情報処理学会論文誌　コンシューマ・デバイス & システム　Vol.8　No.3　1–8 (Oct. 2018)

コンシューマ・デバイス論文

不正抑止効果の高い音声対話型AI帳票の実現に向けた取り組み ——多段階話者適応方式の提案

古明地 秀治[1,a)]　坂口 基彦[1]　田淵 仁浩[1]　服部 浩明[1]　奥村 明俊[1]

受付日 2018年2月28日, 採録日 2018年8月3日

概要：本稿では，検査工程において，不正抑止効果の高い音声対話型 AI 帳票の実現のために，ヒアラブルデバイスとの組合せを提唱する．また，ヒアラブルデバイスとの組合せにおいて課題になる，音声対話型 AI 帳票の音声認識エンジン VoiceDo の認識精度劣化を防止する多段階話者適応方式を提案する．音声対話型 AI 帳票にヒアラブルデバイスを組み合わせることで，検査データに対して「いつ」の情報だけでなく，耳音響認証技術により得られる「誰が」と高精度位置測位技術により得られる「どこで」の情報を付加することができる．これらの情報は検査データの確証になるため，不正を働く心理的障壁を上げる．また，提案する多段話者適応技術により，74%だった VoiceDo の単語認識精度が，97%に改善され，不正抑止効果の高い音声対話型 AI 帳票実現の見通しを得た．

キーワード：AI 帳票，ヒアラブルデバイス，耳音響認証，音声認識，話者適応

An Approach to Realize an Artificial-intelligence Voice-activated Electronic Forms Having Cheat Deterrent Effect — A Proposal of Multi-layer Speaker Adaptation

Shuji Komeiji[1,a)]　Motohiko Sakaguchi[1]　Masahiro Tabuchi[1]　Hiroaki Hattori[1]
Akitoshi Okumura[1]

Received: February 28, 2018, Accepted: August 3, 2018

Abstract: This paper proposes an artificial-intelligence powered, voice-activated electronic forms (AI-forms) having cheat deterrent effect by exploiting hearable device. Besides, this paper also proposes a multi-layer speaker adaptation which covers the defect of automatic speech recognition (ASR) engine, VoiceDo employed by AI-forms with hearable device hearable device. The combination of the AI-forms and the hearable device enables to attach the additional information of not only “when” but also “who” and “where” to inspection data. The information of “who” and “where” can be identified by acoustic ear authentication and high accuracy positioning technology supported by hearable device. These additional information make it more difficult for workers to make falsify data without inconsistencies, and as a result, these enforce a psychological barrier to cheat. Besides, the experiment of multi-layer speaker adaptation achieved 97% ASR accuracy from 74%.

Keywords: AI-forms, hearable device, ear authentication, automatic speech recognition, speaker adaptation

1. はじめに

産業革命以降，人々は様々な製品・サービスを生産・分配してきた．製品・サービスに対する満足度，安全性を保証する品質の低下は，その生産・分配者への信頼を失墜さ

[1] 株式会社 NEC ソリューションイノベータ
NEC Solution Innovator, Ltd., Koto, Tokyo 136–8627, Japan
[a)] s-komeiji@ce.jp.nec.com

せ，経済的に大きな損失を与える．品質低下の防止と品質の証明のために，検査点検の工程は不可欠であり，そこで得られるデータは正しいものでなければならない．

近年，製品・サービスの複雑化，人々の安全性への意識の高まりにより，検査点検の重要度が増している．しかし，昨今，検査データに対する不正の発覚が相次ぎ [1]，大きな社会問題となっている．ジェムコ日本の古谷氏は，検査点検における不正を 4 つに分類している [2]：

- 定められた検査の未実施，あるいは必要な検査項目を一部省略する．
- 実施した検査結果を改竄・捏造する．
- 合格するように検査条件を勝手に変える．
- 合格するまで検査を何度も繰り返す．

これらは管理者の指示ではなく，作業者の意図により行われることが多い．その主な目的は管理者側から提示される納期に間に合わせるための労力や時間の節約にある．不正を防止するためには，検査工程の「効率化」と「見える化」が必要になる．「効率化」により，予期せぬ時間ロスを減らし，作業者が納期に追われる頻度を下げることができる．また，「見える化」により，不正を働く心理的障壁を上げることができる．さらに，「効率化」することで，検査工程に内在する新たな課題が浮かびやすくなり，その課題の検証のために「見える化」が促進される．また，「見える化」することで，課題の本質を見つけることができ，「効率化」が促進される．この「効率化」と「見える化」のループを日常的に回すことで，不正抑止効果の高い検査工程を実現できる．

