社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

未知語とその周辺単語の音声認識誤りを考慮したCRFによる音声認識誤り訂正

中谷　良平† 　岩橋　直人†† 　中野　幹生††† 　滝口　哲也†††† 有木　康雄††††

† 神戸大学システム情報学研究科　〒657–8501 兵庫県神戸市灘区六甲台町 1–1
†† 情報通信研究機構　〒619–0289 京都府相楽郡精華町光台 3–5
††† （株）ホンダ・リサーチ・インスティチュート・ジャパン　〒351–0188 埼玉県和光市本町 8–1
†††† 神戸大学自然科学系先端融合研究環　〒657–8501 兵庫県神戸市灘区六甲台町 1–1
E-mail: †nakatani@me.cs.scitec.kobe-u.ac.jp, ††naoto.iwahashi@nict.go.jp, †††nakano@jp.honda-ri.com,
††††{takigu,ariki}@kobe-u.ac.jp

あらまし　本稿では，未知語モデリングを用いた，Confusion Network 上での音声認識自動誤り訂正手法を提案する．従来の音声認識には，音声認識器が未知語とその周辺単語を誤認識してしまうという問題がある．そこで，未知語認識を可能にし，同時にその周辺単語の認識誤りを軽減するために，hybrid word/syllable recognition を行う．そして，音響特徴や言語特徴など，様々な素性を用いて，CRF による音声認識誤り訂正を行う．この誤り訂正を用いて，未知語の認識誤りだけでなく，未知語周辺の認識誤りも訂正する．

キーワード　confusion network, conditional random fields, 音声認識誤り訂正, 未知語認識

Error Correction Using CRF for Mis-Recognition around OOV Words on Speech Recognition Result

Ryohei NAKATANI†, Naoto IWAHASHI††, Mikio NAKANO†††, Tetsuya TAKIGUCHI††††, and Yasuo ARIKI††††

† Graduate School of System Informatics, Kobe University　1–1 Rokkodai, nada, Kobe 657–8501, Japan
†† National Institute of Information and Communications Technology　3–5 Hikaridai, Sorakugunseikacho, Kyoto 619–0289, Japan
††† Honda Research Institute Japan Co., Ltd.　8–1 Honcho, Wako, Saitama 351–0188, Japan
†††† Organization of Advanced Science and Technology, Kobe University　1–1 Rokkodai, nada, Kobe 657–8501, Japan
E-mail: †nakatani@me.cs.scitec.kobe-u.ac.jp, ††naoto.iwahashi@nict.go.jp, †††nakano@jp.honda-ri.com,
††††{takigu,ariki}@kobe-u.ac.jp

Abstract　This paper presents a fully automatic word-error correction on a confusion network by employing out-of-vocabulary word modeling. In usual speech recognition, there is a problem that speech recognition systems incorrectly recognize OOV words and their neighboring words. In this paper, we add hybrid word/syllable recognition to the speech recognizer in order to make it recognize OOV words and to reduce the recognition error around OOV words. Then, we propose a CRF-based word-error correction method using acoustic and linguistic features. The proposed method can not only recognize OOV words but also correct the words neighboring OOV words.

Key words　confusion network, conditional random fields, word-error correction, out-of-vocabulary word recognition

1. まえがき

現在までに，音声認識技術は目覚しい発展を遂げてきた．アナウンサーが書き言葉で書かれた原稿を読み上げるような場合では，単語正解精度において 95 %程度の性能で認識が可能である [1]．また，学会講演音声のような，話し言葉・自由発話音声であっても，85 %程度の性能が得られるようになってきた．これらの成果から，音声メディアのアーカイブ化において多くの研究がなされている．例えば，World Wide Web 上のポッドキャストを対象に音声メディアのアーカイブ化を行う PodCastle [2] や，MIT の講義映像/音声を対象とした MIT Lecture Browser [3] などがあげられる．これらのシステムでは，Word Error Rate (WER) を低くすることが求められる．言語モデルは音響モデルによって推測された候補に従って，最適な単語列を選択することができるが，現在の音声認識では音声認識誤りを避けることは難しい．

