

## 話者照合のための話者性を考慮した音素情報に基づく ポップノイズ検出法を用いたテキスト依存型声の生体検知

望月 紫穂野\* 塩田 さやか\* 貴家 仁志\*

† 首都大学東京 システムデザイン研究科  
E-mail: †mochizuki-shihono@ed.tmu.ac.jp

**あらまし** 本稿では、話者性を考慮した音素情報に基づくポップノイズ検出法を用いた、テキスト依存型の声の生体検知を提案する。近年、話者照合システムが普及しつつある一方で、登録話者の録音した声をスピーカー再生するなりすまし攻撃によってその認証精度が大幅に低下してしまうことが報告されている。このなりすまし攻撃に対する根本的な解決策の一つとして、入力音声が入力によって実際に発声されたものか否かを識別する声の生体検知が提案されている。入力音声からポップノイズを検出する方法は、その実現法の一つであり、なりすまし攻撃の検出に対して有用であることが報告されている。しかし、なりすまし攻撃からもポップノイズを誤検出してしまうという問題があり、著者らはその精度向上のために、ポップノイズ検出後にポップノイズ区間にかかる音素を考慮して声の生体検知を行う方法（音素情報に基づくポップノイズ検出法）を提案した。先行研究では、話者毎の発話スタイルの違いに着目した話者依存の音素リストや、ポップノイズの発生頻度を考慮したプロンプト文の使用を前提としてきた。一方、近年普及しつつあるスマートフォンや銀行等の話者認証システムは発話内容が固定である場合が多い。そこで本研究では、話者性を考慮した音素情報を用いた声の生体検知に対してさらにテキスト依存の枠組みを導入することを提案し、生体検知実験および話者照合実験による性能評価と考察について報告する。

**キーワード** 話者照合, ポップノイズ検出, 声の生体検知, 音素情報

### Text-dependent voice liveness detection based on pop-noise detector considering speaker-dependent phoneme information for speaker verification

Shihono MOCHIZUKI\*, Sayaka SHIOTA\*, and Hitoshi KIYA\*

† Department of Information and Communication Systems, Tokyo Metropolitan University  
E-mail: †mochizuki-shihono@ed.tmu.ac.jp

**Abstract** This paper proposes a pop-noise (PN) detection method considering speaker-dependent phoneme information for text-dependent voice liveness detection (VLD) systems. Recently, various countermeasures against spoofing attacks have been reported, and PN detection-based VLD frameworks have been proposed as well. The VLD frameworks identify whether an input sample is a genuine speech or a replayed one. However, since spoofing attacks are sometimes regarded as genuine speeches due to the false detection of PN periods, PN detection methods considering speaker-dependent phoneme information have also been proposed for text-independent speaker verification. On the other hand, text-dependent speaker verification systems are usually used for smartphone security and banking systems. Therefore, in this paper, the text-dependent condition is adapted to the proposed PN detection method considering speaker-dependent phoneme information to improve the robustness. To evaluate the effectiveness of the proposed method, VLD and speaker verification experiments are performed. These results show that the proposed method has higher reliability and usability than conventional methods.

