

令和6年度 特別研究報告書

声認証システムの精度向上
についての検討

龍谷大学 先端理工学部 知能情報メディア課程

Y210122 小栗正也

指導教員 三好力 教授

研究梗概

近年、音声を鍵として生体認証を行う声認証の技術の利用が拡大してきている。声認証技術は虹彩認証や指紋認証とは異なり専用のセンサを必要とせずマイクを用意するだけで入力データを入手することができるため容易に実用が可能だという点と、非接触で認証可能なため携帯電話などの通信を経由することでも認証が可能という利点がある。すでにコールセンターでの個人の特定や金融機関での不正検知システムとしてセキュリティシステムとして導入している企業もある。声認証技術は話者が登録者本人か否かを確認する話者照合と、音声に登録されている人のうちのどの人物かを識別する話者識別に分けることができる。主に、話者照合では本人確認で使用され、話者識別では扉の認証鍵のような用途で使用される。しかし、声認証システムには他の生体認証システムと異なり識別精度向上の困難さとなりすましに対して弱いという欠点がある。声認証では音声を入力する周りの環境によって精度が大きく左右され、周りの雑音を取り除いてから識別する必要がある。なりすましに対しても登録された個人の音声を合成で作ってなりすます方法もあるためこれらを解決しないと一般的に普及することが難しいと考える。

本研究ではこれらの問題のうち、周りの雑音を軽減させて識別率の向上を図ることを目的として、実際に音声を収録して雑音を低減し機械学習を用いて識別率が向上するかの検証を行う。

目次

第1章 研究背景.....	1
第2章 既存技術.....	2
2.1 声認証の種類.....	2
2.2 テキスト独立型の話者照合の例.....	2
2.3 ノイズ除去手法.....	2
2.3.1 短時間フーリエ変換.....	2
2.4 メル周波数ケプストラム係数.....	3
2.5 SVM での分類と照合.....	4
第3章 提案手法.....	6
3.2.2 複数マイクを使用することによるノイズ除去.....	6
3.3.1 位置推定によるなりすまし対策.....	6
3.3.2 QDA による機械学習による識別.....	7
第4章 実験.....	9
4.1.1 ノイズ推定.....	9
4.1.2 位置推定.....	10
4.1.3 事前実験の考察.....	10
4.2 本実験概要.....	11
4.3 録音.....	11
4.5 音声の特徴量の抽出.....	13
4.6 機械学習.....	13
4.7 照合精度の検証.....	13
第5章 考察.....	15
第6章 おわりに.....	16
謝辞.....	17

第 1 章 研究背景

近年、指紋や顔などを用いた生体認証技術が普及してきている。その中でも音声を鍵として認証を行う技術が声認証である。声認証は実現するためにマイクを用いるだけでデータを入手することができるため容易に実現可能という点や、指紋認証や虹彩認証とは異なり物理的な接触を必要としない認証技術のため電話越しなどから遠隔で行えるという利点がある。そのため銀行や証券会社などの金融機関での不正検知やコールセンターでの本人確認などで実用化が進んでいる[1]。

一方で声認証の拡大に伴って問題点も存在している。録音した音声を他人が認証を行うリプレイ攻撃や AI や音声合成技術を使って他人の声を模倣する技術などがあり既存の音声認識技術では防ぎきれない可能性ある。さらに声認証は自身の生体情報での認証とはいえ自身の状態や周りの環境によって制度が低下する可能性がある。例えば風邪を引いて声がかすれた時やストレスなどで声のトーンが変わる状態の変化と、公共の場や車内などの周りの環境でノイズが多かった場合には認証が難しくなったりする可能性がある。

本研究では幅広い場面で声認証を普及させるためになりすましや自身の状態、周りの環境での認証精度の低下についての対策を行い、認識制度の向上を目的として研究を行う。

第 2 章 既存技術

2.1 声認証の種類

声認証の方法として大きく分けると入力音声に登録者本人か否かを判断する話者照合と、入力音声に登録されている複数人のうちの人物であることを識別する話者識別になる。さらに認識時の発話内容について、登録時と同じ発話内容に固定するテキスト依存型、システムが発話内容を提示するテキスト指定型、発話内容に制限を用いないテキスト独立型の 3 つに分類することができる。 [2]

