TR-S-0013

対話音声を対象とした日本語連続音声認識システムの試作と評価 (音響モデル：ResearchJ V9)

A continuous speech recognition system for conversational speech (Acoustic model: ResearchJ V9)

内藤 正樹 †
Masaki Naito

松井 知子
Tomoko Matsui

2000.11.27

対話音声を対象とした音声認識システム ATRSPREC の試作・評価を行った．システム構築にあたり，対話音声認識において認識性能劣化の大きな要因となる話者の発話様式の変化に対して頑健な音声認識を実現するため，発話様式依存音響モデルを用い，認識と同時に各発話に対して最適な音響モデルを動的に選択することで，発話様式の変化に対するオンライン適応を実現した．日英音声翻訳システムを通した対話音声を用いた音声認識実験によりシステムの認識性能の評価を行った．対話データの解析の結果，音声認識システム利用者がシステムに慣れるにつれ，発話様式に変化が見られたが，発話様式依存音響モデルの動的選択を行うことで，自然発話，朗読音声用音響モデル各々を単独で用いた場合の誤認識が約 13% 削減され，発話様式の変化に伴う音声認識性能の劣化が改善された．

† 現在、(株) KDD 研究所

目次

1 はじめに

近年，統計的音響・言語モデルの導入，大規模な音声・言語データベースの整備や計算機の処理能力の向上にともない，連続音声認識システムの開発・実用化が進んでいる [1][2]．しかし，これらのシステムの多くは，ディクテーションの様に比較的整った文型で，明瞭に発声された音声の認識を対象として設計されており，自然な対話音声の認識に対しては依然多くの課題を有する．

言語的な面から見た場合，対話音声には，言い直し，省略，間投詞等，書き言葉には見られない現象が多く出現する [3]．このため，大量の言語データが整備されている書き言葉を対象とした言語データベースを，対話用言語モデルの作成に用いる事は困難であり，対話を対象とした言語データベースの整備が必要となる．しかし，実際に大量の対話データを収集，整備することが困難であるため，対話音声の認識に向けた言語モデル作成においては，小量の言語データから，精度の良い言語モデルを構築する必要がある．

音響的な面から見た場合，従来の音声認識システムにおいては，読み上げ音声を基に学習した音響モデルが用いられる場合が多い [1][2]．話者による音声の音響的特徴の差異が音声認識性能に与える影響が大きいため，多数話者による大量の音声データの収集が容易な朗読音声による音響モデルの構築は，不特定話者に対する認識性能の向上を図る上で有効な手段である．しかし，朗読音声と自然発話音声ではその音響的特徴が大きく異なることが指摘されており，対話音声に対して高い認識性能を得るためには，認識対象とする発話様式に適合した音響モデルを用いる必要がある [14]．また，音声認識システムの運用実験において，システム利用者が音声認識システムに慣れるに従い，タスク達成までに要する時間が短縮するとの報告も為されており [5]，話者が音声認識システムに習熟するにつれ，発話様式が変化すると考えられる．音声対話システムの性能向上を図るためには，このような発話様式の変化に対して柔軟に対応可能な認識系を導入する必要がある．

これまで我々は，対話音声データを用いた効率的な音響・言語モデルを構築するため，最ゆう逐次状態分割法 [6]，多重クラス複合 N-gram[7][8] 等の要素技術を提案し，その有効性を示してきた．これら要素技術を統合し音声認識システム ATRSPREC の試作を行った．本システムは，自然な音声の入力を許容する対話システムのフロント・エンドとして設計されたものであり，上記の対話音声向け音響・言語モデルを駆動すると共に，対話音声に特徴的な発話様式の変動に対処するべく，複数の音響モデルを並列に駆動することが可能となっている．具体的には，性別・発話様式に依存した音響モデルを複数用意し，それら複数モデルを用いた探索を並行して行い，各発話に対して，ゆう度を基準とした最適な音響モデルの選択と認識を同時に実行するものである．

本稿では，システムの構成について述べ，認識性能の評価を行う．評価においては，発話様式の異なる種々の音声に対する認識性能を示すため，模擬対話音声と日英音声翻訳システム (ATR-

MATRIX) による音声翻訳対話 [9] を対象とし，対話音声の音響的特徴の分析を行うと共に，認識実験によりシステムの性能の評価を行う．

2 音響モデル

2.1 最ゆう逐次状態分割法による音素コンテキスト依存モデルの生成

音素コンテキストに依存する音素パターン変形のモデル化は，連続音声認識の性能向上を図る上で有効な手段である．本システムで用いる隠れマルコフ網 (HMnet)[10] は，複数の音素コンテキスト間で HMM の状態を共有した音響モデルで，HMM の状態のネットワークとして表現される．HMnet の各状態は複数の音素コンテキストに対応しており，ある音素コンテキストに対する HMM はこのネットワークの始端から終端に至る 1 つの経路上の状態を接続することで作成される．

HMnet の構造決定法として鷹見らにより提案された，逐次状態分割法 (SSS)[10] では，HMM の状態に対する，時間方向または音素コンテキスト方向への 2 分割を繰り返し行うことで HMnet の構造を決定する．その際，出力確率分布の広がりが，音素コンテキストや時間方向への音響的な特徴の多様性を反映するとの仮定に基づき，HMnet 各状態の出力確率分布の分散を基に，被分割状態を決定した後，分割方法を決定し状態分割を行っていた．しかし，SSS を不特定話者の学習データに対して適用する場合，出力確率分布の広がりが，音素コンテキストに起因する音声の多様性と共に話者方向への多様性を反映するため，出力確率分布の分散に基づき被分割状態を選択することが困難であった．

本システムでは，不特定話者の音声データを用いた HMnet の構造決定を可能とするため，最ゆう逐次状態分割法 (ML-SSS)[6] を用いる．ML-SSS では，HMnet の全ての状態に対する全ての分割方法から，分割によるゆう度増加の期待値を最大とする分割方法を選択し，状態分割を行う．その際，Chou により提案された最適分割アルゴリズム [11] を用い，各状態に対する最適な分割方法の決定に要する計算量を削減している．

2.2 話者と発話様式に依存した音響モデルの構築

自然発話音声認識を困難とする要因として，話者や発話様式に起因する音声の音響的特徴の多様性が挙げられる．話者や発話様式による認識性能の劣化を改善する手法として，重回帰写像モデルを用いた話者正規化手法 (MLLR)[12] や，最大事後確立推定法と移動ベクトル場平滑化法を統合した話者適応手法 (MAP-VFS)[13] 等の話者適応手法が提案されている．しかし，これらの手法により高い認識性能を得るためには，多くの適応データが必要となる．不特定の話者が短時間使用する形態の音声認識システムにおいては，利用可能な適応データ量が少ないため，話者クラスタモデルを用いた話者適応手法 [14] のように，複数音響モデルから最適な音響モデルを選択し認識を行なう適応手法が効果的と考える．

よって，本システムでは，性別に起因する話者の音声の特徴や，発話様式の違いに応じて，複数の音響モデルを用意し，認識時にそれらの音響モデルから，最適なモデル選択し，認識を行うことで話者と発話様式への簡易な適応を実現する．具体的には，話者の性別毎に，(a) 旅

表 1: 音響モデル学習条件

音響分析
サンプリング周波数 16kHz, preemphasis 0.98
フレーム周期 10 ms, フレーム長 20 ms (Hamming 窓)
Δlogpower, 12 次 -MFCC,12 次 -ΔMFCC
ケプストラム平均, パワーを正規化
HMnet の構成
音声 1400 状態 5 混合, 無音 3 状態 10 混合
自然発話学習用音声データセット (旅行対話)
男性：167 話者, 総発話時間 約 2 時間
女性：240 話者, 総発話時間 約 3 時間
朗読発話学習用音声データセット (音素バランス文)
男性：165 話者, 総発話時間 約 9 時間
女性：235 話者, 総発話時間 約 14 時間

行に関する模擬対話音声を基に学習した自然発話用音響モデルと，(b) 音素バランス文の朗読音声を基に学習した朗読発話用音響モデルの 2 種類の音響モデル，計 4 種類の音響モデルを用意した．認識時には，4.2節に述べる手法を用い，4 種の音響モデルを並列に使用し認識を行うことで，ゆう度を基準とした最適な音響モデルの選択と認識を同時に実現する．表1に，音響分析条件，及び音響モデルの学習に用いたデータ量を示す．

表 2: 言語モデル学習条件

言語モデル：多重クラス複合 N-gram
クラス数：from- クラス 700, to- クラス 700
学習データ
旅行に関する対話 7,195 片対話
のべ単語数 約 160 万単語, 異なり単語数 約 2 万 7 千語)

3 言語モデル

限られた学習データから，信頼性の高い統計的言語モデルを推定するため，複数の単語を 1 つのクラスとして，単語間の遷移をクラス間の遷移で近似するクラス N-gram[15] が提案されている．クラス N-gram においては，クラスの規定方法がモデルの精度，サイズに影響する重要な問題となる．本システムでは，多重クラス複合 N-gram[8] を用い，単語の効率的なクラス化を行うことで，大量の言語データの整備が困難である対話を対象とした，信頼姓の高い言語モデルの構築を図る．

