TR-IT-0266

連続音声認識用音響モデル (ResearchJ V5)

柘植 覚 Satoru Tsuge
内藤 正樹 Masaki Naito
シンガー・ハラルド Harald Singer
深田 俊明 Toshiaki Fukada
高野 優 Masaru Takano

1998 年 7 月 31 日

本報告では、先に音声翻訳通信研究所に対して正式リリースした音響モデル (TR-IT-0241) よりも認識性能の高い新たな音響モデルを作成することを検討している。本報告では特に、ケプストラム平均減算法 (Cepstrum Mean Subtraction)、VTLN (Vocal Tract Length Normalization) を用いた周波数正規化手法を用いることにより、音響モデルを作成し、音素認識率、単語認識率による比較を行った。比較の結果より、「発話毎に CMS を行った特徴量を用い学習を行った、800 状態 5 混合の (1) 性別依存、(2) 非依存の音響モデル、トポロジー学習のみに発話毎に CMS を行った特徴量を使用し、音響モデルのパラメータ学習には正規化を行わない特徴量を用い学習を行った、800 状態 5 混合の (3) 性別依存、(4) 非依存の音響モデル」を音声翻訳通信研究所に対してリリースする。

目次

1 はじめに

現在、音声翻訳通信研究所にて、音声認識、翻訳、合成を統合し稼働している音声翻訳システムの精度を向上させるためには、音声認識における、音響モデルの性能向上が最も重大な課題の一つであると考えられる。本報告では、先に音声翻訳通信研究所に対しリリースを行った音響モデル Version2 (TR-IT-0241) [1] よりも高精度な音響モデルをリリースするために、単語認識率、音素認識率による性能比較を通して、正規化手法及び、男女別モデルに関する検討を行った。

正規化手法には、種々の正規化手法が提案されている中から、マイク特性等の周波数特性の正規化に有効であるケプストラム平均減算法 [2]、話者によるスペクトルの差異の正規化を行う周波数ワーピングによる話者正規化手法 [3] を用いた。

2 音声データベース

2.1 学習データ

音響モデルの学習データには、ATR の音声データベース (SDB) の男女合計 230 人総音素数約 12 万 (約 200 分) の自然発話音声 (トレーニングセット T_M_0100, T_F_0130) を用いた。表 1に学習データの詳細を示す (表 1中のファイル全体の音素数は、トランスクリプションファイルを用い計算を行い、トポロジー学習、連結学習のサンプル数、フレーム数は、各音響モデル作成時のログファイルを用いて計算を行った。)。本稿では、トポロジー学習、ラベル学習には図 1に示す無音部を含まない **A** の区間を、連結学習には、**A** に 30msec の無音を前後に付加した **B** の区間を用いた。また、全音響モデルは、HMM の最大状態数を 4 としている。そのため、音響モデルのトポロジー学習、ラベル学習時には、自動ラベリングにより 40msec 以上となったデータを用いている[1]。連結学習時には、連結学習が可能なフレーム数をもつ全学習データを用い学習を行っている。

表 1: 学習データの詳細

			音声ファイル全体		トポロジー学習		連結学習	
性別	話者数	発話数	音素数	フレーム数	サンプル数	フレーム数	サンプル数	フレーム数
男性	100	1245	49014	518527	40866	307731	53734	343102
女性	130	1717	68017	756694	60516	466051	74321	504653
男女	230	2962	117031	1275221	101382	773782	128055	847755

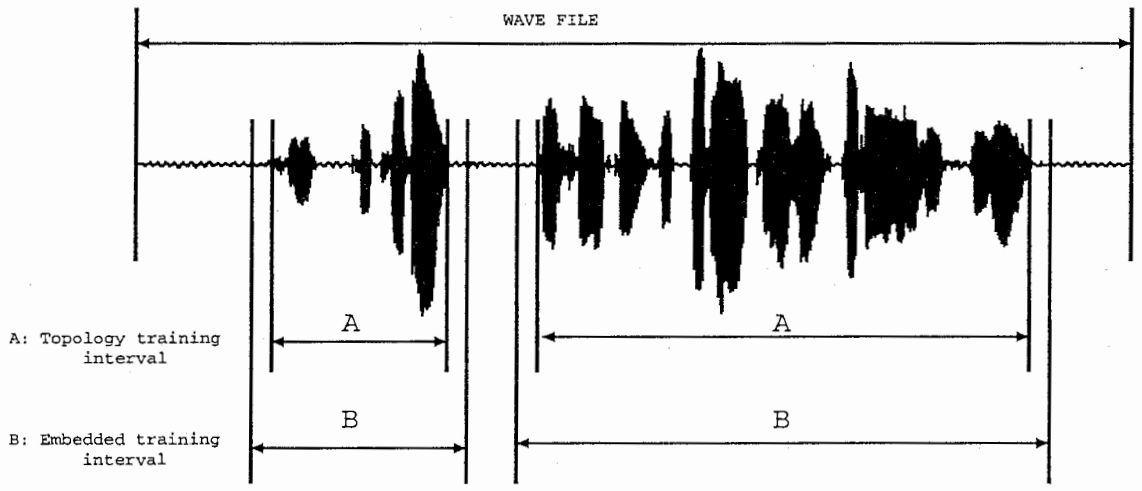

図 1: 各学習に用いる音声データ

2.2 テストデータ

テストデータは、TR-IT-0206[4]、TR-IT-0241[1] と同様に、学習データと同じデータベースより、学習に用いていない男性 17 人と女性 25 人の合計 42 人からなる 551 発声 (テストセット S1, S2, S4)、総音素数約 2 万 (約 40 分) を用いた。

このテストデータに対し、音声認識性能の評価は、日本語の音素接続制限規則を用いた音素タイプライタ [5] による音素認識率と、単語と品詞の複合 n-gram を言語モデル [6] として用いた単語認識率により行った。リリースには、TDMT 体系の言語モデル [7] を用い、評価を行った。

[1] HMM の最大状態数が 4、フレームシフトが 10msec であるため、40msec 以上のデータとなる。

3 音響モデルの作成

音響分析および音響モデルの作成は、ATR-SPREC version r05r05 を用い、計算機は DEC 社の Alpha-Station 500/500 上で行った。

3.1 音響分析

TR-IT-241 の検討により、MFCC (Mel-Frequency Cepstrum Coefficient) を特徴パラメータとして用いる。MFCC は、表 2に示す音響分析条件で求めた。更に

- CMS (Cepstrum Mean Subtraction)
- VTLN (Vocal Tract Length Normalization)

により正規化を行った MFCC を正規化特徴量として用いた。一次回帰係数 (ΔMFCC) は、100msec (9 フレーム) の三角窓を用い計算を行った [8]。それぞれの正規化手法については、4.2節にて詳細を示す。

表 2: 共通の音響分析条件

プリエンファシス	0.98
フレーム周期	10 msec
フレーム長	20 msec
分析窓	ハミング窓
フィルタバンク次数	16
MFCC 次数	12
特徴ベクトル次元数	26 (12 次 MFCC、log パワー、 12 次 Δ MFCC、Δlog パワー)

3.2 音響モデル

(3.2.1) 音素ラベルファイル

TR-IT-0206, TR-IT-0241 にて使用した音素ラベルファイルは、単語データベース (Aset) の 2,620 単語から作成した MHT の 400 状態 1 混合の HMnet を元に、各話者毎にトランスクリプションを用い話者適応を行いポーズ句毎の Viterbi セグメンテーションの結果を用いた。この Viterbi セグメンテーションに用いた HMnet は、LPC ケプストラムを特徴パラメータとして作成されたものであったが、本稿では、自然発話データベースを用い、特徴パラメータとして MFCC を用いているため、この音響モデル作成条件とは異なる。そこで、今回は MFCC を特徴パラメータとした性別依存 800 状態 5 混合モデルを作成し、このモデルを用いて Viterbi セグメンテーションを行い、音素ラベルファイルを作成した。ここで、TR-IT-0206 同様にこのモデルに対して話者適応を行い音素ラベルファイルを作成することもできるが、話者適応後の音素ラベルファイルを用い作成した音響モデルと適応無しの音素ラベルファイルで作成した音響モデルとの認識性能の差がほとんど見られなかった。そのため、本報告では、モデル作成が容易である適応を行わないモデルの Viterbi セグメンテーションの結果を音素ラベルとし用い、音響モデルの作成を行った。これらの音素ラベルファイルの違いによる認識性能の比較を付録 A に示す。

(3.2.2) 音響モデルの作成

本稿で用いる音響モデルは、尤度最大化基準逐次状態分割 (ML-SSS) アルゴリズム [9] により、27 状態の初期モデル [4] から 800 状態各 5 混合の環境依存音素 HMM を作成したものを用いた。この音響モデルの時間方向分割における最大経路長は 4 状態とした。無音モデルは、トランスクリプションファイル (*.TRS) の時間情報に基づいて切り出した無音区間を 3 状態 10 混合の HMM で学習したものを用いた[2]。音響モデルの作成手順を図 2に示す。