2.2 テキスト独立型の話者照合の例

アメリカ国立標準技術研究所(NIST)が開催したテストでは電話での会話内での人物特定という課題では音源に背景雑音や回線のノイズが非常に強い中での識別でこれまでは 89%ほどの精度であったが、日本電気株式会社(NEC)が行った認証の精度が 95%という数値を出すことができた[3]。NEC が精度を上げられたのにはノイズ除去に加えて特徴を抽出してニューラルネットワークを組み込み、識別を行ったからである。

以下ではノイズ除去、特徴抽出、識別について代表的な技術を述べる。

2.3 ノイズ除去手法

ノイズ除去とは音声信号に含まれる不要な成分(ノイズ)を減らして目的とする音声信号をより鮮明にする方法である。以下の方法によってノイズの除去を行う。

2.3.1 短時間フーリエ変換

音声などの時間変化する音声信号の周波数と時間(位相)の解析する際に、音声信号全体をフーリエ変換してしまうと信号全体の周波数構成しか得ることができないため、音声信号を細かく区切ってフーリエ変換を行い時間的変化に伴う周波数の変動を解析する処理のことを短時間フーリエ変換(STFT, Short-Time Fourier Transform)という。 [4]

短時間フーリエ変換の仕組みはまず全体の音声信号を小さな時間間隔(ウィンドウ)で分割して各ウィンドウで区切られた信号にフーリエ変換を行い周波数成分を抽出し、ウィンドウをずらしながら同じ処理を繰り返すものである。これによって時間軸と周波数軸における信号強度の変化を表す情報を得ることができ、これをあらわした図 1 のようなデータをスペクトグラムという。

STFT の計算式

$$\text{STFT}\{x(t)\}(r, \omega) = \int_{-\infty}^{\infty} x(t) \cdot \omega(r-t) \cdot e^{-j\omega t} dt \quad - (1)$$

$x(t)$: 解析対象の音声信号

$\omega(r-t)$: ウィンドウ関数

r : 時間のシフト量

ω : 周波数

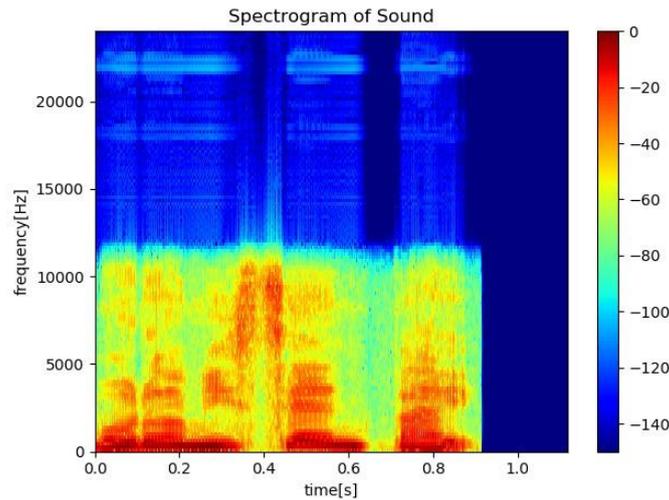


図1：短時間フーリエ変換後のスペクトログラム

短時間フーリエ変換によって得られたスペクトログラムからノイズを特定し除去する。ノイズは特定の周波数に集中している場合が多く、低周波数帯に集中する電源ノイズや高周波数帯に集中する雑音などをスペクトログラムで可視化することによって除去しやすくなる。

特定の周波数を取り除いた後に時間と周波数領域から元の時間領域に戻すために逆短時間フーリエ変換 (ISFTF, Inverse Short-Time Fourier Transform) を行うことによってノイズが除去された音声データに戻すことができる。

2.4 メル周波数ケプストラム係数

メル周波数ケプストラム係数(MFCC, Mel-Frequency Cepstral Coefficients)とは音声信号の特徴を表す表現方法で、音声処理の特徴量の抽出に使われる手法である。メル周波数ケプストラムは音声信号を人間の聴覚に合わせた周波数スケール(メル尺度)に変換することで音声の特徴をより効果的に表現できる。人間が敏感である低周波数成分ほど高い解像度を持ち鈍感である高周波数に向かうほど解像度が粗くなる。