この問題を解決するために，識別的言語モデルを用いて，大語彙連続音声認識によって出力された N-best 候補をリランキングする，音声認識誤り訂正技術が提案されている [4] [5] [6] [7]．これらは間違った単語を含む音声認識結果の単語列を負例，対応する書き起こしデータを正例として識別的に学習することで，N-best 単語列からより誤り特徴の少ない単語列を選び出す．これらの手法は既知語の置換誤り，削除誤り，挿入誤りなどを訂正することは可能だが，辞書に存在しない単語はそもそも N-best リストに出現しないため，未知語の認識誤りを正しく訂正することはできない．また，未知語は未知語そのものを誤認識してしまうだけでなく，助詞のような周辺単語の認識誤りも引き起こすことが知られている [8]．

そこで本稿では，hybrid word/syllable recognition を実装することで，音声認識器に未知語認識機能を追加し，その後に音声認識誤り訂正を行う手法を提案する．これは，既知語は単語で，未知語は音節列で認識することを目的としている．この手法を用いることで，未知語が発音されたときに発生する未知語の置換誤りを軽減することを目的としている．また一方で音声認識には，未知語の周辺単語を誤認識してしまうという問題がある．提案手法では，未知語認識の後に誤り訂正を行うため，未知語周辺単語の認識が誤っていても訂正することができる．音声認識誤り訂正には以前提案していた手法 [9] を使う．これは Confusion Network [10] を競合仮説として扱い，Conditional Random Fields (CRF) [11] を用いて単語ごとに誤り訂正を行う手法である．

以降 2 章では，提案手法の流れについて述べる．3 章では未知語認識手法について，4 章では音声認識誤り訂正手法についてそれぞれ述べる．5 章で評価実験とその結果を示し，6 章でまとめについて述べる．

2. 提案手法の流れ

図 1 は提案手法の流れを示している．点線で囲まれた学習プロセスでは，まず，hybrid word/syllable recognition を行い，認識結果を Confusion Network として出力する．そして対応する書き起こしデータを用いて Confusion Network 内のすべての単語に正誤ラベリングを行い，bigram，trigram，Confusion Network 上の存在確率などを素性として，CRF によって誤り検出モデルを学習する．図 1 下部の誤り訂正プロセスでは，学習プロセスと同様に，音声データに対して hybrid word/syllable recognition を行い Confusion Network を生成する．そして，学習した誤り検出モデルを用いて，Confusion Network 上で単語ごとに正解を探す．

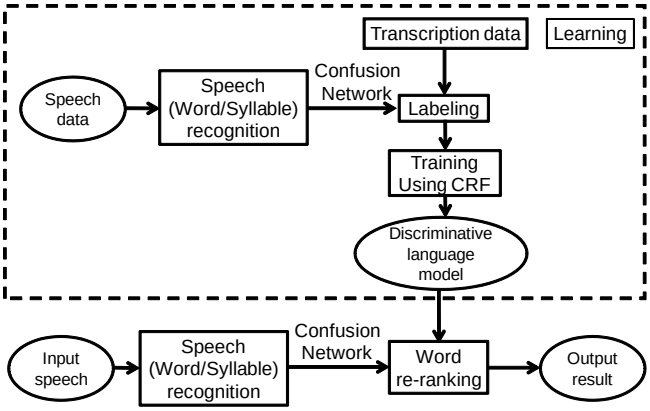

図 1　提案手法の流れ
Fig. 1　Flow of proposed method

3. 未知語認識

本稿では，クラス N-gram によるシンプルな未知語認識手法を用いる．n 個の単語からなる単語列 $\{w_1, \cdots, w_n\}$ が与えられたとき，一般的な単語 N-gram は式 (1) のように定義される．

$$P(w_n|w_1, \cdots, w_{n-1}) = P(w_n|w_{n-N+1}, \cdots, w_{n-1}) \quad (1)$$

一方, クラス N-gram は式 (2) のように定義される．

$$P(w_n|w_1, \cdots, w_{n-1}) = P(c_n|c_{n-N+1}, \cdots, c_{n-1})P(w_n|c_n) \quad (2)$$

