Vol.2013-SLP-99 No.2
2013/12/19

# 音声入力による音声ドキュメント検索における単語重要度を考慮したベイズリスク最小化音声認識

志々見亮[†1]　西田昌史[†1]　南條浩輝[†2]　山本誠一[†1]

これまで我々は，ホームページを対象としたドキュメント検索ならびに講演音声を対象とした音声ドキュメント検索において，単語の重要度を考慮したベイズリスク最小化音声認識を適用し，有効性を示した．それに対して本研究では，音声クエリならびに講演音声に対して，ベイズリスク最小化音声認識におけるリスコアリングにより得られたN-best の上位仮説を用いることで検索精度の改善について検討した．講演音声を対象とした音声ドキュメント検索の評価実験を行った結果，従来の尤度最大化音声認識と比較し，ベイズリスク最小化音声認識により N-best の上位仮説の認識精度の改善ならびに検索精度の改善が得られた．

# Spoken Document Retrieval using Minimum Bayes-Risk Decoding based on Word Significance

RYO SHISHIMI[†1]　MASAFUMI NISHIDA[†1]
HIROAKI NANJO[†2]　SEIICHI YAMAMOTO[†1]

In conventional studies, we applied Minimum Bayes-Risk (MBR) decoding based on word significance to document retrieval for texts and spoken document retrieval for lecture speech and demonstrated that our method was effective. In this study, we evaluated framework using higher rank hypothesis obtained with the MBR decoding for spoken queries and lecture speech. Experimental results showed that our method can improve speech recognition and information retrieval accuracies compared with conventional decoding based on maximum likelihood.

## 1. はじめに

近年の計算機の高性能化や記憶容量の大容量化および低価格化に伴い，情報のマルチメディア化が進行し，音声ドキュメントに対する情報検索技術[1]の開発が進められている．またユーザの情報検索の手法として，情報携帯端末が普及してきた現在，音声入力による情報検索に注目が集まっている．ここで，音声ドキュメント検索においてバックエンドの検索システムがいくら高精度であっても，音声認識部で音声認識誤りが発生した場合，その影響を受けて情報検索精度が低下する．ゆえに，検索システムだけでなく，音声認識精度の問題に対し，これまで様々な研究が行われている．

音声ドキュメント検索の従来研究には，ドキュメント拡張に関する研究[2]，複数のサブワードや言語モデルを使用した検索語検出に関する研究[3]，音声認識と情報検索を一体としてシステム化する研究[4]などがある．ここで検索における単語には，影響の大きい重要語とそうでない単語が含まれるが，従来の一般的な音声認識では全ての単語の認識誤りを同等に扱っている．したがって，検索における音声認識では重要語に重点をおいて，その認識誤りを抑制することが重要であり，重要語がどの程度認識されたかという観点で評価を行う必要があると考えられる．

以上のような背景に基づき，我々は，ベイズリスク最小化（Minimum Bayes-Risk: MBR）の枠組みに基づく音声認識を行う．

これまでこの MBR の枠組みに基づく音声認識は幾つか先行研究[5]～[7]として報告されている．だがこれらは音声認識精度の向上を目的とし，誤り単語の数を減らす実装を行っている．

これに対して我々は，重要語に重点を置いて，その認識誤りの抑制を行うことを目的としたベイズリスク最小化音声認識[8]を用いる．その上で，音声認識の評価尺度として，全ての単語を一様に扱う一般的な「単語誤り率（word error rate: WER）」だけではなく，情報検索の観点から各単語の重要度を考慮した「重み付き単語誤り率（weighted word error rate: WWER）」でも評価を行う．

これまで我々はホームページを検索対象とした音声入力型情報検索に，重要語の認識誤りのリスクを考慮したベイズリスク最小化音声認識を適用し，有効性を示した[9]．また講演音声を対象とした音声ドキュメント検索に対してベイズリスク最小化音声認識を適用し，認識結果の N-best の最上位仮説を用いた評価を行った[10]．

本研究では，講演音声を対象とした音声ドキュメント検索に対し，単語重要度を考慮したベイズリスク最小化音声認識を適用し，認識結果の N-best の上位仮説を用いた検索質問による評価を行った．

