Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

不正抑止効果の高い 音声対話 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, 坂口 基彦, 田淵 仁浩, 服部 浩明, 奥村 明俊[†1]

概要：本稿では，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み合わせを提唱する．また，ヒアラブルデバイスとの組み合わ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「誰により」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．これにより，矛盾のない形でデータの改竄・捏造するのが難しくなり，検査作業者が不正をする心理的障壁を上げることができる．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 powered voice-activated electronic forms having cheat deterrent effect - A proposal of multi-layer speaker adaptation -

SHUJI KOMEIJI, MOTOHIKO SAKAGUCHI, MASAHIRO TABUCHI, HIROAKI HATTORI, AKITOSHI OKUMURA[†1]

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 "by who" and "where" to inspection data. The information of "by 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]：

① 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．
② 実施した検査結果を改竄・捏造する
③ 合格するように検査条件を勝手に変える．
④ 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．納期に対するプレッシャーから行われ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．効率化することにより，作業時間のバラつきを抑え作業予定が立てやすくなり，管理者は，作業者にプレッシャーを与えない納期設定ができるようになる. さらに,「見える化」により，工程や作業者毎の非効率な点の洗い出しができるようになり，さらなる効率化を実現することができる．この「効率化」と「見える化」のループを日常的に回すこと

†1 (株)NEC ソリューションイノベータ
NEC Solution Innovator, Ltd.

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

で，納期に対するプレッシャーから行われる不正だけでなく，この他の各種不正の予防，早期発見が可能になり，不正抑止効果の高い仕組みが完成する．

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“読み書きし易さ”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ーインターフェース(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インターカムを用いて，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 を取り入れることで効率化している．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ている．「見える化」の観点でも音声対話型 AI 帳票は，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成功している．たとえば「いつ」の情報を検査データに付随させることができるため，標準作業手順に潜在する作業時間のバラツキの可視化できる．これにより，更なる効率化の手がかりを得ることができる．

音声対話型 AI 帳票に基づけば，「効率化」と「見える化」のループを作ることができると期待する．しかし，「見える化」の観点で「いつ」の情報だけでなく，「誰により」と「どこで」の情報の取得できれば，さらなる効率化，また，本稿での課題である不正抑止効果が高い検査の実現が可能になる．

たとえば，「誰により」の情報が得られることで，作業者毎の作業時間のバラツキの可視化ができるようになり，効率化の手がかりが得られる．また，従来の音声対話型 AI 帳票では防ぐことのできない，成りすましの防止も可能になる．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．また，検査データの責任の所在が明確になるため，作業員が不正をする心理的障壁を上げることができる．不正抑止の観点から「誰により」の情報は常時判定できる必要がある．ログイン ID や指紋認証のような瞬間的な判定では作業者のすり替えが可能になり，成りすまし防止には役に立たない．これらの認証を作業項目毎に作業者に実施させれば，常時判定が可能になるが，検査点検の効率性を低下させてしまう．

また，「どこで」の情報が得られることで，作業者の動線分析に役立ち，効率化の手がかりとなる．さらに，作業者が検査場所に身を置く必要性が生じるため，検査の未実施，省略といった不正を防止できる．

「誰により」と「どこで」の情報を，NUI を維持した形で取得できるヒアラブルデバイスに置き換えた音声対話型 AI 帳票を提唱する．本稿で対象とするヒアラブルデバイスは，耳音響認証技術と高精度屋内位置測位技術を有する耳着用型のウェアラブルデバイスである[5]．耳音響認証技術[6]は，ヒアラブルデバイスが備える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に向いたマイクによって集音し，この反響音から算出する外耳道特性により，個人を特定する技術である．この技術によれば，「誰により」の情報を取得するために，作業員は検査音を聞くだけでよいため効率性を維持した形で常時取得が可能になる．一方，高精度屋内位置測位技術[7]は，地磁気センサを利用することで，GPS による位置測位のできない屋内において，精度 2 m 程度で「どこで」の情報を常時取得できる．ビーコンを用いるものや Wi-Fi 電波を用いる位置測位のように特別な設備を導入する必要はない．

