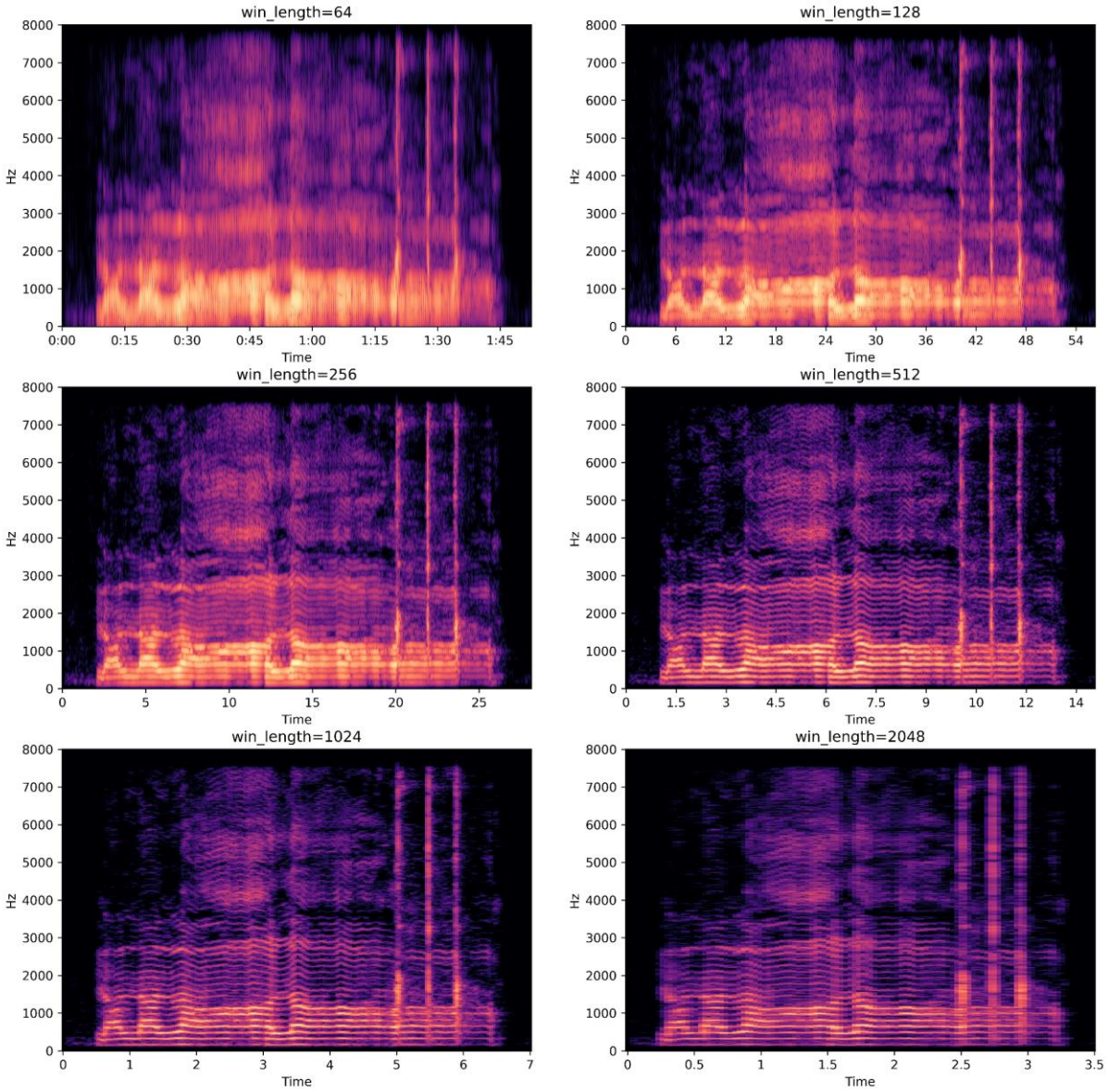