†1 同志社大学理工学研究科
Graduate School of Science and Technology, Doshisha University
†2 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku University

本論文の構成を述べる．**2.**では，ベイズリスク最小化音声認識の枠組みからアルゴリズムについて述べる．**3.**では，本研究で適用した情報検索アルゴリズムや評価尺度に関して述べる．**4.**では，音声ドキュメント検索にベイズリスク最小化音声認識を適用した場合の音声認識評価や情報検索評価について述べる．

## 2. ベイズリスク最小化音声認識

### 2.1 ベイズリスク最小化音声認識の枠組み

**統計的な音声認識は一般的に，与えられた入力音声$X$に対する事後確率$P(W'|X)$が最大となる単語列$\widehat{W}$を見つけるプロセスとして式(1)のように定式化される．**

**ゆえに，統計的な音声認識は音響モデルと言語モデルから得られる確率の積を最大化する尤度最大化音声認識の枠組みとして定式化される．**

$$\widehat{W} = \underset{W'}{\operatorname{argmax}} P(W'|X) \tag{1}$$

ここで単語列$W$を$W'$と誤った時の損失を$l(W,W')$とすると，音声認識は以下のベイズリスク最小化（MBR）の枠組みで式(2)のように記述できる．

$$\widehat{W} = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W') \cdot P(W'|X) \tag{2}$$

式(2)の右辺の事後確率はベイズ則を用いて，$P(W',X)/P(X)$と展開でき，分母の$P(X)$は，式全体の最小化に影響を与えないため省略できる．また，それぞれのスコアに重みパラメータを乗じる手法（式(3)）の有効性が先行研究で示されており，本研究ではこれを用いる．

$$\widehat{W} = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W')^{\lambda_1} \cdot P(W',X)^{\lambda_2} \tag{3}$$

なおここで，式(1)で示されている一般的な音声認識のプロセスは，式(3)において 0/1 損失関数を用いた場合と等価である．

### 2.2 重み付き単語誤り率

**情報検索における音声認識では，各単語は異なる重要度を持つ．ゆえに音声認識の評価尺度として，一般的な「単語誤り率（word error rate: WER）」と情報検索の為の評価尺度である「重み付き単語誤り率（weighted word error rate: WWER）」で評価する．WWER は式(4)で定義する．**

$$\mathrm{WWER} = \frac{V_I + V_D + V_S}{V_N} \tag{4}$$

$$V_N = \sum_{w_i} v_{w_i} \ , \ V_I = \sum_{\hat{w}_i \in I} v_{\hat{w}_i}$$

$$V_D = \sum_{w_i \in D} v_{w_i} \ , \ V_S = \sum_{seg_j \in S} v_{seg_j}$$

$$v_{seg_j} = \max\left(\sum_{\hat{w}_i \in seg_j} v_{\hat{w}_i} \ , \ \sum_{w_i \in seg_j} v_{w_i}\right)$$

ただし$v_{w_i}$，$v_{\hat{w}_i}$はそれぞれ正解文と音声認識結果における単語の重みである．$v_{seg_j}$は置換誤り区間$seg_j$の重みである．$v_{seg_j}$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大きい方とする．全ての単語重みを等しく設定した場合，WWER は WER と一致する．また，一部の単語（キーワード）の重みを等しく設定し，残りの単語の重みを 0 に設定した場合，WWER はキーワード誤り率（KER）と一致する．

図 1 に WWER の計算例を示す．この例での置換誤り区間は，正解文での単語$d'$に対応する部分であり，音声認識結果では，単語列$d$，$e$に対応する．重みの大きい単語，すなわち重要度の高い単語を多く誤った場合, WWER は高く算出される．すなわち，WWER は重要度の高い単語の誤りがどの程度少ないかを表す指標である．

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

*C: correct word, I: inserted word,*
*D: deleted word, S: substituted segment*
$V_N = v_a + v_c + v_{d'} + v_f + v_g$, $V_I = v_b$, $V_D = v_g$,
$V_S = max(v_d + v_e, v_{d'})$, $v_i$ *: weight of word i*

