

TR-S-0003

発話スタイルの異なる音声の認識と誤りの分析
Speech Recognition Error Analysis between
Spontaneous and Read Speech

平野 泉
Izumi Hirano

中嶋 秀治
Hideharu Nakajima

匂坂 芳典
Yoshinori Sagisaka

2000.9.14

現在の大語彙連続音声認識の研究対象は、ディクテーション技術の進展に伴い、読み上げ音声(または朗読発話音声、read speech)から自然で自由な発話(自由発話音声、spontaneous speech)へと移行しつつある。自由発話の音声認識は朗読発話と比べて難しいと言われており、自由発話での認識率向上のためには、朗読発話との差異を明らかにする必要がある。本研究では、自由か朗読かという発話スタイルの違いに着目し、両方の音声の認識実験結果の誤りの分析を通して自由発話と読み上げ発話との間の音声言語的な違いについて考察した。

©ATR 音声言語通信研究所 2000

©2000 by ATR Spoken Language Translation Research Laboratories

目次

1	はじめに	1
2	音声データ概要	2
	2.1 自由発話音声データ	2
	2.2 朗読発話音声データ	3
3	単語認識実験条件	4
	3.1 音響モデル	4
	3.2 言語モデル	4
	3.3 評価データ	5
4	単語認識実験結果	6
5	分析結果	7
6	まとめ	9
	参考文献	11
	付録 A データの収録環境	12
	A.1 自由発話音声収録について	12
	A.2 朗読発話音声収録について	14
	付録 B コンフィギュレーションファイル	17
	付録 C 対話ごとの実験結果	19
	C.1 自由発話音響モデルを使った場合	19
	(C.1.1) 自由発話スタイル	19
	(C.1.2) 朗読発話スタイル	20
	(C.1.3) 演技発話スタイル	21
	C.2 朗読発話音響モデルを使った場合	22
	(C.2.1) 自由発話スタイル	22
	(C.2.2) 朗読発話スタイル	23
	(C.2.3) 演技発話スタイル	24

1 はじめに

音声を入力手段とした音声対話システムや翻訳システムの実現のために、自由発話音声認識の性能向上が求められている。しかし、自由発話の音声認識は読み上げ発話と比べて難しいと言われている。その要因のうち、最も有力な要因と指摘されているのが、発話スタイルである [1]。

文献 [1] では、発声の内容が同じであっても、発話スタイルの違いにより単語認識率が大きく異なることを示し、発話スタイルが大語彙連続音声認識システムの性能を決定する重要な要因であると述べている。しかし、この文献では発話スタイルによってもたらされる具体的な差異を特定するまでには至っていない。

そこで本研究では、発話スタイルの違いによって認識結果に差をもたらす要因を追求することを目的とし、発話スタイルの異なる音声の認識実験を行い、その結果を比較分析した。

2 音声データ概要

単語認識実験を行うために用意した音声データについての説明を行う。まず、自由発話音声データを収録し、次にその書き起こし文を使って読み上げ音声データの収録を行った。

2.1 自由発話音声データ

会話は、ホテルスタッフと顧客との間での日本語による対話を想定したものである。パソコンの画面に相手の顔がほぼリアルタイムで映し出されるテレビ会議システムを使って行われた遠隔対面対話の音声収録された。

双方が電話で話すという状況設定の下で、表1に示すような簡単な状況説明と共に単語が1つ与えられ、被験者はそこから会話の状況や内容をイメージした。顧客側が与えられた単語から状況をイメージして質問や依頼をすることから会話がはじまり、ホテルスタッフがそれに対応を行ないながら進められた。

被験者は、ホテルスタッフ側1名、顧客側1名(共に女性)であった。被験者に与えられた単語の総数は27で、与えられた単語ごとに会話を区切って数えて、収録されたデータは全27対話、総会話時間にして、36分12秒である。

表 1: 収録に与えた状況説明と単語

被験者に与えられた状況の説明	
あなた(顧客役の被験者の方)は、海外のある大都市に、友人と二人で訪れています。宿泊中のホテルの部屋から、フロントに電話をする状況を想定して、話して下さい。	
与えられた単語	
テレビ、隣室、エアコン、タクシー、買い物、ルームキー、中華料理	
ミュージカル、夜景、チェックアウト、ライブハウス、郵便	

収録の後、以上の発話の書き起こし文が作成された。書き起こし文の作成では、間投詞の箇所、および、言い直しを含む言い誤り箇所にマーキングが行なわれた。

今回収録された自由発話データに見られる間投詞や、言い直しを含む言い誤りの出現頻度を表2に示す。本データでは、言い誤りの中に言い直しが含まれている。そのため、ここでの集計では、一発話中に間投詞が、または、言い直しを含む言い誤りが出現する発話の数を数えた。また、ここでの一発話は書き起こし文での一行とした。つまり、ここでの一発話は、発話権を保持している区間と、相槌だけの区間とにほぼ対応する。

表 2: 間投詞と言い誤りの出現頻度

話者名	総発話数	間投詞	言い誤り
Guest	219	128 (58.45%)	28 (12.79%)
Hotel	181	61 (33.70%)	28 (15.47%)

この結果から、話者や役割に相違はあるが、間投詞が全発話の3から5割程度に出現し、言い直しを含む言い誤りが全発話の1割程みられた。

ATRの言語データベース中の国際会議の問い合わせを想定して収録されたデータを使った自由発話の研究[2]には、間投詞が文書全体の40%から60%の文に出現し、また、言い直しが文書全体の2%から4%の文に出現すると報告されており、本研究の自由発話データとの間に違いがみられた。

2.2 朗読発話音声データ

前節で作成された書き起こし文を同じ話者が読み上げることにより、朗読発話の音声データを収録した。

読み上げ音声の収録には、前節の自由発話の全ての書き起こし文の中から、

- 言語モデルの設計上認識することができない言い誤りを含む発話に対応する書き起こし文
- ‘はい’や‘そうですか’だけからなる頻出する発話に対応する書き起こし文

を除いた残りの書き起こし文を利用した。その結果、使用したデータは顧客側85発話、ホテルスタッフ側55発話である。

また、データ収録時には書き起こし文の読み方を被験者に教示した。これは、発声する文があらかじめ決まっても、その文が話されたときの状況を意識したり、発声に感情を込めることによって、読み上げ方にも違いが生じると考えたためである。そのために、今回は以下の二通りの読み方を指定し音声の収録を行った。すなわち、

1. 文脈の影響を排除するため、順序をランダムに並べて変えて書き起こし文を提示し、感情を込めず、淡々とした口調で読み上げる場合(以下、朗読発話)
2. 書き起こし文とともに、その文が発話された会話の状況の説明を提示して、相手に話しかけるように読み上げる場合(以下、演技発話)

の2通りのスタイルで音声を収録した。被験者に与えた読み方の指示を図1に示す。

声に出して読み上げる前に一度文章に目を通し、心の中で一度声に出さずに読んでください。

1. 会話の状況を意識せずに、感情を込めず、淡々と文章を読み上げてください。
2. 会話の状況を意識しながら、相手に話かけるように、文章を読み上げてください。

図1: 被験者に与えた読み方の教示

以上の結果、発声された内容が同じである異なる3つのスタイルの音声を用意した。以下では、それぞれの音声の認識実験を行ない、その結果の比較を行なうことを通じて、発話スタイルの違いによってもたらされるこれらの音声間の相違の分析を行なう。最初に、次章からの単語レベルでの認識実験を行なう。

3 単語認識実験条件

ここでは、単語認識実験で使う諸モデルと評価データについて述べる。

3.1 音響モデル

認識実験には、二種類の音響モデルを使用した。一つは、旅行対話データベース (/DB/SDB/TRA[1234]) 中のデータを使って学習された自然発話音響モデル [3] であり、もう一つは、上記の旅行対話データベース中の音素バランス文 (/DB/SDB/TRA[1234].BLA1) を使って学習された朗読発話音響モデルである。それぞれの、学習条件を表 3 に示す。