クラス $\{c_1, \cdots, c_n\}$ は，単語列 $\{w_1, \cdots, w_n\}$ のそれぞれの単語が属する単語クラスである．クラス N-gram 確率 $P(c_n|c_{n-N+1}, \cdots, c_{n-1})$ はテキストデータから学習される．図 2 はクラス N-gram を用いた未知語認識手法の概要を示している．未知語クラス c_{OOV} に全ての未知語が属すると定義する．c_{OOV} には全ての音節が属していて，音節の連鎖は c_{OOV} の連鎖とすることで擬似的に図 2 の構造を表現する．つまり，未知語は c_{OOV} の連鎖として表現され，c_{OOV} を含むクラス N-gram は未知語が出現しやすい場所を示している．図 2 では c_{OOV} の前後に c_i と c_j を含むクラス N-gram になっているので，c_i は未知語の前に現れやすいクラス，c_j は未知語の後ろに現れやすいクラスとなる．

c_{OOV} には音節しか属していないため，$P(w_n|c_{OOV})$ は c_{OOV} からそれぞれの音節が発生する確率となる．w_n として全ての音節を登録し，$P(w_n|c_{OOV})$ はどの音節に対しても等確率とする．この $P(w_n|c_{OOV})$ をパラメータとして変化させながら実験を

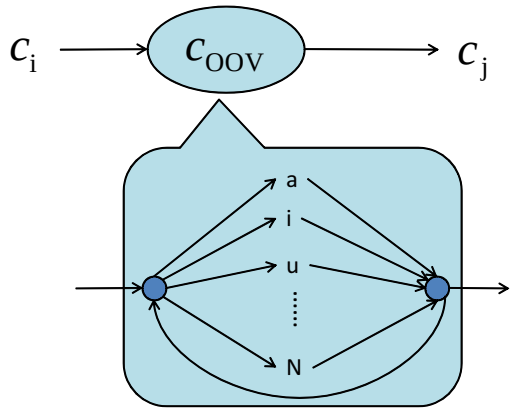

図 2　クラス N-gram による未知語モデリング
Fig. 2　OOV word modeling by class N-gram

表 1　Hybrid word/syllable recognition の例
Table 1　An example of hybrid word/syllable recognition

Input speech: “私たちは Perl スクリプトを使う”
The result of hybrid word/syllable recognition: “私 たち は Syl_pa: Syl_ru スクリプト を 使う”

行う．c_{OOV} を除いたすべてのクラスは 1 クラス 1 単語しか属さないように定義し，既知単語と未知語クラスの N-gram をテキストから学習することで，hybrid word/syllable recognition を実現している．

表 1 は “Perl” を未知語としたときの hybrid word/syllable recognition の例である．“私たちは Perl スクリプトを使う” という発話に対して hybrid word/syllable recognition を行うと，“私 たち は Syl_pa: Syl_ru スクリプト を 使う” という出力が得られた．ここで，“Syl_pa:” と “Syl_ru” はそれぞれ “pa:”, “ru” という発音を表す音節であり，コロン “:” は長音記号を表している．このように，既知語部分を単語列で，未知語部分を音節列で出力するのが hybrid word/syllable recognition の目的である．

4.　音声認識誤り訂正

4.1　Conditional Random Fields

Conditional Random Field (CRF) [11] は，主に自然言語処理やバイオインフォマティクスの分野で用いられているグラフ構造を持つ識別モデルである．文などの構造を持つデータ系列を扱い，モデル式は観測データ系列が与えられたときの出力ラベル系列の条件付確率分布という形をとる．ラベルが与えられた学習データ系列によってモデルを学習し，テストデータ系列を入力すると，モデルが推定するラベル系列が出力される．このとき，データ系列内の各データ一つ一つに最適と推定するラベルを割り当てるのではなく，系列全体として最適と推定するラベルを各データに割り当てる．これは，モデル学習時にデータ間の関係も学習し，ラベル推定時にデータ間の関係を考慮した上で，各データのラベルを推定することで実現する．

