社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2004－ＨＩ－109 (21)
2004－SLP－ 52 (21)
2004／7／17

日英シームレス音声認識技術による航空管制音声認識

小川 厚徳†, 今村 明弘†, 外海 実††, 中村 誠††, 磯部 俊洋†††, 菅原 昌平†††

†日本電信電話株式会社 NTT サイバースペース研究所
株式会社 NTT データ ††第一公共ビジネスユニット, †††技術開発本部

あらまし 本稿では，日英シームレス音声認識技術による航空管制音声認識の検討について報告する．航空管制音声は，航空機の安全な運航を確保するために，航空管制官とパイロットとの間で専用の無線電話を用いてやり取りされる飛行指示等に関する音声であり，世界共通の用語および文型規則に従って発声されている．今回，日英シームレス音声認識技術を用いた航空路管制用シミュレータを新規開発し，日本人英語発声の典型例と考えられる日本国内における管制官の航空路管制発話を対象に，BNF 文法を用いた認識実験による性能評価を行った．実験から，日英シームレス音声認識によれば従来の日本語または英語専用音声認識方式に比べ，高い認識性能が得られることが明らかとなった．また認識結果の分析を通じて，認識辞書中の各単語に付与されている日本語風英語発音またはネイティブ英語発音のいずれかを，発声の日本語らしさまたは英語らしさに応じて適切に選択して認識を行うことで，高い認識精度が得られることが分かった．

Japanese-English Bilingual Speech Recognition of Voice Command in Air Traffic Control Communication

Atsunori Ogawa†, Akihiro Imamura†,
Minoru Tonogai††, Makoto Nakamura††, Toshihiro Isobe††† and Shohei Sugawara†††

†NTT Cyber Space Laboratories, NTT Corporation
††First Public Administration Business Unit and †††R&D Headquarters, NTT Data Corporation

Abstract In this paper, we describe a study of Japanese-English bilingual speech recognition of voice command in Air Traffic Control (ATC) communication. Air traffic controllers communicate with airplane pilots via dedicated radio voice channel according to the world common phraseology and grammatical rule in order to maintain smooth and safe air traffic flows. We developed a new enroute ATC simulator system by using Japanese-English bilingual speech recognition technology, and carried out an assessment of the recognition performance through BNF grammar based recognition experiments of Japanese enroute air traffic controller's voice commands that are the typical examples of "English spoken by Japanese". In the comparative experiments, Japanese-English bilingual speech recognition shows higher performances than the both of conventional Japanese-specific and English-specific speech recognition method. We obtained that an appropriate selection of dictionary word's pronunciation style between Japanese like and native like one according to the trend of the speaker's pronunciation characteristics contributes to high recognition accuracy.

1 はじめに

近年，日本人の英語発話に対して音声認識技術を適用するという試みが盛んに行われている．その目的は，大きく 2 つに分けられる．一つは，日本人の英語発音の評価や修正・習得方法の教示の自動化を目的とするものである[1,2,3,4]．もう一つは，英語を母国語としない日本人話者の英語発声を高精度に認識し，その応用開発の促進を図ることを目的とするものである[5,6]．我々は後者の立場に立ち，単独の音声認識エンジンで，日本人の日本語発話および英語ネイティブ話者の英語発話の認識精度を維持しつつ，日本人の英語発話を高精度に認識する日英シームレス音声認識技術の検討を進めてきた[7,8,9,10,11,12]．

今回我々は，この日英シームレス音声認識技術を，日本人英語発声の典型例である国内の航空管制官による航空路管制発話の認識に適用した．実験の結果，十分に実用に耐える認識精度と処理速度が得られることを確認した．

2 航空管制と取り組みの概要

日本における航空管制(ATC: Air Traffic Control)の概要と，検討の対象とする航空路管制音声の特徴について述べる[13,14,15,16]．さらに，今回新規開発した音声認識機能付き航空路管制用シミュレータについて説明する．

2.1 日本における航空管制の概要

航空機が近代化され，様々な計器が装備されるようになるのと並行して，地上の航行援助無線施設(NAVAID と呼ばれる)の整備が発達したことで，パイロットは地上の目標物を目視しなくても，操縦席の計器を頼りに飛行できるようになった．日本の航空法では，このような計器のみに依存して行う飛行を「計器飛行」と定義している．また，常時，航空管制官(Air Traffic Controller)の指示に従って行う飛行を，目視に基づく「有視界飛行方式(VFR: Visual Flight Rule)」に対して，「計器飛行方式(IFR: Instrument Flight Rule)」と定義している．今日では，日中の好天候下における軽飛行

機の飛行や軍用機の飛行などの特殊なケースを除いて，民間の定期便航空機のほとんどがIFRで飛行しており，管制の対象になっている．

航空管制は，管制対象とする空域によって，2つに大別される(図1)．一つは，空港とその周辺の空域を対象としている「ターミナル管制(Terminal Control)」である．もう一つは，地上の要所毎に設置されている航行援助無線施設や地上および海上に設けられた地点(NAVAIDも含めFIXと総称される)を結んだルート，即ち航空路を対象としている「航空路(エンルート)管制(Enroute Control)」である．今回，我々は航空路管制において管制官からパイロットに向けて送出される管制音声を検討の対象とした．

国内における航空路管制は，国土交通省航空局に属する航空交通管制部(ACC: Area Control Center)が実施しており，札幌，東京，福岡，那覇の各管制部が日本および周辺海域上空を4つに区分して管制業務に当たっている．各管制部の担当管轄空域は，さらに細かい空域(セクタと呼ばれる)に区分され，それぞれにレーダ画面を備えた管制卓が設けられ，24時間体制で管制官が配置されている．

管制官は，管制卓のレーダ画面を見ながら，担当するセクタを飛行する複数の航空機の位置関係を把握・予測し，各パイロットに飛行高度，速度，方位などの飛行制御や管制機関との通信設定などに関する適切な指示(管制指示)を，VHFやUHFの専用周波数の無線電話を用いて送出する．このとき管制官は，専用のヘッドセットマイクとPush-to-Talkボタンを使用し，管制官の管制指示に応答するパイロットと交互に，半2重での音声通信を行っている．

