

TR-IT-0275

音声タイプライタにおける  
子音の認識誤り傾向  
～ ng の脱落現象の解析 ～

中島 雄大 Takehiro Nakajima	山本 博史 Hirofumi Yamamoto	
匂坂 芳典 Yoshinori Sagisaka	高野 優 Masaru Takano	柘植 覚 Satoru Tsuge

1998 年 11 月 9 日

本稿では、音響モデルの高性能化を目指す前段階として、Multi-Class Composite N-gram を用いた音声タイプライタ用言語モデルによる音素認識実験の結果をもとに、音素の認識誤り傾向の調査を行った。特に本稿では、脱落数の多い子音 ng に対象を絞り、HMnet の構造 (使用されている状態、フレーム数)、音素の連鎖 (音素系列)、話者の相違に着目した認識誤り解析を行なった結果を報告する。

## 目次

1	はじめに	1
1.1	音素タイプライタによる音素認識実験	1
1.2	各子音の脱落	2
2	HMnet の構造からの脱落の解析	4
2.1	HMnet の状態	4
2.2	HMnet の状態の位置毎の適用フレーム数	4
2.3	状態のフレーム占有度と前後の音素への吸収	6
3	音素の連鎖からの解析	9
3.1	二項分布による脱落数の評価	9
3.2	音素列からの解析	11
4	話者からの解析	13
5	まとめ	15
	参考文献	16
付録 A	コンフィギュレーションファイル	17
A.1	ATRlattres のコンフィグファイル	17
A.2	atrFrameAlign.py のコンフィグファイル	18
付録 B	全話者の音素認識率	20
付録 C	HMnet の状態の正解数	21
付録 D	二項分布値	22
付録 E	r の音素系列	26
付録 F	全話者の位置毎のフレーム数	27

## 1 はじめに

現在の連続音声認識技術は、大量の音声データを用いて学習を行った音響モデルと、大量のテキストデータを用いて学習を行った言語モデルの融合により成り立っている。それらのモデルを用いて音声認識を行なう際、音響モデルから求められたスコアと、言語モデルから求められたスコアを考慮して認識結果を求めることが一般的である。なぜなら、人間は音声認識の際に、耳から入ってくる音声信号と共に、あらかじめ持っている言語や文法に関する知識を用いて音声の意味内容を理解しているためである。

本稿では、このうち音響モデルの高性能化を目指す前段階として、Multi-Class Composite N-gram[1][2]を用いた音素タイプライタ用言語モデルによる音素認識実験の結果を元に、音素の認識誤り傾向の調査を行なった。特に本稿では、脱落数の多い子音 ng に対象を絞り、HMnet の構造、音素系列、話者の違いに着目した認識誤り解析を行なった結果を報告する。

### 1.1 音素タイプライタによる音素認識実験

音素の誤り傾向としては話者の違いに起因するものとそうでないものが考えられる。そこでまず、話者の違いに起因するものに注目するため、過去に行なわれた単語認識実験 [3] による音素認識率をもとに話者のクラスタリングを行なった。

単語認識実験の際に使用された音響モデルの学習データは、ATR 自然発話旅行会話データベース [4] から男性 100 名、女性 130 名の音声を、評価データは、学習に用いていない男性 17 名、女性 25 名が発声した自然発話音声を利用している。特徴ベクトルは、サンプリング周波数 16kHz、フレーム長 20msec、フレーム周期 10msec で分析を行ない、発話毎に CMS[5] により正規化を行なった MFCC(12 次元) と対数パワーに、各一次回帰係数を加えた計 26 次元である。音響モデルのトポロジーは、ML-SSS アルゴリズム [6] により分割した 800 状態 5 混合の HMnet(音素環境依存 HMM) である。言語モデルは可変長 N-gram(品詞と可変長単語列の複合 N-gram)[7] を用い、単語グラフによるビームサーチ手法 [8] の一位候補を認識結果としている。

(実験結果は /dept1/work1/ResearchJ/V5/result/result.CMS.MFCC.GD を参照)

上記の単語認識実験により、認識性能の悪い男女各 5 名 (表 1) と認識性能の良い男女各 5 名 (表 2) を選択した。

表 1: badspeakers

male	female
TCC70109.A	TCS70004.B
TCC70201.A	TCS70010.A
TCC70212.A	TCS70020.A
TCC71008.A	TCS70047.A
TCS70034.A	TSC71013.A

