論文 / 著書情報
Article / Book Information

論題(和文)	新聞記事データベースに用いた大語い連続音声認識
Title(English)	Large-Vocabulary Continuous Speech Recognition Using a Japanese Buisiness-Newspaper Corpus
著者(和文)	松岡達雄, 大附克年, 森岳至, 古井 貞熙, 白井克彦
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J79-D-II, No. 12, pp. 2125-2131
Citation(English)	, Vol. J79-D-II, No. 12, pp. 2125-2131
発行日 / Pub. date	1996, 12
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1996 Institute of Electronics, Information and Communication Engineers.

論　文

新聞記事データベースを用いた大語い連続音声認識

松岡　達雄[†]　大附　克年[††]　森　岳至[†††]　古井　貞熙[†,†††]
白井　克彦[††]

Large-Vocabulary Continuous Speech Recognition Using a Japanese Business-Newspaper Corpus

Tatsuo MATSUOKA[†], Katsutoshi OHTSUKI[††], Takeshi MORI[†††], Sadaoki FURUI[†,†††], and Katsuhiko SHIRAI[††]

あらまし　近年，大語い連続音声認識の研究がアメリカ英語，イギリス英語，フランス語，ドイツ語，イタリア語などを対象に新聞記事を用いて盛んに行われている．しかしながら，日本語を対象とした，これに類する研究については報告がない．これは，主に，日本語が単語間にスペースなどのデリミタをおくことなく書かれるため，大語い連続音声認識において重要な役割を果たす単語N-gramなどの言語モデルの導入が容易でないためと考えられる．我々は，日本語新聞記事を対象として大語い連続音声認識の研究を進めている．単語N-gramを言語モデルとして用いるため，テキストを形態素解析することにより形態素（単語）にセグメンテーションした．形態素を単語と定義し，約5年分の新聞記事を用いて単語N-gram言語モデルを推定した．認識システムを評価するため，音声データベースを設計し，54名の話者の各100文ずつの音声データを収録した．この音声データベースの最初の10名の音声を用いて大語い連続音声認識の実験を行った．7kの語いサイズに対して，no-grammar言語モデル，音素文脈独立音響モデルを用いた場合には単語誤り率が82.8%であった．単語bigram言語モデルと音素文脈依存音響モデルを用いることにより単語誤り率が20.0%に改善された．

キーワード　大語い連続音声認識，音声データベース，言語モデル，N-gram，新聞記事

1. まえがき

近年，Wall Street Journal，Le Monde，Frankfurter Rundschau，Sole 24 Oreなどの新聞記事を用いて，アメリカ英語，イギリス英語，フランス語，ドイツ語，イタリア語などを対象に大語い連続音声認識の研究が盛んに行われている[1]～[9]．しかしながら，日本語を対象とした，これらに類する研究はこれまで報告がない．これは，主に，日本語の文章が単語区切りなく書かれており，それを自動的に単語に区切ることが容易ではないことによると考えられる．すなわち，日本語の場合，大語い連続音声認識において重要な役割を果たす単語N-gramのような言語モデルの導入が容易ではない．また，新聞記事のような大語いの音声データベースも日本語については存在していない．そこで，我々は，日本経済新聞を用いて新聞記事読上げ音声データベースを設計/構築し，日本語を対象とした大語い連続音声認識の研究を進めている[10]～[13]．

まず，680万文から60万語の単語頻度リストを作成し，この単語頻度リストに従い，WSJタスクの5k，20k，64kとほぼ同等のカバレージとなるよう7k，30k，150kの語いサイズを決定した．音声データベースの音声は54名の話者が各100文ずつ発声した．

このようにして構築した音声データベースを用いて日本語大語い連続音声認識のための音響モデル，言語モデルについて検討を行った．本論文では，音声データベースの設計と，音素文脈依存音響モデル，単語N-gram言語モデルの効果について述べる．

