

研究紹介

中村 友彦

産業技術総合研究所 人工知能研究センター

自己紹介

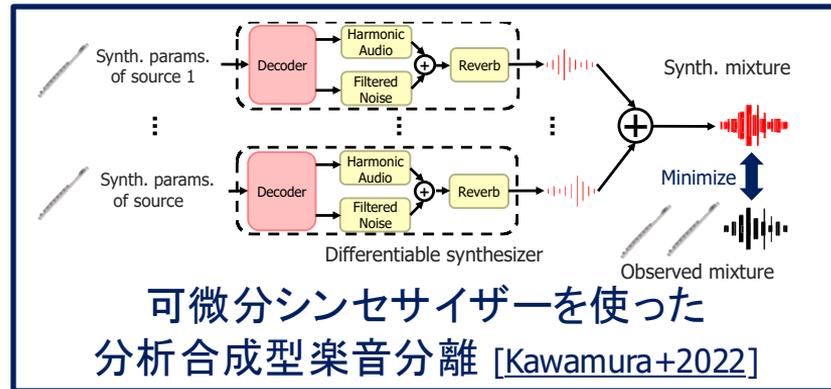
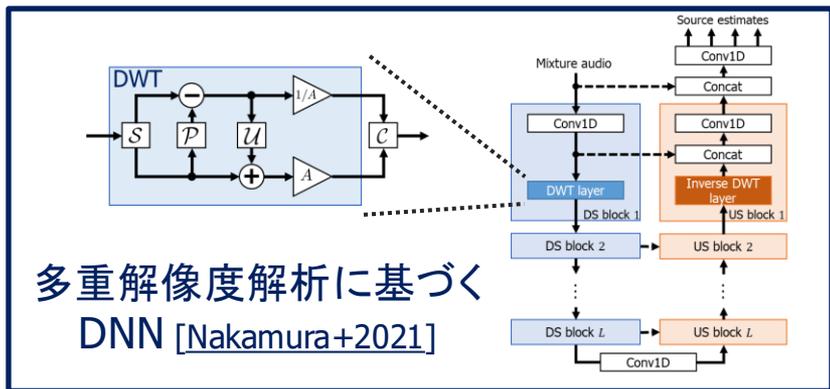
- 名前: 中村 友彦 (Tomohiko Nakamura)
- 経歴
 - 2016年: 東京大学 情報理工学系研究科 博士 (情報理工学)
 - 2016~2019年: セコム(株) IS研究所 研究員
 - 2019~2023年: 東京大学 情報理工学系研究科 特任助教
 - 2023年~現在: 産総研 人工知能研究センター 主任研究員
- 専門: 音響信号処理, 音楽信号処理, (地震波形処理)