表 3: 音響モデルの学習条件

自然発話音響モデル	
	/RR/Recognition/ResearchJ/V9/amodel/BaseLine/AM.F.15.1400.bin
音響分析	サンプリング周波数 16kHz,preemphasis 0.98 フレーム周期 10ms, フレーム長 20ms(Hamming 窓) $\Delta \logpower + MFCC(12) + \Delta MFCC(12)$ ケプストラム平均、パワーを正規化
学習データ	旅行対話データベース 女性 240 話者:音声 1400 状態 5 混合, 無音 3 状態 10 混合
朗読発話音響モデル	
	/RR/Recognition/MatrixJ/amodel/19990602/Adapt/GD/AM.F.1000.3.bin
音響分析	サンプリング周波数 16kHz,preemphasis 0.98 フレーム周期 10ms, フレーム長 20ms(Hamming 窓) $\logpower + \Delta \logpower + MFCC(12) + \Delta MFCC(12)$ ケプストラム平均、パワーを正規化
学習データ	旅行対話データベース中の音素バランス文 女性 235 話者:音声 1000 状態 3 混合, 無音 3 状態 10 混合

3.2 言語モデル

言語モデルには、多重クラス複合 N-gram[4] を利用した。単語辞書には、今回の評価データだけに含まれていた 33 語を含めた 13,732 単語が登録されている。今回の評価データだけに含まれていた単語は、品詞や品詞の持つ種類に基づいて、クラスに追加した。言語モデルの学習条件を表 4 に示す。

表 4: 言語モデルの学習条件

単語辞書	13,699+33(新出語)=13,732
言語モデル	多重クラス複合 N-gram クラス数:from-クラス 700,to-クラス 700
学習データ	旅行に関する対話 7099 片対話

3.3 評価データ

比較のため、収録した二通りの朗読発話音声データすべてと、自由発話音声データから朗読発話音声データと対応するデータを使った。このように書き起こし文が同一のものを使うことによって、言語上の条件を統一し、発話スタイルの違いがもたらす要因をとらえやすくした。

総発話数は140発話であり、総単語数は3,022単語である。また、この評価データに対する Perplexity は62.64であった。

4 単語認識実験結果

前章で述べた実験条件で単語認識実験を行い、発話スタイルの違いによる単語正解率 (word correct) および単語認識精度 (word accuracy) を評価した結果を表5に示す。また、対話毎の認識率を付録Cに示す¹。

表 5: 発話スタイルの違いによる単語認識率 (acc/cor[%])

発話スタイル	読み上げ		自由発話
	朗読発話	演技発話	自由発話
自然発話音響モデル	77.00/80.96	77.69/80.92	53.44/55.55
朗読発話音響モデル	84.69/87.50	84.08/87.46	53.73/62.35

この結果から以下が読み取られる。

- (1) 読み上げのスタイルを朗読と演技の2つに分けたが、違いがみられなかった。
- (2) 読み上げ音声の認識率は、朗読発話音響モデルの場合の方が高かった。
- (3) 自由発話音声の認識率は、両音響モデルの間に差が見られなかった。
- (4) 自由発話は、朗読発話および演技発話と比較して、単語正解率、単語正解精度共に低下する。

以下、上の結果のそれぞれに対して考察を行なう。

読み上げ音声は、あらかじめ決められた文を同じ話者が発話して収録されたが、その文を発話するのにも、話者が発話の状況を意識したり、感情をこめたりするかないかで違いがみられると思われ、二種類に分けて収録した。しかし、認識率からは違いをみることができなかった(1)。これは、話者が発声毎に状況を意識する度合いが一貫していたかどうかは明確でないことに起因すると考えられる。実際に収録音声を聞き比べてみたが、聴感上では明らかな違いをみることができなかった。そのために認識率にも差が生じなかったのではないかと考えられる。

読み上げ音声と朗読発話音響モデルの学習に使われたデータとの間のミスマッチの度合いは小さいと考えられる(2)。

自由発話音声と自由発話音響モデルの学習に使われたデータとの間に、平均的な意味での、あるいは局所的なミスマッチがあると予想される(3)。

また、自由発話音声と読み上げ音声の認識率の差が発話スタイルの違いによって自由発話の音声認識を困難にしている要因があるのではないかと考え(4)、次章では、認識結果から得られた音素系列をもとに誤りの分析を行った。

¹自由発話音響モデルを使って自由発話スタイルの音声の認識を行なった場合に、C18.2.1.1の会話については認識結果を得ることができなかった。しかし、この会話を除いて評価しても全体の認識率に大きな差は生じない

5 分析結果

3種の音声認識実験結果のアライメントを行なった。そして、発話スタイルの違いによってもたらされた認識結果の差異を分析するために、読み上げ音声では正解として認識されているが、自由発話では誤認識となった部分に注目した。自由発話の認識結果の中には、ある単語の誤認識からしばらく連続して誤っている箇所が見られた。そのような箇所は原因の特定が困難であるため、比較分析も困難であるので、分析の対象からは除外した。したがって今回の分析対象は、自由発話音声だけに見られる誤り部分のうち、誤りを特定できそうな2,3単語以内の誤り部分とした。さらに、今回は朗読発話音響モデルを使って認識した結果を用いた。

誤認識の事例の分類結果を以下に示す。分析対象箇所に対して実際の音声データを試聴し、発声されていることを確認した後の結果である。

(1) 語の脱落

文頭、文末に限らず文のあらゆる箇所において、語の脱落がみられた。特に、終助詞、格助詞など1語からなる語は削除誤りとなる現象が多くみられた。これらの認識されなかった音について、音素継続時間長の結果を比較すると、母音の時間長が他と比べて短い傾向がみられた。

(2) 2重母音の脱落

ATRSPRECでの音素転記ファイルにおいて、長母音は母音を2つ重ねて表記される。たとえば、[...しましよ]は、/sh i i m a sh j o o/と表記される。しかし、この2つ重なった母音が1つしか認識されない現象がみられた。

(3) 母音、促音の脱落

語として脱落するものの他に一つの音素が脱落することにより、その前後の音素と共に他の語として認識されるといったものもあった。

(4) 子音間の置換

子音には、いくつか似たものがある。たとえば、鼻音の/m/と/n/などである。このように、似た音へ置換される現象がみられた。

(5) 子音の無声化

本来は有声音で発音されるべきところ、その音が無声化する現象がみられた。特に、/g/が/k/になる箇所がよくみられた。

これらの事例を表6に示す。また、これらの現象は重なって生じることもあった。

表 6: 誤認識の例

1. 脱落			
*助詞の脱落			
終助詞	[か][よ][ね]	*文中の脱落	
格助詞	[を][が][で][に]	-わから+ない	wakaranai
係助詞	[は]	家内	ka nai
連体助詞	[の]	-の+ほう	nohou
並立助詞	[と]	能	no u
		-変わり	kawari
		買い	ka i
2. 二重母音の脱落			
-オーケー	ookei	-し+たら+いい	shitaraii
を+消え	okie	し+たい	shita i
3. 母音、促音の脱落			
-そういう+ツアー+つていう+の+は		sou i u t u a a q t e i u n o w a	
そいつ+あて+の		so i t u a t e n o	
4. 子音間の誤り			
-隣	tonari	-料金	rjooking
泊り	tomari	病気	bjooki
-のみ	nomi	-わかり+まし+た	wakari
の+二	noni	計り	hakari
5. 子音の無声化			
-番号	banggou	-劇場	gekizhjou
観光	kangkou	形状	ke izhjou
			-正解単語列 [上段]
			誤認識単語列 [下段]

6 まとめ

本研究では、自然発話音声認識における問題点を明らかにすることを目的とし、発話スタイルの異なった音声データを使って、単語認識実験の結果から得られた音素系列の比較より、分析を行った。その結果、自由発話スタイルになると以下の現象が見られることがわかった。

- 脱落誤り：単語ごと、母音や促音
- 置換誤り：子音間での置換や無声化

今後さらに、自由発話による問題点を明らかにしていく上で、音響的な差異について調べる必要がある。具体的には、

- 音素継続時間長、発話速度、あるいはそれらの平均や分散といった時間情報による比較
- 音韻間距離や音素認識実験による、変形しやすい音素の特定
- 以上の分析結果をいかし、自由発話音声認識に適した音響モデルの構築や単語辞書作成