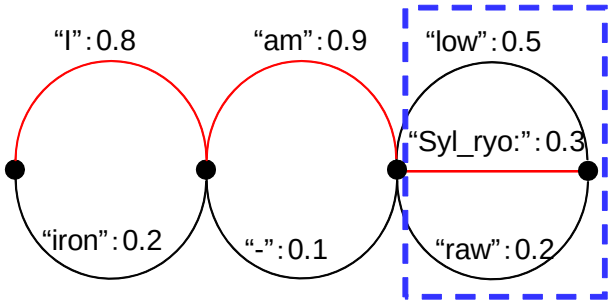

図 3　Confusion Network の例
Fig. 3　Example of Confusion Network

本稿では誤り検出モデルを，認識結果に付与された複数の情報から，各単語に対して正解か誤りかのラベルを付与していく系列ラベリング問題と考え，CRF でモデル化する．CRF を用いた誤り検出モデルは，音声認識結果とそれに対応する書き起こしデータを用いて学習され，入力文書中の不自然な単語を検出することができる．

CRF では，入力記号列 x に対する出力ラベル列 y の条件付確率分布 $P(y \mid x)$ を次式のように定義する．

$$P(y \mid x) = \frac{1}{Z(x)} \exp\Big(\sum_a \lambda_a f_a(y, x)\Big) \tag{3}$$

ここで f_a は素性，λ_a は素性関数に対する重みとなる．$Z(x)$ は分配関数で，次式で与えられる．

$$Z(x) = \sum_y \exp\Big(\sum_a \lambda_a f_a(y, x)\Big) \tag{4}$$

パラメータ λ_a は，学習データ $(x_i, y_i)(1 \leqq i \leqq N)$ が与えられたとき，条件付確率分布 (3) の対数尤度，

$$\mathcal{L} = \sum_{i=1}^{N} \log P(y_i \mid x_i) \tag{5}$$

を最大にするように学習される．これは，正解ラベル列のコストと他のすべてのラベル列のコストとの差が最大になるように学習することに相当する．学習は，準ニュートン法である L-BFGS 法 [12] によって行われる．

識別は学習によって得られた確率分布関数 $P(y \mid x)$ を用いて，与えられた入力記号列 x に対する最適な出力ラベル列 $\hat{y}$ を求める問題となる．$\hat{y}$ は次式をもとに Viterbi アルゴリズムにより効率的に求めることができる．

$$\hat{y} = \operatorname*{argmax}_y P(y \mid x) \tag{6}$$

4.2　Confusion Network

提案しているシステムでは，CRF によって音声認識誤りを検出し，他の競合仮説と置き換えることで誤り訂正を行う．本稿では，競合仮説の表現として Confusion Network を用いる [10]．

Confusion Network は，音声認識器の内部状態を簡潔かつ高精度なネットワーク構造へ変換したもので，単語誤り最小化に基づいた音声認識における中間結果である．図 3 は “I am

Ryo” という発話を認識した際の Confusion Network の例である．点線で囲まれた部分は信頼度が付与された競合単語候補として表現されていて，Confusion Set と呼ばれる．図 3 中には 3 つの Confusion Set が描かれている．信頼度の最も高い候補を選択していくと最尤候補となり，図の例では “I am low” となる．“-” で表された遷移はヌル遷移と呼ばれ，候補単語が存在しないことを意味している．

例えば，図 3 の 3 番目の Confusion Set には，“low”，“Syl_ryo:”，“raw” の 3 つの競合仮説が存在する．ここで，“Syl_ryo:” は “ryo:” という音節を意味する．最も尤度の高い単語列は “I am low” となるが，CRF によって “low” という単語を誤りだと識別することが出来れば，第 2 候補である “Syl_ryo:” と置き換えられる．