図 1 重み付き単語誤り率（WWER）の計算例
Figure 1 Example of WWER calculation

### 2.3 重要語と重要度の定義

**WER の最小化を目的とした場合は損失関数$l(W,W')$として単語誤り率（WER），もしくは WER の定義式の分子に相当する編集距離（Levenshtein Distance）を用いればよい．さらに，重み付き単語誤り率（WWER）の最小化を目的とする場合は，損失関数として，WWER，もしくは WWER の定義式の分子を用いればよい．**

**本研究では，単語の重みを$tf \cdot idf$尺度に基づき，求める．**

**一般的に文書$d$における単語$w$の$tf \cdot idf$値は式(5)のように定義できる．ここで，$tf$とは，ある文書$d$中に出現する単語$w$の頻度であり，$tf(w,d)$で表す．また$idf(w)$は，逆文書頻度である．$N$は検索対象となる文書集合の全文書数，$df(w)$は単語$w$が出現する文書数を表す．**

$$tf \cdot idf(w,d) = tf(w,d) \cdot idf(w) \tag{5}$$

$$idf(w) = log\frac{N}{df(w)} + 1$$

このように，$tf \cdot idf$値は一般的に文書ごとに求めるものである．だがここでは，情報検索に対する影響の大きい単語に高い重要度を与えることが目的であるため，検索対象の文書全体に対して重要度を求める必要がある．その際，文書ごとに求めた$tf \cdot idf$値を単純に全文書で平均する方

法は不適切である．これは平均化処理により，もともと$tf \cdot idf$尺度が表している「ある単語の特定の文書での出現傾向」という情報が生かされないためである．

ゆえに本研究では，$tf \cdot idf$値に基づく単語重みを次の処理で求める．ただし，ベイズリスク最小化音声認識の損失関数における単語の重みは，尤度最大化音声認識結果（baseline）から，評価尺度の WWER における単語の重みは，書き起こし（correct）から求める．

まず検索対象の各文書$d$の各単語（名詞）$w$に対して$tf \cdot idf(w,d)$値を求め，各文書$d$に対して$tf \cdot idf(w,d)$値が高い上位$i$単語をその文書の代表単語とする．そして各単語$w$に対して，それが代表単語となった文書数を単語の重みとする．ただし，この重みが 1 以下の場合は単語の重みを 1 とし，10 以上の場合は 10 とする．本研究では，$i$ = 5 とする．

### 2.4 ベイズリスク最小化音声認識アルゴリズム

本研究では N-best リスコアリングに基づくベイズリスク最小化音声認識のアルゴリズムについて述べる．

まず尤度最大化音声認識を行い，認識スコアの高い順に仮説を N 個求めて N-best を作成し，各仮説$(W_i(i=1,\cdots N))$の評価値$f(W_i)$を計算し，評価値の低い仮説を出力する．なお，仮説$W_i$の評価値$f(W_i)$は以下の式(6)に基づいて計算する．ここで，$P(W_j)$は仮説$W_j$の音声認識スコア，$l(W_i,W_j)$は仮説$W_i$を$W_j$と誤ったときの損失である．

$$f(W_i) = \sum_{W_j \in N-best\ list} l(W_i,W_j)^{\lambda_1} \cdot P(W_j)^{\lambda_2} \tag{6}$$

### 2.5 音声認識エンジン Julius での定式化

本研究では，音声認識エンジンに Julius[11]を利用する．Julius とは，オープンライセンスかつオープンソースの音声認識エンジンである．Julius では，2 パスの探索アルゴリズムを採用し，尤度最大化音声認識の枠組みに基づいて音声認識を行う．すなわち，第 1 パスで単語 bi-gram モデルを用いて荒い照合を行い，その中間結果に対して第 2 パスで単語 tri-gram モデルを適用して，最終的な認識結果を求める．また音響モデルについても，第 1 パスでは単語間については triphone を厳密に適用せず，候補を絞った第 2 パスにおいて正確な尤度を計算する．だが実際の音声認識時には，尤度と事前確率を対数で扱い，音響モデルから得られるスコアと言語モデルから得られるスコアのバランスをとるために，言語モデルのスコアに重み$\alpha$を乗じる．さらに，認識単語列$W$に含まれる単語数$N$に応じてペナルティスコア$\beta N$を与えることで，挿入・削除誤りのバランスを調整している．よって，Julius の音声認識の枠組みは単純な尤度最大化音声認識から拡張が加えられ，式(7)で定式化される．

$$\hat{W} = \underset{W}{\operatorname{argmax}} (\log P(X|W) + \alpha \log P(W) + \beta N) \tag{7}$$

