

画像処理AI技術を用いた 混合音からの特定音除去

神奈川工科大学 情報学部 情報工学科

教授 宮崎 剛

教授 田中 博（共同発明者）

2024年10月15日

研究分野の概要

- 本研究では、混合音から特定の音を除去(抽出)する手法や技術を開発
 - 混合音・・・複数の音が同時に発生している音
- 時期(時間)や状況(状態)、場所、個体に依存する音を対象とする
- 対象とする音を限定する
 - 目的に応じた音除去の実現
 - 高い除去(抽出)精度を目指す

社会的位置づけ

- 動画や音声の録画・録音機会、配信機会の増加
 - － ビデオカメラやWebカメラ、スマートフォン
 - － インターネットやSNS等
 - 新型コロナウイルス発生以降の働き方や授業形態の変化
 - － 在宅ワーク、オンライン会議・授業
 - － 録画コンテンツ作成の増加
 - 動画や音声に周囲の音が混入
 - － 不明瞭、不適切情報、プライバシー、集中力低下
- ▷ 雑音除去に対するニーズは高い

従来技術とその問題点

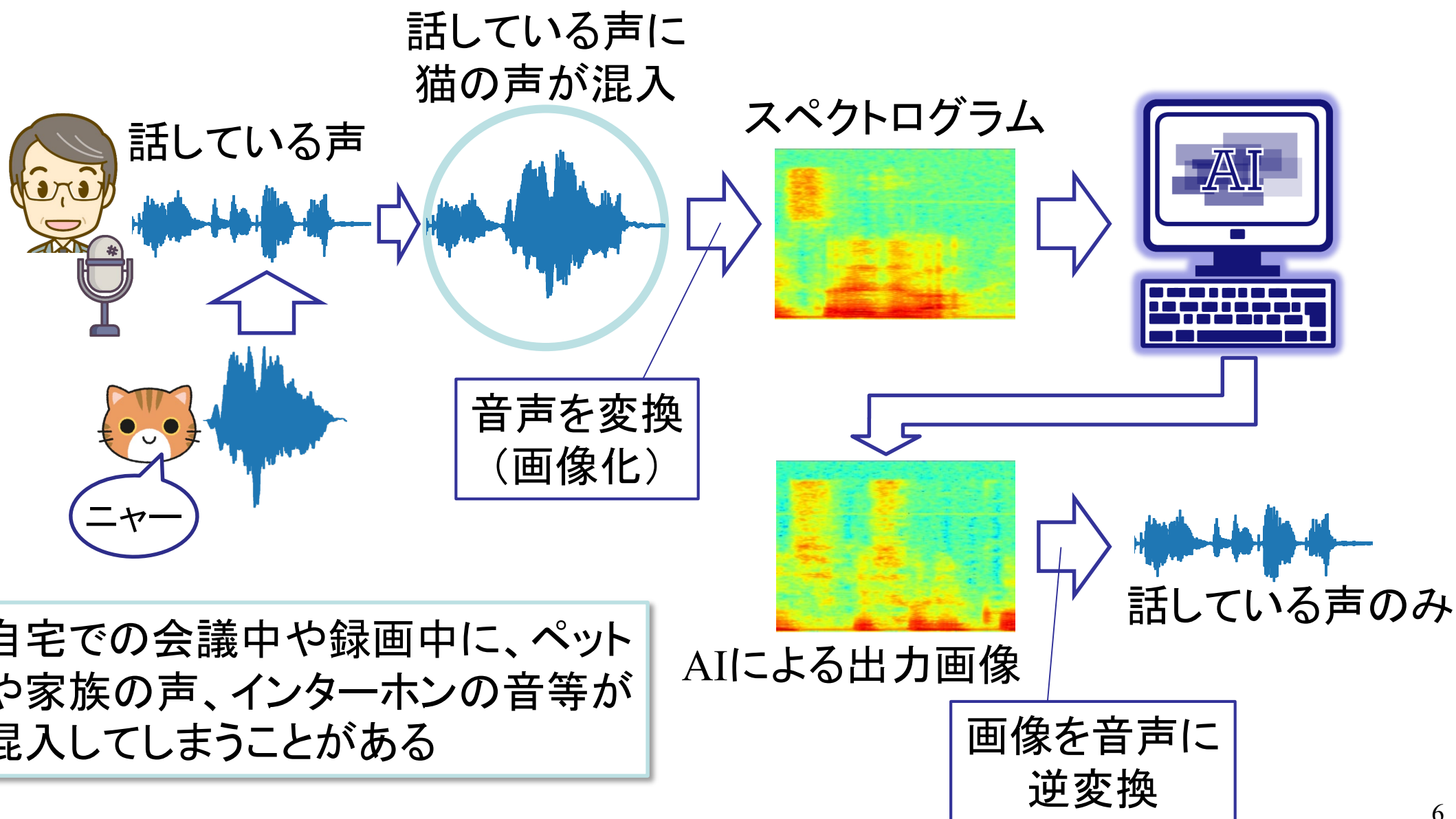
実用化されているノイズ除去

- ヘッドホン、イヤホンのノイズキャンセル
 - リアルタイム性、連続した低周波数帯で効果
 - 時間的に短い、不規則に変化、高周波数帯を含む音には向いていない
 - 専用機器やソフトを用いたノイズ除去
 - GPU搭載機器を使用したリアルタイムノイズ除去
 - 一般的な雑音に対応
- ▷ 環境等に依存する個別雑音では除去が難しい

