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

2004－SLP－54（42）
2004／12／22

音声理解のための音声認識評価尺度とベイズリスク最小化デコーディング

南條　浩輝†　　河原　達也††

† 龍谷大学 理工学部 情報メディア学科
〒520-2194 大津市瀬田大江町横谷 1-5
†† 京都大学 学術情報メディアセンター
〒606-8501 京都市左京区吉田二本松町
E-mail: †nanjo@rins.ryukoku.ac.jp, ††kawahara@i.kyoto-u.ac.jp

あらまし　ドメインを限定しない自然な話し言葉の音声理解を目的とした音声認識の評価尺度とそれに基づくデコーディング手法を提案する．従来，音声認識の一般的な評価尺度として，全ての単語を一様に扱う「単語誤り率（word error rate: WER）」が用いられてきた．これに対して，情報検索の観点から各単語の重要度を考慮した「重みつきキーワード誤り率（weighted keyword error rate: WKER）」を提案する．講演音声からの重要文抽出のタスクにおいて，重みつきキーワード誤り率が重要文抽出の精度と相関が高いことを示す．その上で，ベイズリスク最小化（Minimum Bayes-Risk: MBR）の枠組みに基づいて，重みつきキーワード誤り率の最小化を行う音声認識を実現する．CSJ の学会講演 17 講演を用いて評価を行い，提案する認識手法が重みつきキーワード誤り率及び重要文抽出精度の改善に効果があることを示す．

キーワード　音声認識，音声理解，重要文抽出，ベイズリスク最小化デコーディング

ASR Evaluation Measure and Minimum Bayes-Risk Decoding for Open-domain Speech Understanding

Hiroaki NANJO† and Tatsuya KAWAHARA††

† Faculty of Science and Technology, Ryukoku University
Seta, Otsu 520-2194 Japan
†† Academic Center for Computing and Media Studies, Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan
E-mail: †nanjo@rins.ryukoku.ac.jp, ††kawahara@i.kyoto-u.ac.jp

Abstract　A new evaluation measure of speech recognition and a decoding strategy for keyword-based open-domain speech understanding are presented. Conventionally, WER (word error rate) has been widely used as an evaluation measure of speech recognition, which treats all words in a uniform manner. In this paper, we define a weighted keyword error rate (WKER) which gives a weight on errors from a viewpoint of information retrieval. We first demonstrate that this measure is more appropriate for predicting the performance of key sentence indexing of oral presentations. Then, we formulate a decoding method to minimize WKER based on Minimum Bayes-Risk (MBR) framework, and show that the decoding method works reasonably for improving WKER and key sentence indexing.

Key words　speech recognition, speech understanding, key sentence extraction, Minimum Bayes-Risk decoding

1. はじめに

ディクテーションシステムやドメイン限定の音声対話システムに代表される計算機に対する丁寧な発話に対する音声認識は，90%程度の単語認識率が達成されるに至った．現在，音声認識・理解の主な研究対象は，ドメインを限定しない自然な話し言葉に移ってきている [1] [2]．我々は，大規模な『日本語話し言葉コーパス』(CSJ: Corpus of Spontaneous Japanese) [3] [4] を用いて，話し言葉の音声認識の研究を行っており [5]，また，音声の内容理解を目的として，音声の重要箇所を同定する研究（重要文抽出）も行っている [6]．重要文抽出は，音声の内容を端的に示す語（キーワード）に基づいて行われることが多く [7] [8]，我々も [6] において，そのようなアプローチを採用している．このようなキーワードに基づく重要文抽出（音声理解）を目的とした場合，音声認識時に全ての単語を一様に認識するのではなく，キーワードを優先的に認識するのが望ましい．

これまでの話し言葉の音声認識は，発話全体の忠実なテキスト化を目的とするものが主であり，音声認識の精度評価は単語誤り率（WER）に基づいて行われるのが一般的であった．しかし，この単語誤り率による評価では，音声理解に影響の大きいキーワードとそうでない語（間投詞やつなぎ語など）の認識誤りが同等に扱われており，キーワードに基づく音声理解の観点からは音声認識の評価尺度として必ずしも最適ではない．