2. データベースの設計

新聞記事5年分のうち，4年9か月分を学習用に，残り3か月分を評価用とした．

† NTTヒューマンインタフェース研究所，武蔵野市
NTT Human Interface Laboratories, Musashino-shi, Tokyo, 180 Japan
†† 早稲田大学，東京都
Waseda University, Shinjuku-ku, Tokyo, 169 Japan
††† 東京工業大学，東京都
Tokyo Institute of Technology, Meguro-ku, Tokyo, 152 Japan

2.1 テキスト前処理

形態素解析をする前にテキストに対して前処理を施した．これは，文章の読みやすさを考慮したためと，形態素解析の精度を高め，正確な言語モデルの推定を容易にするためである．我々の目的は大語い連続音声認識であり，いわゆるディクテーションではないので，通常の音声によるコミュニケーションにおいて発音されない記号や括弧は取り除いた．

(1) ○●◇◆□■△▲▽▼☆★◎…＃＠※→←↑↓・／｜！？などは，文書における強調記号であり発音が一意に決まっていないので削除した．

例：◎ECが経済通貨同盟で政府間会議．

(2)「」『』 "" '' [] などは，主に語句の強調に用いられているため削除した．

例：ノイマン型コンピューターの情報は「0」か「1」かで表現される．

(3) () 【】 〔〕 〈〉 《》などは，難しい漢字の読み，語句の説明や見出しなどに用いられており，中身ごと削除しても文の構成が保たれるため，括弧・中身とも削除した．

例：やわらかな日差しがそそぐアトリウム（吹き抜け空間）．

長すぎる文も読みにくいため削除した．文中の単語数の分布を正規分布と仮定した．図1は，テキストデータベース全体に対する，1文中の単語数で測った文の長さの分布である．1文中の平均単語数は，学習セット，評価セットとも25.6単語で，標準偏差は学習セット中では13.8，評価セット中では13.5であった．文の長さが平均単語数±2σ以内のものを単語頻度リストの作成とN-gram言語モデルの学習に用いた．評価セットは，平均単語数±σの長さに収まるものに制限した．テキスト前処理後，学習セットは680万文，1億8000万単語（形態素），評価セットは34万2千文，980万単語（形態素）となった．

2.2 形態素解析

日本語は単語間にスペースなどのデリミタを入れずに表記されるため単語境界は明白でない．単語頻度のような統計量をとるためにはまず，文章を単語単位にセグメンテーションする必要がある．このセグメンテーションのためには高度な形態素解析を必要とする．我々が用いた形態素解析は25万形態素の辞書をもち，日経新聞記事に対する解析精度は95%である．本研究ではこの形態素解析に基づき形態素を単語と定義する．

形態素解析の誤りは，複数の形態素が結合して未知語として解析されたり，辞書にない自立語が助詞などの短い形態素に誤って分割され単語（形態素）数の多

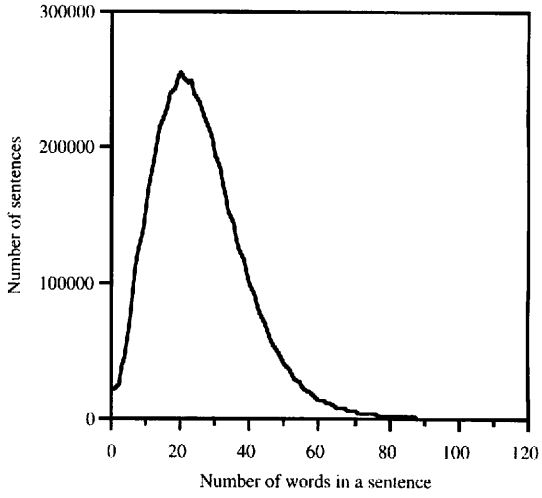

図1 1文中の単語数で測った文の長さの分布
Fig. 1 Distribution of sentence length in terms of the number of words in a sentence.

表1 各言語の語いサイズとカバレージの比較
Table 1 Comparison of lexica and LM training corpora for different languages.