多重クラス複合 N-gram はクラス N-gram を基本として，直前直後の単語の接続性を考慮し，各単語を先行単語として用いる場合と，後続単語として用いる場合とで，複数の異なるクラスを割り当てるモデルである．本モデルでは各単語の予測確率が次式により与えられる．

$$P(w_i|w_{i-1}) = P(w_i|C^t_{w_i})P(C^t_{w_i}|C^f_{w_{i-1}}) \tag{1}$$

ここで，w_{i-1} は単語または単語系列である．$C^f_{w_{i-1}}$ は先行単語 w_{i-1} が属するクラス (from- クラス) であり，$C^t_{w_i}$ は後続単語 w_i が属するクラス (to- クラス) である．

また，本モデルでは，出現頻度の高い連鎖語を N-gram の単位に加えることで，部分的に高次単語 N-gram を導入する [7]．その際，単語系列 $w_h \ldots w_t$ が属するクラス，$C^f_{w_h \ldots w_t}$，$C^t_{w_h \ldots w_t}$ はそれぞれ，$C^f_{w_t}$，$C^t_{w_h}$ と同一であるとする．また，本モデルで使用するクラスは，単語間の接続性に着目したクラスタリング手法を用い，学習データから自動的に決定する [16]．言語モデルの学習条件を表 2に記す．

4 探索手法

4.1 単語グラフを用いた二段探索

本認識システムの探索部は，フレーム同期に単語仮説の生成と単語グラフへの登録を行う第1パスと，単語グラフに含まれる単語仮説の枝刈りを行う第2パスの2つの処理課程により認識を行う [17]．本認識部では，認識における計算量の多くを占める音響ゆう度の計算を，全て第1パスで行い，第2パスでは言語ゆう度の再計算のみを行う．これにより，発声終了から認識結果出力までの遅延時間を短縮することが可能となる．以下に第1パス，第2パスで行われる処理の概要を示す．

・ 第1パス: 単語グラフの生成

第1パスでは，フレーム同期に単語仮説の生成，入力音声との照合を行う．各フレームにおいて，木構造辞書を参照し，HMM の状態毎の仮説（状態仮説）を展開し，各状態仮説の単語内のゆう度と発声開始からのゆう度を計算する．状態仮説は単語の開始時刻，先行単語の違いにより個別に用意し，各状態仮説のゆう度は Viterbi 計算により求めた音響ゆう度と言語ゆう度との重み付き和により定める．つづいて，単語仮説の増加を抑えるため状態仮説の枝刈りを行う．これは，全状態仮説中での文頭からの累積ゆう度の最大値を求め，最大ゆう度とのゆう度差が一定値以上の状態仮説を削除するものである．以上の処理の後，単語の終端状態までのゆう度が計算された単語を単語グラフに登録する．その際，文頭からの累積ゆう度を保持し，後続する単語仮説を生成する際に文頭からの累積ゆう度として使用する．

探索の効率化のため第1パスでは，単語辞書を木構造化することで先頭からの部分音素列が等しい単語の状態仮説を共有し，単語先頭での状態仮説数の増大を防ぐ．また，同音語の単語仮説をマージすることで，後続する単語の仮説生成数の増加を抑える．さらに「単語仮説先頭の発音が同じであれば開始時刻は同じとなる」との近似を与え，単語間の遷移時刻の絞り込みを行う．従来から広く行われている単語対近似 [18] では，先行単語，終了時刻が等しく開始時刻のみが異なる単語仮説の中で，文頭からのゆう度が最大となる単語仮説のみを残し，それ以外の単語仮説を削除することで単語仮説の開始時刻の絞り込みが行われる．その際，ゆう度の比較を行なう条件は，先行単語が同じ場合に限られ，先行単語が異なれば比較の対象とならない．

本認識系では，前述した近似を導入し，図1に示すように，後続する単語仮説先頭の異音を c とする先行単語群 V をマージし，一つの単語仮説に対して複数の先行単語を与える．単語間の遷移時刻の絞り込みを行う際は，時刻 t において，単語仮説先頭の異音を c とする先行単語群 V に後続する単語仮説の，木構造辞書上のノード s，HMM の状態 h に対するゆう度を比較し，最大ゆう度を与える単語仮説のみを残す．

本手法の導入により，時刻 t_e において，単語 w の終端ノード $S(w)$ に到達する単語仮説数を，単語の先頭音素の異音の個数以下に抑えることが可能となり，認識率の低下を招くこと無く，

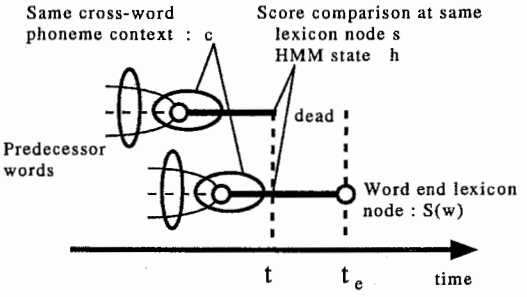

図 1: 単語の先頭異音を用いた単語境界の絞り込みの例

単語仮説数が削減されることが認識実験により確認されている [17].

・ 第 2 パス: 単語グラフの枝刈り

まず，第 1 パスで生成された単語グラフに対して，単語仮説の言語ゆう度と文頭からの累積ゆう度の再評価を行う．これは，第 1 パスで，同音語としてマージされた各単語に対して正確な言語ゆう度を与えるものである．

続いて，単語グラフ上の全ての単語仮説に対して，以下の手順に従い枝刈りを行う．

1. 単語グラフ上でその単語仮説を含む全ての経路中から文頭から文末までの累積ゆう度が最大となるゆう度を求める．

2. その最大値から一定ゆう度幅以下の単語仮説・経路を単語グラフから削除する．

4.2 音響モデルの動的選択によるオンライン適応

前章で述べたように，本認識系では，話者の音声の特徴や発話様式に依存した，複数の音響モデルを用い認識性能の向上を図る．その場合，各発話に対して最適な音響モデルを選択した上で，認識を行う必要がある．本認識システムでは，最適な音響モデルは 1 発話を通じて変化しないとの仮定の下，以下の手順に従い，ゆう度を基準とした最適な音響モデルの選択と，認識を同時に実現する [19].

まず，認識開始時には，各音響モデルごとに異なる仮説を生成し，フレーム同期 Viterbi サーチによる照合を開始する．各フレームでは音響モデル毎に独立に仮説の展開・照合を行う．加えて，各フレームで，全ての音響モデルに基づく状態仮説全体を対象として，全仮説中で最大ゆう度を示す状態仮説とのゆう度差が大きい仮説の枝刈りを行う．音声区間の終了時に最大ゆう度を持つ状態仮説を選択することで，各発話に対する最適な音響モデルの選択と認識が同時に実現される．

本手法では，認識しようとする音声との適合度の低い音響モデルに対する仮説は，認識開始から早い時期に枝刈りされることから認識時間の大きな増加を招くことなく，音響モデルの動的選択によるオンライン適応が実現される．

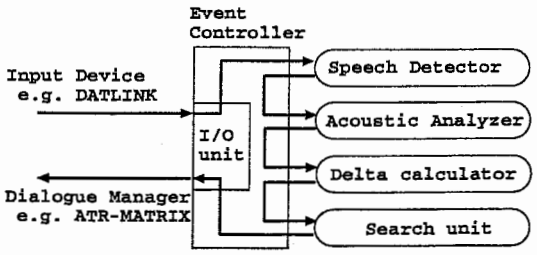

図 2: システムのモジュール構成

5 システム構成

本システムでは，フレーム同期の認識に要する逐次処理を，イベント駆動の形で実現している．図 2にシステムのモジュール構成を示す．図右側に描かれた，音声区間検出部 (Speech Detector)，音響分析部 (Acoustic Analyzer)，Δパラメータ計算部 (Delta Calculator)，探索部 (Search Unit) の各モジュールは，データの入出力をすべてイベントの送信，受信の形で行う．図中央のイベント管理部分 (Event Controller) では，各モジュール間でのイベント送受の制御や，ファイルへの入出力の管理を行う．音声入力や音声認識結果の外部への出力もイベント管理部にある I/O 制御部 (I/O unit) にて行われる．複数モジュールによる逐次処理の実現法としては，パイプライン処理が挙げられるが，イベント駆動の導入により，パイプライン処理において困難であった，(1) 複数のモジュールとの間のデータの受渡しや，(2) すべてのモジュールに対する同一タイミングでの割り込み（発声のキャンセル等）等が容易に実現可能となる．パイプライン処理においては各モジュールが異なるプロセスとして動作するのに対して，本システムにおいては，イベント管理部，及びその管理下にあるモジュールが 1 つのプロセスとして動作する．システムの構成を変更する際は，システムを構成するモジュールの追加，削除やデータの送受先の指定を変更した後に，プログラムの再リンクを行う事で，容易に変更可能である．例えば，イベント管理部を 2 つに分割し，音響分析までを担当するプロセスと探索を行うプロセスに分割し，プロセス間での通信手順を定める事で，サーバー，クライアント型の音声認識システムが構築可能となる [20].

表 3: 模擬対話音声評価データセット

(1) 対面対話
男性 17 名，女性 25 名，のべ 551 発話，6342 単語
(2) 通訳を介した日本語 - 英語の対話 (日本語側のみ)
男性 8 名，女性 15 名，のべ 330 発話，4665 単語
(3) 2. の読み上げ音声
男性 10 名，女性 10 名，のべ 295 発話，4000 単語

6 模擬対話音声による認識性能の評価

6.1 模擬対話音声データ

発話様式の異なる音声に対する本システムの認識性能を評価するため，異なる状況下で収録された以下の 3 種の模擬対話音声データを用い認識実験を行った [21].