**Key words** Speaker verification, Pop-noise detection, voice liveness detection, phoneme information

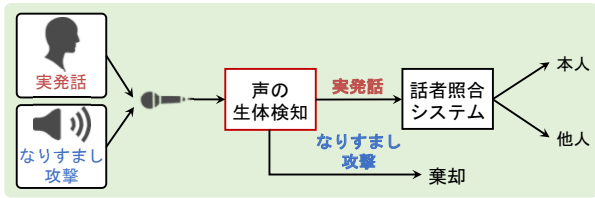


図1 声の生体検知と話者照合システムのフロー

## 1. はじめに

近年、声を用いた生体認証システムである話者照合はネットバンキングや携帯電話などのセキュリティシステムとして、またスマートスピーカーなどの音声対話システムにおけるユーザー個別サービスを実現する技術としての導入が始まってきている。一方で、話者照合システムに登録話者の録音した声をスピーカー再生するなりすまし攻撃によってその認証精度が大幅に低下してしまうことが報告されている [1]。そのため、話者照合システムの課題として精度向上だけでなく、スピーカー再生によるなりすまし攻撃に対する頑健性向上も重要な課題となり、国内外で活発に研究が行われている [2]。スピーカー再生によるなりすまし攻撃に対する根本的な解決策の一つとして、声の生体検知という入力音声が入力によって実際に発声されたものか否かを識別する枠組みが提案された [3]。声の生体検知の実現手法の1つとして、入力音声からポップノイズを検出する方法が有用であることが報告されている。ここでポップノイズとはマイク内部に息や風が入りこむことにより変則的に振動板が揺れるために発生してしまうノイズのことを指す [4]。しかし、なりすまし攻撃においても背景雑音などをポップノイズとして誤検出してしまう場合があり、ポップノイズ検出だけでは精度が十分とはいえなかった。そこで、ポップノイズの発生頻度と音素には依存関係があることを利用し、ポップノイズ検出後にポップノイズ区間にかかる音素を考慮して声の生体検知を行う音素情報を考慮したポップノイズ検出法を提案し、なりすまし攻撃に対するポップノイズ検出の頑健性が向上することを報告してきた [5, 6]。

音素情報に基づくポップノイズ検出法では、話者毎の発話スタイルの違いに着目した話者依存の音素リスト [5] の使用やポップノイズの発生頻度を考慮したプロンプト文 [6] の使用を前提としてきた。一方で、近年普及しつつあるスマートスピーカーやスマートフォン、銀行等の話者認証システムは発話内容が固定であることを前提としている場合が多い。そこで本研究では、テキスト依存型話者照合との統合を想定した話者性を考慮した音素情報に基づくポップノイズ検出法を提案し、生体検知実験および話者照合実験による性能評価と考察について報告する。

## 2. 話者照合のための声の生体検知 [3]

### 2.1 ポップノイズ情報を用いた声の生体検知

声の生体検知とは入力音声が入力によって実際に発声されたものか否かを識別する枠組みのことを指す。特に、図1に示すような話者照合と組み合わせることを想定している。

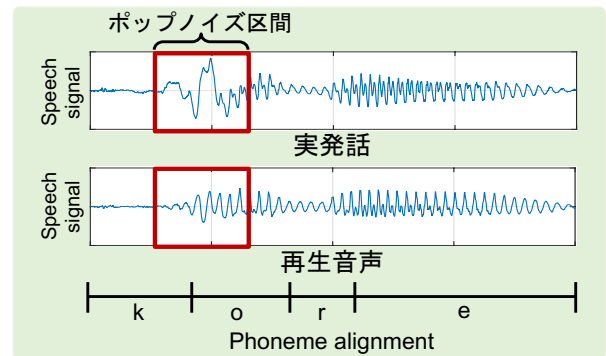


図2 ポップノイズが発生した実発話（上）および収録した音声を再生した音声（下）の波形

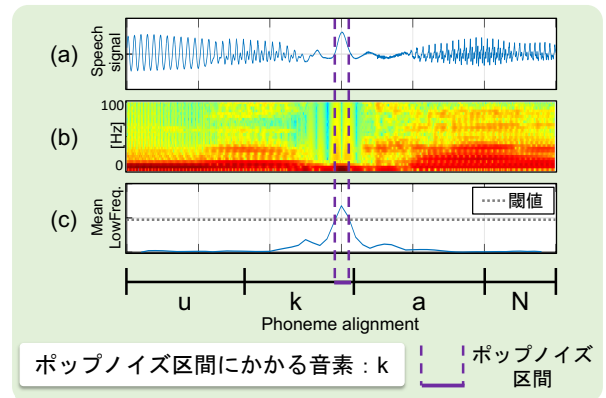


図3 ポップノイズ区間にかかる音素の抽出

図1の例では声の生体検知部で入力された音声信号が実際に人間から発せられたものか否かを識別し、生体であると判定された場合のみ後段の話者照合に入力信号を渡すというフローになっている。これまでに声の生体検知の実現手法として、入力音声にポップノイズが発生しているかを検出する方法が有用であると報告されている。ここでポップノイズとはマイク内部に息や風が入りこむことにより変則的に振動板が揺れるために発生してしまうノイズのことを指す。図2は実際に人間が発話した音声と、収録した実発話をスピーカー再生した音声の波形をそれぞれ示している。図から、実発話で発生したポップノイズによる波形の歪みが再生音声中には発生していないことがわかる。これは人間が発声する際には呼吸を用いるが、スピーカーは振動板による空気の振動で音声を表現しているため、ポップノイズが発生しているような音を再生することはできてもポップノイズの現象を起こすことはできないためである。このことから、発話中のポップノイズを検出することが声の生体検知を実現する手段として適切であると考えられる。