現在，検査工程の「効率化」と「見える化」の両面で成功している事例として，音声対話型 AI 帳票 [3] がある．音声対話型 AI 帳票は，従来の電子帳票では失われがちな “読み書きしやすさ” や “作業引き継ぎなどの運用容易性” を音声対話で実現し，生産性向上（「効率化」）と作業実績収集（「見える化」）を両立する電子帳票である．音声対話型 AI 帳票は，作業内容確認と作業結果入力の手続きをナチュラルユーザインタフェース（Natural User Interface; NUI）[4] の考えに基づき「効率化」している．NUI とは文献 [4] にあるように，「人間の五感や人間が自然に行う動作によって機械を操作する方法」と定義している．音声対話型 AI 帳票においては，長時間利用でも疲労が少ない軽量端末およびヘッドセットを用いて，音声により作業内容を聞き，作業結果を発話により入力するハンズフリー・アイズフリーの音声対話 NUI を実現している．NEC グループの工場での約 2 年間の評価によると，作業者の訓練コストを 1/3 に削減，生産性約 20%向上，作業者のスキル改善サイクルを約 40 倍高速化する効果を実証し，「効率化」の観点で成功している [3]．また，検査結果をサーバで一括管理できる仕組みを整えることで「見える化」の観点でも成功している．さらに，音声対話型 AI 帳票は各作業者のための端末，ヘッドセット，サーバの用意だけで導入できる，導入容易性を持ち合わせる．

上述のような利点から，現在いくつかの事業者が音声対話型 AI 帳票の導入を始めている．しかし，近年問題になっている検査工程における不正防止のために，「いつ」の情報だけでなく，検査データの確証になりうる情報「誰が」と「どこで」の取得による「見える化」強化の要望が高まっている．そこで本稿では，音声対話のために従来用いていたヘッドセットを，NUI を維持した形で「誰が」と「どこで」の情報を取得できるヒアラブルデバイス [5] に置き換えた音声対話型 AI 帳票を提唱し，ヒアラブルデバイスとの組合せにおいて課題になる，音声認識の精度劣化を改善する多段階話者適応方式を提案する．

本稿では，2 章においてヒアラブルデバイスに関して説明する．また，ヒアラブルデバイスを用いた音声対話型 AI 帳票を提唱し，その課題を整理する．3 章において音声対話型 AI 帳票が採用している音声認識エンジン，VoiceDo を説明する．4 章において本稿における提案手法である多段階話者適応について説明する．5 章において多段階話者適応の評価実験について述べ，6 章でまとめる．7 章において，今後の展開を述べる．

2. ヒアラブルデバイス

本稿で対象とするヒアラブルデバイスは，耳にデバイスを装着することで，「ユーザの情報をとらえ続ける」ことと，「UI を意識せず情報取得・操作する」ことの両立を実現する．ヒアラブルデバイスはスピーカ，マイク，地磁気センサ，加速度センサおよび，ジャイロセンサで構成され，様々な機能を提供する [5]．ここでは，本稿で注目している「誰が」と「どこで」の計測技術，耳音響認証技術と高精度屋内位置測位技術を説明する．

2.1 耳音響認証技術

耳音響認証技術は，スピーカから耳内に向けて再生する検査音の反響音を，耳内を向くマイクにより集音することで得られる外耳道音響特性を利用することで，個人を特定する技術である．外耳道音響特性には個人差があることが確認されており [6], [7]，本技術により現在，他人受け入れ率 0.01～0.1%で本人棄却率 2～3% [8] と，ユーザ認証として実用的な精度を確認している．耳音響認証技術によると，指紋認証や虹彩認証のような特別な手続きが不要であるため，効率性を維持しながら「誰が」の情報を常時判定できる．

2.2 高精度屋内位置測位技術

高精度屋内位置測位技術は，地磁気センサを利用することで，GPS によって位置測位のできない屋内において，精度 2 m 程度の測位を実現する [9]．この技術は，屋内に存

在する鉄骨などの影響による地磁気の乱れのデータを事前に測定することで，屋内位置測位を実現する．本技術は，ビーコンや Wi-Fi 電波を用いる方法で必要な設備配置が不要であるという特長を有する．

2.3　ヒアラブルデバイスを利用した音声対話型 AI 帳票

本稿の目的は，不正防止効果の高い検査の実現である．そのために音声対話型 AI 帳票 [3] の「見える化」の観点に着目し，「いつ」の情報だけでなく，「誰が」や「どこで」の情報の取得が求められている．そこで本稿では，音声対話のために従来用いていたヘッドセットを，NUI を維持した形で「誰が」と「どこで」の情報を取得できるヒアラブルデバイス [5] に置き換えた音声対話型 AI 帳票を提唱する．まず，「誰が」の情報により，なりすましを防止できる．たとえば，近年問題になった無資格者による検査点検 [1] の対策になる．ヒアラブルデバイスの耳音響認証技術によれば，作業員が検査音を聞くだけでよいため，効率性を維持したまま「誰が」の情報を取得できる．また，「どこで」の情報により，検査対象によって異なる資格の有無を，検査対象の位置情報から判定することができる．これらの情報を検査データの確証とすることで，矛盾のない形でデータの改竄・捏造するのが難しくなり，検査作業者の不正を働く心理的障壁を上げることができる．また，これらの情報により，作業者の動線の可視化ができるようになり，無駄な動きが生じないよう，物品の配置を最適化することにより，「効率化」が促進される．さらに，高精度屋内位置測位技術では，特別な設備を導入する必要はないため，音声対話型 AI 帳票の導入容易性を維持できる．