表 2: goodspeakers

male	female
TAC70017.A	TAC70019.A
TAC70021.A	TAC70102.A
TAC70023.A	TAC70303.A
TAC70103.A	TCC71007.A
TCS70074.A	TCS70028.A

音素認識性能の良い話者と悪い話者を決定した後に、言語モデルを音素タイプライタ用言語モデルに変更し、全 42 話者に対する音素認識実験を行なった。表 3 に、表 1 と表 2 で選択した話者の会話 ID、音素認識率 (accuracy)、音素正解率 (%Correct)、会話中に出現した総音素数、挿入、脱落、置換誤りの数を示す。

表 3 より、認識性能の良い話者と悪い話者とは、音素認識率においては、性別に関わらず、約 2 ~ 3 倍の誤り率の開きがあることがわかる。また、単語認識実験の音素認識率と比較すると、表 1 の話者に関しては 10 % 前後認識性能が良くなっている。これは、言語モデルの制約による音素の置換・挿入誤りが減少したためであると考えられる。

巻末の付録の表 21 に、全 42 話者の音素認識実験の結果を掲載しておく。

表 3: 音素タイプライタ用言語モデルを使用した時の音素認識率

会話 ID	音素認識率	音素正解率	総音素数	挿入誤り	脱落誤り	置換誤り
TAC70017.A	93.93	94.33	247	1	9	5
TAC70021.A	92.54	93.71	429	5	13	14
TAC70023.A	92.37	94.22	485	9	18	10
TAC70103.A	88.09	89.17	277	3	14	16
TCS70074.A	92.43	94.74	304	7	7	9
good male	91.96	93.40	1742	25	61	54
TAC70019.A	88.72	90.00	390	5	18	21
TAC70102.A	93.35	96.01	526	14	7	14
TAC70303.A	95.62	96.93	457	6	3	11
TCC71007.A	89.47	91.73	617	14	19	32
TCS70028.A	94.32	95.17	352	3	8	9
good female	92.14	93.94	2342	42	55	87
TCC70109.A	82.45	85.64	376	12	25	29
TCC70201.A	67.16	70.71	338	12	52	47
TCC70212.A	77.60	80.36	616	17	57	64
TCC71008.A	78.80	80.26	684	10	76	59
TCS70034.A	82.43	86.26	626	24	38	48
bad male	78.41	81.25	2640	75	248	247
TCS70004.B	78.45	80.07	1109	18	117	104
TCS70010.A	78.77	80.82	391	8	41	34
TCS70020.A	78.75	81.29	433	11	33	48
TCS70047.A	87.50	90.25	400	11	16	23
TSC71013.A	77.20	80.93	886	33	78	91
bad female	79.31	81.83	3219	81	285	300
good	91.82	93.71	4084	77	116	141
bad	78.90	81.57	5859	156	533	547

## 1.2 各子音の脱落

3で示した音素認識の結果から、置換、脱落誤りが支配的であることがわかる。このうち脱落誤りのほうが単語認識結果にはより悪影響を与えることが多いと一般に言われている。そこで、本稿では認識誤り傾向の対象として、置換や挿入ではなく脱落をとりあげる。なお、本章以降で用いる脱落は、正解系列内の任意の対象音素とその前後の数音素(2音素)のフレームを認識系列から抽出した時に、そのフレーム内に対象音素が認識系列に現れていないことを指すものとする。言い替えれば、単純に「その音素が現れているか否か」ということに焦点を置いている。

表4に、対象とした音素と、男女計42名の認識系列における対象音素の脱落数(D)、正解系列における対象音素の出現回数(T)、対象音素の脱落率、N-gram学習時の対象音素の学習サンプル数を示す。また、表中の実線は学習サンプル数2000を区切るものである。

表4より、ng、r、jといった音素の脱落が多いことがわかる。これらの音素は、前後の母音に吸収されているために脱落数が多くなっていると考えられる。

以上の結果より、本稿では子音の中で最も脱落数の多いngに着目し、脱落の傾向の解析を行なう。

表 4: 各子音の脱落数

子音	脱落数 (D)	出現回数 (T)	脱落率 (D/T)	学習サンプル数
ng	110	446	24.66	2665
r	151	740	20.41	4360
g	95	491	19.35	2871
j	99	512	19.34	2993
h	73	387	18.86	2422
n	106	903	11.74	5232
d	85	792	10.73	4746
t	79	758	10.42	4302
m	59	720	8.19	4219
k	68	973	6.99	5696
sh	30	591	5.08	3401
s	25	958	2.61	5502
p	11	17	64.71	171
b	43	92	46.74	619
q	73	168	43.45	897
z	25	82	30.49	438
w	46	190	24.21	1208
zh	37	186	19.89	967
ts	29	146	19.86	760
ch	32	181	17.68	933

