

# 走行中の自動車内環境での音声による個人認証

田本 篤喜

Tamoto Atsuki

法政大学情報科学部デジタルメディア学科

atsuki.tamoto.6u@stu.hosei.ac.jp

## Abstract

Individual authentication using speech is called as speaker verification. Speaker verification which is implemented in portable device is used in many situation. In this study, we work on speaker verification while driving. Noise and long term feature variability are problems of using speaker verification in driving car. In consideration of characteristic of noise in driving car, Spectral Subtraction and cutting low frequency is implemented in noise removal phase. We evaluate long term variability by using speech which is recorded after about 10 months from first enrollment. False Reject Rate(FRR) decreased to 66.6 % on average by implementation of noise removal phase. Increase of FRR by noise removal phase is improved by update template using accepted speech. About long term variability, FRR didn't change for 10 months. This result shows that GMM Posteriorgram is valid to inter-speaker variability. We need to considerate Lombard effect as a problem in the future.

## 1 まえがき

現在、主に個人の認証では、パスワードやカードなどの、ユーザーの所有物に基づく認証が行われてきた。しかしこれらの手法は、盗難・紛失・忘却などの危険を抱えており、これらの危険のためにカード紛失時の再発行や、パスワードの管理が必要となっていた [1]。しかし人間の生体特徴を個人の認証に用いることでこれらの危険は減らすことができる。そのうち、声による人物特定技術(話者認識)は、スマートフォンのロック解除や電話での買い物、電話による送金、個人の秘密に関する情報へのアクセス許可など、多くの応用があり、既に実用化されている例もある。音声を入力に用いることは、手足や目、姿勢の拘束がなく操作することが可能であり、音声の大きな利点である。

上記の利点に着目し本研究では、運転中におけるスマートフォンのロック解除を想定し、走行中の自動車内環境での音声による個人認証精度の向上に取り組む。運転中はハンドルを操作しているために手を離せず、目の視点も自由に変えることはできない。視点は常に遮られずに前を向いていなければならないので、顔認証なども使用できない、などの状況が考えられる。それ以前に、通話のために携帯電話を運転中に保持・操作することは、道路交通法に違反してしまうことにもなる。この場合、声は自由に発することができるため、音声によるロック解除が最も安全である。本研究では、話者照合を使用するにあたり、様々な問題が発生する走行中の自動車内での、話者照合の頑健性を高めることを目的とする。

## 2 話者照合

音声から個人性情報を抽出して個人を認識することを話者認識という。話者認識の形態は、話者識別と話者照合に分けることができる。話者識別とは、未知音声を、あらかじめ登録してある種々の話者の音声(標準パターン)と比較して、最も類似している標準パターンを選び、その登録話者が発声した音声である

と判定するものである。一方話者照合とは、未知音声が発声者が名乗った本人のものであるか否かを、名乗った本人の標準パターンと未知音声との類似の度合いによって判定をするものである [2]。本研究で対象とするシナリオは上記のうちの話者照合に該当する。

話者照合のタスクは、発話内容の違いから3種類に分類できる [3]。一つ目は照合に使用する音声を登録時と同じ発話内容に固定するテキスト依存型である。iPhoneのSiriの声による持ち主登録機能はこのテキスト依存型の話者照合であり、「Hey,siri」という発話を用いて照合を行う [4]。他人が申告者の発声を事前に録音し、認証時にそれを悪用するような詐称方法には脆弱である。主な手法として、HMM [5] や DTW [6] を用いて時間的な情報を踏まえたモデル化をすることで高い精度が得られる。DTWを用いた話者照合は、HMMに比べて少ないデータからテンプレートを作成するという利点がある。このデータ量の少なさは精度低下を招くことがあるが、それを補う手法として、GMM Posteriorgramがある [7]。二つ目は照合に使用する音声を、使用するたびにシステムが提示するテキスト指定型である。これはテキスト依存型の録音問題対策として開発された照合方法である。三つ目は照合に使用する音声に制限を設けないテキスト独立型である。テキスト依存型のように決まったテキストを覚える必要がないという利点がある。主にGMM-UBMを用いるなど、統計的にモデル化を行っている [8]。本研究では、固定の発話内容を照合に用いるため、一つ目のテキスト依存型話者照合である。