変換手順としてまず音声波形を適当な長さに分割して窓関数をかけて高速フーリエ変換を行い時間領域の音声信号を時間 - 周波数領域のスペクトルに変換する。そしてスペクトルをメル尺度に基づいたメルスペクトルに変換する。周波数 f をメル周波数 m に変換する式は

$$m = 2595 \log\left(\frac{f}{700} + 1\right) \quad \text{---(2)}$$

で表すことができる。

周波数とメル周波数の関係をグラフに表すと図2のようになる

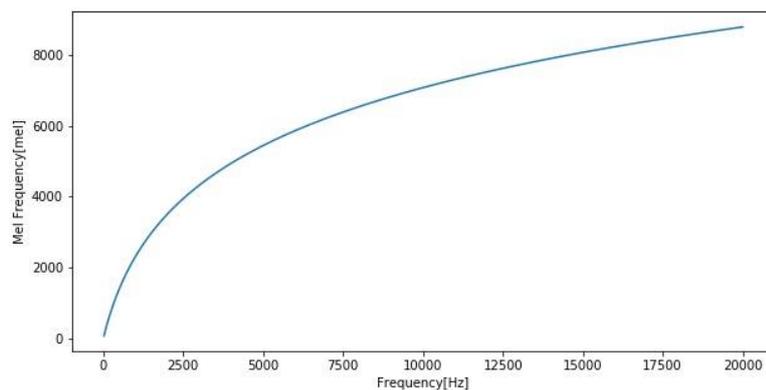


図 2 : 周波数とメル周波数の関係

その後メルフィルタバンクというメルスケールに基づいた三角形のバンドパスフィルタをオーバーラップさせながら並べるフィルタバンクをかける。メルフィルタバンクは周波数領域での N 個の三角窓がメル尺度上で等間隔になるようにならべたものなので図 3 のように低周波ほど幅が狭く、高周波ほど幅が広いフィルタバンクになる。これによって得られた N 個の周波数帯域の特徴量を離散コサイン変換する。これによりメルケプストラム係数を求めることができ低次のケプストラム係数が主要な特徴を表すようになるため 12 次元までのケプストラム係数を抽出する。 [5]

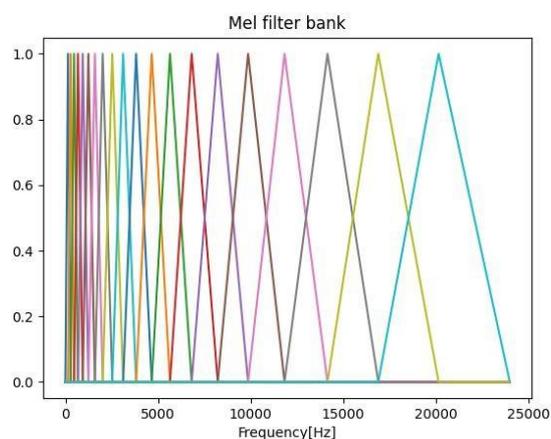


図 3 : メルフィルタバンク

2.5 SVM での分類と照合

SVM(Support Vector Machine)はパターン認識やデータ分類で使われている教師あり学習アルゴリズムで 2 クラス間を区別するための境界線を見つけてデータを分類する学習方法である。カーネル関数を利用することによってデータを高次元へと移すことによって単純な直線で分けられない場合でも分類することが可能となる。

メル周波数ケプストラム係数を SVM を用いて学習することによって特徴量だけで事前に登録された話者が新しく音声が入力すると分類されて識別することが可能になる。 [6]

2.6 既存研究の問題点

既存研究では声認証での識別精度が 95%ほどで意外と高く感じるかもしれないが、他の生体認証の精度と比べてみると誤認識の確率が指紋認証だと 1 万分の 1、顔認証では 10 万分の 1、虹彩認証では 150 万分の 1 の確率と他の生体認証に比べてかなり精度が悪いとわかる。そのため 20 回に一回間違えるようでは広く普及して実用化に至ることができないと考えられる。また、他の人が録音データを用いて認証しようとするリプレイ攻撃に対して対策ができていないことも問題点として挙げられる。

第 3 章 提案手法

3.1 既存技術の改善案

識別率を向上させる方法としてノイズの除去率の上げること、機械学習による識別方法を変えることによって精度向上をできると考える。周りに環境音が入っていても識別できるようなノイズ除去と機械学習の精度向上目的として実験を行う。