4.3 誤り訂正アルゴリズム

前節で述べたように，本稿では CRF を用いて誤り訂正を行う．普通，CRF による誤り傾向の学習には音声認識結果の 1-best 単語列を用いるが，本稿で用いる Confusion Network には特有のヌル遷移が多数存在するため，Confusion Network の第一候補単語列（最尤候補），第二候補単語列，第三候補単語列に正誤ラベリングしたものを，CRF によって学習する．ここで，第三候補がない Confusion Set については，第二候補で補い，第二候補がない Confusion Set については，第一候補で補っている．また，学習に用いる素性は，次章で述べる．誤り検出モデルの学習後，以下のアルゴリズムに従って誤り訂正を行う．

（1） 評価データを音声認識後，Confusion Network を出力する．

（2） Confusion Network の第一候補列のみを抜き出し，CRF による誤り検出を行って，正誤ラベルを付与する．

（3） 入力時系列順に Confusion Set を見ていく．正解と判定された語には何も操作を行わずに次の Confusion Set へ進む．誤りと判定された語は，対応する Confusion Set から次の候補を選び出し，置き換えてもう一度 CRF による誤り検出を行う．

（4） Confusion Set の中に正解と思われる語が存在しなければ，存在確率の最も高い語を選択する．

（5） すべての Confusion Set について順番に (3),(4) を繰り返す．

このアルゴリズムの結果，CRF により誤りと判定された語が，正解と判定された語で訂正される．

また,「入力時系列順に」と述べたのは，CRF によって学習する際の素性として bigram，trigram を用いていることから，前の単語が訂正されると，後ろの単語の正誤判定が変わることがあるためである．例えば，2 単語連続で誤りラベルが付けられている単語列について，1 つ目の単語が訂正されると，bigram 特徴から，2 つ目の単語も正解ラベルに変わることがある．

5. 評 価 実 験

5.1 実 験 条 件

本研究ではベースとなる音声認識システムに，大語彙連続音声認識エンジン Julius-4.1.4 [13] を用いる．

表 2 音響分析条件と HMM の特徴
Table 2 Speech analysis conditions and specifications of HMM

Sampling frequency	16 kHz
Acoustic feature	MFCC (12 dim.) +Δ MFCC (12 dim.)
	+Δ power (total 25 dim.)
Window type	Hamming window
Frame length	25 ms
Frame shift length	10 ms
Acoustic model	Triphone (3,000 states)
The number of mixtures	16
State	5 states and 3 loops

表 3 N-gram エントリー
Table 3 The number of N-gram entries

Unigram	Bigram	Trigram
25,300	731,728	2,611,952

表 4 デ ー タ 数
Table 4 The number of data

	Training	Test
Number of lectures	106	4
Number of speeches	50,780	1,618
Number of words	513,281	17,594

表 5 誤り傾向の学習に用いる素性
Table 5 Features used for error tendency learning

Unigram	w_0
Bigram	w_{-1}/w_0, w_0/w_1
Trigram	$w_{-2}/w_{-1}/w_0$, $w_{-1}/w_0/w_1$,
	$w_0/w_1/w_2$
Confidence of Confusion Network	$CN\ score$

音響モデルは，CSJ の学会講演のうち，953 講演（男性 787 講演+女性 166 講演），計 228 時間分の講演音声から作成した HMM を用いた．音響分析条件と HMM の仕様は表 2 のようになっている．1 状態あたりの混合分布数は 16 としている．サンプリング周波数は 16kHz，音響特徴量は 12 次元 MFCC と対数パワー，12 次元 MFCC の一次微分を加えた 25 次元である．言語モデルは，CSJ の書き起こし文書のうち，2,596 講演の書き起こし文書から学習した N-gram を用いた．N-gram エントリーは表 3 のようになっている．