2.4　ヒアラブルデバイスを利用する場合の課題

音声対話型 AI 帳票のヘッドセットをヒアラブルデバイスに置き換えた場合，耳音響認証用のマイクを用いることになるため，音声認識は外耳道音響特性の影響を受ける．**図 1** は，それぞれ通常のヘッドセットとヒアラブルデバイス未装着状態，ヒアラブルデバイス装着状態で集音した音声のスペクトルを示す．これらのスペクトルは，プリエンファシスを行った同一の内容の発声を含む区間の平均値として算出している．図 1 によると，通常のヘッドセットとヒアラブルデバイス非装着状態のスペクトルはほぼ一致している．一方，ヒアラブルデバイス装着状態でのスペクトルは，前者 2 つと比較して，5 kHz 付近を中心とする谷ができており，違う形状を持つ．この違いは外耳道音響特性の影響によるものである．

図 2 は，男性話者 3 名のヒアラブルデバイス装着状態のスペクトルを示す．図 2 から分かるように，5 kHz 付近に現れる谷の位置，深さには個人差がある．これは，外耳道音響特性が個人により異なるためである．これらの違いは音声認識性能に悪影響を与えると考えられる．

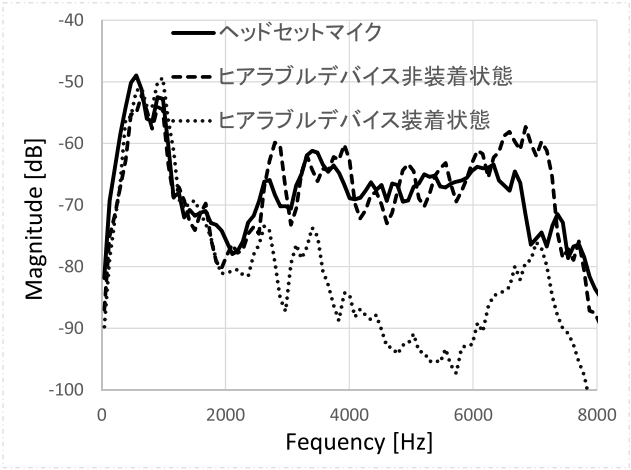


図 1　ヘッドセットマイクとヒアラブルデバイスマイクにおけるスペクトル形状の比較

Fig. 1　The comparison of the spectrum shapes of recorded signal between usual headset microphone and Hearable device microphone.

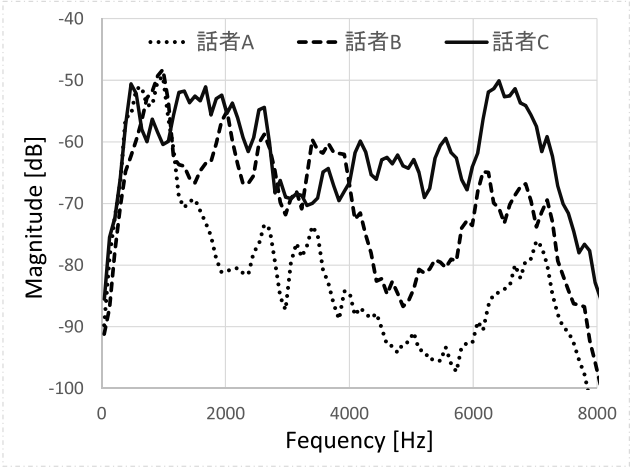


図 2　ヒアラブルデバイスで集音した音声スペクトルの話者による違い

Fig. 2　The differences in speech spectrums recorded by Hearable device among three speakers.

ヒアラブルデバイスが普及していれば，インターネットを介して，大量のデータを容易に収集し専用の音響モデルを作成することができるが，現状では難しい．そこで，話者適応技術に着目する．話者適応技術では，話者の発声から得られる音響特徴を音声認識に反映する技術であり，比較的少量のデータ数で認識精度を向上させることができる．

3.　VoiceDo

本章では，音声対話型 AI 帳票の音声認識エンジンとして採用している VoiceDo [10], [11] について説明する．VoiceDo は高い耐雑音性により，セリ市場やコールセンタなど高騒音の現場 200 カ所以上で採用実績がある [12], [13]．以下では，VoiceDo の概要，音声認識技術および，話者適応技術の説明をする．

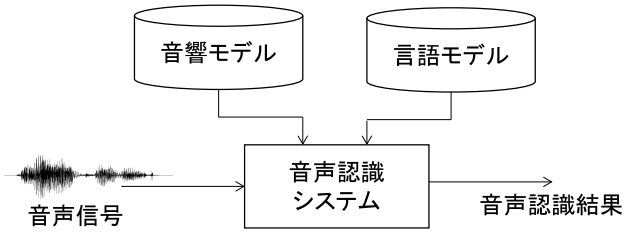


図 3　音声認識のシステム構成図

Fig. 3　An automatic speech recognition system configuration diagram.