音声対話型 AI 帳票のインターカムのヘッドセットをヒアラブルデバイスに置き換える上で，従来と同等の音声認識精度を実現することが課題となる．NUI の機能性を左右する要素は音声認識である．現在，ヘッドセットを用いた認識精度は 95 %程度であり，この精度での音声対話が作業効率化に貢献している．ヘッドセットをヒアラブルデバイスに置き換えた場合，耳認証で使うマイク，つまり，作業者の耳内で集音した発声を認識しなければならない．耳内で集音した発声の音響特徴はヘッドセットで集音した発声と大きく異なる．従って，音声対話型 AI 帳票で採用している音声認識エンジン，VoiceDo[8][9]の音響モデルをヒアラブルデバイスで集音した音声の音響特性に合わせる必要がある．本稿では，低コストで音響モデルを認識対象の音響特性に適応する話者適応技術を，話者ではなく，ヒアラブルデバイスで集音した音に適応できるように改良した多段階話者適応方式を提案する．

本稿では，第 2 章においてヒアラブルデバイスに関して説明する．第 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．第 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し，第 5 章において多段階話者適応の効果を音声認識評価により示し，音声対話型 AI 帳票にヒアラブルデバイスを NUI の機能性を低下させないで組み合わせられることを示す．

2. ヒアラブルデバイス[5]

本稿で対象とするヒアラブルデバイスは，耳着用型のウェアラブルデバイスである．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」ことの両立を可能とするものであ

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

る．ヒアラブルデバイスはスピーカ，マイク，地磁気センサ，加速度センサ及び，ジャイロセンサで構成される．スピーカ，マイクは，「誰により」を測定する耳音響認証技術に用いられ，地磁気センサは，「どこで」を測定する高精度屋内位置測位技術に用いられる．加速度センサとジャイロセンサは，ユーザの活動量計測，姿勢計測，歩行者自立航法に用いられる．これにより，ヒアラブルデバイスによれば，「誰により」と「どこで」の情報の他に「作業者がどんな状態か」の情報もすることができる．本章では，本稿で注目しているヒアラブルデバイスの「誰により」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[6]

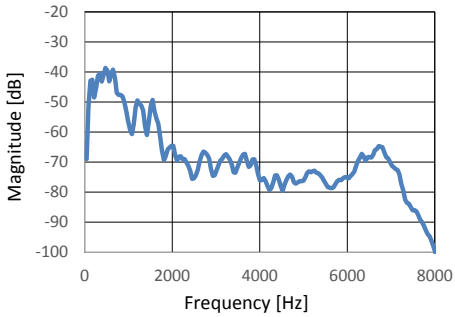
耳音響認証技術は，ヒアラブルデバイスのスピーカから耳内に向けて再生する検査音の反響音を，耳内を向くマイクにより集音する．集音した反響音から算出する外耳道音響特性により，個人を特定する技術である．外耳道音響特性には個人差があることが確認されており[10,11]，本技術により現在，他人受け入れ率 0.01 ~ 0.1 %で本人棄却率 2~3 %と，ユーザ認証として実用的な精度を確認している．耳音響認証技術によると，ユーザは個人認証のために検査音を聞くだけでよく，指紋認証や虹彩認証のようにユーザに特別な手続きを要求しない．このため，任意のタイミングで認証を行えるため，耳認証音響技術は「誰により」の常時判定を可能とする．

2.2 高精度屋内位置測位技術[7]

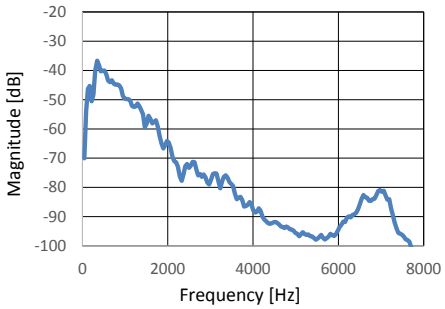
高精度屋内位置測位技術は，地磁気センサを利用することで，GPS によって位置測位のできない屋内において，精度 2 m 程度の測位を実現する．この技術は，屋内に存在する鉄骨などの影響による地磁気の乱れのデータを事前に測定することで，屋内位置測位を実現する．屋内位置測位技術には，従来，ビーコンを用いるものや Wi-Fi 電波を用いるものがある．しかし，本技術は，従来技術のように位置測位用の特別な設備を屋内に配置する必要がない特長を有する．