### 2.2 ポップノイズ検出法

本稿では、入力音声のポップノイズを検出するためにシングルチャンネルポップノイズ検出法 [3] を用いた。ポップノイズは発話内で突発的に起こるノイズのため、局所的に強いエネルギー変動を持つ性質がある。そのため、シングルチャンネルポップノイズ検出法では急激なエネルギー変動を検出している。手順としてはまず、短時間フーリエ変換（分析窓サイズ  $N$ 、窓シフト幅  $N/4$ ）を行い、入力音声の周波数分解を行う（図3 (b)）。

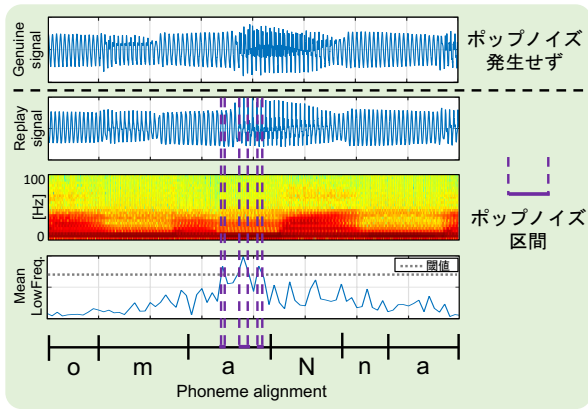


図 4 実発話ではポップノイズが検出されていないにも関わらず再生音声ではポップノイズが誤検出された例

$$X(v, \omega) = \int_{-\infty}^{\infty} x(t)w(t-v)e^{-j\omega t} dt \quad (1)$$

ただし、 $x$  は入力信号、 $w$  は窓関数、 $v$  は時刻、 $\omega$  は角周波数を示す。次にフレーム毎に低周波領域  $[0, F]$  Hz のパワースペクトルの平均を求める (図 3 (c))。この平均が低周波成分のエネルギーの推移を表し、フレーム間でのエネルギー変動が閾値より大きくなる区間をポップノイズが生じている区間として検出する (図 3)。シングルチャンネルポップノイズ検出法は 1 本のマイクで実現可能であり、導入コストが低く、また話者照合システムとの親和性も高いことが利点としてあげられる。

### 3. 音素情報に基づくポップノイズ検出法による声の生体検知 [6]

#### 3.1 ポップノイズと音素の依存性

2.2 節で述べたポップノイズ検出法における誤検出の例を図 4 に示す。図の例では、実発話においてポップノイズが検出されていないにも関わらず、再生音声からポップノイズ区間を検出してしまっている。ポップノイズ検出法はあくまでも低周波領域におけるフレーム毎パワーの急激な変動を検出する手法である。そのため、スピーカー再生時に背景雑音などが同時に収録された場合や、不適切なポップノイズ検出閾値の設定等によりポップノイズを誤検出してしまうためである。そこでポップノイズの誤検出を減らし、ポップノイズ検出精度を向上させる手法について考える。

ポップノイズの発生原理と人の発声器官の仕組みから、ポップノイズを発生させやすい音素と発生させにくい音素があると考えられる。そこでポップノイズ検出後にポップノイズ区間にかかる音素を考慮することで、ポップノイズ検出がより高精度になると期待できる。