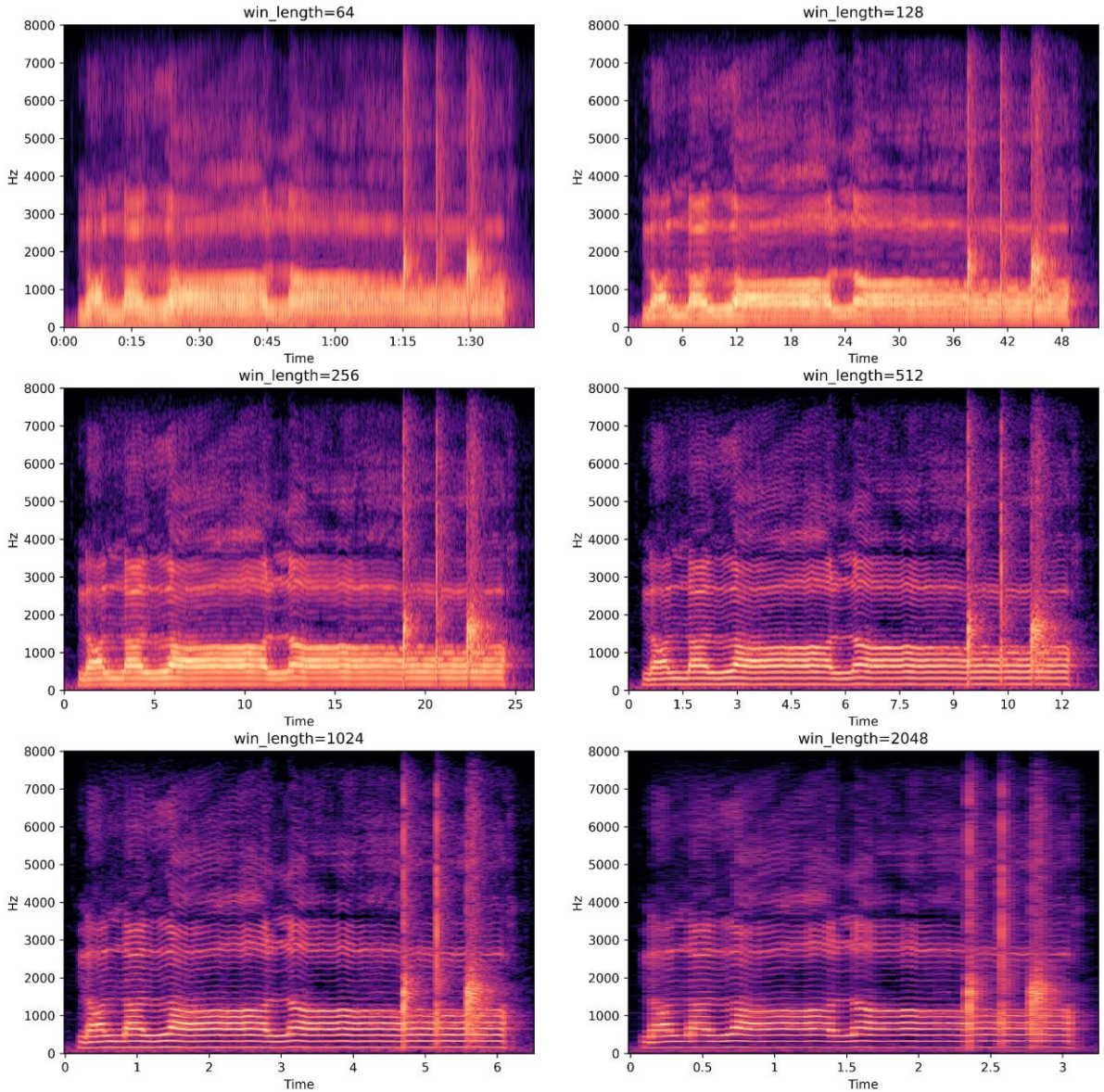
音声再録@PC

