス前の各単音のパワースペクトルとの周波数領域での SNR を用いて 3 つの条件を比較した．(k,l) 番目の単音に関する周波数領域での SNR は，次式で定義する．

$$\mathrm{SNR}_{kl} = \frac{1}{\mathrm{C}(\mathrm{T}_1 - \mathrm{T}_0)} \sum_c \int \mathrm{SNR}_{kl}(c,t)\, dt \tag{38}$$

$$\mathrm{SNR}_{kl}(c,t) = \int \frac{X_{kl}^{(S)}(c,t,f)^2}{\left(X_{kl}^{(S)}(c,t,f) - X_{kl}^{(R)}(c,t,f)\right)^2}\, df \tag{39}$$

ただし，$X_{kl}^{(S)}(c,t,f)$ は (k,l) 番目の単音の分離パワースペクトル，$X_{kl}^{(R)}(c,t,f)$ は (k,l) 番目の単音の参照パワースペクトル，すなわちミックス前のパワースペクトルである．

本来，調波構造モデルでドラム音を表現することは難しいが，本実験では用いたモデル以外の条件を同一にするために，調波構造モデルのみを用いた場合でもドラムパートを含む混合音の分離を行った．また，非調波構造モデルのみを用いた場合は，式 (23) の非調波構造モデルを平滑化する制約を用いるとモデルが調波構造を表現することができなくなるため，この制約は用いていない．

6.2 実験データ

実験には，RWC 研究用音楽データベース：ポピュラー音楽（RWC-MDB-P-2001）[29] から選んだ 10 曲（No.1–10）を用いた．各楽曲は開始から 30 秒の区間を利用した．楽曲ごとの K と L_k の値は，各楽曲の SMF で用いられた楽器数，およびノートオンメッセージ数から求める．本実験では以下の理由により，MIDI 音源から録音した音響信号を分離対象とした．

- 本実験の目的に適した音楽データベースが入手困難である．目的に適したデータベースとは，楽曲の音響信号，それに同期がとられた SMF，さらに各楽器パートの音響信号をミックスする前のマスタートラックのすべてが利用できるものである．これは，ミックス前の音響信号に対する分離後の音響信号の SNR を測定し，定量的評価を行うために必要である．
- 一般的なポピュラー音楽には歌声が含まれているが，1 章で述べたように歌声は本手法で扱う対象には含まれない．

表 2 実験条件
Table 2 Experimental conditions.

Frequency analysis	
sampling rate	44.1 kHz
analysizing method	STFT
STFT window	2,048 points Gaussian
STFT shift	441 points
Parameters	
# of channels: C	2
# of kernels in E_{klm}: M	10
# of partials: N	20
β_v	0.1
β_ω	0.1
$\beta_{\bar{I}}$	3.5
$\beta_{\tilde{I}}$	0.5
MIDI sound generator	
test data	YAMAHA MU2000
template sounds	Roland SD-90

テンプレート音と分離対象となる楽曲とは，異なる楽器メーカの MIDI 音源で生成した．これは，分離対象とは確実に音色（音源波形）が異なるテンプレート音を用いるためである．また周波数解析，ハイパーパラメータ，用いた MIDI 音源に関する条件を**表 2** に示す．この表における β_v などのパラメータは，実験的に最適なものを求めたものである．

6.3 実験結果

表 3 に，統合モデルを用いて分離を行った場合（Integrated），調波構造モデルのみを用いて分離を行った場合（Harmonic），および非調波構造モデルのみを用いて分離を行った場合（Inharmonic）に得た各単音ごとの分離音を周波数領域での SNR で評価し，同種の楽器ごとに平均した結果を示す．Piano, Guitar, ..., Drums は楽器の種類を，1–8 などの番号は楽器に対応する MIDI のプログラムナンバを表す．楽器の種類は，MIDI のプログラムナンバを基準に，ピアノ（プログラムナンバ 1–8），ギター（プログラムナンバ 24–32）のように分類した．ドラムには，プログラムナンバーでは分類しておらず，SMF でドラムトラックであることを指定されたトラックに属する単音を分類した．ただし**表 4** では，実験に用いた 10 曲中の 5 曲以上に用いられている楽器パートについてのみ結果を示している．すべての楽器パートにおいて，Integrated の SNR が最も大きくなっている．

Integrated と Inharmonic に比べて Harmonic の SNR が全体的に小さな値になっており，特に Bass と Drums において SNR の低下が顕著である．Drums に関しては調波構造モデルでドラム音を表現することは

表 3 楽器パートごとの平均 SNR
Table 3 Average SNR of each instrument part.