	Nikkei (Japanese)	WSJ (English)	Le Monde (French)	Frankfurter Randschau (German)	Sole 24 (Italian)
Training text size	180 M	37.2 M	37.7 M	36 M	25.7 M
Distinct words	623 k	165 k	280 k	650 k	200 k
5 k coverage	88.0 %	90.6 %	85.2 %	82.9 %	88.3 %
7 k coverage	90.3 %	-	-	-	-
20 k coverage	96.2 %	97.5 %	94.7 %	90.0 %	96.3 %
30 k coverage	97.5 %	-	-	-	-
40 k coverage	98.2 %	99.2 %	97.6 %	-	98.8 %
65 k coverage	99.0 %	99.6 %	98.3 %	95.1 %	99.0 %
20 k OOV rate	3.8 %	2.5 %	5.3 %	10.0 %	3.7 %

い文として解析されたりする例が多い．未知語に関連する解析誤りを取り除くため，上位15万語（カバレージ99.6%）に含まれない語を含む文は削除した．また，前述のように読みやすさを考慮して長い文を削除することとしたため，誤って細かい形態素に分割され，1文中の単語（形態素）数が多くなっている文も削除することができる．これらの処理により，形態素解析誤りはほとんど排除できた．

語いサイズを決定するため，学習セット中の単語を頻度順に並べた単語頻度リストを作成した．その結果，62万3千語からなる単語頻度リストが得られた．形態素解析の辞書は25万語であるから62万3千語のうち37万3千語は未知語として解析されたことになる．未知語のほとんどは固有名詞や特殊な専門用語であった．

2.3 データベースの設計

表1は各言語に対する新聞記事を用いた大語い連続音声認識タスクの語いサイズとカバレージを示している[9]．WSJタスクの5k，20kと同等のカバレージを維持するため7k，30kの語いサイズを決定した．日本語とドイツ語は複合語が多いため異なり語い数が多くなっている．更に，日本語では活用形により見掛けの異なる語い数は多くなる．日経タスクでの未知語の割合は英語とドイツ語の中間であった．

表2は，辞書中の同音異義語の数を示す[9]．homophone-class size =1は，同音異義語のない単語を意味し，homophone-class size=2は同音異義語が一つあることを示している．日本語とフランス語において辞書中の同音異義語の割合が高い．辞書中に同音異義語が多い場合，表記のみで単語N-gramを推定すると，意味の異なる単語を同一単語として統計量を求めることとなり，統計量としての精度が低下する可能性がある．意味まで正確に用いて言語モデルを推定するためには意味的な知識源が必要であり，また，認識時の処理も複雑となる．

大語い連続音声認識システムを評価するため，語いサイズと1文中の未知語数を設定し，学習セット，評価セットごとに5種類のサブセットを定義した．各サブセットの定義を表3に示す．未知語についても単語頻度リストの上位150kに現れたものに限っているため，我々のタスクは150kで閉じていることになる．

表4は，各サブセットに定義された語いで表現可能な文章数の割合を示している．サブセット30k++は，1文中に三つ以上の未知語を許すにもかかわらず，その語いで表現できる文章数は全体の64.2%にとどまっている．

表2　同音異義語の種類と数の比較
Table 2　Entry items corresponding to the number of homophone classes.

Corpus	Rate in Lexicon	Homophone class size			
		1	2	3	≥4
Nikkei (30 k)	20 %	24.1 k	2438	716	565
BREF (10 k)	35 %	6686	1329	215	73
BREF (40 k)	45 %	23.7 k	5361	1264	1039
WSJ (9 k)	6 %	8453	237	22	1
WSJ (65 k)	15 %	60.4 k	3689	890	291
FR (64 k)	10 %	58.1 k	2769	221	57
So24 (10 k)	1.7 %	9872	85	3	0

表3　各サブセットの定義
Table 3　Description of subsets.