次に，Julius でのベイズリスク最小化の定式化について述べる．まず，式(3)で定式化されるベイズリスク最小化音声認識の枠組みにおけるスコア$P(W',X)$は，Julius の尤度最大化音声認識から得られるものを利用する．すなわち，Julius で音声認識を行う際に算出される仮説スコアを用いる．なお，Julius は仮説スコアを対数で保持するため，MBR の評価値を計算する際には真数に変換する（式(8)）．

$$\hat{W} = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W')^{\lambda_1} \cdot S(W')^{\lambda_2} \tag{8}$$

$$S(W') = P(X|W) \cdot P(W')^{\alpha} \cdot 10^{\beta N}$$

その際，MBR の評価値を算出する際にアンダーフローが起きることを防ぐため，最尤候補$W_1$のスコア$S(W_1) = P(W_1,X)$で全候補のスコアを正規化する（式(9)）．なお，この$S(W_1)$での除算は式(6)を最小にする$W$を求める過程に影響を及ぼさない．

$$\hat{W} = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W')^{\lambda_1} \cdot \left(\frac{S(W')}{S(W_1)}\right)^{\lambda_2} \tag{9}$$

また Julius における N-best リスコアリングに基づくベイズリスク最小化音声認識の評価値は式(10)に基づいて計算する．

$$f(W_i) = \sum_{W_j \in N-best\ list} l(W_i,W_j)^{\lambda_1} \cdot P\left(\frac{S(W_j)}{S(W_1)}\right)^{\lambda_2} \tag{10}$$

## 3. 情報検索

### 3.1 検索モデル

索引語の集合で表現された文書と検索質問の比較によって検索を行う検索モデルには，これまでに多くのモデルが提案されているが，本研究ではベクトル空間モデル（*vector space model*）を採用する．

ベクトル空間モデルでは，索引語の重みを要素とするベクトルで文書を表現する．検索対象となる文書を$D_1,D_2,\cdots D_n$とし，これら文書集合全体を通して全部で$m$個の索引語$w_1,w_2,\cdots w_m$があるとする．このとき，文書$D_j$は，式(11)のように表現される．ここで，$d_{ij}$は索引語$w_i$の文書$D_j$における重みである．本研究では，索引語に名詞を用い，$tf \cdot idf$尺度に基づき重み付けを行う．

$$d_j = \begin{bmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{mj} \end{bmatrix} \tag{11}$$

また，文書集合全体は，次のような$m \times n$行列$D$によって式(12)のように表現することができる．

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \tag{12}$$

検索質問も，文書と同様に索引語の重みを要素とするベクトルで表現することができる．検索質問文に含まれる索引語$w_i$の重みを$q_i$とすると，検索質問ベクトル$q$は式(13)のように記述できる．

$$q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix} \tag{13}$$

実際の文書検索においては，与えられた検索質問文と類似した文書を見つけ出す必要があるが，ベクトル空間モデルでは，これを検索質問ベクトル$q$と各文書ベクトル$d_j$の間の類似度を計算することにより行う．ベクトル間の類似度の定義としては様々なものが考えられるが，文書検索においてよく用いられているものはコサイン類似度であり，式(14)のように計算する．

$$Sim(d_j, q) = \frac{d_j \cdot q}{\|d_j\|\|q\|} = \frac{\sum_{i=1}^{m} d_{ij} q_i}{\sqrt{\sum_{i=1}^{m} d_{ij}^2}\sqrt{\sum_{i=1}^{m} q_i^2}} \tag{14}$$