1. 2 人の日本語話者が対面して行った旅行に関する対話音声 (**対面対話**)
2. 日本語話者と英語話者の間で通訳を介して行われた対話中の日本語話者（通訳者は除く）の発話 (**通訳あり**)
3. (2) の対話書き起こしテキストを読み上げた音声 (**明瞭発話**)

(3) では，音声対話システム利用時に近い発話様式の音声を用いた評価を行うため，後述する音声翻訳対話システム (ATR-MATRIX) の利用経験者を話者とし，普段各自がシステムに向け発声する場合と，同様な様式で発話するよう依頼し音声収録を行った.

各評価データセットを構成する話者数等の情報を表 3に示す.

6.2 模擬対話音声の特徴分析

発話速度，母音の継続時間を基に，2.2節に示した音響モデルの学習データセットと前節に示した 3 種の評価データセットの特徴の比較を行う．表 4に各データセット別に，発話速度 (モーラ / 秒) と母音の継続時間長の平均 (msec)，母音の継続時間の標準偏差を示した．なお，発話速度は人手により付与された音声区間の時刻情報と音素書き起こしを元に算出した．母音の継続時間は音素書き起こしと HMM を用い，Viterbi アルゴリズムにより求めた音素境界の時刻情報を元に算出した．学習データセットについて見ると，対話音声から成る (a) 自然発話と比較して，音素バランス文から成る (b) 朗読発話の発話速度は約 3% 遅い．朗読発話は自然発話と比較して，母音の継続時間の分散が小さい点に特徴があり，発話速度の変動が小さいことが分かる.

表 4: 音響モデル学習用音声データセット・評価用模擬対話音声データセットの発話速度と母音の継続時間長

	データセット	発話速度 (モーラ / 秒)	母音の平均継続時間 [標準偏差](msec)
学習	(a) 自然発話	7.99	75.4[52.3]
データ	(b) 朗読発話	7.78	77.9[35.6]
評価	(1) 対面対話	8.02	76.3[55.6]
データ	(2) 通訳あり	8.52	71.7[43.8]
	(3) 明瞭発話	6.41	82.0[48.6]

続いて評価データセットの特徴について述べる．学習データセット (a) 自然発話と，評価データセット (1) 対面対話とは，発話者は異なるが，同一の収録条件で模擬対話を実施した音声である．このため，両者の発話速度，継続時間長は近い値となっている．評価データセット (2) 通訳ありは，(1) 対面対話と比較して発話速度が 6% 程度速い一方で，母音の継続時間の分散は小さく，発話速度の変動が少ない点に特徴がある．評価データセット (3) 明瞭発話は，他の学習データセット，評価データセットと比較して発話速度が非常に遅く，また，(1) 対面対話と比較すると母音の継続時間の分散は小さく，発話速度の変動が少ない．

6.3 模擬対話音声に対する認識性能

まず，(a) 自然発話音響モデル (**spontaneous**) を単独で使用し，枝刈りのビーム幅を数種変更した上で，評価データセット (1) 対面対話の認識を行い，認識速度と認識率の関係を調べた．認識時間と単語誤認識率を表 5に示す．実験は，メインメモリ 1.5G バイトを搭載した Compaq 社製 Alpha Station500 上で行い，表中の認識時間は，認識に要した時間を発話時間で正規化した値である．音声対話システムを構築する際には，これら認識時間と認識率の関係を基に，本認識システムを実行する計算機の処理能力と，要求される応答速度に応じて，ビーム幅を調整する．表 5右端の，認識時間 4.1，単語誤認識率 12.5% を示した実験は，上記メモリ容量で動作可能な範囲で，ビーム幅を最大とした際の結果である．以降の実験においては，このビーム幅を共通に使用し，上記の計算機構成で最も高い認識率が得られる設定の下，認識性能の比較を行っている．

続いて，発話様式の異なる評価データセットを対象とした認識実験の結果を示す．実験は，(a) 自然発話音響モデル (**spontaneous**)，(b) 朗読発話音響モデル (**read**) の 2 種類の音響モデルそれぞれ単独に用いた場合と，(c) 音響モデルの動的選択により，2 種類の発話様式依存音響モデルを同時に使用し認識を行った場合 (**auto-select**) の 3 通りの実験を行った．なお，

表 5: 評価データセット: 対面対話の認識における認識時間と単語誤認識率

認識時間	1.1	1.6	2.0	3.1	4.1
単語誤認識率 (%)	28.9	18.4	15.8	13.0	12.5

発話様式依存音響モデルは，各性別毎に学習されており，認識の際は，性別依存音響モデルの動的選択も同時に行われる．このため，(a),(b) に対しては 2 つ，(c) に対して 4 つの音響モデルを並列に使用し認識が行われる．

実験の結果得られた単語誤認識率を図 3に示す．まず，評価データセット毎に，自然発話音響モデルと朗読発話音響モデルをそれぞれ単独で用いた場合の認識性能を比較する．評価データセット (1) 対面対話の認識を行った際 (**face-to-face speech**) には，認識対象となる音声と等しい発話様式の音声で学習した自然発話音響モデルを用いた場合に，単語誤認識率 12.5% が得られた．これに対して，発話様式が異なる朗読発話音響モデルを用いた場合には，単語誤認識率 15.6% と誤認識が約 25% 増加する．

評価データセット (2) 通訳ありの認識 (**bilingual speech**) に対しては，2 つの音響モデルの間に大きな認識性能の差はないが，(3) 明瞭発話の認識 (**machine-friendly speech**) においては，朗読発話音響モデルを用いた場合の単語誤認識率 4.8% に対して，自然発話音響モデルを用いた場合の単語誤認識率は 7.8% と誤認識が約 6 割増加している．これらの結果から明らかなように，高い認識性能を得るためには，認識対象となる音声の発話様式に適した音響モデルを用いる必要がある．

続いて，音響モデルの動的選択により，発話様式依存音響モデルを並列に用いた場合の認識性能を見る．この結果，発話様式依存音響モデル各々を単独で用いた場合に見られた評価データの発話様式による，認識性能の大きな劣化は見られず，3 種の評価データセットいずれに対しても平均して高い認識性能が得られた．

なお，音響モデルの動的選択を行う場合，照合を行う音響モデル数が増加するため認識速度の低下が予想される．評価データセット (1) 対面対話を用い，Compaq 社製 Alpha Station500 上で認識に要した時間の比較を行った．その結果，自然発話音響モデルを単独で用い認識を行った際の認識時間が実時間の 4.1 倍であるのに対して，発話様式依存音響モデルの動的選択を行った場合の認識時間は実時間の 4.9 倍であり，認識時間の増加は約 20% であった．

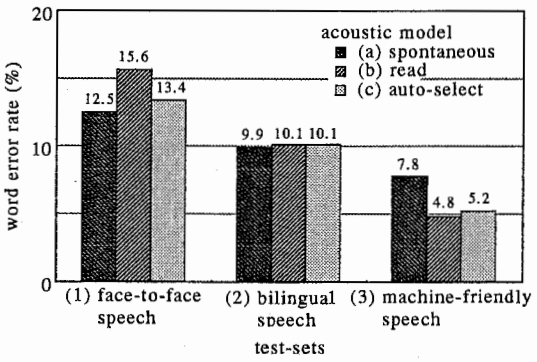

図 3: 模擬対話音声に対する単語誤認識率 (%)

図 4: 音声翻訳システム ATR-MATRIX の日本人話者側画面表示

7 音声翻訳システムへの入力音声による認識性能の評価

7.1 音声翻訳対話実験

人間は発声を行う環境に応じてその振舞を変化させることから，音声対話システムに向けて行われる発話は，前章で評価に用いた模擬対話とは異なる特徴を有すると予想される．この点から，音声認識システム使用時の，実環境下で収録した対話を対象として，対話の特徴分析，及び本システムの性能評価を行う．評価には，音声翻訳システム (ATR-MATRIX) を用いた日英音声翻訳対話実験 [9] の際に収集された音声データを用いる．

対話実験は，日本語側，英語側各 1 台の音声翻訳システムを用い行われた．日本語側システムでは，音声認識システム (ATRSPREC) を用い日本語音声の認識を行い，強調融合翻訳部 (TDMT)[22] において認識結果を英語に翻訳した後相手側システムへ送信する．また，相手側システムから送られる英日の翻訳結果は，波形接続型音声合成部 (CHATR)[23] により合成され，日本語音声としてシステム利用者に伝えられる．英語側システムではこれと逆の処理が行われる．

なお，対話実験当時の認識部の諸元は，前章までに述べた現在の認識部の諸元とは異なる．主な相違点として，まず音響モデルに，性別依存・発話環境適応モデルが使用された点が挙げられる．これは，ATR-MATRIX の音声入力装置を介して収録した複数話者の音素バランス文音声を適応データとし，MAP-VFS 法により，自然発話音響モデルを初期モデルとして，朗読発話への発話様式適応，音声の入力特性，背景雑音等への適応を行ったモデルである [24]．また，言語モデルには，語彙数約 3000 語の品詞及び可変長単語列の複合 N-gram [7] が使用された．

図 4に音声翻訳システムの日本側画面を示した．画面上には，テレビ会議システムを通した相手話者の様子と，日本語の音声認識結果が表示される．音声翻訳システムは常時音声入力を受け付ける状態にあり，音声が入力されると，画面に認識結果を表示し英語への翻訳を行う．日本語話者の生音声は英語側には伝えられず相手側には翻訳結果が英語の合成音のみで伝えら

