Internal Use Only (非公開)

TR-IT-0032

自由発話音声認識における音響的および言語的な問題点の検討
A Discussion of Acoustic and Linguistic Problems in Spontaneous Speech Recognition

村上仁一
Jin'ichi Murakami

1993.12

概要

自由発話の音声認識は、今後の大きな課題であるが、従来の朗読発声の認識と比較すると、さまざまな困難が予想される。本論文では、自由発話と朗読発声の差を音声認識の立場から、音響的な特徴と言語的な特徴にわけて調べた。この結果、自由発話は朗読発話と比較すると、融合ラベルは約5割増加し、音素認識誤り率は約2倍増加していることが示された。また冗長語は、自由発話の文章全体の約4割、言い直しは約1割を占めていることが示された。

©ATR 音声翻訳通信研究所
©ATR Interpreting Telecommunications Research Labs.

自由発話音声における音響的・言語的な問題点の検討

A Discussion of Acoustic and Linguistic Problems in Spontaneous Speech Recognition

村上仁一
Jin'ichi Murakami
ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Labs.

概要

本論文では、音声認識の立場から自由発話の特性を調べるために、音響的な特徴と言語的な特徴の両面について、朗読発声と比較して調べた結果を述べる。
自由発話の音声認識は今後の大きな課題であるが、従来の朗読発声の認識と比較すると、さまざまな困難が予想される。したがって、まず初めに自由発話の持つ特徴を把握するための予備的な調査を行なった。音響的な特徴について調べたところ、発話速度は朗読発声でも自由発話でも大きな差が見られなかったが、視察によるラベリング作業において決定することができなかった音素境界(融合ラベル)は、朗読発声と比較すると、約2割増加した。そしてHMMによる音素認識実験を行なったところ、誤り率は約3割増加した。また、言語的な特徴について調べた結果、自由発話に特徴的な冗長語は自由発話の文章全体の約4割に、言い直しは約1割に現れた。

abstract

In this paper, we present a preliminary study of spontaneous speech recognition, describing both the acoustic and linguistic characteristics of spontaneous speech.
Recognition of spontaneous speech is one of the hardest problems in the speech recognition area, it seems for example to be much more difficult than read speech recognition. To get a better understanding of spontaneous speech, a preliminary study was done to compare spontaneous and read speech. In hand-labeled spontaneous speech, the labeling uncertainty increased by about 20% in number. A phoneme recognition experiment resulted in a factor of 30% increase in the error rate. Filled pauses appeared in 40% of 11,000 sentences of spontaneous speech utterances and false starts were found in 10% of the sentences.

1 まえがき

近年、連続音声認識の研究が盛んに行なわれ、いくつかの研究機関で文音声システムが構築されている[1],[2]。しかし、これらのシステムの多くは、朗読音声のような丁寧に発声された音声を入力対象にしている。しかし、人間同士のコミュニケーションでは、「あのー」、「えーと」などに代表される冗長語や、言い淀みや言い誤りおよび言い直しなどが頻繁に見受けられる。このような音声でも認識が可能な、いわゆる自由発話の音声認識が、今後の重要な研究課題になると思われる。

現在、自由発話の認識に関する報告としては、日本では自由発話に対してスポッティングの方法を適用して認識率を報告した例がある[3]。また、海外では、自由発話のデータベースを作成して自由発話の特徴を報告し、従来の音声認識で使用されたアルゴリズムを利用して、認識率を報告した例がある[4],[5]。

ここでは、自由発話の音声認識の第一歩として、自由発話の音声の特徴と言語的な特徴を調べた。そこで自由発話と従来の朗読発声の差を見るために、

表 1: 調査に用いた自由発話の音声データの収録条件

話者	ナレータ 2 名 (通称 MTK,FKN)
収録環境	遮音室
発話内容	国際会議の申し込みに関する参加者と事務局の対話 「トピック」(質問項目と、その背景に関する情報) や「バックグラウンド」(会話の前提になる背景) を詳細に設定して対話したもの。
入力系	マイクロフォン、DAT録音
データ量	26 対話 437 文 (MTK) 26 対話 569 文 (FKN)
音素数	約 8919 音素 (MTK) 約 14867 音素 (FKN)
発話様式	自由発話

表 2: 調査に用いた自由発話の言語データの収録条件

発話内容	国際会議の申し込みに関する参加者と事務局の対話
データ量	3178 対話、11054 文
発話様式	自由発話
発話環境	
1 通常の部屋	大部分が家庭用のカセットテープレコーダで録音。外来雑音も混在。
2 スタジオ録音 (遮音室)	DATで録音。明瞭。
話者	
1 事務局員役	当該分野の専門家
2 申し込み者役	ナレータ + 一般話者 (複数話者)

まず融合ラベルの付与率と発話速度と HMM による音素認識誤り率を調査した。次に言語的な面から、従来の朗読発声では出現しない言語現象である冗長語と言い直しの出現頻度を調べた。本論文では、これらの結果について報告する。

2 調査に用いたデータベース

現在ATRでは、各種言語現象を調査するために対話文を中心とする言語データベースの作成を進めているが [7]、この目的のために録音された音声は、同時に、音声データベースとして利用できる。本来は、自由発話音声の収録は、話者に録音していることを気づかれずに録音することが好ましいが、通信の守秘義務などの問題の他にも、話題が次々に移行するため、会話の語彙が膨大な数になるという問題も発生する。このため現実には、会話にある程度の制約を入れた模擬会話で収録している。ATR では、役割を与えられた 2 名の話者により行なわれた会話音声を録音収録し、後に文字化して言語データベースを作成している [6]。現在、発話内容で 5 種類、収録環境で 2 種類、話者で 2 種類、発話様式で 2 種類の variety を含むデータベースを収集中である [7]。

今回の調査に使用した音声データは、このデータベースのなかから、電話による国際会議の問い合わせに関する音声データを使用した。これの収録条件を表 1に示す。音声データは、遮音室でナレータが発声したもので、ドアの開閉音などの日常雑音や話者の舌打ちの音などは含まれていない。また両話者は、完全に分離されて録音されているため、音声区間の重畳はない。この意味で、この音声データは、自由発話音声としてはかなり clean な音声であると言ってよい。また、調査は 2 名の話者で行なった。ただし、両者の発話内容は異なっている。

一方、今回の調査に使用した言語データは、国際会議の問い合わせに関する音声データを文字化したもので、申し込み者役にはナレータの他に一般の話者も含まれているが、対応する事務局員役は、この分野の専門家が演じている。また収録は、遮音室の他に通常の部屋でも行なっている。これの収録条件を表 2に示す。

3 自由発話の音響的な特徴

ここでは、自由発話の音響的な特徴を知るために、融合ラベルの付与率と発話速度と HMM による音素認識誤り率を調査した。

3.1 ラベリング作業からみた自由発話

表 3に、音声データのラベリングの作業において見受けられた、自由発話の音素の定性的な特徴を挙げる。なお、ラベリングの基準は文献 [8] にしたがった。これらから、自由発話では音素境界がかなり曖昧になっていることや、従来の朗読発声には見られない音素が現れていることがわかる。

表 3: 自由発話の音素の定性的特徴

1	文の語尾の音素が不明瞭になる。 （例: 「なんですか」の「か」がほとんど聞こえない。）
2	母音 /a,i,u,e,o/ 全てが無声化する。 （朗読発声では /a,e,o/ は、あまり無声化しない。）
3	2重に解釈できる音素がある。 （例:「んー」（考え込むとき発声している音）は /N/ あるいは /uN/ の 両者に解釈できる。）
4	子音 /r/ をともなう音節の発音が全体的に弱い。 （例:「そうすると」の「る」がほとんど聞こえない。）
5	母音（特に、文末の母音『a』）の第1フォルマントが あらわれないことがある。 しかし、第1フォルマント以外の構造は明確で、 波形も母音の特徴を備えている。 （話者 MTK においてのみ）

3.2 融合ラベルの付与率から見た自由発話

ＡＴＲでは、発話テキストを参照しながら人手で音素境界を決定するラベリング作業において、音素境界が不明瞭な音素区間に対して付与するラベルのことを融合ラベルと呼んでいる。融合ラベルの例を図9に載せる。この例では、「はい、それで結構」の発話に対し、/o/,/r/,/e/ と /o/,/u/ に融合ラベルが付与されている。

この融合ラベルの付与率を、同一話者の4種類の発話様式において調査した。これを図1および表4に示す。また、各音素ごとの融合ラベルの付与率を表および表に示す。この結果からわかることを以下に示す。

1. 自由発話と文の朗読発声を比較すると、融合ラベルの付与率は話者 MTK では 33%(23.9% → 31.7%)、話者 FKN では 15% (23.0% → 26.4%)、増加する。

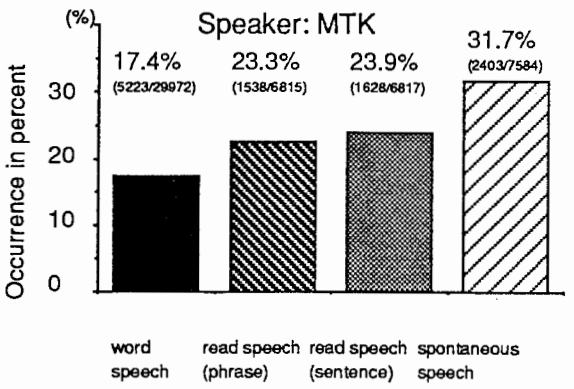
2. 音素別に自由発話と文の朗読発声を比較すると、母音では /a/ の増加が顕著である (MTK:4.0% → 13.3%, FKN:3.9% → 8.1%)。子音では、/m/ の増加が著しい (MTK:1.0% → 18.1%, FKN:1.6% → 10.8%)。

3. 自由発話では、全音素の約 1/4 以上が融合ラベルになる。

4. 単語発声・文節単位の朗読発声・文単位の朗読発声では融合ラベルの付与率に話者の相違は見られない。しかし、自由発話では話者の相違が見られる (MTK:31.7%, FKN:26.4%)。