- 16kHz
- 32bit



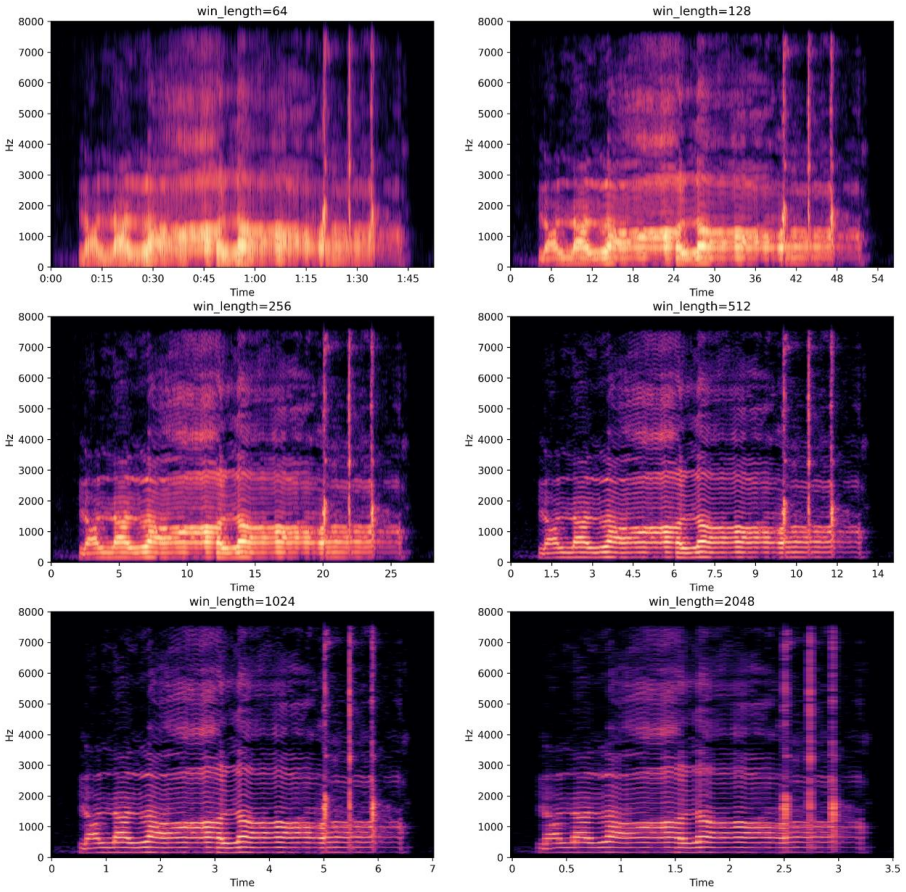
音声録音@スマホ

- 16kHz
- 32bit

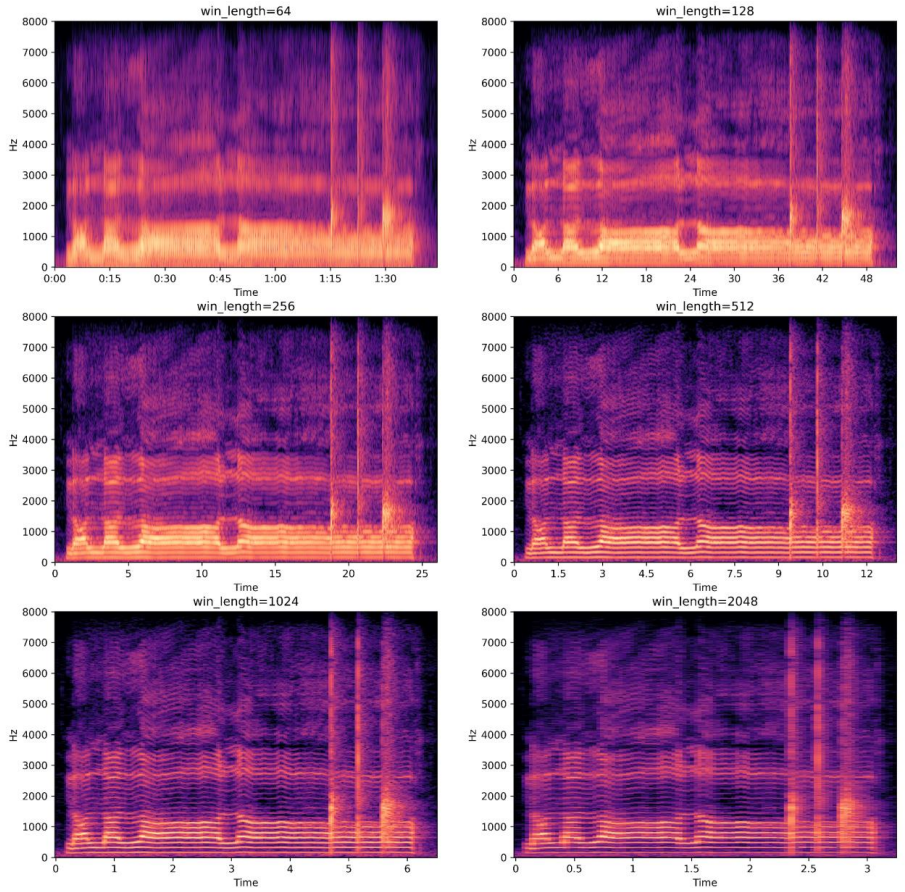


PC録音とスマホ録音の比較

PC



スマホ



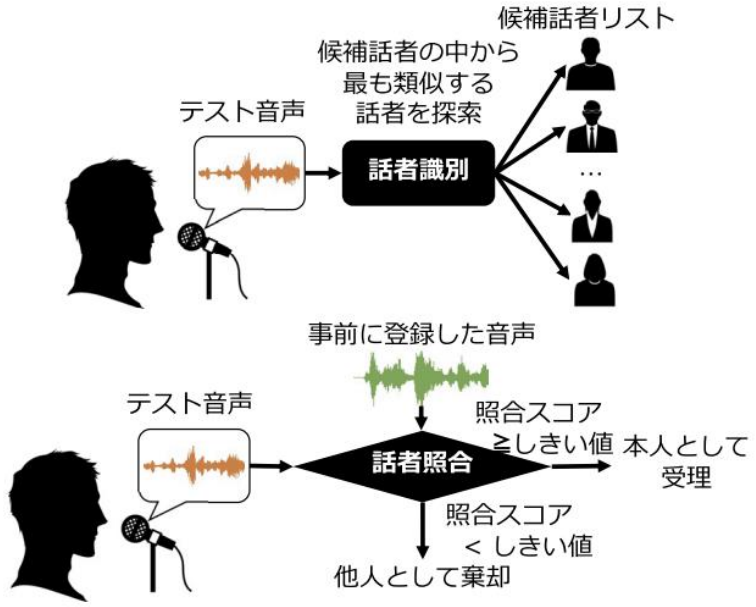
話者認識システムとなりすまし対策

- 話者認識
- 話者照合システムの原理
  - システム構成
  - フロントエンド・バックエンド型の処理の流れ
- 深層話者埋め込みの最新技術
- 課題と今後の展望

https://github.com/tenk-9/ShiotaLab_seminar/blob/master/3rd/paper.md

話者認識

- 音声を生体認証の鍵として用いる
- 定義による分類
  - 話者識別
    - 複数の登録者から提示された音声の話者を探索
    - 1対n
  - 話者照合
    - 提示された二つの音声が同一話者によるものか否かを判定
    - 1対1
- 音声内容による分類