3.2 ノイズ除去の改善案

3.2.1 窓関数を用いた短時間フーリエ変換によるノイズ除去

ノイズ除去の主流な方法としてフーリエ変換を行って除去する方法が挙げられていたがこの方法では特定の周波数帯域への除去は可能になるがホワイトノイズなどのすべての周波数帯域にわたるノイズの除去や瞬間的な変動が大きいノイズへの除去には不向きという点がある。そこで、人の音声周波数が集中しているため 300Hz~3400Hz の範囲以外の周波数帯域に対して窓関数をかけることによってノイズ除去を行う。

3.2.2 複数マイクを使用することによるノイズ除去

音声入力の主マイクとそこから離れた場所に置いたマイクの 2 つのマイクを使用して録音することにより、2 つの音声データを比較して雑音部分を差し引き、発声部分を強調することによって人の音声周波数帯である 300Hz~3400Hz の範囲でも雑音部分の除去を行う。発生源から近くのマイクには話者の音声が大きく、遠くのマイクは小さくなり、ノイズ部分の信号は両方のマイクにも一定に記録されるため、近いマイクのデータから遠いマイクのデータを減算することで音声部分の強調とノイズ部分の除去を行う。

3.3 識別方法の改善案

3.3.1 位置推定によるなりすまし対策

マイクを 2 本使用することで 2 つのマイクに入力された音声データの時間差から発話者の音源位置を求める。本人が発声する場合、音声源の位置は前後の方向ではマイクまでの距離、上下の方向では口の高さとなるが、なりすまして機械から音声を入力する場合には本人発声と異なる位置から発声すると考えられる。発生源の位置を推定することで実際に人が発声しているか、機械で流しているかを見分けて、なりすましへの対策を行う。

発生源から 2 本のマイクへの到達時間差を求める方法としてクロスコリレーション(Cross-Correlation)[7]という 2 つの信号間の類似性のある時間遅延に基づいて評価する指標があり、クロスコリレーションの結果に対して音速をかけることによって音源の位置推定を行う。

クロスコリレーションの求め方は以下の式で求めることができる

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) \cdot y(t + \tau) dt \quad - (3)$$

- $x(t)$:信号 1(基準信号)
- $y(t)$:信号 2(比較信号)
- τ :シフト量(時間遅延)
- $R_{xy}(\tau)$:シフト量 τ における相関値

3.3.2 QDA による機械学習による識別

識別を行うための機械学習として SVM が代表的であるが今回実験で行う手法では多くの特徴量を分類する必要があるためサンプル数が多いときに識別がしやすい二次判別分析(QDA Quadratic Discriminant Analysis)を用いて識別を行う方法を提案する。QDA は線形判別分析(LDA Linear Discriminant Analysis)の拡張版であり、QDA も LDA もデータを使ってクラスを区別するという点では同じだが、区別するための境界が LDA は直線や平面の線形であるのに対し、QDA では二次曲線や曲面の非線形で境界を判別する違いがある。また、LDA ではクラス間で同じ共分散行列を仮定するのに対し QDA ではクラスごとに異なる共分散行列を仮定することができるためクラスの分布が複雑な場合でも識別が可能である。QDA の方が計算コストとしては高く学習データが少ないと過学習するというデメリットもある。

機械学習のアルゴリズムとしては訓練データの特徴量分布をモデル化して計算関数を決定し、クラスを判別するというものである[8]。

QDA の基本概念として、各クラスがガウス分布に従うと仮定し、それぞれのクラスの分布をモデル化する。この仮定の下で、ベイズの定理を用いてデータが属するクラスを予測するというものである。

QDA を求めるための準備として特徴量ベクトル x_i ,クラスラベル y_i を設定し、各クラス k のデータの平均ベクトルを μ_k ,共分散行列を Σ_k ,クラス k に属するデータの割合をクラス事前確率 π_k ,特徴量の次元 d とすると、クラスごとの確率密度関数は以下の式で表すことができる

$$p(x|y = k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right) \quad - (4)$$