キーワードの誤り率を音声認識の評価尺度として用いる研究もみられるが，これらは主にタスク限定の対話システムを対象としており，内容理解に必要なキーワードリストを事前に定義できることが前提である．これに対して，本研究では，講演や講義のような “ドメイン非限定” の発話を対象としており，事前にキーワードリストを用意することは困難である．したがって，従来の単純なキーワード認識・評価の枠組をそのまま適用することはできない．

このような背景に基づいて，キーワードに基づく講演音声の理解を目的とした，音声認識の新しい評価尺度と認識手法を提案する．具体的には，情報検索の観点から各単語の認識誤りに異なる重み（tf-idf 値）を与える「重みつきキーワード誤り率（WKER）」を評価尺度として定義し，次に，ベイズリスク最小化（Minimum Bayes-Risk: MBR）デコーディングの枠組に基づいて，重みつきキーワード誤り率の最小化を行う音声認識手法を実現する．この概要を図 1 に示す．

本稿では，CSJ の学会講演 17 講演を用いて評価を行い，キーワードに基づく重要文抽出（音声理解）に対するベイズリスク最小化音声認識手法の有効性を調べる．

2. 講演からの重要文抽出

本研究では，ドメイン非限定の話し言葉の音声理解，具体的には講演音声の音声理解を目的として，音声認識及び重要文抽出を行う．本章では，まず使用する話し言葉コーパスについて述べ，次にベースラインとなる音声認識システムと重要文抽出手法について述べる．

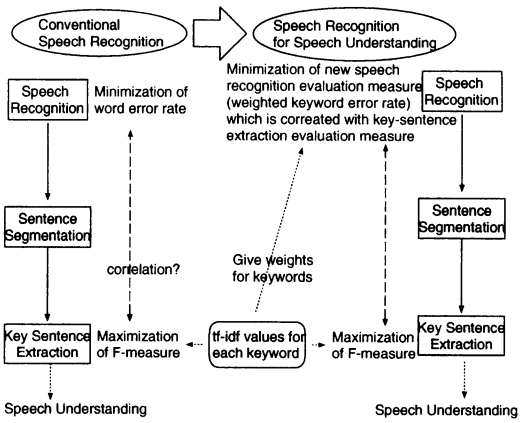

図 1 音声理解を目的とした音声認識

Fig. 1 Speech Recognition for Speech Understanding

2.1 日本語話し言葉コーパス（CSJ）

CSJ は主に学会講演と模擬的な講演からなるコーパスであり，音声データと人手による書き起こしテキストから構成される．CSJ の講演のうち，コアとよばれる一部の講演には複数人による重要文タグが付与されている．講演ごとに 3 人の作業者が割り当てられており，全体の文の 50%と 10%を目安に重要文の抽出が行われている．作業者は研究者であり，学会講演の発表スタイルに精通しているが，必ずしもそれぞれのテストセットの分野における専門家ではない．

2.2 ベースライン音声認識システム

本研究では，以下に示す音響・言語モデルと認識エンジン Julius で構成した音声認識システムを用いる．音響モデルには，CSJ の 781 講演（106 時間）で学習した状態共有型 triphone モデル（3000 状態）を教師なし話者適応したモデル [5] を用いる．言語モデルには，CSJ の 2592 講演（6.7M 形態素）から学習した単語 3-gram モデル（語彙サイズ: 24437）[9] を話者・話題適応したモデル [5] を用いる．

本研究では形態素単位として，国立国語研究所で定義された短単位 [10] を使用し，形態素解析システムは，通信総合研究所で最大エントロピー法により CSJ を用いて統計的に学習されたものを用いる [11]．

2.3 ベースライン重要文抽出システム

2.3.1 キーワードの統計情報に基づく重要文抽出

本研究では，話題と関連のある単語（キーワード）の統計情報を利用する重要文抽出手法を用いる [12]．これは，式 (1) で定義される単語の tf*idf 値を用いて重要文抽出を行うものである．

$$KW_m = tf_m * \log\left(\frac{N_d}{df_m}\right) \quad (1)$$

tf_m は単語 m の当該講演内での出現回数を表わし，df_m は単語 m が出現した講演数を表す．全講演数 N_d を df 値で除したものが idf 値である．各文 j に含まれる単語の tf*idf 値の合計値（式 (2)）をキーワードに基づく重要度とし，この値の大き