3.1　VoiceDo の概要

VoiceDo は，音声用マイクと雑音用マイクの 2 つのマイクを持つ 2ch ヘッドセットを用いることを特徴とし，耐雑音性能の高い音声認識を実現する．2 つのマイクに入力された音声は，スペクトル領域において引き算することで（2ch スペクトルサブトラクション）雑音の影響が低減される [11]．今回は，1ch 入力のヒアラブルデバイスを用いるため，1ch スペクトルサブトラクションにより雑音抑圧を実施する．なお，ヒアラブルデバイスの場合，集音用マイクは耳内にあるため，雑音の影響は物理的に低減されている．

近年，深層学習のような潤沢な計算リソースをインターネットを介して利用するクラウド型音声認識が用いられる．しかし，音声対話型 AI 帳票が利用される作業現場では，工場や倉庫などで，インターネット接続ができない状況にしばしば直面する．また，作業効率化のためには，速い応答速度が求められる．VoiceDo は，インターネットを介さないため，各作業者が持つ端末で処理が完結し，応答速度も速いため，音声対話型 AI 帳票に適している．また，公共のインターネットに載せられない機密情報を含む発声を安全に扱うことができる．

3.2　音声認識の仕組み

音声認識は，音響モデルと言語モデルを参照して，マイクで集音した音声信号をテキスト（単語や文章）に変換する技術である（**図 3**）．音声認識ではまず，マイクで集音した音声信号を特徴量列に変換する．次に，それぞれ特徴量ごとに音響モデルが記録する各音素の音響特徴と比較し，類似度を計算する．最後に，言語モデルが記録する単語やそのつながり方の法則，つながりやすさに基づいて，特徴量列の合計類似度が最大となる音素列を算出する．この音素列が構成する単語や文章が，音声認識結果として出力される．

3.3　話者適応技術

上述のように，音声認識では，音声信号の音響特徴と音響モデルの類似度を基にしていため，それらの音響特徴が異なるときに認識性能が低下する．話者適応は，話者が事前に発声した音声データを用いて音響モデルをその話者に適応することで，認識対象話者の認識精度を向上させる技術である．昨今では，深層学習に関する話者適応が提案されている [14], [15]．

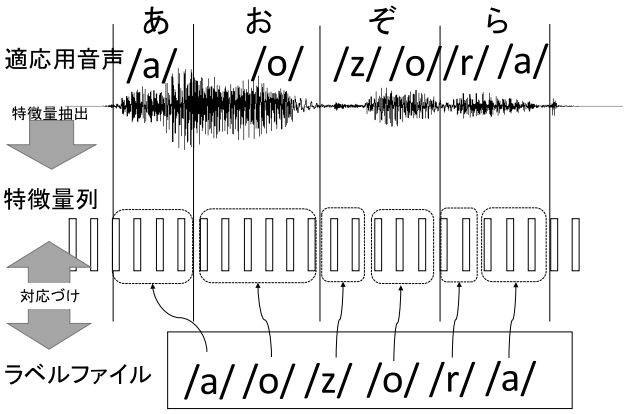


図 4　話者適応における音素対応づけ（成功例）

Fig. 4　Phoneme alignment for speaker adaptation (Successful example).

話者適応技術は，発声内容を記述したラベルファイルを必要としない教師なし適応と必要とする教師あり適応に分類される．教師なし適応の場合は，話者の自由な発声を適応に用いることができるが，たくさんの発声を必要とする．一方，教師あり適応の場合には，話者に決められた内容の発声を強いることになるが，教師なし話者適応と比較して少量の発声での適応が可能である．音声対話型 AI 帳票は，業務目的での利用になるため，利用者への負担を考慮して，より少量の発声でも話者適応できる教師あり話者適応を採用している．

教師あり話者適応では，発話内容を記述したラベルファイルおよび，ラベルに従った発声と，適応対象の音響モデルが必要である．まず，ラベルファイルの発声内容に従って構成された音響モデルと話者の発声から抽出された音素の特徴量列を対応づける（**図 4**）．音素ごとに対応づけられた特徴量をまとめあげ，元の音響モデルの音素特徴との特徴量差分を求め，これを用いて音響モデルを適応する．

教師あり話者適応は，対応づけが正しいことを前提としている．そのため，ラベルファイルに従わない言い間違えやたまたま大きな雑音が入ってしまった発声を，適応に用いるべきではない．そこで VoiceDo では，発声した音声データから，実際に話者適応に用いるデータを，事前の音声認識で正解する発声のみに絞り，正しい対応づけが行われないと思われる発声（**図 5**）を取り除く，事前音声認識を用いた発声選別を行っている．

深層学習を用いない VoiceDo では，認識の応答速度向上のために各音素の音響特徴を木構造化した音響モデルを用いており [16]，この木構造音響モデルに着目した自律的モデル複雑度制御法（Autonomous Model Complexity

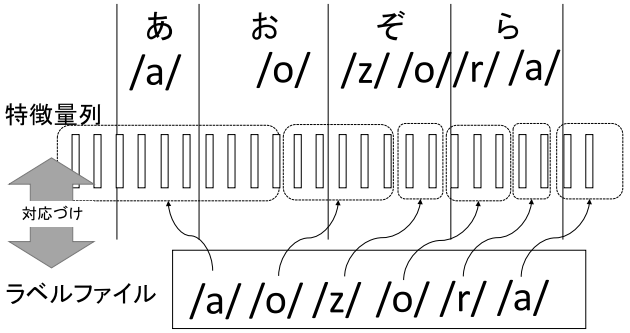


図 5　話者適応における音素対応づけ（失敗例）

Fig. 5　Phoneme alignment for speaker adaptation (Failed example).