  - テキスト依存型
    - 登録時と照合時に同じ内容の音声を用いる
  - テキスト独立型
    - 登録時と照合時で異なる内容の音声を用いる



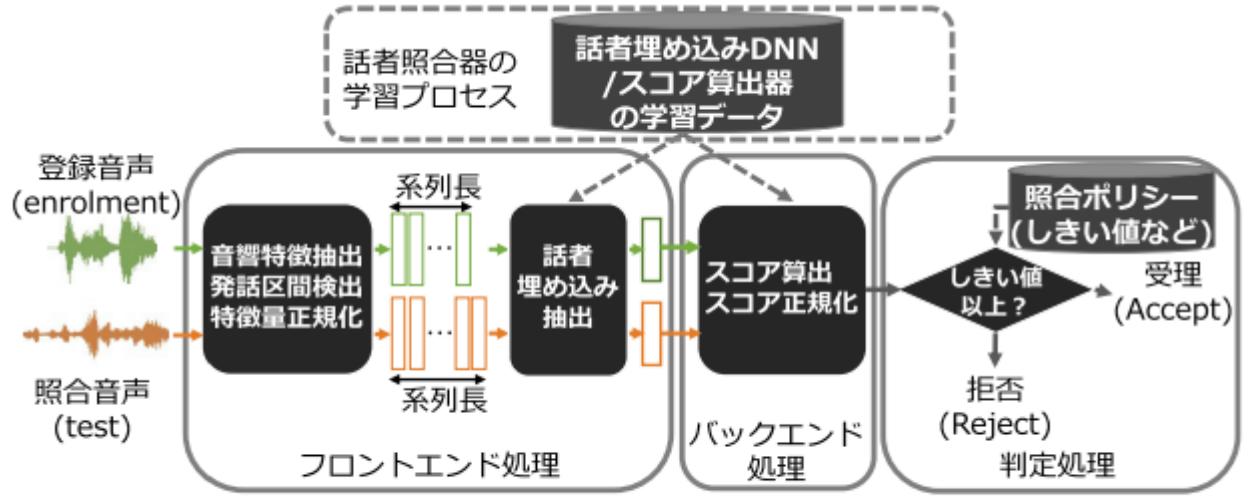
図–2　（上図）話者識別と（下図）話者照合の違い

話者照合システム

- エンドツーエンド型
  - 登録，照合音声のペアから照合スコアを直接計算する
    - 動的時間短縮(DTW)に基づく方法はこれ
- フロントエンド・バックグラウンド型
  - 音声から特徴量を抽出
  - 特徴量に基づき照合スコアを計算
  - 現在主流

フロントエンド・バックグラウンド型の処理の流れ

1. 特徴量抽出
2. 話者埋め込み抽出
3. バックエンド処理



図–3　フロントエンド・バックエンド型話者照合システム

特徴量抽出

- スペクトル包絡に基づく特徴量
  - MFCC(mel-frequency Cepstral coefficients)
  - PLP(perceptual linear prediction)
- フィルタバンク特徴量
  - より生の信号に近い