Subset	Description
7k	Sentences composed solely from 7k vocabulary
7k+	Sentences composed from 7k vocabulary and up to two OOV words
30k	Sentences composed solely from 30k vocabulary
30k+	Sentence composed from 30K vocabulary and up to two OOV words
30k++	Sentences composed from 30K vocabulary and more than two OOV words

表4　各サブセットの語いで完成される文の割合
Table 4　Percentage of complete sentences made by each subset.

Subset	Training set	Testing set
7k	10.7	11.1
7k, 7k+	40.8	41.8
30k	42.0	42.9
30k, 30k+	62.0	62.2
30k, 30k+, 30k++	64.2	63.9

表5　1文の平均単語数と平均継続時間
Table 5　Average number of words and duration for each sentence.

Subset	Number of words	Duration [s]
7k	20.9	6.3
7k+	23.3	7.0
30k	23.4	7.1
30k+	25.7	7.9
30k++	27.4	8.6

る．

54名の話者が，各サブセットから20文ずつ，計100文ずつ文を発声した．学習セットと評価セットに対する認識性能の比較から言語モデルの頑健性を評価するため，50文は学習セットから，残り50文は評価セットから選んだ．音声は，接話マイク（ゼンハイザーHMD-410）と，バウンダリマイク（クラウンPCC-160）を用いて同時2ch録音した．以下の実験ではすべて接話マ

イクで収録した音声を用いている．表5は，1文中の平均単語数と1文の平均継続時間である．語いサイズが大きくなるにつれて文の長さも長くなるがそれほど顕著ではない．

3. 音響モデル

大語い連続音声認識では，単語単位の音響モデルを用いることは現実的には不可能であるから，音素単位などのsub-word音響モデルを用いるのが妥当である．表6に我々が用いた音素体系を示す．無音を含んで42種類の音素を用いた．

音素文脈独立モデル，音素文脈依存モデル（diphone/triphone context）について評価を行った．各モデルは，ATR Bセットを用いて，ラベル情報を使い初期モデルを学習し，日本音響学会連続音声データベース中の503文と模擬対話を用い連結学習を行った．合計，58名の発声した15,000文を学習に用いた．評価には，今回収録した新聞記事読上げデータベースの最初の5名分，500文を用いた．学習音声と評価対象のタスクが異なる上に，収録マイクが，通常のスタンドマイクと接話マイクで異なっているため，同一タスクからの音声を音響モデルの学習に用いている他言語の大語い連続音声認識タスクよりやや困難な条件となっている．

各音素HMMの状態数は3状態とした．各状態は，4混合の混合ガウス分布で表現した．音声のサンプリングレートは12kHz，量子化は16bitである．フレーム長32ms，フレームシフト8 msでLPC分析した16次のLPCケプストラムと正規化対数パワーとそれぞれの1次時間微分を特徴量として用いた．diphone-context-dependentモデルでは前側（左側）コンテキストを考慮している．表7に，音素認識実験結果を示す．この実験では，音素を認識単位として，任意の音素に任意の音素が接続可能な，いわゆるno-grammarネットワークを用いて連続音声認識を行った．正解精度（*Accuracy*）は次のように計算した．

$$Accuracy = \left(1 - \frac{S + D + I}{N}\right) \cdot 100,$$

ここで，S，D，Iは，それぞれ置換，脱落，挿入誤りである．Di2000は，diphone-context-dependentモデルで，学習サンプルが2000個以上学習データ中に観測されたモデルのセットを意味する．Tri600は，triphone-context-dependentモデルで，同様に学習サンプルが600個以上観測されたモデルのセットを意味する．Di700とTri300と文脈独立モデルを組み合わせて用いたときに最良の結果が得られ，61.6%の音素正解精度であった．

表6 音素体系
Table 6 Phonemes.