### 3.2 評価尺度

検索精度の評価尺度には，式(15)に示す補間 11 点平均精度（Interpolated 11-points Average Precision, “11ptAP”と記す）を用いる．これは各検索クエリ$Q$に対して 0.0 から 1.0 まで 0.1 刻みでの各再現率レベル$x$における補間精度$IP_Q(x)$を求め，それらの平均$AP(Q)$を全検索クエリで平均をとったものである．今回は，1 つのクエリに対して上位 1000 件を検索して出力している．ここで$R_{qi}$と$P_{qi}$は，それぞれクエリ$Q$に対する検索結果の上位$i$番目までの検索結果の再現率と適合率である．

また再現率は，検索対象の文書集合の中の検索質問に適合する文書のうち，実際に検索された文書の割合を示すもので，検索漏れの少なさを示す尺度である．適合率は，検索された文書集合の中で，検索質問に適合する文書の割合を示すもので検索ノイズの少なさを示す尺度である．

$$\text{11ptAP} = \frac{1}{N}\sum_{k=1}^{N} AP(Q) \tag{15}$$

$$AP(Q) = \frac{1}{11}\sum_{i=0}^{10} IP_Q\left(\frac{i}{10}\right)$$

$$IP_Q(x) = \max_{x \le R_{qi}} P_{qi}$$

ここで，実際に検索性能評価を計算するためには，検索対象となる文書集合中の各文書に対して検索質問ごとの適合性が与えられている必要がある．適合情報は，検索質問集合の中の各検索質問文に対し，文書集合中のどの文書が適合しているか，または不適合であるかという情報である．また適合か不適合かという 2 元的な情報だけでなく，部分適合などの情報が与えられている場合もある．一般的に，テスト・コレクションと呼ばれる評価用データを用いる．

## 4. 評価実験

### 4.1 音声認識実験

検索質問には，情報処理学会，音声言語情報処理研究会の「音声ドキュメント処理ワーキンググループ」の活動として作成された，日本語話し言葉コーパス（CSJ）[12]を対象とした「CSJ 音声ドキュメント内容検索テスト・コレクション」[13]を使用した．これは後述する CSJ の学会講演 987 講演と模擬講演 1715 講演の合計 2702 講演を検索対象としており，39 件の検索質問が設定されている．検索質問は複数人によって作成されており，39 件の検索質問の例として，「翻訳手法にはどのようなものがあるか」や「ＯＳの役割または種類についての解説を見たい」などがある．これらの検索質問を 20 名に読み上げてもらった．

検索対象となる音声ドキュメントには，上記の CSJ の 2702 講演を使用した．学会講演，模擬講演どちらも独話で自発発話であり，両方の講演を合わせると 600 時間を越える．ここで講演音声を人手で付与されたラベル情報に基づいて一定の無音区間によって分割し，発話毎に音声認識を行った．

検索質問ならびに検索対象の講演音声を認識する際に，認識エンジンとして「Julius-4.1.5」を用いた．ただし，ベイズリスク最小化音声認識機能を実装してある．

音響モデルには CSJ 付属の CSJ の 2496 講演から学習した性別非依存 PTMtriphone モデル，また言語モデルには，CSJ の 2702 講演から学習した「順向き単語 2-gram」と「逆向き単語 3-gram」（語彙 26K）を用いた．

また本研究では，N-best 文数は 100，損失関数の重みパラメータ$\lambda_1 = 1.0$，事後確率に対する重みパラメータ$\lambda_2 = 0.10$とした．これらのパラメータの値は，先行研究[14]によって導き出されたものである．また式(7)及び式(10)の$S(W)$算出のための言語モデルスコアのパラメータ$\alpha$及び単語挿入ペナルティのパラメータ$\beta$は，それぞれ 8, -2 とした．これらの値はいずれも「Julius-4.1.5」のデフォルト値であり，その他の Julius のオプションも変更せずデフォルト値を用いている．

ここで，ベイズリスク最小化音声認識の損失関数における重要語の重みの頻度分布を図 2 に示す．

重要語の重みは前述の通り，代表単語となった文書数としている．この重みが 10 以上の場合は 10 と定義したが，重みが 10 より大きい重要語はそれほど多く見られず，重みが 1 から 10 の間に集約していることが確認できる．また，重みが 1 の重要語が多いことも見て取れる．