表 6: 音声翻訳対話音声の発話速度と母音の継続時間長

	発話速度 (モーラ / 秒)	母音の平均継続時間 [標準偏差](msec)
初回	6.64	91.1[57.6]
2 回目	6.68	87.8[55.8]
練習後	6.13	87.2[48.5]

れる．話者は自分が発声した音声の認識結果を目で確認できるため，誤認識された場合など，発声しなおすことが可能である．

音声翻訳対話実験は，日本人旅行者が，音声翻訳システムを介して，アメリカ人のフロントに電話をし，ホテルの予約を行うという設定で行った．ただし，音声の入力は電話からではなくマイクからの入力音声を使用している．対話中には，空室の照会，値段の確認，連絡先電話番号の確認，クレジットカード番号等支払方法の確認等の会話が行われた．

実験は，日本人話者 (男性 3 名，女性 2 名) とアメリカ人話者 (女性 1 名) の間でなされた．話者の音声認識システムへの習熟度を考慮するため，同一の話者について，(1) 音声翻訳システムの GUI の基本操作の説明の直後に行った対話 (**初回**), (2) 初回の対話に引続き行われた対話 (**2 回目**)，(3) 例文を用いた発声練習の後に行った対話 (**練習後**) の 3 度の対話実験を行った．

(3) の収録前に行われる発声練習の際には，各話者は対話内容とは直接関係のない 11 文を用い，認識結果を見ながら発声を行うことで，認識しやすい発声を練習している．

7.2 音声翻訳対話音声の特徴分析

上記の音声翻訳対話を対象に，発話の特徴分析を行った．なお，音声認識システム利用時に，認識誤りを生じた場合には，話者がシステムに対して，同一の内容を複数回繰り返し発声する場合がある．発話の特徴の分析，及び認識実験には，これらの内最初に発声された 1 発話のみを対象とした．評価データセット (1) 初回，(2)2 回目，(3) 練習後に含まれるデータ量はそれぞれ，110 発話 (955 単語)，105 発話 (877 単語)，110 発話 (838 単語) である．

表 7に，各評価データセットに対する発話速度 (モーラ / 秒) を母音の継続時間の平均値 (msec) と標準偏差を示した．前章に示した模擬対話音声 (1) 対面対話，(2) 通訳ありと比較すると，発話速度が 2 割程度低下しており，(3) 明瞭発話に近い発話速度となっている．音声翻訳対話 (1) 初回と (2)2 回目の対話を比較すると，発話速度には大きな差は見られないが，母音の継続時間の分散が小さくなっており，発話速度の変動が若干小さくなっている．また，(3) 発声練習後には，発話速度は，初回，2 回目と比較して約 8% 遅くなり，発話速度の変動が小さくなる事が分かる．

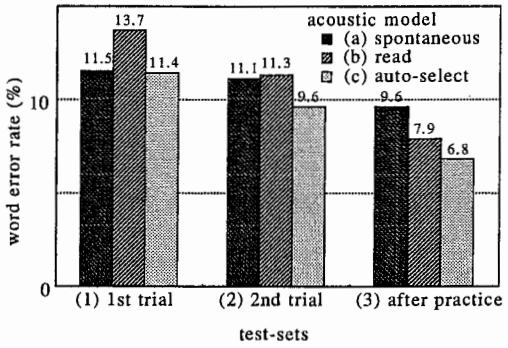

図 5: 音声翻訳対話音声に対する単語誤認識率 (%)

7.3 音声翻訳対話における認識性能

音声翻訳対話データを対象とした認識実験を行った．実験は模擬対話音声による実験と同様に，(a) 自然発話音響モデル (**spontaneous**)，(b) 朗読発話音響モデル (**read**) の 2 種類の音響モデルそれぞれ単独に用いた場合と，(c) 音響モデルの動的選択により，2 種類の発話様式依存音響モデルを同時に使用し認識を行った場合 (**auto-select**) の 3 通りの実験を行った．実験により得られた単語誤認識率を図 5に示す．

各評価データセットについて，自然発話音響モデルと朗読発話音響モデルをそれぞれ単独で用いた場合の認識性能を比較すると，(1) 初回の対話 (**1st trial**) においては，自然発話音響モデルの単語誤認識率 11.5% と比較して朗読発話音響モデルの単語誤認識率 13.7% と認識性能が低く，(2)2 回目の対話 (**2nd trial**) においては 2 つの音響モデルの認識性能は同程度となる．(3) 練習後の対話 (**after practice**) においては朗読発話音響モデルの認識性能が高く単語誤認識率 7.9% を示した．このように，話者がシステムに慣れるに従い生じた発話様式に変化により認識に適した音響モデルが変化することが確認された．話者がシステムに慣れるまでの初期の段階での認識性能向上を図るためには，自然発話に対応した音響モデルが必要であり，話者がシステムに慣れ，明瞭な発声を心がける事でシステムの認識性能も向上し，朗読発話音響モデルの利用が効果的となる．反面，この対話実験では朗読発話に発話様式適応した音響モデルを認識に用いており，被験者がシステムの認識性能の不足を補うため明瞭な発話を行うよう学習した結果，このような発話様式の変化が生じたものと考えられる．今後，話者に発話様式の矯正を要求しない音声対話システムの構築を進めるためには，自然発話に対応した音響モデルの導入・性能向上が必要となる．

また，発話依存音響モデルの動的選択を用いることで，全評価データセットに対して発話様式依存音響モデルを単独で用いた場合と同等以上の認識性能が得られた．評価データセット全体に対する平均で見ると，発話依存音響モデルの動的選択により，自然発話音響モデルを用いた際の誤認識が約 13%，朗読発話発話音響モデルを用いた際の誤認識が約 19% 削減され，本手法により発話様式の変化に伴う音声認識性能の劣化が改善される事が確認された．

8 むすび

最ゆう逐次状態分割法，多重クラス複合 N-gram 等，対話音声データを用いた音響・言語モデル作成のための要素技術を統合し，対話音声を対象とした音声認識システム ATRSPREC の試作・評価を行った．システム構築にあたり，対話音声認識において認識性能劣化の大きな要因となる話者の発話様式の変化に対して頑健な音声認識を実現するため，発話様式依存音響モデルを用い，認識と同時に各発話に対して最適な音響モデルを動的に選択することで，発話様式の変化に対するオンライン適応を実現した．日英音声翻訳システムを通した対話音声を用いた音声認識実験によりシステムの認識性能の評価を行った．対話データの解析の結果，音声認識システム利用者がシステムに慣れるにつれ，話者の発話様式に変化が生じることが観察され，認識実験においても，発話様式の変化に応じて，認識に適した音響モデルが変化することが確認された．話者がシステムに慣れるまでの初期の段階での認識性能向上を図るためには，自然発話に対応した音響モデルが必要であり，話者がシステムに慣れ，明瞭な発声を心がける事でシステムの認識性能も向上し，朗読発話音響モデルの利用が効果的となる．反面，このような現象は，被験者がシステムの認識性能の不足を補うため明瞭な発話を行うよう学習した結果生じたものと考えられる．今後，話者に発話様式の矯正を要求しない音声対話システムの構築を進めるためには，自然発話に対応した音響モデルの導入・性能向上が必要となる．また，発話様式依存音響モデルの動的選択を行うことで，自然発話音響モデルを用いた際の誤認識が約 13%，朗読発話発話音響モデルを用いた際の誤認識が約 19% 削減され，本手法により発話様式の変化に伴う音声認識性能の劣化が改善される事が確認された．

なお，本システムを認識部に用いた音声翻訳システム (ATR-MATRIX) は，1999 年 7 月に行われた C-STAR 音声翻訳国際共同実験等を通じてその動作を確認している．また，MAP-VFS 法を用い，音声入力装置の入力特性への，音響モデルの適応を行う事で，携帯電話を介した音声翻訳への応用等も検討を進めている [25].

9 謝辞

研究の機会を与えて頂いたATR音声翻訳通信研究所山本誠一社長に感謝いたします．熱心な御討論と有益な御助言をいただいた第一研究室の諸氏，ならびに音声データベースの整備等に御協力頂きました音声翻訳通信研究所の皆様に感謝します．

参考文献

[1] 黒岩 眞吾, 内藤 正樹, 武田 一哉, 谷戸 文廣, 山本 誠一: "N ベスト意味探索と再評価法を用いた大規模内線受付装置の試作," 信学論 (D-II), **J79-D-II**, No. 12, pp. 2132–2138, (1996).

[2] 西村 雅史, 伊藤 伸泰: "単語を単位とした日本語ディクテーションシステム," 信学論 (D-II), **J81-D-II**, No. 1, pp. 10–17, (1998).

[3] 村上 仁一, 嵯峨山 茂樹: "自由発話音声における音響的な特徴の検討," 信学論 (D-II), **J78-D-II**, No. 12, pp. 1741–1749, (1995).

[4] 小坂 哲夫, 松永 昭一: "発話 / 話者適応による自由発話中の音素認識," 音講論集, **2–5–4**, pp. 37–38, (1995–3).

[5] 黒岩眞吾, 武田一哉, 井ノ上直己, 山本誠一: "機械との対話における発話分析 – 内線電話受付システムにより収集した対話データの分析 –," 信学技報, **SP94-30**, pp. 57–64, (1994).

[6] M. Ostendorf and H. Singer: "HMM topology design using maximum likelihood successive state splitting," Computer Speech and Language, Vol. 11, No. 1, pp. 17–41, 1997.

[7] 政瀧 浩和, 松永 昭一, 匂坂 芳典. "品詞および可変長単語列の複合 n-gram の自動生成. 信学論, Vol. J81-D-II, No. 9, pp. 1929–1936, (1998).