2.3 ヒアラブルデバイスを利用した音声対話型 AI 帳票

音声対話型 AI 帳票[3]は，長時間利用でも疲労が少ない軽量インターカムを用いて，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 を取り入れることで作業の効率化に成功している．一方，「見える化」の観点では，「いつ」の情報だけでなく，「誰により」や「どこで」の情報が取得できれば，さらなる効率化や，本稿での課題である不正抑止効果が高い検査の実現が可能になる．そこで本稿では，「誰により」と「どこで」の情報を，NUI を維持した形で取得できるヒアラブルデバイスに置き換えた音声対話型 AI 帳票を提案する．

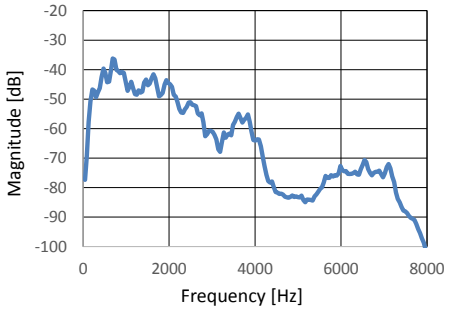


(a) ヘッドセットで集音

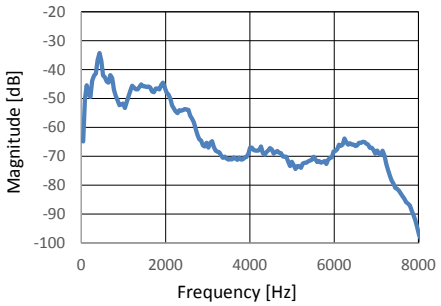


(b) 外耳道内部で集音
(5kHz 付近を中心に谷ができる)

図 1 ヘッドセットと外耳道内部で集音したときの音声スペクトルの比較



話者 B



話者 C

図 2 外耳道内部で集音した音声スペクトルの話者による違い
(話者 A(図 1(b)を参照)，話者 B(上)，話者 C(下))

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

ヒアラブルデバイスに置き換えた場合，音声認識はヒアラブルデバイスのマイクを利用することになる．しかし，ヒアラブルデバイスのマイクで集音した音声は，ヘッドセットマイクで集音した音声と音響特徴が異なる．これは，ヒアラブルデバイスのマイクは耳音響認証で使われるものであり，外耳道内部の音を集音しているためである．ヘッドセットマイクで集音した音声はユーザの口元から空気を伝わって計測される一方，ヒアラブルデバイスのマイクのように外耳道内部で集音した音声は，ユーザの体内を通過して計測される.この違いが音響特徴の違いを生じさせる．

図 1 は, ヘッドセットマイクで集音したとき(a)と外耳道内部で集音したとき(b)のスペクトル形状の違いを示す．外耳道内部で集音した音声スペクトルには 5 kHz 付近を中心とする谷ができていることがわかる．また図 2 は，他の話者 2 人の外耳道内部で集音した音声スペクトルを示す．これによると，話者によって谷の深さ，谷の位置が異なることもわかる．

このように，外耳道内部で集音した音声は通常のマイクで集音した音声とは異なる．このため，音声対話 AI 帳票にヒアラブルデバイスを組み合わせる上で，この音声の差異を吸収する音声認識技術が必要になる．

3. VoiceDo

音声対話型 AI 帳票の NUI 実現のカギの一つは音声認識技術である．我々は音声認識エンジンに，セリ市場やコールセンターなど高騒音の現場 200 か所以上で採用実績[12][13]がある VoiceDo[8][9]を採用し，AI 帳票のアイズフリー・ハンズフリー対話を実現している．

本章ではまず，音声認識技術及び，音声認識精度を改善する一手法として研究・開発されている話者適応技術の説明をする．その後で，ヒアラブルデバイスで集音した発声を VoiceDo で認識する際の問題点を指摘する．

3.1 音声認識の仕組み

音声認識は，あらかじめ用意した多量の音声データから作成する各音素の音響特徴を記録する音響モデルと，認識単語と単語間のつながり方やつながり易さを記録する言語モデルを参照することで，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，それぞれ特徴量毎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，つながり易さに基づいて，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．