を行っていく。

また、自由発話音響モデルでの自由発話の認識率が読み上げ音声の認識率から大きく劣化している点から、認識対象である自由発話データが自由発話音響モデル学習に使われたデータ集合の外れ値に近い恐れがある。そのため、自由発話音響モデルの学習時にこの自由発話データを含めて学習し、クローズで実験を行ない、発話スタイルの違いによる相違を明確にしていきたい。

謝辞

本研究で利用した「自由発話スタイルの音声」は菅谷史昭氏が予備的にあつめたものを使わせて頂きました。また、音声収録には柴田葉子氏、森田千秋氏にご協力して頂きました。本研究を進めるにあたり、伴敏雄氏、林輝昭氏をはじめとする多くの方に協力して頂き、深く感謝致します。

参考文献

- [1] Mitch Weintraub, Kelsey Taussig, Kate Hunicke-Smith, Amy Snodgrass. Effect of Speaking Style on LVCSR Performance, ICSLP'96, pp16-19
- [2] 村上仁一, 嵯峨山茂樹. 自由発話音声における音響的な特徴の検討, 電子情報通信学会論文誌, pp1741-1759, 1995.12
- [3] 内藤正樹, Harald Singer, 山本博史, 中嶋秀治, 松井知子, 塚田元, 中村篤, 匂坂芳典. 旅行会話タスクにおける ATRSPREC の性能評価, 音講論, pp196-197, 1999.9
- [4] 山本博史, 匂坂芳典. 接続の方向性を考慮した多重クラス複合 N-gram 言語モデル, 信学技報, Vol.98, No.462, pp-49-54, 1998.12

付録 A データの収録環境

収録した音声データの詳細を示す。

A.1 自由発話音声収録について

・データ保存場所:

8月1日現在, 音声区間切り出し前のものを, UNIX上の以下に仮置している.

/home/pxn060/evaluate/TestWave/

000613morita.16k

000613sibata.16k

20000613.16k

これが, オリジナルで, 切り出したものは,

/home/ytsuyama/dialogue13June2000/kiridashi/wave_file/

・ENV

このデータは, ホテル対話を模擬した会話の予備収録に当たる。

発声形態は制約を与えない自然な形態である。

収録日時:2000年6月13日

目的: 現行言語モデルの補強を目指しての試行錯誤の中, ある音声収録の準備とテストを兼ねた収録。

総会話時間:36分12秒

有効会話時間:NIL

ノイズレベル:NIL<オフィスルーム>

発話形態:自由発話

ドメイン:ホテル対話

トピック:NIL

言語パターン:日本語-日本語

発声方法:NIL

対面・非対面:非対面

発呼者:柴田・葉子_女性_NIL_京都_NIL

被呼者:森田・千秋_女性_NIL_京都_NIL

発呼者発話ラベル:申込者

被呼者発話ラベル:担当者

コメント:

実験構成:

パソコンの画面に, ほぼリアルタイムで相手の顔が映し出される,

テレビ会議システムを使い, 離れた居室間での対話を行なった。

相互の通信には, windows98ベースのノートPCでnetmeetingを使用。

収録内容:

疑似ホテル対話(予約タスク以外)。プロットは, 或る簡単な状況説

明と共に単語を一つ与え, そこからの被験者のイメージで, ホテル

対話を求めた。

収録時にプロットの為に与えた単語:

あなた(=日本人客役の被験者の方)は, 海外のある大都市に, 友人

とふたりで訪れています。宿泊中のホテルの部屋から, フロントに

電話をする状況を想定して, 話してください。

1.(1)-7 「テレビ」)

2.(1)-8 「隣室」)

3.(1)-9 「エアコン」)

4.(1)-10 「タクシー」 その1)

5.(1)-10 「タクシー」 その2)

6.(1)-11 「買い物」)

7.(1)-12 「ルームキー」その1)

8.(1)-12 「ルームキー」その2)

9.(1)-13 「ミュージカル」)

10.(1)-14 「夜景」)

11. ((1) - 15 「禁煙」)
12. ((1) - 16 「盗難」 その1)
13. ((1) - 16 「盗難」 その2)
14. ((1) - 17 「郵便」)
15. ((1) - 18 「ライブハウス」)
16. ((1) - 19 「中華料理」)
17. ((1) - 20 「チェックアウト」)

あなたは、家族（夫婦および3歳と5歳の子供）で、海辺のあるリゾートに行くことを計画しています。予約を済ませたホテルに、日本から電話をしている状況を想定して、話してください。

18. ((2) - 1 「眺望」)
19. ((2) - 2 「割引」)
20. ((2) - 3 「送迎バス」)
21. ((2) - 4 「エクストラベッド」)
22. ((2) - 5 「チェックイン」)
23. ((2) - 6 「子ども」)
24. ((2) - 7 「日本語」)
25. ((2) - 8 「市内観光ツアー」)
26. ((2) - 9 「プール」)
27. ((2) - 10 「日本円」)

収録形態： パソコン TV 会議システムを使った遠隔対面対話形式.

収録場所： 申込者：客員室（旧 ITL 資産置き場）

担当者：実験室（工事前の新しい音声収録実験室）

ユーザインターフェイスマシン：

申込者側：pxn068 windows98 netmeeting
 担当者側：pxn067 windows98 netmeeting
 2台のマシン間は、所内共有の100MLANを接続
 ノートPC FUJITSU FMV-6500MF6/X
 CPU モバイルPIII 500MHz
 Memory 128MB
 LAN card 3Com 10/1000
 モバイルビデオカメラ付き

音響機器：

DigitalRecordingConsole YAMAHA 02R
 DigitalRecordingConsole YAMAHA 01v
 8 track digital recoder ALESIS adat xt
 AD/DA converter SONY SRP-DC5
 MIC Amp YAMAHA HA-8
 MIC SENNHEISER M-410

接続と記録方法：

```

                                     pxp068(netmeeting)
                                     /
  申込者側：MIC - 01v              /
                                     \
                                     02R - adat
                                     /
  担当者側：MIC - HA-8 - SRP-DC5  /
                                     \
                                     pxp067(netmeeting)
  
```

マルチトラックレコーダーの adaxt に、二人の話者の声と netmeeting の音を、メインミキサー 02R を経由して 4ch で同時収録。

ADコンバータ：01v 48KHz 20bits linear 128x oversampling

SRP-DC5 48KHz 20bits linear

ミキサー内部処理：48KHz 20bits

adat 内部処理: 48KHz 20bits --> 48KHz 18bits
 ミキサー 01v と 02R 間の接続には、ビデオ用同軸ケーブルを使用、
 民生規格の S/PDIF フォーマットのデジタル信号で通信。
 居室間が離れている為、同軸ケーブルは 10m 2本接続の 20m のものを
 使用。

コンピューターへの取り込み方法:

DATLink を使ってコンピューターへのデータ取り込みを行なった。
 DATLink には、ミキサー 02R を接続しており、adat の記録テープに
 一時保存したデータを再生 DATLink へ送っている。

データ取り込みコマンド (DATLink 添え付け)

```
narecord -q -o left -t ??? -p 16 -e linear -f raw -s 16000 outputfile
```

DATLink のバージョン

Host Software: Revision 2.28
 NetAudio Protocol: Revision 4.1
 Server Hardware: DAT-Link:T1055, Version 16053

環境整備について:

上記実験環境の準備時間:

実験環境の検討と場所の調査及び実験の収録について考察 (半日)

音響機器の準備

居室の整理, 実験空間の確保及び機材運び込みと設置 (半日)

機器間の接続と初期設定及び音だしテストとノイズ対策 (半日)

PC の準備

設置とネットワーク接続 (1 時間)

netmeeting の諸設定テスト (半日)

通信帯域不足の為か? ビデオのクオリティーと音声のクオリティー

の妥協点を探す調査とテスト。また、以前のパナソニックの

Let's note での不具合などのこともあり、夜間の長時間稼動

テストなども行なった。

音響機器とコンピューターの接続とテスト (半日)

音響機器と PC を接続するとコンピューターのノイズが音響機器に入って
 しまいその対策。

netmeeting の無音検出の調整と、ミキサー側のノイズゲートを使って、
 netmeeting の通信経由の聞き取り音声に、不快なノイズがのらないよ
 うに調整。

総合テスト及びリハーサル (1 日)