[8] 山本 博史, 匂坂 芳典: "単語の方向性を考慮した多重クラス複合 N-gram 言語モデル," 信学技報, **SP98-102**, pp. 49–54, (1998–12).

[9] 菅谷 史昭, 竹澤 寿幸, 横尾 昭男, 山本 誠一: "日英双方向音声翻訳システム (ATR-MATRIX) の対話実験," 音講論集, **3–1–6**, pp. 107–108, (1999–3).

[10] 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法による隠れマルコフ網の自動生成," 信学論 (D-II), **J76-D-II**, No 10, pp. 2155–2164 Oct 1993.

[11] P. Chou. "Optomal partitioning for classification and regression trees," IEEE trans. PAMI, 13(4) pp. 340–354, 1991.

[12] C.L. Leggetter and P.C. Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," Computater Speech and Language, Vol. 9, pp. 171–185, (1995).

[13] M. Tonomura, T. Kosaka and S. Matsunaga: "Speaker Adaptation Based on Transfer Vector Field Smoothing Using Maximum a Posteriori Probability Estimation," Computater Speech and Language, Vol. 10, No. 2, pp. 117–132, (1996).

[14] 小坂哲夫, 松永昭一, 嵯峨山茂樹: "木構造話者クラスタリングを用いた話者適応," 信学論, Vol. **J78-D-II**, No. 1, pp. 1–9, (1995).

[15] P.F. Brown, V.J.D. Pietra, P.V. de Souza, J.C. Lai, and R.L. Mercer: "Class-based n-gram models of natural language," Computational Linguistics, Vol. 18, No. 4, pp. 467–479, (1992).

[16] Shuanghu Bai, Haizhou Li, and Baosheng Yuan: "Building class-based language models with contextual statistics," Proc. ICASSP, pp. 173–176, (1998).

[17] 清水徹, 山本博史, 政瀧浩和, 松永昭一, 匂坂芳典: "大語い連続音声認識のための単語仮説数削減," 信学論, Vol. **J79-D-II**, No. 12, pp. 2117–2124, (1996).

[18] H. Ney and X. Aubert: "A word graph algorithm for large vocabulary, continuous speech recognition," Proc. ICSLP, pp. 1355–1358, (1994).

[19] K. Yamaguchi, H. Singer, S Matsunaga, S. Sagayama; "Speaker-Consistent Parsing for Speaker-Independent Continuous Speech Recognition," IEICE trans. INF.&SYST, Vol. **E78-D**, No. 6, pp. 470–475, (1995).

[20] H. Singer, R. Gruhn, and Y. Sagisaka: "Speech translation anywhere: Client-server based ATR-MATRIX," 音講論集, pp. 165–166, (1999).

[21] T. Takezawa, T. Morimoto, Y. Sagisaka: "Speech and Language Databases for Speech Translation Research in ATR," Proc. of EALREW, pp.148-155, (1998).

[22] E. Sumita, S. Yamada, K. Yamamoto, M. Paul, H. Kashioka, K. Ishikawa and S. Shirai: "Solutions to Problems Inherent in Spoken-Language Translation: the ATR-MATRIX Approach," Proc. of MT Summit VII, pp.229-235, (1999).

[23] W.N. Campbell: "CHATR: A high-definition speech re-sequencing system," Proc. of ASA/ASJ Joint Meeting, pp.1223-1228, (1996).

[24] 内藤 正樹, 政瀧 浩和, シンガー ハラルド, 塚田 元, 匂坂 芳典: "日英音声翻訳システム ATR-MATRIX における音声認識用音響・言語モデル," 音講論集, **2–Q–20**, (1998–3).

[25] シンガー ハラルド, ライナー グルーン, 塚田 元, 内藤 正樹, 西野 敦士, 中村 篤, 匂坂 芳典: "Cellular-Phone based Speech Translation System ATR-MATRIX," 音講論集, **1–Q–26**, (2000–3).

付録 A 音素書き起こしファイル作成法による比較

従来, ATR においては人手により付与された音素書き起しに従い、音響モデルの学習を行っていた。しかし、音素書き起こしには多くの人手を要することから、日本語書き起こしや、形態素情報など、より上位の情報のみを用いて音響モデルの学習が可能となれば、音響モデルの学習に用いる音声データベースの作成作業の効率化が可能となる。この観点から、音響モデルの学習に用いる音素書き起こしとして、(0) 人手により付与された音素書き起こしを用いた場合と、(1) 手動でなされた形態素と認識辞書を元に、各音声ファイルに対応する音素系列の候補を FSA で表し、Viterbi アルゴリズムにより最適な音素系列を選択することで作成した音素書き起こしを用いた場合、さらに、(2) 1 で学習された音響モデルを用い, 1 と同様な処理を行い作成した音素書き起こしを用いた場合について認識実験を行い性能の比較を行った。この結果、形態素情報と認識辞書を用い作成した音素書き起こしを用い音響モデルを学習した場合でも認識性能の大きな劣化は見られなかった。同様な、音素書き起こしの作成法を、自動形態素解析手法と組み合わせることで音響モデル学習用音声データベース整備の更なる効率化が可能と期待される事から、今後更なる検討が必要である。

表 7: 音素書き起しファイルの作成法が音響モデルの認識性能に与える影響

	単語認識率
(0) 手動音素書き起こし	87.66%
(1) 手動形態素 + 認識辞書+音声認識	87.05%
(2) (1) のモデルで再度音素書き起こし	87.29%

音響モデルの格納ディレクトリ

1. 手動音素書き起こしを用いた場合
   /RR/Recognition/ResearchJ/amodel/19991111/trainSLFfromTRS/
2. 手動形態素 + 認識辞書+音声認識
   /RR/Recognition/ResearchJ/amodel/19991111/trainSLFfromJMOR
3. (1) のモデルで再度音素書き起こし
   /RR/Recognition/ResearchJ/amodel/19991111/trainSLFfromJMOR.itr1

認識結果の格納ディレクトリ

1. 手動音素書き起こしを用いた場合
   /RR/Recognition/ResearchJ/V9/result/HowToMakeSLF/recSLFfromTRS

2. 手動形態素 + 認識辞書＋音声認識
   /RR/Recognition/ResearchJ/V9/result/HowToMakeSLF/recSLFfromJMOR

3. (1) のモデルで再度音素書き起こし
   /RR/Recognition/ResearchJ/V9/result/HowToMakeSLF/recSLFfromJMOR.itr1

付録 B　音響モデル格納ディレクトリ一覧

トップディレクトリ：/RR/Recognition/ResearchJ/amodel/19991111

ディレクトリ名	Topology の学習	出力確率分布の学習法
1) trainSLFfromTRS/	自然発話音声	自然発話音声で学習
2) trainSLFfromTRS.TRA2BLA/	自然発話音声	自然発話音響モデルを初期モデルとして、朗読発話音声で連結学習 (30 回)
3) trainSLFfromTRS.TRA2BLA.fromLBL/	自然発話音声	朗読発話音声でラベル学習から行う
4) trainSLFfromTRS.TRA+BLA/	自然発話音声	状態ごとに自然発話音響モデルと朗読発話音響モデルを合成
5) trainSLFfromTRS.retrainTRA+BLA/	自然発話音声	自然発話音声と朗読発話音声を併せてラベル学習から行う
a) trainSLFfromTRS.BLA/	朗読発話音声	朗読発話音声で学習
b) trainSLFfromTRS.BLA2TRA/	朗読発話音声	朗読発話音響モデルを初期モデルとして、自然発話音声で連結学習 (30 回)
c) trainSLFfromTRS.BLA2TRA.fromLBL/	朗読発話音声	自然発話音声でラベル学習から行う
d) trainSLFfromTRS.BLA.retrainTRA+BLA/	朗読発話音声	自然発話音声と朗読発話音声を併せてラベル学習から行う

なお、先の ATRSPREC の性能評価には音響モデルは (a),(b) を用いた。

使用推奨モデルは以下の通りである。

- 自然発話音声の認識： b) trainSLFfromTRS.BLA2TRA/
- 朗読発話音声の認識： a) trainSLFfromTRS.BLA/
- 自然・朗読発話が混在する音声の認識： d) trainSLFfromTRS.BLA.retrainTRA+BLA/

予備実験の結果、Topology 学習に朗読発話音声を用いた方 (a ～ d) が、自然発話音声を用いる (1 ～ 5) よりも高い認識率が得られることを確認している。また、自然・朗読発話が混在する音声の認識においては (d) の方が (c) よりも高い認識率が得られることを確認している。

付録 C　音響モデルに関する音響分析条件

音響分析条件については予備実験を通じて、V8 以前のものと比べ、主に次の点に変更がある。

- パワーを使用しない（V8 以前は 26 次元）。
- デルタの窓を前後 2 フレームにした（V8 以前は前後 4 フレーム）。
- デルタの窓を三角窓から方形窓にした。
- フィルターバンクの次数を 20 にした（V8 以前は 16）。

以下に configuration file の例を示す。

```
I/Ocontrol:inputFormat=NULL
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NULL
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:rpcNumber=3

ATRwave2cep:CentroidFreqType=linear
ATRwave2cep:CentroidFreqGamma=0.5
ATRwave2cep:CentroidFreqOrder=0
ATRwave2cep:AnalysisType=fft
ATRwave2cep:CutoffHighFrequency=8000.0
ATRwave2cep:CutoffLowFrequency=0.0
ATRwave2cep:FilterBankOrder=20
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:LpcOrder=16
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:SamplingFrequency=16000
```