そして各クラスに対して事後確率を以下の式で計算する。

$$P(y = k|x) = \frac{p(x|y=k)\pi_k}{\sum_j p(x|y=j)\pi_j} \quad - (5)$$

この値が最大になる事後確率を持つクラスが予想クラスとなる。

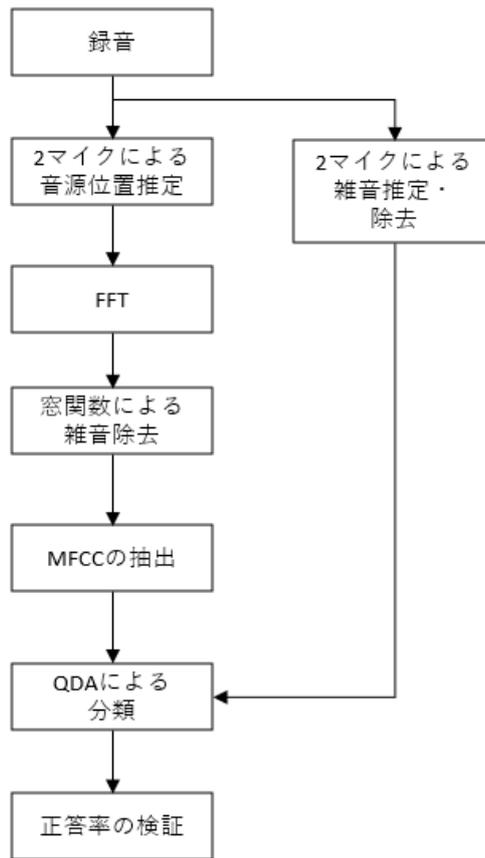


図4：提案システムにおける処理のフローチャート

第4章 実験

4.1 事前実験

本実験に入る前に提案した手法が上手く作用するかを確認するための事前実験を行う。

マイクを2本使用して録音し、ノイズの推定と位置の推定を行えるかを実験する。図5のように発話者の立ち位置から20cm正面にマイクを一つ設置して、さらにその1メートル下の位置にもう一つのマイクを設置する。実際に録音した正面のマイクの音声データを図6、下のマイクの音声データを図7に示す。

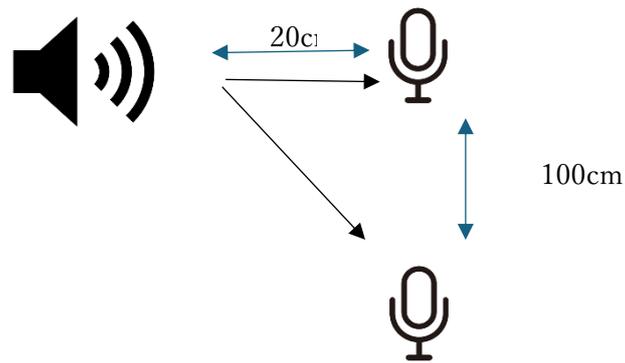


図5：実験のイメージ

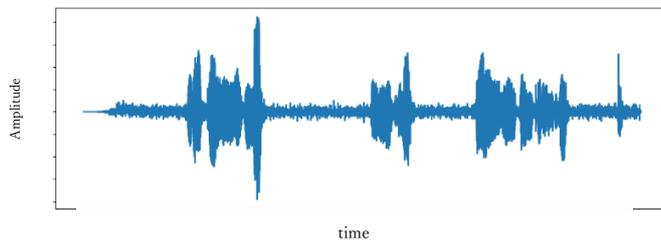


図6：正面の音声データ

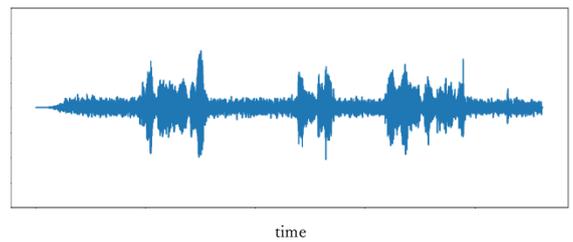


図7：下の音声データ

4.1.1 ノイズ推定

得られた2つのデータから発音源から近い方の振幅が大きくなり遠くなると振幅が小さくなるのが分かり、近い位置の信号から遠い位置の信号を差し引いた結果を図8に示す。

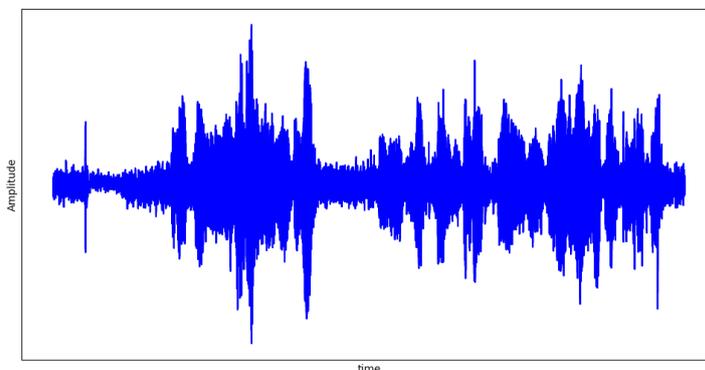


図8：差分計算した音声データ

結果より、2 マイク間の振幅差を利用して音声のノイズの除去を行おうとするとほとんど効果を得ることができず、かえって悪化してしまう結果となった。原因としては 2 マイク間の距離を離れたことで近いマイクと遠いマイクでの発生源から届くまでの時間とノイズ源の位置の違いからノイズ部分と音声部分の区別があいまいになりノイズの除去が上手くできなかったと考える。