問題が出るたびに、元の段階に戻っての修正と調整、そのあと

再テストの繰り返し。

その他:

netmeeting の送信側と受信側での音質の差の調整 (2 時間)

netmeeting の設定は同じでも、Win98 側の設定の差で、音質に差が
 出てしまっていた。ネットワークのスピードダウンの原因調査と回
 避 (1 日)

netmeeting 使用テスト中に、突然通信スピードがダウンしてしまい、
 実験を行なえる状態ではなくなる。

何度か再現するまで動作させ、その時にネットワークチェック。

非常に時間効率の悪い作業となる。現在も、はっきりした原因は掴め
 ていないが、同一スイッチの同一グループのポート間での通信では、
 ほぼ起こらないようなので、それで回避対応している。

A.2 朗読発話音声収録について

・データ保存場所:

```
/home/pxp143/ban/atrcollect_20000613/Wave/  

  Cust_01/ Cust_02/ Hotel_01/ Hotel_02/
```


収録時の状態のままの仮置き

/data2/pxn035/ban/DecoderTest1/Wave/
01/ 02/
認識実験などに使ったもの

・ENV:

総会話時間:
有効会話時間: NIL
ノイズレベル: NIL <オフィスルーム>
発話形態: 読み上げ
ドメイン: ホテル対話 (2000年6月13日収録の書きお越しを使って)
トピック: NIL
言語パターン: 日本語
発声方法: NIL
対面・非対面: 非対面
発呼者: 柴田・葉子__女性__NIL__京都__NIL
被呼者: 森田・千秋__女性__NIL__京都__NIL
発呼者発話ラベル: 申込者
被呼者発話ラベル: 担当者
収録内容: 6月13日収録の時と同じ被験者で, その時のままの内容の読み上げ.
収録場所: 打ち合わせ室3 (旧ITL資産置き場)
収録日時: 2000年8月24日29日
実験構成: ノートPCを使いコンピューターへ直接音声収録を行なった.
収録形態: ATRcollectを使い, コンピューター画面に對面した形.
ユーザインターフェイスマシン:
pxn068 windows98 DATLink naserver, Xvision
X11アプリケーションの表示
サーバーマシン:
pxp143 Linux ATRcollect
X11を使っている画面をインターフェイスマシンに表示

音響機器接続と記録方法:

24日・森田千秋さん

MIC : SENNHEISER M-410
MIC Amp : YAMAHA HA-8
AD converter : SONY SRP-DC5 48KHz 20bits linear
DD converter : TOWNSHEND DATLINK+ down sampling
PC : pxn068 16KHz 16bits linear raw data

29日・柴田葉子さん

MIC : SENNHEISER M-410
AD converter : YAMAHA 01v 48KHz 20bits linear 128xoversampling
digital console : YAMAHA 01v
DD converter : TOWNSHEND DATLINK+ down sampling
PC : pxn068 16KHz 16bits linear raw data

DATLink+データ受け取り: 48KHz 20bits S/PDIF フォーマット

コンピューターへの取り込み:

DATLinkを使ってコンピューターへのデータ取り込みを行なった.
DATLinkには, 音響機器より同軸ケーブルで民生規格のS/PDIFで,
音声信号を受け取った.

データ取り込みコマンド (DATLink 添え付けのもの)

```
narecord -q -o left -p 16 -e linear -f raw -s 16000
outputfile
(ATRcollect 内部で呼び出している)
```

DATLink のバージョン

Host Software: Revision 2.28

NetAudio Protocol: Revision 4.1

Server Hardware: DAT-Link:T1055, Version 16053

付録 B コンフィギュレーションファイル

単語認識実験の結果を求めるために使用したコンフィギュレーションファイルを以下に示す。
音響モデルの違いにより、以下の項目を変更する。

* 自由発話音響モデル使用時

```
-ATRwave2cep:MeanInFile=/RR/Recognition/ResearchJ/amodel/19991111/CM/CM.TRS.TRA407.mean
-ATRcep2para:OutputParameter=cep(12)+dpow+dcep(12)
-ATRLattice:amname=/RR/Recognition/ResearchJ/amodel/19991111/BaseLine/AM.F.15.1400.bin
-ATRLattice:dimension=25
```

* 朗読発話音響モデル使用時

```
-ATRwave2cep:MeanInFile=/RR/Recognition/MatrixJ/MatrixJ/amodel/19990917/CM/CM.TRS.MDB11.mean
-ATRcep2para:OutputParameter=pow+cep(12)+dpow+dcep(12)
-ATRLattice:amname=/RR/Recognition/MatrixJ/MatrixJ/amodel/19990917/CM/AM.F.1000.3.bin
-ATRLattice:dimension=26
```

* 単語辞書と使用言語モデル

```
-ATRLattice:lexicon=/home/pxn112/xihirano/recg/MatrixJV5/NEW.LEX.W.comp
-ATRLattice:ngram=/home/pxn112/xihirano/recg/MatrixJV5/Lang.Model.700.bin
```

#ATRI/Ocontrol config :

```
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=NULL
I/Ocontrol:inputEOFexit=OFF
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NULL
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:rpcNumber=5
```

#ATRwave2cep config :

```
ATRwave2cep:CutoffHighFrequency=8000.0
ATRwave2cep:CutoffLowFrequency=0.0
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:inputParameter=waveRaw
ATRwave2cep:AnalysisType=fft
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:SamplingFrequency=16000
ATRwave2cep:FrameShift=10
ATRwave2cep:FrameLength=20
ATRwave2cep:Preemphasis=0.98
ATRwave2cep:DebuggingLevel=0
```

CMS

```
ATRwave2cep:Subtract=logpow+cep
ATRwave2cep:MeanInFile=
```

#ATRcep2para config :

```
ATRcep2para:OutputParameter=
ATRcep2para:rho=1.0
ATRcep2para:DDCepstrumPadding=zero
```

```
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:CepstrumOrder=12
ATRcep2para:DebuggingLevel=10
```

```
#ATRresult config :
```

```
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:Lattice_out=/dev/null
ATRresult:merge_list=../MatrixJV5/merge2.list
```

```
#ATRLattice config :
```

```
ATRLattice:lexicon=
ATRLattice:ngram=
```

```
## amname ##
```

```
ATRLattice:active_amname=all
ATRLattice:amname=
ATRLattice:lmscale=7.000000,12.000000
```

```
## beam ##
```

```
ATRLattice:beam=100.0,100.0
```

```
# for amodel RJ
```

```
ATRLattice:wdpenalty=10000,10000
ATRLattice:work_area=7000,200
ATRLattice:frame_shift=10
ATRLattice:pause_symbol=-
ATRLattice:dimension=
ATRLattice:phone_boundary=OFF # recog test 11/30
ATRLattice:word_merge=all
ATRLattice:UTT_START=5
ATRLattice:UTT_END=6
ATRLattice:backward_frame=-1
ATRLattice:amscale=1.000000
ATRLattice:FSA=
ATRLattice:null_trans=OFF
ATRLattice:UTT_END_delay=70
ATRLattice:word_boundary_skip=2
ATRLattice:state_skip=ON,75000
```

```
# r05r05 以降
```

```
# add by yama (edit by taka)
```

```
#EOF
```