#### 3.2 ポップノイズ区間にかかる音素の選択

ポップノイズを発生させやすい音素と発生させにくい音素を選択するための手順は以下に示す通りである。

- 1: 音声データに対して音声認識を行い、音素アライメントを取得。
- 2: 音声データに対してシングルチャンネルポップノイズ検

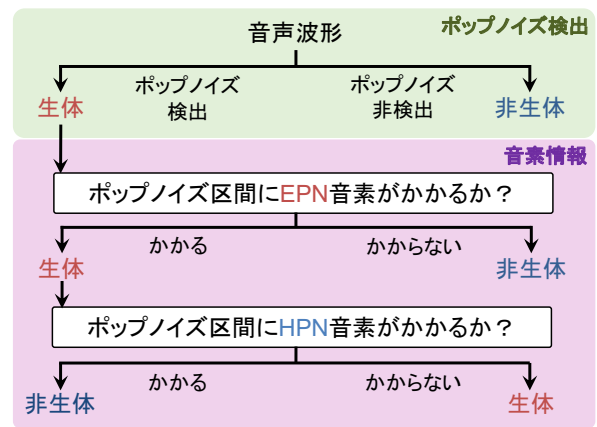


図 5 音素情報を考慮したポップノイズ検出法のフロー

出法を用い、ポップノイズ区間のアライメントを取得。

3: 手順 1, 2 で得られたアライメント情報を比較して、ポップノイズを発生させやすい音素を選択 (図 3)。

4: ポップノイズ検出の閾値を十分に下げた際にもポップノイズとして検出されない音素を選択。

ここで、手順 3 で選択された音素をポップノイズを発生させやすい (Easily caused pop-noise; EPN) 音素、手順 4 で選択された音素をポップノイズを発生させにくい HPN (Hardly caused pop-noise; HPN) 音素とする。

#### 3.3 音素情報に基づくポップノイズ検出法

3.2 節の手順により得られた EPN 音素および HPN 音素を用いた音素情報に基づくポップノイズ検出法について説明する。フローを図 5 に示す。はじめにシングルチャンネルポップノイズ検出法を用いて入力音声のポップノイズを検出する。入力音声にポップノイズが発生しているならばその音声を生体による音声として受理し、発生していないならばなりすまし攻撃として棄却する。次に検出されたポップノイズ区間に EPN 音素がかかるかを判定し、かかる場合にはそれは人間の発声によって発生したポップノイズと想定されるため生体として受理し、逆に EPN 音素がかからないならば、なりすまし攻撃と想定されるため非生体として棄却する。最後に EPN 音素情報で生体として受理された音声のポップノイズ区間に、HPN 音素がかかるかどうかで生体検知を行う。実発話の場合、HPN 音素の場所ではポップノイズが非常に発生しづらいため、ポップノイズ区間に HPN 音素がかかることは人間の発声としては不自然である。そこでポップノイズ区間に HPN 音素がかかる場合はなりすまし攻撃として棄却し、HPN 音素がかからない場合は生体による音声として受理する。HPN 音素まで確認するのは、詐称者が音声を再生中に故意にポップノイズを発生させた場合にも適切な区間でポップノイズが発生していないと受理できないようにするためである。

#### 3.4 テキストを考慮した音素情報の選択

これまで EPN 音素および HPN 音素リストは、複数の話者から調査した一般的なポップノイズ区間にかかる音素の出現傾向から作成してきた。また、生体検知性能の改善のために話者毎に音素リストを設定する方法やポップノイズの発生頻度を考

表 1 ポップノイズ検出条件

データベース	VLD,VLD2
周波数帯域	[0,50] Hz
分析フレーム長 (N)	20 msec
フレームシフト (N/4)	5 msec

表 2 音素リスト作成データ

データベース		VLD (2 セッション)
共通の音素リスト (従来法)	話者数	17 名
	発話数	30 文/話者
	サンプル数	実発話 510 文
話者毎の音素リスト (従来法)	話者数	8 名
	発話数	120 文/話者
	サンプル数	実発話 960 文
テキスト毎の音素リスト	話者数	17 名
	発話数	20 文/話者
	サンプル数	実発話 340 文
話者毎かつテキスト毎の音素リスト	話者数	8 名
	発話数	120 文/話者
	サンプル数	実発話 960 文