2.2 航空路管制音声の特徴

管制官とパイロットの交信で用いられる用語は国連の専門機関の一つである「国際民間航空機関(ICAO: International Civil Aviation Organization)」が作成した文書("DOC.4444/RAC501, Procedure for Air Navigation Services, Rules of the Air and Air Traffic Services")で定められており，各国ともこの文書に準拠した標準用語を用いている．日本においても，この文書に基づいて国土交通省が監修した管制方式基準があり，これによって定められた標準用語および標準発話文型が使用されている．また航空管制では各国とも英語の管制用語を使うことが原則となっており，日本でも管制方式基準において原則として英語を使うことを規定している．

管制官とパイロットの間の英語による交信では通常の英会話とはかなり異なる特殊な用語が多く使われている．図2に，管制官とパイロットの航空路管制における交信の一例を示す．航空路管制で用いられる単語としては，"ANA"，"JAL"等の航空会社名や，便名(コールサイン)，飛行高度・速度・方位，通信周波数等を表す数字や，"Tokyo(ACC)"や"Toyota (NAVAID)"等の日本の地名などがある．

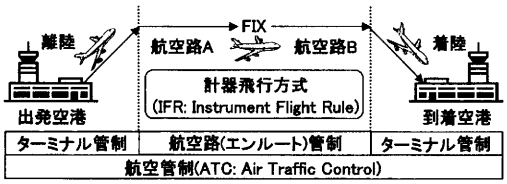

図1 計器飛行方式(IFR)における航空管制の概要

無線電話通信において雑音や混信がある場合などに聞き取りづらくなる可能性がある数字などには特殊な発音も使用されており，例えば，"9"は"ナイナー"，"thousand"は"タウザンド"と発音されるなどの例がある．また，"climb(descend) and maintain flight level [高度]"や"cleared via present position direct [地名]"など，いくつかの標準発話文型が組み合わせて用いられる場合もある．さらに，航空路管制業務に当たる管制官は国家公務員であることから，そのほとんどが英語を母国語としない日本人である．従って，その発話にしばしば登場する日本の地名部分はもちろんのこと，一般の英単語部分にも日本語風の発音が含まれる場合がある．

このような特徴を有する航空路管制発話の認識には，日本人英語発話の高精度認識が可能な日英シームレス音声認識技術が有効と考えられる．また発話が，管制方式基準で規定された文型に従って行われるということから，そのほとんどはBNF等のネットワーク文法で記述可能である．

2.3 音声認識機能付き航空路管制用シミュレータ

管制官の訓練に用いられる航空路管制用シミュレータは，実際の航空路管制卓と同様に，担当セクタ中の複数の航空機の位置関係を表示する模擬レーダ画面を備えている．訓練では訓練生である管制官と教官に，それぞれ専用の操作卓が用意され，同一セクタの模擬レーダ画面を見ながらシミュレータ操作を行う．航空機のパイロット役を兼ねる教官卓には，シミュレーション飛行するセクタ中の各航空機を制御するコマンドが入力可能となっている．

図3中の既存シミュレータを用いた訓練では，教官は，(1)訓練生の管制発話を聞き取り，(2)それを基に航空機の

管制官：ANA 347, Tokyo control, climb and maintain flight level 240.
(全日空347便，こちら東京管制部，上昇して高度24,000フィートを維持せよ)
ANA347: ANA 347, climbing to flight level 240.
(こちら全日空347便，24,000フィートに上昇する)
管制官：JAL 936, cleared via present position direct Toyota, rest of route unchanged.
(日本航空936便，豊田に直行せよ．その後の経路は変更なし)
JAL936: Direct Toyota, JAL 936.
(豊田へ直行する．こちら日本航空936便)
管制官：KAL 514, turn left heading 200, reduce speed to 190 knots, then descend and maintain flight level 170.
(大韓航空514便，磁方位200度に左旋回し，速度を190ノットまで落とした後，17,000フィートまで降下せよ)
KAL514: KAL 514, heading 200, descend to 170, speed 190.
(こちら大韓航空514便，磁方位200度で17,000フィートまで降下，速度は190ノットに落とす)
管制官：ANA 336, contact Fukuoka control, 133 decimal 15.
(全日空336便，周波数133.15MHzで福岡管制部と交信せよ)
ANA336: ANA 336, 133 decimal 15.
(こちら全日空336便，133.15MHzに切り替える)

図2 航空管制官とパイロットの交信例

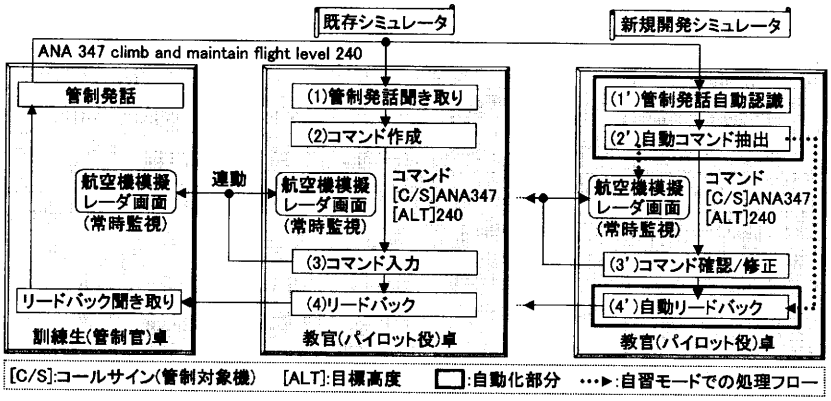

図 3 既存および新規開発航空路管制用シミュレータにおける処理フロー

飛行制御コマンドを作成し，(3) 作成したコマンドをキーボードで入力し，最後に，(4) 訓練生に対してリードバック(復唱=パイロットの応答)を行う．次々と発声される訓練生の管制発話に対応して，教官はこのようなシミュレータ操作を素早く行う必要があり，大きな作業負担となっていた．