本研究では，重要語の重みを$tf \cdot idf$尺度に基づいて求めたが，重要語とその重要度をどう定義するかにより，精度面に影響すると思われる．ゆえに，様々な手法で試していく必要があると考えられる．

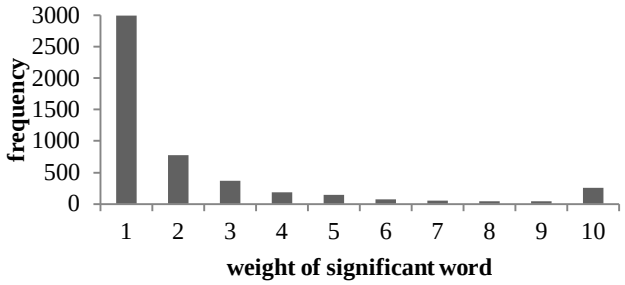


図 2　重要語の重みの頻度分布
Figure 2　Frequency of weight of significant word

次に，尤度最大化音声認識（baseline）とベイズリスク最小化音声認識（MBR），それぞれの N-best 中の最上位仮説に対し，認識精度を求めた．結果を表 1 に示す．

Query，CSJ ともに尤度最大化音声認識と比較し，ベイズリスク最小化音声認識により，一般的な評価尺度の WER，そして単語の重要度を加味した WWER 両方の認識精度の改善が得られた．

表 1　音声認識精度（%）
Table 1　ASR Results (%)

| | | WER | WWER |
|---|---|---|---|
| Query | baseline | 21.41 | 21.89 |
| | MBR | 21.36 | 21.77 |
| CSJ | baseline | 24.14 | 30.24 |
| | MBR | 24.03 | 29.96 |

次に，検索質問が音声の場合，単語が正しく認識されるとは限らない．ゆえに検索においては，誤認識した際のリスクを考慮に入れ，最上位仮説のみだけでなく，複数仮説を用いて検索を行う必要性が考えられる．

また，ベイズリスク最小化音声認識により，正解仮説がたとえ最上位仮説でなくとも，尤度最大化音声認識と比較し，リスコアリングにより上位仮説に上がってくる可能性がある．

本研究では，1-best から 10-best までの検索質問の尤度最大化音声認識精度とベイズリスク最小化音声認識精度をそれぞれ求めた．WER の結果を図 3 に，WWER の結果を図 4 に示す．baseline と比較し，MBR では，1-best の認識精度はほぼ同等であるものの，2-best 以降の認識精度は WER，WWER ともに大幅に改善している．ここで特に WWER の結果においては，2-best 以降の MBR の認識精度が一定の水準を保っていることから，baseline と比較し，リスコアリングにより上位仮説における重要語の認識精度の向上を確認できる．

以上から，MBR 結果の複数仮説を検索に用いることは，baseline 結果よりも検索精度の向上の可能性を示している．

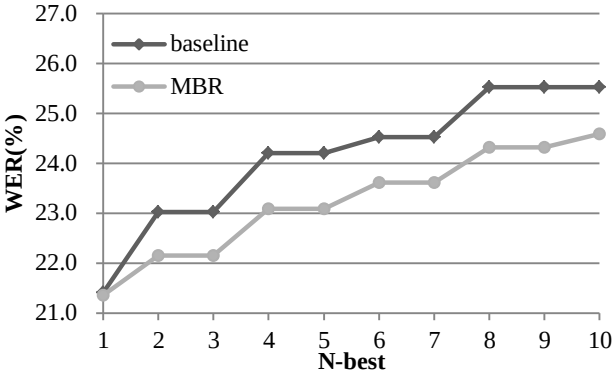


図 3　検索質問に対する単語誤り率（N-best）
Figure 3　WER of spoken queries (N-best)