慮したプロンプト文の使用について提案してきた。一方で、近年スマートスピーカーやネットバンキングなどに用いられる話者照合システムでは発話内容が固定である場合が多い。そこでテキスト依存型話者照合との統合を想定した音素リストの設定についても考える。テキスト依存型話者照合では、事前に入力文章が固定されているため、プロンプト文のようにポップノイズの発生頻度などを考慮できない一方で、フレーズ毎に音素リストを設定することが可能である。そこで、文章固有の音素リストを設定することにより提案法のポップノイズ検出性能が向上すると考えられる。また、本報告ではさらに話者性を考慮した文章毎の音素リストを作成することも行った。これは話者毎に音素リストを設定することで生体検知性能が改善したことを踏まえ、話者性およびテキストの両情報をを用いることで生体検知精度がより向上すると期待されるためである。

## 4. 評価実験

### 4.1 実験条件

EPN 音素および HPN 音素リストを話者毎およびテキスト毎に設定することの有効性を確認するため、VLD データベース [3] および VLD2 データベース [6] を用いて生体検知実験および話者照合との統合実験を行った。ここで、VLD および VLD2 データベースはポップノイズ検出法のために収録された日本語データベースで、マイク (AKG P170) に風防カバーを装着しないで収録した音声データが収録されている。またマイクと口およびスピーカーとの距離は約 7cm となるようにした。収録内容は女性話者 15 名、各話者 100 発話の合計 1500 文章の実発話およびその実発話をスピーカー (ELECOM LBT-SPP300) で再生し収録した再生音声 1500 文章となっている。VLD および VLD2 データベースの発話内容は同じである。表 1 にポップノイズ検出条件を示す。ポップノイズ検出に用いる閾値については、実発話を全て受理する最大値で固定した。音素アライメントの抽

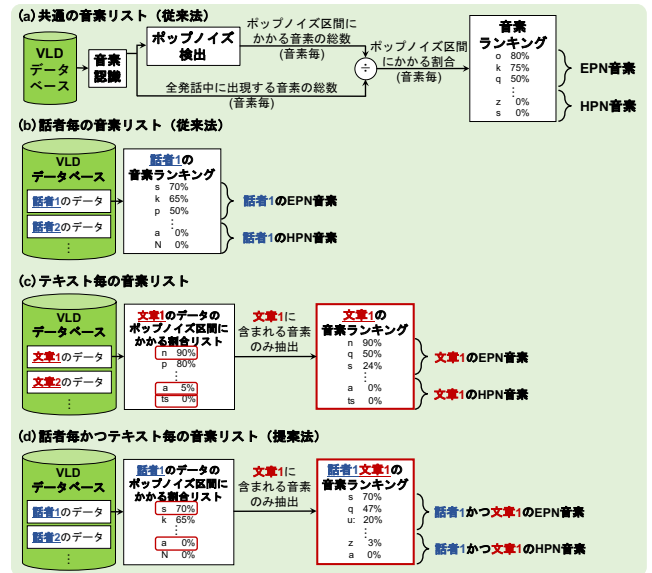


図 6 音素リスト毎の作成方法

出には汎用大語彙連続音声認識エンジン Julius (Ver.4.3.1) [7] およびディクテーションキット (Ver.4.4, DNN-HMM 版) の音響モデルと言語モデルを使用した。

提案法および比較手法における音素リストの作成手順は図 6 および以下に示す。また、各音素リストを作成するために用いたデータの条件は表 2 にまとめた。

(a) **共通の音素リスト**： データベース全体からポップノイズ検出によって検出したポップノイズ区間にかかる音素とその音素の全発話中に出現する総数からその音素がポップノイズ区間にかかる割合を求め、ランキングを作成した。ランキング作成手順は以下も同様である。EPN 音素にはランキング上位 11 個の音素を選択し、HPN 音素にはランキング下位 3 個の音素を選択した。

(b) **話者毎の音素リスト**： 表 2 の音素リスト作成データに対し、話者毎にランキングを作成した。EPN 音素にはランキング上位 11 個の音素を選択し、HPN 音素にはランキング最下位の音素を選択した。

(c) **テキスト毎の音素リスト**： 表 2 の音素リスト作成データに対し、テキスト毎にポップノイズ区間にかかる音素の割合を算出した。その後テキストに含まれる音素のみを抜粋しテキスト毎のランキングを作成した。EPN 音素にはランキング上位 11 個の音素を選択し、HPN 音素にはランキング最下位の音素を選択した。