新技術の特徴・従来技術との比較

- 単一マイクを使用したノイズ除去
 - 複数マイクを使用した構成に比べてシンプル
- 音声信号を画像に変換
 - 画像を用いた深層学習モデル
 - 画像に現れる雑音の特徴を学習して除去
 - SN比やピッチ変更によるデータ拡張
- 選択的雑音除去モデルの構築
 - 除去したい音声に応じた学習モデルの構築

AIを使った特定音除去



AIの学習プロセス

①音の収録（残したい音、除去したい音）

残したい音→**対象音**、除去したい音→**背景音**



②データ拡張（SN比、ピッチ）



③学習用データの作成

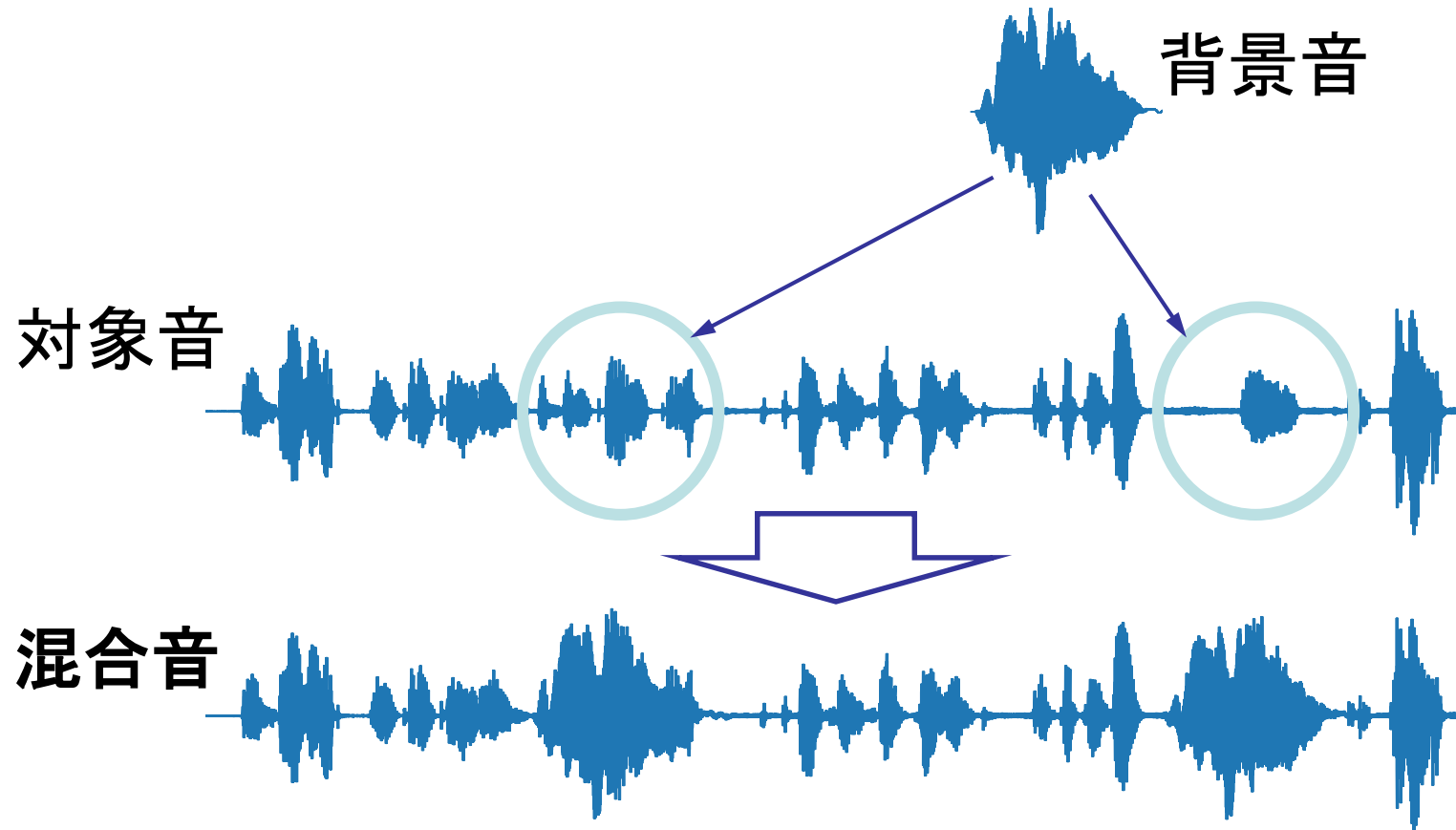


④AIの学習

②データ拡張(混合音)

1. 混合音作成

対象音の複数箇所背景音を混入



データ拡張 (SN比)