付録 C 対話ごとの実験結果

表5で示した単語認識実験結果を対話毎に示す。

C.1 自由発話音響モデルを使った場合

-ATRLattice:lmscale = 7,12,-ATRLattice:beam = 110

(C.1.1) 自由発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	46.67/ 46.67	68.89/ 68.89	45	0	15	9	4.39	0	0
C11_1_15_1	100.00/100.00	100.00/100.00	6	0	0	0	2.62	0	0
C12_1_16_1	72.34/ 72.34	82.98/ 82.98	47	0	6	7	3.09	0	0
C13_1_16_2	41.74/ 41.74	65.22/ 65.22	115	0	22	45	4.04	0	0
C14_1_17_1	46.88/ 50.00	64.06/ 65.62	64	2	15	17	3.24	0	0
C15_1_18_1	50.00/ 50.00	70.69/ 70.69	58	0	14	15	5.06	0	0
C16_1_19_1	57.53/ 58.90	76.71/ 76.71	73	1	18	12	2.57	0	0
C17_1_20_1	71.43/ 76.19	83.33/ 85.71	42	2	4	6	2.41	0	0
C19_2_2_1	55.56/ 55.56	77.78/ 77.78	18	0	6	2	2.90	0	0
C1_1_7_1	60.26/ 65.38	75.64/ 76.92	78	4	14	13	3.91	0	0
C20_2_3_1	56.72/ 56.72	82.09/ 82.09	67	0	13	16	3.01	0	0
C21_2_4_1	62.86/ 71.43	77.14/ 80.00	35	3	3	7	4.77	0	0
C22_2_5_1	76.60/ 76.60	87.23/ 89.36	47	0	7	4	3.66	0	0
C23_2_6_1	55.26/ 55.26	68.42/ 68.42	38	0	13	4	3.78	0	0
C24_2_7_1	47.37/ 52.63	76.32/ 76.32	38	2	6	12	4.77	0	0
C25_2_8_1	60.78/ 60.78	74.51/ 76.47	51	0	7	13	3.23	0	0
C26_2_9_1	75.00/ 75.00	87.50/ 87.50	8	0	1	1	6.17	0	0
C27_2_10_1	59.38/ 59.38	71.88/ 71.88	32	0	7	6	8.19	0	0
C2_1_8_1	60.00/ 62.22	80.00/ 82.22	45	1	11	6	3.66	0	0
C3_1_9_1	52.38/ 52.38	64.29/ 66.67	42	0	9	11	2.85	0	0
C4_1_9_2	39.58/ 39.58	56.25/ 56.25	48	0	15	14	2.69	0	0
C5_1_10_1	62.50/ 62.50	70.83/ 70.83	24	0	6	3	4.96	0	0
C6_1_11_1	61.39/ 63.37	81.19/ 82.18	101	2	26	11	3.07	0	0
C7_1_12_1	46.67/ 48.33	60.00/ 60.00	60	1	18	13	3.32	0	0
C8_1_12_2	33.33/ 33.33	47.62/ 47.62	42	0	14	14	2.93	0	0
C9_1_13_1	53.03/ 53.03	66.67/ 69.70	66	0	20	11	4.28	0	0
H10_1_14_1	46.43/ 50.00	71.43/ 75.00	56	2	12	16	6.69	0	0
H11_1_15_1	40.00/ 40.00	62.50/ 62.50	40	0	10	14	5.75	0	0
H12_1_16_1	82.86/ 82.86	91.43/ 91.43	35	0	3	3	2.46	0	0
H13_1_16_2	47.06/ 50.98	66.67/ 68.63	102	4	18	32	4.61	0	0
H14_1_17_1	53.12/ 53.12	78.12/ 78.12	32	0	8	7	3.91	0	0
H15_1_18_1	58.59/ 59.60	78.79/ 78.79	99	1	12	28	4.67	0	0
H16_1_19_1	52.87/ 54.02	73.56/ 74.71	87	1	26	14	4.78	0	0
H17_1_20_1	41.67/ 41.67	66.67/ 66.67	24	0	7	7	3.37	0	0
H18_2_1_1	53.33/ 56.19	74.29/ 77.14	105	3	17	29	5.29	0	0
H1_1_7_1	39.29/ 41.07	51.79/ 58.93	56	1	16	17	4.12	0	0
H22_2_5_1	55.56/ 58.73	80.95/ 80.95	63	2	9	17	5.10	0	0
H23_2_6_1	43.24/ 43.24	64.86/ 64.86	37	0	10	11	3.58	0	0
H24_2_7_1	18.52/ 25.93	44.44/ 48.15	27	2	5	15	3.20	0	0
H25_2_8_1	72.00/ 72.00	89.33/ 90.67	75	0	7	14	2.41	0	0
H26_2_9_1	45.83/ 58.33	75.00/ 75.00	24	3	3	7	3.71	0	0
H27_2_10_1	66.67/ 66.67	83.33/ 83.33	18	0	2	4	3.29	0	0
H2_1_8_1	36.84/ 42.11	68.42/ 73.68	38	2	0	22	4.49	0	0
H5_1_10_1	71.91/ 73.03	79.78/ 79.78	89	1	12	12	5.93	0	0
H6_1_11_1	42.72/ 52.43	62.14/ 67.96	103	10	15	34	3.59	0	0
H8_1_12_2	33.33/ 33.33	33.33/ 33.33	6	0	3	1	1.73	0	0
H9_1_13_1	44.95/ 47.71	68.81/ 73.39	109	3	10	47	5.15	0	0
Total	53.44/ 55.55	72.05/ 73.56	2515	53	495	623	4.03	0	0

(C.1.2) 朗読発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	75.56/ 82.22	93.33/ 95.56	45	3	4	4	2.04	0	0
C11_1_15_1	100.00/100.00	100.00/100.00	6	0	0	0	1.38	0	0
C12_1_16_1	85.11/ 85.11	89.36/ 89.36	47	0	2	5	1.50	0	0
C13_1_16_2	74.78/ 80.00	88.70/ 90.43	115	6	6	17	1.72	0	0
C14_1_17_1	78.12/ 84.38	90.62/ 92.19	64	4	2	8	1.66	0	0
C15_1_18_1	58.62/ 63.79	89.66/ 94.83	58	3	12	9	1.71	0	0
C16_1_19_1	83.56/ 84.93	94.52/ 95.89	73	1	4	7	1.36	0	0
C17_1_20_1	78.57/ 85.71	88.10/ 90.48	42	3	3	3	1.41	0	0
C18_2_1_1	80.00/ 83.53	92.94/ 94.12	85	3	3	11	1.72	0	0
C19_2_2_1	88.89/ 88.89	94.44/ 94.44	18	0	2	0	1.22	0	0
C1_1_7_1	83.33/ 87.18	93.59/ 93.59	78	3	0	10	1.66	0	0
C20_2_3_1	85.07/ 86.57	95.52/ 97.01	67	1	1	8	1.65	0	0
C21_2_4_1	68.57/ 80.00	88.57/ 88.57	35	4	2	5	2.22	0	0
C22_2_5_1	87.23/ 91.49	97.87/ 97.87	47	2	1	3	1.44	0	0
C23_2_6_1	76.32/ 81.58	89.47/ 92.11	38	2	1	6	1.69	0	0
C24_2_7_1	52.63/ 68.42	78.95/ 78.95	38	6	2	10	1.92	0	0
C25_2_8_1	88.24/ 90.20	98.04/ 98.04	51	1	1	4	1.89	0	0
C26_2_9_1	87.50/ 87.50	100.00/100.00	8	0	0	1	1.67	0	0
C27_2_10_1	81.25/ 87.50	100.00/100.00	32	2	0	4	1.64	0	0
C2_1_8_1	77.78/ 84.44	97.78/ 97.78	45	3	0	7	1.45	0	0
C3_1_9_1	88.10/ 90.48	95.24/ 95.24	42	1	1	3	1.45	0	0
C4_1_9_2	68.75/ 72.92	83.33/ 85.42	48	2	2	11	1.66	0	0
C5_1_10_1	95.83/ 95.83	100.00/100.00	24	0	1	0	1.96	0	0
C6_1_11_1	79.21/ 82.18	96.04/ 96.04	101	3	9	9	1.63	0	0
C7_1_12_1	81.67/ 83.33	93.33/ 93.33	60	1	3	7	1.54	0	0
C8_1_12_2	76.19/ 78.57	92.86/ 95.24	42	1	1	8	2.24	0	0
C9_1_13_1	77.27/ 81.82	89.39/ 92.42	66	3	4	8	2.08	0	0
H10_1_14_1	71.43/ 76.79	85.71/ 89.29	56	3	3	10	2.72	0	0
H11_1_15_1	82.50/ 82.50	92.50/ 95.00	40	0	2	5	2.05	0	0
H12_1_16_1	77.14/ 77.14	85.71/ 85.71	35	0	2	6	1.61	0	0
H13_1_16_2	75.49/ 76.47	86.27/ 89.22	102	1	6	18	1.89	0	0
H14_1_17_1	96.88/ 96.88	100.00/100.00	32	0	1	0	1.37	0	0
H15_1_18_1	70.71/ 73.74	85.86/ 88.89	99	3	5	21	2.19	0	0
H16_1_19_1	81.61/ 83.91	94.25/ 94.25	87	2	3	11	1.41	0	0
H17_1_20_1	79.17/ 79.17	95.83/ 95.83	24	0	1	4	1.22	0	0
H18_2_1_1	75.24/ 80.95	91.43/ 91.43	105	6	5	15	1.56	0	0
H1_1_7_1	73.21/ 75.00	89.29/ 91.07	56	1	6	8	2.40	0	0
H22_2_5_1	71.43/ 76.19	85.71/ 85.71	63	3	1	14	1.80	0	0
H23_2_6_1	72.97/ 81.08	86.49/ 91.89	37	3	2	5	1.44	0	0
H24_2_7_1	62.96/ 66.67	77.78/ 85.19	27	1	2	7	1.96	0	0
H25_2_8_1	82.67/ 82.67	92.00/ 93.33	75	0	6	7	1.70	0	0
H26_2_9_1	58.33/ 66.67	83.33/ 83.33	24	2	1	7	1.79	0	0
H27_2_10_1	66.67/ 72.22	83.33/ 88.89	18	1	1	4	1.64	0	0
H2_1_8_1	84.21/ 86.84	89.47/ 89.47	38	1	0	5	1.01	0	0
H5_1_10_1	86.52/ 91.01	94.38/ 96.63	89	4	4	4	2.05	0	0
H6_1_11_1	66.02/ 71.84	77.67/ 81.55	103	6	1	28	1.97	0	0
H8_1_12_2	83.33/ 83.33	100.00/100.00	6	0	1	0	0.82	0	0
H9_1_13_1	66.97/ 74.31	86.24/ 89.91	109	8	4	24	1.94	0	0
Total	77.00/ 80.96	90.42/ 92.04	2600	103	124	371	1.74	0	0