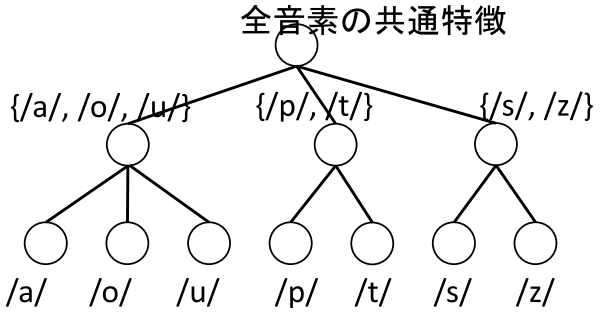


図 6　木構造音響モデル

Fig. 6　Tree structure acoustic model.

Control; AMCC）を教師あり話者適応方法として採用している [17]．AMCC は適応データ数に応じて音響モデルの適応の複雑度を制御する教師あり話者適応方式であり，適応データ数が少ない場合にも多い場合にも，適切に話者適応が行われる．木構造音響モデルとは各リーフノードが各音素特徴に相当，親ノードの音素特徴は，子ノードの特徴量の共通特徴に相当，ルートノードは全音素の共通音素特徴に相当する音響モデルである（図 **6**）．

木構造音響モデルを用いて教師あり話者適応を行った場合，前述の特徴量との対応づけが，リーフノード，親ノード，ルートノードに対してそれぞれ行われる（図 **7**）．したがって，先に述べた特徴量差分がそれぞれのノードで求められる．特徴量差分の算出に用いられるデータ量は，リーフノードからルートノードへ，階層が浅くなるにつれて多くなる．特徴量差分は，データ量が多い方が，信頼度が高いと考えられる．そこで，十分なデータ量が集まらない信頼度が低いノードに対して，その親ノードで求めた信頼度が高い特徴量差分を用いてそのノードの適応を行うこととする．つまり，AMCC は，適応データ数が少ない場合に大局的で大まかな適応，多い場合には，局所的で詳細な適応を行う方式である．

適応データ数の違いによる，適応の様子を特徴量空間で眺めたものを図 **8** に示す．適応データ数が多い場合には，音響モデル中の各音素特徴が適応される一方（詳細な適応），少ない場合には，まとまった単位で音素特徴が適応される（大まかな適応）．

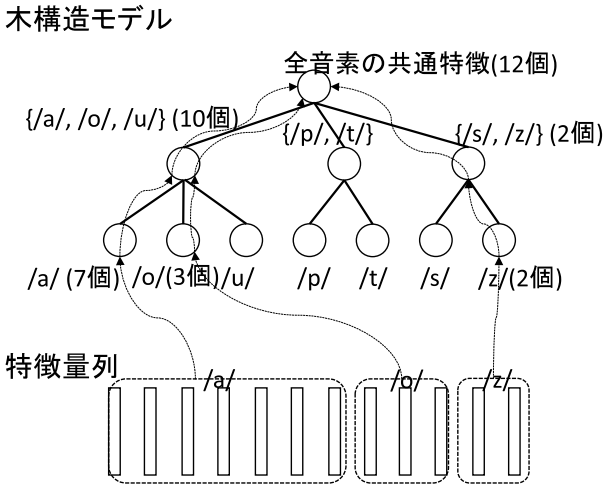


図 7　木構造音響モデルにおける音素対応づけ．音素/a/，/o/，/z/ からなる音素列に図 6 の木構造音響モデルを適応する様子を示す

Fig. 7　Phoneme indexing to tree structure acoustic model. The tree structure acoustic model in Fig. 6 is adapted to a feature sequence composed of phonemes /a/, /o/, and /z/.

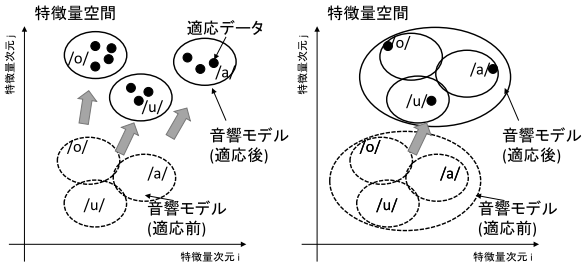


図 8　AMCC における音響モデル適応の概要（左：適応データが多い場合 → 詳細な適応，右：適応データが少ない場合 → 大まかな適応）

Fig. 8　The overview of model adaptation with AMCC (The left shows fine adaptation with many data, the right shows coarse adaptation with few data).