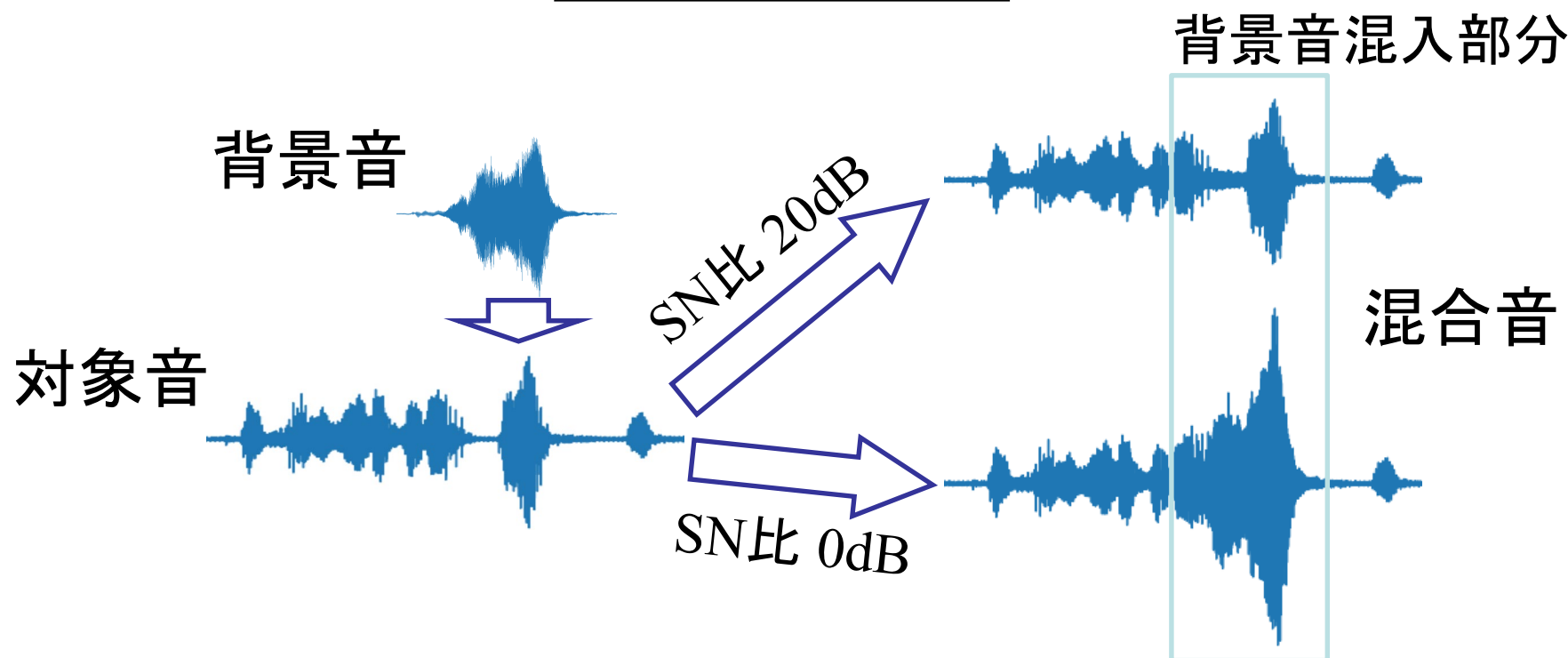
2. SN比

信号(**S**ignal)に対する
雑音(**N**oise)の比

SN比と雑音の関係

SN比	小	大
雑音	大	小

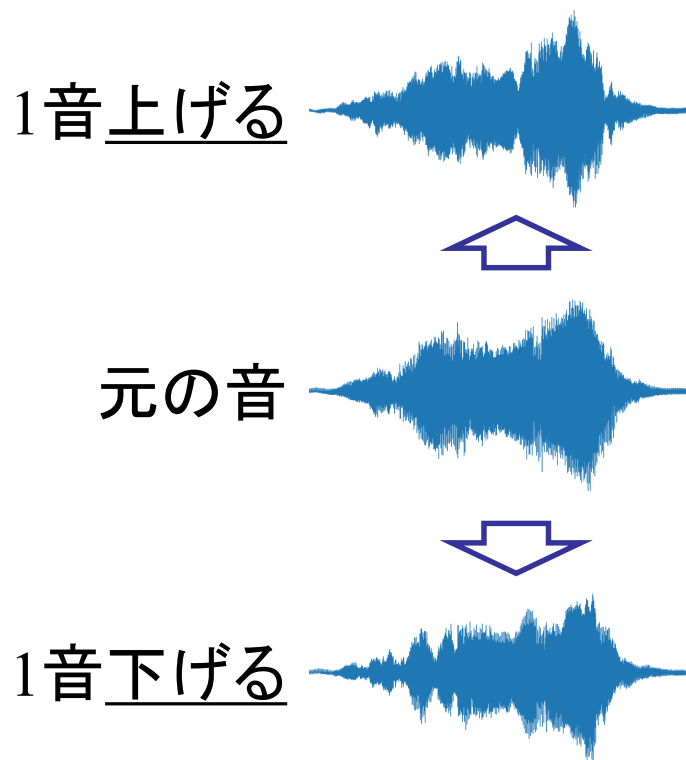
⇒ SN比を変えて複数の混合音を作成



データ拡張(ピッチ)