[2]音素モデル作成の連結学習時には、1 状態 10 混合の無音モデルを用いている。

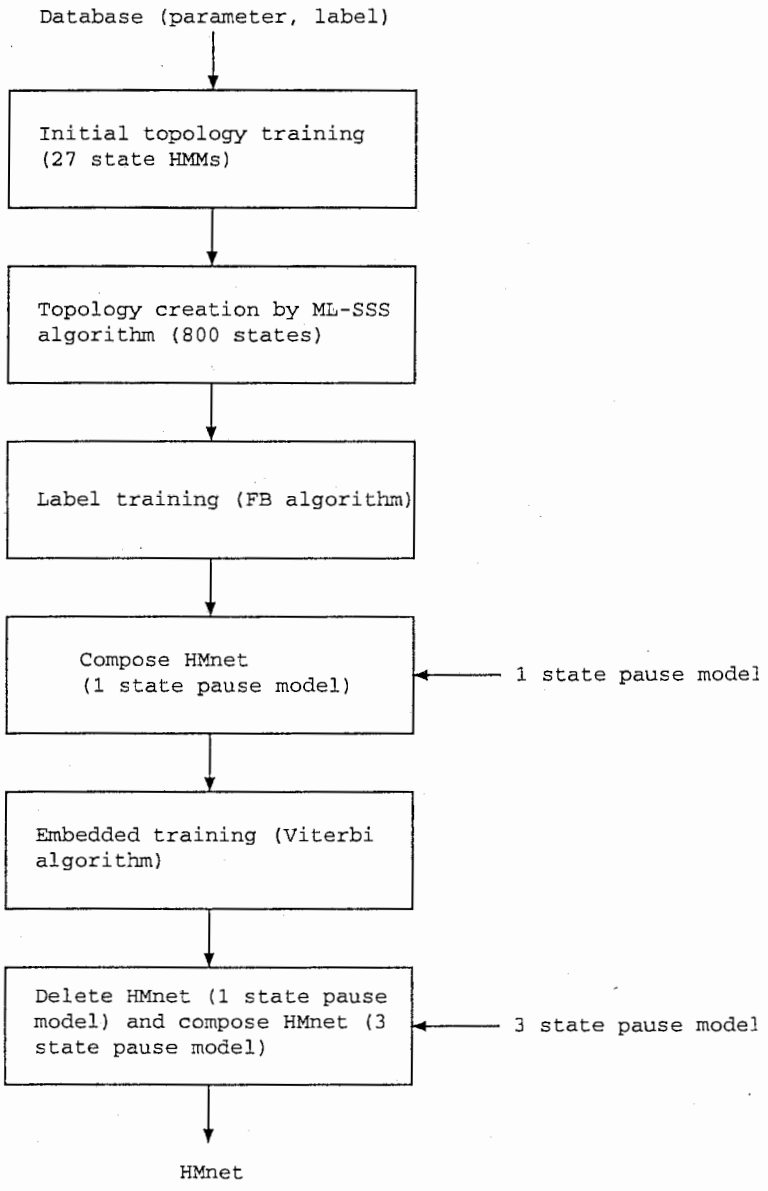

図 2: 音響モデル作成手順

4 音声認識実験

4.1 ResearchJ V3 との異なる点

ResearchJ V3 (TR-IT-0241) では、特徴パラメータ作成時に、16kHz の音声波形に srconv コマンドを用い 16kHz から 16kHz になるようにダウンサンプリング (フィルタリング) を行っていた。しかし、本稿では、この途中のフィルタリングを省き、16kHz の音声波形から直接特徴パラメータを計算を行った。そのため、ResearchJ V3 と同じ特徴量 (MFCC) であっても若干の異なりが生じる。また、認識時間削減のため、forward 探索時に同音異義語を区別せずに認識を行った。これは、認識時のコンフィグレーションのパラメータを

`ATRlattice:word_merge=all`

とすることにより設定ができる。MFCC を特徴パラメータとした音響分析のコンフィグレーションファイルと、単語、音素認識時の認識コンフィグレーションファイルを付録 B.1、B.2、B.3にそれぞれ示す。

4.2 正規化手法の比較

(4.2.1) ケプストラム平均減算法

ケプストラム平均減算法 (Cepstrum Mean Subtraction: CMS) [2] は、ある時間長のケプストラムの平均の値を現時刻のケプストラム値から減算をすることにより正規化を行う。CMS は、ケプストラムの時間軌跡の直流成分を取り除くことができ、マイクの周波数特性等の乗算性雑音を軽減することができる。CMS は以下の式で表すことができる。

$$\hat{\mathbf{C}}(\mathbf{t}) = \mathbf{C}(\mathbf{t}) - \bar{\mathbf{C}} \tag{1}$$

ただし、$\hat{\mathbf{C}}(\mathbf{t})$ は CMS を行ったケプストラム、$\mathbf{C}(\mathbf{t})$ は観測されたケプストラム、$\bar{\mathbf{C}}$ はケプストラムの平均値である。本節において、CMS の有効性を示し、ケプストラムの平均の計算方法が認識に与える影響を調べる。

本節で、比較検討を行ったケプストラムの平均の計算方法は、

1. 各話者の全発声の有音区間のケプストラムから平均を計算する (CMS_話者毎)
2. 各発話毎に発声の有音区間のケプストラムから平均を計算する (CMS_発話毎)

である。ここで、ケプストラムの平均を計算する有音区間は、トランスクリプションファイルに示される有音部分とした。CMS は、計算した平均値を無音区間も含む全発声から減算し行った。また、対数パワー項にも同様に、対数パワーの平均を計算し減算をすることにより正規化を行った。

認識実験を行った結果を表 3に示す。この表より、CMS を行うことによる正規化は、メモリ (ワークエリア) 不足のため、認識不可能であったエラー発話を削減し、認識性能が向上することが分かった。ケプストラムの平均の計算方法による認識性能の顕著な差はみられない。そのため、実環境で使用した場合に計算が容易な発話毎にケプストラムの平均を計算し減算を行う CMS_発話毎を以下の実験で用いることにする。

表 3: CMS による正規化実験の結果 (err: メモリ (ワークエリア) 不足のため、認識不可能であった発話数)

正規化手法	音素認識率 (%) male/female/both	単語認識率 (%) male/female/both(err)
正規化なし	70.83/76.12/74.20	66.02/75.59/72.04(1)
CMS_話者毎	73.70/79.36/77.31	68.96/76.48/73.71(0)
CMS_発話毎	73.50/79.17/77.12	**68.79/77.49/74.28(0)**

(4.2.2)　VTLN による周波数ワーピング

周波数ワーピングによる声道長正規化は、話者毎に音声スペクトルを周波数軸方向に伸縮することにより、話者の正規化を行う手法である [3]。音声スペクトルの伸縮には、TR-IT-0257[10] の 2 節に示される周波数ワーピング関数 2 (一次関数による周波数ワーピング) を用いた。また、周波数ワーピングを行うためには、音声スペクトルを伸縮するためのワーピング係数を決めなければならない。このワーピング係数を話者毎に最適に決定することは、この正規化手法の重要な問題の一つとなる。本報告では、ワーピング係数の決定には、TR-IT-0257 で有効であった、各話者の母音の第二フォルマント周波数から、ワーピング係数を求め、周波数ワーピングを行った。周波数ワーピング係数を求めるための各話者の第二フォルマント計算には、xwaves/ESPS の formant コマンドを使用した。このコマンドにより、標本化周波数 16kHz の音声データを 10kHz にダウンサンプリングを行い、ダウンサンプリング後の帯域に男性 4 個、女性 3 個のフォルマントが存在すると仮定をし、各フォルマント周波数を計算した。

本報告では、計算を行った各フレームの第二フォルマント周波数から、母音 (a, i, u, e, o) のみのフレームを用い、以下の 2 種類の計算法

1. 母音部の全フレームをから平均値を計算し、平均フォルマント周波数とする (全母音の平均)
2. 各母音毎のフレームから平均フォルマント周波数を計算し、それらの平均値の平均を平均フォルマント周波数とする (各母音の平均)

で各話者の平均第二フォルマント周波数を決めた。この計算を行った話者毎の平均第二フォルマント周波数と学習話者全体の平均第二フォルマント周波数の比を各話者の周波数ワーピング係数とし、周波数ワーピングを行い特徴パラメータを計算した。この計算で求めたテストセットの周波数ワーピング係数を付録 C に示す。

VTLN (Vocal Tract Length Normalization) による話者正規化手法を用いた認識実験結果を表 4に示す。この表より、母音部の全フレームから平均を求め、周波数ワーピング係数を決定した全母音の平均は、正規化を行う効果がみられないが、各母音毎の平均を平均した各母音の平均は、正規化の効果がみられる。これは、本稿では、自由発話データベースを用いているため、各話者毎に発声された母音の種類、個数が異なる。そのため、全母音フレームから平均を求め、ワーピング係数を計算する方法では、ある母音に偏った周波数ワーピングが行われる可能性があり、正規化の効果が少なかったと考えられる。それに対し、各母音の平均からさらに平均をもとめ、ワーピング係数を計算し、正規化を行った場合には、母音の偏りがなくなり有効に正規化が行えたため認識性能が向上したと考えられる。よって、以下の実験では、各母音の平均から周波数ワーピング係数を求め、正規化を行う方法 (各母音の平均) を用いる。

表 4: VTLN による話者正規化手法を用いた実験結果 (err: メモリ (ワークエリア) 不足のため、認識不可能であった発話数)

正規化手法	音素認識率 (%) male/female/both	単語認識率 (%) male/female/both(err)
正規化なし	70.83/76.12/74.20	66.02/75.59/72.04(1)
全母音の平均	73.05/75.64/74.70	66.32/75.21/71.94(0)
各母音の平均	73.57/76.10/75.18	**67.01/76.97/73.28(1)**

4.3　正規化手法の併用

CMS は、マイク特性などの回線特性等の正規化を行い、周波数ワーピングは、話者によるスペクトルの広がりを抑える話者正規化を行う。つまり、この 2 種類の正規化手法は異なる方向への正規化であるため、併用することによる有効性があると考えられる。本節では、(4.2.1)節、(4.2.2)節で検討を行った 2 種類の正規化手法を併用した場合の影響を調べた。この実験に用いた正規化手法は、先に有効であった、各発話毎にケプストラムの平均を計算し減算を行う CMS_発話毎、各母音の平均第二フォルマント周波数の平均より計算したワーピング係数を使用し正規化を行う各母音の平均を用いた。正規化手法併用による実験の結果を表 5に示す。表より、2 種類の正規化を併用した場合は正規化する方向が異なるために、各々単独で使用した場合よりも効果があること分かる。