## 3 走行中の自動車内における話者照合

話者照合は、スマートフォンのようなウェアラブル端末に実装される場合、多くの場面での使用が想定される。本研究ではその一つとして、運転中の話者照合に取り組む。走行中の自動車内で話者照合する場合の問題点を以下に示す。

1. 雑音
2. 発声時期の違いによる話者内特徴変動
3. ロンバード効果

雑音 - 走行中の自動車内は、静かな室内と違い、走行音や風切り音、エンジン音などの雑音がある。話者照合では主に周波数領域で特徴抽出を行うが、雑音成分が音声区間に混じっていると周波数領域で雑音の周波数成分も加わってしまうため、精度低下・音声としての特徴抽出を妨げる。雑音は、走行車内で話者照合システムを使用する問題点のうち最も主要な問題である。

発声時期の違いによる話者内特徴変動 - 話者照合システムは一度の登録から長期的な使用が見込まれる。一方で、人間の発話は、同じ人が同じ言葉を繰り返し発声しても、ある程度時期をおくと、発声の仕方が変わり、音声のスペクトルが変化する。登録者本人としての特徴が重要な話者照合ではこの特徴変動が認識系の性能に影響を及ぼす。話者照合においては、標準テンプレート作成のための学習サンプルと未知入力音声が発声される時期に必ず隔たりがある。さらに、個人性情報は音韻性情報に比べて細かい特徴であるため、この発声時期の違いによる変動の影響を受けやすい [9]。

ロンバード効果 - 人間の耳に入る背景雑音が多くなればなるほど、それに伴って無意識に声が大きくなる現象のことである。

本研究で想定している車内の場合は、背景雑音が走行音に対応する。

本研究では、雑音、発声時期の違いによる話者内特徴変動、の二つの問題に対して対策手法を導入し、頑健な話者照合の構築を目指す。評価においても、雑音と発声時期の違いによる話者内特徴変動のみを考慮したデータを使用し評価実験を行う。

### 3.1 処理の流れ

DTW を用いたテキスト依存型の話者照合は、5 発話程度の正解テンプレートを登録しておき、入力されたときには、テンプレートと入力音声の特徴を比較し、予め決めたしきい値に基づいて照合を行う。本研究では、登録音声発話全ての特徴量で学習した GMM を用いてテンプレートを構成することで、正解話者の話者内変動を考慮したテンプレート表現ができ、また GMM で多くの発話の情報を考慮できるという利点に着目し、GMM Posteriorgram をテンプレート表現として導入する。一般的な話者照合の構成に加えて、音声が入力されたときの車内雑音抑圧処理を加える。特徴抽出を行う前の音声データに対して雑音抑圧処理を施すことで、雑音の影響を低減させた特徴抽出ができることをねらう。

### 3.2 個人性情報に関する特徴量

音声から得られる情報は、音韻性情報と個人性情報の二つがあり、話者照合では個人性情報が使用される。さらに個人性情報は先天的個人性情報と後天的個人性情報に分けられる。先天的個人性情報は、話者の身体的特徴に起因する個人性であり、固有の声道に関する情報として、スペクトル包絡の外形などに差が現れる。後天的個人性情報は、話者が母語獲得過程またはそれ以降に身に付けた行動的な特徴であり、発話速度や地方方言独自のアクセントとして基本周波数に現れる [10]。

話者照合に用いる特徴量としては、人間が真似することが困難な、先天的個人性が効果的であるとされている。実際に話者照合の研究で、人間の声道特性の特徴量である MFCC が先天的個人性情報として用いられている [10]。アクセントの上下にもある程度個人性が現れることから、基本周波数に関する特徴量も有効的であるとされている。他にも、スペクトル包絡における一つ目のピークである第一フォルマントと二つ目のピークである第二フォルマントなどを導入している研究もある [11]。