(C.1.3) 演技発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	68.89/ 75.56	91.11/ 93.33	45	3	3	8	3.06	0	0
C11_1_15_1	100.00/100.00	100.00/100.00	6	0	0	0	1.20	0	0
C12_1_16_1	72.34/ 72.34	78.72/ 78.72	47	0	4	9	1.59	0	0
C13_1_16_2	73.91/ 73.91	87.83/ 88.70	115	0	11	19	2.28	0	0
C14_1_17_1	76.56/ 81.25	90.62/ 90.62	64	3	1	11	2.23	0	0
C15_1_18_1	65.52/ 68.97	86.21/ 86.21	58	2	7	11	2.04	0	0
C16_1_19_1	84.93/ 86.30	95.89/ 95.89	73	1	5	5	1.41	0	0
C17_1_20_1	88.10/ 88.10	95.24/ 95.24	42	0	2	3	1.20	0	0
C18_2_1_1	72.94/ 74.12	92.94/ 94.12	85	1	13	9	2.13	0	0
C19_2_2_1	94.44/ 94.44	100.00/100.00	18	0	1	0	1.41	0	0
C1_1_7_1	79.49/ 84.62	93.59/ 93.59	78	4	1	11	2.10	0	0
C20_2_3_1	82.09/ 82.09	92.54/ 92.54	67	0	3	9	1.68	0	0
C21_2_4_1	82.86/ 88.57	91.43/ 94.29	35	2	1	3	2.33	0	0
C22_2_5_1	87.23/ 87.23	93.62/ 95.74	47	0	0	6	1.72	0	0
C23_2_6_1	68.42/ 76.32	92.11/ 97.37	38	3	5	4	1.68	0	0
C24_2_7_1	68.42/ 71.05	84.21/ 84.21	38	1	3	8	2.07	0	0
C25_2_8_1	72.55/ 74.51	94.12/ 94.12	51	1	5	8	2.10	0	0
C26_2_9_1	87.50/ 87.50	100.00/100.00	8	0	0	1	2.14	0	0
C27_2_10_1	81.25/ 87.50	96.88/ 96.88	32	2	0	4	1.93	0	0
C2_1_8_1	82.22/ 82.22	95.56/ 95.56	45	0	0	8	1.82	0	0
C3_1_9_1	80.95/ 80.95	92.86/ 92.86	42	0	3	5	1.98	0	0
C4_1_9_2	70.83/ 77.08	93.75/ 93.75	48	3	4	7	1.79	0	0
C5_1_10_1	87.50/ 87.50	95.83/ 95.83	24	0	2	1	2.66	0	0
C6_1_11_1	78.22/ 79.21	95.05/ 95.05	101	1	13	8	2.37	0	0
C7_1_12_1	81.67/ 83.33	96.67/ 98.33	60	1	4	6	2.45	0	0
C8_1_12_2	59.52/ 64.29	85.71/ 85.71	42	2	5	10	5.38	0	0
C9_1_13_1	59.09/ 62.12	80.30/ 81.82	66	2	8	17	3.98	0	0
H10_1_14_1	75.00/ 87.50	87.50/ 91.07	56	7	0	7	2.40	0	0
H11_1_15_1	92.50/ 92.50	95.00/ 95.00	40	0	1	2	1.84	0	0
H12_1_16_1	74.29/ 74.29	85.71/ 85.71	35	0	2	7	1.47	0	0
H13_1_16_2	81.37/ 84.31	91.18/ 92.16	102	3	6	10	1.79	0	0
H14_1_17_1	90.62/ 90.62	90.62/ 90.62	32	0	1	2	1.47	0	0
H15_1_18_1	74.75/ 76.77	87.88/ 89.90	99	2	4	19	2.20	0	0
H16_1_19_1	78.16/ 80.46	95.40/ 96.55	87	2	4	13	1.51	0	0
H17_1_20_1	75.00/ 79.17	87.50/ 87.50	24	1	1	4	1.26	0	0
H18_2_1_1	80.95/ 87.62	94.29/ 96.19	105	7	4	9	1.54	0	0
H1_1_7_1	69.64/ 69.64	85.71/ 87.50	56	0	5	12	1.79	0	0
H22_2_5_1	80.95/ 85.71	95.24/ 96.83	63	3	0	9	1.61	0	0
H23_2_6_1	89.19/ 89.19	89.19/ 94.59	37	0	1	3	1.34	0	0
H24_2_7_1	66.67/ 81.48	85.19/ 92.59	27	4	0	5	1.62	0	0
H25_2_8_1	85.33/ 88.00	89.33/ 92.00	75	2	3	6	1.43	0	0
H26_2_9_1	66.67/ 75.00	83.33/ 87.50	24	2	0	6	1.52	0	0
H27_2_10_1	83.33/ 88.89	100.00/100.00	18	1	0	2	1.30	0	0
H2_1_8_1	89.47/ 92.11	92.11/ 94.74	38	1	0	3	1.17	0	0
H5_1_10_1	92.13/ 93.26	94.38/ 94.38	89	1	2	4	1.60	0	0
H6_1_11_1	68.93/ 79.61	85.44/ 91.26	103	11	3	18	1.59	0	0
H8_1_12_2	100.00/100.00	100.00/100.00	6	0	0	0	0.85	0	0
H9_1_13_1	74.31/ 78.90	90.83/ 92.66	109	5	1	22	1.82	0	0
Total	77.69/ 80.92	91.08/ 92.42	2600	84	142	354	1.97	0	0