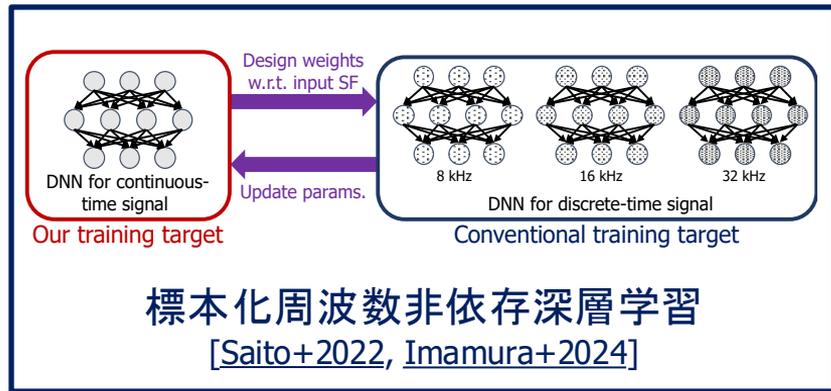
@TomohikoNakamu2

研究テーマ俯瞰

- 信号処理と深層学習の“良いとこどり”を目指した方法論を構築



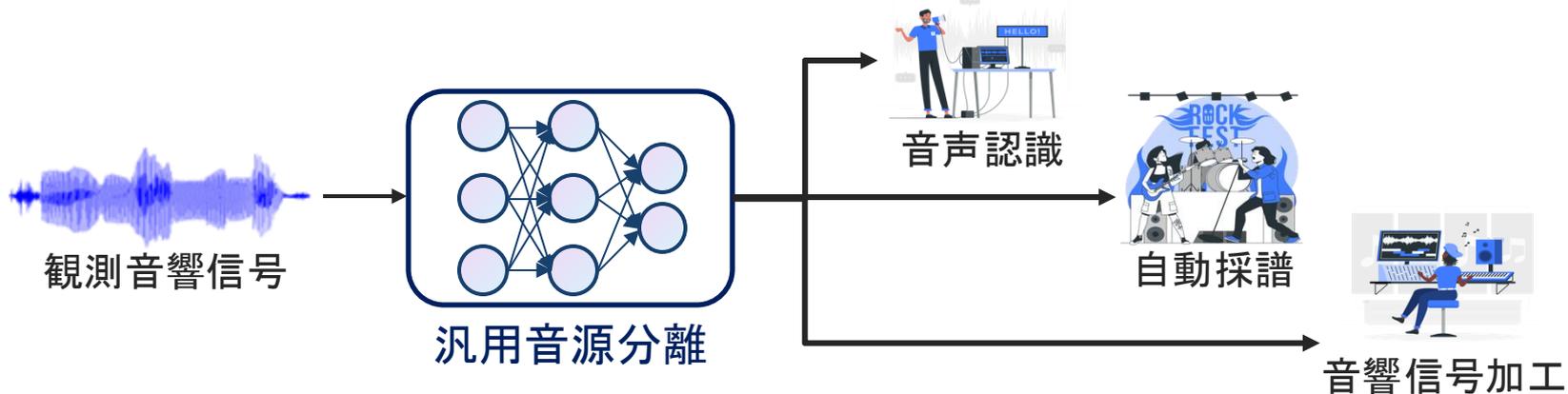
jaCappellaコーパス: 日本語重唱コーパス
[Nakamura+2023]



研究紹介1: 標本化周波数非依存深層学習

背景：汎用音源分離の実現に向けて

- 音源分離：観測音響信号を各音源信号に分解する技術
 - 深層ニューラルネットワーク(DNN)の導入により、分離性能が飛躍的に向上 [Dubey+2022, Uhlich+2023]
 - 様々な音響信号処理タスクの**前処理として有用** ⇒ 性能と**汎用性**の両立が重要



- 様々なタスクに対し汎用的な音源分離(汎用音源分離)の実現には、「**何に関する汎用性をいかに備えさせるか？**」が肝要

汎用音源分離と標本化周波数

- 汎用音源分離の実現には、録音条件に関する汎用性が重要
- 代表的な録音条件：標本化周波数(SF)
 - タスクによって典型的なSFは異なる.

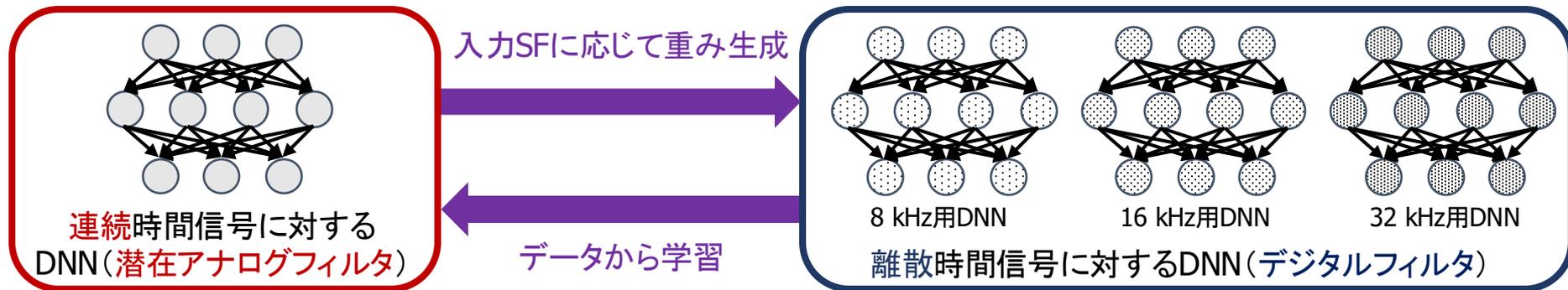
タスク	典型的なSF
リミキシング, 楽音加工	44.1, 48 kHz
音声認識, 話者ダイアリゼーション	8, 16 kHz

–利用可能なSFは、録音デバイスやシステムにも依存

汎用音源分離は様々なSFを扱える必要あり。
しかし、従来のDNNが未学習のSFを直接扱うのは困難 🙄

標本化周波数非依存深層学習 [Saito+2022]

- 連続時間/周波数領域(SF非依存)で定義された関数から, 入力SFに応じて重みを生成する過程を導入 ⇒ **未学習のSFも対応可能!!**
 - 重み生成はアナログフィルタからのデジタルフィルタ設計と解釈可能
⇒ **デジタルフィルタ設計手法を援用可能**