Inst. Part	Prog. # in MIDI	SNR (dB) Integrated	Harmonic	Inharmonic
Piano	1 – 8	48.87	25.67	48.17
Guitar	25 – 32	47.22	13.59	46.88
Bass	33 – 40	40.93	−22.19	40.53
Ensemble	49 – 56	48.11	31.45	47.45
Pipe	73 – 80	49.28	26.78	49.16
Drums	—	43.31	−18.02	42.56

表 4 楽器パートごとの平均 KLD
Table 4 Average KLD of each instrument part.

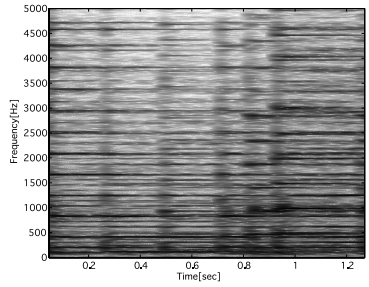
Inst. Part	KLD Initial	Estimated	Ideal
Piano	402.82	21.25	12.68
Guitar	342.92	28.12	13.72
Bass	1152.2	44.93	25.45
Ensemble	520.67	32.98	13.22
Pipe	504.57	23.73	13.24
Drums	842.32	56.41	25.39

困難なことが，Bass に関しては低音部を担当する楽器が多く，その基本周波数に対して周波数解像度が粗いために正しく調波構造を表現することができなかったことが，それぞれ原因であると考えられる．これは Integrated と Inharmonic の場合にも同様にあてはまる．この問題に対しては，STFT よりも周波数解像度を詳細に制御可能な連続ウェーブレット変換などを周波数解析手法に採用することで対処できると考えられる．また，Bass と Drums 以外の楽器パートでも Harmonic の SNR が低下しているが，上記 2 つの楽器パートを正しく分離することができなかった結果，それ以外のパートにドラム音などが混入したためと考えられる．

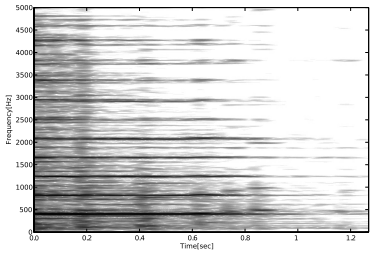
楽器パートごとに，Integrated と Harmonic，Integrated と Inharmonic のそれぞれの組において SNR の平均値が等しいという帰無仮説を立てて t 検定を行ったところ，Integrated と Inharmonic 間の Pipe 以外の SNR に関してはすべて有意水準 0.05 で帰無仮説を棄却することができたため，この結果は SNR の向上という点において統合モデルを用いることの有用性を示している．SNR に優位な差が現れなかった原因としては，Pipe のパートの多くの単音がフルートでかつフルートがメロディ（ボーカル）に割り当てられており，単音の音量が大きくかつ他の楽器パートとの高調波成分の重複が起こりにくいため，統合モデルを用いなくても単音のパワースペクトルを容易に推定できたことがあげられる．

表 4 に，テンプレートでパラメータを初期化した直後のモデル（Initial），それを混合音に適応させたモデル（Estimated），テストデータの各単音でパラメータを初期化したモデル（Ideal）に関して，テストデータの各単音とモデルとの KLD を同種の楽器ごとに平均した結果を示す．楽器の分類は表 3 の場合と同様である．Estimated の KLD が Ideal のものに近ければ推定されたパラメータが「良い」と考えられるが，それぞれの単音に関しての KLD が非負となることは保証されていないため，必ずしも KLD は小さいほど良いというわけではなく，KLD の大小はあくまで良し悪しの目安であることに注意する必要がある．表 4 を見ると，Estimated の KLD は Initial のものよりもかなり Ideal に近い値を示しており，ある程度は妥当なパラメータ推定が行われたと考えられる．詳細なデータは省略するが，モデルとテストデータの単音間の KLD が 150–200 を超える，もしくは 0 を下回ると SNR が大きく減少する傾向が見られた．

