

非負値行列因子分解を用いた音楽信号分離

北村 大地 (総研大), 小野 順貴 (NII/総研大), 澤田 宏 (NTT),
亀岡 弘和 (東大/NTT), 猿渡 洋 (東大)

どんな研究？

市販のCDやバンド演奏の録音データ等のように、いろんな楽器や歌声が混ざった音から、個々の音源を高品質で分離抽出する音源分離アルゴリズムを研究しています。

何のために？

複数の音源（ギターやボーカル等）がハーモニーとして混ざった音楽信号を個々の音源毎に分離できれば、楽器の練習補助や音楽教育に活用できる他、ユーザによる既存音楽のリミックスなどが容易となり、音楽の楽しみ方がこれまで以上に広がります。

状況設定

演奏の一発録音データ



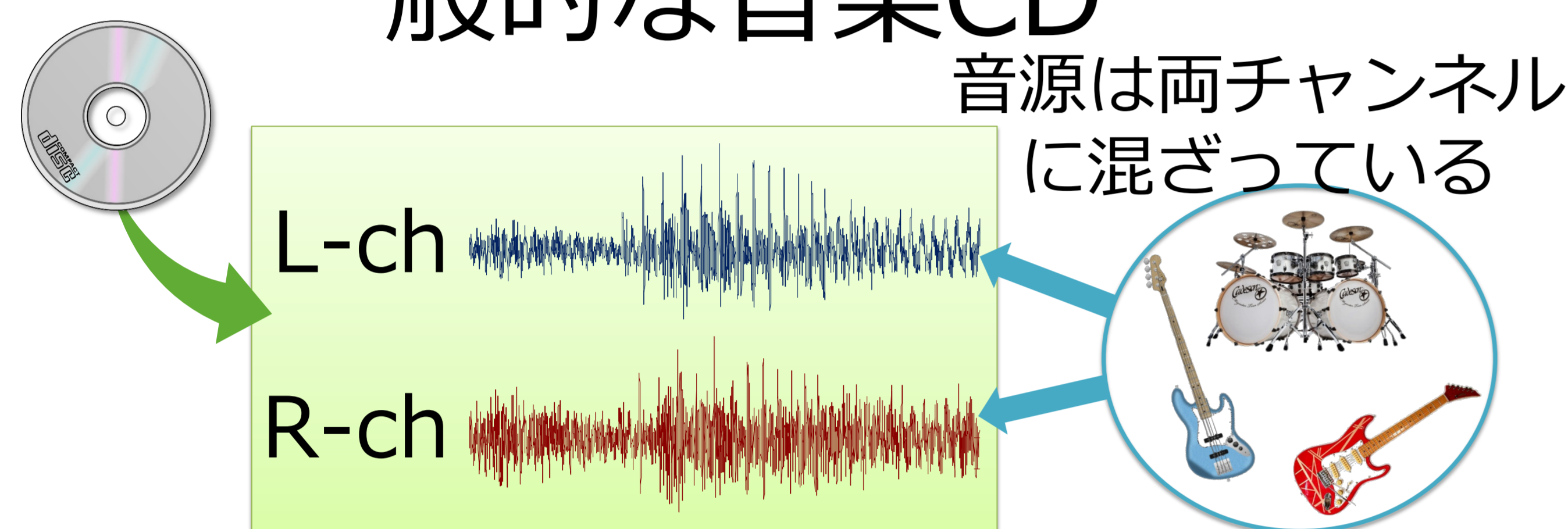
マイクアレイ等で録音すれば、たくさんのチャンネルで録音できる！音源の数以上のチャンネル数も可能