今回我々は，このような教官の作業負担を削減し，より効率的な訓練フローを実現することを目的として，日英シームレス音声認識技術を用いた航空路管制用シミュレータを開発した．具体的には，図 3 中の新規開発シミュレータにおいて，(1') 訓練生の航空路管制発話の認識と，(2') 認識結果からのコマンド抽出を日英シームレス音声認識により自動化し，教官は，(3') 自動抽出されたコマンドを確認して正解であれば OK ボタンを押し，誤っている場合にのみキーボードで修正すれば済む構成とした．また (4') リードバックも音声合成によって自動化した．さらに，音声認識による自動コマンド抽出の精度が高い場合には，(3') の教官による確認・修正なしに，飛行制御コマンド自動確定と (4') 自動リードバックまでを行うことができる訓練生の自習モードも用意した．

さらに，訓練時に音声認識機能を最も良い状態で利用できるよう訓練前に訓練生が自らの声量レベルにシステムを合わせこむことができるマイク音量自動調整機能と話者適応(後述)機能も実装した．また，訓練生はヘッドセットマイクを装着して Push-to-Talk ボタンで音声入力を行う方式を採用しており，実際の航空路管制卓での業務時とほぼ同様のインタフェース仕様となっている．

3 日英シームレス音声認識技術の航空路管制音声への適用

日英シームレス音声認識技術[7,8,9,10,11,12]は，文献[5,6]のように日本人の英語発声データベースを用いて構築されるものではなく，既存の日本人の日本語発声データベースを用いて構築された日本語専用音声認識技術と英語ネイティブ話者の英語発声データベースを用いて構築された英語専用音声認識技術を統合したものであり，単独の音声認識エンジンで，日本人の日本語発話および英語ネイティブ話者の英語発話の認識精度を維持しつつ，日本人英語発話の高精度認識を可能にする．以下では，航空路管制音声へ特化した点を中心に，日英シームレス音声認識技術の各構成要素を説明する．

3.1 発音辞書

各単語には，日本語風英語発音(カタカナ表記)とネイティブ英語発音(SAMPA : Speech Assessment Methods Phonetic Alphabet[17]準拠の表記)が併記されており，認識時にはこれらが音素列に変換される．基本の音素体系は，日本語 31 音素および英語 43 音素であり，両者を単純に統合して日英シームレス音素体系を構成する．ただし，無音については日英で共通化するため，73 音素となる．

表 1 に，実験で用いた航空路管制音声認識用日英シームレス発音辞書の記述例を示す．表1には，発音定義として，カタカナ表記による日本語風英語発音と SAMPA 準拠表記によるネイティブ英語発音が示されており，それぞれの発音の後ろに続く括弧内には，それらを変換して得られる音素列も表示されている(日本語音素体系に由来する音素列には "J" を，英語音素体系に由来する音素列には "E" を，それぞれ先頭に付与している．以降の音素列および音素モデルの表記も同じ)．"maintain" などの一般的な英単語に対しては，管制官により発音の仕方は様々であると考え，ネイティブ英語発音としてはネイティブ話者の通常の発音を付与し，日本語風英語発音としては日本人のいわゆるカタカナによる読み表記を発音として付与した．

一方; "Toyota" などの日本の地名に対しては，管制官が日本人であるため発音の揺れは小さいと考え，まず，カタカナ表記による日本語風英語発音として日本語の通常の発音を付与し，それを変換して得られる日本語音素列を参

表 1 航空路管制音声認識用日英シームレス発音辞書の記述例

単語定義	発音定義
ANA	オールニッポン(J:o o r u n i q p o ng) 0-InIpAn(E:ao l n ih p aa n)
Toyota	トヨタ(J:t o j o t a) t+0I+oUt@(E:t oy ow t ax)
climb	クライム(J:k u r a i m u) kl+aIm(E:k l ay m)
and	アンド(J:a ng d o) &nd(E:ae n d)
maintain	メインテイン(J:m ei ng t ei ng) m+eInt+eIn(E:m ey n t ey n)
9	ナイン(J:n a i ng) ナイナー(J:n a i n a a) n+aIn(E:n ay n) n+aInR(E:n ay n er)
thousand	サウザンド(J:s a u z a ng d o) タウザンド(J:t a u z a ng d o) T+aUzInd(E:th aw z ih n d) t+aUzInd(E:t aw z ih n d)

表 2 日本語専用・英語専用・日英シームレス音響モデルの仕様

言語	日本語		英語		日英シームレス	
性別	男性	女性	男性	女性	男性	女性
状態数	2,198	2,191	3,492	3,496	5,687	5,684
学習データ	日本人男性の日本語単語および文章発声計40時間	日本人女性の日本語単語および文章発声計38時間	英語ネイティブ男性の英単語および文章発声40時間	英語ネイティブ女性の英単語および文章発声41時間	――――	
音素体系	日本語31音素		英語43音素		日英73音素	
音素モデル構造	3状態 混合ガウス分布 状態共有型 triphone および monophone					
音響分析条件	音声帯域：100-5200Hz，分析窓：ハミング窓，フレーム長/シフト：30ms/10ms					
特徴パラメータ	12次元MFCC + 12次元ΔMFCC + 1次元Δlogパワー ： 計25次元					

考に，近いと思われる英語音素列を作成し，さらにそれを逆変換したものをSAMPA準拠表記によるネイティブ英語発音として付与した．

3.2 音響モデル

日本語音響モデルと英語音響モデルを個別に学習した後，両者を統合して日英シームレス音響モデルを作成する．今回の実験では，性別依存モデルを用いた．表 2 に，各音響モデルの仕様について示す．なお，日本語音響モデルと英語音響モデルを個別に学習する段階で，それぞれ個別の無音 monophone モデルが得られるが，統合の際には，日本語音響モデルに由来する無音 monophone モデルのみを残した．