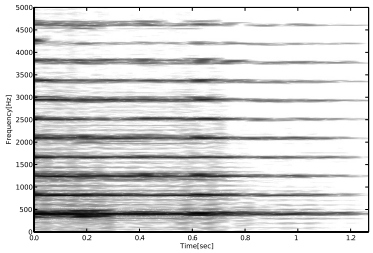
また，**図 2**，**図 3** に，統合モデル，調波構造モデル，非調波構造モデルのそれぞれを用いて分離されたピアノ単音，スネアドラム単音のパワースペクトル（それぞれ図 {2, 3}-{b, c, d}），該当する区間の混合音およびミックス前のピアノ単音のパワースペクトル（それぞれ図 {2, 3}-{a, e}）を示す．これらの図を見ると，いずれのモデルを用いて分離を行った場合でも調波構造はある程度表現されていることが分かる．しかし，調波構造モデルを用いた場合では発音時刻付近に現れている非調波成分が少ない点，非調波構造モデルを用いた場合では調波構造がミックス前のものよりも早く減衰している点，本来の調波構造以外の周波数成分が含まれている点などが，ミックス前の単音とは異なっている．これに対して，統合モデルを用いた場合では，これら 3 つのモデルによる分離音の中では上記の点のようなミックス前の単音とのずれは最も小さい．調波構造の減衰が非調波構造モデルを用いた場合よりも遅い原因は，調波構造モデルの高調波成分相対強度が時間の経過によって変化しないように定義されており，高次の高調波成分だけが先に減衰しないことがあげられる．また，統合モデルや非調波構造モデルを用いた場合には，発音時刻以外の時刻においてミックス前の単音には見られない非調波成分が存在している．この非調波成分は本来は他のピアノ音やドラム音の非調波成分であるが，非調波構造モデルが混合音に過剰に適応した結果混入したと考えられる．これは非調波構造モデルの自由度の高さに由来しているため，非調波構造モデルに構造上の制約を持たせる，もしくは何らかの事前分布を用いるなどによって解決されると考えられる．



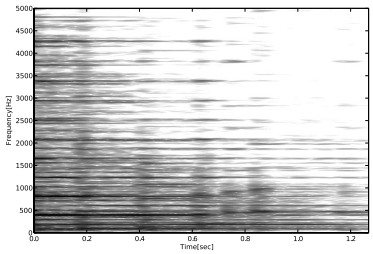
a：混合音



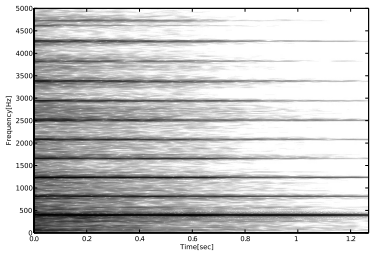
b：統合モデルを用いて分離したもの



c：調波構造モデルを用いて分離したもの

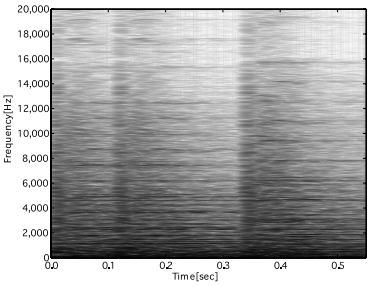


d：非調波構造モデルを用いて分離したもの

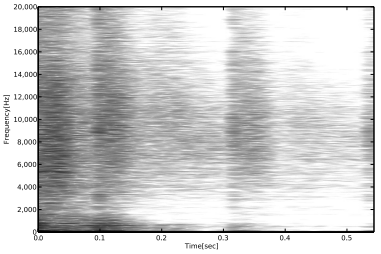


e：ミックス前のもの

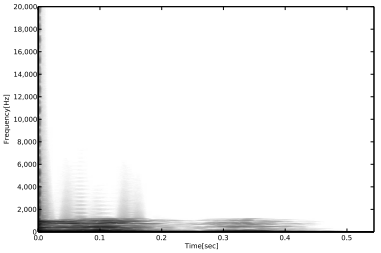
図 2 ピアノ単音のパワースペクトル
Fig. 2 Power spectrogram of a piano sound.