表 5: 正規化手法併用による効果

正規化手法	音素認識率 (%) male/female/both	単語認識率 (%) male/female/both(err)
正規化なし	70.83/76.12/74.20	66.02/75.59/72.04(1)
CMS_発話毎	73.50/79.17/77.12	68.79/77.49/74.28(0)
VTLN_各母音	73.57/76.10/75.18	67.01/76.97/73.28(1)
併用	76.19/79.07/78.03	**71.25/78.82/76.03(0)**

4.4 正規化特徴量を用いたトポロジー学習

前節までの正規化実験では、正規化を行った特徴量を音響モデル作成の全学習 (トポロジー学習、音響モデルのパラメータ学習) に用いていた。しかし、VTLN に基づく話者正規化手法は、発声された音声の母音区間の検出、その区間の平均フォルマント周波数の導出、さらに、そのフォルマント周波数を用い周波数ワーピング係数の計算を行わなければならないため、実環境での認識への利用は、使用が困難であると思われる。そのため、正規化を行った特徴量を音響モデルのトポロジー学習のみに用い、音響モデルのパラメータ学習には、正規化を行っていない特徴量を用い、音響モデルを作成した。これより、各正規化手法が音響モデルのトポロジー学習へ与える影響を調べた。このとき、話者正規化により話者間への状態の分割が低減し、コンテキスト方向への分割が増加することによる認識性能向上が期待できる。

トポロジー学習のみに正規化手法を用いた結果を表 6に示す。表より、正規化を行った特徴量によるトポロジー学習が認識に与える影響は比較的少なかった。以下の実験では、単語認識率が最も良い CMS_発話毎を共通のトポロジーとして使用する。

表 6: トポロジー学習のみに正規化手法を用いた結果

正規化手法	音素認識率 (%) male/female/both	単語認識率 (%) male/female/both(err)
正規化なし	70.83/76.12/74.20	66.02/75.59/72.04(1)
CMS_話者毎	71.88/76.67/74.94	67.53/75.57/72.59(1)
CMS_発話毎	71.48/76.80/74.87	**67.10/76.18/72.81(1)**
各母音の平均 (VTLN)	71.48/76.48/74.66	66.97/75.24/72.17(1)
正規化手法の併用	72.17/76.51/74.94	67.96/75.62/72.78(1)

4.5 性別依存モデル

本節では、4.4節の検討より、単語認識率が最も高かった CMS_発話毎の特徴量で作成したトポロジーを用い、CMS_発話毎により正規化を行った特徴量と、正規化を行わない特徴量でパラメータ学習を行った男女の性別依存モデルを作成した。これらの性別依存モデルを用い、認識実験を行った音素認識結果を表 7に、単語認識結果を表 8にそれぞれ示す。ここで、表中の male は、男性モデルを用いテストセットの男性のみを認識した結果、female は、同様に女性モデルを用い女性のみを認識した結果、parallel は、男女の性別依存モデルを並列に使用し認識を行った結果である。ここで、parallel では、無音モデルが異なることによる、モデル選択の誤りを無くすために、性別依存モデルの無音モデルを、性別非依存モデルのそれと変更し、それぞれのモデルとした。GI は、男女の全学習データを用い学習を行った性別非依存モデルで、全テストセットを認識した結果である。

表 7、8より、性別依存モデルの使用、並列使用は、性別非依存モデルより認識性能が高いことが確認できる。性別依存モデルを並列に使用した場合、モデルの選択後には、モデルの精度が性別非依存モデルより向上するため、単語仮説数が減少し、認識速度が速くなると考えられる。表 9に認識時の認識速度を示す。この表より、性別依存モデルの並列使用は、認識性能向上とともに、認識速度の高速化に有効であることが分かる。

音声翻訳システム MATRIX[11] では、認識時に最もゆう度の高かった仮説を示す音響モデルより、男女を判別し翻訳後の合成音を出力している。そのため、性別依存モデルを並列に使用した場合、ゆう度が最大となる仮説

を示す音響モデルの性別が大きな問題となる。表10にモデルを並列に使用した場合のゆう度が最大となる仮説を示す音響モデルの性別と、話者の性別との合致の度合を調べた性別識別率を示す。この表より、CMS によって、正規化を行った場合においても、正規化無しのモデルと同等の性別識別率であることがわかる。よって、CMS の話者方向への正規化の効果は少く、MATRIX 用の音響モデルとして、CMS を行った音響モデルを使用しても性別識別に対する悪影響が無いことが確認された。

表 7: 性別依存モデルを使用した音素認識結果 (音素認識率 %)

正規化手法	male	female	parallel (male/female/both)	GI (male/female/both)
正規化なし	73.91	77.58	72.71/77.07/75.49	71.48/76.80/74.87
CMS_発話毎	75.36	79.58	74.31/79.06/77.34	73.50/79.17/77.12

表 8: 性別依存モデルを使用した単語認識結果 (単語認識率 %)

正規化手法	male	female	parallel (male/female/both)	GI (male/female/both)
正規化なし	69.39	77.61	68.96/77.31/74.23	67.10/76.18/72.81
CMS_発話毎	69.22	77.41	69.30/78.27/74.96	68.79/77.49/74.28

表 9: 性別依存モデルを使用した認識速度 (CPU 時間 / 実時間)

正規化手法	GI モデル	GD モデル並列使用	削減率 (%)
正規化なし	3.04	1.98	34.9
CMS_発話毎	2.32	1.63	29.7

表 10: 性別依存モデルを使用した時の性別識別率

正規化手法	モデルの性別識別率
正規化なし	98.2% (男性 97.7%, 女性 98.5%)
CMS_発話毎	98.2% (男性 98.6%, 女性 97.9%)

5 言語モデルの変更

本報告の実験中に新しい言語モデルのリリースが行われた (ResearchJ V4)[7]。そのため、前節までの検討に基づき、ResearchJ V5 として、リリース予定である発話毎にケプストラムの平均を計算し減算を行った特徴量により全学習を行った音響モデル (CMS_発話毎)、その音響モデルのトポロジーを用い正規化を行っていない特徴量で音響モデルのパラメータを学習した音響モデル (正規化無し) に対し、新しくリリースされた言語モデルを用い認識実験を行った。ReseachJ V4 で新しくリリースされた言語モデルは、以前の言語モデルの品詞体系であった SLDB 体系とは異なり、TDMT 品詞体系を用いている。詳しくは、文献 [7] を参考にされたい。また、言語モデルの変更により、認識パラメータの変更が行われた。前節で比較を行った認識パラメータからの変更点は、

- 言語重みの変更
- 音響モデルの状態スキップが可能
- 単語の終端を各フレーム毎に調べず、2 フレーム毎に調べる
- 複数の単語を一つの単語で代表する、単語のマージを行う
- 音響モデルの無音状態が一定時間 (700msec) 以上続いた場合に、発声の自動切り出しを行い、認識対象範囲の決定を行う

である。

この認識パラメータを用い、各音響モデル (正規化無し、CMS_発話毎) の性別非依存モデルと、TDMT 体系の言語モデルに対し、認識実験を行った結果を表 11に示す。表中の Ins は挿入誤り数、Del は削除誤り数、Sub は置換誤り数を表す。また、No/Wo は、ラティスのノード数を認識結果の単語数で除した値、cpu/u は認識に要した時間を発声時間で除した値、err-utt は、ワークエリア不足などにより、認識が正常に行うことが不可能であった発話数である。この表より、正規化無しの音響モデルは、エラー発話が存在するが、認識速度を抑えほぼ同等の認識性能を得ることができる。また、正規化を行った CMS_発話毎の音響モデルでは、V4 に対して計算時間が 23.9% 削減で、6.5% の誤り改善率を得ることができた。単語認識に用いたコンフィグレーションファイルを付録 B.4に示す。

本稿では、V4 とは異なり新しい音響モデルを用いているため、新たに言語重み、ビーム幅の変更が必要であると考えられる。付録 D にて、比較を行った結果、表 12に示す結果がそれぞれの音響モデルと TDMT 言語モデルを用いた場合に最適と考えられる言語重み、ビーム幅である。これらの結果の男女別、話者別に対する認識率、認識速度などに関する詳細は、付録 E を参考にされたい。また、本稿で最適と考えられる言語重み、ビーム幅は、認識時間がほぼ同等であるところの認識性能の比較により決定した。そのため、使用目的に応じて言語重み、ビーム幅の最適値の変更が必要であると考えられる。参考のため、言語重み、ビーム幅に対する比較を行った結果を、付録 D に、マージリスト使用による認識性能の変化を付録 F に示す。

表 11: V4 の認識パラメータにおける比較

	単語認識率 (%)	全単語	Ins	Del	Sub	No/Wo(cpu/u)	err-utt.
正規化無し	78.77	4961	228	240	585	4.92 (2.11)	1
CMS_発話毎	80.30	4990	193	224	566	4.53 (1.85)	0
V4	78.92	4990	200	238	614	5.05 (2.43)	0

表 12: 認識パラメータ

音響モデル		言語重み		ビーム幅	単語認識率 (%)
		1st	2nd		
正規化無し	性別非依存モデル	7	13	100	79.32
	性別依存モデル	7	11	100	80.14
CMS_発話毎	性別非依存モデル	8	11	110	81.52
	性別依存モデル	8	13	110	82.06

6 まとめ

本稿では、高精度の音響モデルの構築を目的としてCMSによるケプストラム正規化、VTLNに基づく話者正規化を行った特徴量で音響モデルを構築し、音素認識率、単語認識率による評価を行った。比較の結果より、発話毎にケプストラムの平均を計算し、減算を行い正規化を行った特徴量で作成したトポロジーを用い、正規化を行わない特徴量、発話毎のCMSを行った特徴量で音響パラメータを学習したモデルが高い性能を示すことが分かった。

よって、発話毎にCMSを行った特徴量から作成したトポロジーを用い、

- 正規化を行わないMFCCを特徴パラメータとする800状態5混合の性別非依存モデル
- 正規化を行わないMFCCを特徴パラメータとする800状態5混合の性別依存モデル
- 発話毎にCMSを行ったMFCCを特徴パラメータとする800状態5混合の性別非依存モデル
- 発話毎にCMSを行ったMFCCを特徴パラメータとする800状態5混合の性別依存モデル

を音声翻訳通信研究所に対する音響モデルとしてリリースする。リリースを行う各音響モデル、コンフィグレーションファイル等のディレクトリは、付録Gを参照にされたい。

参考文献

[1] 深田俊明, 柘植覚, H. Singer, 内藤正樹. 連続音声認識用音響モデル (version 2.0). Technical Report TR-IT-0241, ATR, 1997.