3.3 BNF 文法・音素モデルネットワークおよび探索処理

2.2 で述べたように，今回対象としている航空路管制音声は，BNF 文法でほぼ記述可能である．例として，図 4 に，模式化した航空路管制音声認識用 BNF 文法と，その一部である “climb and maintain” の単語境界部分を拡大した音素モデルネットワークの展開図を示す．

図 4 の音素モデルネットワーク展開図に示すように，単語境界において日英間の音素モデル遷移が可能である．このため，カタカナによる読み表記で発声された単語については日本語風英語発音を通る仮説に高いスコアが与えられ，ネイティブに近い発音で発声された単語については，ネイティブ英語発音を通る仮説に高いスコアが与えられるというように，単語単位での柔軟な日英発音（音素モデル）選択に基づく探索処理を行うことができる．

ただし，単語境界において日英間の発音の遷移を行う場合は，先行単語末および後続単語頭の音素モデルに triphone モデルが適用できず，精度の劣る monophone モデルが使用されることになる．このため，単語境界における日英間の発音の遷移があまりにも多い場合は，認識精度の低下をまねくという副作用も考えられる．

3.4 話者適応

2.2 や 3.1 で述べたように，航空路管制音声は特殊な用語および文型規則に基づく英語発声である．さらに管制官は日本人であることから，その発声の癖は多様である．また，表 2 に示したように，今回の実験で用いた日英シームレス音響モデルの学習データには，航空路管制音声は含まれていない．このような状況で，航空路管制音声を高精度に認識するためには，話者適応が不可欠であるため，教師あり話者適応を検討し，新規開発した航空路管制用シミュレータにもこの機能を実装した．

3.4.1 十分統計量の蓄積およびパラメータ更新方法

図 5 の例を用いて，日英シームレス音響モデルに対する十分統計量の蓄積およびパラメータ更新方法を説明する．

まず，話者の “9 thousand” という一つの発声に対して，表 1 の航空路管制音声認識用日英シームレス発音辞書を用いて，カタカナ表記による日本語風英語発音トランスクリプションとSAMPA準拠表記によるネイティブ英語発音トランスクリプションの 2 つを付与する．発音辞書に示されるように，“9” には，“ナイン” または “ナイナー”，“thousand” には，“サウザンド” または “タウザンド” という複数のカタカナ表記による日本語風英語発音が付与されているため，一旦，適応前の音響モデルを用いて発音を選択するための BNF 文法認識を行い，各単語に一つの日本語風英語発音が付与されるようにして，日本語風英語発音トランスクリプションを得る．ネイティブ英語発音トランスクリプションを得る場合も同様の処理を行う．また，この BNF 文法認識では，単語間の無音区間の検出も行い，必要であれば，トランスクリプションに無音記号を挿入する．

次に，上記 2 つのトランスクリプションをそれぞれ triphone モデル列および monophone モデル列に変換する．したがって，話者の一発声に対して適応すべき 4 つのモデル列が与えられることになる．これらに対する十分統計量の蓄積をそ

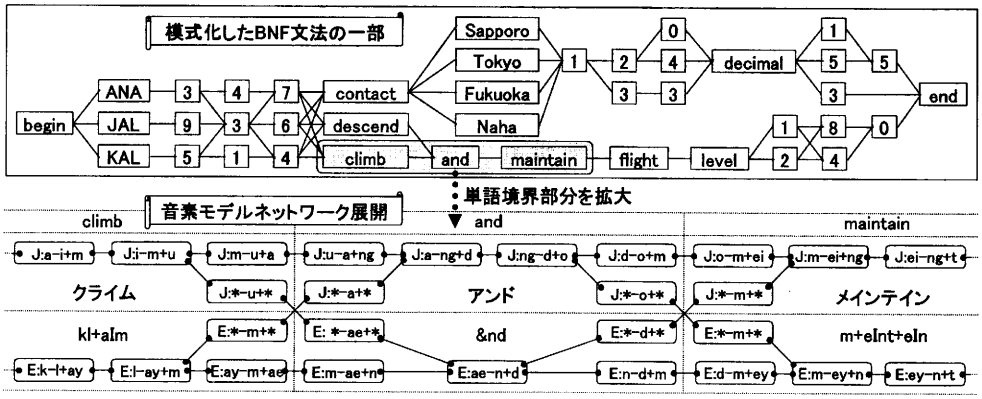

図 4 航空路管制音声認識用 BNF 文法と日英シームレス音素ネットワーク展開の例

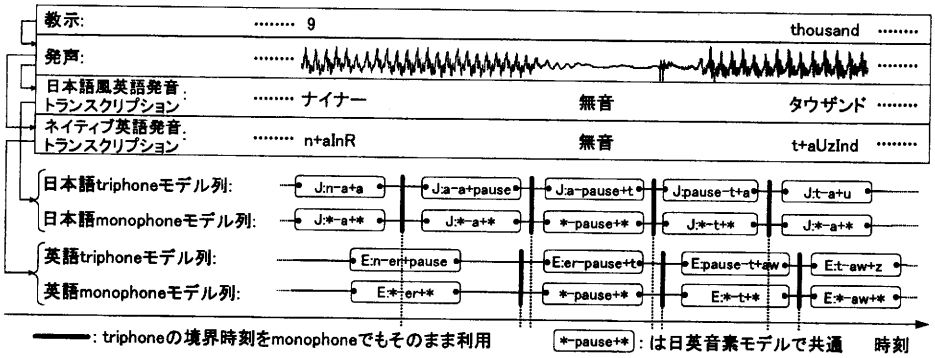

図 5 日英シームレス話者適応の十分統計量蓄積時におけるモデル境界時刻の取得例

れぞれ独立に行うことには大きな処理量を要するため，文献[18]の方法を用いて効率化を図った．具体的には，図 5 に示すように，triphone モデル列の十分統計量蓄積時に得られるビタビ状態アライメントからモデル境界時刻を取得し，それをそのまま monophone モデル列の十分統計量蓄積におけるモデル境界時刻として再利用する．この効率化により，4 つのモデル列の十分統計量蓄積をそれぞれ独立に行う場合と比較して，処理量を半分程度に抑えることができる．