Vowels	a, e, i, o, u aa, ee, ei, ii, oo, ou,uu
Consonants	b, d, g, p, t, k ch, j, sh, ts, f, h, s, z N, m, n, r, w, y by, gy, hy, ky, my, ny, py, ry
Double consonant	Q
Silence	#

表7 音素認識実験結果：音素正解精度(%)
Table 7 Phoneme recognition accuracy. (%)

	CI	Di2000	Di1000	Di700	Di500	Di300	Di100
CI	49.2	58.0	58.4	58.2	57.9	57.2	56.7
Tri600	58.4	60.4	60.6	60.6	57.9	60.1	-
Tri500	59.4	61.0	60.9	61.0	60.5	60.6	-
Tri400	58.9	61.0	60.9	61.2	61.1	60.8	-
Tri300	60.6	61.5	61.2	61.6	61.5	61.3	-
Tri200	60.4	60.9	60.6	60.9	60.9	60.8	-
Tri100	60.9	61.3	60.8	61.0	61.1	60.9	-
Tri50	60.9	-	-	-	-	-	-

4. 言語モデル

大語い連続音声認識では言語モデルが不可欠である．言語モデルとしては，ネットワーク文法，文脈自由文法，N-gramなどが考えられるが，大語いの場合には簡潔かつ正確なネットワーク文法や文脈自由文法を人手で書くことはほとんど不可能である．一方，N-gram言語モデルは，大量のテキストデータがあれば比較的容易に統計的に推定することができるため，大語い連続音声認識によく用いられている．

表8に，学習セット中に観測されたunigram，bigram，trigramの種類数と平均出現頻度を示す．bigram，trigramは，そのほとんどが学習テキスト内に1回だけしか観測されないsingletonであった．bigram，trigramの出現頻度が低く，明らかに言語モデルに対し

表8 学習セット中のN-gramの種類数（上段）と平均出現頻度（下段）
Table 8 Number of N-grams in the training set (upper), and average number of occurences of each N-grams. (lower)

	Unigram	Bigram	Trigram
7 k	7000 24388.0	2.1 M 65.1	17.1 M 7.2
30 k	30000 6121.9	4.9 M 33.8	30.5 M 5.1

表9　テストセットパープレキシティ
Table 9　Test-set perplexity.

Nikkei			WSJ			
Vocabulary size	Language model	Test-set perplexity	Vocabulary size	Language model	Test-set perplexity	
					VP	NVP
7 k	Unigram	597	5 k	Unigram	-	-
	Bigram	82		Bigram	80	118
	Trigram	38		Trigram	44	68
30 k	Unigram	693	20 k	Unigram	-	-
	Bigram	124		Bigram	158	236
	Trigram	64		Trigram	101	155

表10　大語い連続音声認識実験結果
Table 10　Large-vocabulary continuous-speech recognition experimental result.

Acoustic model		Language model	Training set		Test set	
HMM	Features		% Correct	Accuracy	% Correct	Accuracy
CI	cepstrum delta cesptrum	NG	19.6	18.0	18.1	17.2
CD		NG	24.5	23.0	23.1	22.1
CI		BG	67.0	65.0	65.8	63.7
CD		BG	77.2	73.5	76.0	72.5
CI	+ log-energy delta log-energy	BG	75.3	73.4	73.2	71.5
CD		BG	82.4	80.3	82.3	80.0

て何らかのスムージングが必要である．我々は，Katzによるback-offスムージングを適用した[14]．スムージングを行った言語モデルを用いて，評価セットの新聞記事に対するテストセットパープレキシティを評価した．表9に日経タスクとWSJタスクのテストセットパープレキシティを示す[8]．日経タスクでは，記号類は基本的に読まないこととしたため，言語モデルの計算でも記号は含まれていないが，句読点だけは，文の切れ目として残している．そのため，パープレキシティの比較では，WSJのVP (Verbalized Punctuation: 句読点などの記号類を発音)の場合と比較するのが妥当と考えられるが，引用符の有無や，単語の定義自体の違いなどがあるため厳密な比較はできない．これらの違いはあるが，およその値が近いことは興味深い．