ATRwave2cep:FrameShift=10

ATRwave2cep:FrameLength=20

ATRwave2cep:Preemphasis=0.98

ATRwave2cep:DebuggingLevel=0

ATRcep2para:OutputParameter=cep(12)+dpow+dcep(12)

ATRcep2para:rho=1.0

ATRcep2para:DDCepstrumPadding=zero

ATRcep2para:DDCepstrumWindow=5

ATRcep2para:deltaCepstrumPadding=zero

ATRcep2para:DeltaCepstrumWindow=5

ATRcep2para:CepstrumOrder=12

ATRcep2para:DebuggingLevel=10

ATRcep2para:WindowType=rectangular

付録 D　主な認識実験結果格納ディレクトリ一覧

トップディレクトリ：/RR/Recognition/ResearchJ/V9/result/SpeechStyle

認識対象データ別サブディレクトリ

TRA	自然発話 (TRA テストセット S1, S2, S4)
SLTA1	自然発話 (SLDB テストセット SLTA1)
MDB	SLTA1 の読み上げ音声
CDB05	ATR-MATRIX 対話実験音声 (/DB/MDB/CDB05)
CDB09	ATR-MATRIX 対話実験音声 (/DB/MDB/CDB09)

なお、ATR-MATRIX 対話実験による性能評価においては CDB05 を用いた。

各ディレクトリには、認識に使用した音響モデルごとにサブディレクトリを作成し認識結果を格納している。

以下に、本文中に示した認識実験結果 + α を格納したサブディレクトリを示す。

模擬対話音声による認識実験 (TRA,SLTA,MDB)

(a) 自然発話音響モデル	$ 認識タスク名 /ModelBLA2TRA.itr30.P=TRA
(b) 朗読発話音響モデル	$ 認識タスク名 /ModelBLA
(c) (a),(b) から動的選択	$ 認識タスク名 /ModelBLA2TRA+BLA.itr30
(d) 自然発話と朗読で学習したモデル	$ 認識タスク名 /ModelBLA.retrainTRA+BLA

ATR-MATRIX 対話実験による評価 (CDB05)

(a) 自然発話音響モデル	CDB05/ModelBLA2TRA.itr30.TRS.mrg
(b) 朗読発話音響モデル	CDB05/ModelBLA.TRS.mrg
(c) (a),(b) から動的選択	CDB05/ModelBLA2TRA+BLA.itr30.TRS.mrg
(d) 自然発話と朗読で学習したモデル	CDB05/ModelBLA.retrainTRA+BLA

各サブディレクトリの下には,M5mixF5mix, というディレクトリがあり、この下に、認識結果の log ファイル、ラティス等が存在している。なお、ATR-MATRIX 対話実験による評価については、各テストセット（初回、2 回目、練習後）それぞれに対して認識結果の log ファイル

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL1

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL2

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL3

が作成されている。また 1 発話目のみに対する認識率は、各サブディレクトリの下の

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL1-NEW1ST

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL2-NEW1ST

log.lm1-6.lm2-9.bm-90.wp-5000.TRIAL3-NEW1ST

を参照してください。

付録 E　音声翻訳対話テストセット

初回の対話:/RR/Recognition/ResearchJ/V9/result/SpeechStyle/CDB05/ModelBLA2TRA+BLA.itr30.TRS.mrg/select/list1

TAM00721.0010.A : ニューヨークシティーホテル です か
TAM00721.0040.A : 二 月 七 日 に 一 泊 ホテル の 方 予約 し たい の です が
TAM00721.0150.A : はい シングルルーム 一 名 で お 願い し ます
TAM00721.0190.A : 一 泊 です
TAM00721.0230.A : ちがい ます
TAM00721.0310.A : 一 名 です
TAM00721.0330.A : 百 五 十 ドル の 部屋 は 在り ません か
TAM00721.0350.A : お 願い し ます
TAM00721.0390.A : 一 泊 百 四 十 ドル で お 願い し ます
TAM00721.0430.A : はい スズキ カズコ です
TAM00721.0510.A : はい 電話番号 は 八 一 の 二 三 五 三 七 の 一 九 二 八 です
TAM00721.0550.A : はい けっこう です
TAM00721.0570.A : クレジットカード で お 願い し ます
TAM00721.0610.A : ビザカード です
TAM00721.0630.A : 番号 は 八 二 九 四 零 九 八 二 一 九 四 八 一 二 七 八 です
TAM00721.0700.A : ちがい ます もう 一 度 言い ます
TAM00721.0710.A : 八 二 九 四 零 九 八 二 一 九 四 八 一 二 七 八 です
TAM00721.0830.A : もう 一 度 最初 から 言っ ていただけ ます か
TAM00721.0860.A : 前半 の 八 桁 が わかり ません
TAM00721.1050.A : 番号 が ちがい ます
TAM00721.1230.A : はい そう です
TAM00721.1250.A : 千 九 百 九 十 九 年 十 月 です
TAM00721.1300.A : はい そう です
TAM00721.1320.A : ゆうがた 六 字 頃 です
TAM00721.1360.A : ありがとうございます
TAM00731.0010.A : はい もしもし すみません ニューヨークシティーホテル でしょう か
TAM00731.0070.A : 二 月 七 日 の お 部屋 が 開い ている かどうか 確認 し たい の です けれども
TAM00731.0110.A : はい 名前 は スズキ カズコ です
TAM00731.0150.A : はい そう です
TAM00731.0160.A : はい 八 一 の 六 七 三 の 三 零 六 七 三 八 です

TAM00731.0210.A : もしもし もう 一 度 言え ば よい の でしょう か
TAM00731.0390.A : 一 名 です
TAM00731.0440.A : 一 泊 です
TAM00731.0490.A : はい そう です
TAM00731.0520.A : 予算 は 百 五 十 ドル くらい な の です が
TAM00731.0540.A : クレジットカード で お 願い し ます
TAM00731.0560.A : ビザカード です
TAM00731.0590.A : 五 一 九 三 七 七 二 零 六 三 一 五 一 二 九 七 です
TAM00731.0690.A : 千 九 百 九 十 九 年 の 十 月 です
TAM00731.0770.A : はい お 願い し ます
TAM00731.0780.A : だいたい ゆうがた の ゆうがた の 六 字 頃 の 予定 です
TAM00731.0820.A : はい 正しい です
TAM00731.0840.A : 会っ てい ます
TAM00731.0860.A : ゆうがた の 六 字 くらい です
TAM00731.0920.A : 予算 が 十 わたし 予算 が 百 五 十 ドル ほど で お 願い し たい の です が
TAM00731.1020.A : はい
TAM00741.0010.A : ニューヨークシティーホテル です か
TAM00741.0030.A : 宿泊 の 予約 を し たい の です が
TAM00741.0050.A : はい スズキ タケシ と 申し ます
TAM00741.0100.A : 二 月 七 日 金曜日 一 泊 で お 願い し ます
TAM00741.0140.A : シングル で お 願い し ます
TAM00741.0170.A : 一 名 です
TAM00741.0190.A : もう 少し 安い 部屋 は 在り ます か
TAM00741.0240.A : その 部屋 で お 願い し ます
TAM00741.0300.A : 金曜日 お 願い し ます
TAM00741.0460.A : 二 月 八 日 土曜日 です
TAM00741.0500.A : はい お 願い し ます
TAM00741.0520.A : はい スズキ タケシ です
TAM00741.0580.A : はい 八 一 七 四 二 四 五 三 九 一 三 です
TAM00741.0820.A : はい よろしい です
TAM00741.0840.A : カード で お 願い し ます
TAM00741.0860.A : クレジットカード で お 願い し ます
TAM00741.0920.A : ビザカード です
TAM00741.0940.A : 三 七 四 八 零 九 一 二 四 九 二 一 三 六 八 二 です

TAM00741.1060.A : はい よろしい です
TAM00741.1080.A : はい 千 九 百 九 十 九 年 十 月 です
TAM00741.1100.A : ゆうがた 六 字 頃 です
TAM00741.1120.A : お 願い し ます
TAM00751.0010.A : もしもし ニューヨークシティーホテル さま でしょう か
TAM00751.0030.A : 予約 の 金曜日 です が 二 月 の 七 日 金曜日 一 泊 二 日 お 願い し たい の です が よろしい でしょう か
TAM00751.0050.A : 金曜日 です
TAM00751.0070.A : 一 泊 お 願い し ます
TAM00751.0090.A : 人数 です が 一 名 で 一 名 で お 願い し たい の です が
TAM00751.0130.A : シングル の 部屋 を お 願い し たい の です けれども よろしい でしょう か
TAM00751.0150.A : 予算 の 方 です けれども 百 ドル で お 願い し たい 部屋 を 捜し ている の です けれども 在り ます でしょう か
TAM00751.0200.A : もう 一 度 お 願い し ます ちょっと 聞こえ ない ので お 願い
TAM00751.0220.A : シングル 百 四 十 ドル でしょう か
TAM00751.0250.A : 一 名 で シングル の 部屋 は ございます でしょう か もう 一 度 お 願い し ます
TAM00751.0270.A : それ は シングル でしょう か ダブル の 部屋 な の でしょう か
TAM00751.0280.A : シングル の 部屋 で お 願い し ます
TAM00751.0310.A : 二 月 八 日 土曜日 の 一 泊 二 日
TAM00751.0320.A : じゃ その 部屋 で お 願い でき ます でしょう か
TAM00751.0350.A : わたくし は スズキ タケシ と 申し ます が あなた の 名前 は
TAM00751.0370.A : はい お 願い し ます
TAM00751.0390.A : 電話番号 です けれども よろしい でしょう か
TAM00751.0420.A : 申しあげ ます 七 一 六 一 七 三 零 六 八 三 九 です
TAM00751.0460.A : クレジットカード で お 願い し たい の です が よろしい でしょう か
TAM00751.0500.A : 使用 する カード です が ビザカード で お しはらい に なり ます が よろしい でしょう か
TAM00751.0560.A : では ビザカード の 登録 し ている ナンバー 申しあげ ます
TAM00751.0570.A : はい わかり まし た 申しあげ ます よろしい でしょう か 五 一 九 三 七 七 二 零 六 三 一 五 一 二 九 七 です
TAM00751.0600.A : もしもし 番号 が ちょっと ちがい ます ので もう 一 度 申しあげ ます けれども よろしい でしょう か
TAM00751.0670.A : カード の 有効 期限 です 九 十 九 年 の 十 月 です