以上のようにして 4 つのモデル列の十分統計量蓄積を行った後，全音素モデルのパラメータ更新を一斉に行う．日英シームレス音響モデルにおける日本語音響モデル部分と英語音響モデル部分のそれぞれにおける十分統計量の蓄積とパラメータ更新は独立して行われる．しかし，無音 monophone モデルに関しては，上記の 4 つのモデル列の全てに登場する可能性があるため，4 つのモデルにおいて蓄積された十分統計量を全て反映したパラメータ更新が行われる．今回の実験では，パラメータ更新アルゴリズムとして MAP 推定[19]を採用した．

3.4.2 段階的話者適応方式

新規開発した航空路管制用シミュレータでは，管制官の定期的な訓練を通して，段階的に話者適応を行うことにより，認識性能を維持・向上させることを想定している．図 6 に，訓練前に 1 セット 30 文章の適応用発声を毎回行うことを想定した段階的話者適応方式の流れを示す．

初回の話者適応では，適応用 30 発声を得て，まず，適応なし音響モデルに対して十分統計量蓄積を行い，30 発声分の十分統計量データを得る．次に，蓄積された 30 発声分の十分統計量を基に，適応なし音響モデルをベースとしたパラメータ更新を行い，30 発声適応音響モデルを得る．

二回目の話者適応でも，適応用 30 発声を得て十分統計量蓄積を行うが，これを初回で得た 30 発声分に上乗せする．この処理により，累積 60 発声分の十分統計量データを得る．その後，蓄積された 60 発声分の十分統計量を基に，適応なし音響モデルをベースとしたパラメータ更新を行い，60 発声適応音響モデルを得る．

三回目以降も同様に，それ以前に得たものに上乗せする形で十分統計量蓄積を行い，それを基に，常に話者適応なし音響モデルをベースとしたパラメータ更新を行うことで，話者適応あり音響モデルを得る．

このようにして得られた各段階の話者適応後の音響モデルを，訓練を行う管制官毎に個別に管理・運用することにより，定期的な訓練における認識性能の維持・向上が可能となっている．

4 航空路管制音声認識実験

以上のように航空路管制音声に特化した日英シームレス音声認識技術を，音声認識エンジン VoiceRex[20]に実装し，管制官による読み上げ航空路管制発話およびシミュレータ操作音声に対する認識実験を行い，評価した．

4.1 【実験 1】読み上げ音声の認識

4.1.1 実験条件

実験では，男女各 5 名の管制官が，指定された同一の航空路管制 273 文章をそれぞれ読み上げた音声を使用した．273 発声の内，155 発声(平均文長 14.0 単語，平均発声長 6.27 秒)を話者適応に，残りの 118 発声(平均文長 13.4 単語，平均発声長 5.88 秒)を評価に用いた．

話者適応用の 155 発声に関しては，20 発声，20 発声(累積 40 発声)，40 発声(累積 80 発声)，75 発声(累積 155 発声)と 4 セットに分割した．4 セットへの分割は，全管制官で共通である．管制官ごとに，これら 4 セットの話者適応用発声を用いて，3.4.2 で述べた段階的話者適応方式により，話者適応ありの音響モデルを 4 つ作成し，話者適応なしの音響モデルとともに実験で用いた．

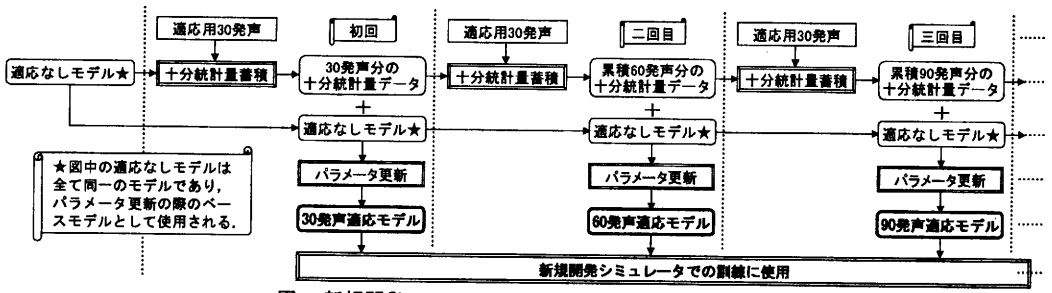

図 6 新規開発シミュレータでの段階的話者適応のフロー

表 3 使用した BNF 文法の仕様

	【実験1】読み上げ音声の認識	【実験2】シミュレータ操作音声の認識
語彙数	530	548
発音数	1,513	1,799
日本語風英語発音数	727	905
ネイティブ英語発音数	786	894
平均分岐数	5.7	6.6

使用した BNF 文法(表 3)は，航空路管制で用いられる文型規則の約 80%をカバーするもので，評価用の 118 発声を全て受理可能である．表 3 において，平均分岐数は，一つの単語に後続する平均単語数を表し，ディクテーションにおける perplexity に相当する．

実験では，日英シームレス音声認識と，その構成要素である日本語専用音声認識および英語専用音声認識の 3 方式を比較した．日本語専用音声認識では，表 1 に示す日英シームレス発音辞書の日本語風英語発音のみ，および表 2 に示す日英シームレス音響モデルの日本語音響モデル部分のみを使用し，話者適応および認識実験を行った．英語専用音声認識についても同様である．探索処理におけるビーム幅は 3 つの認識方式で共通とした．

4.1.2 評価尺度(コマンド正解精度)