音源数とチャンネル数の関係はとっても重要

チャンネル数が多い程音源分離は簡単になりますが、音源数がチャンネル数よりも多いととても難しい問題になってしまいます。

一般的な音楽CD

音源は両チャンネルに混ざっている



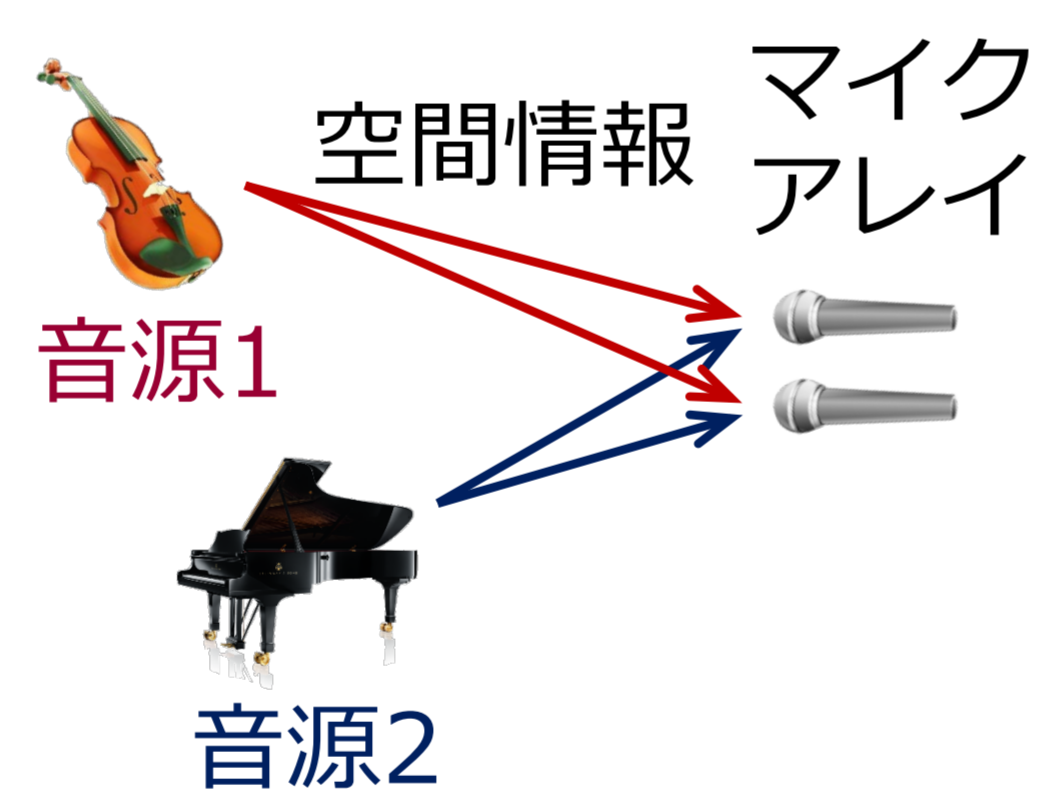
ステレオ（LとRの2チャンネル）信号混ざっている音源の数は普通3つ以上（ギター、ベース、ボーカル、ドラム…）