5. 文節単位の朗読発声と文単位の朗読発声を比較すると、融合ラベルの付与率にあまり差がない。

6. 単語発声・文節単位の朗読発声・文単位の朗読発声・自由発話の順に融合ラベルの付与率が増加する。



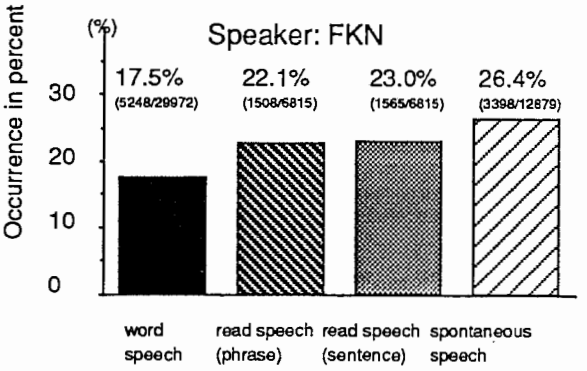


図 1: 発話様式の違いによる融合ラベルの付与率の変化

表 4: 発話様式の違いによる融合ラベルの付与率の変化

(融合ラベルを付与した音素数 / 調査音素数)

話者	MTK	FKN
単語発声	17.4% (5223/29972)	17.5% (5248/29972)
朗読発声 (文節)	22.3% (1538/6815)	22.1% (1508/6815)
朗読発声 (文)	23.9% (1628/6817)	23.0% (1565/6815)
自由発話	31.7% (2403/7584)	26.4% (3398/12879)

3.3 発話速度からみた自由発話音声

発話様式における発話速度の違いを調べるために、同一話者における4種類の発話様式（単語発声、文

節単位の朗読発声、文単位の朗読発声、自由発話）におけるモーラ速度の差を調査した。これを図2および表5に示す。ただし、調査の際、息つぎなどの長いポーズ区間は除去した。また融合ラベルを付与された音素は音素継続時間の計算から除いた。なお朗読発声の文節単位と文単位の発話内容は同一であるが、単語発声と自由発話の内容は朗読発声と異なる。

また、各話者の各発話様式における各音素の平均音素継続時間を表17および表18に示す。

この結果から以下のことが示される。

1. 自由発話の発話速度は、話者 MTK では文単位の朗読発声より早いが、話者 FKN では文単位の朗読発声より遅い。

2. 自由発話における母音の平均音素継続時間は朗読発声より長い。しかし子音の平均音素継続時間は朗読発声より短かい。また、自由発話の音素継続時間の分散は朗読発声より大きい。これは、自由発話では音素の音素継続時間に大きなバラツキがあることを意味する。

3. 発話速度は単語発声・文節単位の朗読発声・文単位の朗読発声の順に早くなる。

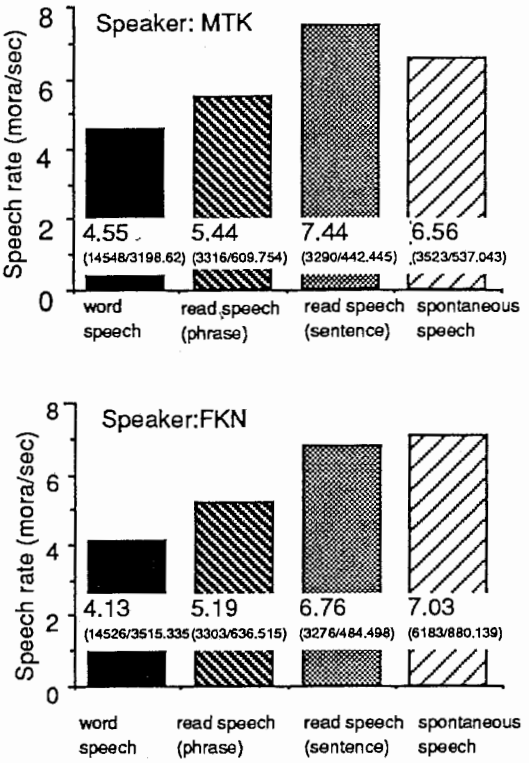


図 2: 発話様式の違いによる発話速度の変化

表 5: 発話様式の違いによる発話速度の変化
モーラ /sec (総モーラ数 / 発話時間)

話者	MTK	FKN
単語発声	4.55 (14548/3199)	4.13 (14526/3515)
朗読発声 (文節)	5.44 (3316/609.8)	5.93 (3315/559.0)
朗読発声 (文)	7.44 (3290/442.4)	6.76 (3276/484.5)
自由発話	6.56 (3523/537.0)	7.03 (6183/880.1)

3.4 音素認識誤り率から見た自由発話

ここでは音素認識率で各発話様式の差を調べた。認識アルゴリズムには混合連続分布型 HMM を用い、音素認識誤り率で評価した。ただし、融合ラベルを付与された音素は実験では用いなかった。

また学習データとして単語発声から視察によって切り出した音素を使用した場合と、同一発話様式の音声データから視察によって切り出した音素を使用した場合の、2種類の実験を行なった。

その他の実験条件を表6に示す。

表 6: 音素認識の実験条件

認識対象	32 音素
サンプリング周波数	12kHz
話者	男性のナレータ
学習データ	単語音声
音響パラメータ	log power + 16 次 LPCcepstrum
	+ Δlog power + 16 次 Δcepstrum
フレーム窓長	20ms
フレーム周期	5ms
音素モデル	4-state 3-loop 10 mixture
(単語学習)	Gaussian continuous HMM
音素モデル	4-state 3-loop 3 mixture
(同一発話)	Gaussian continuous HMM

学習データに単語発声を使用した場合の、各発声様式における音素認識誤り率を、図3および表7に、各音素毎の音素認識誤り率を表20および表?? に示す。

また、同一発話様式の音声データを2つにわけ、一方を学習データとし、一方をテストデータとして実験した場合の音素認識誤り率を、図4および表8に、各音素毎の音素認識誤り率を表?? および表22に示す。

これから次のような結果が示される。

1. 学習データが単語発話のとき、自由発話の音素認識誤り率は高い。朗読発声の音素認識誤り率と比較すると、ナレータ MTK は約160% 程度増加し (21.6% → 37.6%) ナレータ FKN では約240% も増加している (18.8% → 44.4%)。

2. 学習データに自由発話の音声を利用することにより、音素認識誤り率は大きく低下する (MTK:37.6% → 16.0%, FKN:44.4% → 15.0%)。学習データが単語発声のときの文の朗読発声の音素認識誤り率 (MTK:21.6%, FKN:18.8%) より低くなる。

3. 自由発話を学習データとした場合、母音の中では /u/ の認識誤り率が高い (MTK:43.9%, FKN:27.9%)。また、子音では /w/ の認識誤り率が高いが (MTK:78.9%, FKN:66.7%)、調査音素の数が少ないため、明確には言えない。

4. 単語発声、文節単位の朗読発声、文単位の朗読発声、自由発話の順に音素認識誤り率が増加する。

5. 学習データが同一発話様式の場合、各発話様式において話者の相違はあまり見られないが、学習データが単語発話のとき、話者の相違が見られる。

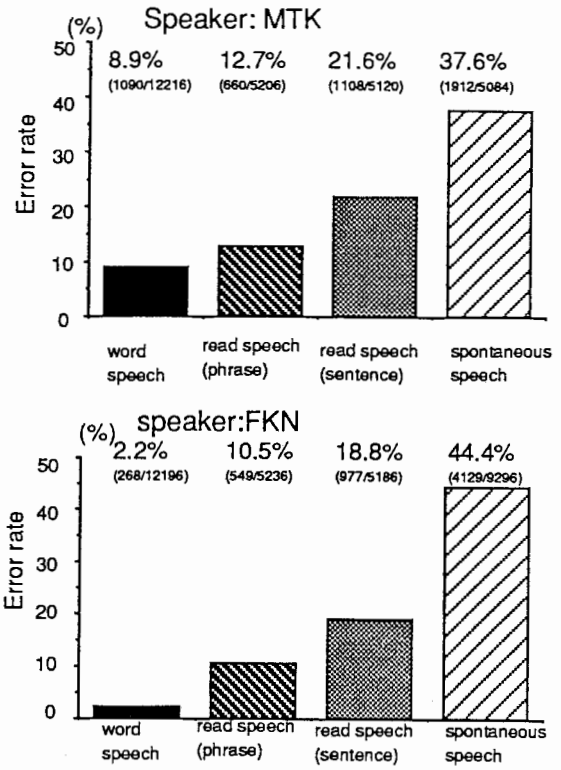


図3 音素認識誤り率 学習データ単語発話 (融合ラベルを除く)

Fig.3 Phony error rate (without compound labels)

図 3: 音素認識誤り率 (%) 学習データ・単語発声

表 7: 音素認識誤り率 (%) 学習データ・単語発声 (誤認識した音素数 / 調査音素数)

話者	MTK	FKN
単語発声	8.9% (1090/12216)	2.2% (268/12196)
朗読発声 (文節)	12.7% (660/5206)	10.5% (549/5236)
朗読発声 (文)	21.6% (1108/5120)	18.8% (977/5186)
自由発話	37.6% (1912/5084)	44.4% (4129/9296)