- 発話区間検出による無音区間の除去も行われる
- 伝送経路等の影響を除去するため，特徴量の各次元について3秒程度の移動窓ごとに平均や分散を正規化する

話者埋め込み抽出

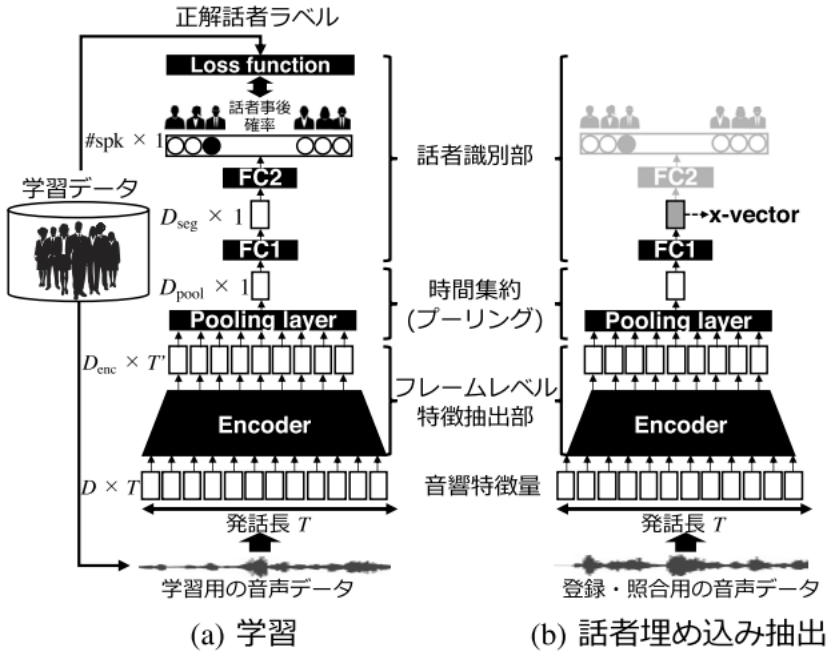
- 発話内容，収録環境，伝送路等の違いにより話者内で特徴量は大きく変わる
- 特徴量の系列長は入力音声の長さに依存する
  - 系列長の違うを吸収する仕組みが必要
- **生成モデルに基づく話者埋め込み**
- **深層話者埋め込み**

生成モデルに基づく話者埋め込み

- 先週の機械学習の部分
- 音声が話者固有のGMMに従うと仮定
- UBMと話者に適合させたGMMの差(GMMスーパベクトル)を話者埋め込みとする
  - 因子分析によってこの次元を削減して得たi-vectorは良い性能を得ていた

深層話者埋め込み

- i-vectorに置き換わる形で近年急速に発展
- 右の構造を持つDNNで話者埋め込みを得る
  - Encoderで時間フレーム単位で特徴抽出
  - プーリング層で時間方向に集約
  - 話者認識部で，集約したベクトルから話者を推定
  - 大量の学習データでパラメータを推定する
- 話者認識部の中間出力(ベクトル)は，話者識別に重要な情報が詰まってるんじゃないの?
- 深層話者埋め込み(x-vector)



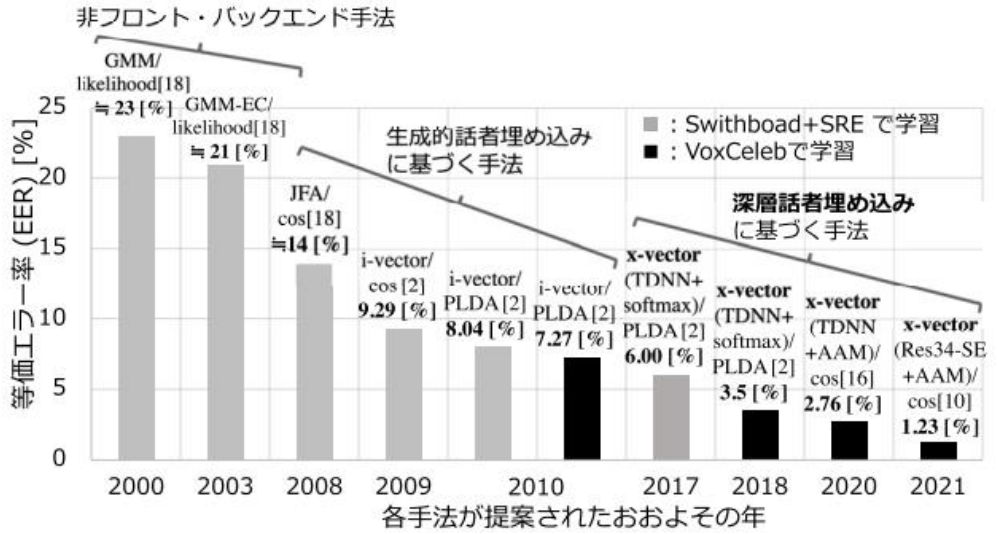
図–4　深層話者埋め込みの (a) 学習と (b) 推論の概念図

バックエンド処理

- 登録音声，照合音声からそれぞれ話者埋め込み(ベクトル)を得，照合スコアを算出
  - コサイン類似度
  - PLDA(Probabilistic linear discriminant analysis)
- 異なる話者において，発話間の照合スコア分布が標準正規分布になるようにスコアを正規化
- 閾値で切って受理or棄却