また，学習と評価に用いたデータ数を表 4 に示す．学習には 106 講演分，評価には 4 講演分の音声データをそれぞれ用いた．コーパスは CSJ を用いている．学習には，Julius が出力した Confusion Network を用いた．

次に，誤り傾向を学習するための素性を表 5 に示す．表層単語 N-gram，Confusion Netwrok 上の存在確率を素性として学習した．識別対象の単語を w_0，そこから n 個前の単語を w_{-n}，

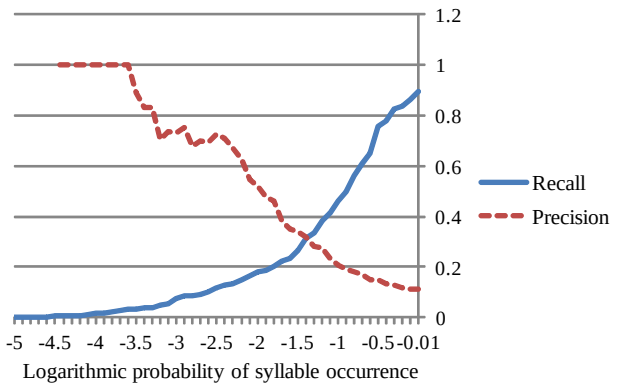

図 4　Recall-Precision 曲線

Fig. 4　Recall-Precision curve

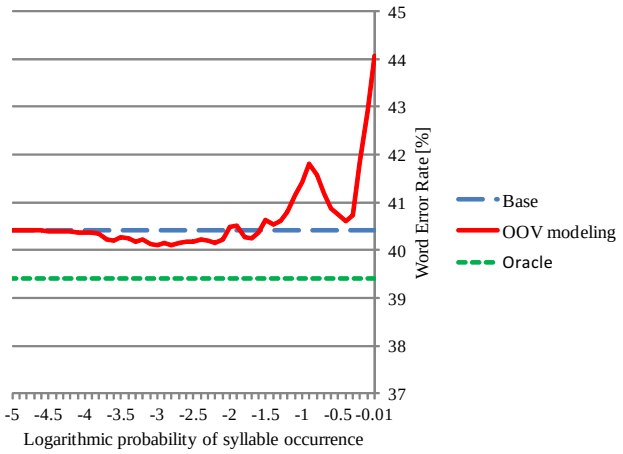

図 5　単語誤り率

Fig. 5　Ward Error Rate

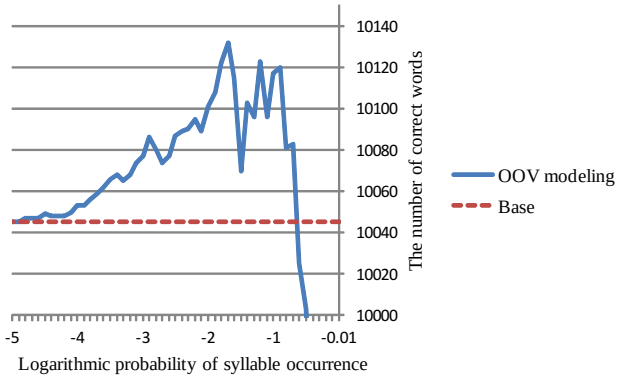

図 6　正解単語数

Fig. 6　The number of correct words

n 個後ろの単語を w_n と表記している．

実験は，最適なパラメータ値を調査し，hybrid word/syllable recognition の精度をより良くするために，3 章で述べた音節の発生確率を制御するパラメータ $P(w_n|c_{OOV})$ を，-0.01 から -5.0 まで変化させながら行う．また，未知語は単語辞書に含まれる名詞からランダムに選ばれており，選ばれた単語を辞書から削除した．テストデータに含まれる未知語の割合は 2 %である．