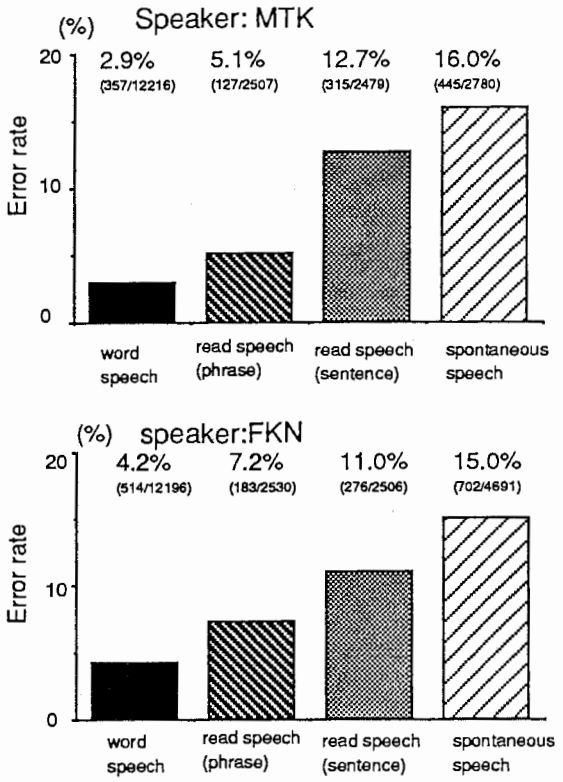


図　音素認識誤り率　学習データ同一発話様式
(融合ラベルを除く)

Fig. Phony error rate (without compound labels)

図 4: 音素認識誤り率 学習データ 同一発話様式

表 8: 音素認識誤り率 学習データ 同一発話様式
(誤認識した音素数 / 調査音素数)

話者	MTK	FKN
単語発声	2.9% (357/12216)	4.2% (514/12196)
朗読発声 (文節)	5.1% (127/2507)	7.2% (183/2530)
朗読発声 (文)	12.7% (315/2479)	11.0 % (276/2506)
自由発話	16.0% (445/2780)	15.0 % (702/4691)

学習データに自由発話の音声データを使用したときの、自由発話の母音の音素認識誤り傾向を表 9および表 10に示す。この結果を見ると /u/ を中心とした誤りが目立つ。例えば、/i/ を /u/ とする誤りが話者 MTK では 7.6% 話者 FKN では 5.4%、/u/ を /o/ とする誤りが話者 MTK では 25.4% 話者 FKN で 10.0%、/o/ を /u/ とする誤りが話者 MTK では 11.3% 話者 FKN で 8.9% であった。

表 9: 音素認識誤り傾向 話者 MTK
(認識音素数 / 調査音素数)

	出力				
	a	i	u	e	o
a	85.3 %	1.1 %	0.9 %	4.9 %	7.7 %
	(401/470)	(5/470)	(4/470)	(23/470)	(36/470)
i	1.8 %	87.6 %	0.0 %	7.6 %	0.4 %
入	(5/275)	(241/275)	(0/275)	(21/275)	(1/275)
u	4.4 %	1.8 %	56.1 %	4.4 %	25.4 %
力	(5/114)	(2/114)	(64/114)	(5/114)	(29/114)
e	3.8 %	8.5 %	4.4 %	80.9 %	0.7 %
	(11/293)	(25/293)	(13/293)	(237/293)	(2/293)
o	5.8 %	0.7 %	11.3 %	3.1 %	78.1 %
	(17/292)	(2/292)	(33/292)	(9/292)	(228/292)

表 10: 音素認識誤り傾向 話者 FKN
(認識音素数 / 調査音素数)

	出力				
	a	i	u	e	o
a	82.4 %	0.1 %	2.8 %	9.0 %	3.9 %
	(734/891)	(1/891)	(25/891)	(80/891)	(35/891)
i	3.8 %	82.4 %	1.2 %	5.4 %	0.0 %
入	(19/500)	(412/500)	(6/500)	(27/500)	(0/500)
u	1.4 %	1.8 %	72.1 %	10.5 %	10.0 %
力	(3/219)	(4/219)	(158/219)	(23/219)	(22/219)
e	0.7 %	2.4 %	5.3 %	87.6 %	2.4 %
	(3/418)	(10/418)	(22/418)	(366/418)	(10/418)
o	3.2 %	0.2 %	8.9 %	1.6 %	84.6 %
	(14/436)	(1/436)	(39/436)	(7/436)	(369/436)

3.5 自由発話の音響的な特徴の考察

ここでは、自由発話の音響的な特徴を調査するために、主に融合ラベルの付与率および発話速度および HMM における音素認識誤り率でと朗読発声と比較した。その結果、自由発話は文単位の朗読発声と比較すると、発話速度に大きな差がないが、音素の継続時間長にバラツキが大きいこと、融合ラベルの出現頻度は約 2 割近く増加すること、音素認識誤り率は、単語発声の音声データを学習に使用したときの 2 倍以上になるが、自由発話の音声データを学習に使用した場合は約 3 割増加することが示された。

このような結果は、見方によっては自由発話の認識はさほど困難ではないとも考えられる。例えば、学習データが自由発話のときの、自由発話の約 15% という音素認識誤り率は、学習データが単語発話のときの朗読発声の文認識よりも低い。したがって少

なくとも音素モデルに関しては、自由発話と朗読認識において大きな差はないように思われる。

ただし、これらの調査した値は話者によって差がある。特に融合ラベルの付与率が自由発話のとき話者 MTK と FKN で差がある。したがって自由発話の認識は話者によって大きく異なる可能性が残っている。

4 自由発話の言語的な特徴

自由発話における言語の特徴については、自然言語処理の立場からどのような言語表現があるのかを調べた報告が既にある [9]。また、山本ら [10] は実際の対話文約 1800 文を名詞文節の助詞落ちや倒置の点から解析している。ここでは、自由発話と朗読発声の言語現象の差を調べる立場から、朗読発話では見られない言語（発話）現象、特に言い直しと冗長語に焦点をあてて、それぞれの出現頻度を調べた。今回調査した会話文は、国際会議の問い合わせの対話文 11054 文である。

4.1 自由発話の文の長さ

図 5に、文の長さと出現回数を示す。自由発話においては「文」の明確な定義が困難である事例が多いため、文字化する際に意味的にまとまっていると判断できる単位を「文」とした。したがって文の定義には文に書き起こした人の主観が含まれる。

漢字仮名混じり表記に書き下したテキストデータを調査した結果、自由発話の 1 文の平均文字数は 25.4 文字、もっとも短い文は「あ」の 1 文字、もっとも長い文は 311 文字であった。最も出現頻度の高い文は、「はい」の 2092 文であった。また、自由発話の文の 73% は 32 文字以下であった。長い文の例を表 11に示す。

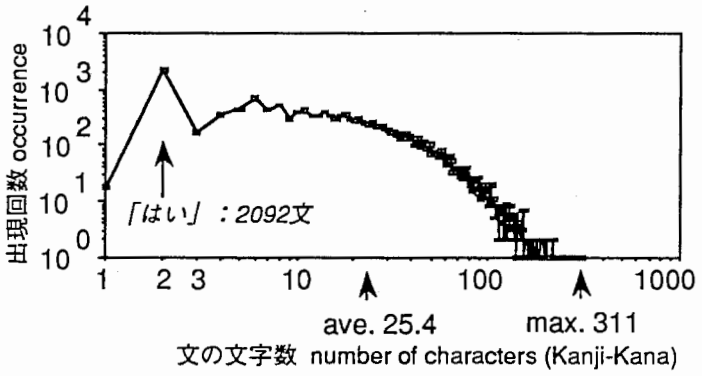


図 5: 自由発話における文の長さの分布

表 11: 長い文の例

えー、松下の場合にはですね、もうすでに、まー、あのー、見学コースっていうのが設定されておりまして、えー、会議の参加者のみならず、いろんな興味のある方々、これは日本人の方も外国人の方も見れる訳ですが、そういった、松下電器が、今までどの様な製品を作り、現在どの様なシステムで、えー、いろんな製品を作っておるか、そして、今後将来、松下がどういう方向性を目指してるか、という過去現在未来といった様な、製品の製作展開等のコースを見て頂くことになります。

4.2 自由発話における言い直しや冗長語の出現頻度

自由発話の音声には、「あのー」や「えーと」などの冗長語や、言葉の言い直しおよび言い誤りなどがある。これらの言語現象は、朗読発声では通常出現しないため、従来の文法の枠組では、あまり考慮されていない。そこで、自由発話における冗長語や言い直しの出現頻度を調べた。

この結果を図 6および表 12に示す。この結果において、冗長語も言い直しも共にない文は、全体の約 5 割であった。これらの多くは「はい」「いいえ」「もしもし」「どうぞ」「わかりました」「失礼します」などの定型文で、この種類の 8 割の文は 14 文字以下の短い文であった。

自由発話の文の約 5 割は冗長語を含み、多くの単語が続く文の多くは冗長語を含んでいた。ただし、冗長語には個人差が多く、冗長語を多く話す話者とあまり話さない話者がいた。また、一人の話者が話す冗長語の種類は限られていた。言い直しがある文は自由発話全体の約 1 割であった。そして、「はい」「もしもし」「あ」などの独立語も冗長語に含めた場合、全体の文の 83%(9121 文) は冗長語があった。この中で文頭に冗長語があるものは、全体の文の 65.8%(7303 文) であった。また、その多くの場合、言い直しの前後に冗長語が付加されていた。

なお、今回調査した自由発話のデータは、ナレータや実際の事務局員など、言葉の対応に慣れた話者が発話した音声である。したがって、言葉の応対に慣れていない一般の話者では、冗長語や言い直しの出現頻度が増加する可能性がある。

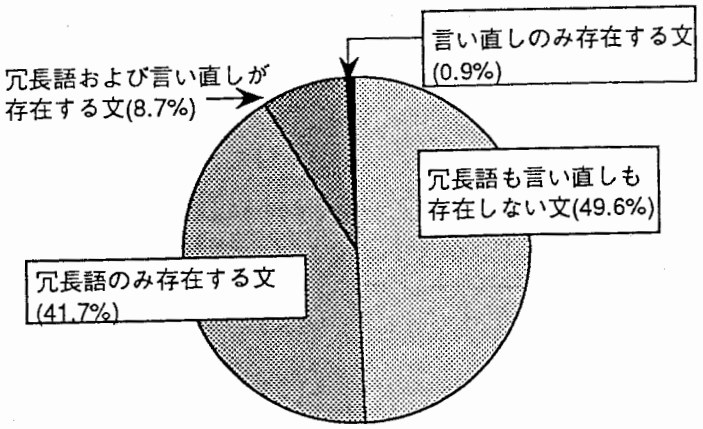


図 6: 冗長語の出現頻度

表 12: 冗長語の出現頻度
(冗長語が出現した文 / 調査した文)