4.　多段階話者適応

前述のように，ヒアラブルデバイス装着時の発声は，外耳道音響特性の影響により，通常のマイクで集音した発声と音響特徴が大きく異なる．そのため，従来の VoiceDo の教師あり話者適応を行っても，事前音声認識による発声選別の過程でほとんどの発声が除外されてしまう．しかし，AMCC によれば少量のデータでも話者適応の効果が得られるため，適応後の音響モデルを用いて再度話者適応した場合，事前音声認識の精度が向上し，適応に用いられる発声を増やすことができる．また，発声が増えることにより，AMCC ではより詳細な話者適応が可能になる．このような AMCC の性質を利用して繰り返し話者適応を実施することで，話者適応の効果を向上させる方法を，多段階話者

適応として提案する．

話者適応を N 回繰り返した場合の N 段階話者適応のアルゴリズムを**図 9** に示す．また，適応の繰返しの様子を特徴量空間で眺めたものを**図 10** に示す．図 10 は 1 回適応を増やすことにより，適応データ数，音響モデルがどのように変化するかを示す．左図は n 回の適応を実施した場合を示し，右図は $n+1$ 回の適応を実施した場合を示す．適応回数が少ないうちは，適応に用いられる発声数が少なく，大まかな適応になる（図 10 左）．一方，適応回数が増えると，発声数が増加し，詳細な適応になる（図 10 右）．

なお，多段階で話者適応を行う方法として，文献 [18] があげられるが，この方法は音響モデルを話者のスペクトル空間にクラスタリング技術を用いて階層を明示的に指定して適応する技術であり，話者適応を繰り返すことにより変動する適応データ量に応じて適切な階層で適応を行う多段階話者適応とは異なる．

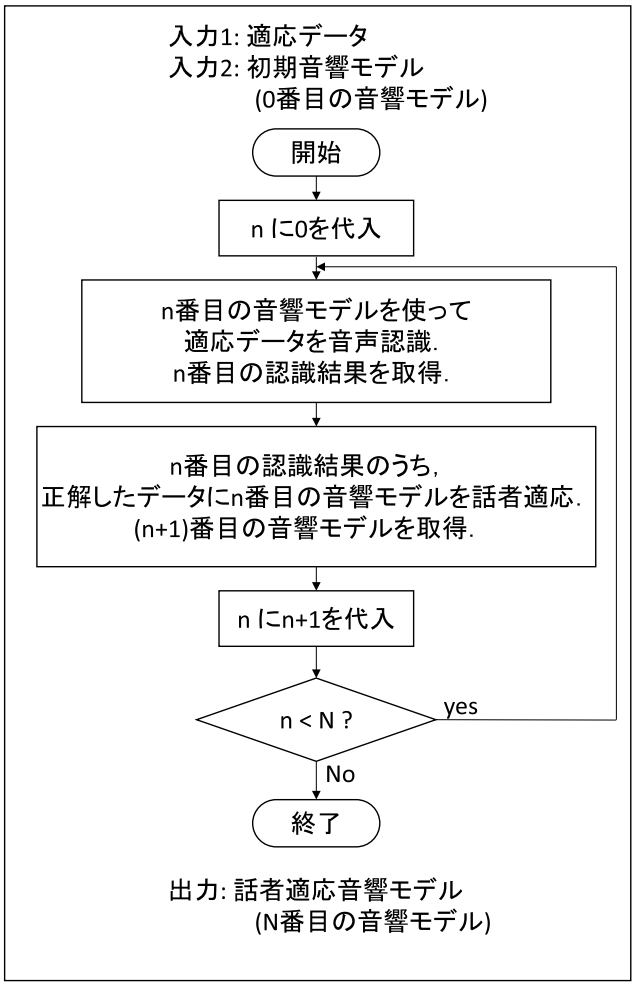


図 9　N 段階話者適応のアルゴリズム

Fig. 9　Adaptation algorithm for N-layer speaker adaptation.

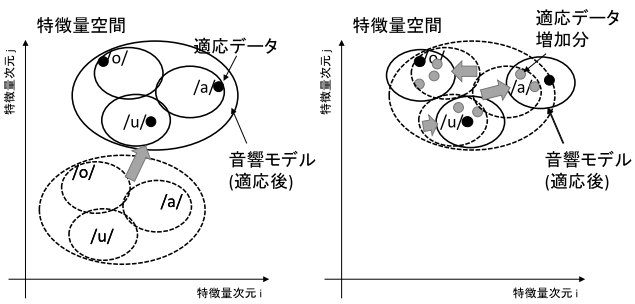


図 10　多段階話者適応における音響モデルの適応過程（左：n 段目の適応，右：$n+1$ 段目の適応）

Fig. 10　Adaptation process of multi-layer speaker adaptation.

5. 評価実験

提案法の多段階話者適応の評価実験を行った．音声は 3 人の男性話者の発声を用いた（話者 A，話者 B，話者 C）．これらの話者は，2.3 節における図 1，図 2 に示すスペクトルの話者に相当する．適応データは音素バランスを考慮した 250 単語とした．また，テストのタスクは 5 桁数字棒読み 20 発声とした．多段階話者適応における段数 N は 3 とした．

評価結果を**表 1** に示す．表 1 は，男性 3 人の平均単語認識精および 3 人の話者各々の認識精度を示す．各行は話者の単語認識精度を，各列はそれぞれ，未適応，従来法，提案法の認識精度を示す．従来法と提案法の括弧内の数字は，適応データ 250 発声のうち，発声選別のための事前音声認識で正解し，実際に適応に使った発声数の割合である．まず，3 話者の平均単語認識精度を見ると，未適応時の認識精度は 74%であり，従来法の適用により 93%に改善する．さらに，提案法により 97%にまで改善する．また，発声選別による発声数の割合も従来法では 54%であったところ，繰り返すことで，90%に増加していることが分かる．