### 3.3 MFCC

MFCC とは、人間の感覚尺度を考慮したフィルタバンク分析により調音フィルタを表現する特徴量である [10]。メル周波数軸において等間隔に配置される三角窓により、帯域ごとのパワーを求め、それらの対数値の離散コサイン変換により求められる。本研究では、短い時間ごとに MFCC を得るために、以上の処理を、フレーム長 20ms、フレームシフト 10ms で抽出を行い、13 次の MFCC と  $\Delta$ MFCC、 $\Delta\Delta$ MFCC を特徴量として使用する。

### 3.4 GMM Posteriorgram

発話テンプレートを表現する特徴ベクトルとして、GMM Posteriorgram を導入する [7]。GMM posteriorgram とは、異なる GMM コンポーネントによって生成される、出力確率を要素に持つ特徴ベクトルの確率表現である。GMM posteriorgram の生成には、以下の処理を施す。p 個のフレームを含む、長さ  $\mathbf{S}$  の音声データを考える。 $\mathbf{S}$  は、

$$\mathbf{S} = [s_1, s_2, \dots, s_p] \quad (1)$$

と表せる。i 番目のフレームに対応する GMM posterior vector は、

$$\mathbf{x}_i = [P(\lambda_1|s_i), P(\lambda_2|s_i), \dots, P(\lambda_N|s_i)]^T \quad (2)$$

で与えられる。ここで、 $\lambda_j$  は j 番目のガウシアンコンポーネントを表し、N はガウシアンコンポーネントの総数を表す。GMM posteriorgram は、1 発話中の各フレームの posterior vector を連結させ、

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p] \quad (3)$$

とする。混合数 3 の GMM を用いた実際の登録話者テンプレートの例と詐称者テンプレートの例を図 1 に示す。

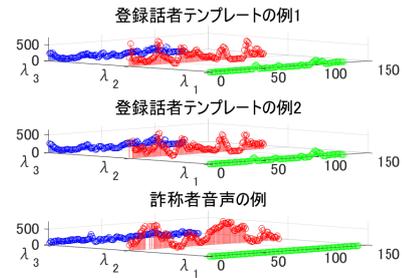


図 1. テンプレートの例

同じ話者のテンプレートである上の二つは似たテンプレート表現になっているのがわかる。

GMM Posteriorgram の利点は、登録話者の音声のみで GMM を構成することで、その話者の特徴空間でのテンプレート表現を可能にし、GMM でカバーされた範囲内の特徴 (主に登録話者の特徴) であれば変動を抑圧できることである。しかし一方で、登録発話外 (大きな環境変化や雑音など) のような、カバーされない音声の特徴に関しては脆弱である。この GMM Posteriorgram を用いることで発声時期の違いによる特徴変動を抑圧することを狙う。

### 3.5 照合

照合では、登録音声 5 発話の GMM Posteriorgram を用いて決定したしきい値と入力音声・登録音声間の距離を比較することで照合を行う。事前に決定するしきい値を式 (4) で示す [12]。

$$T = \frac{\mu_1\sigma_2\omega_2 + \mu_2\sigma_1\omega_1}{\sigma_1\omega_1 + \sigma_2\omega_2} \quad (4)$$

$\mu_1, \sigma_1$  は正解話者音声内の距離の平均と標準偏差、 $\mu_2, \sigma_2$  は正解話者・詐称者音声間の距離の平均と標準偏差、 $\omega_1, \omega_2$  は足して 1 になるような重みであり、しきい値を調整する際に変更する値である。正解話者音声内の距離とは、登録 5 発話それぞれの距離である。また、正解話者・詐称者音声間の距離とは、登録 5 発話と詐称者として用意した音声とのそれぞれの距離である。申告話者テンプレートと入力音声の平均距離が式 (4) のしきい値よりも小さいときに申告話者、大きいときに詐称話者と決定する。