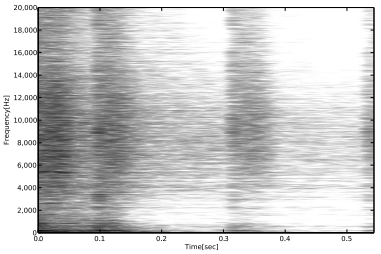
a：混合音



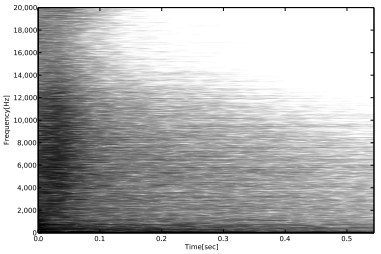
b：統合モデルを用いて分離したもの



c：調波構造モデルを用いて分離したもの



d：非調波構造モデルを用いて分離したもの



e：ミックス前のもの

図 3 スネアドラム単音のパワースペクトル
Fig. 3 Power spectrogram of a snare drum sound.

7. おわりに

本論文では，調波的な音と非調波的な音の混合音をすべての楽器パートに分離するという問題に取り組んだ．具体的な手法として，調波構造モデルと非調波構造モデルを統合したモデルを用いた多重奏の音源分離手法と，そのためのモデル適応手法について述べた．また，調波構造モデルと非調波構造モデルを統合する際の問題点を非調波構造モデルへの制約という形でコスト関数最小化によるパラメータ推定の枠組みを崩すことなく解決した．本手法の性能を示すための評価実験では，統合モデルを用いることの有効性を確認した．

また我々は，本手法で生成した楽器パートごとの分離音を用いるアプリケーションとして，1 章で述べた楽器音イコライザを開発した．楽器音イコライザを用いることで，ユーザは楽曲の楽器パートごとの音量バランスを自由に操作することができ，その結果として能動的な音楽鑑賞[1)] が可能となる．

本論文で設計した調波・非調波統合モデルは，音源分離への利用のみに限定されるものではない．たとえば，本モデルを用いて，多重音解析や楽器音認識へと応用範囲を広げることが考えられる．また，任意のパワースペクトルを表現できる非調波構造モデルの性質上，扱う対象も音楽音響信号に限定されるものではなく，音声や環境音へとその対象を広げることも可能である．今後は，歌声の分離など，扱える信号の対象を増やし分離性能を改善することと，SMF から得られる事前情報（各単音の音高，音長，音量，発音時刻，楽器）が一部利用できない，もしくは利用はできるが信頼性が低いといった状況における分離手法を構築することで，分離技術の汎用的を高めることに取り組んでいく予定である．

謝辞 本研究の一部は，科学研究費補助金（基盤研究 (S)，特定領域「情報爆発 IT 基盤」)，21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」，科学技術振興機構 CrestMuse プロジェクトによる支援を受けた．