4.1.2 位置推定

次に、距離の離れた 2 本のマイクに音声が届く時間差を利用して発生源の推定を行う。正面のマイクから 20cm 離れた場所から発声した場合と 50cm 離れた場所から発生した場合の 2 通りで録音し、正確に距離が推定できるかを確認する。発生源から 2 本のマイクへ到達するまでに時間差から音速を用いて計算することによって距離を推定した実験結果を表 1 に示す

	音源の距離	
	20cm	50cm
到達時間差	0.001542秒	0.00289秒
距離差	52.87cm	99.32cm

表 1：距離の推定結果

実験結果から 2 マイク間の到達する時間差から音源までの距離を推測すると実際の音源の距離から 30 cm から 50cm ほどのずれが生じる結果となった。原因としては音源からマイクまでの距離が小さすぎるため誤差の範囲が大きくなってしまふ可能性があるのと、音源に対して遠いマイクへの角度が立体的であるため床や壁からの反射音まで入ってしまい正確に時間差を測定することができずに誤差が生じてしまっていると考えられる。

4.1.3 事前実験の考察

以上の事前実験から、2 本のマイクを使用してもノイズ除去や位置推定に対して有効性がないと結論付けるため、本実験では使用するマイクを 1 つにして実験を行うこととする。本実験で行う手順を以下のフローチャート図 9 に示す。



図9：本実験で行う実験手順

4.2 本実験概要

被験者 8 人の音声を録音し、その音声のノイズの除去・特徴量を抽出し識別してその声を誰が発声したかを当てる精度を測定する実験を行う。

4.3 録音

被験者 8 人に「おはようございます」「開け、ゴマ」「こだわり酒場のレモンサワー」「こんにちは、今日はいい天気ですね」「生麦、生米、生卵」の 5 つのフレーズの録音を行う。録音するための機材は図 10 で行い、録音環境はカラオケルームの中で行い、裏で電車の音[9]を流して雑音を作り出した中で録音を行った。



図 10: 録音機材 H4essential Handy Recorder

4.4 ノイズの除去

録音した音声を時間短領域から周波数帯領域に変換するために窓関数をかけながら高速フーリエ変換を行う。それによって求められたスペクトルからノイズに当たる周波数成分を特定して閾値を計算し、マスクすることによってノイズの除去を行う。その後逆フーリエ変換を行うことによって周波数帯領域から元の時間帯領域に戻して音声を復元する。

「こんにちは、今日はいい天気ですね」発声した波形が以下の図 11 で、振幅が大きいところは実際に発声している部分だが、小さな振幅がずっと続いている部分がノイズの部分だと推測できる。

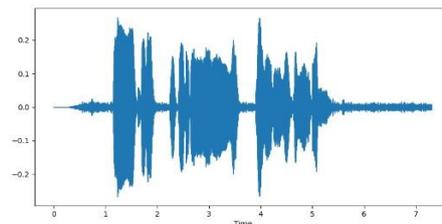


図 11： 雑音が混じった音声波形

これに窓関数をかけて高速フーリエ変換を行うと図 12 のようなスペクトルを得ることができる。図から見て取れるように低周波帯には大きな波ができていますが、高周波帯には小さな波が少しずつあることが確認できる。これがノイズの部分であると推測できる。そのため 3400Hz をしきい値としてそれ以上の周波数帯をマスクすると図 13 のようになる