5. 連続音声認識実験

7k語いについて，今回収録した音声データベース中のチェックの終わった最初の10名の話者の音声を用いて，連続音声認識実験を行った．音響モデルは，文脈独立モデル（CI），および，音素認識実験において最もよい性能を示した音響モデルのセット，すなわち，文脈依存モデルと文脈独立モデル，計748モデル（CD）を用い比較を行った．CDの場合も単語間にわたる部分では，文脈独立モデルを用いた．探索には，Viterbiビームサーチを採用し，予備実験から定めた重みで音響モデルと言語モデルのゆう度を加算したスコアを評価尺度とした．ビームサーチでは，予備実験より決定したゆう度幅によるしきい値を設定した．表10に大語い連続音声認識結果を示す．正解率と正解精度は以下のように計算した[15]．

$$\%Correct = \left(1 - \frac{S+D}{N}\right) \cdot 100,$$

$$Accuracy = \left(1 - \frac{S+D+I}{N}\right) \cdot 100,$$

ここで，S, D, Iは，それぞれ，置換，脱落，挿入誤り，また，Nは，全単語数である．

文脈独立音響モデルとno-grammar言語モデルを用いたベースラインの正解精度は，17.2%であった．bigram言語モデルの導入により正解精度は63.7%まで改善された．更に，文脈依存音響モデル，そして，特徴量として対数パワーを用いることで，80.0%まで改善された．すなわち，言語モデルの導入により誤り率が1/2となり，音響モデルを改善することで，誤り率を更に1/2とすることができた．単語bigramを用いた場合も，学習セットと評価セットに対する性能にははとんど差がなく，頑健な言語モデルが推定されていると考えられ

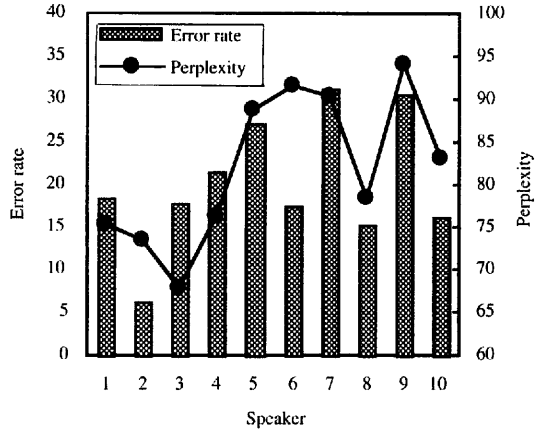

図2 話者ごとのテストセットパープレキシティと単語誤り率
Fig. 2 Test-set perplexity and word-error rate for each speaker.

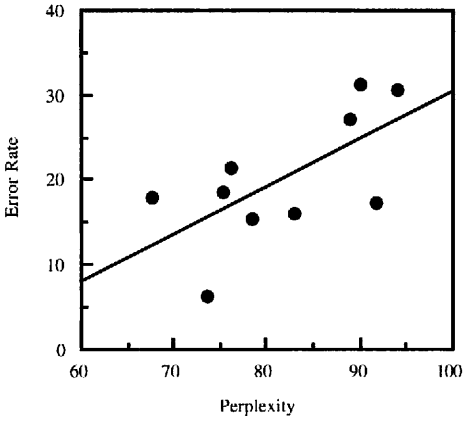

図3 テストセットパープレキシティと単語誤り率の関係
Fig. 3 Relation between perplexity and word-error rate.

る.