参考文献

1) Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. ICASSP*, Vol.IV, pp.1441–1444 (2007).
2) Yoshii, K., Goto, M. and Okuno, H.G.: INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, *Proc. EWIMT*, pp.205–212 (2005).
3) Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Drumix: An Audio Player with Real-time Drum-part Rearrangement Functions for Active Music Listening, *IPSJ Journal*, Vol.48, No.3, pp.134–144 (2007).
4) ヤマハミュージックイークラブ．http://www.music-eclub.com/
5) internet MIDILINK. http://www.midilink.com/
6) Cano, P., Loscos, A. and Bonada, J.: Score-Performance Matching Using HMMs, *Proc. ICMC*, pp.441–444 (1999).
7) Dannenberg, R.B. and Hu, N.: Polyphonic Audio Matching for Score Following and Intelligent Audio Editors, *Proc. ICMC*, pp.27–33 (2003).
8) Adams, N., Marquez, D. and Wakefield, G.: Iterative Deepening for Melody Alignment and Retrieval, *Proc. ISMIR*, pp.199–206 (2005).
9) Dixon, S. and Widmer, G.: MATCH: A Music Alignment Tool Chest, *Proc. ISMIR*, pp.492–497 (2005).
10) Cont, A.: Realtime Audio to Score Alignment for Polyphonic Music Instruments Using Sparce Non-negative Constraints and Hierarchical HMMs, *Proc. ICASSP*, Vol.II, pp.641–644 (2006).
11) Virtanen, T. and Klapuri, A.: Separation of Harmonic Sound Sources Using Sinusoidal Modeling, *Proc. ICASSP*, Vol.II, pp.765–768 (2000).
12) Virtanen, T. and Klapuri, A.: Separation of Harmonic Sounds Using Linear Models for the Overtone Series, *Proc. ICASSP*, Vol.2, pp.1757–1760 (2002).
13) Every, M. and Szymanski, J.: A Spectral-filtering Approach to Music Signal Separation, *Proc. DAFx*, pp.197–200 (2004).
14) Kameoka, H., Nishimoto, T. and Sagayama, S.: Harmonic-temporal Structured Clustering via Deterministic Annealing EM Algorithm for Audio Feature Extraction, *Proc. ISMIR*, pp.115–122 (2005).
15) Viste, H. and Evangelista, G.: A Method for Separation of Overlapping Partials Based on Similarity of Temporal Envelopes in Multichannel Mixtures, *IEEE Trans. Speech and Audio Processing*, Vol.14, No.3, pp.1051–1061 (2006).
16) Woodruff, J., Pardo, B. and Dannenberg, R.: Remixing Stereo Music with Score-informed Source Separation, *Proc. ISMIR*, pp.314–319 (2006).
17) Smaragdis, P. and Brown, J.C.: Non-negative Matrix Factorization for Polyphonic Music

Transcription, *Proc. WASPAA*, pp.177–180 (2003).
18) Fitzgerald, D., Cranitch, M. and Coyle, E.: Sound Source Separation using Shifted Non-negative Tensor Factorization, *Proc. ICASSP*, Vol.V, pp.653–656 (2006).
19) Uhle, C., Dittmar, C. and Sporer, T.: Extraction of Drum Tracks from Polyphonic Music Using Independent Subspace Analysis, *Proc. ICA*, pp.834–848 (2003).
20) Helen, M. and Virtanen, T.: Separation of Drums from Polyphonic Music Using Non-negative Matrix Factorization and Support Vector Machine, *Proc. EUSIPCO* (2005).
21) Barry, D., Fitzgerald, D., Coyle, E. and Lawlor, B.: Drum Source Separation Using Percussive Feature Detection and Spectral Modulation, *Proc. ISSC*, pp.13–17 (2005).
22) Casey, M. and Westner, A.: Separation of Mixed Audio Sources by Independent Subspace Analysis, *Proc. ICMC*, pp.154–161 (2000).
23) Dubnov, S.: Extracting Sound Objects by Independent Subspace Analysis, *Proc. AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio* (2002).
24) FItzGerald, D., Coyle, E. and Lawlor, B.: Independent Subspace Analysis Using Locally Linear Embedding, *Proc. DAFx*, pp.13–17 (2003).
25) Brown, J.C. and Smaragdis, P.: Independent Component Analysis for Automatic Note Extraction from Musical Trills, *J. Acoust. Soc. Am.*, Vol.115, No.5, pp.2295–2306 (2004).
26) Barry, D., FitzGerald, D. and Lawlor, B.: Single Channel Source Separation Using Short-time Independent Component Analysis, *Proc. AES* (2005).
27) Yoshii, K., Goto, M. and Okuno, H.G.: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates with Harmonic Structure Suppression, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.15, No.1, pp.333–345 (2007).
28) Goto, M.: A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication* (*ISCA Journal*), Vol.43, No.4, pp.311–329 (2004).
29) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical and Jazz Music Databases, *Proc. ISMIR*, pp.287–288 (2002).

付　　録

A.1 パラメータ更新式の導出

式 (35) のコスト関数を各パラメータで偏微分したものの零点を求めることで，J が極小になるようにパラメータを更新する式を導出する．

A.1.1 $r_{kl}(c)$：各チャネルの相対強度

$$\frac{\partial J}{\partial r_{kl}(c)} = \sum_{m,n} \iint \left(-\frac{X^{(H)}_{klmn}}{r_{kl}(c)} \right) dt\, df + \iint \left(-\frac{X^{(I)}_{kl}}{r_{kl}(c)} \right) dt\, df + w_{kl} - \lambda^{(r)}_{kl} = 0 \quad (40)$$

$$\frac{\partial J}{\partial \lambda^{(r)}_{kl}} = \sum_{c} r_{kl}(c) - \mathrm{C} = 0 \quad (41)$$