研究内容

● どうやって混ざった音源を分離するの？

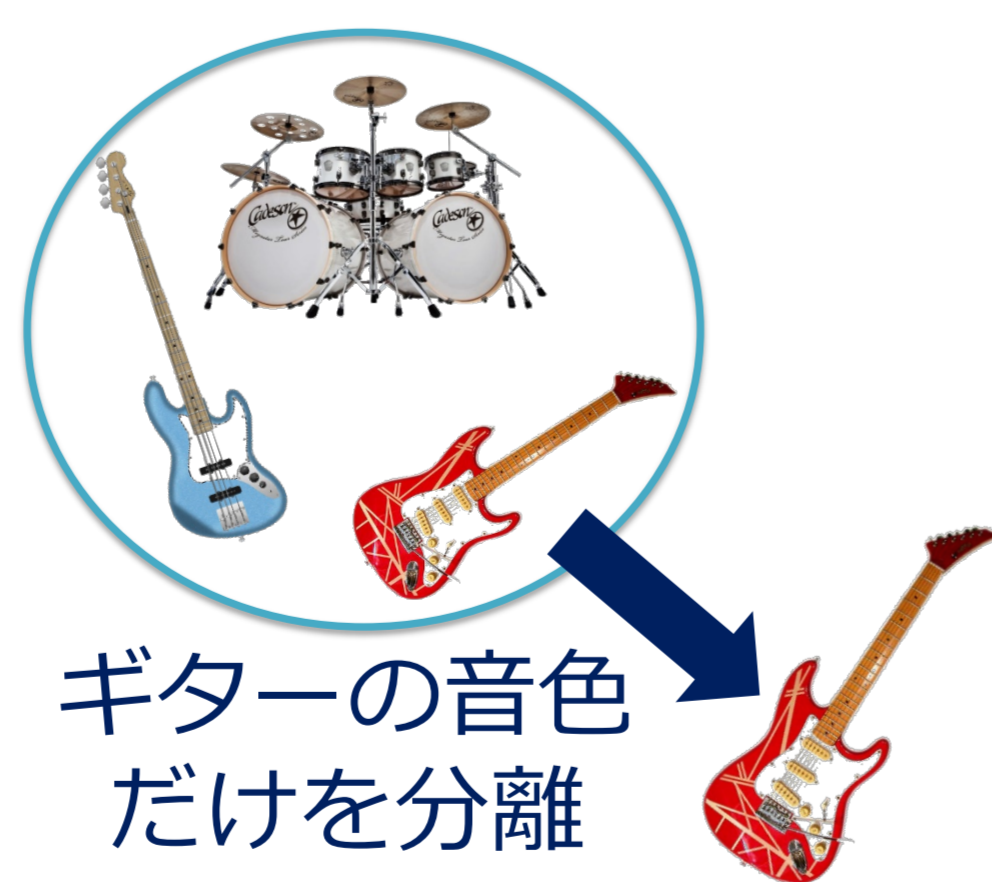
・音源の空間情報を用いた音源分離

音源分離は昔から独立成分分析（ICA）という手法が有名でした。ICAはチャンネルの数が十分多い場合に適用でき、**音源の空間的な性質**と**音源自身の統計的な性質**を用いて音源を分離する手法です。つまり、マイクアレイに対して、どの方向に音源があるのかという空間情報を手掛かりにしています。空間情報には、音量比と位相差（伝達時間差）が含まれていますが、CDなどのように人工的に空間情報を作り出している信号では、うまく分離できないことがあります。また、一般的なCDはステレオなので2チャンネルしかありません。音源は3つ以上ある場合が多いので、ICAでは分離できない問題になります。



・音源の音色を用いた音源分離

ICAに代わる新しい音源分離として、非負値行列因子分解（NMF）を用いた手法が近年提案され、非常に有名になりました。音響信号にNMFを適用すると、**その信号の中に含まれているいろいろな音色**を取り出すことができます。つまり、**音源毎の音色の違い**を用いて音源を分離することができます。この手法は、チャンネルの数が1つでも適用できるため、ICAではできなかった条件での音源分離がNMFで可能となりました。