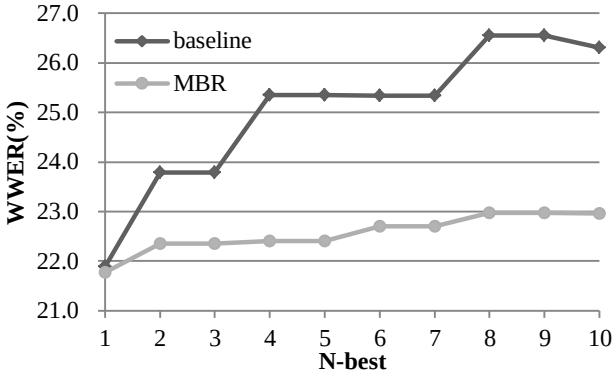


図 4　検索質問に対する重み付き単語誤り率（N-best）
Figure 4　WWER of spoken queries (N-best)

また，検索質問における baseline 認識結果と MBR 認識結果の N-best 中に正解仮説が含まれる割合を図 5 に示す．baseline と比較し MBR の方が，正解仮説が N-best 中に含まれる割合が高いことが分かる．

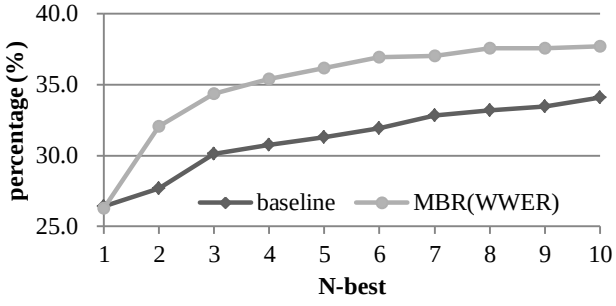


図 5　N-best 中に正解仮説が含まれる割合
Figure 5　Rate of correct hypothesis in N-best

### 4.2 情報検索実験

**検索精度の評価尺度には，補間 11 点平均精度（11ptAP）を用いる．ここでテスト・コレクションにおける適合情報として，検索質問に適合（R; *Relevant*），部分適合（P; *Partially Relevant*）する講演音声 ID に関する情報が与えられており，本研究では正解判定に R+P を用いる．**

Query，CSJ ともに書き起こし（correct），尤度最大化音声認識結果（baseline），ベイズリスク最小化音声認識結果（MBR）の場合での，それぞれの N-best 中の最上位仮説を用い，検索精度を算出した．結果を表 2 に示す．

文書ベクトルの重みに$tf$と比較し，$tf \cdot idf$を用いた場合，検索精度が改善した．また baseline 結果と比較し，MBR 結果を文書ベクトルに用いた場合，検索精度が改善した．

表 2　情報検索精度（1-best）
Table 2　Information Retrieval Performance (1-best)

| weight of vector | Query | CSJ | 11ptAP |
|---|---|---|---|
| $tf$ | correct | correct | 0.245 |
| $tf \cdot idf$ | correct | correct | 0.325 |
| $tf \cdot idf$ | baseline | baseline | 0.230 |
| $tf \cdot idf$ | MBR | MBR | 0.232 |

次に，検索質問の複数仮説を用いた検索結果について示す．本研究では，3-best，5-best，10-best までを検索質問とした場合で評価を行った．文書ベクトルの重みは$tf \cdot idf$とし，結果を表 3 に示す．baseline，MBR ともに 1-best と比較し，複数仮説を用いた場合，N の値が大きすぎると検索精度が低下した．しかし MBR を用いた場合，baseline よりも N-best 上位仮説の認識精度が高いため，N の値が大きくても検索精度は baseline ほど下がらなかった．

これまで先行研究として，講演音声の書き起こしに対して検索質問の複数仮説を用いた音声検索に関する研究[15]が行われており，似通った傾向となった．

表 3　情報検索精度（N-best）
Table 3　Information Retrieval Performance (N-best)

| N-best | Query CSJ | 11ptAP | Query CSJ | 11ptAP |
|---|---|---|---|---|
| 1-best | baseline | 0.230 | MBR | 0.232 |
| 3-best | | 0.229 | | 0.232 |
| 5-best | | 0.226 | | 0.231 |
| 10-best | | 0.222 | | 0.230 |