いものを重要文として抽出する．

$$S_j^{KW} = \sum_i KW_{m_i} \tag{2}$$

2.3.2 重要文抽出の評価尺度と評価方法

2.1節で述べた通り，CSJの一部の講演の書き起こしには，3名の作業者による10%及び50%重要文抽出が行われている．本研究では，それらのうちの学会講演17講演（表1）をテストセットとして用いる．表2にテストセットにおける人間の重要文抽出の一致度を示す．10%抽出においては，人間同士でも抽出される文セットの一致率が低いことがわかる．したがって本研究では，10%抽出を行わず，50%重要文抽出の評価を行う．

次に，評価を行う際に正解とする重要文の設定と評価法について述べる．重要文として抽出される文は3名の作業者で異なるため，正解とする重要文は，3名から任意の2名を選び，その2名がともに50%の重要文抽出において抽出した文とする．評価は，3通りの正解それぞれに対する再現率（recall），適合率（precision），F値，κ値を算出し，それらの平均で行う．ここで，F値は再現率と適合率の調和平均，κ値は，二つのデータの間で偶然に一致する割合を除いた一致度を示す指標であり，それぞれ式（3），式（4）により定義される．

$$\text{F-measure} = \frac{2 * recall * precision}{recall + precision} \tag{3}$$

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{4}$$

この2名一致による正解重要文に対して，他の1名が抽出したものを照合することで，人間の重要文抽出精度が算出できる．人間の重要文抽出精度は表1に示されてある．講演ごとにばらつきが大きいこと及び人間同士でも重要文として抽出するものにかなり不一致があることがわかる．再現率，適合率，F値，κ値の平均は，それぞれ，81.9%，60.6%，0.697，0.480であった．これらは，重要文抽出システムの評価の際に，目標となる数値である．

2.3.3 キーワードに基づく重要文抽出の結果

書き起こし及び音声認識結果からのキーワードに基づく重要文抽出の結果を表3に示す．また表3には，人間の重要文抽出精度も示してある．表3の(1)と(2)を比較することにより，書き起こしからの重要文抽出では，人間の精度の約85%の精度が達成できていることがわかる．次に，表3の(2)と(3)を比較することにより，音声認識誤りによって，重要文抽出の精度（特に適合率）が低下していることがわかる．なお，重要文抽出の評価は文の区切りが一致しないと行えないため，音声認識結果からの重要文抽出の際には，音声認識結果のある文を抽出した場合に，それが対応する書き起こしにおける文(注1)を抽出したとみなして評価を行っている．

我々は[13]で，音声認識時における文分割が重要文抽出に影響が大きいことを示し，重要文抽出の精度改善を目的とした文分割手法により，重要文抽出の精度改善が行えることを示した．

(注1)：複数の文に対応する場合もあるし，どの文にも対応しない場合もある．

表1 人間による重要文抽出精度（50%抽出）

Table 1 Human performance of key sentence indexing (50% indexing)

Presentation ID	F-measure	κ-value
A01M0007	0.756	0.566
A01M0035	0.725	0.521
A01M0056	0.605	0.321
A01M0074	0.657	0.400
A01M0097	0.622	0.355
A01M0110	0.773	0.583
A01M0137	0.742	0.552
A01M0141	0.653	0.390
A03M0016	0.585	0.303
A03M0106	0.635	0.384
A03M0112	0.821	0.669
A03M0156	0.569	0.291
A04M0051	0.748	0.551
A04M0121	0.584	0.303
A04M0123	0.688	0.467
A05M0011	0.750	0.555
A05M0031	0.758	0.566
Average	0.697	0.480

表2 人間同士の重要文抽出一致率

Table 2 Agreement among subjects in key sentence extraction

	by 2 persons	by 3 persons
50% extraction	71.5%	56.7%
10% extraction	41.6%	25.1%

Estimated with 17 test-set presentations listed in Table 1.