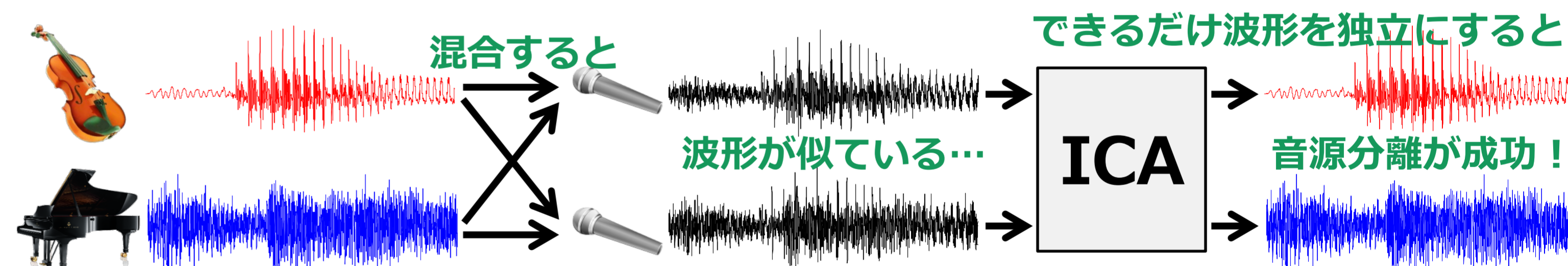
・より高性能な音源分離を目指して

我々のグループでは、ICAとNMFを組み合わせた新しい音源分離手法を提案しています。これは、音源の空間的な性質をICAで学習しつつ、各音源の音色をNMFで捉える手法で、従来のどの手法よりも高い音源分離性能を高速な演算で実現しています。また、チャンネル数が音源数よりも少ない状況においては、分離したい特定の音源の音色をあらかじめNMFに入力しておく「教師あり音源分離手法」も提案しています。

●各音源分離手法の詳しい解説

・独立成分分析 (independent component analysis: ICA)

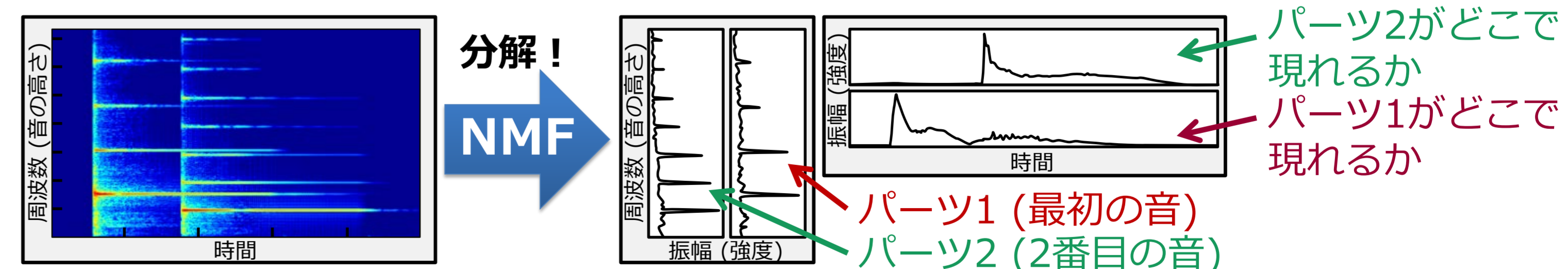
チャンネル数が音源数と同じか、チャンネル数の方が多い場合、ICAと呼ばれる統計的信号処理が適用できます。今、二つの音源が混ざって二つのチャンネルで与えられたと仮定しましょう。



音源が混ざった二つの信号は強い相関関係があります。ICAでは、これらの信号を統計的に独立にするような逆空間情報を求めることができるため、出力される信号は混ざっていた信号が個々に分離されたものになります。これは、ある混合系（行列）の逆行列を求める問題とみることができますが、元の混合系は分からないので、音源間の独立性という別の基準を活用して求めています。

・非負値行列因子分解 (nonnegative matrix factorization: NMF)

チャンネル数より多い数の音源がある場合には、NMFと呼ばれるアルゴリズムを応用します。これは、マイナスの値がない2次元データを、そのデータの中の特徴的なパーツに分解する手法です。NMFによる音源分離では、周波数の時間的な変化を表すスペクトログラムを分解します。



ここで、分解されて得られたパーツとは音のスペクトルであり、例えばピアノのドの音やレの音などを表します。また、パーツがどこで現れるかという情報は、各音のタイミングを表すので、楽譜のようなものです。それぞれのパーツがどの音源の一部なのか分かれば、パーツを取捨選択することで音源分離された音を再構成することができます。