5.2　実験結果

まず始めに，図 4 に未知語検出の Recall-Precision 曲線を示す．横軸の値は 3 章で述べた未知語音節の発生確率である．このパラメータは未知語の発生しやすさを制御している．つまり，この値が大きいほど認識器が未知語を出力しやすくなる．Recall と Precision は次のように定義する．

$$Recall = \frac{The\ number\ of\ truly\ recognized\ OOV\ words}{The\ number\ of\ true\ OOV\ words\ in\ the\ reference\ file} \quad (7)$$

$$Precision = \frac{The\ number\ of\ truly\ recognized\ OOV\ words}{The\ number\ of\ words\ recognized\ as\ OOV\ words} \quad (8)$$

Recall 曲線と Precision 曲線は，値の変動が急激になってしまっているが，パラメータの値が増加するとともに Recall が増大し Precision が減少しているため，意図した通りの挙動となっている．用いた hybrid word/syllable recognition が単純すぎたため，あまり高いパフォーマンスを得ることができていない。

図 5 は，音節の発生確率を制御するパラメータを変化させた場合の音声認識結果を示している．評価指標としては，単語誤り率（Word Error Rate：WER）を用いている．未知語については，音節系列が異なっていても，場所さえ合っていれば正解とした．“Base” は言語モデルとして単語 N-gram を用いて得られた従来通りの音声認識結果で，2 %の未知語を含んでいる．“OOV modeling” は同じく 2 %の未知語を含んでいて，3 章で述べた hybrid word/syllable recognition を用いて得られた認識結果である．また，“Oracle” は音声データに現れるすべての単語を既知語にした場合の音声認識結果である．横軸のパラメータは未知語の発生しやすさを制御するため，その値が小さくなればなるほど未知語を検出しにくくなり，“OOV modeling” は徐々に “Base” に近づいていく．パラメータが -4.9 以下のとき，hybrid word/syllable recognition によって未知語が出力されなくなり，“Base” と “OOV modeling” は一致する．また，パラメータが -0.01 に近づくほど未知語が過剰に発生するようになり，未知語の挿入誤りが増えるために WER は高くなっていく．パラメータの値が -1.7 付近のときと -2.1 から -4.9 の間のとき，“OOV modeling” の WER が “Base” を下回り，良くなっている．また，図 6 はパラメータを変化させた場合の正解単語数を示している．“OOV modeling” の正解単語数はほぼ “Base” を上回っていて，最も正解数が多いのは値が -1.7 のときとなっている．

表 6 は，単語誤り率と誤りタイプごとの誤り数となっている．それぞれ，“SUB” は置換誤り，“DEL” は削除誤り，“INS” は挿入誤り，“COR” は正解単語の数である．図 6 で，パラメータの値が -1.7 のとき正解単語数が最も多かったので，この点を表 6 の “OOV modeling” として固定した．“Base corrected” は 4 章で述べた誤り訂正手法に基づいて “Base” の誤り訂正を行った結果である．同様に，“Proposed method” は “OOV modeling” の誤りを訂正した場合の結果となっている．提案手法の置換誤りと挿入誤りの数は最も小さくなっていて，結果と

表 6　誤りタイプごとの評価
Table 6　Evaluation with each error type

	SUB	DEL	INS	COR	WER
Base	4,213	813	2,086	10,045	40.42
OOV modeling	4,174	796	2,112	10,132	40.25
Base corrected	4,114	927	1,797	10,030	38.86
Proposed method	4,069	905	1,776	10,128	38.37

して，単語誤り率も最も小さくなっている．“Base” と比較すると，40.42 %から 38.37 %まで低下し，トータルで 2.05 ポイント改善した．“Base corrected” と比較しても 0.49 ポイントの改善となった．“Base corrected” と “Proposed method” を比較すると，どの誤りも改善しているが，置換誤りにおいて特に効果が大きいのがわかる．これは，本来未知語を既知語として認識誤りしていた部分を正しく認識できただけでなく，“OOV modeling” の時点で誤って未知語として認識してしまった部分を既知語に訂正することもできたためである．しかし，この実験において，Confusion Set には未知語ではなくある 1 音節が出現するため，1 音節 1 単語の Confusion Network に対して誤り訂正が行われてしまっている．連続する音節列を 1 単語として扱い Confusion Network を作成することで，誤り訂正後の WER がもっと改善されるのではないかと考えられる．