今回，評価尺度としては，コマンド正解精度を採用した．

航空路管制発話は，航空機の飛行を制御するものであるため，必ずしも発話の全てを正確に認識する必要はなく，飛行制御に関わるコマンドのみ正確に認識できればよい．例えば，“ANA 3 4 7 climb and maintain flight level 2 4 0”という発声に対して，コマンドを抽出すると，“[C/S]ANA347(全日空 347 便)[ALT]240(24,000 フィート)”となる(図 3 参照)．即ち，この発声例の場合，“climb(上昇せよ)”を誤って“descend(降下せよ)”と認識しても，コマンド抽出結果は同じとなり，飛行制御上は問題ない．

今回使用したコマンド正解精度は，一発声に対して，コマンドが完全に抽出できたときのみ 100%とし，一部でも誤っていれば 0%と定義した．ディクテーションの評価尺度として従来から用いられている単語正解精度のように，置換・挿入・脱落誤りなどは考慮しない．また，今回採用したコマンド正解精度は，文全体の正解との一致性評価尺度として一般的な文章正解精度よりも高い値を示すものとなっている．

コマンド抽出は，BNF 文法中で単語を定義する際に，同時にその属性(C/S, ALT, FIX 等)を付与し，意味属性情報付きの認識結果を得ることにより実現した．

4.1.3 実験結果

3 つの認識方式のそれぞれにおいて，管制官ごとに，話者適応なしおよび 4 段階の話者適応あり音響モデルをそれぞれ用いて，評価用 118 発声に対して認識実験を行った．

図 7 および図 8 は，3 つの認識方式それぞれにおける全管制官の平均コマンド正解精度および認識処理時間の実時間比(RTF)を，音響モデルの 5 つの適応段階毎に示したものである．

図 7 では，話者適応を進めることにより，3 つの認識方式の全てにおいて段階的に認識精度が向上するとともに，日英シームレス音声認識において，他の 2 方式よりも大幅な認識精度の改善が得られている．具体的な認識結果例では，特に，便名，飛行高度・速度・方位，通信周波数等を表す数字の認識精度が改善されている．なお，この実験では文章正解精度もコマンド正解精度に近い高い数値を示しており，文全体としての認識精度も高いものであった．

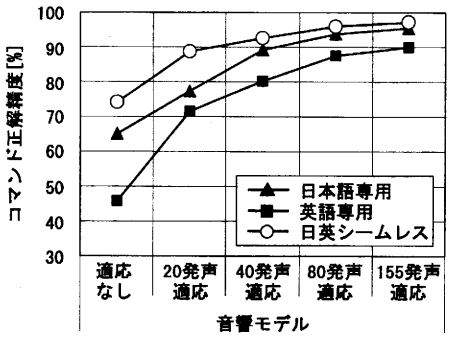

図 7【実験 1】読み上げ音声の認識におけるコマンド正解精度 (管制官男女 10 名の平均)

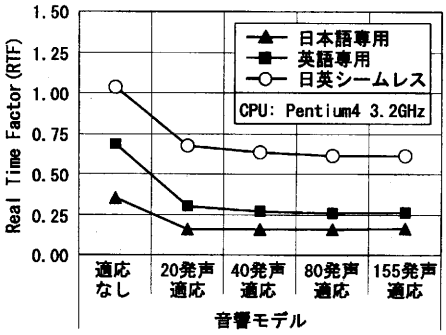

図 8【実験 1】読み上げ音声の認識における RTF (管制官男女 10 名の平均)

図 8 においても，話者適応を進めることにより，3 つの認識方式の全てにおいて段階的に RTF が小さくなり，認識処理時間が短くなっているが，日英シームレス音声認識では，他の 2 方式に比べ，RTF が大きい．これは，日英シームレス音声認識が他の方式の約 2 倍の音素モデルネットワーク上で探索処理を行っていることが原因である．しかし，今回用いた処理系での RTF の絶対値は，話者適応なしの場合で 1.0 付近，さらに適応発声数が 20 以上の場合では 0.7 以下であり，ほぼリアルタイムでの認識となっている．

4.2 【実験 2】シミュレータ操作音声の認識

4.2.1 実験条件

新規開発した航空路管制用シミュレータを男性管制官(実験 1 の 10 名の管制官とは異なる)1 名が実際に操作した音声を収録し，オフラインで認識実験を行った．

シミュレータ操作開始前に管制官の話者適応用発声を収録した．画面の表示を読み上げる形で，1 セット 30 文章として，3 セット合計 90 文章を発声した(平均文長 9.2 単語，平均発声長 3.32 秒)．このとき，教師あり話者適応インタフェースへの慣れを考慮し，各セットでは，先頭の 5 つは単語，次の 5 つは部分文章，その次の 5 つは短文章と，徐々に文章を長くして教示するようにした．このため，話者適応用発声の量は実験 1 と比較して少なく，累積 90 発声の合計時間長としては実験 1 の累積 47 発声程度に相当する．

これら 3 セットの話者適応用発声を用いて，3.4.2 で述べた段階的話者適応方式により，話者適応ありの音響モデルを 3 つ作成し，話者適応なしの音響モデルとともに実験で用いた．