C.2 朗読発話音響モデルを使った場合

-ATRlattice:lm scale = 8,10,-ATRlattice:beam = 100

(C.2.1) 自由発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	37.78/ 46.67	77.78/ 84.44	45	4	11	13	2.71	0	0
C11_1_15_1	33.33/ 83.33	66.67/ 83.33	6	3	0	1	1.84	0	0
C12_1_16_1	61.70/ 74.47	85.11/ 89.36	47	6	7	5	1.68	0	0
C13_1_16_2	49.57/ 56.52	71.30/ 74.78	115	8	16	34	2.50	0	0
C14_1_17_1	57.81/ 68.75	75.00/ 78.12	64	7	6	14	2.24	0	0
C15_1_18_1	53.45/ 62.07	86.21/ 89.66	58	5	4	18	2.13	0	0
C16_1_19_1	58.90/ 68.49	83.56/ 84.93	73	7	5	18	1.77	0	0
C17_1_20_1	57.14/ 64.29	83.33/ 85.71	42	3	7	8	2.42	0	0
C18_2_1_1	50.59/ 51.76	76.47/ 76.47	85	1	11	30	2.68	0	0
C19_2_2_1	61.11/ 61.11	88.89/ 88.89	18	0	4	3	2.04	0	0
C1_1_7_1	53.85/ 74.36	79.49/ 84.62	78	16	4	16	2.60	0	0
C20_2_3_1	50.75/ 59.70	80.60/ 82.09	67	6	6	21	2.35	0	0
C21_2_4_1	68.57/ 82.86	77.14/ 91.43	35	5	1	5	2.11	0	0
C22_2_5_1	80.85/ 87.23	93.62/ 95.74	47	3	2	4	1.88	0	0
C23_2_6_1	60.53/ 63.16	76.32/ 78.95	38	1	2	12	2.43	0	0
C24_2_7_1	39.47/ 71.05	73.68/ 92.11	38	12	5	6	2.56	0	0
C25_2_8_1	74.51/ 76.47	86.27/ 86.27	51	1	4	8	2.16	0	0
C26_2_9_1	87.50/ 87.50	100.00/100.00	8	0	0	1	2.48	0	0
C27_2_10_1	62.50/ 75.00	78.12/ 81.25	32	4	1	7	2.10	0	0
C2_1_8_1	60.00/ 73.33	95.56/ 95.56	45	6	3	9	1.67	0	0
C3_1_9_1	54.76/ 64.29	71.43/ 78.57	42	4	6	9	2.95	0	0
C4_1_9_2	41.67/ 52.08	72.92/ 75.00	48	5	3	20	3.65	0	0
C5_1_10_1	66.67/ 79.17	91.67/ 95.83	24	3	2	3	3.45	0	0
C6_1_11_1	67.33/ 75.25	85.15/ 89.11	101	8	7	18	2.30	0	0
C7_1_12_1	50.00/ 53.33	63.33/ 70.00	60	2	13	15	2.57	0	0
C8_1_12_2	40.48/ 45.24	54.76/ 64.29	42	2	2	21	2.69	0	0
C9_1_13_1	63.64/ 66.67	95.45/ 95.45	66	2	7	15	3.08	0	0
H10_1_14_1	23.21/ 55.36	66.07/ 71.43	56	18	2	23	4.78	0	0
H11_1_15_1	62.50/ 70.00	67.50/ 72.50	40	3	2	10	2.83	0	0
H12_1_16_1	45.71/ 48.57	77.14/ 82.86	35	1	5	13	2.57	0	0
H13_1_16_2	38.24/ 48.04	66.67/ 71.57	102	10	13	40	2.92	0	0
H14_1_17_1	31.25/ 31.25	46.88/ 50.00	32	0	8	14	2.16	0	0
H15_1_18_1	57.58/ 62.63	77.78/ 80.81	99	5	9	28	3.04	0	0
H16_1_19_1	58.62/ 63.22	74.71/ 75.86	87	4	9	23	2.88	0	0
H17_1_20_1	33.33/ 33.33	58.33/ 58.33	24	0	5	11	2.17	0	0
H18_2_1_1	39.05/ 45.71	68.57/ 70.48	105	7	9	48	3.34	0	0
H1_1_7_1	41.07/ 53.57	60.71/ 66.07	56	7	6	20	1.89	0	0
H22_2_5_1	66.67/ 68.25	84.13/ 84.13	63	1	4	16	2.01	0	0
H23_2_6_1	51.35/ 59.46	83.78/ 83.78	37	3	3	12	2.10	0	0
H24_2_7_1	37.04/ 48.15	62.96/ 70.37	27	3	4	10	3.37	0	0
H25_2_8_1	65.33/ 72.00	82.67/ 86.67	75	5	4	17	1.54	0	0
H26_2_9_1	70.83/ 75.00	75.00/ 79.17	24	1	1	5	1.46	0	0
H27_2_10_1	55.56/ 55.56	83.33/ 83.33	18	0	3	5	1.34	0	0
H2_1_8_1	44.74/ 60.53	71.05/ 76.32	38	6	1	14	2.05	0	0
H5_1_10_1	70.79/ 74.16	85.39/ 87.64	89	3	4	19	2.39	0	0
H6_1_11_1	53.40/ 62.14	74.76/ 79.61	103	9	8	31	2.05	0	0
H8_1_12_2	33.33/ 33.33	50.00/ 50.00	6	0	2	2	2.05	0	0
H9_1_13_1	47.71/ 60.55	66.97/ 75.23	109	14	6	37	5.11	0	0
Total	53.73/ 62.35	76.35/ 80.15	2600	224	247	732	2.60	0	0

(C.2.2) 朗読発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	95.56/ 97.78	100.00/100.00	45	1	1	0	0.86	0	0
C11_1_15_1	100.00/100.00	100.00/100.00	6	0	0	0	0.80	0	0
C12_1_16_1	82.98/ 87.23	93.62/ 93.62	47	2	2	4	1.12	0	0
C13_1_16_2	80.00/ 82.61	91.30/ 91.30	115	3	10	10	1.03	0	0
C14_1_17_1	89.06/ 92.19	93.75/ 93.75	64	2	2	3	0.98	0	0
C15_1_18_1	86.21/ 86.21	98.28/ 98.28	58	0	5	3	0.91	0	0
C16_1_19_1	86.30/ 89.04	98.63/ 98.63	73	2	4	4	0.84	0	0
C17_1_20_1	88.10/ 90.48	95.24/ 97.62	42	1	3	1	0.82	0	0
C18_2_1_1	92.94/ 92.94	96.47/ 96.47	85	0	4	2	1.05	0	0
C19_2_2_1	94.44/ 94.44	100.00/100.00	18	0	1	0	0.74	0	0
C1_1_7_1	91.03/ 92.31	94.87/ 94.87	78	1	1	5	0.94	0	0
C20_2_3_1	82.09/ 83.58	95.52/ 95.52	67	1	1	10	0.94	0	0
C21_2_4_1	85.71/ 88.57	94.29/ 94.29	35	1	1	3	1.03	0	0
C22_2_5_1	91.49/ 93.62	100.00/100.00	47	1	1	2	0.84	0	0
C23_2_6_1	84.21/ 92.11	100.00/100.00	38	3	1	2	0.95	0	0
C24_2_7_1	86.84/ 94.74	97.37/100.00	38	3	1	1	0.99	0	0
C25_2_8_1	86.27/ 88.24	96.08/ 96.08	51	1	1	5	1.00	0	0
C26_2_9_1	87.50/ 87.50	100.00/100.00	8	0	0	1	0.88	0	0
C27_2_10_1	93.75/ 93.75	100.00/100.00	32	0	0	2	0.80	0	0
C2_1_8_1	93.33/ 93.33	100.00/100.00	45	0	1	2	0.94	0	0
C3_1_9_1	90.48/ 92.86	95.24/ 95.24	42	1	1	2	0.99	0	0
C4_1_9_2	75.00/ 81.25	87.50/ 89.58	48	3	3	6	0.85	0	0
C5_1_10_1	87.50/ 87.50	100.00/100.00	24	0	2	1	1.08	0	0
C6_1_11_1	82.18/ 85.15	96.04/ 96.04	101	3	9	6	0.90	0	0
C7_1_12_1	90.00/ 90.00	96.67/ 96.67	60	0	2	4	1.01	0	0
C8_1_12_2	76.19/ 78.57	97.62/ 97.62	42	1	1	8	0.79	0	0
C9_1_13_1	80.30/ 83.33	98.48/ 98.48	66	2	6	5	0.97	0	0
H10_1_14_1	73.21/ 76.79	85.71/ 85.71	56	2	3	10	1.10	0	0
H11_1_15_1	90.00/ 95.00	92.50/ 95.00	40	2	0	2	0.88	0	0
H12_1_16_1	77.14/ 77.14	85.71/ 85.71	35	0	3	5	0.89	0	0
H13_1_16_2	84.31/ 87.25	93.14/ 94.12	102	3	3	10	1.03	0	0
H14_1_17_1	96.88/ 96.88	96.88/ 96.88	32	0	0	1	0.87	0	0
H15_1_18_1	76.77/ 83.84	87.88/ 92.93	99	7	2	14	1.33	0	0
H16_1_19_1	82.76/ 88.51	90.80/ 94.25	87	5	2	8	0.88	0	0
H17_1_20_1	79.17/ 79.17	91.67/ 91.67	24	0	1	4	0.94	0	0
H18_2_1_1	87.62/ 88.57	95.24/ 95.24	105	1	1	11	0.98	0	0
H1_1_7_1	83.93/ 83.93	91.07/ 94.64	56	0	3	6	0.96	0	0
H22_2_5_1	80.95/ 82.54	90.48/ 92.06	63	1	0	11	0.88	0	0
H23_2_6_1	81.08/ 89.19	100.00/100.00	37	3	1	3	0.71	0	0
H24_2_7_1	66.67/ 74.07	85.19/ 88.89	27	2	0	7	0.93	0	0
H25_2_8_1	74.67/ 80.00	89.33/ 90.67	75	4	5	10	1.19	0	0
H26_2_9_1	79.17/ 83.33	100.00/100.00	24	1	0	4	1.06	0	0
H27_2_10_1	72.22/ 77.78	94.44/ 94.44	18	1	1	3	0.89	0	0
H2_1_8_1	89.47/ 92.11	94.74/ 97.37	38	1	0	3	0.90	0	0
H5_1_10_1	94.38/ 95.51	96.63/ 98.88	89	1	0	4	1.06	0	0
H6_1_11_1	83.50/ 84.47	95.15/ 95.15	103	1	1	15	0.97	0	0
H8_1_12_2	83.33/ 83.33	83.33/ 83.33	6	0	0	1	0.86	0	0
H9_1_13_1	84.40/ 89.91	95.41/ 95.41	109	6	1	10	1.20	0	0
Total	84.69/ 87.50	94.50/ 95.31	2600	73	91	234	0.97	0	0