## 5. おわりに

本研究では講演音声を対象とした音声入力型音声ドキュメント検索に対してベイズリスク最小化音声認識を適用し，評価を行った．その結果，従来の尤度最大化音声認識と比較し音声認識精度が改善した．特に，リスコアリングにより N-best の上位仮説の認識精度の改善が大きかった．そして情報検索においては，尤度最大化音声認識結果を用いるよりも，ベイズリスク最小化音声認識結果を用いた場合，情報検索精度の改善を得ることができた．特に複数仮説を用いた場合，検索精度の差が顕著となった．

**謝辞**　本研究では，大語彙連続音声認識エンジン Julius を利用した．Julius の開発・公開に携わる関係各位に感謝する．また本研究は科研費基盤研究(B)(21300066)の助成を受けたものである．

## 参考文献

1)　大淵康成, 神田直之: 音声検索実用化の現状と課題, 情報処理学会研究報告, 2011-SLP-88, No.5, pp.1-4 (2011).
2)　杉本樹世貴, 西崎博光, 関口芳廣: 音声ドキュメント検索における Web ページを用いたドキュメント拡張の効果, 情報処理学会研究報告, 2009-SLP-76, No.11, pp.1-7 (2009).
3)　小野寺悠二, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭: 複数のサブワード・言語モデルを用いた音声中の検索語検出の高精度化, 第 4 回音声ドキュメント処理ワークショップ講演論文集, No.14 (2010).
4)　翠輝久, 駒谷和範, 清田陽司, 河原達也: 音声対話によるソフトウェアサポートのための効率的な確認戦略, 電子情報通信学会論文誌, Vol. J88-DII, No. 3, pp. 499‐508 (2005).
5)　L. Mngu, E. Brill, and A. Stolcke: Finding consensus in speech recognition: word error minimization and other applications of confusion networks, Computer Speech and Language, Vol.14, pp.373-400 (2000).
6)　V. Goel, W. Byrne, and S. Khudanpur: LVCSR rescoring with modified loss functions: A decision theoretic perspective, Proc. IEEE-ICASSP, Vol.1, pp.425-428 (1998).
7)　A. stolcke, Y. Konig, and M. Weintrub: Explicit word error minimization in N-best list rescoring, Proc. EUROSPEECH, pp163-165 (2007).
8)　南條浩輝, 河原達也, 七里崇: 音声理解を指向したベイズリスク最小化枠組みに基づく音声認識, 電子情報通信学会論文誌, Vol.J91-D, No.5 (2008).
9)　志々見亮, 西田昌史, 南條浩輝, 山本誠一: 音声入力型情報検索に対する単語信頼度によるリスコアリングを適用したベイズリスク最小化音声認識, 日本音響学会研究発表会講演論文集（秋季）, 3-P-32, pp.205-206 (2012).
10)　志々見亮, 西田昌史, 南條浩輝, 山本誠一: ベイズリスク最小化音声認識を適用した音声入力型音声ドキュメント検索, 日本音響学会研究発表会講演論文集（秋季）, 2-P-27, pp.221-222 (2013).
11)　河原達也, 李晃伸: 連続音声認識ソフトウェア Julius, 人工知能誌, Vol.20, No.1, pp.41-49 (2005).
12)　K. Maekawa: Corpus of Spontaneous Japanese: Its design and evaluation, Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003), pp.7-12 (2003).
13)　T. Akiba, K. Aikawa, Y. Itoh, T. Kawahara, H. Nanjo, H. Nishizaki, N. Yasuda, Y. Yamashita, and K. Itou: Construction of a test collection for spoken document retrieval from lecture audio data, IPSJ-Journal, Vol.50. No.2, pp.82-94 (2009).
14)　南條浩輝, 古谷遼, 西田昌史: オープンソース音声認識エンジン Julius へのベイズリスク最小化機能の実装と評価, 電子情報通信学会論文誌（D）, Vol.J96-D, No.10, pp.2530-2539 (2013).
15)　南條浩輝, 古谷遼: ベイズリスク最小化音声認識の複数仮説を用いた音声検索, 情報処理学会研究報告, 2013-SLP-97, No.7, pp.1-8 (2013).