言い直しのみがある文	0.9% (94/11054)
冗長語も言い直しもある文	8.7% (960/11054)
冗長語のみがある文	41.7% (4607/11054)
冗長語も言い直しもない文	49.6% (5487/11054)

4.3 自由発話における冗長語の種類と出現確率

自由発話において観測された冗長語の種類のなかで、出現頻度の高いものを図 7 および表 13に示す。また、観測された全ての冗長語を表 23に示す。この表から、冗長語の種類はかなり多いが、上位 4 種類で冗長語全体の出現頻度の約 7 割を占めていることがわかる。

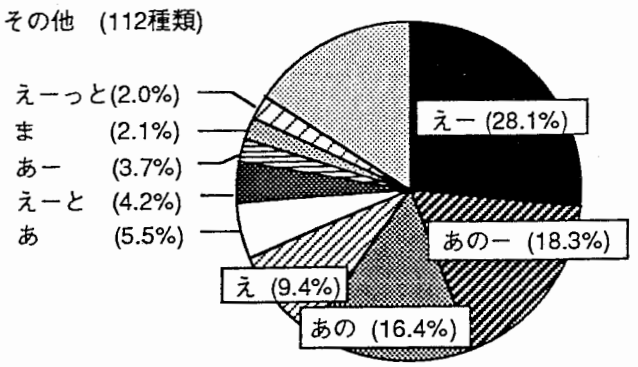


図 7: 自由発話における冗長語の出現率

表 13: 自由発話における冗長語の出現率

冗長語	出現率	出現頻度
「えー」	28.1%	3105
「あのー」	18.3%	2025
「あの」	16.4%	1809
「え」	9.4%	1040
「あ」	5.5%	604
「えーと」	4.2%	466
「あー」	3.7%	268
「ま」	2.1%	263
「えーっと」	2.0%	256
その他合計		11064

自由発話中では、しばしば冗長語と言い淀みの区分が不明確になる。例えば「100 パーセント、え 日本語と英語で行われます。」における「え」は、冗長語とも「英語」の言い淀みとも解釈できる。また、語尾音の継続時間には、話者間に大きなバラツキがあるため、語尾の伸びる語と伸びない語（例えば「えー」と「え」）の決定は、文字化した人の判断に依存している。ここで示した冗長語の出現頻度のデータには、このような意味で曖昧さがある。

なお、話し相手と対面して話す自由発話に対して、電話のような音声のみによる対話では、冗長語は相手の注意を促す役割を持つ場合がある。このため、今回調査した冗長語の出現頻度は、高めに評価されている可能性がある。

4.4 自由発話における言い直しの種類と出現頻度

自由発話において特有な、言い誤りや言い淀みなど言語現象は、テキストデータ中では、話者が言い直さないかぎり、検出するのは困難である。なぜなら言い誤りであるかどうかの判定は、文法的、意味的な前後関係を考慮して決定する必要がある。また、言い淀みは、音声を注意深く聞いて決定する必要がある。これらの問題があるため、本論文では言い直しの出現頻度のみを調査した。調査は200文に対して行なった。この言い直しの分類と出現頻度を、図8および表14に示す。また例文を以下に示しめす。例文中においてアンダーラインは言い直しを意味する。

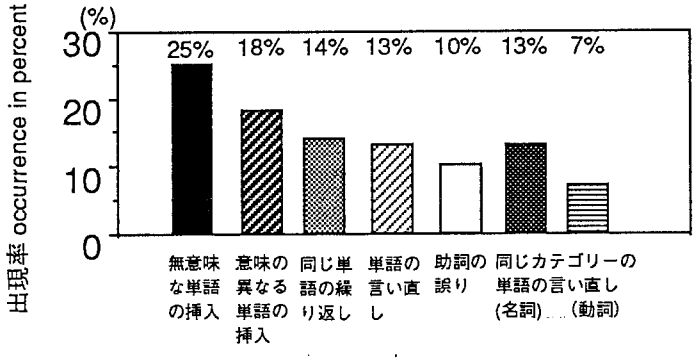


図 8: 自由発話における言い直しの種類別頻度

表 14: 自由発話における言い直しの種類別頻度

無意味な単語の挿入	25%
意味の異なる単語の挿入	18%
同じ単語の繰り返し	14%
単語の言い直し	13%
助詞の誤り	10%
丁寧な単語を用いた言い直し (名詞)	13%
丁寧な単語を用いた言い直し (動詞)	7%

ただし、この言い直しの分類には問題が多い。例えば、単語の意味の違いは明確でないため『意味の異なる単語の挿入』と『丁寧な単語での言い直し』の区別の差は明確でない。また、日本語では単語の概念が曖昧なため、『同じ単語の繰り返し』と『単語の言い直し』の区別の差も明確でない。したがって、ここで示した分類はかなり主観的である。

自由発話における言い直しの例文

1. 無意味な単語の挿入　25%

- 日本語から英語へというように、と、翻訳を、す、あ、 通訳をするコンピュータを開発している
（「通訳」と言おうとして「翻訳」と言い間違いをし、これに気がついて直そうとして言い淀んでいる。）
- えーっと、あのー、こ、 会議期間中は特にあのー、バスを運行しておりまして、土曜ダイヤでバスが、あのー、運行するようになっております。
（原因が不明、「こ」は、無意味な音の発声であるため、冗長語と判断される可能性がある。）
- 最終的な、えーっと、草稿、原、 えーとスピーチ原稿を提出していただきたいと思います。
（「原稿」と言おうとして「草稿」と言い間違いをし、これに気がついて直そうとして言い淀んでいる。）
- パンフレットの方を拝、 見ていただきましたら
（「拝見」と言おうとして敬語の間違いをして言い淀んでいる。）

2. 意味の異なる単語の挿入　18%

- あの、そのようなことが、あの、そちらの方にお教え、 お知らせできないんです。
（「知らせる」を「教える」に言い間違えている。）
- タクシーに、あのー、京都駅からお乗りになれば、大体３５分か４０分位で着きますし、旅費、料金としては、大体１５００円位になります。
（「旅費」と「料金」は、意味的にはほとんど同じであるため、『丁寧な言葉への言い直し』とも分類できる。）
- この件に関しましては、えーっと、大阪まで、あのー、新幹線で来られますと、飛行機で来られますと４５分間位で参ります。
（「飛行機」を「新幹線」と言い間違えている。文全体の挿入の誤り。）

3. 同じ単語の繰り返し　14%

- えーっと、その、その中でちょっと、あの、クレジットカードをね書類の方は、（「その中」を１つの単語と捉えたならば『単語の言い直し』とも解釈できる。）
- 会議の内容なんかをかいつまんでお話、お話し下さればと思うんですが。

4. 単語の言い直し　13%

- あの、この、クレ、クレジットカードというのは本来外国人のゲストの方
- 従いまして、２、あ、２時間半位で東京から国際会議の行なわれる場所まで行けるわけですから、

5. 助詞の誤り　10%

- まだ割引を私の方で、あのー、することに、はできないんですが
- はい、それで、はそうですね。
- コンピュータによる同時通訳を、に関する、あのー会議を開こうということです。
- オーバーヘッドープロジェクタと２インチ×２インチのスライドを、と使えるようになっています。

6. 丁寧な単語を用いた言い直し（名詞）　13%

- えーっと、郵送でＶＬＤＢ８６の、えーと、会議事務局、国際会議事務局宛にお送りいただきたいと思います。（意味的には『同じ単語の繰り返し』ともみなせる。）
- それで、えーっと、受領の通知は、受け取りの通知は１２月３１日までに出させていただきます。
- これは現在の為替でいきますと、レートでいきますと、大体16,000円程になりますので
- その次に日本の総理大臣中曽根首相から挨拶を、スピーチをすることになってます。

7. 丁寧な単語を用いた言い直し（動詞）　7%

- はがきででも来られない、参加できないという風に、御通知いただければ、
- ええ、外国人の申し込みの方は、現在までで１３名であり、ございます
- そうですか、という、といいますと、それは英語でしなければいけないわけでしょうか。

4.5　自由発話の言語的な特徴の考察

本報告では、自由発話の言語的な特徴を見るために、文の長さおよび冗長語および言い直しの出現頻度を調べた。その結果、自由発話では平均して25.4文字、短い文が多いが、非常に長い文もあることがわかった。また、自由発話全体の約１割に言い直しが出現することがわかった。そして冗長語は種類は多いが、４種類で冗長語全ての出現数の７割を占めていることがわかった。

音声認識に使用される文法の枠組には、有限状態オートマトンや文脈自由文法、単一化文法などがあるが、これらの文法で、言い直しや言い誤りなどに対応するには、かなりの困難が予想される。しかし、冗長語については、独立語として従来の文法の枠組で扱える可能性が考えられるため、自由発話全体の約９割の文章は、従来の文法の枠組で扱える可能性がある。ただし、冗長語を考慮した文法のperplexityは増大しがちなため、音声認識はさらに困難になるかも知れない。

5　まとめ

ここでは自由発話の認識にむけて、自由発話と朗読発声の音響的および言語的な差を調べた。この結果、自由発話の音声は朗読音声と比較すると、発話速度ではあまり差がないが、融合ラベルの付与率が約２割増加すること、また音素認識誤り率は約３割増加することがわかった。しかし融合ラベルを除いた音素認識誤り率は約15%と、著しく高い値ではないことや、融合ラベルの増加の問題は連結学習により解消されることが期待できるため、少なくとも音素モデルに関しては、自由発話音声認識には問題がないように思われる。また、言語モデルの面から見ると、自由発話に特有な冗長語は自由発話全体の９割に出現するが、独立語として扱うことによって従来の文法の枠組でカバーできる可能性がある。したがって、言い直しの言語現象を除けば、自由発話音声認識は、ある程度の実現可能性を持っていると言って良いように思われる。