### 3.6 雑音対策

一般的な話者照合システムに加えて、運転中に生じるエンジン音や路面からの音、風切り音など [13] の雑音の影響が重大であり、対策をする必要がある。一般的に研究で用いられる雑音対策としては、主に以下の 3 点がある。

- マイクによる指向性
- 信号処理による雑音除去
- 雑音環境に対する適応処理

マイクによる指向性とは、マイクアレイを用いて、欲しい方向からの音のみを収集することで雑音を收音しない手法である。信号処理による雑音除去とは、フィルタリングによって雑音成分を除去する手法である。雑音環境に対する適応処理とは、モデルを雑音環境に適応させることで、雑音の加わった音声でもモデルとの比較が有効になる手法である。この中でも本研究では、周波数領域で雑音の特徴がみられる場合の音声入力時における音声へのフィルタリング処理を導入する。マイクによる指向性は別の分野であるため、対象としない。また、申告話者音声テンプレートの登録は車内ではないことが想定されるので、雑音環境に対する適応処理も対象としない。

本研究では 3.6.1 項と 3.6.2 項で述べる二つの処理を施す [14]。

#### 3.6.1 低周波数成分の除去

車内雑音が低周波成分に集中するという特徴から、100Hz 以下の低周波成分を除去する。100Hz という数字は、スペクトル領域での個人の特徴をつぶさない程度で決定した値である。

### 3.6.2 定常雑音成分の除去

定常雑音除去の手法として SS 法を実装する。SS 法とは、雑音が重畳した音声のスペクトルから雑音のスペクトルの推定値を周波数領域で減算する、スペクトル減算に基づく雑音抑圧法である。

### 3.7 テンプレート更新

始めの登録テンプレートは 5 発話であるが、実際に使用する中で申告話者と正しく照合した音声新しくテンプレートとして加え、GMM Posteriorgram を再構成する。3.5 節で述べた、しきい値の計算と照合時に使用する正解話者音声内距離と正解話者・詐称者音声間距離の計算には、追加した音声も含めすべてのテンプレートで計算する。3.4 節で示した GMM Posteriorgram の特徴に基づき、テンプレートを更新することで、雑音抑圧処理による雑音も含めた学習データで登録話者 GMM を構成できるため、雑音抑圧処理による雑音にも適応したテンプレート表現を可能にする。

## 4 評価

一般に、話者照合の研究の評価には等誤り率 (EER) が用いられる。等誤り率とは、ふたつのトレードオフの関係を持つ誤り率が等しくなるしきい値での誤り率のことをいう。本研究でのトレードオフのグラフと等誤り率の関係を図 2 に示す。

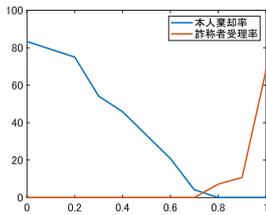


図 2. 誤り率と EER(交点)

図 2 の二つの誤り率の交点が EER である。本研究では、雑音抑圧処理の評価、発声時期の違いによる話者内特徴変動の影響の評価ともに、この EER となるしきい値 ( $\omega_1 = 0.7$ ) で実験を行う。

雑音対策に関する評価としては、登録・照合ともに同じ時期に発話したデータを用いて精度評価を行う。また、長期的話者内特徴変動対策に関する評価としては、以下の二つの実験を行う。

- テンプレートの更新をせずに初期登録から期間を空けて録音した発話で実験
- テンプレートの更新処理を加えて初期登録から期間を空けて録音した発話で実験

上記の二つの実験で得られた精度を比較し、評価を行う。

#### 4.1 使用するデータ

本研究は、使用者が運転中におけるスマートフォンの使用を想定した話者照合であるので、評価データに使用するデータには運転中の走行音が加わった音声データを使用することを考えている。また、話者照合のみの評価だけでなく、走行音の雑音のレベルにも着目した評価をする必要がある。先行研究 [15] では、擬似雑音環境下における話者照合システムの評価に用いる音声として、目的の音声信号と雑音レベルの比 (SNR) が指定した SNR になるような雑音を付与し、いくつかの雑音レベルで実験・評価を行っている。