(d) **話者毎かつテキスト毎の音素リスト (提案法)**： 表 2 の音素リスト作成データに対し、話者毎にポップノイズ区間にかかる音素の割合を算出した。その後テキストに含まれる音素のみを抜粋し話者毎かつテキスト毎のランキングを作成した。EPN 音素にはランキング上位 11 個の音素を選択し、HPN 音素にはランキング最下位の音素を選択した。ただし、リストによっては EPN 音素が 11 個以下となる場合もあった。

実際に用いた音素リストの例を表 3 に示す。図 7 に実験フローを示す。実験に用いた各検出手法の詳細は以下の通りである：

表 3 使用した音素リスト（話者毎，テキスト毎および話者毎かつテキスト毎の音素リストは一例を掲載）

音素リスト種類	音素	共通音素リストと同じ音素	共通音素リストと違う音素
共通（従来法）	EPN	b,e:,hy,k,ky,o,o:,s,sh,t,u:	—
	HPN	i:,m,ry	—
話者毎（従来法）	話者 1	EPN	b,hy,o: a:,d,gy,n,N,py,ry,z
		HPN	— i
	話者 2	EPN	e:,ky,o: by,my,n,p,py,q,y,z
		HPN	— i
テキスト毎	文章 1	EPN	k,o,o:,t d,e,i,j,m,N,sh
		HPN	— a
	文章 2	EPN	o,sh,t,u: a:,N,py,r,t,ts,u
		HPN	m —
話者毎かつテキスト毎	話者 1 文章 1	EPN	o,o:,sh a,d,e,j,m,n,N,py
		HPN	— e:
	話者 2 文章 1	EPN	k,e:,o,o:,s,t a,i,n,N,py
		HPN	d —

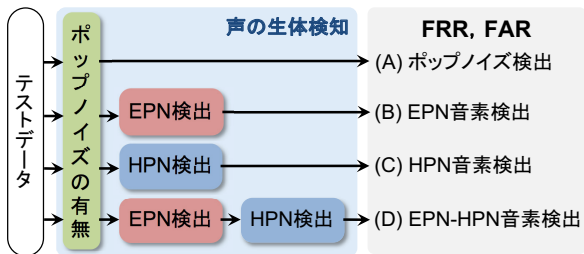


図 7 生体検知実験のフロー

(A) ポップノイズ検出：発話内におけるポップノイズの有無のみで判定。

(B) EPN 音素検出：ポップノイズ検出後に EPN 音素情報のみを用いて判定。

(C) HPN 音素検出：ポップノイズ検出後に HPN 音素情報のみを用いて判定。

(D) EPN-HPN 音素検出：ポップノイズ検出後に EPN 音素情報を用いて生体検知し，その後 HPN 音素情報を用いて判定。

また，テストデータには VLD2 データベースを用いた。女性話者 15 名それぞれに対し 10 発話を用意し，実発話，再生音声共に 150 文章用いて実験を行った。ただし発話内容は音素リスト作成データと同じであり，全話者共通となっている。生体検知実験の評価尺度には生体による音声を非生体として誤棄却した生体拒否率（False rejection rate; FRR）と非生体による音声を生体として誤受理した非生体受入率（False acceptance rate; FAR）を用いた。

話者照合実験には GMM-UBM に基づく話者照合システムを用いた。特徴量抽出およびモデル学習に用いた実験条件は表 4 にまとめてある。UBM の学習データとしては，JNAS の原音声および，JNAS の音声に電子協騒音データベース [8] の展示会場の雑音を SN 比が 0, 5, 10, 15, 20, 30dB となるよう重畳した音声をを用いて学習した。話者照合の評価尺度には等価エ

表 4 GMM-UBM の実験条件

UBM	
データベース	JNAS（女性のみ）
話者数	153 名
学習データ	165,599 文
混合数	1,024
特定話者モデル	
データベース	VLD
話者数	15 名
学習データ	60 文人
特徴量抽出	
サンプリング周波数	16 kHz
量子化ビット数	16 bit
分析フレーム	25 msec
フレームシフト	10 msec
特徴量	MFCC19+E+Δ+ΔΔ (60 次元)