最後に、言い直しの言語現象に対応するには、認識アルゴリズムとしてスポッティングなどが考えられる。あるいは、より精密な文法を作成することによって、従来ＡＴＲで試みられてきているHMM-LR

法も考えられる。また、統計的な言語モデルで言い直しの現象を扱える可能性もある。以上の長所や短所を考えながら、自由発話の音声認識システムを構築する必要があろう。

ただし、ここで扱った自由発話は、言葉の対応に慣れた人たちが限定した条件の下で発話したデータであるため、かなり clean な音声データと言うべきである。したがって、一般の話者が、雑音下で制約の少ない状態で話した音声では、この論文で調査した結果と若干異なる可能性がある。

謝辞

研究で使用した自由発話の言語のデータベースの解析は、匂坂氏や現ＮＨＫの江原氏や指示のもとになされたものです。また、自由発話の音素ラベリングには、木田嬢を始めとする多くのラベラーの方々の協力によってなされました。また、ATR 自動翻訳電話研究所の森元室長や飯田室長の他、音声とデータと言語の各研究員に多くの協力をいただきました。これらの方々に厚くお礼を申しあげます。

参考文献

[1] Kai-Fu Lee, "Large-Vovabulary Speaker Independent Continuous Speech Recognitton: The SPHINX System," Computer Science Department Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 CMU-CS-88-148 (April 18, 1988).

[2] 永井 明人, 他, "HMM-LR 連続音声認識装置の開発と性能評価," 日本音響学会平成 3 年度秋季研究発表会, 1-5-23 pp.45-46, (Oct. 1991).

[3] 坪井 宏之, 橋本 秀樹, 竹林 洋一, "連続音声理解のためのキーワードラティスの解析," 日本音響学会平成 3 年度秋季研究発表会, 1-5-11, pp.21-22, (Oct. 1991).

[4] Victor Zue, James Glass, David Goodine, et al. "The MIT ATIS System: Preliminary Development, Spontaneous Speech Data Collection,And Performance Evaluation," Eurospeech 91, pp.537-540, Genova, Italy, (Sep. 1991)

[5] Victor Zue, Nancy Daly, et al, "The Collection and Preliminay Analysis of a Spontaneous Speech Database," DARPA Workshop 1989, pp.126-134 (1989).

[6] 篠崎 直子, 小倉 健太郎, 森元 逞, "言語データベース作成のためのシュミレーション会話," 第 37 回情報処理全国大会, pp.1000-1001 (1988).

[7] 江原 暉将, 小倉 健太郎, 森元 逞, "電話対話データベースの構築," 第 40 回情報処理全国大会, pp.486-491 (1990).

[8] 武田 一哉, 匂坂 芳典, 片桐 滋, 桑原 尚夫, "音韻ラベルの持つ日本語音声データベースの構築"SP87-19(1987).

[9] 有田 英一, 小暮 潔, 野垣内 出, 飯田 仁, "メディアに依存する会話の様式," NL61-5 (1987).

[10] 山本 幹雄, 小林 聡, 中川 聖一, "音声対話文における助詞落ち・倒置の分析と解析手法," 情報処理学会論文誌 Vol.11, No11, pp.1322-1330 (1992).

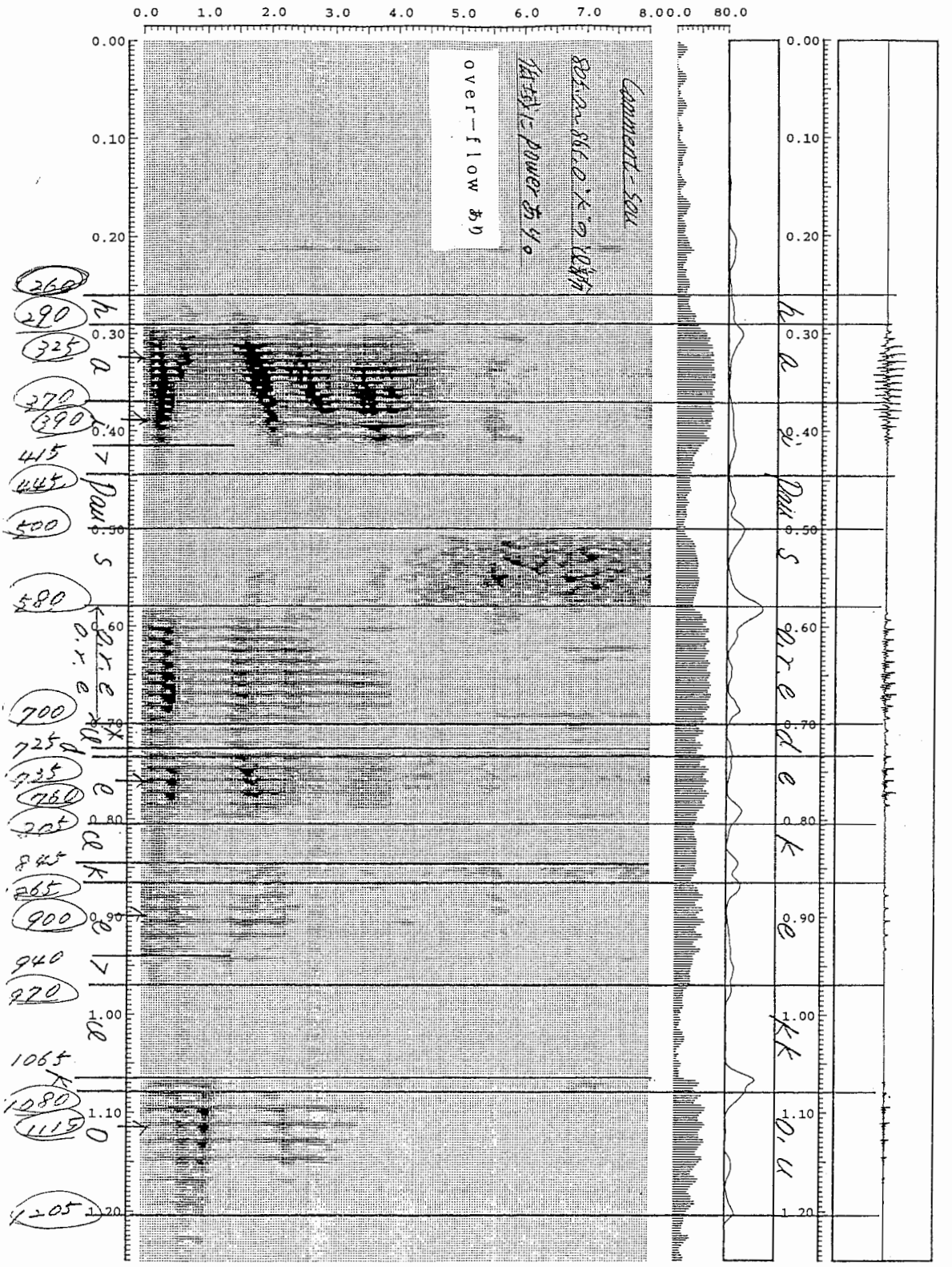


図 9: 融合ラベルの例「はい、それで結構」

表 15: 融合ラベルの出現頻度 話者・ MTK

(融合した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	1.4% (52/3523)	3.3% (30/902)	4.0% (36/903)	13.3% (135/1018)
i	24.0% (818/3407)	28.7% (216/752)	30.4% (228/751)	36.9% (298/ 808)
u	39.1% (1869/4777)	67.1% (516/769)	69.0% (531/770)	76.1% (632/ 830)
e	19.9% (332/1661)	8.8% (39/443)	9.5% (42/443)	25.3% (175/ 693)
o	36.7% (1013/2760)	29.9% (236/790)	31.4% (248/790)	38.1% (321/ 842)
p	6.2% (3/ 48)	0.0% (0/ 2)	0.0% (0/ 2)	0.0% (0/ 4)
t	18.7% (0/ 865)	0.0% (0/251)	0.4% (1/251)	3.9% (10/ 257)
k	6.8% (168/2446)	10.5% (55/522)	11.7% (61/522)	10.8% (41/ 378)
b	7.6% (34/ 446)	14.6% (7/ 48)	14.6% (7/ 48)	18.0% (9/ 50)
d	7.3% (28/ 382)	11.0% (26/236)	11.0% (26/236)	28.0% (97/ 347)
g	7.5% (42/ 554)	16.9% (24/142)	17.6% (25/142)	27.8% (42/ 151)
m	1.9% (19/ 976)	1.0% (2/192)	1.0% (2/192)	18.1% (40/ 221)
n	4.1% (23/ 548)	5.8% (17/294)	7.5% (22/294)	26.3% (88/ 335)
N	20.2% (226/1118)	54.1% (98/181)	55.2% (100/181)	74.2% (155/ 209)
s	6.0% (66/1083)	46.5% (120/258)	47.9% (124/259)	31.7% (116/ 366)
h	10.6% (52/ 487)	8.5% (13/153)	8.5% (13/153)	12.5% (25/ 200)
f	45.3% (69/ 152)	70.0% (7/ 10)	70.0% (7/ 10)	63.6% (7/ 11)
z	9.4% (24/ 253)	4.5% (1/ 22)	4.5% (1/ 22)	6.3% (1/ 16)
j	11.3% (45/ 395)	1.7% (1/ 59)	1.7% (1/ 59)	15.2% (12/ 79)
r	0.0% (0/1484)	0.0% (0/235)	0.9% (2/235)	9.8% (20/ 205)
w	9.5% (16/ 168)	58.3% (21/ 36)	66.7% (24/ 36)	50.0% (19/ 38)
y	10.3% (39/ 378)	16.2% (16/ 99)	18.2% (18/ 99)	50.0% (36/ 72)
ts	0.0% (91/ 485)	15.5% (9/ 58)	20.7% (12/ 58)	22.2% (12/ 54)
sh	18.6% (147/ 787)	40.4% (69/171)	40.9% (70/171)	50.5% (92/ 182)
hy	0.0% (0/ 19)	0.0% (0/ 8)	0.0% (0/ 8)	
ch	12.6% (36/ 284)	14.0% (6/ 43)	14.0% (6/ 43)	9.4% (8/ 85)
ky	0.0% (0/ 100)	0.0% (0/ 34)	5.8% (2/ 34)	0.0% (0/ 16)
by	12.5% (1/ 8)			
gy	10.7% (3/ 28)	0.0% (0/ 1)	0.0% (0/ 1)	
my	11.1% (1/ 9)			
ny	7.6% (1/ 13)	0.0% (0/ 2)	0.0% (0/ 2)	0.0% (0/ 2)
ry	0.0% (0/ 81)	0.0% (0/ 14)	0.0% (0/ 14)	0.0% (0/ 6)
cch	0.0% (0/ 9)	0.0% (0/ 1)	0.0% (0/ 1)	
dd	0.0% (0/ 1)			
he		0.0% (0/ 2)	0.0% (0/ 2)	
kk	0.0% (0/ 83)	0.0% (0/ 8)	0.0% (0/ 8)	0.0% (0/ 8)
kky	0.0% (0/ 9)			
pp	0.0% (0/ 26)	0.0% (0/ 3)	0.0% (0/ 3)	0.0% (0/ 6)
ppy	0.0% (0/ 1)	0.0% (0/ 13)	0.0% (0/ 13)	0.0% (0/ 4)
py	0.0% (0/ 1)			
ss	0.0% (0/ 35)	0.0% (0/ 1)	0.0% (0/ 1)	0.0% (0/ 2)
ssh	15.3% (4/ 26)			0.0% (0/ 2)
tt	0.0% (0/ 50)	0.0% (0/ 19)	0.0% (0/ 19)	0.0% (0/ 65)
tts	16.6% (1/ 6)	0.0% (0/ 1)	0.0% (0/ 1)	
wo		42.5% (17/ 40)	47.5% (19/ 40)	54.5% (12/ 22)