3. ピッチ変更

同一人物や同一個体でも、状況で声は変化する
音の高さを変更した複数の対象音、背景音を作成



ピッチ変更した対象音と
背景音を異なるSN比で混合

多くのバリエーションの
混合音を生成できる

↑AIの学習に利用

AIの学習プロセス

①音の収録（残したい音、除去したい音）

残したい音→**対象音**、除去したい音→**背景音**



②データ拡張（SN比、ピッチ）



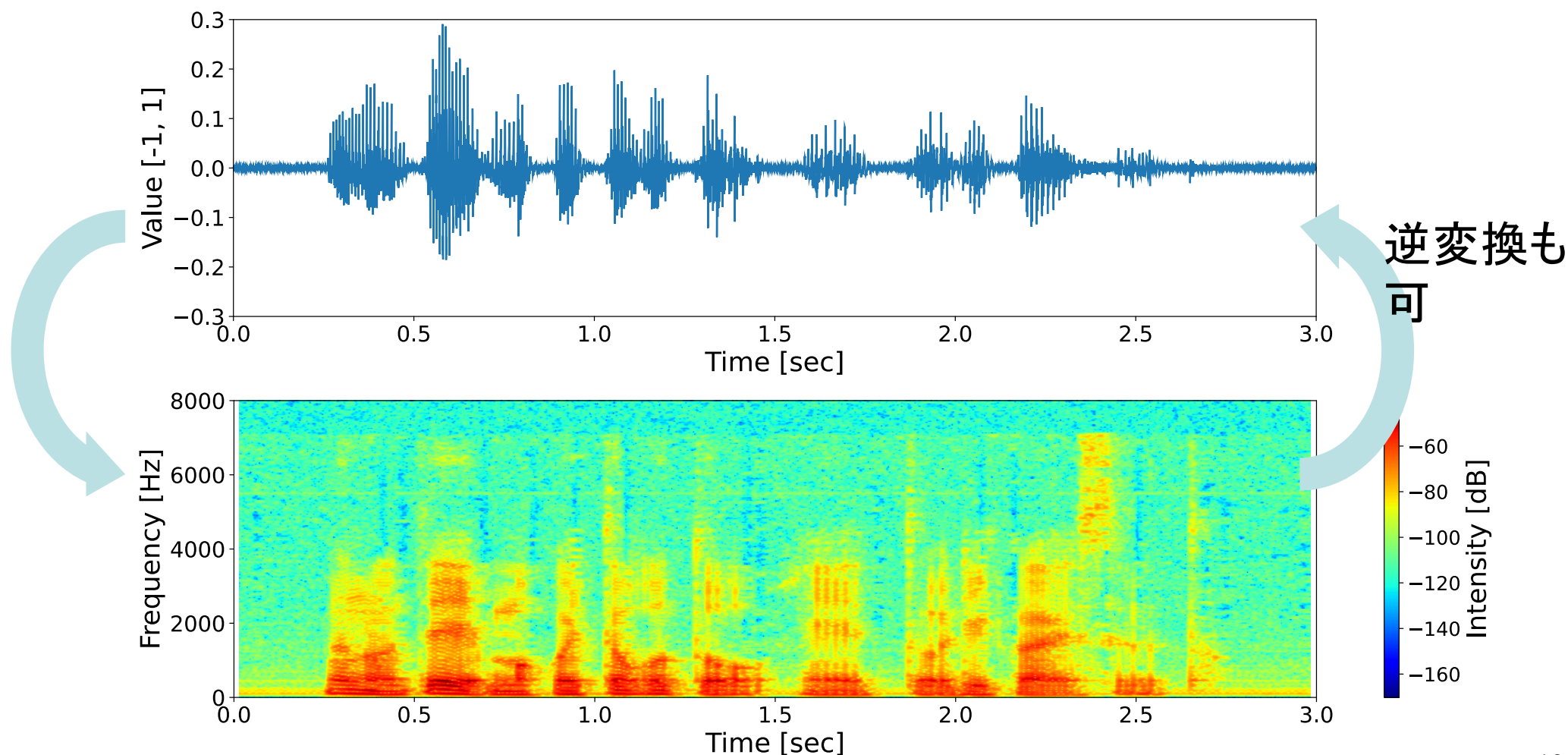
③学習用データの作成



④AIの学習