実際の車内で録音したデータを用いた実験を行ったが、直線の道路走行時にもふらつくなど、危険であり、多くの音声を録音するのは現実的ではないため、本研究においても、雑音を付与することで、多くのデータを集める。使用する雑音は、実際に車内で録音した走行音を用いる。走行音のみを用いるため、録音時はカーステレオや話し声がなく、エアコンなども消した環境で、運転席と助手席の間にマイクを設置して録音を行った。雑音抑圧処理の評価データは、ほぼ同時期に録音された音声を用いて SNR を変えた雑音を付与したデータとして行う。音声の録音に

はコンデンサーマイクを用いた。

話者照合システムの問題点のうちの、時期の違いによる特徴変動を考慮するために、1 週間に 1 日、朝昼夜の録音を 1 年間行った音声コーパスである、AWA 長期間収録音声コーパス (AWA-LTR) を用いて評価を行う。雑音抑圧処理に関する評価、長期的特徴変動の影響の評価、どちらも詐称者音声として 14 音声を使用する。この 14 音声は、年齢や性別に偏りがないよう意識し、録音を行った音声である。

#### 4.2 実験結果

そもそもの精度として、雑音を付与したテンプレートを用いている従来研究 [15] と比較するために、本研究でも雑音を付与したテンプレートを用いて EER での予備実験を行った。その結果、SNR 10dB での誤り率は 4.2 % となり、誤り率 20 % の従来研究よりも良い結果であることが確認できた。

雑音抑圧処理の結果として、SNR ごとの本人棄却率 (FRR) と詐称者受率率 (FAR) を表 1 に示す。

表 1. 雑音抑圧処理の結果 (%)

SNR		抑圧あり	抑圧なし
ノイズなし	FRR	0	0
	FAR	0	0
10dB	FRR	16.7	83.3
	FAR	0	0
0dB	FRR	100	100
	FAR	0	0

SNR 10dB での、抑圧ありと抑圧なしの FRR を比較すると、抑圧ありの方が FRR が改善される結果となり、雑音抑圧処理の有効性が示される結果となった。SNR 0dB においては、本研究で導入した照合の枠組みと雑音抑圧処理では、精度の改善が見られない結果となった。SNR 10dB で誤り率が改善した一方で、表 1 より、ノイズなしの誤り率より SNR 10dB での誤り率が高くなっているのがわかる。これはスペクトルサブトラクションによる雑音成分が影響したと考えられる。ここで、雑音抑圧処理を施した、受理された音声をテンプレートに加え 6 テンプレートとし、SNR 10dB でもう一度実験を行った結果を表 2 に示す。

表 2. 雑音抑圧後音声による更新後の結果 (%)

SNR		抑圧あり
10dB	FRR	8.3
	FAR	3.6

表 1 の雑音抑圧ありの結果と比べると、FAR がやや上昇したが、FRR が 8.4 % 改善する結果となり、雑音環境下で照合した音声をテンプレートに加えることで、誤り率を減少させる結果が得られた。これは、3.4 節で示した GMM Posteriorgram の脆弱性に対して、雑音環境下での発話をテンプレートに追加し、GMM でカバーする範囲を雑音環境に適応させたことが要因として挙げられる。

#### 4.2.1 初期登録発話数の考察

初期テンプレート数の違いによる誤り率への影響を調べるために、テンプレートとして 5 発話より少ない 3 発話を登録し、雑音抑圧処理の評価実験と更新処理の評価実験を行った。初期登録発話の数について考察するために、混合数は 5 発話での実験と同じ 3 混合とした。雑音抑圧処理の実験結果を以下の表 3 に示す。

表 3. 雑音抑圧処理の結果 (%)

SNR		抑圧あり	抑圧なし
ノイズなし	FRR	0	0
	FAR	7.1	0
10dB	FRR	29.2	66.7
	FAR	3.6	3.6
0dB	FRR	100	100
	FAR	0	0