表 16: 融合ラベルの出現頻度 話者・ FKN

(融合した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	1.5% (52/3523)	2.2% (20/902)	3.9% (35/902)	8.1% (154/1893)
i	23.8% (810/3407)	27.6% (207/751)	28.6% (215/752)	33.6% (485/1444)
u	39.3% (1877/4777)	66.5% (512/770)	66.6% (512/769)	67.8% (951/1402)
e	19.9% (331/1661)	8.8% (39/443)	8.8% (39/443)	15.6% (154/ 990)
o	36.9% (1018/2760)	30.0% (237/790)	32.7% (259/791)	34.3% (457/1334)
p	4.2% (2/ 48)	0.0% (0/ 2)	0.0% (0/ 2)	0.0% (0/ 8)
t	0.0% (0/ 865)	0.0% (0/251)	0.0% (0/251)	0.2% (1/ 407)
k	7.0% (170/2446)	10.0% (52/522)	10.3% (54/522)	10.0% (77/ 768)
b	7.4% (33/ 446)	14.6% (7/ 48)	14.6% (7/ 48)	32.9% (25/ 76)
d	7.3% (28/ 382)	11.0% (26/236)	11.0% (26/236)	23.6% (121/ 512)
g	7.8% (43/ 554)	16.9% (24/142)	17.6% (25/142)	14.6% (36/ 246)
m	1.9% (19/ 976)	1.0% (2/192)	1.6% (3/192)	10.8% (50/ 465)
n	4.4% (24/ 548)	5.8% (17/294)	6.8% (20/294)	16.2% (87/ 537)
N	19.9% (223/1118)	54.1% (98/181)	54.1% (98/181)	70.9% (231/ 326)
s	8.2% (89/1083)	43.0% (111/258)	39.5% (102/258)	42.2% (227/ 538)
h	11.5% (56/ 487)	9.6% (14/153)	8.5% (13/153)	11.1% (36/ 323)
f	44.1% (67/ 152)	70.0% (7/ 10)	70.0% (7/ 10)	25.0% (4/ 16)
z	8.7% (22/ 253)	9.1% (2/ 22)	18.2% (4/ 22)	6.5% (3/ 46)
j	11.6% (46/ 395)	0.0% (0/ 59)	1.7% (1/ 59)	5.8% (7/ 121)
r	0.0% (0/1484)	0.0% (0/235)	0.0% (0/235)	3.0% (13/ 428)
w	10.1% (17/ 168)	58.3% (21/ 36)	66.7% (24/ 36)	54.4% (37/ 68)
y	10.3% (39/ 378)	16.2% (16/ 99)	18.2% (18/ 99)	13.0% (15/ 115)
ts	17.3% (84/ 485)	15.5% (9/ 58)	19.0% (11/ 58)	21.7% (23/ 106)
sh	20.1% (158/ 787)	36.3% (62/171)	36.3% (62/171)	46.2% (178/ 385)
hy	0.0% (0/ 19)	25.0% (2/ 8)	12.5% (1/ 8)	
ch	12.0% (34/ 284)	14.0% (6/ 43)	14.0% (6/ 43)	8.3% (9/ 108)
ky	0.0% (0/ 100)	0.0% (0/ 34)	0.0% (0/ 34)	0.0% (0/ 52)
by	12.5% (1/ 8)			
gy	10.7% (3/ 28)	0.0% (0/ 1)	0.0% (0/ 1)	0.0% (0/ 6)
my	11.1% (1/ 9)			
ny	7.7% (1/ 13)	0.0% (0/ 2)	0.0% (0/ 2)	0.0% (0/ 4)
ry	0.0% (0/ 81)	0.0% (0/ 14)	0.0% (0/ 14)	0.0% (0/ 4)
cch	0.0% (0/ 9)	0.0% (0/ 1)	0.0% (0/ 1)	0.0% (0/ 2)
dd	0.0% (0/ 1)			
he		0.0% (0/ 2)	0.0% (0/ 2)	
kk	0.0% (0/ 83)	0.0% (0/ 8)	0.0% (0/ 8)	0.0% (0/ 22)
kky	0.0% (0/ 9)			
pp	0.0% (0/ 26)	0.0% (0/ 3)	0.0% (0/ 3)	0.0% (0/ 2)
ppy	0.0% (0/ 1)	0.0% (0/ 13)	0.0% (0/ 13)	0.0% (0/ 10)
py	0.0% (0/ 1)			0.0% (0/ 8)
ss	0.0% (0/ 35)	0.0% (0/ 1)	0.0% (0/ 1)	0.0% (0/ 8)
ssh	0.0% (0/ 26)			0.0% (0/ 10)
tt	0.0% (0/ 50)	0.0% (0/ 19)	0.0% (0/ 19)	
tts	0.0% (0/ 6)	0.0% (0/ 1)	0.0% (0/ 1)	
wo		42.5% (17/ 40)	59.0% (23/ 39)	53.3% (16/ 30)

表 17: 各音素ごとの平均音素継続時間 話者 MTK

発話様式	単語発声		朗読発声 (文節)		朗読発声 (文)		自由発話	
音素	平均 (ms)	分散	平均 (ms)	分散	平均 (ms)	分散	平均 (ms)	分散
a	120.4	513.2	108.8	821.4	90.4	1528.56	96.4	2877.6
i	115.1	808.7	95.1	1140.5	69.8	1209.1	76.6	1766.6
u	108.8	758.0	82.4	1574.5	56.8	1044.5	80.2	3069.9
e	126.0	501.7	107.4	920.3	75.3	1134.9	92.9	4331.0
o	110.6	599.3	108.0	1253.0	80.4	1339.9	88.6	3138.7
p	103.1	2060.9	112.5	756.3	51.3	264.1	64.4	19.9
t	71.5	1495.4	78.7	1193.7	50.5	401.2	38.2	401.8
k	88.4	1389.1	76.3	749.4	55.8	384.7	45.9	399.6
b	67.4	464.2	71.1	534.2	52.0	452.3	50.5	835.8
d	65.6	439.9	53.4	314.0	42.7	484.3	43.3	316.9
g	79.8	366.9	72.6	415.6	50.2	247.3	45.8	677.5
m	64.7	300.0	59.1	180.3	46.9	198.6	53.3	311.8
n	63.5	236.8	61.5	219.6	44.1	167.2	47.8	287.0
N	165.1	571.4	126.7	1025.6	77.7	329.9	88.6	2026.3
s	133.9	703.4	114.9	704.3	80.3	354.0	96.3	798.9
h	64.8	928.3	65.2	548.1	56.2	337.4	57.1	1298.9
f	90.8	1122.9	115.0	1616.7	65.0	5.0	60.0	325
z	76.5	326.8	70.7	434.0	54.9	289.6	59.0	664
j	99.8	1315.2	96.4	962.7	80.3	708.3	110.5	2630.8
r	31.0	170.7	25.9	216.3	20.0	112.2	24.6	226.4
w	66.3	308.7	50.3	468.2	43.8	333.8	42.9	392.9
y	85.8	601.1	73.7	437.1	68.9	467.1	66.4	381.4
ts	126.0	553.6	113.5	1126.5	80.9	423.2	77.9	517.7
sh	166.4	1169.8	142.3	646.2	102.7	327.5	97.0	1230.9
hy	148.9	759.4	109.4	352.7	91.3	473.4		
ch	130.6	1290.5	99.3	537.8	70.7	718.2	68.1	746.6
ky	142.8	2630.1	111.4	1015.5	84.5	441.6	82.7	249.6
by	151.4	267.6						
gy	131.2	872.6						
my	136.9	1105.9						
ny	170.8	316.0	135	225	102.5	306.3	85	25.0
ry	102.0	699.1	99.1	301.4	92.3	382.6	72.5	306.25