## 2 HMnet の構造からの脱落の解析

### 2.1 HMnet の状態

ng の脱落の現象を解析するために、まず、ng を表現している HMnet の状態に着目した。表 5 に、正解系列との Viterbi マッチングにおいて、ng がラベリングされているフレームにおける、HMnet の状態、その状態の HMM における位置 (pos)、正解系列での出現フレーム数 (T)、認識系列において正解系列と同一フレームに出現したフレーム数 (C) とその正解率を示す。本稿で用いた HMnet は、ng を表す HMM をすべて 4 状態で表現している。また、この HMnet において、ng に割り当てられていた状態数は 29 であった。その状態中で T が 100 以上のものを表 5 に掲げた。

また、表 6 に正解系列での状態 (CS) と同一フレームにおいて、誤認識された状態 (RS) の中で、フレーム数 (F) が 10 以上の状態と triphone の誤り例を示す。

表 5: HMnet の状態の正解数

状態番号	pos	C	T	rate
772	4	165	182	90.66
657	2	129	148	87.16
749	4	101	119	84.87
523	1	174	207	84.06
491	1	158	190	83.16
782	2	151	183	82.51
526	3	129	160	80.63
513	4	191	248	77.02
507	3	133	180	73.89
188	3	167	235	71.06
263	2	74	107	69.16
394	2	113	168	67.26
395	4	77	173	44.51

表 6: triphone の誤り例

CS	pos	RS	F	誤り例
507	3	395	20	ng d → ng n, ng m
395	4	386	13	ung n → une, uno, uni
263	2	732	10	ing n → ina, ini, ino
264	1	742	10	ong n → ona, ono
706	2	541	10	- ng - → ng n -
782	2	340	10	ang → an

表 6 より、ng と性質の近い n や m といった鼻音との誤りが多いことがわかる。これらの誤りが、表 5 での正解率を大きく劣化させているといえる。また、ng 内における状態の位置から、脱落数の多い状態として ng の中心を構成するものが多い傾向にある。

巻末の付録の表 22 に、ng に与えられている HMnet の全状態の正解数を掲載しておく。

### 2.2 HMnet の状態の位置毎の適用フレーム数

ng, n, m といった鼻音はフォルマントが現れることから、鼻音をモデル化する際、母音と同程度のフレーム数が必要であると考えられる。本章では、1 音素は 4 つの HMnet の状態を持つことに着目し、各状態の占有フレーム数から ng の脱落の現象を解析を行なう。

まず、全 42 話者の評価データ中に出現する ng 446 個を、先行音素、もしくは後続音素が - (pause) であるもの (17 個, 脱落 4 個) と - を含まないもの (429 個, 脱落 106 個) に分別した。そして、正解音素系列において ng を

構成している HMnet の4つの状態を抜きだし、それぞれの状態の位置毎に継続フレーム数と総フレーム数の平均を求めた。また、平均は認識結果において脱落が生じているものと、それ以外のものに分けて算出した。

表7に、認識系列で ng の脱落を生じている場合 (D) と脱落を生じていない場合 (not D) の、ng の前後に - を含む HMnet の状態の継続フレーム数と総フレーム数の平均を、表8に、ng の前後に - を含まない場合の結果を示す。

表7: pause を含む際の位置毎のフレーム数

	1	2	3	4	total
D	1.00	1.00	15.25	0.75	18.00
not D	16.00	23.31	5.54	0.00	44.85
average	12.47	18.06	7.82	0.18	38.53

表8: pause を含まない際の位置毎のフレーム数

	1	2	3	4	total
D	1.58	1.95	1.38	1.42	6.33
not D	1.51	1.56	1.33	1.42	5.83
average	1.53	1.66	1.34	1.42	5.95

表7より、認識の際に脱落を起こしている場合 (D) は、ng を構成する3番目の状態が、脱落を起こしていない場合 (not D) は2番目の状態が最もフレーム数が長いことがわかる。また、D では ng が第3 state に大部分を占有されているが、not D では単一状態に占有されていないことがわかる。