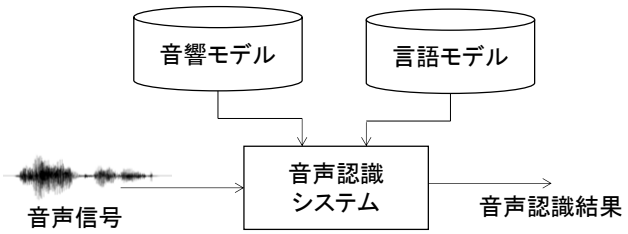


図 3　音声認識のシステム構成図

音声認識の精度は，音響モデルと言語モデルが認識対象の音声とその発声内容に，いかに一致しているかに依存する．音声対話型 AI 帳票では，用途に合わせて語彙を設定するため, 言語モデルの認識精度への影響は小さい. しかし，様々な現場における様々な話者の発声を想定しているため, あらかじめ用意した音響モデルが，環境や話者の声質の音響特性をカバーしきれない可能性がある．そのため，音響モデルと認識対象の音声との差異をいかに埋めるかが課題になる．

音声対話型 AI 帳票で採用する音声認識エンジン VoiceDo は，音響モデルと認識対象の音声との間の差異を埋めるため，認識エンジンに入力される発声から加算性雑音を取り除く雑音抑圧機能と，認識対象の話者の声に音響モデルを適応する話者適応機能を備える．これにより，VoiceDo は様々な騒音環境で様々な話者において，高い認識性能を実現している．

次に，ヒアラブルデバイスとの組み合わせにおいてカギになる話者適応技術について説明する．

3.2 話者適応技術

話者適応は，認識対象の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．話者適応技術は，発声内容を示す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には，ラベルファイルが不要であるため，話者の自由な発声を適応に用いることができる．その反面，たくさんの発声を必要とする．逆に，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，業務効率化の観点で，少量の発声でも話者適応できることが望まれる．VoiceDo は教師あり話者適応を採用しており，この目的に適している．

教師あり話者適応では，適応する話者の発声とラベルファイル及び，適応対象の音響モデルが必要である．適応ではまず，ラベルファイルに記載された発声内容の音素列及び，適応対象の音響モデルの音素特徴を参照することで，話者の発声から抽出する特徴量列の各特徴量について，それらに相当する音素の対応づけを行う(図 4)．音素毎にこ

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

れらの特徴量をまとめあげ，音響モデルが記録する音素特徴を作り直す．これが適応音響モデルとなる．

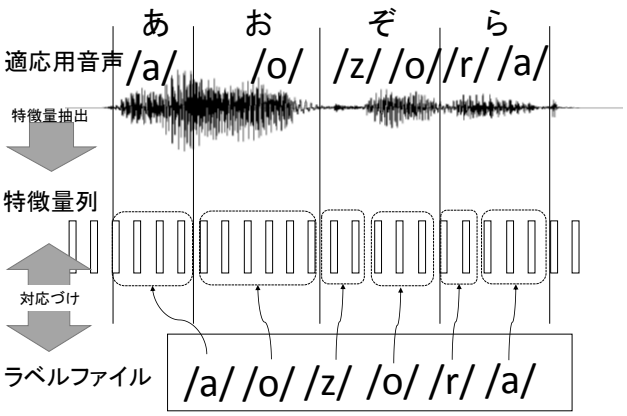


図 4　話者適応における音素対応づけ(成功例)

なお，VoiceDo の音声認識では，認識速度向上のために各音素の音響特徴を木構造化して記録する音響モデルを用いている[14]．この木構造音響モデルに着目し, VoiceDo が備える話者適応に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方式，自律的モデル複雑度制御法(Autonomous Model Complexity Control; AMCC)を採用している[15]．AMCC により，通常不利とされる適応データ数が少ない場合にも認識精度が改善される．

AMCC の知識は，本稿の提案法，多段階話者適応の次章での説明で必要となるので，ここで AMCC を概説する．まず，木構造音響モデルを説明する．木構造音響モデルにおける，各リーフノードは音響モデルが記録する各々の音素特徴に相当する．これらのノードの音響特徴をクラスタリングして得られるいくつかのクラスタが，リーフノードの親ノードになる．これらのノードはクラスタに属するノードの音響特徴の共通特徴を保持する．さらに，ノードのクラスタリングを繰り返すと，最終的に，全音素の共通特徴を保持するルートノードが生成される(図 5).