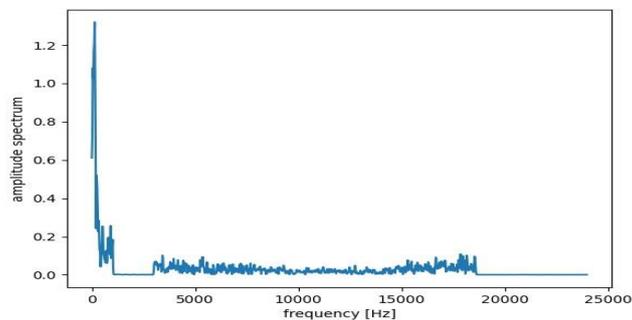


図 12： フーリエ変換を行って得たスペクトル

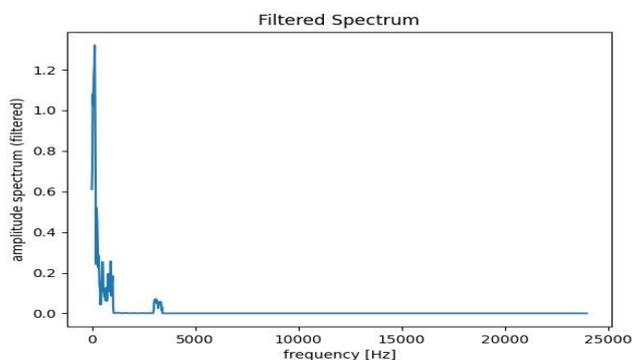


図 13： 閾値処理を行ったスペクトル

4.5 音声の特徴量の抽出

音声の特徴量を求めるためにメル周波数ケプストラム係数を求める。先ほどノイズを取り除いたスペクトルに対してメル尺度に基づいたメルスペクトルに対応するように以下の式をかけて変換を行う。

その後メルスケールに基づいた三角形のバンドパスフィルタをオーバーラップさせながら並べるメルフィルタバンクをかける。これによって得られた周波数帯域の特徴量を離散コサイン変換することでメルスペクトラム係数が求めることができる。実験では1次から12次までのメルスペクトラム係数を抽出して特徴量とする。表2がノイズ除去を行わずに抽出したメルスペクトラム係数で表3がノイズ除去を行ってから抽出したものである。

mfcc_1	mfcc_2	mfcc_3	mfcc_4	mfcc_5	mfcc_6	mfcc_7	mfcc_8	mfcc_9	mfcc_10	mfcc_11	mfcc_12	target
94.66985	-2.4224	3.961415	7.361192	-11.4625	-6.61693	11.30879	-15.9592	-1.91456	-9.65497	-12.0624	-3.84441	1
75.69906	3.871368	8.144995	6.488169	-12.2699	-3.80769	13.05622	-17.485	-0.07065	-10.4919	-11.3082	-2.0011	1
72.86626	6.698624	7.763494	-2.33796	-11.4549	-5.14659	12.6262	-16.0542	-0.53202	-7.76774	-11.098	-2.29212	1
84.0347	4.178397	1.682091	-0.41016	-12.5369	-2.94634	14.34065	-16.3539	-4.20287	-9.80306	-12.6269	0.032268	1
77.2935	4.878837	11.04998	8.203699	-7.28073	-9.23124	14.14019	-14.3597	-4.06587	-10.9877	-12.5767	-5.48775	1

表2 ノイズ除去を行わずに抽出したメルスペクトラム係数

mfcc_1	mfcc_2	mfcc_3	mfcc_4	mfcc_5	mfcc_6	mfcc_7	mfcc_8	mfcc_9	mfcc_10	mfcc_11	mfcc_12	target
83.79829	1.46748364	7.314268112	7.54068	-9.85428	-5.67695	11.30896	-15.3314	-0.7602	-11.4233	-11.6689	-2.34736	1
75.02148	8.207076073	6.494332314	6.925714	-5.82864	-5.46148	13.47572	-10.2425	-5.04244	-6.86541	-8.83647	-1.77974	1
57.36061	1.692656159	1.900591016	0.139094	-10.8475	-8.03707	10.51373	-23.7669	-1.29805	-8.55037	-13.8489	-4.85869	1
74.40749	1.372265816	12.17408466	3.539077	-13.3178	-3.74161	13.616	-15.8678	-1.08297	-10.8189	-11.7665	-1.42556	1
68.33314	6.771344185	13.48001766	5.789091	-10.5958	-4.83383	14.48937	-15.621	-3.47785	-11.6146	-13.6369	-2.50594	1

表3 ノイズ除去を行って抽出したメルスペクトラム係数

4.6 機械学習

抽出したメルスペクトラム係数を機械学習を行い分類する。機械学習の方法は二次判別分析(QDA Quadratic Discriminant Analysis)と SVM(Support Vector Machine)の2つの方法で行う。