表 18: 各音素ごとの平均音素継続時間 話者 FKN

発話様式	単語発声		朗読発声 (文節)		朗読発声 (文)		自由発話	
音素	平均 (ms)	分散	平均 (ms)	分散	平均 (ms)	分散	平均 (ms)	分散
a	127.5	782.8	96.7	424.4	87.4	416.0	97.4	1266.1
i	118.8	1188.7	75.5	571.8	63.8	517.1	77.8	2035.3
u	118.0	984.4	71.9	918.6	53.7	481.8	68.6	2107.8
e	136.9	834.8	97.7	545.3	80.6	574.6	77.0	1174.6
o	117.7	927.3	93.6	689.1	82.6	608.3	85.8	1405.5
p	153.2	7652.4	110	2500	73.8	1701.6	66.9	2099.6
t	85.3	4777.8	89.8	2090.3	71.7	731.1	40.9	1239.2
k	115.7	4366.8	88.1	1514.3	86.6	757.6	60.1	775.3
b	66.3	294.5	55.3	289.1	49.4	181.0	45	188.2
d	59.5	295.1	45.8	156.4	40.1	142.8	37.4	242.8
g	85.4	471.6	69.3	351.4	58.2	219.4	43.8	264.6
m	73.5	387.6	57.1	182.5	55.8	174.6	51.6	222.9
n	65.3	400.0	58.9	268.0	48.9	210.9	45.6	284.2
N	157.5	1579.7	88.8	1097.3	69.4	363.9	84.5	1427.4
s	154.6	1007.9	108.6	532.8	99.6	484.6	102.9	733.3
h	83.7	1338.9	70.6	331.2	66.4	292.3	65.6	742.7
f	100.9	1558.4	90	50	83.3	572.2	55.4	347.7
z	81.9	416.1	63.8	154.7	62.5	242.4	54.1	474.7
j	106.5	1338.5	92.5	1049.5	87.7	767.4	90.7	1301.0
r	37.6	245.8	24	110.0	21.5	101.2	19.8	54.8
w	64.8	474.6	43.7	328.2	47.3	684.9	34.7	235.4
y	89.9	699.0	67.4	367.0	69.7	404.3	57.8	305.2
ts	152.9	1869.0	107.9	1022.3	101.2	834.0	67.5	547.8
sh	190.9	1827.7	136.1	491.7	120.6	464.3	106.6	632.3
hy	186.1	1391	130.8	1028.5	93.6	148.0		
ch	158.8	4977.6	97.9	993.1	90.9	842.7	94.7	2707.5
ky	168.9	8962.2	101.6	1698.5	97.2	1151.8	100.3	2141.7
by	138.6	90.8						
gy	157	866					89.2	245.1
my	155	1612.5						
ny	157.5	1102.1	97.5	6.3	87.5	6.3	90	650
ry	112.3	915.7	90.7	170.0	86.8	241.5	63.8	17.2

表 19: 音素認識誤り率 話者・ MTK 学習データ・単語発声

(誤認識した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	1.6% (29/1758)	3.8% (33/880)	8.5% (74/867)	29.2% (258/883)
i	8.4% (107/1278)	18.7% (100/536)	36.5% (191/523)	48.0% (245/510)
u	9.7% (139/1428)	12.6% (32/253)	24.3% (58/239)	36.4% (72/198)
e	7.0% (47/ 672)	9.9% (40/404)	25.2% (101/401)	44.6% (231/518)
o	5.0% (43/ 861)	11.2% (62/554)	26.0% (141/542)	47.2% (246/521)
p	33.3% (5/ 15)	100.0% (2/ 2)	100.0% (2/ 2)	50.0% (2/ 4)
t	7.5% (33/ 440)	8.4% (21/251)	19.2% (48/250)	33.2% (82/247)
k	8.5% (97/1139)	7.7% (36/467)	5.4% (25/461)	13.4% (45/337)
b	18.8% (39/ 207)	22.0% (9/ 41)	41.5% (17/ 41)	41.5% (17/ 41)
d	7.1% (12/ 170)	10.5% (22/210)	25.7% (54/210)	41.2% (103/250)
g	9.1% (23/ 253)	22.9% (27/118)	40.2% (47/117)	76.1% (83/109)
m	15.8% (76/ 482)	12.6% (24/190)	13.2% (25/190)	54.1% (98/181)
n	6.8% (18/ 265)	7.2% (20/277)	25.4% (69/272)	51.4% (127/247)
N	6.0% (26/ 435)	24.1% (20/ 83)	40.7% (33/ 81)	27.8% (15/ 54)
s	26.0% (140/ 539)	30.4% (42/138)	28.9% (39/135)	26.8% (67/250)
h	4.2% (9/ 215)	50.0% (70/140)	52.1% (73/140)	37.1% (65/175)
f	13.3% (6/ 45)	0.0% (0/ 3)	100.0% (3/ 3)	100.0% (4/ 4)
z	7.2% (8/ 111)	0.0% (0/ 21)	0.0% (0/ 21)	13.3% (2/ 15)
j	13.1% (22/ 168)	3.4% (2/ 58)	5.2% (3/ 58)	9.0% (6/ 67)
r	2.2% (16/ 727)	6.4% (15/235)	8.6% (20/233)	9.7% (18/185)
w	17.7% (14/ 79)	26.7% (4/ 15)	58.3% (7/ 12)	78.9% (15/ 19)
y	6.7% (12/ 178)	19.3% (16/ 83)	23.5% (19/ 81)	58.3% (21/ 36)
ts	17.0% (31/ 182)	24.5% (12/ 49)	30.4% (14/ 46)	54.8% (23/ 42)
sh	20.9% (67/ 321)	28.4% (29/102)	24.8% (25/101)	42.2% (38/ 90)
hy	80.0% (8/ 10)	100.0% (8/ 8)	100.0% (8/ 8)	0.0% (0/ 0)
ch	26.4% (32/ 121)	16.2% (6/ 37)	18.9% (7/ 37)	27.3% (21/ 77)
ky	5.8% (3/ 52)	5.9% (2/ 34)	3.1% (1/ 32)	12.5% (2/ 16)
by	66.7% (2/ 3)	0.0% (0/ 0)	0.0% (0/ 0)	0.0% (0/ 0)
gy	84.6% (11/ 13)	100.0% (1/ 1)	100.0% (1/ 1)	0.0% (0/ 0)
ny	100.0% (8/ 8)	100.0% (2/ 2)	100.0% (2/ 2)	100.0% (2/ 2)
ry	17.1% (7/ 41)	21.4% (3/ 14)	7.1% (1/ 14)	66.7% (4/ 6)
total	8.9% (1090/12216)	12.7% (660/5206)	21.6% (1108/5120)	37.6% (1912/5084)

表 20: 音素認識誤り率 話者・ MTK 学習データ・同一発話様式

(誤認識した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	0.5% (8/1758)	1.0% (4/405)	11.1% (45/404)	14.7% (69/470)
i	1.5% (19/1278)	3.4% (9/262)	10.5% (27/256)	12.4% (34/275)
u	4.5% (64/1428)	9.4% (12/128)	29.8% (37/124)	43.9% (50/114)
e	0.1% (1/ 672)	1.5% (3/204)	12.4% (25/202)	19.1% (56/293)
o	1.5% (13/ 861)	4.9% (13/264)	14.7% (38/259)	21.9% (64/292)
p	33.3% (5/ 15)			100.0% (2/ 2)
t	2.0% (9/ 440)	1.7% (2/116)	1.7% (2/116)	2.7% (4/148)
k	5.8% (66/1139)	10.4% (24/231)	7.4% (17/229)	6.4% (11/171)
b	6.8% (14/ 207)	23.8% (5/ 21)	66.7% (14/ 21)	40.9% (9/ 22)
d	7.1% (12/ 170)	5.6% (6/107)	16.8% (18/107)	12.1% (17/140)
g	5.9% (15/ 253)	8.5% (4/ 47)	19.6% (9/ 46)	21.0% (13/ 62)
m	6.4% (31/ 482)	6.5% (7/107)	9.3% (10/107)	24.8% (25/101)
n	6.0% (16/ 265)	0.8% (1/128)	9.5% (12/126)	8.1% (11/135)
N	3.2% (14/ 435)	7.5% (3/ 40)	32.5% (13/ 40)	27.6% (8/ 29)
s	0.4% (2/ 539)	0.0% (0/ 69)	2.9% (2/ 68)	2.0% (3/148)
h	1.4% (3/ 215)	8.1% (5/ 62)	19.4% (12/ 62)	22.6% (19/ 84)
f	6.7% (3/ 45)			
z	5.4% (6/ 111)	40.0% (4/ 10)	60.0% (6/ 10)	7.2% (8/111)
j	3.6% (6/ 168)	3.7% (1/ 27)	3.7% (1/ 27)	5.1% (2/ 39)
r	3.4% (25/ 727)	9.2% (11/119)	4.2% (5/118)	10.0% (10/100)
w	3.8% (3/ 79)			100.0% (10/ 10)
y	5.1% (9/ 178)	2.2% (1/ 45)	18.2% (8/ 44)	55.6% (10/ 18)
ts	0.0% (0/ 182)	4.2% (1/ 24)	18.2% (4/ 22)	5.0% (1/ 20)
sh	0.3% (1/ 321)	1.9% (1/ 53)	3.8% (2/ 53)	6.4% (3/ 47)
hy	20.0% (2/ 10)			
ch	0.8% (1/ 121)			
ky	0.0% (0/ 52)	0.0% (0/ 12)	0.0% (0/ 12)	58.3% (7/ 12)
by	0.0% (0/ 3)	29.4% (5/ 17)	23.5% (4/ 17)	2.5% (1/ 40)
gy	15.4% (2/ 13)			
ny	62.5% (5/ 8)			
ry	4.9% (2/ 41)	55.6% (5/ 9)	44.4% (4/ 9)	100.0% (2/ 2)
total	2.9% (357/12216)	5.1% (127/2507)	12.7% (315/2479)	16.0% (445/2780)