SNR 10dB の結果を見ると、5 発話より少ない 3 発話を用いて GMM を構成し、テンプレートを構成したことで、5 発話での結果より、FRR が 12.5 % 悪い結果となった。

SNR 10dB で更新処理を施して実験をした結果を表 4 に示す。

表 4. 雑音抑圧後音声による更新後の結果 (%)

SNR		抑圧あり
10dB	FRR	4.2
	FAR	3.6

初期登録 5 発話での実験では FRR が 8.4 % 改善しているのに対し、表 3, 表 4 を見ると、3 発話では更新処理によって FRR が 25.0 % 改善しているのがわかる。これは、初期登録 5 発話で 1 発話をテンプレートに追加したときの影響に対し、初期登録 3 発話で 1 発話をテンプレートに追加したときの影響の方が大きいことが要因として考えられる。この実験結果から、雑音抑圧処理に関しては、初期登録 5 発話の方が、より有効に働く結果が得られるが、更新処理に関しては、1 発話をテンプレートに追加する影響がより大きい、初期登録 3 発話の方が、誤り率の改善が見られるという結果が得られた。実際にスマートフォンなどに実装されることを考えて初期登録発話数の違いに言及すると、3 発話の場合には、更新前の FRR が低いいため車内では使いやすく感じる事が考えられるため、5 発話を初期登録とする方が、運転者によってより安全に使用できると考える。

テンプレートとして 3 発話、5 発話としたどちらの実験においても、実行することの若干の変動が見られた。本研究ではすべての実験において、4 桁連続数字の読み上げ音声を使用した。この変動を改善するために、適切な発話内容や発話長についての検討をする必要がある。また、この検討後の混合数の決定方法として、ベイズ情報量基準などを用いることで、変動が少なくなった後の混合数の検討でも、モデルとしての有効性を検討できると考える。

#### 4.2.2 長期的特徴変動に関する実験結果と考察

テンプレートの更新処理をせずに、登録直近の 2 ヶ月で発話された音声で実験した結果、FRR は表 5 より 0 % となった。登録から 10 ヶ月の期間を空けて発話された音声で実験した結果、FRR は 0 % という結果が得られ、本研究で実装した手法では時期差による精度への影響がなかった。MFCC を直接テンプレートとして用いた本人棄却率と、特徴量 MFCC の GMM Posteriorgram (GP) を用いた本人棄却率を表 5 に示す。なお、どちらも EER となるしきい値 (MFCC, GP, それぞれ  $\omega_1 = 0.7, 0.85$ ) である。

表 5. 特徴変動に関する結果のテンプレートごとの比較 (%)

FRR	DTW-MFCC	DTW-GP
直近 2 ヶ月	4.2	0
10 ヶ月後	12.5	0

表 5 を見ると、GMM Posteriorgram をテンプレート表現とした場合は精度に変化がないのに対し、MFCC を直接テンプレートとして用いた場合は、登録から期間を空けることで精度が悪化している。これは、テンプレート表現として話者内変動を考慮できる GMM Posteriorgram を導入したことにより、発声時期の違いによる精度への影響はなかったと言える。

## 5 あとがき

本研究では、走行中の自動車内での話者照合として、走行車内での問題 (雑音)、話者照合の問題 (特徴変動) の二つを問題視し、評価を行った。結果、雑音抑圧処理とテンプレート更新によって、SNR 10dB において合計で FRR 75.0 % 誤り率が改善した。雑音抑圧処理のみでは、FRR 66.6 % 改善する結果が得られた。実際の走行車内で録音した音声を用いた実験でも FRR 57.1 % 改善する結果が得られ、3 章で示した 3 つ目の問題点である、ロンバード効果はあまり精度に影響しないということが言える。テンプレート更新処理では、雑音抑圧処理の誤り

率からさらに、FRR 8.4 % 改善する結果が得られた。また、テンプレートを 3 発話とした実験も行ったことで、初期登録発話数の違いによる、雑音抑圧処理とテンプレート更新処理の有効性について検討することができた。