[2] F. H. Liu, R. M. Stern, X. Huang, and A. Acero. Efficient cepstral normalization for robust speech recognition. In *Proc. DARPA Workshop*, pp. 69–74, March 1993.

[3] E. Eide and H. Gish. A parametric approach to vocal tract length normalization. In *Proc. ICASSP*, pp. 346–348, 1996.

[4] H. Singer, M.Tonomura, Q. Huo, J. Ishii, T. Fukada, and M. Schuster. Baseline acoustic models for the spoken language database(SDB/SLDB). Technical Report TR-IT-0206, ATR, 1997.

[5] 大脇浩, Harald Singer, 鷹見淳一, 樽松明. 音素配列構造の制約を用いた音素タイプライタ. 信学技報, Vol. SP93-113, pp. 71–78, December 1993.

[6] 政瀧浩和, 松永昭一, 匂坂芳典. 連続音声認識のための可変長連鎖統計言語モデル. 信学技報, Vol. SP95-73, pp. 1–6, November 1995.

[7] 塚田元, 中嶋秀治, 伴敏雄, 山本博史. 音声翻訳システムのための日本語音声認識言語モデル. Technical Report TR-IT-00265, ATR, 1998.

[8] 嵯峨山茂樹. 音声認識のための音声分析とラベル変換. Technical Report TR-I-0347, ATR, 1993.

[9] 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.

[10] 杉田記男, 内藤正樹. 周波数ワーピングに基づく話者正規化手法の検討. Technical Report TR-IT-0257, ATR, 1997.

[11] B. Reaves, A. Nishino, and T. Takezawa. ATR-MATRIX: Implementation of a speech translation system. 音講論, pp. 53–54, March 1999.

付録 A　音素ラベルファイルによる認識性能の比較

音素ラベルファイルの違いによる音素認識率を表13に、単語認識率を表14にそれぞれ示す。表中の **TR-IT-0206** は、TR-IT-0206 で作成された音素ラベルファイルを用いた認識結果、話者適応無しは、MFCC を特徴パラメータとした 800 状態 5 混合の性別依存モデルで作成した音素ラベルファイルを用いた結果、話者適応ありは、話者適応無しモデルを各話者毎に話者適応を行い作成した音素ラベルファイルを用いた結果である。表中の acc は accuracy, cor は %correct を表し、best phone は 1 位候補の結果を、net phone は正解系列を与えた場合にラティス内で最高性能を与える結果である。total は正解の総音素数、Ins は挿入誤り数、Del は脱落誤り数、Sub は置換誤り数である。また、No/Wo はラティスのノード数を認識結果の音素 (単語) 数で除した値、cpu/u は認識に要した計算時間を発声時間で除した値、error utt. はワークエリア不足、ビーム幅不足などに起因して認識を正常に行うことができなかった発声数である。

表 13: 音素ラベルファイルの比較 (音素認識率)

音響モデル	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u)	utt.
TR-IT-0206	74.20/ 81.14	85.68/ 90.66	21014	1457	1130	2834	3.95(0.35)	0
話者適応あり	75.05/ 82.04	85.87/ 90.77	21014	1467	1178	2597	3.64(0.33)	0
話者適応無し	74.20/ 81.40	85.45/ 90.36	21014	1512	1184	2725	3.78(0.33)	0

表 14: 音素ラベルファイルの比較 (単語認識率)

音響モデル	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u)	utt.
TR-IT-0206	71.73/ 78.23	81.21/ 86.31	6200	403	252	1098	2.46(3.81)	2
話者適応あり	71.87/ 78.39	80.79/ 86.01	6242	407	264	1085	2.43(3.17)	1
話者適応無し	72.04/ 78.56	80.62/ 85.53	6242	407	292	1046	2.44(3.28)	1

付録 B　コンフィギュレーションファイル

正規化を行っていない MFCC の特徴量を作成するためのコンフィグレーションファイルを付録 B.1に、各正規化手法の検討に用いた、音素認識時のコンフィグレーションファイルを付録 B.2に、単語認識時のコンフィグレーションファイルを付録 B.3にそれぞれ示す。また、リリースを行うために認識実験を行ったリリースのための単語認識時のコンフィグレーションファイルを付録 B.4に示す。

B.1　プリプロセシング (正規化無し)

```
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NoHeader
I/Ocontrol:outputParamSize=26
I/Ocontrol:outputParamType=float
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian

ATRwave2cep:Preemphasis=0.98
ATRwave2cep:FrameLength=20
ATRwave2cep:FrameShift=10
ATRwave2cep:SamplingFrequency=16000
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:LpcOrder=16
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:CutoffLowFrequency=0
ATRwave2cep:CutoffHighFrequency=8000
ATRwave2cep:AnalysisType=fft
ATRwave2cep:DebuggingLevel=0
ATRwave2cep:CentroidFreqOrder=0
ATRwave2cep:CentroidFreqGamma=0.5
ATRwave2cep:CentroidFreqType=linear
ATRwave2cep:VTLWarpingFactor=1.0

ATRcep2para:CepstrumOrder=12
ATRcep2para:LDA=
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:DDCepstrumWindow=9
ATRcep2para:DDCepstrumPadding=zero
ATRcep2para:rho=1.0
ATRcep2para:OutputParameter=pow+cep(12)+dpow+dcep(12)

ATRexpand:samplingFrequency=16000
ATRexpand:frameShift=10
ATRexpand:outputParamSize=26
ATRexpand:inputFd
ATRexpand:outputFd
ATRexpand:debuggingLevel=ON
ATRexpand:htkFlag=0
ATRexpand:outputFormat=SSSData
ATRexpand:exec="execute.cmd -config=config.file"
```

B.2 音素認識のコンフィグファイル

```
I/Ocontrol:rpcTable=
I/Ocontrol:rpcNumber=3
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputParamType=
I/Ocontrol:outputParamSize=
I/Ocontrol:outputFormat=Lattice
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputParamType=float
I/Ocontrol:inputParamSize=26
I/Ocontrol:inputFormat=FrameSync

ATRresult:merge_list=
ATRresult:answer=/dept1/work1/ResearchJ/V3/data/SP_ID.TRS
ATRresult:dp_weight=1.0,1.0,1.0
ATRresult:pause_symbol=-
ATRresult:UTT_END=6
ATRresult:UTT_START=5
ATRresult:re_beam=
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:lattice_out=stdout

ATRlattice:lexicon=/dept1/work1/ResearchJ/V3/lmodel/LEX.P
ATRlattice:amname=amodel_name
ATRlattice:active_model=1
ATRlattice:lmscale=4.000000,8.000000
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=Class-2,/dept1/work1/ResearchJ/V3/lmodel/LM.P
ATRlattice:beam=30.000000,30.000000
ATRlattice:work_area=200,50
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=-
ATRlattice:dimension=26
ATRlattice:max_allophone=5000
ATRlattice:phone_boundary=OFF
ATRlattice:word_boundary_skip=OFF
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:amscale=1.000000
```

B.3　単語認識のコンフィグファイル

```
I/Ocontrol:rpcTable=
I/Ocontrol:rpcNumber=3
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputParamType=
I/Ocontrol:outputParamSize=
I/Ocontrol:outputFormat=Lattice
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputParamType=float
I/Ocontrol:inputParamSize=26
I/Ocontrol:inputFormat=FrameSync

ATRresult:merge_list=
ATRresult:answer=/dept1/work1/ResearchJ/V3/data/SP_ID.ANS,/dept1/work1/ResearchJ/V3/data/SP_ID.TRS
ATRresult:dp_weight=1.0,1.0,1.0
ATRresult:pause_symbol=-
ATRresult:UTT_END=6
ATRresult:UTT_START=5
ATRresult:re_beam=
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:lattice_out=stdout

ATRlattice:lexicon=/dept1/work1/ResearchJ/V3/lmodel/LEX.W
ATRlattice:amname=amodel_name
ATRlattice:active_model=1
ATRlattice:lmscale=8.000000,20.000000
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=Class-2,/dept1/work1/ResearchJ/V3/lmodel/LM.W
ATRlattice:beam=85.000000,85.000000
ATRlattice:work_area=1800,200
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=-
ATRlattice:dimension=26
ATRlattice:max_allophone=5000
ATRlattice:phone_boundary=OFF
ATRlattice:word_boundary_skip=OFF
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:amscale=1.000000
```