表 21: 音素認識誤り率 話者・ FKN 学習データ・単語発声

(誤認識した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	0.1% (2/1759)	3.4% (30/882)	8.5% (74/867)	39.9% (690/1729)
i	0.5% (6/1277)	14.0% (76/544)	33.1% (178/537)	62.9% (600/ 954)
u	1.1% (15/1425)	12.0% (31/258)	18.7% (48/257)	17.2% (77/ 448)
e	0.3% (2/ 672)	4.2% (17/404)	12.6% (51/404)	68.6% (571/ 832)
o	0.0% (0/ 860)	12.5% (69/553)	25.8% (137/532)	50.4% (440/ 873)
p	40.0% (6/ 15)	100.0% (2/ 2)	100.0% (2/ 2)	87.5% (7/ 8)
t	1.8% (8/ 440)	1.6% (4/251)	3.2% (8/251)	29.2% (118/ 404)
k	3.4% (39/1137)	4.5% (21/470)	3.6% (17/468)	5.4% (37/ 688)
b	2.4% (5/ 207)	19.5% (8/ 41)	14.6% (6/ 41)	43.1% (22/ 51)
d	6.5% (11/ 170)	15.7% (33/210)	26.2% (55/210)	60.4% (235/ 389)
g	9.9% (25/ 253)	12.7% (15/118)	27.4% (32/117)	51.7% (108/ 209)
m	3.7% (18/ 482)	7.4% (14/190)	6.9% (13/189)	16.6% (68/ 410)
n	6.8% (18/ 265)	22.0% (61/277)	40.9% (112/274)	75.1% (338/ 450)
N	1.8% (8/ 439)	22.9% (19/ 83)	28.9% (24/ 83)	23.2% (22/ 95)
s	0.2% (1/ 524)	0.7% (1/147)	8.3% (13/156)	24.7% (76/ 308)
h	1.4% (3/ 215)	54.0% (75/139)	59.3% (83/140)	54.2% (154/ 284)
f	17.0% (8/ 47)	100.0% (3/ 3)	66.7% (2/ 3)	90.9% (10/ 11)
z	7.1% (8/ 113)	0.0% (0/ 20)	11.1% (2/ 18)	35.7% (15/ 42)
j	2.4% (4/ 168)	5.1% (3/ 59)	13.8% (8/ 58)	56.1% (64/ 114)
r	2.3% (17/ 727)	7.7% (18/235)	12.8% (30/235)	32.1% (133/ 414)
w	21.5% (17/ 79)	53.3% (8/ 15)	91.7% (11/ 12)	100.0% (31/ 31)
y	7.9% (14/ 178)	24.1% (20/ 83)	44.4% (36/ 81)	67.0% (67/ 100)
ts	0.6% (1/ 181)	2.0% (1/ 49)	2.1% (1/ 47)	29.3% (24/ 82)
sh	0.0% (0/ 315)	7.3% (8/109)	11.9% (13/109)	78.0% (160/ 205)
hy	50.0% (5/ 10)	100.0% (6/ 6)	100.0% (7/ 7)	0.0% (0/ 0)
ch	0.0% (0/ 121)	0.0% (0/ 37)	10.8% (4/ 37)	43.4% (43/ 99)
ky	5.8% (3/ 52)	0.0% (0/ 34)	2.9% (1/ 34)	9.6% (5/ 52)
by	100.0% (3/ 3)	0.0% (0/ 0)	0.0% (0/ 0)	0.0% (0/ 0)
gy	61.5% (8/ 13)	100.0% (1/ 1)	100.0% (1/ 1)	100.0% (6/ 6)
ny	87.5% (7/ 8)	100.0% (2/ 2)	100.0% (2/ 2)	100.0% (4/ 4)
ry	14.6% (6/ 41)	21.4% (3/ 14)	42.9% (6/ 14)	100.0% (4/ 4)
total	2.2% (268/12196)	10.5% (549/5236)	18.8% (977/5186)	44.4% (4129/9296)

表 22: 音素認識誤り率 話者・ FKN 学習データ・同一発話様式

(誤認識した音素数 / 調査音素数)

音素	単語発声	朗読発声 (文節)	朗読発声 (文)	自由発話
a	0.2% (4/1759)	3.7% (15/407)	8.4% (34/407)	17.6% (157/891)
i	2.3% (29/1277)	2.6% (7/269)	7.5% (20/265)	17.6% (88/500)
u	2.3% (33/1425)	6.9% (9/130)	22.7% (29/128)	27.9% (61/219)
e	0.7% (5/ 672)	8.3% (17/204)	10.3% (21/204)	12.4% (52/418)
o	0.0% (0/ 860)	9.5% (25/264)	12.0% (30/250)	15.4% (67/436)
p	40.0% (6/ 15)			100.0% (4/ 4)
t	4.1% (18/ 440)	3.4% (4/116)	2.6% (3/116)	6.7% (13/194)
k	13.4% (152/1137)	6.0% (14/233)	9.4% (22/233)	6.0% (20/336)
b	5.3% (11/ 207)	28.6% (6/ 21)	19.0% (4/ 21)	35.0% (7/ 20)
d	10.6% (18/ 170)	6.5% (7/107)	12.1% (13/107)	15.3% (30/196)
g	13.8% (35/ 253)	27.7% (13/ 47)	23.9% (11/ 46)	29.7% (35/118)
m	7.3% (35/ 482)	5.6% (6/107)	8.5% (9/106)	4.2% (8/190)
n	9.4% (25/ 265)	5.5% (7/128)	8.7% (11/126)	9.7% (23/237)
N	4.8% (21/ 439)	32.5% (13/ 40)	27.5% (11/ 40)	25.6% (11/ 43)
s	0.4% (2/ 524)	0.0% (0/ 73)	1.3% (1/ 76)	3.8% (6/159)
h	2.8% (6/ 215)	12.9% (8/ 62)	14.5% (9/ 62)	24.1% (33/137)
f	4.3% (2/ 47)			
z	6.2% (7/ 113)	55.6% (5/ 9)	87.5% (7/ 8)	61.9% (13/ 21)
j	4.8% (8/ 168)	7.1% (2/ 28)	7.1% (2/ 28)	13.0% (9/ 69)
r	7.2% (52/ 727)	6.7% (8/119)	11.8% (14/119)	9.9% (20/202)
w	11.4% (9/ 79)			66.7% (12/ 18)
y	6.7% (12/ 178)	6.7% (3/ 45)	25.0% (11/ 44)	38.7% (24/ 62)
ts	3.9% (7/ 181)	4.2% (1/ 24)	4.3% (1/ 23)	7.0% (3/ 43)
sh	0.6% (2/ 315)	0.0% (0/ 59)	3.4% (2/ 59)	2.1% (2/ 97)
ch	2.5% (3/ 121)	29.4% (5/ 17)	23.5% (4/ 17)	5.3% (3/ 57)
ky	0.0% (0/ 52)	0.0% (0/ 12)	0.0% (0/ 12)	0.0% (0/ 23)
by	66.7% (2/ 3)			
gy	7.7% (1/ 13)			
hy	0.0% (0/ 10)			
ny	87.5% (7/ 8)			
ry	4.9% (2/ 41)	88.9% (8/ 9)	77.8% (7/ 9)	100.0% (1/ 1)
total	4.2% (514/12196)	7.2% (183/2530)	11.0 % (276/2506)	15.0 % (702/4691)

表 23: 自由発話における冗長語の一覧

冗長語	出現回数	冗長語	出現回数	冗長語	出現回数
「あ」	604	「えーっとお」	1	「ぞの」	115
「あー」	268	「えーっとですね」	8	「そのー」	48
「あーっと」	2	「えーと」	466	「だか」	1
「あーと」	1	「えーとー」	4	「ちょっと」	8
「あーん」	5	「えーとですね」	3	「つ」	2
「ああ」	7	「えーまあ」	3	「で」	61
「あっ」	151	「えーん」	2	「でー」	13
「あっと」	1	「ええ」	13	「でぃ」	1
「あと」	1	「ええー」	1	「と」	77
「あなー」	1	「ええっと」	1	「とー」	11
「あの」	1809	「えっ」	22	「ねー」	1
「あのー」	2025	「えっーと」	4	「のー」	1
「あのーえー」	1	「えっと」	62	「は」	4
「あのう」	77	「えっとー」	11	「はあ」	1
「あのうー」	3	「えっとおー」	1	「はあー」	2
「あのと」	1	「えと」	47	「ははぁーん」	1
「あれ」	1	「えとー」	13	「ひ」	1
「あん」	1	「えへへっ」	1	「ふーん」	2
「い」	26	「えん」	1	「ま」	263
「いー」	58	「お」	59	「まー」	8
「いゃー」	1	「おー」	196	「まぁ」	186
「いやー」	2	「おーえー」	1	「まぁね」	1
「う」	23	「おっ」	2	「まあ」	176
「うー」	71	「ぐっ」	1	「まああのう」	1
「うーん」	26	「こう」	9	「まあまあ」	1
「うーんと」	2	「この」	9	「まっ」	5
「うっ」	1	「このー」	4	「も」	2
「うん」	7	「じゃ」	4	「もう」	1
「え」	1040	「じゃー」	1	「よ」	1
「えー」	3105	「じゃあ」	1	「りー」	1
「えーえ」	1	「す」	8	「わあ」	1
「えーちょっと	1	「すー」	2	「わっ」	1
「えーっ」	1	「すい」	1	「ん」	27
「えーって」	1	「すっ」	2	「んー」	19
「えーっと」	256	「せ」	1	「んっ」	1
「えーっとー」	2	「そ」	2	「んっと」	1
「えーっとえー」	1	「そう」	1	「んで」	1
				「んと」	2