表3 音声認識結果からの重要文抽出の結果

Table 3 Results of key sentence indexing from ASR results

	transcript.	indexing	recall	precision	F-measure
(1)	manual	manual	81.9%	60.6%	0.697
(2)	manual	auto	70.8%	52.5%	0.603
(3)	auto	auto	71.1%	44.2%	0.545

これに対し，本稿では，重要文抽出の精度改善を目的とした新しい音声認識手法について述べる．

3. 重要文抽出のための音声認識の評価尺度

3.1 単語誤り率の一般化

音声認識の評価尺度で最も一般的なものは単語誤り率（WER）であり，これは式（5）で定義される．

$$\text{WER} = \frac{I + D + S}{N} * 100 \tag{5}$$

ここで，Nは正解文における単語の数，Sは置換誤り単語の数，Dは削除誤り単語の数，Iは挿入誤り単語の数である．

式（5）からも明らかなように，この評価尺度（単語誤り率）では，内容理解に影響が大きい重要語（キーワード）とそうでない単語が区別されておらず，全ての単語の音声認識誤りが等しく扱われている．キーワードの認識誤りは，そうでない単語の認識誤りに比べて，内容理解に対する影響が大きいため，内

容理解を目的とした場合は，各単語の音声認識誤りに差異（重み）を与えることが重要である．

このような背景に基づき，ここでは，単語誤り率（WER）の一般化を行う．具体的には，各単語の音声認識誤りに差異（重み）を与える重みつき単語誤り率（Weighted Word Error Rate: WWER）について述べる．重みつき単語誤り率は，以下の式で定義する．

$$\mathrm{WWER} = \frac{V_I + V_D + V_S}{V_N} * 100 \quad (6)$$

$$V_N = \Sigma_{w_i} \quad v_{w_i} \quad (7)$$

$$V_I = \Sigma_{\hat{w}_i \in I} \quad v_{\hat{w}_i} \quad (8)$$

$$V_D = \Sigma_{w_i \in D} \quad v_{w_i} \quad (9)$$

$$V_S = \Sigma_{seg_j \in S} \quad v_{seg_j} \quad (10)$$

$$v_{seg_j} = \max(\Sigma_{\hat{w}_i \in seg_j} v_{\hat{w}_i}, \Sigma_{w_i \in seg_j} v_{w_i}) \quad (11)$$

ここで，v_{w_i} は正解文における i 番目の単語（w_i）の重みであり，$v_{\hat{w}_i}$ は音声認識結果における i 番目の単語（$\hat{w}_i$）の重みである．また，seg_j は j 番目の置換誤り区間を指し，v_{seg_j} は誤り区間 seg_j の重みである．この誤り区間の重み v_{seg_j} は，当該区間 seg_j に含まれる正解系列の単語の重みの合計値と認識結果の単語の重みの合計値のうち，大きいほうとする．参考のため，図 2 に重みつき単語誤り率の計算例を示す．

なお本研究では，誤り単語を同定する際に，単語誤り率を求める際に用いた DP マッチングの結果を用いている．したがって，全ての単語の重みを等しく設定した場合，重みつき単語誤り率は単語誤り率と等しくなる．

3.2 重みつきキーワード誤り率

本研究では音声認識は，重要文抽出を目的として行う．本研究での重要文抽出は，2.3 節で述べたように tf-idf 尺度に基づいて行うものであり，各単語 m に tf-idf 値（KW_m）を与えている．したがって，この tf-idf 値(注2)を各単語の重みとして重みつき単語誤り率の算出に用いることで，重要文抽出を目的とした音声認識の評価尺度となると期待できる．

ベースラインの重要文抽出手法では，固有名詞，代名詞，数詞を除く名詞をキーワードと定義し，キーワードには tf-idf 値（単語の重み）を与え，キーワード以外の単語には重み 0 を与えている．このようにキーワード以外に重み 0 を与える重みつき単語誤り率を，本稿では重みつきキーワード誤り率（WKER）と定義する．なお，全てのキーワードの重みを等しく設定（キーワード以外には重み 0）する場合をキーワード誤り率（KER）と定義し，このキーワード誤り率も比較のために用いる．