提案アプローチでの学習対象

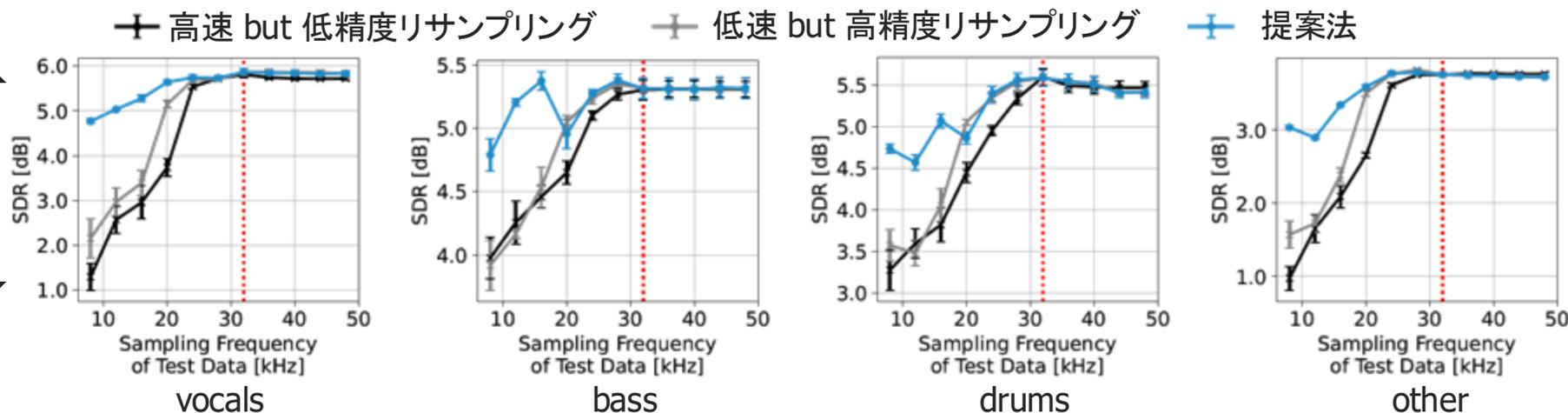
従来の学習対象

- 様々な拡張も可能: 非整数ストライドへの対応 [Imamura+2023], ニューラルアナログフィルタ [Imamura+2024, in press], etc.

提案法 vs リサンプリング

- 4音源の楽音分離タスクで、提案法をリサンプリングと比較
 - MUSDB18 [Rafii+2018] の学習・テストデータを利用
 - 学習時のSFは32 kHzのみ, テスト時は8~48 kHz(4 kHz毎)のSFを使用
- 提案法は、未学習のSFでリサンプリングよりも高い分離性能を達成
 - 未学習のSFへの適応方法のみ異なる(i.e., 学習済みモデルは同一)

↑
音
分
離
性
能
↓



研究紹介2: 可微分信号処理を利用した 分析合成型音源分離

分析合成型音源分離モデル

- 近年の音声・楽音合成手法の品質は飛躍的に向上
⇒ 合成技術を分離に援用できるかも？
- 分離問題を「混合音からの合成パラメータ推定問題」と解釈
 - 推定した合成パラメータから目的音を合成
⇒ 他の音源の成分が分離音に入りにくい(入らない)
- 様々な深層生成モデルを用いた方法が提案
 - e.g., 敵対的生成ネットワーク (GAN) [Narayanaswamy+2020],
正規化フロー [Zhu+2022], 可微分信号処理 (DDSP) 混合モデル [[Kawamura+2022](#)],
拡散モデル [Serra+2022, Mirania+2024]

信号処理を援用した楽器音合成用DNN

- 可微分信号処理 (DDSP) [Engel+2020]
 - シンセサイザーなどで使われる信号処理技術をDNNのモジュールとして利用
- DDSP自己符号化器: 入力音響信号をシンセサイザを用いて再構成