図2は，話者ごとの単語誤り率をテストセットパープレキシティと共に示している．話者による正解精度のばらつきが大きいことがわかる．それぞれの話者は異なる文章を読み上げているため，文自体の難しさ，つまり，パープレキシティが認識精度の違いに影響していると考えられる．各話者が読み上げた文章を話者ごとに一つのテストセットと考え，テストセットパープレキシティを求めた．図3は，テストセットパープレキシティと単語誤り率（100.0-正解精度）の関係をプロットしたものである．この結果より，パープレキシティが高いほど単語誤り率が高くなっており，これらには相関があると考えられる．実線は，1次の回帰曲線である．この線からのばらつきは，各話者の音響的な特徴の違いに起因するばらつきと考えることができるであろう．

6. むすび

大語い連続音声認識のための音声データベースの設計について述べ，その音声データベースを用いた大語い連続音声認識システムの評価について述べた．

学習セット中の単語頻度に基づき7k，30kの語いセットを決定し，約5年分の新聞記事から読上げ文を選択した．

構築した音声データベースの最初の10名の話者の音声を用いて大語い連続音声認識システムを評価した．文脈独立音響モデルとno-grammar言語モデルを用いたベースラインシステムの単語誤り率は82.8%であった．文脈依存音響モデルとbigram言語モデルを導入することにより単語誤り率は20.0%に改善された．誤り削減率は76%である．bigram言語モデルも文脈依存音響モデルも単語誤り率の削減に大変効果的であった．bigram言語モデルは単語誤り率を1/2にし，更に，文脈依存音響モデルは残る単語誤りを1/2にした．

今後は，音響モデルに関しては，単語間にわたる音素文脈を考慮したモデルを導入する予定である[16]．言語モデルは，trigramや，品詞などの単語クラスを基準にしたN-gramにより精度向上を図っていきたい．

謝辞 形態素解析プログラムを提供して下さったNTTヒューマンインタフェース研究所田中一男主幹研究員に感謝します．新聞記事日経CD-ROM90-94の本研究への利用を許諾して下さった日本経済新聞社に感謝します．また，音声データベース構築に協力して下さった発声者の皆様，音声データのチェックをして下さった波方祐子氏に感謝します．

文　献

[1] D. B. Paul and J. M. Baker, "The design for the Wall Street Journal-based CSR corpus," Proc. ICSLP-92, pp. 899-902, Oct. 1992.

[2] J. L. Gauvain, L. Lamel, and M. Eskenazi, "Design considerations and text selection for BREF, a large French read-speech corpus," Proc. ICSLP-90, pp. 1097-1100, Oct. 1990.

[3] T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, "WSJCAM0: A British English speech corpus for large vocabulary continuous speech recognition," Proc. ICASSP-95, pp. 81-84, May 1995.

[4] H. J. M. Steeneken and D. A. van Leenwen, "Multi-lingual assessment of speaker independent large vocabulary speech-recognition systems: SQUALE-project," Proc. EUROSPEECH-95, pp. 1271-1274, Sept. 1995.

[5] P. C. Woodland, C. J. Leggetter, J. J. Odell, V. Valtchev, and S. J. Young, "The 1994 HTK Large vocabulary speech recognition system," Proc. ICASSP-95, pp. 73-76, May 1995.

[6] D. Pye, P. C. Woodland, and S. J. Young, "Large vocabulary

multilingual speech recognition using HTK," Proc. EUROSPEECH-95, pp. 181-184, Sept. 1995.

[7] L. Lamel, M. Adda-Decher, and J. L. Gauvain, "Issues in large vocabulary, multilingual speech recognition," Proc. EUROSPEECH-95, pp. 185-188, Sept. 1995.

[8] D. B. Paul, and B. F. Necioglu, "The Lincoln large-vocabulary stack-decoder HMM CSR," Proc. ICASSP-93, pp. 660-663, April 1993.

[9] L. Lamel and R. De Mori, "Speech recognition of European languages," Proc. IEEE Automatic Speech Recognition Workshop, pp. 51-54, Snowbird, Dec. 1995.

[10] 大附克年，森　岳至，松岡達雄，古井貞熙，白井克彦，"新聞記事を用いた大語彙連続音声認識の検討，" 信学技報，SP95-90，pp. 63-68，Dec. 1995.