3.3 音声認識の評価尺度と重要文抽出精度との相関

前節で定義した種々の音声認識の評価尺度と重要文抽出精度との相関について分析を行った．ここでも，表 1 に示す 17 講演を用いて分析を行った．種々の言語モデル，音響モデル(注3)，デコーディングパラメータ（単語挿入ペナルティ）を用いて音声認識を行い，講演ごとに 10 通りの音声認識結果（合計 170 の音声認識結果）を生成し，相関を分析した．

ここでは，重要文抽出の評価尺度として正規化した F 値（$N(F)$）及び κ 値（$N(\kappa)$）を用いている．この，$N(F)$ 及び $N(\kappa)$ は，システムによる重要文抽出精度を人間の重要文抽出精度で除したものと定義する．すなわち，各講演において，人間の重要文抽出精度が 1 になるように正規化している．

結果を表 4 に示す．重みつきキーワード誤り率（WKER）が重要文抽出精度と最も相関が高いことがわかる．一方，単語誤り率（WER）やキーワード誤り率（KER）と重要文抽出精度との相関は相対的に低く，特に正規化 F 値（$N(F)$）との間の相関には有意性が見られなかった．この結果は，重みつきキーワード誤り率（WKER）が重要文抽出の精度を予測するのに適していることを示している．

ASR result	:	a	b	c	d	e	f
Correct transcript	:	a		c	d'	f	g
DP result	:	*C*	*I*	*C*	*S*	*C*	*D*

$$\mathrm{WWER} = (V_I + V_D + V_S)/V_N * 100$$

$$V_N = v_a + v_c + v_{d'} + v_f + v_g$$
$$V_I = v_b$$
$$V_D = v_g$$
$$V_S = \max(v_d + v_e, v_{d'})$$

v_i: weight of word i.

図 2 重みつき単語誤り率 (WWER) の計算例
Fig. 2 Example of weighted word error rate (WWER) calculation

表 4 音声認識の種々の評価尺度と重要文抽出の評価尺度との相関
Table 4 Relation between ASR evaluation measures and indexing evaluation measures

	$N(F)$	$N(\kappa)$
Word Error Rate (WER)	0.00	0.28**
Weighted WER (WWER)	0.09	0.29**
Keyword Error Rate (KER)	0.14	0.37**
Weighted KER (WKER)	0.20**	0.40**

$N(F)$: normalized F-measure
$N(\kappa)$: normalized κ-value
**: significantly correlated (1%)

4. 音声理解のためのベイズリスク最小化デコーディング

前章では，重みつきキーワード誤り率（WKER）と重要文抽出の精度に相関があることを示した．本章では，この重みつきキーワード誤り率を最小化する音声認識手法（デコーディング手法）を提案し，本提案手法が重要文抽出の精度改善に効果があることを示す．なお本提案手法は，ベイズリスク最小化（Minimum Bayes-Risk: MBR）枠組み [14] に基づく音声認識手法である．

(注2)：ただし，tf_j を N-best リストでの出現回数として算出するため，式 (1) で求める KW_m とこの点でのみ異なる．
(注3)：タスクマッチモデル（CSJ 音響・言語モデル）／非マッチモデル（JNAS 音響モデル，新聞記事言語モデル），不特定話者／特定話者モデル

4.1 ベイズリスク最小化音声認識の概念

統計的な音声認識は一般的に，与えられた入力音声信号 X を最もよく説明する単語列 $\hat{W}$ を求めるプロセスとして定式化される．これを式 (12) に示す．

$$\hat{W} = \underset{W'}{\operatorname{argmax}}\, P(W'|X) \tag{12}$$

ベイズ決定理論に基づくと，音声認識は決定規則 ($\delta(X) : X \rightarrow \hat{W}$) と記述できる．ここで，損失関数を $l(W, \delta(X)) = l(W, W')$ とすると，音声認識は以下のベイズリスク最小化の枠組みで記述できる [14].