ラー率（Equal error rate; EER）を用いた。話者照合実験での比較手法の詳細は以下の通りである。

**なりすまし攻撃なし**：なりすまし攻撃を含まないテストデータに対し，声の生体検知を行わずに話者照合し，EER を算出。

**なりすまし攻撃あり**：なりすまし攻撃を含むテストデータに対し，声の生体検知を行わずに話者照合し，EER を算出。

声の生体検知と組み合わせる手法は，なりすまし攻撃を含むテストデータに対して各検出手法（A）～（D）を行った後，話者照合を行い EER を算出した。

#### 4.2 実験結果

図 8, 9 に音素リスト毎の FRR および FAR をそれぞれ示す。まず共通の音素リストの結果（塗りつぶし）に着目すると，（A）ポップノイズ検出に比べて音素情報（（B） EPN 音素検出，（C） HPN 音素検出，（D） EPN-HPN 音素検出）を用いることで FRR が増加し，FAR が減少している。これは音素情報を用いることでポップノイズ検出のみの場合と比べ実発話を誤棄却することが増えてしまう一方で，再生音声の誤受理を減らしていることを示している。次に話者毎の音素リストの結果（ドット）に着目する。共通の音素リスト使用時と比べると，全検出手法で FRR が減少している。これは話者性を考慮することで共通の音素リストでは対応できなかった再生音声を棄却できたことを示している。しかしながら実発話も多く棄却してしまっている。以上より，従来の音素リストを用いることでなりすまし攻撃に対する頑健性は向上するもののユーザーの利便性が低下してしまうことがわかる。そのため，音素リストの設定方法を工夫することでなりすまし攻撃の誤受理率を維持しつつ，実発話の誤棄却数の減少を目指す必要がある。次にテキスト毎の音素リストの結果（斜線）に着目する。共通の音素リストに比べ FRR が大幅に減少している。これは共通の音素リストは複数文章を用いて作成したポップノイズ区間にかかる音素のランキングから選択されているため，文章によっては音素リストにかかる音素が文中に 2, 3 個しか含まれていない等，音素情報の判定に用いることができる音素数が少ない場合があった。し

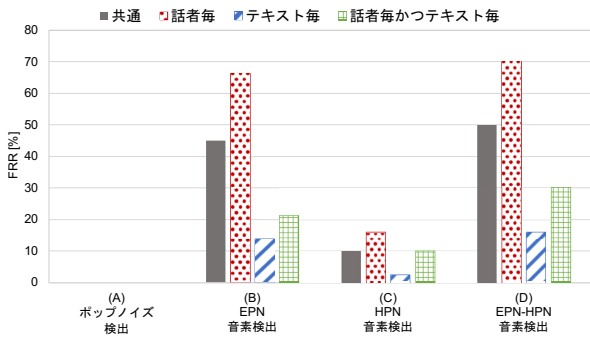


図 8 音素リスト毎の FRR

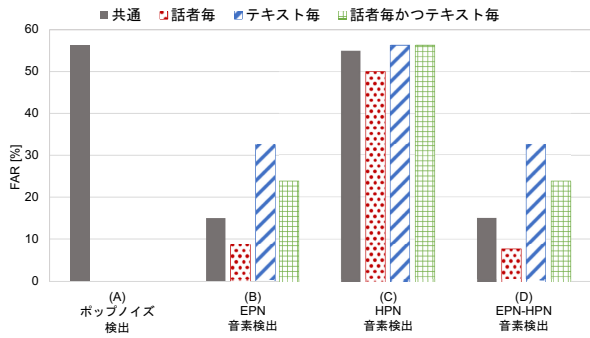


図 9 音素リスト毎の FAR

かしテキスト毎の音素リストはその選択方法から、音素リストに含まれる全ての音素が判定に利用可能なため FRR が改善したと考えられる。一方で FAR も増加している。これは共通の音素リストに比べ、テキスト毎の音素リストは判定に利用できる音素数が増えるため、検出されたポップノイズ区間に EPN 音素が被りやすくなったためである。最後に、話者毎かつテキスト毎の音素リストの結果 (格子) に着目する。EPN 音素検出および EPN-HPN 音素検出では、共通の音素リストと比べ大幅に FRR が減少した。これは音素リストの作成にテキストを考慮した効果であると考えられる。同時に FAR も増加しているが、テキスト毎の音素リスト使用時に比べると低い FAR となっている。これは、テキストの考慮だけでは対処できなかった再生音声を話者性の考慮により棄却できたためである。以上より、話者性およびテキストを考慮した音素リストを用いることで再生音声の誤受率率は若干増加するものの、実発話の誤棄却率を大幅に減少させることができるという。