長期的特徴変動に関する実験の結果は、登録から 10 ヶ月空いた発話では精度への影響が出ない結果となった。原因として、実際に MFCC のみをテンプレートとして用いた話者照合と比較し、GMM Posteriorgram が長期的変動にも有効であることが確認できた。

本研究で導入した雑音対策手法は定常雑音にのみ有効な手法である。マイクで取音された走行中の車内雑音が定常であることを想定したデータで実験を行ったが、実際はカーステレオやラジオを流していたり、複数人で乗車している場合は話し声が非定常的に発生することが考えられる。実用レベルではこのような音声も含めた環境を想定し対策すべきである。車内は席が固定であり、マイクの設置位置によってはすべての席に対して違う角度をつけられるため、3.6 節、雑音対策の「マイクによる指向性」でも示したように、マイクアレイを用いて欲しい方向からの音声を集めることで対策ができると考えている。

本研究では音声を周波数領域で扱う上で代表的な、MFCC のみを用いた。誤り率をさらに改善させるための一つの手段として特徴量の改良が挙げられる。

## 参考文献

- [1] 市野将嗣, 吉浦裕, “テキスト指定型話者認識のための音声個人認証方式”, 電子情報通信学会論文誌 A, Vol.J98-A, No.12, pp.632-645, 2015.
- [2] 古井貞熙, “話者認識”, 日本音響学会誌, vol.37, no.5, 1981.
- [3] 塩田さやか, “音声を用いた生体認証技術, 話者照合ソフトウェア入門”, システム制御情報学会, vol.62, no.2, pp.1-6, 2018.
- [4] Siri Team, <https://machinelearning.apple.com/2018/04/16/personalized-hey-siri.html>, “Personalized Hey Siri”
- [5] 岩野公司, 堀畑拓斗, “プロの物真似タレントの声真似が話者照合に与える影響と音響特徴の分析”, 信学技報 IEICE Technical Report, 2017, pp.55-60.
- [6] Medha Pandit, Josef Kittler, “Feature Selection for a DTW-based Speaker Verification system”, IEEE International Conference on Acoustics, Speech and Signal Processing, May 1998.
- [7] Sarfaraz Jelil, Rohan Kumar Das, R.Sinha, S.R.Mahadeva Prasanna, “Speaker Verification Using Gaussian Posteriorgrams on Fixed Phrase Short Utterances” INTERSPEECH-2015, pp. 1042-1046, Oct.2015.
- [8] Douglas A. Reynolds, Thomas F. Quatieri, Robert B. Dunn, “Speaker Verification Using Adapted Gaussian Mixture Models”, Digital Signal Processing 10(2000), 2000, pp.19-41.
- [9] 古井貞熙, “デジタル音声処理”, デジタルテクノロジーシリーズ 6, Sep.1985.
- [10] 王 龍標, 西田昌史, 柘植寛, 網野加苗, “話者認識におけるロバストネス”, 日本音響学会誌, vol. 69, no. 7, pp. 357-364, 2013.
- [11] Shi-Huang Chen, Yu-Ren Luo, Rodrigo Capobianco Guido, “Speaker Verification Using Line Spectrum Frequency, Formant, and Support Vector Machine”, 2009 11th IEEE International Symposium on Multimedia, Dec. 2009, pp.562-566.
- [12] Suman Paul, Tushar Kanti Das, Partha Saha, Rabul Hussain, Ujwala Baruah, “Comparative Analysis of Two Different System’s Framework for Text Dependent Speaker Verification” ICCPCT-2015, march.2015.
- [13] 星野博之, 小沢義彦, “車内音を構成する音の要素とその評価”, 豊田中央研究所 RD レビュー, vol. 30, no. 3, Sep. 1995.
- [14] 武田一哉, “自動車の中での音声認識”, IPSJ Magazine, vol.45, no.10, Oct.2004.
- [15] 鎌田敏明, 峯松信明, 長内隆, 時苗久則, 谷本益巳, “雑音環境下における話者照合”, 信学技報 TECHNICAL REPORT IEICE, pp.55-60, March. 2007.