次に，この木構造音響モデルの各ノードに関する話者適応を説明する．ここに，適応データとして一発声の音声データがある．この音声データを用いて，ルートノードを適応する場合，この一発声から得られる全ての特徴量を適応に使うことができる．しかし，リーフに近いノードの適応程，一発声から得られる特徴量が減っていく．一般的に，各ノードの音響特徴を適応する場合，適応データが少ないほど，適応精度は劣化する．そこで，AMCC は，適応データ量に応じて，適応するノードの階層を制御している．たとえば，あるノードを適応するデータ数が集まらない場合は，その親ノードの適応に留めるといった制御である．したがって, AMCC によれば, 適応データ数が多い場合には，リーフに近いノードが適応される．これは，音響モデルの細かい適応になる. 一方, 適応データ数が少ない場合には，ルートに近いノードが適応される．これは音響モデルの大雑把な適応になる． このように，AMCC によると，適応データが少ない場合にも適応による劣化を防止し，精度改善を望める．

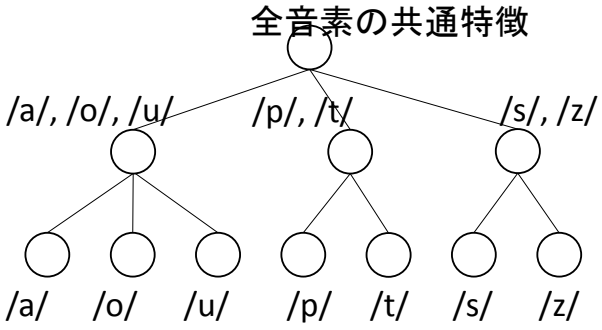


図 5　木構造音響モデル

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 6 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，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．

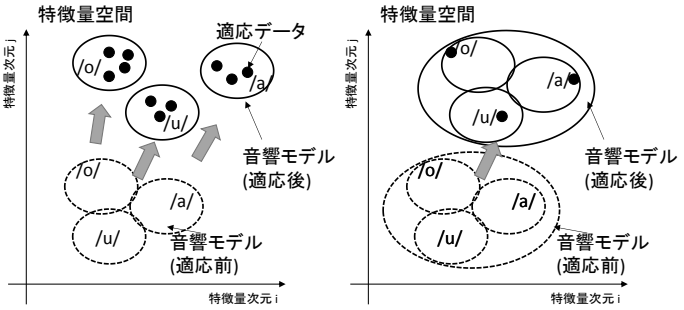


図 6　AMCC における音響モデル適応の概要
(左：適応データが多い場合→細かい適応,
右：適応データが少ない場合→大雑把な適応)

3.3 VoiceDo とヒアラブルデバイスの組み合わせにおける課題

2.3 で説明したように，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上で，音声認識は外耳道内部で集音した音声に対応する必要がある．しかし，VoiceDo の音響モデルはヘッドセットマイクで収録した音声で学習されているため，ヒアラブルデバイスで集音した音声を VoiceDo で認識させるのは難しい．この問題を解決するために，まず考えられる方法は VoiceDo の話者適応である．しかし，従来の VoiceDo の話者適応方式 AMCC では，多少の改善効果は期待できるものの, 音声対話型 AI 帳票で NUI の機能性を維持する 95%の認識精度は難しい．これは，ヒアラブルデバイスで集音した音声が音響モデルの音響特徴と大きく異なるため，図 7 に示すような，音素の対応づけ

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

の失敗が起こりやすくなるためである．音素の対応づけに失敗すると，特にリーフに近いノードになるほど，間違った音素に適応してしまう．これにより，話者適応本来の性能を発揮できなくなってしまう．そこで，本課題を解決する方式として，本稿で提案する多段階話者適応を次章にて説明する．