深層話者埋め込みの最新技術

- エンコーダ
  - TDNN
  - ResNet
- プーリング層
- 目的関数
  - ソフトマックス交差エントロピー
  - AAM softmax
  - 話者埋め込み間の距離
- 学習アルゴリズム
  - 深層話者埋め込みは学習データを増やしたときの性能改善が大きい
    - 学習データを増やしてデータ数を増やす(データ拡張)

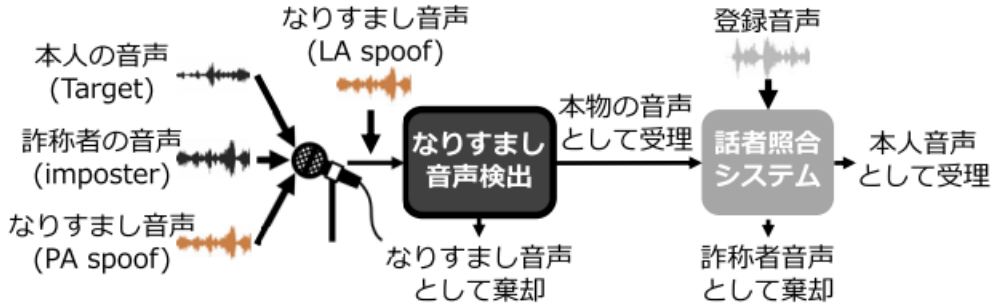


図–6　代表的な手法を SITW core-core タスクで評価した結果
ただし灰色と黒色で示した結果はそれぞれ異なるデータセットで学習した結果であることに注意。

課題と今後の展望

- 音声波形の直接利用
  - 学習可能なフィルタをエンコーダに組み込む
  - 特徴量抽出自体も話者識別の意味で最適化する
- エンドツーエンド話者照合
- モデルサイズ
  - モバイル端末や低資源環境での実行可能性の向上
- 他タスクとの結合学習
  - 方向推定
  - DNNベースの音声強調と同時に最適化
- 環境に対する頑健性
  - 雑音，収録チャネル，言語等の音環境
  - どの環境変数がどの程度影響するかはっきり分かっていない

なりすまし攻撃の種類



図–7　直列型なりすまし検出システムの概要

- 論理的アクセス
  - LA, logical access
  - マイクを介さず直接なりすまし音声を入力
  - 音声合成や声質変換で生成した音声
- 物理的アクセス
  - PA, phisical acess
  - なりすまし音声を照合システムのマイクを通じて入力
  - リプライ攻撃
    - 事前に収録した音声を再生する

なりすまし検出の研究動向

- 音響特徴量の抽出
  - スペクトラム包絡に基づく特徴が広く用いられる
    - MFCC
    - LFCC
    - メルスペクトログラムが特に有効と報告
    - Constant-Q変換
      - 低周波巣帯域でより詳細な周波数解像度
      - 合成音声の種類によっては有効
  - sincNetのような，特徴量抽出自体も最適化の対象とする枠組みも積極的に導入
- バックエンド

なりすまし検出の研究動向

- バックエンド
  - 実音声と，なりすまし音声で学習したGMMに対する尤度比
    - 広く用いられていた
  - DNNを用いて検出スコアを直接計算
    - LFCCを直接入力とするLight CNN
    - 音声波形を入力とするRawNet2を，なりすましか否かの識別に対するソフトマックス交差エントロピー基準で最適化する
  - 様々なデータ拡張を組み合わせることで，未知のコーデックに対する頑健性を確保している

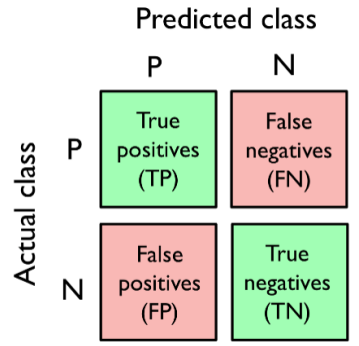
課題と今後

- 合成・リプライ音声のどちらにおいても，学習時と評価時で条件が異なる未知の攻撃に対しては精度が低い.
  - 汎化性能の向上が課題
- 積極的な攻撃に対する対策も求められる
  - 話者認識システムを騙すように最適化された合成音声
  - 攻撃対象の話者照合システムの誤受理率を上げるようなノイズを付加した敵対的サンプルに基づく攻撃
- 呼気がマイクロホンに入ることによるポップノイズを用いた生体検知に基づく方法もある