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W, W') \cdot P(W'|X) \tag{13}$$

なお，式 (12) で示されている一般的な音声認識のプロセスは，式 (13) において 0/1 損失関数を用いる場合と等価であり，文誤り率 (Sentence Error Rate) を最小化するプロセスであるといえる．本研究でのベースライン音声認識システムではこのデコーディング手法を用いている．

次に単語誤り率 (WER) 最小化音声認識について述べる．単語誤り率はテキスト編集における最小編集距離 (Levenshtein distance) と等価であり，式 (5) で与えられるものである．単語誤り率最小化を目的とした場合は，損失関数 $l(W, W')$ としてこの単語誤り率を用いればよいことが知られており，これまで実際に研究も行われてきた [14] [15]．これに対し本研究では，重要文抽出精度の改善を目的とした音声認識を行うため，重要文抽出の精度と最も相関が大きかった音声認識の評価尺度である「重みつきキーワード誤り率 (WKER)」の最小化を考える．すなわち，重みつきキーワード誤り率に基づく損失関数を利用して以下の式 (14) で音声認識を定式化する．

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} \mathrm{WKER}(W, W') \cdot P(W'|X) \tag{14}$$

ここで，右辺の事後確率 $P(W'|X)$ はベイズ則により $P(W', X)/P(X)$ と記述でき，分母の $P(X)$ は式全体の最小化に影響を及ぼさないため，式 (14) は次のように変形できる．

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} \mathrm{WKER}(W, W') \cdot P(W', X) \tag{15}$$

さらに文献 [14] で有効性が主張されているように，スコア正規化のためのパラメータ λ を導入することにより，決定規則は最終的に次のように記述できる．

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} \mathrm{WKER}(W, W') \cdot P(W', X)^{\frac{1}{\lambda}} \tag{16}$$

なお，デコーディング時に最適な単語列 W を直接求めるのは困難であるため，本研究では，まずベースラインの音声認識手法により N-best リストを生成し，次に生成された N-best リストのリスコアリングを行うことで，ベイズリスク最小化デコーディングを行う．

表 5 重みつきキーワード誤り率 (WKER) 最小化デコーディングの結果
Table 5 Result of WKER minimization decoding

ID	WKER (%) 1-best→ MBR	Key sentence indexing accuracy (F-measure) 1-best→ MBR
A04M0123	42.19 → 42.20	0.555 → 0.535
A04M0121	41.19 → 41.16	0.482 → 0.498
A04M0051	9.38 → 8.85	0.630 → 0.630
A01M0056	10.66 → 9.73	0.529 → 0.532
A01M0035	41.90 → 40.67	0.533 → 0.526
A01M0007	8.91 → 8.76	0.600 → 0.600
A01M0110	35.15 → 35.19	0.629 → 0.629
A01M0141	26.31 → 26.39	0.508 → 0.484
A01M0137	43.56 → 42.93	0.537 → 0.545
A01M0074	34.47 → 33.89	0.508 → 0.522
A01M0097	4.09 → 3.15	0.547 → 0.533
A03M0112	15.90 → 14.95	0.506 → 0.507
A03M0016	38.03 → 36.99	0.564 → 0.551
A03M0156	39.76 → 38.11	0.446 → 0.452
A03M0106	53.64 → 52.60	0.485 → 0.479
A05M0011	49.83 → 49.39	0.558 → 0.585
A05M0031	22.57 → 22.48	0.621 → 0.639
Average	25.57 → 24.96	0.545 → 0.548

表 6 種々の音声認識手法の比較
Table 6 Comparison of decoding methods

minimization target	WER	WKER	key sentence indexing accuracy (F-measure)
WER	25.69	25.00	0.545
WKER	26.10	24.96	0.548
baseline	25.94	25.57	0.545

4.2 実験結果

テストセット (学会講演 17 講演) を用いて，キーワード誤り率最小化デコーディング (N-best リスコアリング) 及びそれに基づく重要文抽出の実験を行った．本研究では，発話ごとに N-best リスト ($N = 1000$) を生成して実験を行った．リスコアリングパラメータ λ は予備実験に基づき 18 に設定した．

表 5 に重みつきキーワード誤り率最小化デコーディングの結果を示す．提案手法により，重みつきキーワード誤り率が 25.57%から 24.96%に改善され，提案手法が適切に機能していることがわかった．

表 6 には，重みつきキーワード誤り率最小化デコーディング (提案手法) と単語誤り率最小化デコーディングを用いた場合の音声認識精度及び重要文抽出精度を示す．重みつきキーワード誤り率最小化デコーディングを用いた場合は，重みつきキーワード誤り率の改善にともない，重要文抽出精度 (F 値) も 0.548 に改善された．一方，単語誤り率最小化デコーディングを用いた場合は，単語誤り率の改善は見られたが，重要文抽出精度の改善は見られなかった．このことは，講演からのキーワードに基づく重要文抽出に対して，提案手法である重みつき