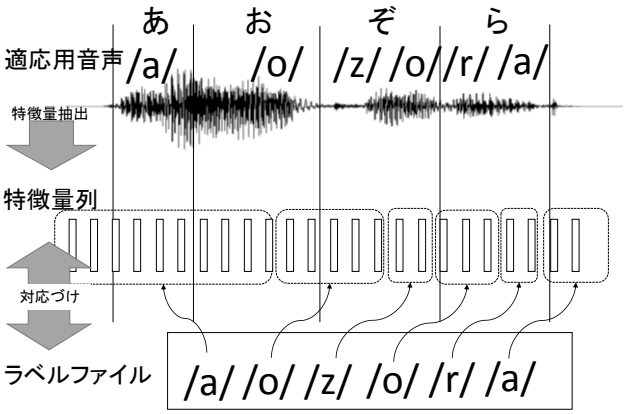


図 7　話者適応における音素対応づけ(失敗例)

4. 多段階話者適応

本稿で提案する多段階話者適応は，AMCC による話者適応を，適応データ数を増やしながら繰り返す方式である．繰り返しの初期段階では適応データ数を少なく絞るため，AMCC の効果により大雑把な適応になる．これにより，ヒアラブルデバイスで集音した音声が音響モデルと音響特徴が大きく異なる場合の，音素対応づけの失敗を吸収できる．繰り返しを続けるに従い，音響モデルの音響特徴がヒアラブルデバイスの音声に徐々に近づいていく．また，適応データ数も増やしていくため，正しい対応づけの下，AMCC の効果により細かい適応ができるようになる．これにより，適応する音声が適応する音響モデルと音響特徴が大きく異なる場合でも，精度よく話者適応ができるようになる．

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 8 に示す．本稿では適応データ数は，事前に実施する音声認識の正解数とした．つまり一回の話者適応で使う適応データは事前の音声認識で正解したものとした．

適応の繰り返しの様子を特徴量空間で眺めたものを図 9 に示す．図 9 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，適応データ数が少なく，大雑把な適応になる(図 9 左)．一方，適応回数が増えると，適応データ数が増え，細かい適応になる(図 9 右)．

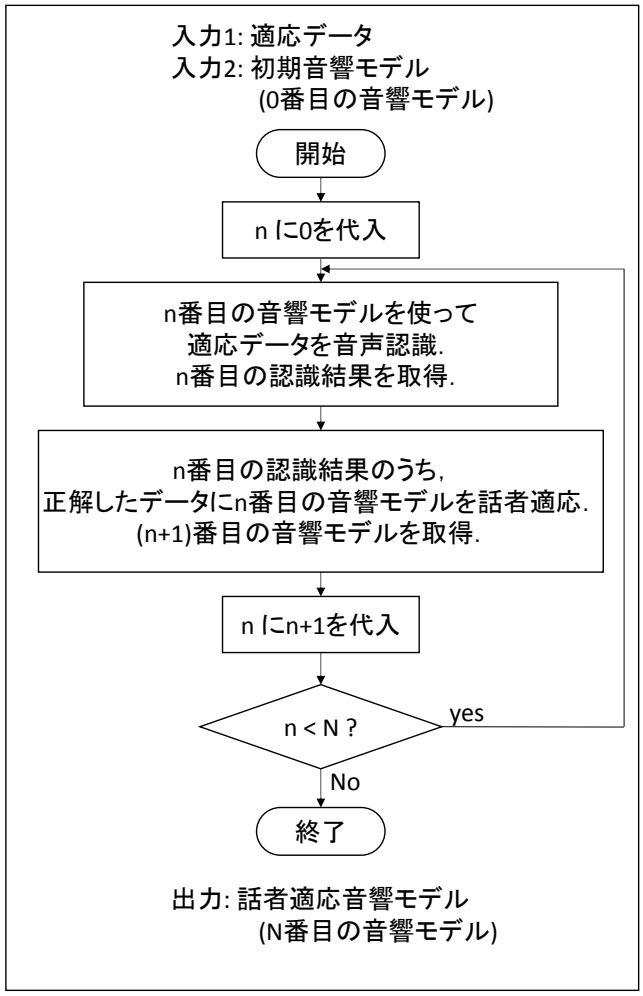


図 8　N 段階話者適応のアルゴリズム

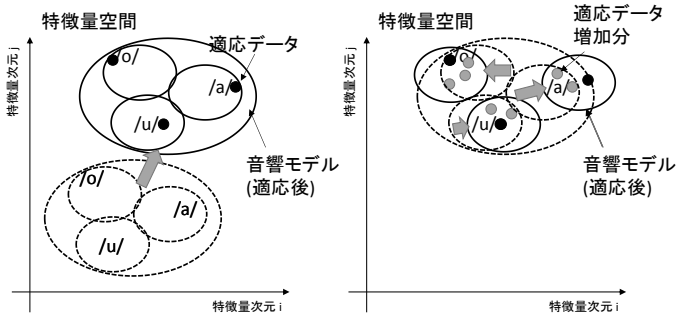


図 9　多段階話者適応における音響モデルの適応過程
(左：n 段目の適応，右：n+1 段目の適応)