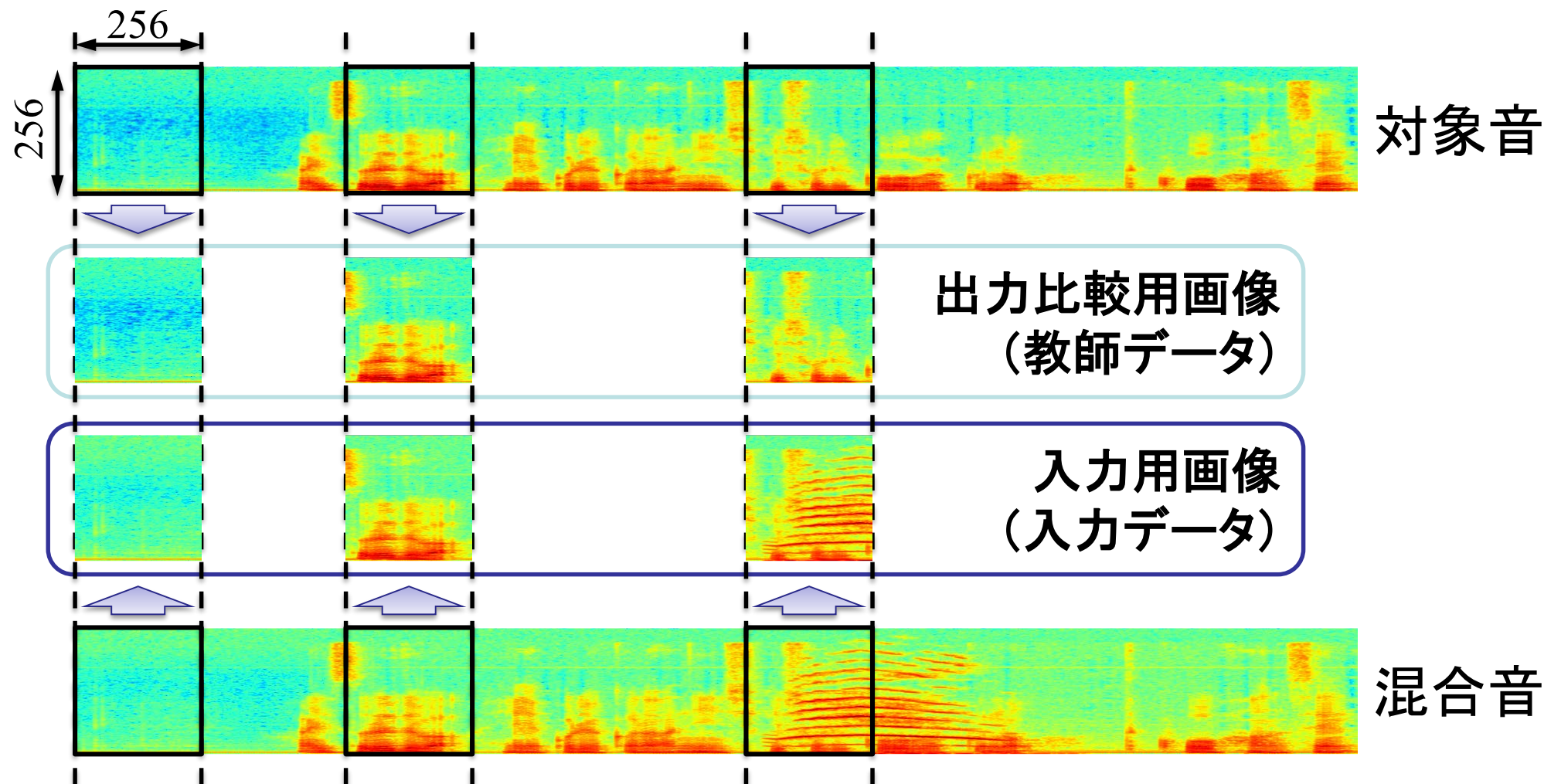
③学習用データの作成(スペクトログラム)

1. 音声信号をスペクトログラムに変換



学習用データの作成(データセット)

2. データセットの作成



AIの学習プロセス

①音の収録（残したい音、除去したい音）

残したい音→**対象音**、除去したい音→**背景音**



②データ拡張（SN比、ピッチ）



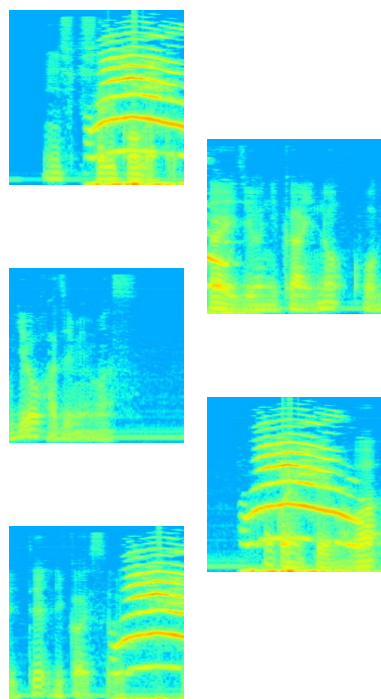
③学習用データの作成



④AIの学習

④AIの学習

入力画像
(混合音)