キーワード誤り率最小化デコーディングが有効であることを示している．

5. おわりに

講演などのドメインを限定しない話し言葉音声の音声理解を目的とした音声認識の評価尺度と認識手法について報告を行った．まず，音声認識の種々の評価尺度と重要文抽出の評価尺度との関係を調べ，重要文抽出に影響の大きい単語に情報検索の観点に基づく重みを付与した「重みつきキーワード誤り率」が，重要文抽出の精度と相関が大きいことを確認した．次に，この重みつきキーワード誤り率を最小化するデコーディング手法を提案・実装し，評価を行った．提案した重みつきキーワード誤り率最小化デコーディングが，キーワードに基づく重要文抽出（音声理解）に効果があることを示した．

今後は，リスコアリング時の重みの自動設定や，種々のリスコアリング手法の比較・検討を行っていく予定である．

文　献

[1] S.Furui, K.Maekawa, and H.Isahara. Toward the realization of spontaneous speech recognition – introduction of a Japanese priority program and preliminary results –. In *Proc. ICSLP*, Vol. 3, pp. 518–521, 2000.

[2] S.Furui. Recent advances in spontaneous speech recognition and understanding. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 1–6, 2003.

[3] 前川喜久雄. 言語研究における自発音声. 音講論, 1-3-10, 春季 2001.

[4] 小磯花絵, 前川喜久雄. 『日本語話し言葉コーパス』の概要と書き起こし基準について. 情報処理学会研究報告, 2001-SLP-36-1, 2001.

[5] H.Nanjo and T.Kawahara. Language model and speaking rate adaptation for spontaneous presentation speech recognition. *IEEE Trans. Speech & Audio Process.*, Vol. 12, No. 4, pp. 391–400, 2004.

[6] T.Kawahara, M.Hasegawa, K.Shitaoka, T.Kitade, and H.Nanjo. Automatic indexing of lecture presentations using unsupervised learning of presumed discourse markers. *IEEE Trans. Speech & Audio Process.*, Vol. 12, No. 4, pp. 409–419, 2004.

[7] 野畑周, 関根聡, 内元清貴, 井佐原均. 話し言葉コーパスにおける文の切り分けと重要文抽出. 「話し言葉の科学と工学」ワークショップ予稿集, pp. 93–100, 2002.

[8] 菊池智紀, 古井貞熙, 堀智織. 重要文抽出と文圧縮による音声自動要約. 信学技報, SP2002-158, 2002.

[9] T.Kawahara, H.Nanjo, T.Shinozaki, and S.Furui. Benchmark test for speech recognition using the Corpus of Spontaneous Japanese. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 135–138, 2003.

[10] 小椋秀樹. 話し言葉コーパスの単位認定基準について. 「話し言葉の科学と工学」ワークショップ講演予稿集, pp. 21–28, 2001.

[11] 内元清貴, 井佐原均. 話し言葉コーパスの形態素解析. 「話し言葉の科学と工学」ワークショップ講演予稿集, pp. 33–38, 2002.

[12] 南條浩輝, 北出祐, 河原達也. 談話標識の統計的学習に基づいた講演からの重要文抽出. 日本音響学会研究発表会講演論文集, 2-6-18, 秋季 2003.

[13] 南條浩輝, 北出祐, 河原達也. 談話標識の統計的選択に基づいたCSJ の講演からの重要文抽出. 電子情報通信学会技術研究報告, SP2003-125, NLC2003-62 (SLP-49-13), 2003.

[14] V.Goel, W.Byrne, and S.Khudanpur. LVCSR rescoring with modified loss functions: A decision theoretic perspective. In *Proc. IEEE-ICASSP*, Vol. 1, pp. 425–428, 1998.

[15] A.Stolcke, Y.Konig, and M.Weintraub. Explicit word error minimization in N-best list rescoreing. In *Proc. EUROSPEECH*, pp. 163–165, 1997.