この連立方程式を解き，以下を得る．

$$r_{kl}(c) = \frac{\mathrm{C} \iint \left(\sum_{m,n} X^{(H)}_{klmn} + X^{(I)}_{kl} \right) dt\, df}{\sum_{c} \iint \left(\sum_{m,n} X^{(H)}_{klmn} + X^{(I)}_{kl} \right) dt\, df} \quad (42)$$

A.1.2 w_{kl}：統合モデル全体の重み

$$\frac{\partial J}{\partial w_{kl}} = \sum_{c,m,n} \iint \left(-\frac{X^{(H)}_{klmn}}{w_{kl}} \right) dt\, df + \sum_{c} \iint \left(-\frac{X^{(I)}_{kl}}{w_{kl}} \right) dt\, df + \mathrm{C} - \lambda^{(w)} = 0 \quad (43)$$

$$\frac{\partial J}{\partial \lambda^{(w)}} = \sum_{k,l} w_{kl} - X_0 = 0 \quad (44)$$

この連立方程式を解き，以下を得る．

$$w_{kl} = \iint \left(\sum_{m,n} X^{(H)}_{klmn} + X^{(I)}_{kl} \right) dt\, df \quad (45)$$

A.1.3 $w^{(H)}_{kl}$，$w^{(I)}_{kl}$：調波・非調波構造モデルの重み

$$\frac{\partial J}{\partial w^{(H)}_{kl}} = \sum_{c,m,n} \iint \left(-\frac{X^{(H)}_{klmn}}{w^{(H)}_{kl}} \right) + \mathrm{C} w_{kl} - \lambda^{(wHI)}_{kl} = 0 \quad (46)$$

$$\frac{\partial J}{\partial w^{(I)}_{kl}} = \sum_{c} \iint \left(-\frac{X^{(I)}_{kl}}{w^{(I)}_{kl}} \right) dt\, df + \mathrm{C} w_{kl} - \lambda^{(wHI)}_{kl} = 0 \quad (47)$$

$$\frac{\partial J}{\partial \lambda^{(wHI)}_{kl}} = w^{(H)}_{kl} + w^{(I)}_{kl} - 1 = 0 \quad (48)$$

この連立方程式を解き，以下を得る．

$$w_{kl}^{(H)} = \frac{\sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df}{\sum_c \iint \left(\sum_{m,n} X_{klmn}^{(H)} + X_{kl}^{(I)} \right) dt\, df} \tag{49}$$

$$w_{kl}^{(I)} = \frac{\sum_{c} \iint X_{kl}^{(I)}\, dt\, df}{\sum_c \iint \left(\sum_{m,n} X_{klmn}^{(H)} + X_{kl}^{(I)} \right) dt\, df} \tag{50}$$

A.1.4 $\omega_{kl}(t)$：F0 の軌跡

$$\frac{\partial J}{\partial \omega_{kl}(t)} = \sum_{c,m,n} \int -\frac{n\,(f - n\omega_{kl}(t))\, X_{klmn}^{(H)}}{\sigma_{kl}^2}\, df + \beta_\omega \left(-\frac{\tilde{\omega}_{kl}(t)}{\omega_{kl}(t)} + 1 \right) = 0 \tag{51}$$

$$\Rightarrow \quad a_\omega \omega_{kl}(t)^2 + b_\omega \omega_{kl}(t) + c_\omega = 0 \tag{52}$$

$$\begin{cases} a_\omega = \sum_{c,m,n} \int n^2 X_{klmn}^{(H)}\, df \\ b_\omega = \sigma_{kl}^2 \beta_\omega - \sum_{c,m,n} \int n f X_{klmn}^{(H)}\, df \\ c_\omega = -\sigma_{kl}^2 \beta_\omega \tilde{\omega}_{kl}(t) \end{cases} \tag{53}$$

この方程式を解き，以下を得る．

$$\omega_{kl}(t) = \frac{-b_\omega + \sqrt{b_\omega^2 - 4a_\omega c_\omega}}{2a_\omega}$$