4.7 照合精度の検証

機械学習を行って分類した学習モデルに検証用データ音声を10個入力して正答率の結果を表3、表4に示す。表4はQDAを用いた検証結果で、表5はSVMを用いた検証結果である。

正答率		
	ノイズ除去	
被験者	有り	無し
A	90%	80%
B	70%	70%
C	80%	60%
D	100%	100%
E	80%	80%
F	90%	80%
G	90%	90%
H	70%	80%

表 4 QDA を用いた検証結果

正答率		
	ノイズ除去	
被験者	有り	無し
A	60%	60%
B	60%	60%
C	70%	60%
D	90%	90%
E	50%	40%
F	80%	70%
G	60%	50%
H	40%	40%

表 5 SVM を用いた検証結果

正答率を見てみると平均値が QDA を用いた場合、ノイズ除去有りが 83.75% で除去無しが 80.0 %、SVM を用いた場合、ノイズ除去有りが 63.75% で除去無しが 58.75% という結果になった。

全体的に QDA を用いた場合の方が正答率は高く、ノイズ除去を行った方がわずかに正答率は高くなるという結果が得られた。

第5章 考察

結果より提案手法である2本のマイクを使用する手法は有効なノイズ除去や位置推定を行うことができないという結果となった。また、録音した音声の短時間フーリエ変換を行ってノイズの周波数帯を特定し、その周波数帯にフィルタをかけて抑制するノイズ除去を行ってからメルフィルタをかけて12個のメルスペクトラム係数を抽出し、二次判別分析で分類を行うという手法は約83%の精度で照合できるという結果になった。この結果は先行研究である NEC による声認証技術の精度の95%と比べると下回る結果となった。本研究が既存技術の精度を超えられなかった要因としては、2本のマイクを利用するアイデアが失敗した事や、登録した音声データの数が少なかったことが原因であると考えられる。先行研究である NEC の技術では1つの学習データに対してノイズや改変を加えることによって別の音声データをつくるなどしてデータ量を20倍に行っていたのに対し、本研究では1人に対して5つの音声データで識別をしていたため学習するためのデータが不足していたため十分な結果が得ることができなかったと考える。

第6章 おわりに

本研究では話者照合システムの精度向上を目的として音声に含まれるノイズの除去と機械学習による微量の識別を行う手法について提案した。結果としては2本のマイクを使用した場合には有効的なノイズ除去ができず、発生源推定についても十分な推定は行えなかった。一本のマイクを使用した場合のノイズの除去については約3%の精度を向上させることに成功したが、それでも精度が約83%と目標としていた95%以上の精度を出すことができなかった。

今後の課題として、本研究で使用した学習音声データを増やすことによる機械学習の改善や照合まで時間がかかってしまうという点で実用化には程遠い点を改善していくことが考えられる。また、本研究では行えなかったがなりすまし攻撃の対策として音声以外の認証システムとの併用や合成音声を見抜いてはじくようなシステムの導入を検討することが今後の課題であると考えられる。

謝辞

本論文を作成するにあたり、ご指導していただきました三好教授に深く感謝いたします。
また、研究の協議をしていただいた三好研究室の皆様、実験に協力していただいた皆様に感謝いたします。

参考文献

[1] Pindrop Security,Inc.による声紋認証システムの導入例

https://www.accrete-inc.com/news_information/20230602/

[2]話者照合

https://www.jstage.jst.go.jp/article/isciesci/62/2/62_63/_pdf/-char/ja

[3]NEC による声認証技術の強化

https://jpn.nec.com/press/201902/20190219_01.html

[4] 滝口哲也(2021)「日本音響学会編 音響講座 vol.6 音声(上)」 pp.81-85 コロナ社

[5]岩野公司(2023)「日本音響学会編 音響講座 vol.7 音声(下)」 pp.14-16 コロナ社

[6]SVM

<https://qiita.com/renesisu727/items/964005bd29aa680ad82d>

[7]クロスコリレーション

<https://bsd.neuroinf.jp/wiki/%E7%9B%B8%E4%BA%92%E7%9B%B8%E9%96%A2%E8%A7%A3%E6%9E%90>

[8]QDA

https://toukei-lab.com/discriminant-analysis#google_vignette

[9]電車の音

<https://youtu.be/cctUr8ioNxs?si=Ne0qo6iukYENGZeH>