評価方法

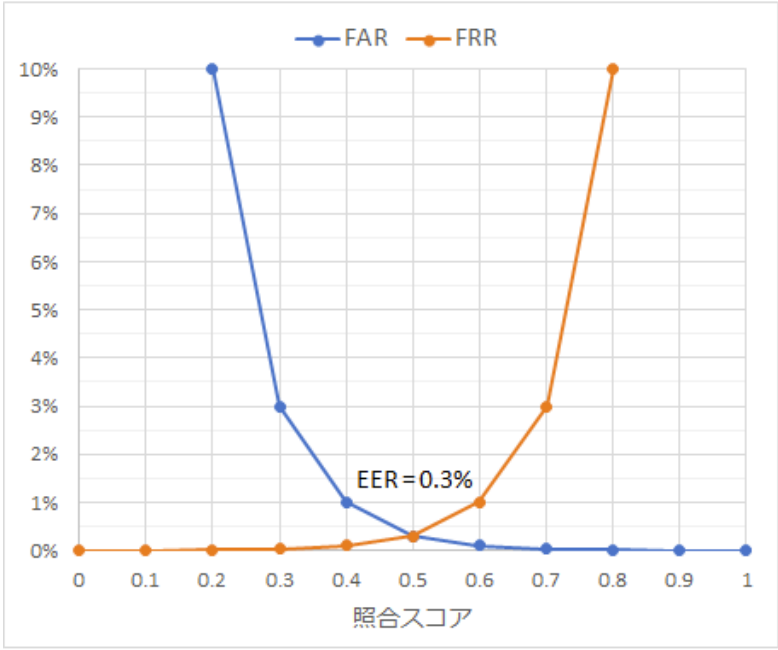
- EER
- DCF
- ROC曲線

EER



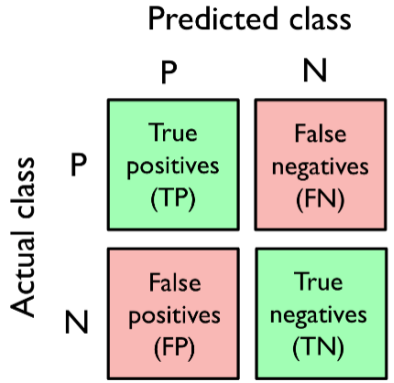
https://github.com/rasbt/machine-learning-book/blob/main/ch06/figures/06_08.png

- 等価エラー率
- FARとFRRが一致するように閾値を調整した際のエラー率
  - FAR
    - 他人受け入れ率
    - False Acceptance Rate
    - FP/(TN+TP)
  - FRR
    - 本人拒否率
    - False Rejection Rate
    - FN/(FN+TP)
  - FARとFRRは逆相関



https://51takahashi.hatenablog.com/entry/2019/04/01/231544

DCF

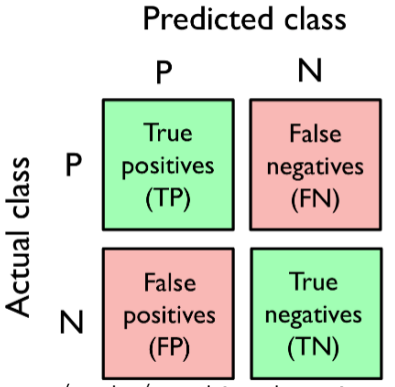


https://github.com/rasbt/machine-learning-book/blob/main/ch06/figures/06_08.png

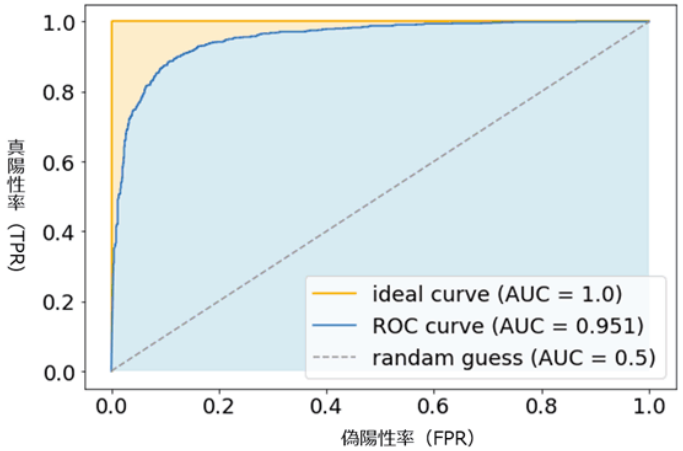
- Detection Cost Function
- 偽陽性(False positive)と偽陰性(False negative)に関連するコストを考慮
- 閾値を変化させ，DCFを最小化する
- DCFが最小となる閾値がベスト_?