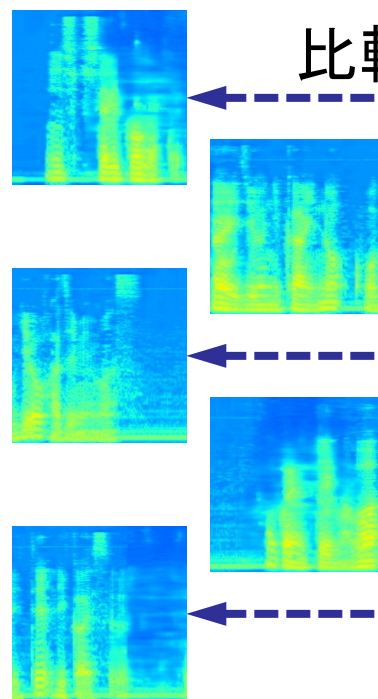


学習



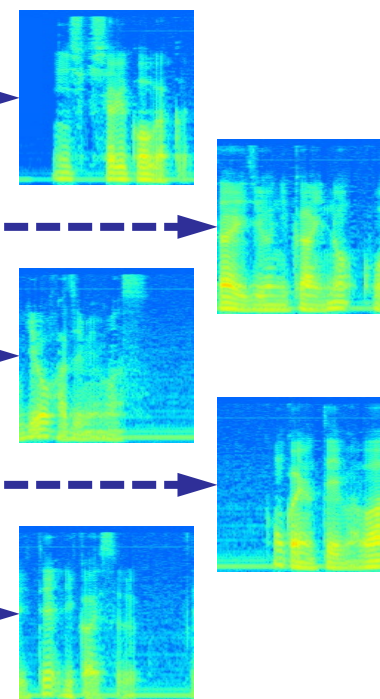
U-Net

出力画像

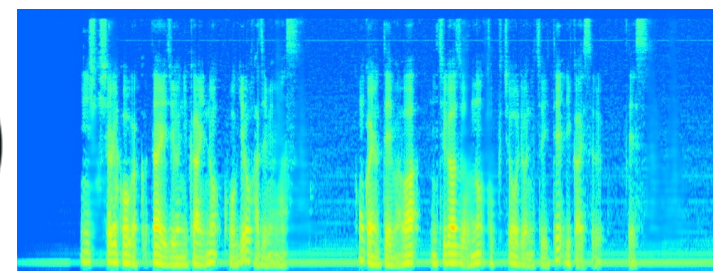
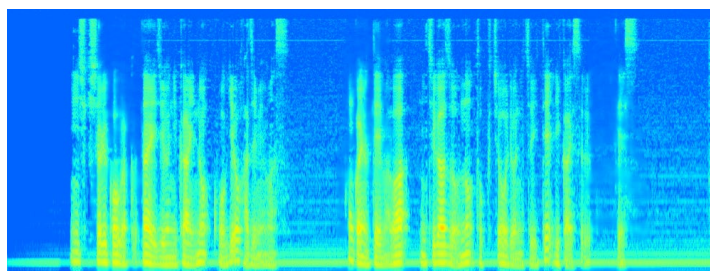
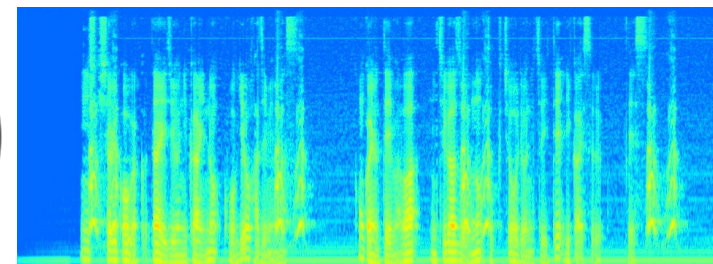
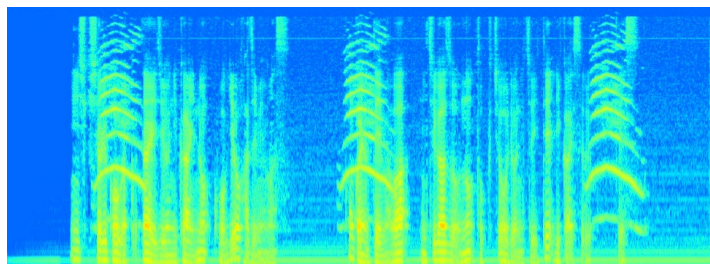
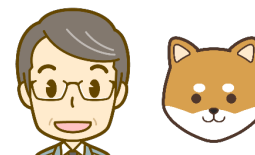
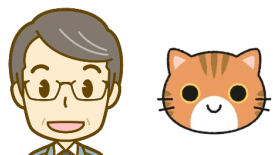


比較

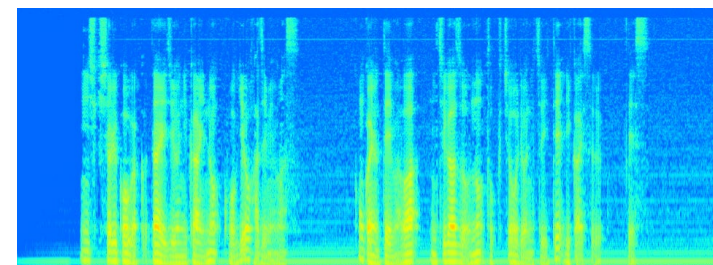
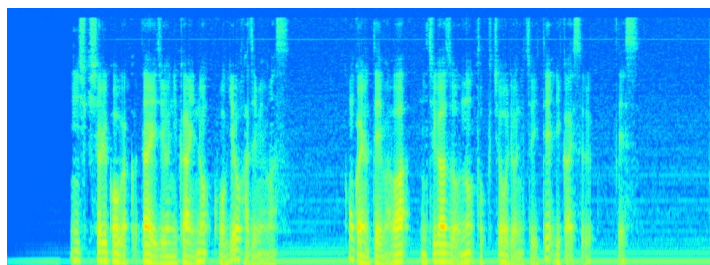
教師画像
(対象音)