話者 B，C に関しては，従来法の適用でも十分な効果が得られている．これは，図 2 に示したように，耳内集音音声のスペクトルにおける 5 kHz 付近の谷が話者 B，C に関しては浅く，外耳道音響特性の影響が小さかったためと思われる．

次に，多段話者適応の効果を示す他の例として，防塵マスク着用時の音声認識評価結果を示す．防塵マスクでは，口元全体がマスクに覆われるため，ヒアラブルデバイス以上に，音響特徴が通常想定されているものと大きく異なる．**表 2** に防塵マスク着用時における多段階話者適応の評価結果を示す．適応データやテストのタスクはヒアラブルデバイス評価のものと同じとした．まず，4 話者の平均単語

表 1　耳内集音音声の単語認識精度括弧内は，実際に適応に使った発声数の割合

Table 1　Word accuracies of speech signals recorded in ear.

	未適応	従来法	提案法
3 話者平均	74 %	93 % (54 %)	97 % (90 %)
話者 A	36 %	81 % (23 %)	95 % (72 %)
話者 B	92 %	97 % (57 %)	97 % (98 %)
話者 C	95 %	100 % (81 %)	100 % (99%)

表 2　防塵マスク着用時の単語認識精度括弧内は，実際に適応に使った発声数の割合

Table 2　Word accuracies when speakers are wearing dust protective masks.

	未適応	従来法	提案法
4 話者平均	41 %	81 % (24 %)	95 % (93 %)
話者 A’	23 %	54 % (9 %)	85 % (80 %)
話者 B’	67 %	97 % (51 %)	99 % (97 %)
話者 C’	41 %	97 % (26 %)	100 % (100%)
話者 D’	33 %	75 % (9 %)	95 % (95 %)

認識精度を見ると，未適応時の認識精度は 41%であり，従来法の適用により 81%に改善する．さらに，提案法により 95%にまで改善する．また，発声選別による発声数の割合も従来法では 24%であったところ，適応を繰り返すことで，93%に増加することが分かる．個々の話者についても全話者で多段階話者適応の効果が得られている．

表 1 と表 2 から，多段階話者適応により単語認識精度が向上することが分かった．また，発声選別のための事前音声認識の精度も段数を増やすに従い向上することが分かる．これは，少量の発声数でも適応効果が得られる AMCC を活用する多段階話者適応に期待する性質であり，本評価により多段階話者適応の効果が示された．

6.　まとめ

本稿では，不正抑止効果が高い音声対話型 AI 帳票の実現を目指して，検査データに「いつ」だけでなく「誰が」と「どこで」の情報も付加するために，ヒアラブルデバイスとの組合せを提唱した．さらに，音声対話型 AI 帳票が採用している音声認識エンジン VoiceDo をヒアラブルデバイスが集音する音声の音響特徴に適応させる多段階話者適応方式を提案した．評価実験により多段階話者適応の有効性を確認した．これにより不正防止効果の高い音声対話型 AI 帳票の実現の見通しを得た．

7.　今後の展開

今後の取り組みとして，音声認識エンジン VoiceDo の話者適応機能に本稿で提案した多段階話者適応を実装し，現場への試験導入を行っていく．試験導入において，本稿で提唱している音声対話型 AI 帳票の有効性を実証する．その後，より多くの現場に導入し，「効率化」と「見える化」のループを日常的に回すことで不正防止効果の高い検査工程を実現していく．

また，本稿で対象とするヒアラブルデバイスには，ユーザの活動量計測，姿勢計測の機能をあわせ持つ．これらを活用することで，不正防止の観点に加えて，作業者の健康状態に配慮する音声対話型 AI 帳票も実現可能と考えられ，今後，検証していく．