A.1.5 u_{klm}：パワーエンベロープの概形

$$\frac{\partial J}{\partial u_{klm}} = \sum_{c,n} \iint -\frac{X_{klmn}^{(H)}}{u_{klm}}\, dt\, df - \lambda_{kl}^{(u)} = 0 \tag{54}$$

$$\frac{\partial J}{\partial \lambda_{kl}^{(u)}} = \sum_m u_{klm} - 1 = 0 \tag{55}$$

この連立方程式を解き，以下を得る．

$$u_{klm} = \frac{\sum_{c,n} \iint X_{klmn}^{(H)}\, dt\, df}{\sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df} \tag{56}$$

A.1.6 v_{kln}：n 次倍音成分の相対強度

$$\frac{\partial J}{\partial v_{kln}} = \sum_{c,m} \iint -\frac{X_{klmn}^{(H)}}{v_{kln}}\, dt\, df + \beta_v \left(-\frac{\bar{v}_{kn}}{v_{kln}} + 1 \right) - \lambda_{kl}^{(v)} = 0 \tag{57}$$

$$\frac{\partial J}{\partial \lambda_{kl}^{(v)}} = \sum_n v_{kln} - 1 = 0 \tag{58}$$

この連立方程式を解き，以下を得る．

$$v_{kln} = \frac{\beta_v \bar{v}_{kn} + \sum_{c,m} \iint X_{klmn}^{(H)}\, dt\, df}{\beta_v + \sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df} \tag{59}$$

A.1.7 τ_{kl}：発音時刻

$$\frac{\partial J}{\partial \tau_{kl}} = \sum_{c,m,n} \iint -X_{klmn}^{(H)} \frac{t - \tau_{kl} - m\phi_{kl}}{\phi_{kl}^2}\, dt df = 0 \tag{60}$$

この方程式を解き，以下を得る．

$$\tau_{kl} = \frac{\sum_{c,m,n} \iint (t - m\phi_{kl})\, X_{klmn}^{(H)}\, dt\, df}{\sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df} \tag{61}$$

A.1.8 $\mathbf{Y}\phi_{kl}$：音長

$$\frac{\partial J}{\partial \phi_{kl}} = \sum_{c,m,n} \iint X_{klmn}^{(H)} \cdot \frac{(t - \tau_{kl})\,(t - \tau_{kl} - m\phi_{kl}) - \phi_{kl}^2}{\phi_{kl}^3}\, dt\, df = 0 \tag{62}$$

$$\Rightarrow \quad a_\phi \phi_{kl}^2 + b_\phi \phi_{kl} + c_\phi = 0 \tag{63}$$

$$\begin{cases} a_\phi = \sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df \\ b_\phi = \sum_{c,m,n} \iint m\,(t - \tau_{kl})\, X_{klmn}^{(H)}\, dt\, df \\ c_\phi = -\sum_{c,m,n} \iint (t - \tau_{kl})^2\, X_{klmn}^{(H)}\, dt\, df \end{cases} \tag{64}$$

この方程式を解き，以下を得る．

$$\phi_{kl} = \frac{-b_\phi + \sqrt{b_\phi^2 - 4a_\phi c_\phi}}{2a_\phi} \tag{65}$$

A.1.9 σ_{kl}：周波数方向の広がり

$$\frac{\partial J}{\partial \sigma_{kl}} = \sum_{c,m,n} \iint -X_{klmn}^{(H)} \cdot \frac{-n^2\sigma_{kl}^2 + (f - n\omega_{kl}(t))^2}{n^2\sigma_{kl}^3}\, dt\, df = 0 \tag{66}$$

この方程式を解き，以下を得る．

$$\sigma_{kl} = \sqrt{\frac{\sum_{c,m,n} \iint (f - n\omega_{kl}(t))^2 X_{klmn}^{(H)}\, dt\, df}{\sum_{c,m,n} \iint X_{klmn}^{(H)}\, dt\, df}} \tag{67}$$