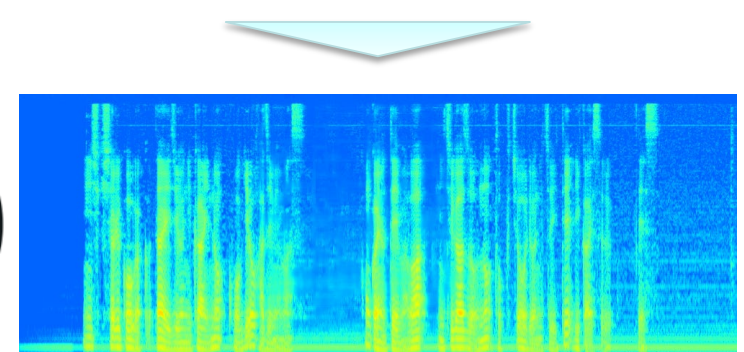
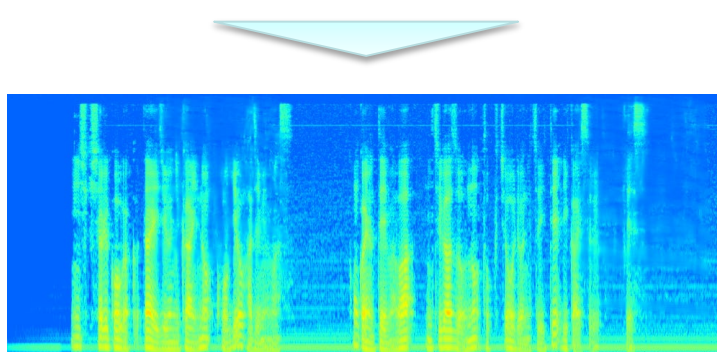
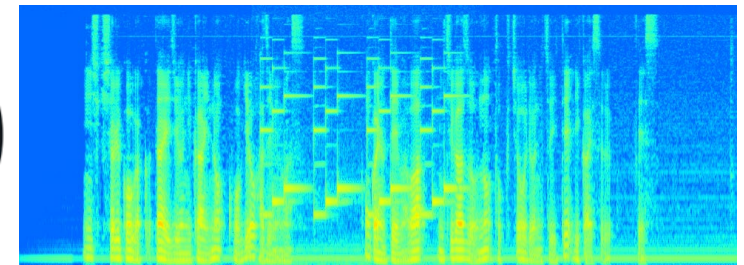
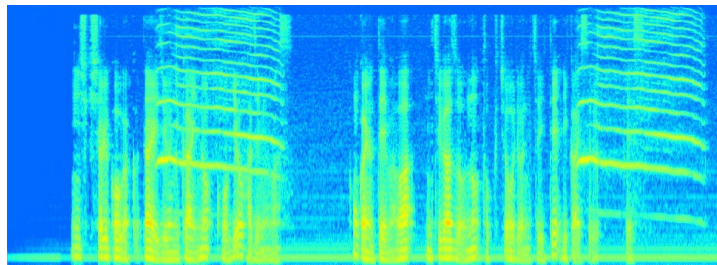
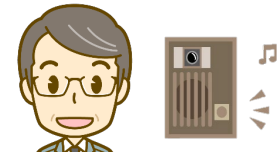
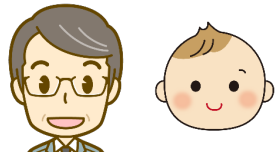
背景音除去の例(猫／犬)



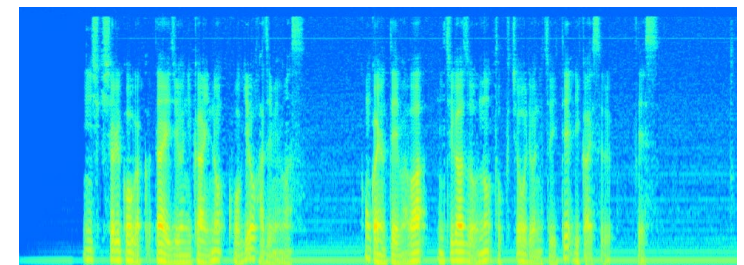
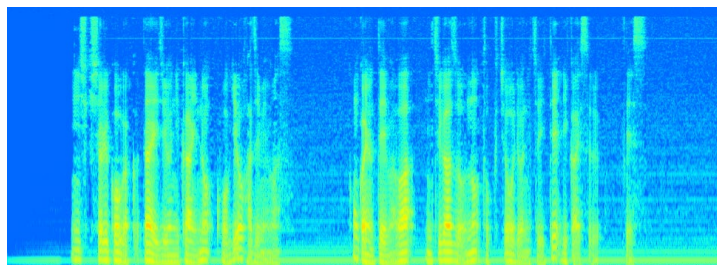
教師
画像