B.4 リリースのための単語認識コンフィグファイル

このコンフィグレーションファイルでは、認識時のメモリ不足により、正常に認識が行えない発話 (エラー発話) を減少させるために、ATRlattice:work_area=4000,200 としてあるため、確保するメモリ量が非常に大きい。そのため、認識を行う環境によっては、この値の変更が必要となると考えられる。

```
I/Ocontrol:rpcTable=
I/Ocontrol:rpcNumber=3
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputParamType=
I/Ocontrol:outputParamSize=
I/Ocontrol:outputFormat=Lattice
I/Ocontrol:inputEOFexit=ON
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:inputParamType=float
I/Ocontrol:inputParamSize=26
I/Ocontrol:inputFormat=FrameSync
I/Ocontrol:inputFd=stdin

ATRlattice:lexicon=/dept1/work1/ResearchJ/V5/lmodel/LEX.W
ATRlattice:amname=/dept1/work1/ResearchJ/V5/amodel/amodel_name
ATRlattice:active_model=1
ATRlattice:lmscale=7.000000,15.000000
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=Class-2,/dept1/work1/ResearchJ/V5/lmodel/LM.W
ATRlattice:beam=95.000000,95.000000
ATRlattice:work_area=4000,200
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=-
ATRlattice:dimension=26
ATRlattice:state_skip=ON,75000
ATRlattice:phone_boundary=ON
ATRlattice:word_boundary_skip=2
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:amscale=1.000000
ATRlattice:UTT_END_delay=70

ATRresult:minimum_utt=0
ATRresult:dp_unit=FILE
ATRresult:dp_weight=1.0,1.0,1.0
ATRresult:pause_symbol=-
ATRresult:UTT_END=6
ATRresult:UTT_START=5
ATRresult:re_beam=
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:lattice_out=SP_ID.latt
ATRresult:merge_list=/dept1/work1/ResearchJ/V5/result/merge.list
ATRresult:answer=/dept1/work1/ResearchJ/V5/data/SP_ID.ANS,/dept1/work1/ResearchJ/V5/data/SP_ID.TRS
```

付録 C　テストセットのワーピング係数

VTLN に基づく周波数ワーピングに使用する、テストセットのワーピング係数を表 15、表 16に示す。表 15は、母音部の全フレームをから平均値を計算し、周波数ワーピング係数を計算した値、表 16は、各母音毎に平均フォルマント周波数を計算し、それらの平均値から周波数ワーピング係数を計算した値である。

表 15: 全母音の平均から計算を行った認識話者のワーピング係数

男性		女性	
話者名	ワーピング係数	話者名	ワーピング係数
MMAHA	1.0167	FTOAR	0.930073
MMAUC	1.04145	FYOOO	0.930817
MSAHA	1.09534	FMASZ	1.00297
MKEWA	1.12885	FYUKI	0.946824
MNACH	1.13146	FTAAO	0.967203
MMINA	1.09262	FYOAZ	1.08499
MJUHO	1.13144	FKOTS	0.916923
MHITA	1.04753	FTOHO	1.00727
MYUYO	1.00757	FRIYU	0.868625
MMAOX	1.15986	FKAAR	0.979681
METYA	1.16581	FASSA	0.960382
MSHSZ	1.12012	FASHI	0.881715
MMAOK	1.02725	FJUSA	0.920894
MMAMU	1.11407	FYUNZ	0.88701
MKENA	1.0333	FAKMX	0.909282
MKEOO	1.13151	FKYYA	0.936675
MKEWZ	1.14654	FSAWA	0.918187
		FNOTA	0.906106
		FKENA	0.872327
		FAYYA	0.934354
		FKITA	0.907283
		FYOSU	0.862527
		FHAOK	0.974605
		FYUNI	1.02584
		FYUKO	0.977224

表 16: 各母音の平均から計算を行った認識話者のワーピング係数

男性		女性	
話者名	ワーピング係数	話者名	ワーピング係数
MMAHA	1.03819	FTOAR	0.942415
MMAUC	1.05223	FYOOO	0.938743
MSAHA	1.11762	FMASZ	0.973427
MKEWA	1.16042	FYUKI	0.950754
MNACH	1.13565	FTAAO	0.971116
MMINA	1.10007	FYOAZ	1.07688
MJUHO	1.13909	FKOTS	0.927954
MHITA	1.09115	FTOHO	1.00362
MYUYO	1.03975	FRIYU	0.883111
MMAOX	1.10962	FKAAR	0.954465
METYA	1.15255	FASSA	0.938581
MSHSZ	1.11036	FASHI	0.873641
MMAOK	1.02366	FJUSA	0.912687
MMAMU	1.08048	FYUNZ	0.920746
MKENA	1.06788	FAKMX	0.902603
MKEOO	1.10493	FKYYA	0.918764
MKEWZ	1.15875	FSAWA	0.91345
		FNOTA	0.913273
		FKENA	0.887497
		FAYYA	0.897734
		FKITA	0.919367
		FYOSU	0.869543
		FHAOK	0.946316
		FYUNI	1.02243
		FYUKO	0.951597

付録 D　リリースのための認識パラメータの検討

リリースを行う正規化無しの性別依存、非依存モデル、発話毎にケプストラムの平均を計算し、減算を行うCMS_発話毎の性別依存、非依存モデルの各音響モデルとTDMT体系言語モデルを用い、認識を行う場合の認識パラメータ言語重み、ビーム幅に対する比較検討を付録D.1、D.2、D.3、D.4でそれぞれ行った。表中のaccは、単語accuracy、No/Woは、ラティスのノード数を認識結果の単語数で除した値、cpu/uは認識に要した計算時間を発声時間で除した値、Li/Woは、ラティスのリンク数を認識結果の単語数で除した値である。また、errは、ワークエリア不足などで認識が正常に行えなった発話数を示す。

D.1　正規化無し、性別非依存モデル

正規化無しの性別非依存モデルの、1stパスの言語重みとビーム幅による認識性能の比較の結果を表17、2ndパスの言語重みによる認識性能の比較の結果表18をそれぞれ示す。

表17より、この音響モデルを用いた場合、ほとんどの認識結果で同じエラー発話が存在するため、そのエラー発話を含んだ認識結果で最適であると考えられる言語重み、ビーム幅を検討する。表中では、1stパスの言語重み8、ビーム幅110が最も高い認識性能を示しているが、認識時間が必要のため、ほぼ同等の認識時間でもっとも認識性能が高い、1stパスの言語重み7、ビーム幅100がこの音響モデルに適した認識パラメータであると考えられる。

また、表18より、2ndパスの言語重みは、13が適した値であると考えられる。2ndパスの言語重みは、認識時間にほとんど影響を与えないことが分かるが、ノード数、リンク数の増加により、ラティスのサイズが大きくなるため、保存時のディスク容量を注意する必要がある。

表 17: 1stパスの言語重み、ビーム幅の検討 (単語認識率, No/Wo (cpu/u, Li/Wo))

ビーム幅	言語重み (1st, 2nd)					
	6,15	err	7,15	err	8,15	err
85	78.19, 3.78(1.70,5.06)	1	76.93, 3.58(1.43,4.75)	0	72.65, 3.35(1.12,4.35)	0
90	78.71, 4.49(2.19,6.30)	1	77.87, 4.23(1.67,5.87)	1	75.29, 3.94(1.42,5.36)	0
95	78.77, 5.24(2.84,7.64)	1	78.77, 4.92(2.11,7.04)	1	77.32, 4.66(1.66,6.62)	1
100			**79.04, 5.78(2.72,8.60)**	1	77.95, 5.45(2.07,8.06)	1
105					78.75, 6.43(2.64,9.90)	1
110					79.36, 7.54(3.52,12.02)	1

表 18: 2ndパスの言語重みの検討 (1stパスの言語重み = 7, ビーム幅 = 100)

2ndパスの言語重み	acc	No/Wo (cpu/u, Li/Wo)	err
11	79.10	11.25(2.76,20.77)	1
13	**79.32**	**7.81(2.73,12.79)**	1
15	79.04	5.78(2.72,8.60)	1
17	78.53	4.55(2.72,6.26)	1

D.2 正規化無し、性別依存モデル

正規化無し、性別依存モデルの 1st パスの言語重みとビーム幅による認識性能の比較を行った結果を表 19に示す。この表より、1st パスの言語重み 6、ビーム幅 95 と、1st パスの言語重み 7、ビーム幅 100 は、ほぼ同等の認識性能であるため、性別非依存モデルで最適と考えられた 1st パスの言語重 7 を同様にこのモデルの最適 1st パスの言語重みとする。この、1st パスの言語重み、ビーム幅を用い、2nd パスの言語重みの比較を行った結果を表 20に示す。この結果より、エラー発話が存在せず、最も認識性能が高い 11 がこの音響モデルに適した 2nd パスの言語重みであると考えられる。

表 19: 1st パスの言語重み、ビーム幅の検討 (単語認識率, No/Wo (cpu/u, Li/Wo))

ビーム幅	言語重み (1st, 2nd)					
	6,15	err	7,15	err	8,15	err
90	79.02, 3.92(1.80,5.31)	0	78.18, 3.76(1.45,5.05)	0	75.75, 3.53(1.19,4.67)	0
95	79.58, 4.50(2.23,6.34)	0	79.24, 4.35(1.76,6.09)	0	77.62, 4.12(1.43,5.68)	0
100	79.70, 5.24(2.84,7.66)	0	**79.50, 5.08(2.17,7.41)**	0	78.22, 4.87(1.72,7.05)	0
105					79.02, 5.64(2.12,8.47)	0

表 20: 2nd パスの言語重みの検討 (1st パスの言語重み = 7, ビーム幅 = 100)

2nd パスの言語重み	acc	No/Wo (cpu/u, Li/Wo)	err
9	80.28	13.69(2.24,27.12)	1
11	**80.14**	**9.21(2.19,16.09)**	0
13	79.82	6.60(2.18,10.44)	0
15	79.50	5.08(2.17,7.41)	0
17	79.50	4.12(2.17,5.63)	0

D.3　CMS_発話毎による正規化、性別非依存モデル

CMS_発話毎による正規化、性別非依存モデルに対し、1st パスの言語重みとビーム幅による認識性能の比較を行った結果を表 23に示す。この表より、1st パスの言語重み 8、ビーム幅 110 が最も高い単語認識率であることが分かる。よって、1st パスの言語重みを 8、ビーム幅を 110 とし、2nd の言語重みの比較を行った。比較の結果を表 22に示す。表 21、表 22より、1st パス及び 2nd パスの言語重みをそれぞれ 8, 11、ビーム幅 110 が CMS_発話毎による正規化、性別非依存モデルに適した認識パラメータであると考えられる。