[11] 森　岳至，大附克年，松岡達雄，古井貞熙，白井克彦，"新聞読み上げタスクを用いた大語彙連続音声認識における言語モデルの検討，" 日本音響学会春季研究発表会，3-8-7，pp. 159-160，March 1996.

[12] 大附克年，森　岳至，松岡達雄，古井貞熙，白井克彦，"新聞読み上げタスクを用いた大語彙連続音声認識における音響モデルの検討，" 日本音響学会春季研究発表会，3-8-8，pp. 161-162，March 1996.

[13] T. Matsuoka, K. Ohtsuki, T. Mori, S. Furui, and K. Shirai, "Large-vocabulary continuous-speech recognition using a Japanese business newspaper (Nikkei)," Proc. of ARPA Speech Recognition Workshop, Feb. 1996.

[14] S. M. Katz, "Estimation of probabilities from sparse data for the language model component of a speech recognizer," IEEE Trans. vol.ASSP-35, pp. 400-401, March 1987.

[15] F. Kubala, Y. Chow, A. Derr, M. Feng, O. Kimball, J. Makhoul, P. Price, J. Rohlicek, S. Roucos, R. Schwartz, and J. Vandegrift, "Continuous speech recognition results of the BYBLOS system on the DARPA 1000-word resource management database," Proc. ICASSP-88, pp. 291-294, May 1988.

[16] W. Chou, T. Matsuoka, B.-H. Juang, and C.-H. Lee, "An algorithm of high resolution and efficient multiple string hypothesization for continuous speech recognition using inter-word models," Proc. ICASSP-94, vol. II, pp. 153-156, April 1994.

（平成８年５月２日受付，８月16日再受付）

松岡　達雄　（正員）

1982早大・理工・電子通信卒．1984同大大学院修士課程了．同年日本電信電話公社（現NTT）入社．横須賀電気通信研究所においてディジタル電話端末の研究開発に従事．1987よりNTTヒューマンインタフェース研究所において統計的手法をベースとした音声認識の研究に従事．1992～1993AT&T ベル研究所（Murray Hill）において客員研究員として主に話者適応化，N-best探索法の研究に従事．現在，NTTヒューマンインタフェース研究所古井特別研究室主任研究員．日本音響学会，IEEE各会員．

大附　克年

1993早大・理工・電気工卒．1995～1996NTTヒューマンインタフェース研究所古井特別研究室において実習生として大語い連続音声認識の研究に従事．1996同大大学院修士課程了．同年，日本電信電話株式会社入社．現在，NTTヒューマンインタフェース研究所勤務．日本音響学会会員．

森　　岳至

1994東工大・情報工卒．1995～1996NTTヒューマンインタフェース研究所古井特別研究室において実習生として大語い連続音声認識の研究に従事．1996同大大学院修士課程了．同年日本電信電話（株）入社．現在，音声信号符合化の研究に従事．

古井　貞熙　（正員）

1968東大・工・計数卒．1970同大大学院修士課程了．同年NTT電気通信研究所入社．以後，同研究所において，音声認識，話者認識，音声知覚などの研究に従事．1978～1979ベル研究所客員研究員．現在NTTヒューマンインタフェース研究所古井特別研究室長．東京工業大学客員教授．工博．1975米沢賞，1988，1993論文賞，1989科学技術庁長官賞，IEEE ASSP Society Senior Award，1990著述賞など受賞．著書「ディジタル音声処理」，「Digital Speech Processing, Synthesis, and Recognition」，「音響・音声工学」，編著「Advances in Speech Signal Processing」など．IEEE Fellow.

白井　克彦　（正員）

1963早大・理工・電気工卒．1968同大大学院博士課程了．工博．同年，同大学理工学部講師．1970同助教授．1975同教授（電気工学科）．1982～1990同大学情報システムセンター所長．1991同教授（情報学科）．音声認識・合成技術，自然言語処理，信号処理向けアーキテクチャ設計，CAI等を中心にヒューマンインタフェースの研究に従事．情報処理学会，日本音響学会，IEEE等各会員．