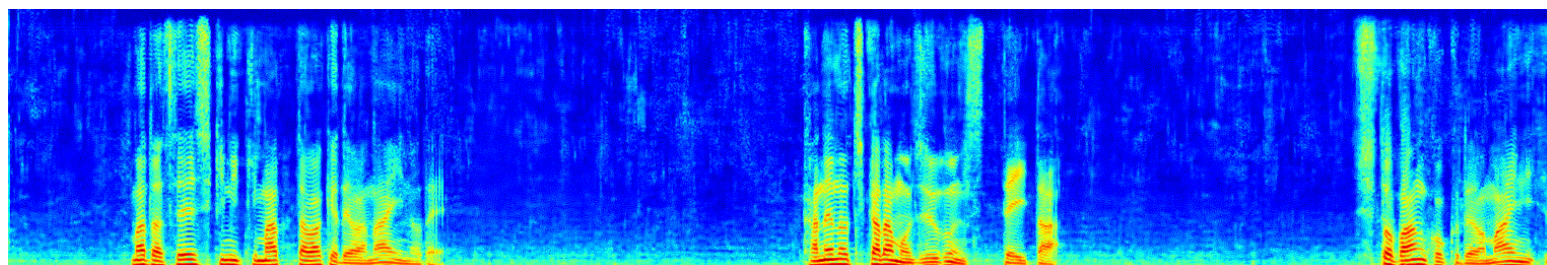
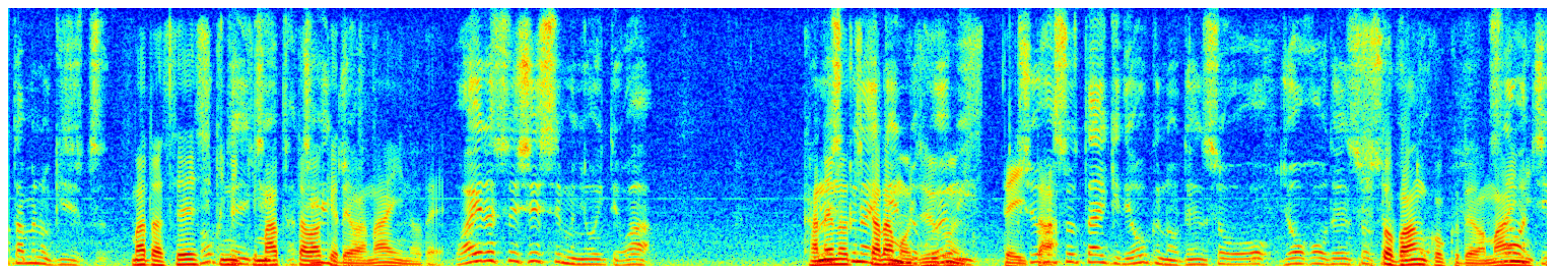
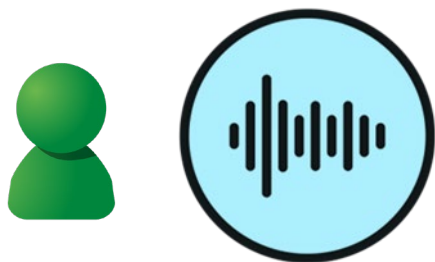
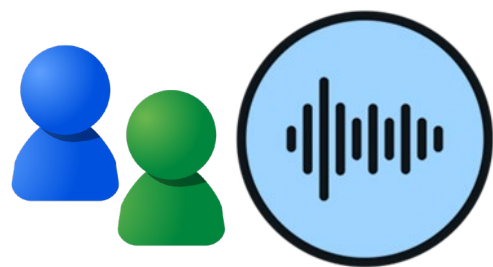
背景音除去の例(赤ちゃん／インターホン)



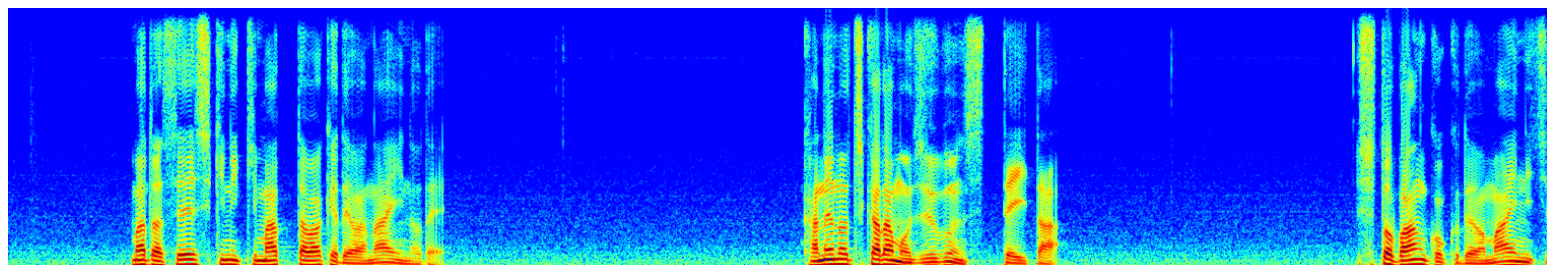
教師
画像



背景音除去の例(人物二人→一人)



教師
画像



想定される用途

- 動画からの雑音・特定音除去
 - YouTube、アーカイブ映像、講義動画
- 話者分離
 - 会議録音データから特定人物の発言抽出
- 騒音環境での音声認識
 - 音声認識時のプレフィルタ
- 混入した複数の時系列信号から特定の信号を除去・抽出の可能性

実用化に向けた課題

- パラメータチューニングによる背景音除去精度の向上
- 残存対象音の音質向上
- 処理の高速化
- 適用シーンの明確化
- 実フィールド(様々な機器や環境)での検証
- サービス提供の仕組み・システム

企業への期待

- 共同研究
 - ノイズ除去技術を保有している企業
 - ノイズ除去をサービスや事業として考えている企業
 - 電氣的信号や振動等から情報の除去・抽出を必要とする企業
- 実証実験
 - 企業で保有するデータを使った評価実験

企業への貢献、PRポイント

- 評価用のデータをいただければ、直ちに学習モデル構築から性能評価の実施が可能
- 技術的なノウハウの提供が可能
- 対象や環境に合わせたカスタマイズ
 - 現在数社からの問い合わせあり

本技術に関する知的財産権

- 発明の名称 : 学習モデル作成方法、学習モデル作成装置、背景音除去方法、および背景音除去装置
- 出願番号 : 特願2023-081092
- 特許番号 : 特許7414214
- 出願人 : 学校法人幾徳学園
- 発明者 : 田中 博、宮崎 剛

お問い合わせ先

神奈川工科大学
研究推進機構 研究支援部門

T E L 0 4 6 – 2 9 1 – 3 2 7 7
e-mail ken-shien@ccml.kanagawa-it.ac.jp