TAM00751.0690.A : 二 月 八 日 土曜日 の ゆうがた 六 字 頃 そちら の 方 へ 参り ます けれども よろしい でしょう か
TAM00751.0710.A : では もう 一 度 確認 の ため 申し上げ ていただけ ます でしょう か
TAM00751.0730.A : はい ありがとうございました
TAM00761.0010.A : もしもし ニューヨークシティーホテル さん でしょう か
TAM00761.0060.A : こんど の 二 月 七 日 金曜日 の ご 予約 予約 の 方 し たい の です けれども
TAM00761.0070.A : お 部屋 の 方 は 開い て ます でしょう か
TAM00761.0140.A : 百 五 十 ドル の 部屋 は 在り ません か
TAM00761.0160.A : そしたら その 百 四 十 ドル の 部屋 を お 願い し ます
TAM00761.0200.A : スズキ タケシ です
TAM00761.0230.A : 七 一 の 六 一 七 の 三 零 の 六 八 三 九 です
TAM00761.0290.A : クレジットカード で お 願い し ます
TAM00761.0310.A : ビザカード です けれども よろしい でしょう か
TAM00761.0330.A : 五 一 九 三 七 七 二 零 六 三 一 五 一 二 九 七 です
TAM00761.0350.A : すみません 番号 間違え まし た もう 一 回 言い ます ので もう 一 回 メモ お 願い し ます
TAM00761.0370.A : すみません
TAM00761.0410.A : 千 九 百 九 十 九 年 十 月 です
TAM00761.0430.A : ゆうがた の 六 字 頃 です
TAM00761.0460.A : ありがとうございます 失礼します

2 回目の対話:/RR/Recognition/ResearchJ/V9/result/SpeechStyle/CDB05/ModelBLA2TRA+BLA.itr30.TRS.mrg/select/list2

TAM00722.0010.A : ニューワシントンホテル です か
TAM00722.0030.A : 四 月 十 五 日 に 一 泊 ホテル を 予約 し たい の です が
TAM00722.0090.A : ツインルーム で 三 百 ドル で は 在り ません か
TAM00722.0180.A : それでは ダブル の 三 百 ドル で お 願い でき ます か
TAM00722.0220.A : ダブル の 部屋 が 三 百 ドル です ね
TAM00722.0240.A : ツインルーム は 三 百 六 十 ドル です ね
TAM00722.0260.A : それでは ダブル の 部屋 を お 願い し ます
TAM00722.0280.A : 一 泊 です
TAM00722.0300.A : 二 名 です
TAM00722.0320.A : タナカ カズコ です
TAM00722.0400.A : 電話番号 です か

TAM00722.0420.A : 八 一 六 一 三 五 四 零 一 九 零 です
TAM00722.0500.A : マスターカード で お 願い し ます
TAM00722.0520.A : 一 三 九 九 五 零 三 九 一 九 零 零 三 三 九 零 です
TAM00722.0560.A : はい その 通り です
TAM00722.0580.A : 二 千 一 年 三 月 です
TAM00722.0740.A : 三 字 頃 に なり ます
TAM00722.0770.A : ありがとうございました
TAM00732.0010.A : ニューワシントンホテル でしょう か
TAM00732.0050.A : ホテル の 部屋 を 予約 し たい の です けれども
TAM00732.0070.A : はい タナカ カズコ と 申し ます
TAM00732.0090.A : まず 電話番号 が 八 一 の 六 一 三 五 四
TAM00732.0110.A : 二 人 です
TAM00732.0140.A : そう です
TAM00732.0150.A : はい 電話番号 は 八 一 六 一 三 五 四 の 零 一 九 零 です
TAM00732.0190.A : 正しい です
TAM00732.0320.A : ツインルーム が 希望 な の です が ツインルーム は 開い てい ます か
TAM00732.0360.A : 日にち は 四 月 十 五 日 の 水曜日 です
TAM00732.0380.A : はい そう です
TAM00732.0410.A : 部屋 は 開い てい ます か
TAM00732.0500.A : 四 月 十 五 日 水曜日 一 泊 を 希望 し ます
TAM00732.0520.A : ツイン が 希望 です
TAM00732.0610.A : はい そう です
TAM00732.0640.A : 三 百 ドル くらい の 部屋 は 在り ません か
TAM00732.0670.A : それでしたら ダブル の 部屋 の 方 に し ていただけ ます か
TAM00732.0700.A : はい それ で けっこう です
TAM00732.0720.A : わたし の カード は マスターカード です
TAM00732.0740.A : はい 一 三 九 九 五 零 三 九 一 九 零 零 三 三 九 零 です
TAM00732.0820.A : はい そう です
TAM00732.0840.A : はい 有効 期限 は 二 零 零 一 年 の 三 月 です
TAM00732.0920.A : はい そう です
TAM00732.0940.A : だいたい 三 字 頃 に なる と おもい ます
TAM00742.0010.A : すみません ここ は ニューワシントンホテル です か
TAM00742.0040.A : 宿泊 の 予約 を 申込み たい の です が
TAM00742.0060.A : はい 名前 は タナカ カズオ と 言い ます 電話番号 は 七 一 六 一

三 五 四 零 一 九 零 です
TAM00742.0130.A : すみません 七 一 六 一 三 五 四 零 一 九 零 です
TAM00742.0160.A : いいえ 在り ません
TAM00742.0220.A : よろしい です
TAM00742.0240.A : 四 月 十 五 日 です
TAM00742.0260.A : 一 泊 で お 願い し ます
TAM00742.0280.A : 二 名 です
TAM00742.0300.A : ツイン で お 願い し ます
TAM00742.0320.A : もう 少し 安い 部屋 は 在り ます か
TAM00742.0360.A : では その 部屋 を お 願い し ます
TAM00742.0390.A : 二 百 五 十 ドル の 部屋 を
TAM00742.0410.A : はい お 願い し ます
TAM00742.0460.A : ツイン で 二 百 五 十 ドル の 部屋 は 在り ます か
TAM00742.0480.A : クレジットカード で お 願い し ます
TAM00742.0500.A : マスターカード です
TAM00742.0520.A : 一 三 九 九 五 零 三 九 一 九 零 零 三 三 九 零 です
TAM00742.0550.A : すみません
TAM00742.0560.A : 一 九 零 零 三 三 九 零 です
TAM00742.0910.A : はい よろしい です
TAM00742.0930.A : 千 九 百 九 十 九 年 三 月 です
TAM00742.0950.A : 三 字 頃 です
TAM00742.1050.A : はい 正しい です よろしい です
TAM00742.1070.A : よろしく お 願い し ます
TAM00752.0020.A : もしもし ワシントンホテル さま でしょう か
TAM00752.0060.A : ホテル の 予約 の 方 お 願い し たい の です けれども よろしい でしょう か
TAM00752.0110.A : 一 泊 二 日 で ホテル の 予約 の 方 お 願い し たい の です けれども よろしい でしょう か
TAM00752.0150.A : 日にち の 方 です が 四 月 十 五 日 水曜日 一 泊 二 日 で 二 名 で お 願い し たい です が
TAM00752.0170.A : 部屋 の 種類 です けれども よろしい です か
TAM00752.0300.A : 二 名 で ツイン で お 願い し たい の です けれども
TAM00752.0340.A : ダブル で 三 百 ドル で は どう でしょう か ございます でしょう か
TAM00752.0360.A : では ダブル で 二 百 五 十 ドル で お 願い はい その 部屋 で お 願い でき ます でしょう か