5. 評価

外耳道内部で集音した音声を用いて，提案法の多段階話者適応を評価する．音声は三人の男性声を収録した(話者 A，話者 B，話者 C)．これらの話者は，2.3 章における図 1，図 2 に示すスペクトルの話者に相当する．適応用データは音

Vol.2018-GN-103 No.3
Vol.2018-CDS-21 No.3
Vol.2018-DCC-18 No.3
2018/1/26

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，男性三人の平均単語認識精及び三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(1 段)，提案法(3 段)の認識精度を示す．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．従来法の適用だけでは達成できていない，音声対話型 AI 帳票の NUI の機能性を下げない 95 %の認識精度を，提案法の適用により三話者全員について達成できている．

表 1　ヒアラブルデバイス収録音声の単語認識精度

	未適応	従来法(1 段)	提案法(3 段)
三話者平均	74 %	93 %	97 %
話者 A	36 %	81 %	95 %
話者 B	92 %	97 %	97 %
話者 C	95 %	100 %	100 %

以上より，提案する多段階話者適応により，音声対話型 AI 帳票に NUI の機能性を維持しつつ，「いつ」の情報だけでなく，「誰により」と「どこで」の情報を付加することができるようになった．すなわち，不正抑止効果の高い音声対話型 AI 帳票の実現の見通しを得た．

6. おわりに

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「誰により」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み合わせを提唱した．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．音声認識評価により，多段階話者適応方式で音声対話型 AI 帳票の NUI の機能性を低下させない単語認識精度 95 %を達成することを確認した．これにより，音声対話型 AI 帳票に「誰により」と「どこで」の情報を付加することができるようになり，不正抑止効果の高い音声対話型 AI 帳票の実現の見通しを得た．

今後，提案した多段階話者適応を音声認識エンジン VoiceDo の話者適応機能に組み込む．その後，実際の現場にヒアラブルデバイスを組み合わせた不正抑止効果の高い音声対話型 AI 帳票の導入を進め，「効率化」と「見える化」のループを作っていく．そのループから課題を洗い出し，将来さらなる改善を目指す．これにより不正抑止効果の高い仕組みの完成を目指す． また，本稿で対象とするヒアラブルデバイスでできるユーザの活動量計測，姿勢計測，歩行者自立航法を活用した検査点検の検討も行う．

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”, 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” 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 Vol.2017-MLB-84 No.2. Vol.2017-CDS-20 No.2, 2017/8/29.
[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” 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”, http://jpn.nec.com/techrep/journal/g17/n01/170110.html, (参照 2017-12-04)
[6] 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. 2016/03.
[7] "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～", http://jpn.nec.com/press/201610/20161028_02.html, (参照 2017-12-04)
[8] 塚田 聡, 耐雑音音声認識装置 VoiceDo, NEC 技報 Vol.63 No.1 http://jpn.nec.com/techrep/journal/g10/n01/pdf/100118.pdf, 2010/2
[9] 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, 2013/10/15.
[10] A. H. M. Akkermans, T.A. M. Kevenaar, and D. W. E.Schobben. Acoustic ear recognition for person identification. Proc AutoID'05, pp. 219-223, 2005.
[11] S. Yano, H. Hokari, and S. Shimada. A Study on the Personal Difference in the Transfer Functions of Sound Localization Using Stereo Earphones. IEICE Trans. Funda, vol.E83-A,No.5. pp.877-887, 2000.
[12] "VoiceDo 活用シーン", 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), http://www.itmedia.co.jp/enterprise/articles/0909/15/news039.html, (参照 2017-12-04)
[14] T. Watanabe, K. Shinoda, K. Takagi, and E. Yamada, "Speech Recognition using tree-structured probability density function, " Proc. Of ICSLP94, pp.223-226, 1994.
[15] 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, 1996/12