表 21: 1st パスの言語重み、ビーム幅の検討 (単語認識率, No/Wo (cpu/u, Li/Wo))

ビーム幅	言語重み (1st, 2nd)					
	6,15	err	7,15	err	8,15	err
85	80.04, 3.56(1.54,4.67)	0	78.74, 3.38(1.24,4.38)	0	74.37, 3.24(0.97,4.17)	0
90	80.42, 4.10(1.92,5.61)	0	79.70, 3.89(1.49,5.25)	0	77.21, 3.75(1.22,5.02)	0
95	80.86, 4.71(2.47,6.66)	0	80.30, 4.53(1.85,6.36)	0	78.90, 4.31(1.48,6.02)	0
100	80.84, 5.48(3.29,8.04)	0	80.56, 5.23(2.37,7.62)	0	79.62, 5.01(1.82,7.25)	0
105			80.88, 6.06(3.08,9.14)	0	80.18, 5.83(2.30,8.72)	0
110					**80.98, 6.77(2.95,10.46)**	0

表 22: 2nd パスの言語重みの検討 (1st パスの言語重み = 8, ビーム幅 = 110)

2nd パスの言語重み	acc	No/Wo (cpu/u, Li/Wo)	err
9	81.02	20.86(3.07,45.56)	0
11	**81.52**	**13.20(2.98,24.84)**	0
13	81.14	9.11(2.92,15.33)	0
15	80.98	6.77(2.95,10.46)	0

D.4 CMS_発話毎による正規化、性別依存モデル

CMS_発話毎による正規化、性別依存モデルに対する、1st パスの言語重み、ビーム幅の比較の結果を表 23、2nd パスの言語重みの比較の結果を表 24にそれぞれ示す。表 23、表 24より、1st パス及び 2nd パスの言語重みをそれぞれ 8, 13、ビーム幅 110 を CMS_発話毎による正規化、性別非依存モデルの認識パラメータとする。

表 23: 1st パスの言語重み、ビーム幅の検討 (単語認識率, No/Wo (cpu/u, Li/Wo))

ビーム幅	言語重み (1st, 2nd)					
	6,15	err	7,15	err	8,15	err
90	81.14, 3.64(1.60,4.84)	0	80.38, 3.55(1.33,4.75)	0	78.18, 3.38(1.11,4.43)	0
95	81.48, 4.09(1.97,5.62)	0	81.04, 3.94(1.58,5.39)	0	79.86, 3.77(1.31,5.08)	0
100			81.62, 4.48(1.90,6.33)	0	80.58, 4.34(1.54,6.11)	0
105			81.46, 5.18(2.34,7.60)	0	81.22, 4.96(1.87,7.23)	0
110					**81.62, 5.72(2.27,8.60)**	0

表 24: 2nd パスの言語重みの検討 (1st パスの言語重み = 8, ビーム幅 = 110)

2nd パスの言語重み	acc	No/Wo (cpu/u, Li/Wo)	err
9	81.36	16.43(2.31,35.17)	0
11	81.72	10.70(2.28,19.76)	0
13	**82.06**	**7.58(2.29,12.52)**	0
15	81.62	5.72(2.27,8.60)	0
17	81.40	4.58(2.25,6.43)	0

付録 E　認識結果の詳細

リリースを行う各音響モデルを用いた、対話毎の認識結果を表 25、26、27、28にそれぞれ示す。

表 25: 正規化無し、性別非依存モデル、言語モデル重み 7,13、ビーム幅 100 の認識結果

conversation ID	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u, Li/Wo)	utt.
TAC70016.A	80.68/ 84.09	92.05/ 93.18	88	3	6	8	7.11(3.74, 10.73)	0
TAC70017.A	85.94/ 93.75	95.31/ 96.88	64	5	2	2	6.34(2.41, 7.80)	0
TAC70021.A	91.26/ 96.12	97.09/ 97.09	103	5	3	1	4.69(2.15, 5.73)	0
TAC70022.A	72.31/ 74.62	90.00/ 90.77	130	3	8	25	8.84(3.31, 14.82)	0
TAC70023.A	89.29/ 97.32	97.32/ 98.21	112	9	0	3	5.04(2.53, 6.90)	0
TAC70103.A	91.89/ 95.95	95.95/ 97.30	74	3	1	2	3.92(2.61, 5.32)	0
TAC70202.A	79.45/ 79.45	92.47/ 93.15	146	0	13	17	12.26(5.25, 23.36)	0
TAC70304.A	70.49/ 88.52	88.52/ 91.80	61	11	0	7	11.62(5.57, 18.81)	0
TCC70109.A	69.88/ 78.31	86.75/ 91.57	83	7	2	16	12.29(3.65, 19.23)	0
TCC70201.A	54.43/ 63.29	77.22/ 81.01	79	7	9	20	15.96(4.61, 30.76)	0
TCC70212.A	63.23/ 67.10	80.00/ 81.94	155	6	13	38	11.23(3.79, 18.77)	0
TCC70307.A	86.05/ 87.60	93.02/ 95.35	129	2	2	14	9.62(3.53, 16.52)	0
TCC71008.A	61.90/ 66.67	85.12/ 85.71	168	8	16	40	11.02(2.81, 19.39)	0
TCS70034.A	71.33/ 82.00	93.33/ 97.33	150	16	8	19	10.26(2.79, 16.63)	0
TCS70055.A	81.37/ 81.37	91.93/ 93.17	161	0	13	17	5.45(2.65, 8.95)	0
TCS70070.A	84.85/ 89.39	93.94/ 93.94	66	3	2	5	11.08(2.69, 17.27)	0
TCS70074.A	97.44/ 98.72	100.00/100.00	78	1	0	1	2.97(1.44, 3.24)	0
TAC70015.A	85.71/ 88.57	98.10/ 99.05	105	3	3	9	7.24(2.51, 11.83)	0
TAC70019.A	86.73/ 89.80	96.94/ 96.94	98	3	2	8	4.95(1.89, 7.41)	0
TAC70101.A	86.55/ 89.08	100.00/100.00	119	3	7	6	3.80(1.37, 4.52)	0
TAC70102.A	92.59/ 97.04	97.04/ 98.52	135	6	0	4	3.76(1.38, 4.95)	0
TAC70201.A	80.95/ 82.54	94.44/ 96.03	126	2	8	14	5.16(1.44, 7.73)	0
TAC70203.A	78.36/ 83.58	88.06/ 89.55	134	7	6	16	6.22(1.89, 9.12)	0
TAC70301.A	83.62/ 88.79	95.69/ 96.55	116	6	7	6	4.53(1.98, 6.74)	0
TAC70303.A	93.69/ 93.69	98.20/ 98.20	111	0	3	4	3.12(1.33, 3.49)	0
TCC70103.A	74.36/ 80.77	89.74/ 91.03	78	5	5	10	4.77(1.75, 6.49)	0
TCC71001.B	83.78/ 85.95	94.59/ 95.68	185	4	8	18	5.03(2.28, 8.30)	0
TCC71007.A	94.52/ 96.58	96.58/ 97.26	146	3	1	4	4.02(1.52, 4.47)	0
TCC71016.A	73.91/ 82.61	92.75/ 96.38	138	12	2	22	6.96(2.10, 10.72)	0
TCC71035.A	74.16/ 80.90	96.63/ 96.63	89	6	3	14	8.27(2.60, 12.94)	0
TCS70004.B	62.61/ 63.48	83.48/ 84.35	230	2	30	54	15.50(6.54, 32.64)	1
TCS70010.A	69.39/ 76.53	81.63/ 85.71	98	7	7	16	7.82(2.47, 12.64)	0
TCS70013.A	90.12/ 95.06	96.30/ 97.53	81	4	1	3	6.52(2.09, 10.23)	0
TCS70020.A	69.52/ 74.29	93.33/ 94.29	105	5	13	14	11.07(4.77, 20.02)	0
TCS70023.A	81.58/ 85.79	96.84/ 98.42	190	8	11	16	6.87(2.00, 11.60)	0
TCS70025.A	77.14/ 80.95	93.33/ 96.19	105	4	6	14	4.93(1.50, 7.31)	0
TCS70028.A	94.12/ 98.82	100.00/100.00	85	4	0	1	4.58(1.56, 6.23)	0
TCS70047.A	70.93/ 81.40	89.53/ 95.35	86	9	8	8	7.38(1.79, 11.42)	0
TCS70059.A	82.28/ 83.54	93.67/ 94.94	79	1	7	6	7.37(2.09, 12.38)	0
TCS70082.A	85.33/ 88.67	96.00/ 97.33	150	5	1	16	7.19(2.39, 12.51)	0
TSC71005.B	86.09/ 88.70	98.26/ 98.26	115	3	4	9	7.93(2.18, 12.88)	0
TSC71013.A	72.86/ 79.52	86.19/ 89.52	210	14	7	36	12.50(4.79, 23.36)	0
male	77.15/ 81.97	90.74/ 92.37	1847	89	98	235	8.94(3.22, 14.64)	0
female	80.60/ 84.65	93.42/ 94.89	3114	126	150	328	7.13(2.45, 11.68)	1
both	79.32/ 83.65	92.42/ 93.95	4961	215	248	563	7.81(2.73, 12.79)	1