次に声の生体検知および話者照合を統合したシステム全体の評価について考察する。表 5 に示す通り、なりすまし攻撃なしと攻撃ありの話者照合システムでは、後者の方が約 6.5 ポイントも EER が増加している。このことから、なりすまし攻撃によって話者照合システムの認証精度が大幅に低下することがわかる。次にポップノイズ検出の EER をみると、なりすまし攻撃ありに比べ EER が低下していることがわかる。これは話者照合の前の段階で再生音声棄却されたためである。ただし、実発話を 100% 受理するような閾値に設定しているため、図 9 にも示す通りなりすまし攻撃の半分以上は話者照合システムに入力されてしまっており、なりすまし攻撃なしの EER までは戻っていない。次に音素情報を用いたポップノイズ検出法との統合結果についてみると、どの手法でもポップノイズ検出単体

表 5 話者照合システムの EER

	攻撃	EER [%]			
		共通	話者毎	テキスト毎	話者毎かつテキスト毎
話者照合	無	4.46			
話者照合	有	10.9			
ポップノイズ検出	有	7.83			
		音素リスト			
		共通	話者毎	テキスト毎	話者毎かつテキスト毎
EPN 音素	有	-	-	13.8	21.6
HPN 音素	有	12.5	17.5	8.75	12.5
EPN-HPN 音素	有	-	-	16.3	30.0

の結果よりも EER が悪くなっていることがわかる。特に共通の音素リスト、話者毎の音素リストを用いた際の EPN 音素および EPN-HPN 音素の結果においては、声の生体検知の段階で実発話を多く棄却しすぎてしまったために、話者照合システム全体としての EER が測れなかった。これは図 8 に示す通り、声の生体検知の段階で本人の音声が多く棄却されてしまっていることが大きな要因である。しかし、話者照合において重大な問題である誤受率 FAR を低下させていることが大事であり、図 9 のように提案法はポップノイズ検出よりも大幅に FAR を低下させることが可能であるため、本人の受率率が上がれば十分に効果が出ると期待できる。

## 5. おわりに

本稿では話者性を考慮したテキスト依存型の声の生体検知について提案した。生体検知実験から、従来の音素リスト使用時に比べ話者毎およびテキスト毎に音素リストを設定することで音素情報に基づくポップノイズ検出法のユーザーの利便性が向上することを示した。一方で話者照合との統合実験から、実発話誤棄却数の更なる減少が必要であることがわかった。今後の課題として、提案法の統計に基づいた手法の提案などが挙げられる。

謝辞 本研究の一部は科学研究比基盤 (B) 2628006 による。

## 文献

- [1] Z. Wu, *et al.*, "Spoofing and countermeasures for speaker verification: a survey," *Speech Communication*, Vol. 66, pp.130-153, 2015.
- [2] T. Kinnunen, *et al.*, "The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection," *Proc. INTERSPEECH*, pp.2-6, 2017.
- [3] S. Shiota, *et al.*, "Voice liveness detection algorithms based on pop noise caused by human breath for automatic speaker verification," *Proc. INTERSPEECH*, pp.239-243, 2015.
- [4] G. W. Elko, *et al.*, "Electronic pop protection for microphones," *Proc. WASPAA*, pp.46-49, 2007.
- [5] 望月ら, "話者照合のための話者性を考慮した音素情報に基づくポップノイズ検出法による声の生体検知," *日本音響学会 秋季大会*, no.2-Q-16, pp.165-168, 2017.
- [6] 望月ら, "話者照合のための音素情報を考慮したポップノイズ検出法による声の生体検知," *電子情報通信学会 論文誌*, vol.J101-D, no.3, 2018.
- [7] 汎用大語彙連続音声認識エンジン Julius, <http://julius.osdn.jp/>
- [8] 音声資源コンソーシアム 電子協騒音データベース, <http://research.nii.ac.jp/src/JEIDA-NOISE.html>