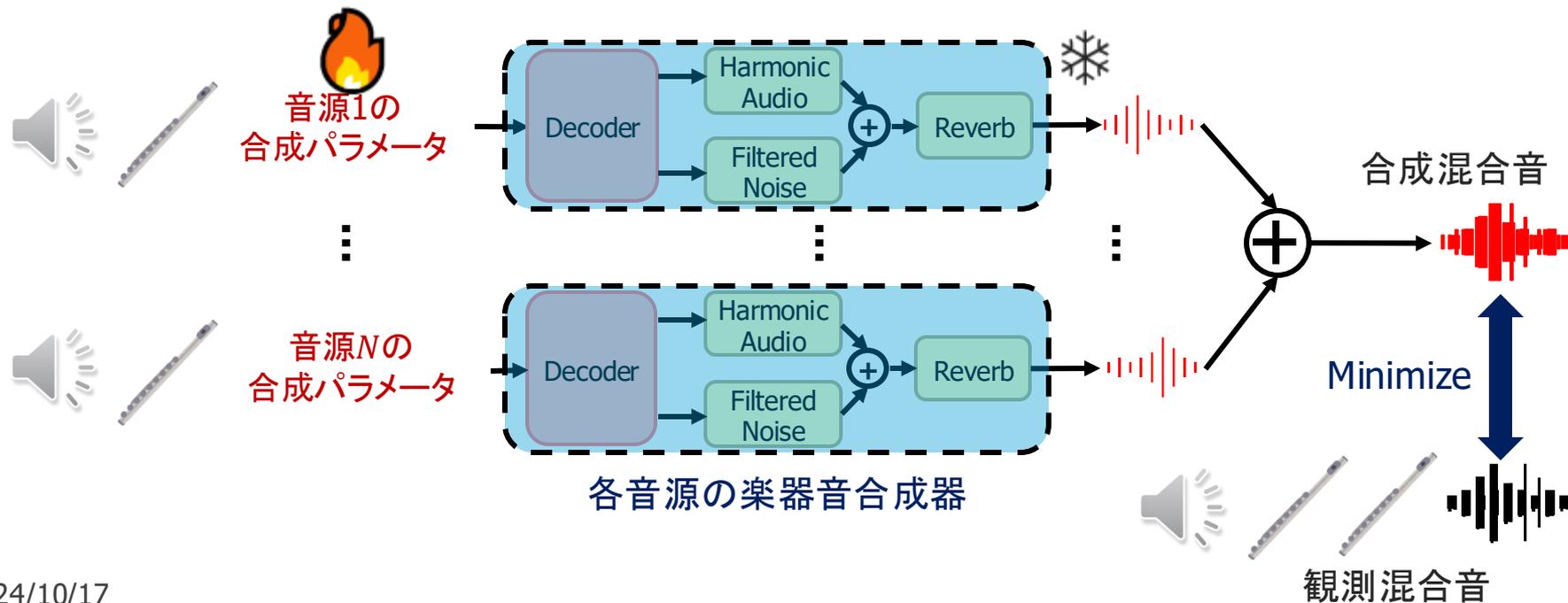
楽器変換
(歌声 to Violin)



DDSP自己符号化器は単旋律の楽器音しか扱えない 😞

DDSP混合モデルを用いた音源分離 [Kawamura+2022]

- 混合音から各音源の基本周波数, 音色パラメータ, ラウドネスを推定
 - 事前学習したDDSP自己符号化器の一部を楽器音合成器として利用
- 楽譜を援用することで, 同一楽器同士での分離でも高性能に動作!!



研究紹介3: 重唱信号・情報処理に向けて

jaCappellaコーパス [Nakamura+2023]

- 日本語重唱コーパス (**Japanese a Cappella Corpus**) を構築
 - YouTube, TikTok で出現してきた新たなアカペラストایلを模擬

異なる10個のジャンルの重唱曲 (計50曲)

- 著作権保護期間の過ぎた童謡・唱歌を編曲
- 声部: 6声 (リードボーカル, ボーカルパーカッション含む)
- ジャンル: ジャズ, パンクロック, ボサノバ, EDM etc.

譜面や各声部の音源信号も収録

- 譜面: MusicXML形式
- 音源: 48 kHz, 24 bitリニアPCM

研究利用(無償)・商用利用(有償)も可能

- 著作権や隣接権もコーパス作成チームで処理済み



Corpus



Demo



Code

jaCappellaコーパス:トレーラー

◆
JACAPPELLA
CORPUS
◆
Version 2

応用例：各声部への分離（重唱分離）

入力信号（どんぐりころころ）：

Voice part	Groundtruth	X-UMX [Sawata+2020]	DPTNet [Sarkar+2021]	MRDLA [Nakamura+2021]
Lead vocal				
Soprano				
Alto				
Tenor				
Bass				
Vocal percussion				

● 他の応用例

－ 重唱歌声合成 [Hyodo+2024], 歌唱者ダイアライゼーション [Suda+2022], etc.

まとめ

- 研究紹介1: 標本化周波数非依存深層学習
 - 汎用音源分離を実現するには, 様々な標本化周波数に扱える必要あり.
 - 層の重みとデジタルフィルタのアナロジーから, 標本化周波数に非依存な領域のパラメータを学習する方法論を構築
- 研究紹介2: 可微分信号処理を利用した分析合成型音源分離
 - 分析合成型音源分離: 合成手法を援用し, 混合音から各音源信号を合成
 - 事前学習したDDSP自己符号化器の一部を楽器音合成器として利用した, DDSP混合モデルを構築
- 研究紹介3: 重唱信号・情報処理に向けて
 - 日本語重唱コーパス(jaCappellaコーパス)を構築
 - 応用例: 重唱分離, 重唱歌声合成, 歌唱者ダイアライゼーション