(C.2.3) 演技発話スタイル

	top acc/cor	network acc/cor	total	Ins	Del	Sub	cpu/ utt	err utt	skp utt
C10_1_14_1	91.11/ 93.33	100.00/100.00	45	1	1	2	1.12	0	0
C11_1_15_1	100.00/100.00	100.00/100.00	6	0	0	0	0.81	0	0
C12_1_16_1	87.23/ 89.36	91.49/ 93.62	47	1	2	3	1.08	0	0
C13_1_16_2	79.13/ 81.74	92.17/ 93.04	115	3	10	11	1.14	0	0
C14_1_17_1	87.50/ 87.50	92.19/ 92.19	64	0	2	6	1.19	0	0
C15_1_18_1	79.31/ 81.03	96.55/ 96.55	58	1	4	7	1.07	0	0
C16_1_19_1	89.04/ 90.41	97.26/ 97.26	73	1	4	3	0.85	0	0
C17_1_20_1	88.10/ 88.10	100.00/100.00	42	0	3	2	0.91	0	0
C18_2_1_1	80.00/ 85.88	95.29/ 96.47	85	5	5	7	1.15	0	0
C19_2_2_1	94.44/ 94.44	100.00/100.00	18	0	1	0	0.99	0	0
C1_1_7_1	85.90/ 87.18	96.15/ 96.15	78	1	2	8	1.30	0	0
C20_2_3_1	79.10/ 82.09	95.52/ 95.52	67	2	3	9	1.16	0	0
C21_2_4_1	77.14/ 85.71	94.29/ 97.14	35	3	1	4	1.06	0	0
C22_2_5_1	91.49/ 95.74	100.00/100.00	47	2	0	2	1.01	0	0
C23_2_6_1	84.21/ 92.11	100.00/100.00	38	3	1	2	0.89	0	0
C24_2_7_1	78.95/ 89.47	92.11/ 97.37	38	4	2	2	1.42	0	0
C25_2_8_1	88.24/ 90.20	98.04/ 98.04	51	1	2	3	1.23	0	0
C26_2_9_1	87.50/ 87.50	100.00/100.00	8	0	0	1	0.88	0	0
C27_2_10_1	90.62/ 90.62	100.00/100.00	32	0	0	3	0.95	0	0
C2_1_8_1	88.89/ 95.56	100.00/100.00	45	3	0	2	1.15	0	0
C3_1_9_1	78.57/ 80.95	92.86/ 92.86	42	1	1	7	1.04	0	0
C4_1_9_2	72.92/ 83.33	89.58/ 91.67	48	5	0	8	1.22	0	0
C5_1_10_1	91.67/ 91.67	100.00/100.00	24	0	2	0	1.39	0	0
C6_1_11_1	76.24/ 79.21	97.03/ 98.02	101	3	11	10	1.15	0	0
C7_1_12_1	86.67/ 88.33	96.67/ 98.33	60	1	1	6	1.10	0	0
C8_1_12_2	83.33/ 88.10	90.48/ 90.48	42	2	1	4	1.51	0	0
C9_1_13_1	83.33/ 87.88	96.97/ 96.97	66	3	4	4	1.47	0	0
H10_1_14_1	80.36/ 89.29	96.43/ 96.43	56	5	1	5	1.36	0	0
H11_1_15_1	92.50/ 95.00	92.50/ 95.00	40	1	0	2	0.96	0	0
H12_1_16_1	88.57/ 88.57	97.14/ 97.14	35	0	0	4	1.01	0	0
H13_1_16_2	81.37/ 85.29	92.16/ 95.10	102	4	4	11	1.51	0	0
H14_1_17_1	93.75/ 93.75	93.75/ 93.75	32	0	0	2	0.89	0	0
H15_1_18_1	77.78/ 80.81	92.93/ 92.93	99	3	3	16	1.32	0	0
H16_1_19_1	75.86/ 79.31	90.80/ 91.95	87	3	4	14	1.17	0	0
H17_1_20_1	83.33/ 83.33	91.67/ 91.67	24	0	2	2	1.01	0	0
H18_2_1_1	84.76/ 86.67	92.38/ 92.38	105	2	2	12	1.12	0	0
H1_1_7_1	82.14/ 87.50	94.64/ 96.43	56	3	1	6	1.08	0	0
H22_2_5_1	88.89/ 88.89	95.24/ 95.24	63	0	1	6	1.06	0	0
H23_2_6_1	91.89/ 94.59	100.00/100.00	37	1	0	2	0.76	0	0
H24_2_7_1	85.19/ 96.30	92.59/ 96.30	27	3	0	1	1.00	0	0
H25_2_8_1	84.00/ 86.67	98.67/ 98.67	75	2	2	8	1.03	0	0
H26_2_9_1	83.33/ 87.50	91.67/ 91.67	24	1	0	3	0.98	0	0
H27_2_10_1	94.44/ 94.44	100.00/100.00	18	0	0	1	0.83	0	0
H2_1_8_1	81.58/ 92.11	84.21/ 92.11	38	4	0	3	1.18	0	0
H5_1_10_1	89.89/ 92.13	96.63/ 96.63	89	2	2	5	1.13	0	0
H6_1_11_1	86.41/ 90.29	93.20/ 94.17	103	4	1	9	0.95	0	0
H8_1_12_2	100.00/100.00	100.00/100.00	6	0	0	0	0.61	0	0
H9_1_13_1	85.32/ 88.99	91.74/ 93.58	109	4	1	11	1.14	0	0
Total	84.08/ 87.46	94.85/ 95.69	2600	88	87	239	1.13	0	0

被験者：_____様

会話実験について

- 海外ホテルスタッフと日本人の顧客が電話でやり取りをする会話を文章にしたものを読み上げていただきます。
- 収録は2回おこなっていただきます。
- 発声時間は、1回20分ぐらいになります。
- 記号[]の中の文字も読み上げてください。

—[あ] ヒルトップクラブですね。はい。ビクトリアホテルでしたよね。

[あ] はい。[え]それは、そのホテルは、どこにあるんですか。

—どの辺に。

—[あ] そうですか。[あ] [あ] そうですか。[あ] いや、ヒルトンホテルは、
[その] どうやっていけばいいのかなと思って。[あの]

—[あ] はい。はい。わかりました。ありがとうございます。

読み上げ文章例

(1) 1回目の収録

声に出して読み上げる前に一度文章に目を通し、心の中で一度声に出さずに読んでください。

読み方の指定をさせていただきます。

会話の状況を意識せずに、感情を込めず、淡々と文章を読み上げてください。

(2) 2回目の収録

声に出して読み上げる前に一度文章に目を通し、心の中で一度声に出さずに読んでください。

読み方の指定をさせていただきます。

会話の状況を意識しながら、相手に話しかけるように、文章を読み上げてください。