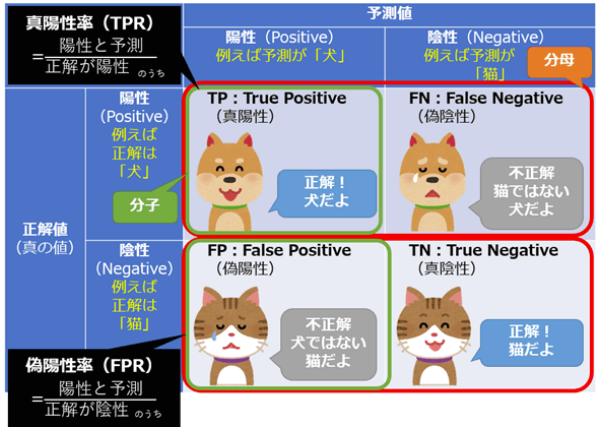
ROC曲線

- 受信者操作特性
- Receiver Operating Characteristic
- 偽陽性率を横軸に，真陽性率を縦軸にとる
  - 偽陽性率(FPR)
    - 間違えて陽性だと判断した割合
    - FalsePositive/Negatives = PF/(FP+TN)
  - 真陽性率(TPR)
    - 正しく陽性と判断できた割合
    - TruePositive/Positives = TP/(TP+FN)



https://github.com/rasbt/machine-learning-book/blob/main/ch06/figures/06_08.png



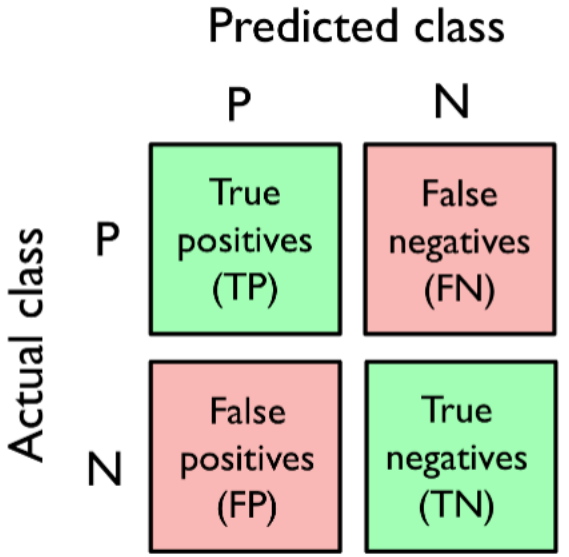


https://atmarkit.itmedia.co.jp/ait/articles/2211/24/news019.html

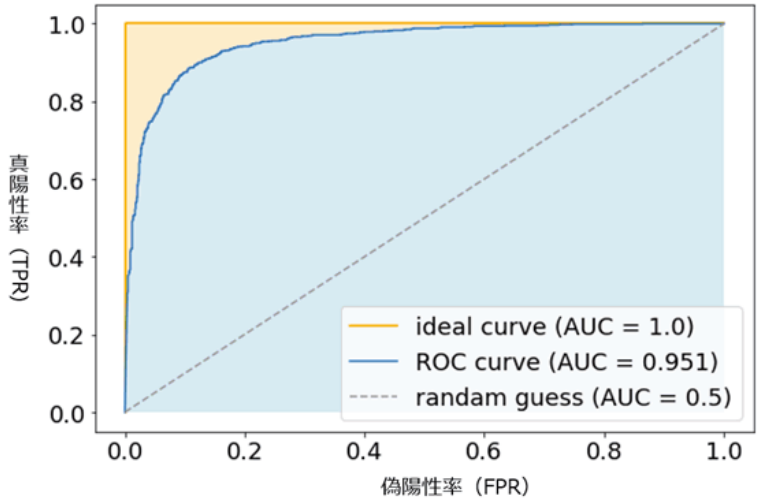
ROC曲線

- 閾値を変えると，偽陽性率，真陽性率が変わる
  - 閾値を上げるほどFPR-, TPR+
  - ランダムだと対角線上に収束
  - 良い分類器ほど左上に膨らむ
  - 理論値は左上張り付き

- AUC
  - Area Under the Curve, 曲線下面積
  - 膨らみ具合を数値化したもの
  - 1に近いほど精度がいい



https://github.com/rasbt/machine-learning-book/blob/main/ch06/figures/06_08.png



https://atmarkit.itmedia.co.jp/ait/articles/2211/24/news019.html