表 26: 正規化無し、性別依存モデル、言語モデル重み 7,11、ビーム幅 100 の認識結果

conversation ID	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u, Li/Wo)	utt.
TAC70016.A	82.95/ 85.23	93.18/ 93.18	88	2	4	9	5.15(2.13, 7.75)	0
TAC70017.A	89.06/ 95.31	96.88/ 96.88	64	4	0	3	7.12(1.71, 9.64)	0
TAC70021.A	90.29/ 93.20	97.09/ 97.09	103	3	4	3	7.05(2.21, 9.77)	0
TAC70022.A	74.62/ 79.23	89.23/ 91.54	130	6	5	22	9.22(2.37, 16.17)	0
TAC70023.A	89.29/ 96.43	95.54/ 97.32	112	8	0	4	6.24(2.16, 9.45)	0
TAC70103.A	90.54/ 91.89	97.30/ 97.30	74	1	2	4	5.07(2.33, 6.72)	0
TAC70202.A	81.51/ 81.51	94.52/ 94.52	146	0	9	18	14.05(3.61, 27.44)	0
TAC70304.A	68.85/ 83.61	91.80/ 95.08	61	9	0	10	12.41(4.85, 20.22)	0
TCC70109.A	73.49/ 80.72	91.57/ 93.98	83	6	2	14	16.42(2.91, 27.59)	0
TCC70201.A	56.96/ 62.03	82.28/ 84.81	79	4	7	23	18.99(3.89, 40.70)	0
TCC70212.A	66.45/ 70.32	82.58/ 86.45	155	6	8	38	15.05(2.78, 25.98)	0
TCC70307.A	82.17/ 89.15	96.12/ 97.67	129	9	3	11	11.06(3.53, 19.87)	0
TCC71008.A	66.07/ 71.43	89.29/ 89.88	168	9	11	37	12.20(2.02, 23.44)	0
TCS70034.A	69.33/ 80.00	92.67/ 96.00	150	16	6	24	12.82(2.36, 23.56)	0
TCS70055.A	80.12/ 84.47	95.65/ 96.27	161	7	7	18	7.29(2.12, 13.63)	0
TCS70070.A	83.33/ 89.39	95.45/ 95.45	66	4	2	5	12.92(2.38, 22.49)	0
TCS70074.A	97.44/ 98.72	98.72/ 98.72	78	1	0	1	3.37(1.50, 3.80)	0
TAC70015.A	89.52/ 93.33	99.05/100.00	105	4	2	5	8.20(2.14, 13.17)	0
TAC70019.A	87.76/ 88.78	95.92/ 95.92	98	1	3	8	6.10(1.72, 9.67)	0
TAC70101.A	88.24/ 89.92	99.16/ 99.16	119	2	7	5	3.66(1.24, 4.53)	0
TAC70102.A	95.56/ 98.52	97.78/ 98.52	135	4	0	2	4.07(1.37, 5.49)	0
TAC70201.A	80.16/ 83.33	94.44/ 96.03	126	4	7	14	7.27(1.34, 12.53)	0
TAC70203.A	77.61/ 81.34	87.31/ 89.55	134	5	6	19	9.34(1.81, 15.36)	0
TAC70301.A	87.07/ 91.38	96.55/ 97.41	116	5	5	5	6.25(1.64, 10.57)	0
TAC70303.A	94.59/ 95.50	98.20/ 98.20	111	1	3	2	4.32(1.25, 5.60)	0
TCC70103.A	82.05/ 84.62	93.59/ 94.87	78	2	5	7	5.81(1.53, 8.56)	0
TCC71001.B	80.54/ 82.70	94.05/ 94.05	185	4	7	25	6.71(1.68, 12.34)	0
TCC71007.A	92.47/ 96.58	95.89/ 97.26	146	6	1	4	4.82(1.35, 5.87)	0
TCC71016.A	78.26/ 85.51	93.48/ 96.38	138	10	3	17	8.51(1.72, 14.93)	0
TCC71035.A	79.78/ 86.52	96.63/ 96.63	89	6	2	10	9.31(2.11, 14.43)	0
TCS70004.B	68.34/ 69.50	86.87/ 86.87	259	3	23	56	15.88(4.50, 34.02)	0
TCS70010.A	76.53/ 81.63	84.69/ 88.78	98	5	5	13	10.18(1.93, 17.94)	0
TCS70013.A	87.65/ 95.06	95.06/ 97.53	81	6	1	3	5.41(1.52, 8.07)	0
TCS70020.A	66.67/ 76.19	88.57/ 92.38	105	10	9	16	13.01(3.71, 24.11)	0
TCS70023.A	82.63/ 86.32	97.37/ 98.42	190	7	13	13	8.53(1.57, 15.85)	0
TCS70025.A	76.19/ 81.90	92.38/ 95.24	105	6	7	12	6.10(1.39, 9.91)	0
TCS70028.A	92.94/ 98.82	100.00/100.00	85	5	0	1	4.80(1.53, 6.67)	0
TCS70047.A	70.93/ 79.07	87.21/ 94.19	86	7	8	10	8.48(1.52, 13.24)	0
TCS70059.A	81.01/ 83.54	96.20/ 98.73	79	2	6	7	7.08(1.50, 13.66)	0
TCS70082.A	84.67/ 88.00	95.33/ 96.67	150	5	4	14	8.65(1.88, 15.75)	0
TSC71005.B	80.00/ 86.09	94.78/ 95.65	115	7	2	14	9.58(1.88, 17.66)	0
TSC71013.A	74.29/ 81.90	87.62/ 91.43	210	16	4	34	13.75(3.08, 27.08)	0
male	77.86/ 83.00	92.53/ 93.94	1847	95	70	244	10.61(2.60, 18.62)	0
female	81.48/ 85.71	93.51/ 95.07	3143	133	133	316	8.39(1.96, 14.59)	0
both	80.14/ 84.71	93.15/ 94.65	4990	228	203	560	9.21(2.19, 16.09)	0

表 27: CMS による正規化、性別非依存モデル、言語モデル重み 8,11、ビーム幅 110 の認識結果

conversation ID	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u, Li/Wo)	utt.
TAC70016.A	82.95/ 86.36	97.73/100.00	88	3	5	7	10.27(3.56, 18.66)	0
TAC70017.A	90.63/ 96.88	96.88/ 98.44	64	4	1	1	12.28(3.11, 17.49)	0
TAC70021.A	90.29/ 95.15	100.00/100.00	103	5	3	2	8.19(2.10, 12.13)	0
TAC70022.A	75.38/ 80.00	90.77/ 90.77	130	6	5	21	15.78(3.89, 31.57)	0
TAC70023.A	92.86/ 99.11	99.11/ 99.11	112	7	0	1	9.33(3.69, 13.36)	0
TAC70103.A	95.95/ 97.30	100.00/100.00	74	1	1	1	6.22(2.80, 9.90)	0
TAC70202.A	78.08/ 78.08	97.26/ 97.26	146	0	13	19	24.92(6.94, 56.88)	0
TAC70304.A	75.41/ 86.89	90.16/ 91.80	61	7	0	8	16.93(4.81, 33.74)	0
TCC70109.A	73.49/ 80.72	92.77/ 93.98	83	6	2	14	21.94(4.32, 37.91)	0
TCC70201.A	58.23/ 64.56	81.01/ 87.34	79	5	3	25	25.95(4.83, 58.46)	0
TCC70212.A	67.10/ 71.61	89.68/ 89.68	155	7	8	36	27.23(5.00, 54.80)	0
TCC70307.A	86.82/ 89.15	95.35/ 96.90	129	3	3	11	16.95(3.81, 34.13)	0
TCC71008.A	61.31/ 67.26	90.48/ 90.48	168	10	14	41	20.44(3.21, 41.09)	0
TCS70034.A	74.67/ 82.00	94.00/ 97.33	150	11	4	23	17.45(3.61, 32.80)	0
TCS70055.A	84.47/ 86.34	96.89/ 96.89	161	3	8	14	8.88(2.62, 16.79)	0
TCS70070.A	77.27/ 81.82	95.45/ 95.45	66	3	2	10	25.17(4.14, 52.01)	0
TCS70074.A	97.44/ 98.72	100.00/100.00	78	1	0	1	4.64(1.65, 5.83)	0
TAC70015.A	91.43/ 93.33	100.00/100.00	105	2	3	4	11.50(2.66, 23.23)	0
TAC70019.A	89.80/ 91.84	97.96/ 97.96	98	2	2	6	9.54(2.48, 19.57)	0
TAC70101.A	91.60/ 93.28	100.00/100.00	119	2	6	2	4.82(1.37, 6.12)	0
TAC70102.A	93.33/ 97.78	97.78/ 98.52	135	6	0	3	5.06(1.54, 7.01)	0
TAC70201.A	82.54/ 83.33	95.24/ 96.03	126	1	8	13	7.11(1.66, 11.49)	0
TAC70203.A	78.36/ 83.58	89.55/ 91.79	134	7	7	15	12.17(2.21, 22.59)	0
TAC70301.A	87.93/ 91.38	96.55/ 97.41	116	4	7	3	7.04(1.97, 11.38)	0
TAC70303.A	94.59/ 95.50	98.20/ 98.20	111	1	3	2	5.14(1.34, 6.76)	0
TCC70103.A	82.05/ 84.62	94.87/ 96.15	78	2	5	7	6.87(1.88, 10.66)	0
TCC71001.B	85.41/ 86.49	95.14/ 95.14	185	2	6	19	9.43(2.28, 18.85)	0
TCC71007.A	93.15/ 95.89	97.95/ 98.63	146	4	1	5	7.34(1.63, 9.83)	0
TCC71016.A	76.81/ 83.33	97.83/ 97.83	138	9	4	19	11.94(2.10, 21.68)	0
TCC71035.A	78.65/ 83.15	97.75/ 97.75	89	4	2	13	13.65(2.64, 24.31)	0
TCS70004.B	75.29/ 76.83	91.12/ 91.51	259	4	22	38	16.56(4.31, 33.89)	0
TCS70010.A	75.51/ 81.63	84.69/ 88.78	98	6	6	12	17.66(3.48, 36.00)	0
TCS70013.A	87.65/ 93.83	92.59/ 96.30	81	5	1	4	8.65(2.35, 14.01)	0
TCS70020.A	71.43/ 77.14	96.19/ 97.14	105	6	13	11	19.74(6.06, 39.43)	0
TCS70023.A	83.68/ 86.32	98.42/ 98.95	190	5	12	14	10.56(2.10, 20.08)	0
TCS70025.A	78.10/ 81.90	95.24/ 97.14	105	4	8	11	8.28(1.79, 15.16)	0
TCS70028.A	91.76/ 97.65	100.00/100.00	85	5	0	2	6.29(1.78, 9.53)	0
TCS70047.A	67.44/ 77.91	90.70/ 96.51	86	9	8	11	10.07(1.95, 15.53)	0
TCS70059.A	82.28/ 83.54	94.94/ 97.47	79	1	6	7	10.34(2.05, 18.53)	0
TCS70082.A	86.00/ 86.67	98.67/ 98.67	150	1	2	18	11.77(2.67, 22.43)	0
TSC71005.B	81.74/ 85.22	99.13/ 99.13	115	4	3	14	12.97(2.47, 23.34)	0
TSC71013.A	76.67/ 81.43	91.43/ 93.81	210	10	4	35	22.12(4.87, 46.97)	0
male	78.94/ 83.38	94.42/ 95.34	1847	82	72	235	16.54(3.78, 31.99)	0
female	83.04/ 86.41	95.51/ 96.53	3143	106	139	288	11.24(2.52, 20.59)	0
both	81.52/ 85.29	95.11/ 96.09	4990	188	211	523	13.20(2.98, 24.84)	0