A.1.10 $I'_{kl}(t,f)$：非調波構造モデル

$$\frac{\partial J}{\partial I'_{kl}(t,f)} = \sum_c \left(-\frac{X_{kl}^{(I)}}{I'_{kl}(t,f)} + r_{kl}(c) \right) + \beta_{\bar{I}} \left(-\frac{\bar{I}_k}{I'_{kl}(t,f)} + 1 \right) + \beta_{\tilde{I}} \left(-\frac{\tilde{I}_{kl}}{I'_{kl}(t,f)} + 1 \right) = 0 \tag{68}$$

$$\frac{\partial J}{\partial \lambda_{kl}^{(I)}} = \iint I'_{kl}(t,f)\,dt\,df - 1 = 0 \quad (69)$$

この方程式を解き，以下を得る．

$$I_{kl} = \frac{\mathrm{C}X_{kl}^{(I)} + \beta_{\bar{I}}\bar{I}_k + \beta_{\tilde{I}}\tilde{I}_{kl}}{\mathrm{C}\iint X_{kl}^{(I)}\,dt\,df + \beta_{\bar{I}} + \beta_{\tilde{I}}} \quad (70)$$

(平成 19 年 6 月 21 日受付)
(平成 19 年 12 月 4 日採録)

糸山 克寿（学生会員）

2006 年京都大学工学部情報学科卒業，現在，京都大学大学院情報学研究科知能情報学専攻修士課程に在籍中．2008 年より日本学術振興会特別研究員（DC1）．音楽情報処理，音楽鑑賞インタフェース等の研究に従事．

後藤 真孝（正会員）

1998 年早稲田大学大学院理工学研究科博士後期課程修了．博士（工学）．同年電子技術総合研究所（2001 年に産業技術総合研究所に改組）に入所し，現在，主任研究員．2000 年から 2003 年まで科学技術振興事業団さきがけ研究 21「情報と知」領域研究員，2005 年から筑波大学大学院准教授（連携大学院）を兼任．音楽情報処理，音声言語情報処理等に興味を持つ．2000 年 WISS2000 論文賞・発表賞，2001 年日本音響学会粟屋潔学術奨励賞・ポスター賞，2003 年インタラクション 2003 ベストペーパー賞，2005 年情報処理学会論文賞，2007 年第 6 回ドコモ・モバイル・サイエンス賞基礎科学部門優秀賞等 21 件受賞．電子情報通信学会，日本音響学会，日本音楽知覚認知学会各会員．

駒谷 和範（正会員）

1998 年京都大学工学部情報工学科卒業．2000 年同大学院情報学研究科知能情報学専攻修士課程修了．2002 年同大学院博士後期課程修了．京都大学博士（情報学）．同年京都大学情報学研究科助手．2007 年より助教．音声対話システムの研究に従事．情報処理学会平成 16 年度山下記念研究賞，FIT2002 ヤングリサーチャー賞等受賞．電子情報通信学会，言語処理学会，人工知能学会，ISCA，ACL 各会員．

尾形 哲也（正会員）

1997 年日本学術振興会特別研究員，1999 年早稲田大学大学院博士後期課程単位修得退学，同年早稲田大学助手．2000 年情緒交流ロボットの開発研究により博士（工学）．2001 年理化学研究所脳科学総合研究センター研究員，2003 年京都大学大学院情報学研究科講師，2005 年同大学助教授を経て，2007 年より同大学准教授．この間，2001〜2003 年早稲田大学ヒューマノイド研究所客員講師，2005〜2007 年同研究所客員助教授，2007 年より同研究所准教授．2005 年より理化学研究所脳科学総合研究センター客員研究員．研究分野は神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学．2000 年日本機械学会論文賞，2005 年国際会議 IEA-AIE 最優秀論文賞等受賞．著書は，身体性とコンピュータ（共立出版），メカノクリーチャ（コロナ社）等．

奥乃 博（正会員）

1972 年東京大学教養学部基礎科学科卒業．日本電信電話公社，NTT，JST，東京理科大学を経て，2001 年より京都大学大学院情報学研究科知能情報学専攻教授．博士（工学）．この間，スタンフォード大学客員研究員，東京大学工学部客員助教授．人工知能，音環境理解，ロボット聴覚，音楽情報処理の研究に従事．1990 年度人工知能学会論文賞，IEA/AIE-2001，2005 最優秀論文賞，IEEE/RSJ IROS-2001，IROS-2005 Best Paper Nomination Finalist，第 2 回船井情報科学振興賞等受賞．JSAI，RSJ，ACM，IEEE，AAAI 等各会員．本学会英文図書出版委員．