TAM00752.0380.A : では 確認 し ます よろしい でしょう か
TAM00752.0410.A : はい では これ の 部屋 で お 願い でき ます でしょう か
TAM00752.0440.A : はい 予約 の 方 お 願い し ます
TAM00752.0450.A : はい お 願い し ます
TAM00752.0480.A : わたくし は タナカ カズオ と 申し ます
TAM00752.0500.A : 電話番号 を 申しあげ ます
TAM00752.0530.A : 七 一 六 一 三 五 四 零 一 九 零
TAM00752.0590.A : 番号 が ちょっと 違っ てい ます ので もう 一 度 申しあげ ます よろしい でしょう か
TAM00752.0730.A : クレジットカード で お 願い し たい の です けれども よろしい でしょう か
TAM00752.0750.A : では マスターカード で マスターカード です
TAM00752.0800.A : 一 三 九 九 五 零 三 九 一 九 零 零 三 三 零 失礼しました 三 三 九 零 です
TAM00752.0910.A : では カード の 有効 期限 を
TAM00752.0950.A : 九 十 九 年 の 三 月 です
TAM00752.0990.A : まちがい ございません 会っ とり ます
TAM00752.1000.A : ゆうがた の 三 字 頃 お 願い し たい の です が よろしい でしょう か
TAM00752.1040.A : わかり まし た ありがとうございました
TAM00762.0010.A : もしもし ワシントンホテル さん でしょう か
TAM00762.0070.A : 予約 を し たい の です けれども よろしい でしょう か
TAM00762.0090.A : 二 名 で ツイン お 願い し ます
TAM00762.0120.A : 四 月 十 五 日 の 水曜日 一 泊 お 願い し ます
TAM00762.0140.A : 三 百 ドル より 安い の 在り ません か
TAM00762.0160.A : そしたら そちら の 方 の 部屋 を お 願い し ます
TAM00762.0190.A : タナカ カズオ です
TAM00762.0230.A : 七 一 六 一 三 五 四 零 一 九 零 です
TAM00762.0260.A : クレジットカード で お 願い し ます
TAM00762.0310.A : マスターカード です けれども よろしい でしょう か
TAM00762.0330.A : 一 三 九 九 五 零 三 九 一 九 零 零 三 三 九 零 です
TAM00762.0360.A : 九 十 九 年 三 月 です
TAM00762.0380.A : 三 字 頃 で お 願い し ます
TAM00762.0400.A : ありがとうございます さようなら

練習後の対話:/RR/Recognition/ResearchJ/V9/result/SpeechStyle/CDB05/ModelBLA2TRA+BLA.itr30.TRS.mrg/select/list3

TAM00723.0010.A : ペニンシュラホテル です か
TAM00723.0030.A : 四 月 三 十 一 日 に 部屋 を 予約 し たい の です が
TAM00723.0070.A : ツインルーム を お 願い し ます
TAM00723.0170.A : 海 の 見える 部屋 の ねだん を もう 一 度 聞か せ てください
TAM00723.0190.A : ちがい ます
TAM00723.0290.A : それでは 街 の 見える 部屋 を お 願い し ます
TAM00723.0330.A : 一 泊 です
TAM00723.0370.A : それでは ツインルーム を お 願い し ます
TAM00723.0390.A : はい サトウ ヨウコ です
TAM00723.0530.A : はい 八 一 三 五 六 八 八 二 三 五 七 です
TAM00723.0580.A : 最初 の 数字 が 聞きとれ ません
TAM00723.0600.A : はい そう です
TAM00723.0630.A : 二 名 です
TAM00723.0650.A : ツインルーム に 二 名 です
TAM00723.0670.A : アメックス で お 願い し ます
TAM00723.0690.A : 一 八 九 三 五 六 七 零 三 七 三 一 四 八 八 七 です
TAM00723.0780.A : はい そう です
TAM00723.0800.A : 有効 期限 は 二 千 二 年 六 月 です
TAM00723.0830.A : はい そう です
TAM00723.0850.A : 四 字 頃 に なり ます
TAM00723.0870.A : はい そう です
TAM00723.0890.A : ありがとうございます
TAM00733.0010.A : ペニンシュラホテル です か
TAM00733.0030.A : はい ホテル の 予約 を し たい の です けれども
TAM00733.0050.A : はい タナカ カズコ です
TAM00733.0070.A : 四 月 二 十 七 日 の 火曜日 に 一 泊 滞在 し たい の です が
TAM00733.0090.A : 一 泊 です
TAM00733.0110.A : はい 一 泊 です
TAM00733.0130.A : 二 名 です
TAM00733.0160.A : 予算 が 三 百 五 十 ドル ほど な の です が
TAM00733.0190.A : 海 の 見える 部屋 は
TAM00733.0260.A : はい じゃ お 願い し ます
TAM00733.0330.A : 他 に 三 百 五 十 ドル ほど で 泊まれる 部屋 は 在り ます か

TAM00733.0350.A : そちら の 部屋 は いくら です か
TAM00733.0370.A : じゃ そちら で お 願い し ます
TAM00733.0400.A : はい 三 百 十 ドル の 部屋 で お 願い し ます
TAM00733.0430.A : はい クレジットカード で お 願い し ます
TAM00733.0450.A : はい アメックス です
TAM00733.0480.A : はい ナンバー は 一 八 九 三 五 六 七 零 三 七 三 一 四 八 二 七 です
TAM00733.0530.A : もう 一 度 お 願い し ます
TAM00733.0710.A : はい そう です
TAM00733.0750.A : はい なん でしょ
TAM00733.0770.A : はい 千 はい 千 九 百 九 十 九 年 六 月 まで です
TAM00733.0790.A : はい 電話番号 は 八 一 三 五 六 八 八 五 六 五 七 です
TAM00733.0810.A : その 通り です
TAM00733.0830.A : 四 字 頃 に なる と おもい ます
TAM00733.0860.A : 予約 は これ で オーケー です か
TAM00733.0910.A : それでは よろしく お 願い し ます
TAM00733.0930.A : はい ありがとうございました
TAM00743.0010.A : すみません ここ は ペニンシュラホテル です か
TAM00743.0030.A : 宿泊 の 予約 を し たい の です が
TAM00743.0070.A : タナカ タケシ です
TAM00743.0120.A : 電話番号 は 七 一 三 五 六 八 八 五 六 五 七 です
TAM00743.0170.A : はい よろしい です
TAM00743.0190.A : 四 月 二 十 七 日 一 泊 で お 願い し ます
TAM00743.0210.A : 二 名 で お 願い し ます
TAM00743.0260.A : ツイン で お 願い し ます
TAM00743.0280.A : はい はい よろしい です
TAM00743.0340.A : 海 の 見える 部屋 を お 願い し ます
TAM00743.0390.A : もう 少し 安い 部屋 は 在り ます か
TAM00743.0420.A : じゃ お 願い し ます
TAM00743.0440.A : 海 の 見える 部屋 で お 願い し ます
TAM00743.0700.A : はい お 願い し ます
TAM00743.0720.A : はい よろしい です
TAM00743.0750.A : 二 名 で お 願い し ます
TAM00743.0790.A : はい お 願い し ます
TAM00743.0800.A : よろしい です

TAM00743.0820.A : クレジットカード で お 願い し ます
TAM00743.0860.A : アメックス です
TAM00743.0880.A : はい 一 八 九 三 五 六 七 零 三 七 三 一 四 八 二 七
TAM00743.1030.A : はい 千 九 百 九 十 九 年 六 月 です
TAM00743.1100.A : 四 字 頃 です
TAM00743.1180.A : よろしく お 願い し ます
TAM00753.0030.A : ペニンシュラホテル さま でしょう か
TAM00753.0080.A : ホテル の ホテル の 予約 を お 願い し たい の です けれども よろしい でしょう か
TAM00753.0120.A : では 四 月 二 十 四 月 二 十 七 日 火曜日 一 泊 二 日 で お 願い し たい の です けれども よろしい でしょう か
TAM00753.0140.A : 部屋 です けれども デラックスツイン で 二 名 お 願い し たい の です が よろしい でしょう か
TAM00753.0170.A : では ツイン スタンダードツイン は いくら に なり ます でしょう か もう ひとつ です けれども ツイン スタンダードツイン は いくら でしょう か
TAM00753.0200.A : では では 確認 です よろしい でしょう か
TAM00753.0250.A : では 四 月 二 十 七 日 火曜日
TAM00753.0270.A : では スタンダードツイン を お 願い し ます
TAM00753.0320.A : はい そう です お 願い し ます
TAM00753.0340.A : はい タナカ タケシ と 申し ます
TAM00753.0390.A : 電話番号 を 申しあげ ます 七 一 三 五 六 八 八 五 六 五 七 です
TAM00753.0410.A : はい まちがい ございません お 願い し ます
TAM00753.0430.A : では お しはらい の 方法 です が どのように すれ ば よろしい アメックス で お 願い し ます
TAM00753.0450.A : 一 八 九 三 五 六 七 零 三 七 三 一 四 八 二 七 です
TAM00753.0480.A : もう 一 度 お 願い し ます
TAM00753.0500.A : 九 十 九 年 の 六 月 です
TAM00753.0540.A : ゆうがた の 四 字 頃 お 願い し ます
TAM00753.0560.A : はい そう です
TAM00753.0580.A : では よろしく
TAM00753.0600.A : はい よろしく お 願い し ます
TAM00753.0620.A : ありがとうございました
TAM00763.0010.A : もしもし ペニンシュラホテル でしょう か
TAM00763.0070.A : 予約 を し たい の です けれども よろしい でしょう か
TAM00763.0090.A : 四 月 二 十 七 日 火曜日 です

TAM00763.0110.A : 二 名 です
TAM00763.0130.A : 予算 が 三 百 五 十 ドル です ので その なか で もの 在り ません か
TAM00763.0150.A : そしたら それ で お 願い いたし ます
TAM00763.0170.A : タナカ タケシ です
TAM00763.0200.A : はい お 願い し ます
TAM00763.0210.A : 七 一 三 五 六 八 八 五 六 五 七 です
TAM00763.0240.A : もう 一 回 言い ます
TAM00763.0270.A : クレジットカード で お 願い し ます
TAM00763.0290.A : アメリカンエキスプレス です
TAM00763.0310.A : 一 八 九 三 五 六 七 零 三 七 三 一 四 八 二 七 です
TAM00763.0370.A : 九 十 九 年 六 月 です
TAM00763.0390.A : 四 字 頃 で お 願い し ます
TAM00763.0410.A : ありがとうございます 失礼します