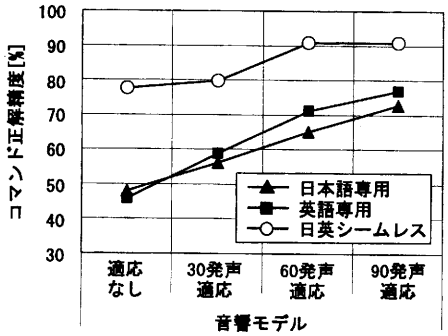

図 9【実験 2】シミュレータ操作音声の認識におけるコマンド正解精度（実験 1 とは別の男性管制官 1 名）

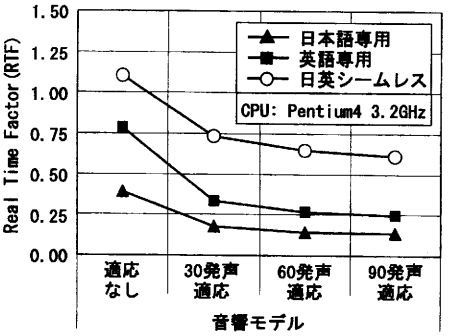

図 10【実験 2】シミュレータ操作音声の認識における RTF（実験 1 とは別の男性管制官 1 名）

実験で使用した BNF 文法は，シミュレータに搭載のものである．その仕様は表 3 の通りであり，文型規則のカバー率は実験 1 で用いたものよりも大きい．

実際のシミュレータ操作を通じて評価用 187 発声を収録した．操作は，2.3 で述べた訓練生の管制官とパイロット役の教官が対峙して行うモードで行われ，全発声が使用した BNF 文法で受理可能であった．平均文長は 13.1 単語で，実験 1 の 13.4 単語と大きく変わらないが，平均発声長は 3.98 秒と実験 1 の 5.88 秒よりもかなり短くなっており，シミュレータ操作音声は，より実際の管制業務シチュエーションに近い話速となっていた．この評価用 187 発声に対し，実験 1 と同様，3 つの認識方式を比較する形でオフラインでの認識実験を行った．

4.2.2 実験結果

音響モデルの適応段階と，コマンド正解精度および RTF の関係を，図 9 および図 10 に示す．両図から，シミュレータ操作音声に対しても読み上げ音声に対する実験 1 と同様の傾向が見て取れ，日英シームレス音声認識によって，リアルタイムの高精度認識が実現されていることが分かる．

また別途，実験1および2とは全く異なる数名の管制官にシミュレータを操作してもらったところ，2.3 で述べた訓練生の自習モードも含め，実験結果と同様のリアルタイムでの高精度認識が実現されていることが確認された．

4.3 認識結果の分析

実験 1 で得られた男女各 5 名の管制官 (M1-5, F1-5) それぞれに対する認識結果に対し，以下の 3 つの側面から分析を行った．

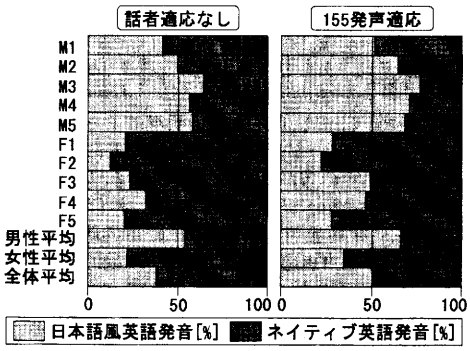

図 11 日本語風英語発音およびネイティブ英語発音の割合（実験1）

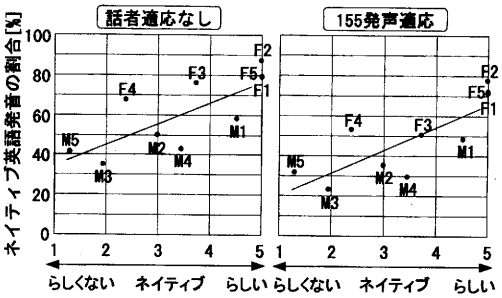

図 12 ネイティブ英語らしさの主観評価値とネイティブ英語発音と認識された単語の割合の相関（実験1）

4.3.1 日本語風英語発音とネイティブ英語発音の割合

認識結果中の各単語の音素列を見て，日本語音素体系に属するもの（日本語風英語発音として認識された単語）と，英語音素体系に属するもの（ネイティブ英語発音として認識された単語）の割合を調査した．調査対象は，話者適応なし音響モデルによる認識結果（総単語数 15,940）と，155 発声適応音響モデル（全適応用発声を使用）による認識結果（総単語数 15,836）である．

図 11 に，各管制官における日本語風英語発音とネイティブ英語発音の割合と男女それぞれの平均および全体平均を示す．図から管制官によって傾向がかなり異なることが分かるが，平均すると，男性よりも女性の発声の方がネイティブ英語発音として認識される単語の割合が多くなっている．

また，話者適応なしよりも 155 発声適応後の方が日本語風英語発音として認識される単語の割合が多くなっている．これは，管制官の発話に日本語風の発音が多く含まれるために，日英シームレス音響モデルの英語音響モデル部分においてよりも，日本語音響モデル部分において，より精度の高い話者適応が行われているためと考えられる．

4.3.2 ネイティブ英語らしさの主観評価値とネイティブ英語発音の割合の相関

英語のヒヤリングに熟練した日本人女性 1 名（495 点満点の TOEIC ヒアリングパートで 480 点以上の実績あり）が，各管制官の評価用発声を一つずつ検聴し，5 段階の主観評価を行った（評点が 5 に近いほどよりネイティブ話者に近い発声と判定）．主観評価結果と，図 11 で示した日英シームレス音声認識結果においてネイティブ英語発音が選択される割合の相関関係をプロットしたものが図 12 である．

表 4 日英間発音遷移頻度(一発声あたりの平均遷移回数)

	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5	男性平均	女性平均	全体平均
話者適応なし	4.8	5.6	5.3	5.5	5.5	4.1	2.9	4.3	4.6	4.0	5.3	4.0	4.7
155発声適応	2.6	2.9	2.9	3.2	3.2	2.5	3.0	3.9	3.2	3.5	3.0	3.2	3.1

平均的に見て，男性よりも女性の発声に高い主観評価値が与えられていることが分かるが，これは，先に述べた男性よりも女性の発声の方がネイティブ英語発音として認識される単語の割合が多いことと一致している．ネイティブ英語発音の割合と主観評価値の相関係数は，話者適応なしの場合で0.78，155発声適応の場合で0.81と比較的高い値が得られた(10 点の標本数の相関係数が 0.80 の場合，母集団の相関係数の 95%信頼区間は 0.36～0.94 である)．

したがって，日英シームレス音声認識においては，話者の発声の日本語らしさまたは英語らしさにより，各単語に付与された日本語風英語発音またはネイティブ英語発音のいずれかを適切に選択して認識を行うことで，高い認識精度が得られるものと考えられる．