表 28: CMS による正規化、性別依存モデル、言語モデル重み 8,13、ビーム幅 110 の認識結果

conversation ID	best phone	net phone	best phone accuracy				speed	error
	acc/cor	acc/cor	total	Ins	Del	Sub	No/Wo(cpu/u, Li/Wo)	utt.
TAC70016.A	85.23/ 88.64	100.00/100.00	88	3	5	5	4.61(2.68, 6.05)	0
TAC70017.A	95.31/ 98.44	100.00/100.00	64	2	1	0	4.36(2.28, 5.01)	0
TAC70021.A	89.32/ 93.20	97.09/ 98.06	103	4	5	2	5.22(2.16, 7.13)	0
TAC70022.A	73.08/ 78.46	89.23/ 90.77	130	7	6	22	8.58(2.69, 15.08)	0
TAC70023.A	94.64/ 97.32	97.32/ 98.21	112	3	0	3	6.82(2.97, 9.25)	0
TAC70103.A	91.89/ 93.24	95.95/ 98.65	74	1	2	3	3.84(2.57, 4.97)	0
TAC70202.A	80.82/ 80.82	95.89/ 95.89	146	0	10	18	9.58(3.18, 17.83)	0
TAC70304.A	78.69/ 86.89	91.80/ 95.08	61	5	0	8	11.72(4.21, 21.94)	0
TCC70109.A	72.29/ 81.93	86.75/ 89.16	83	8	1	14	12.10(3.17, 17.49)	0
TCC70201.A	56.96/ 68.35	81.01/ 83.54	79	9	5	20	17.94(4.61, 35.66)	0
TCC70212.A	65.16/ 70.97	89.03/ 89.68	155	9	8	37	12.11(3.18, 20.78)	0
TCC70307.A	82.17/ 86.82	97.67/ 99.22	129	6	3	14	9.60(3.51, 16.94)	0
TCC71008.A	70.24/ 75.00	90.48/ 91.67	168	8	15	27	11.10(2.32, 20.47)	0
TCS70034.A	72.00/ 81.33	90.67/ 94.00	150	14	7	21	11.09(2.89, 19.20)	0
TCS70055.A	88.82/ 90.68	96.89/ 97.52	161	3	6	9	6.13(2.36, 11.03)	0
TCS70070.A	83.33/ 86.36	95.45/ 95.45	66	2	2	7	13.94(3.10, 23.73)	0
TCS70074.A	97.44/ 98.72	98.72/ 98.72	78	1	0	1	3.55(1.63, 4.13)	0
TAC70015.A	89.52/ 92.38	100.00/100.00	105	3	1	7	6.01(2.09, 9.59)	0
TAC70019.A	91.84/ 93.88	97.96/ 97.96	98	2	1	5	6.52(1.82, 11.52)	0
TAC70101.A	90.76/ 92.44	99.16/ 99.16	119	2	6	3	3.73(1.36, 4.47)	0
TAC70102.A	94.07/ 97.78	98.52/ 98.52	135	5	0	3	4.10(1.38, 5.67)	0
TAC70201.A	81.75/ 84.13	92.86/ 96.03	126	3	7	13	5.14(1.56, 7.78)	0
TAC70203.A	81.34/ 85.82	92.54/ 93.28	134	6	7	12	6.64(1.93, 9.88)	0
TAC70301.A	87.07/ 92.24	96.55/ 97.41	116	6	6	3	4.84(1.70, 7.50)	0
TAC70303.A	94.59/ 95.50	98.20/ 98.20	111	1	3	2	3.91(1.28, 4.72)	0
TCC70103.A	83.33/ 85.90	93.59/ 97.44	78	2	5	6	3.64(1.56, 4.88)	0
TCC71001.B	87.57/ 89.73	96.22/ 96.76	185	4	4	15	5.61(1.77, 9.79)	0
TCC71007.A	95.89/ 97.95	97.26/ 98.63	146	3	1	2	4.95(1.55, 6.11)	0
TCC71016.A	79.71/ 86.23	94.93/ 97.10	138	9	2	17	7.43(1.81, 12.48)	0
TCC71035.A	83.15/ 87.64	96.63/ 96.63	89	4	1	10	8.12(2.04, 12.32)	0
TCS70004.B	74.52/ 76.06	89.19/ 89.19	259	4	23	39	8.29(2.62, 14.86)	0
TCS70010.A	77.55/ 81.63	86.73/ 88.78	98	4	6	12	8.67(2.20, 14.53)	0
TCS70013.A	90.12/ 93.83	93.83/ 96.30	81	3	1	4	5.36(1.70, 8.25)	0
TCS70020.A	72.38/ 78.10	91.43/ 94.29	105	6	11	12	11.14(4.45, 20.83)	0
TCS70023.A	83.68/ 86.32	98.42/ 98.95	190	5	14	12	6.73(1.67, 11.46)	0
TCS70025.A	80.95/ 83.81	95.24/ 97.14	105	3	8	9	4.88(1.48, 7.32)	0
TCS70028.A	91.76/ 97.65	100.00/100.00	85	5	0	2	4.91(1.55, 7.27)	0
TCS70047.A	70.93/ 79.07	89.53/ 95.35	86	7	9	9	7.59(1.70, 11.03)	0
TCS70059.A	81.01/ 82.28	93.67/ 96.20	79	1	6	8	6.23(1.81, 9.70)	0
TCS70082.A	86.00/ 87.33	97.33/ 98.00	150	2	3	16	6.36(1.90, 10.71)	0
TSC71005.B	79.13/ 82.61	93.91/ 95.65	115	4	3	17	8.72(1.87, 14.21)	0
TSC71013.A	70.00/ 77.14	85.71/ 89.52	210	15	2	46	12.33(3.26, 23.85)	0
male	79.86/ 84.46	93.56/ 94.80	1847	85	76	211	9.07(2.88, 15.30)	0
female	83.36/ 86.83	94.46/ 95.83	3143	109	130	284	6.71(1.95, 10.87)	0
both	82.06/ 85.95	94.13/ 95.45	4990	194	206	495	7.58(2.29, 12.52)	0

付録 F　単語のマージによる認識性能の比較

単語のマージを行うことによる認識性能の比較を行った。認識を行う認識パラメータは、単語のマージの有無以外は同じである。表 29に、単語のマージの効果を示す。表中の MFCC は、正規化を行わない特徴量の音響モデル、CMS は、CMS_発話毎の特徴量の音響モデル、GI は性別非依存モデル、GD は、性別依存モデルを示す。また、acc は、単語 accuracy、No/Wo は、ラティスのノード数を認識結果の単語数で除した値、cpu/u は認識に要した計算時間を発声時間で除した値、Li/Wo は、ラティスのリンク数を認識結果の単語数で除した値である。また、err は、ワークエリア不足などで認識が正常に行えなった発話数を示す。

表 29: マージリストの効果

	単語マージなし				単語マージあり			
	acc	No/Wo	(cpu/u,Li/Wo)	err	acc	No/Wo	(cpu/u,Li/Wo)	err
MFCC, GI	75.33	7.43	(2.74,12.29)	1	79.32	7.81	(2.73,12.79)	1
MFCC, GD	75.69	8.77	(2.27,15.46)	0	80.14	9.21	(2.19,16.09)	0
CMS, GI	77.21	12.57	(3.02,23.86)	0	81.52	13.20	(2.98,24.84)	0
CMS, GD	77.81	7.22	(2.30,12.03)	0	82.06	7.58	(2.29,12.52)	0

付録 G　リリースディレクトリ

以下のディレクトリに各音響モデル、コンフィグレーションファイル、認識に用いた音響パラメータ、比較実験の結果等を置く。

```
/dept1/work1/ResearchJ/V5/
        |- TR/: テクニカルレポート (TR-IT-0266)
        |
        |- amodel/: 音響モデル
        |
        |- config/: 認識に用いたコンフィグレーションファイル
        |
        |- data/: 認識用音響パラメータファイル
        |       |- MFCC/: 正規化を行っていない MFCC
        |       |- CMS_MFCC/: 発話毎に CMS を行った MFCC
        |
        |- lmodel/: 言語モデル
        |
        |- result/: 認識結果
                |- MFCC.GI/: 正規化を行っていない MFCC を性別非依存モデルで認識を行った結果
                |- MFCC.GD/: 正規化を行っていない MFCC を性別依存モデルで認識を行った結果
                |- CMS_MFCC.GI/: 発話毎に CMS を行った MFCC を性別非依存モデルで認識を行った結果
                |- CMS_MFCC.GD/: 発話毎に CMS を行った MFCC を性別非依存モデルで認識を行った結果
                |- script/: 実行を行ったスクリプト
```