6. ま と め

本稿では，未知語の認識誤りと未知語周辺単語の認識誤りも考慮した音声認識誤り訂正手法を提案した．誤り訂正の前に，hybrid word/syllable recognition を行い Confusion Network として出力することで，提案手法を実装した．単語誤り率は 40.42 %から 38.37 %まで改善した．これは，ベースラインを誤り訂正した場合と比較しても 0.49 ポイントの改善である．実験結果は提案手法の有効性を示した．

今回用いた未知語認識手法はシンプルで実装が簡単だが，効果はあまり大きくなかったため，今後の課題として，Rastrow らが提案している未知語検出手法 [14] など，もっと効果的な未知語認識手法を導入することが挙げられる．また，CRF で学習する際の素性として，文脈を考慮した N-gram よりも広範囲な長距離言語情報などを用いることも有効であると考えられる．そして，CRF を改善した手法である Conditional Neural Fields [15] を利用することも考えたい．

文　献

[1] 中川聖一, “音声ディクテーションから音声ドキュメント処理へ”, 日本音響学会講演論文集（秋）, pp. 1–4, 2007.

[2] M. Goto, J. Ogata, K. Eto, “Podcastle: A Web2.0 Approach to Speech Recognition Research,” in *Proc. Interspeech2007*, pp. 2397–2400, 2007.

[3] J. Glass, T.J. Hazen, S. Cypher, I. Malioutov, D. Huynh, and R. Barzilay, “Recent progress in the mit spoken lecture processing project,” in *Proc. Interspeech2007*, pp. 2553–2556, 2007.

[4] B. Roark, M. Saraclar, M. Collins, and M. Johnson, “Discriminative language modeling with conditional random fields and the perceptron algorithm,” in *Proc. ACL*, pp. 47–54, 2004.

[5] T. Oba, T. Hori, and A. Nakamura, “A study of efficient discriminative word sequences for reranking of recognition results based on n-gram counts,” in *Proc. Interspeech2007*, pp. 1753–1756, 2007.

[6] Z. Zhou, J. Gao, F. K. Soong, and H. Meng, “A comparative study of discriminative methods for reranking lvcsr n-best hypotheses in domain adaptation and generalization,” in *Proc. ISCA*, pp. 1574–1577, 2006.

[7] A. Kobayashi, T. Oku, S. Homma, S. Sato, T. Imai, and T. Takagi, “Discriminative rescoring based on minimization of word errors for transcribing broadcast news,” in *Proc. ISCA*, pp. 1574–1577, 2008.

[8] C. Parada, M. Dredze, D. Filimonov, F. Jelinek, “Contextual Information Improves OOV Detection in Speech,” in *Proc. HLT-NAACL*, pp. 216–224, 2010.

[9] 中谷良平, 滝口哲也, 有木康雄, “文脈特徴を用いた CRF による音声認識誤り訂正”, 日本音響学会講演論文集（秋）, pp. 189–190, 2011.

[10] L. Mangu, E. Brillx, and A. Stolcke, “Finding consensus in speech recognition: word error minimization and other applications of confusion networks,” in *Computer Speech and Language*, pp. 373–400, 2000.

[11] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proc. ICML*, pp. 282–289, 2001.

[12] J. Nocedal, “Updating quasi-newton matrices with limited storage,” in *Mathematics of Computation*, pp. 773–782, 1980.

[13] “Julius,” http://julius.sourceforge.jp/.

[14] A. Rastrow, A. Sethy, and B. Ramabhadran, “A new method for oov detection using hybrid word/fragment system,” in *ICASSP2009*, pp. 3953–3956, 2009.

[15] J. Xu J. Peng, L. Bo, “Conditional neural fields,” in *NIPS2009*, pp. 1419–1427, 2009.