謝辞　データ収集のサポート，評価に対するコメントをいただいた NEC ソリューションイノベータ田中大介氏，小田英司氏に感謝いたします．

参考文献

[1] フジサンケイ危機管理室, 入手先 ⟨http://www.fcg-r.co.jp/research/incident/⟩ （参照 2017-12-12）.
[2] 現場はこうしてデータを偽装する，入手先 ⟨http://techon.nikkeibp.co.jp/atcl/feature/15/122200045/102300193/⟩ （参照 2017-12-12）.
[3] 田淵仁浩，坂口基彦，服部浩明，奥村明俊：音声対話型 AI 帳票を実現する現場作業支援ソリューションの提案，情報処理学会論文誌コンシューマ・デバイス&システム（CDS），Vol.8, No.2, pp.13–23 (2018).
[4] 東京工芸大学，ナチュラルユーザーインターフェースに関する調査，入手先 ⟨https://www.t-kougei.ac.jp/static/file/nui.pdf⟩ （参照 2017-12-20）.
[5] ヒアラブル技術によるヒューマン系 IoT ソリューションの取り組みと展望，入手先 ⟨http://jpn.nec.com/techrep/journal/g17/n01/170110.html⟩ （参照 2017-12-04）.
[6] Akkermans, A.H.M., Kevenaar, T.A.M. and Schobben, D.W.E.: Acoustic ear recognition for person identification, *Proc. AutoID2005*, pp.219–223 (2005).
[7] Yano, S., Hokari, H. and Shimada, S.: A Study on the Personal Difference in the Transfer Functions of Sound Localization Using Stereo Earphones, *IEICE Trans. Fundamentals*, Vol.E83-A, No.5, pp.877–887 (2000).
[8] 荒川隆行，矢野昌平，越仲孝文，入澤英毅，今岡　仁：外耳道音響特性を用いた高精度個人認証，音講論集，pp.841–842 (2016).
[9] NEC，地磁気を活用して屋内の対象者の位置を正確に測定する技術を開発—ヒアラブルデバイス向け事業を推進，入手先 ⟨http://jpn.nec.com/press/201610/20161028_02.html⟩ （参照 2017-12-04）.
[10] 塚田　聡：耐雑音音声認識装置 VoiceDo，NEC 技報，Vol.63, No.1 (2010), 入手先 ⟨http://jpn.nec.com/techrep/journal/g10/n01/pdf/100118.pdf⟩
[11] 服部浩明，辻川剛範：耐雑音音声認識エンジン VoiceDo の応用，情報処理学会，Vol.2013-SLP-98, No.3 (2013).
[12] VoiceDo 活用シーン，入手先 ⟨http://jpn.nec.com/voicedo/jirei.html⟩ （参照 2017-12-04）.
[13] 「音声受注入力システム」を導入注文の処理時間を 6 割減，日本酒類販売（株），入手先 ⟨http://www.itmedia.co.jp/enterprise/articles/0909/15/news039.html⟩ （参照 2017-12-04）.
[14] Abdel-Hamid, O. and Jiang, H.: Fast Speaker Adaptation of Hybrid NN/HMM Model for Speech Recognition Based on Discriminate Learning of Speaker Code, *Proc. ICASSP2013*, pp.7942–7946 (2013).
[15] Liao, H.: Speaker Adaptation of Context Dependent Deep Neural Networks, *Proc. ICASSP2013*, pp.7947–7951 (2013).
[16] Watanabe, T., Shinoda, K., Takagi, K. and Yamada, E.: Speech Recognition using tree-structured probability density function, *Proc. ICSLP1994*, pp.223–226 (1994).
[17] 篠田浩一，渡辺隆夫：音声認識における自律的なモデル複雑度制御を用いた話者適応化，電子情報通信学会論文誌 D-II，Vol.J79-D-II, No.12, pp.2054–2061 (1996).
[18] 古井貞熙：スペクトル空間の階層的クラスタ化による音声認識，音響学会音声研資，SP88-21 (1988).

古明地 秀治

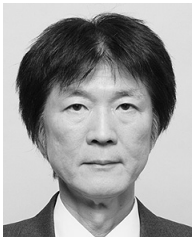
NEC ソリューションイノベータ（株）．2007 年東京農工大学工学部電気電子工学科卒業．2009 年東京大学大学院情報理工学系研究科修了．同年 NEC 入社．耐雑音音声認識の研究に従事．2015 年 NEC 情報システムズ出向，音声認識の研究，開発に従事．日本音響学会会員．

坂口 基彦

NEC ソリューションイノベータ（株）．1997 年東京農工大学大学院工学研究科修了．同年 NEC 入社．MBA in Technology Management．UI/UX の研究開発，先端技術を活用した新事業創出に従事．人工知能学会 2016 年現場イノベーション賞受賞．

田淵 仁浩　（正会員）

NEC ソリューションイノベータ（株）．1987 年早稲田大学理工学部電子通信学科卒業．1993 年同大学大学院理工学研究科電気工学専攻博士後期課程修了．1989～1993 年同大学情報科学研究教育センター助手．1993 年日本電気（株）C&C 研究所入社．現在，NEC ソリューションイノベータ（株）で認知科学や人工知能を用いた人間機能拡張の事業開発に従事．博士（工学）．1988 年情報処理学会第 35 回全国大会学術奨励賞，1994 年情報処理学会平成 6 年度山下記念研究賞，人工知能学会 2016 年現場イノベーション賞等受賞．電子情報通信学会会員．



服部 浩明　（正会員）

NEC ソリューションイノベータ（株）．1985 年北海道大学大学院工学研究科修了．同年 NEC 入社．2008 年 NEC 情報システムズ出向，現職に至る．工学博士．音声認識・合成，話者照合の研究，開発に従事．人工知能学会 2016 年現場イノベーション賞受賞．日本音響学会，電子情報通信学会各会員．

奥村 明俊　（正会員）

NEC ソリューションイノベータ（株）．1986 年京都大学大学院工学研究科修士課程修了．同年日本電気（株）入社．自然言語処理，音声翻訳，コミュニケーションロボット等の研究開発に従事．1992～1994 年南カリフォルニア大学客員研究員として DARPA 機械翻訳 PJ 参加．現在，NEC ソリューションイノベータ（株）執行役員．工学博士．情報処理学会平成 20 年度喜安記念業績賞，2007 年度独創性を拓く先端技術大賞経済産業大臣賞，人工知能学会 2010 年，2015 年，2016 年現場イノベーション賞，情報処理学会 2017 年度山下記念研究賞，情報処理学会 2017 年度業績賞等受賞．