ただし今回の分析については，上記主観評価が一つの発声全体に対してなされたものであって，単語毎に対してなされたものではないという点，もともと航空路管制音声は特殊な英語発声(個々の管制官毎に極端な癖のある発音も多い)であり，そのネイティブ英語発声らしさを評価するのは困難な作業であったという点にも留意が必要である．

4.3.3 日英間発音遷移頻度

認識結果において，どの程度の頻度で発音の日英間遷移(単語境界における日本語風英語発音とネイティブ英語発音との間での遷移)が生じているのかを調べた．調査対象は実験1と同じ，男女各5名の管制官の認識結果であり，話者適応なし音響モデルで得られたものと，155 発声適応音響モデルで得られたものである．

表4の日英間発音遷移頻度は，一発声あたり平均して何回，発音の日英間遷移があったかを示すものであり，分析した認識結果の平均文長は，管制官および話者適応段階の違いではほとんど変わらず，約 13.5 単語であった．

表 4 より，話者適応なしと比較して，155 発声適応では，日英間発音遷移回数が減少している．これは，話者を特定して話者適応を進めることにより，日本語音響モデル部分と英語音響モデル部分の統計的性質が近づくためであると考えられる．また，日英間発音遷移回数の減少は，探索処理中に単語境界において triphone モデルより精度の劣る monophone モデルを使用する回数が減少することも意味している．これは 3.3 で述べた日英間の発音遷移に対して適切な triphone モデルが適用できないという本方式の課題を克服することになっており，話者適応後に認識精度の大きな向上が得られる一因であると考えられる．

5 まとめ

日英シームレス音声認識技術を用いた航空路管制用シミュレータを新規に開発し，日本人英語発声の典型例と考えられる日本国内における管制官の航空路管制発話の認識方法を検討した．

管制官による読み上げ音声とシミュレータ操作音声に対する認識実験を行い，いずれの実験でも，段階的な話者適応が効果的に作用し，日英シームレス音声認識により，リアルタイムでの高精度認識を実現できることが確認された．

また，認識結果の分析を通じて，話者の発音の日本語らしさまたは英語らしさにより，単語に付与されている日本語風英語発音またはネイティブ英語発音のいずれかを，数単語単位程度で適切に選択して認識することで，高い認識精度を得られることが分かった．

謝辞 実験にご協力いただきました国土交通省航空局の皆様ならびに東京航空交通管制部の航空管制官の皆様に感謝いたします．

参考文献

[1] 坪田康，壇辻正剛，河原達也，“日本人の誤りパターンの対判別を利用した英語発音教示システム，” 音学講論，3-8-8, pp.153-154, 2001.3.

[2] 河合剛，石田朗，広瀬啓吉，“2 言語の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価，” 日本音響学会誌 57 巻 9 号, pp.569-580, 2001.

[3] 小橋川哲，峯松信明，廣瀬啓吉，ドナ・エリクソン，“英語文リズム学習支援を目的とした文強勢音節のモデル化とその検出，” 信学技報，SP2001-100, pp.99-104, 2001.12.

[4] 峯松信明，“音声の音響的普遍構造の歪みに着眼した外国語発音の自動評定，”信学技報，SP2003-180, pp.31-36, 2004.1.

[5] 阿部一彦，田中和世，河原達也，清水政明，壇辻正剛，“対話型英語学習システムにおける日本人英語音声認識精度の検討，” 音学講論，2-5-20, pp.113-114, 2002.3.

[6] 倉田岳人，峯松信明，広瀬啓吉，“発音習熟度に着眼した適応処理に基づく非母国語音声認識の高精度化，” 信学技報，SP2002-38, pp.13-18, 2002.6.

[7] 松永昭一，小川厚徳，山口義和，今村明弘，“日本人及び母国語話者英語文音声における認識手法の比較，” 音学講論，3-Q-22, pp.197-198, 2002.9.

[8] 松永昭一，小川厚徳，山口義和，今村明弘，“日本人英語音声認識における話者適応の検討，” 電子情報通信学会総合大会講演論文集，D-14-15, p.182, 2003.3.

[9] S. Matsunaga, A. Ogawa, Y. Yamaguchi and A. Imamura, “Non-native English speech recognition using bilingual English lexicon and acoustic models,” Proc. ICASSP2003, vol.1, pp.340-343, 2003.4.

[10] S. Matsunaga, A. Ogawa, Y. Yamaguchi and A. Imamura, “Non-native English speech recognition using bilingual English lexicon and acoustic models,” Proc. ICME2003, vol.3, pp.625-628, 2003.7.

[11] S. Matsunaga, A. Ogawa, Y. Yamaguchi and A. Imamura, “Speaker adaptation for non-native speakers using bilingual English lexicon and acoustic models,” Proc. EUROSPEECH2003, vol.4, pp.3113-3116, 2003.9.

[12] 松永昭一，小川厚徳，“日本人英語音声認識における発話者英語能力別の効果，” FIT2003 講演論文集，F-023, pp.253-254, 2003.9.

[13] 中野秀夫，航空管制のはなし，交通ブックス 303，成山堂書店，2001.

[14] 園山耕司，航空管制の科学－飛行ラッシュの空をどうコントロールするか－，ブルーバックス B-1399，講談社，2003.

[15] 管制方式基準（追録第 19 号），鳳文書林出版販売，2001.

[16] 航空用語辞典，鳳文書林出版販売，2003.

[17] SAMPA computer readable phonetic alphabet, http://www.phon.ucl.ac.uk/home/sampa/home.htm

[18] 大附克年，山口義和，高橋敏，“様々な音素環境依存性を持つ音響モデルの効率的学習に関する検討，” 音学講論，1-9-24, pp.47-48, 2002.9.

[19] J.L. Gauvain and C.H. Lee, “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,” IEEE Trans. on Speech and Audio Processing, vol.2, no.2, pp.291-298, Apr. 1994.

[20] 野田喜昭，山口義和，大附克年，小川厚徳，中川聡，今村明弘，“音声認識エンジン VoiceRex の開発，” 音学講論，2-1-19, pp.91-92, 1999.9-10.