表9の上部に、男性話者17名(178個, 脱落48個)の位置毎の継続フレーム数と総フレーム数を、表9の下部に、女性話者25名(235個, 脱落55個)の結果を示す。以下では、-を含むデータを除いている。

表9: 性別における位置毎のフレーム数

male	1	2	3	4	total
D	1.50	1.38	1.35	1.38	5.60
not D	1.48	1.63	1.48	1.45	6.05
average	1.49	1.56	1.45	1.43	5.93
female	1	2	3	4	total
D	1.69	2.47	1.42	1.40	6.98
not D	1.54	1.50	1.25	1.41	5.69
average	1.57	1.73	1.29	1.40	6.00

表9からは男女間の特徴を表すような顕著な傾向は見受けられない。しかし、男女ともに第2, 3stateの継続時間長の差がDとnot Dでは大きいということが分った。

表 10 に、(1) 表 1 の話者 10 名 (計 133 個, 脱落 43 個)、(2) (1) 以外の男性 12 名、女性 20 名の計 32 名の平均値 (計 296 個, 脱落 63 個)、(3) 表 1 の男性 5 名の平均値 (計 64 個, 脱落 20 個)、(4) (3) 以外の男性 12 名の平均値 (計 94 個, 脱落 28 個)、(5) 表 1 の女性 5 名の平均値 (計 69 個, 脱落 23 個)、(6) (5) 以外の女性 20 名の平均値 (計 166 個, 脱落 32 個) についての結果を示す。各条件の 1 行目は認識系列において ng の脱落が生じた時の、2 行目は ng の脱落が生じていない時の継続フレーム数と総フレーム数の平均値である。

表 10: 話者クラスにおける位置毎のフレーム数

	1	2	3	4	total
(1)	1.56	1.95	1.40	1.51	6.42
	1.41	1.58	1.26	1.36	5.60
(2)	1.63	1.97	1.38	1.30	6.28
	1.56	1.55	1.39	1.45	5.95
(3)	1.60	1.15	1.35	1.50	5.60
	1.55	1.55	1.34	1.45	5.89
(4)	1.43	1.54	1.36	1.29	5.61
	1.45	1.67	1.56	1.45	6.14
(5)	1.52	2.65	1.43	1.52	7.13
	1.28	1.61	1.17	1.26	5.33
(6)	1.81	2.34	1.41	1.31	6.88
	1.63	1.46	1.28	1.46	5.82

表 10 より、ng の脱落を起こす際の第 2state について、男性話者に関しては正解系列よりフレーム数が短く、逆に女性話者に関してはフレーム数が長くなっていることがわかる。

### 2.3 状態のフレーム占有度と前後の音素への吸収

ng を含む評価データを認識する際、評価データ中に ng に特徴的な音響が現れていなければ完全に脱落して認識されると思われるが、正解系列で ng とラベリングされているフレームが認識系列において先行または後続の音素の一部としてラベリングされていたり、該当フレームにおいて正解系列の尤度と認識系列の尤度間に差がない場合は、脱落というよりむしろ前後の音素に吸収されている、もしくは同一の音素とみなされているといえる。

そこで、ng の脱落 (110 個) の中で、正解系列と認識系列を比較したときに認識系列において ng のみが脱落しているもの (53 個) を抽出し、正解系列において ng に占有されているフレームが、認識系列において先行・後続のどちらの音素により多く占有されているかを集計した。表 11 に先行音素ごとに集計した結果を、表 12 に後続音素ごとに集計した結果を示す。

表 11: 先行音素と音素境界の分類

先行音素	先行	後続	中間
-	1	0	0
a	3	7	1
i	8	4	3
u	1	4	1
e	3	2	3
o	7	5	0



表 12: 後続音素と音素境界の分類

後続音素	先行	後続	中間
o	1	0	1
k	1	0	0
s	2	0	0
sh	1	0	0
t	5	1	0
h	4	0	0
n	1	16	1
m	2	4	0
w	3	0	2
g	2	1	0
d	1	1	3

また、図1~図3に、正解系列と認識系列のフレーム毎の音響尤度値を示す。図中の縦の実線は正解系列における音素境界を、縦の破線は音素内の HMnet の状態の境界を表す。図1は先行音素 i に ng のフレームが吸収されている場合を、図2は後続音素 n に ng が吸収されている場合を、図3は e と w に ng が吸収されている場合を示している。

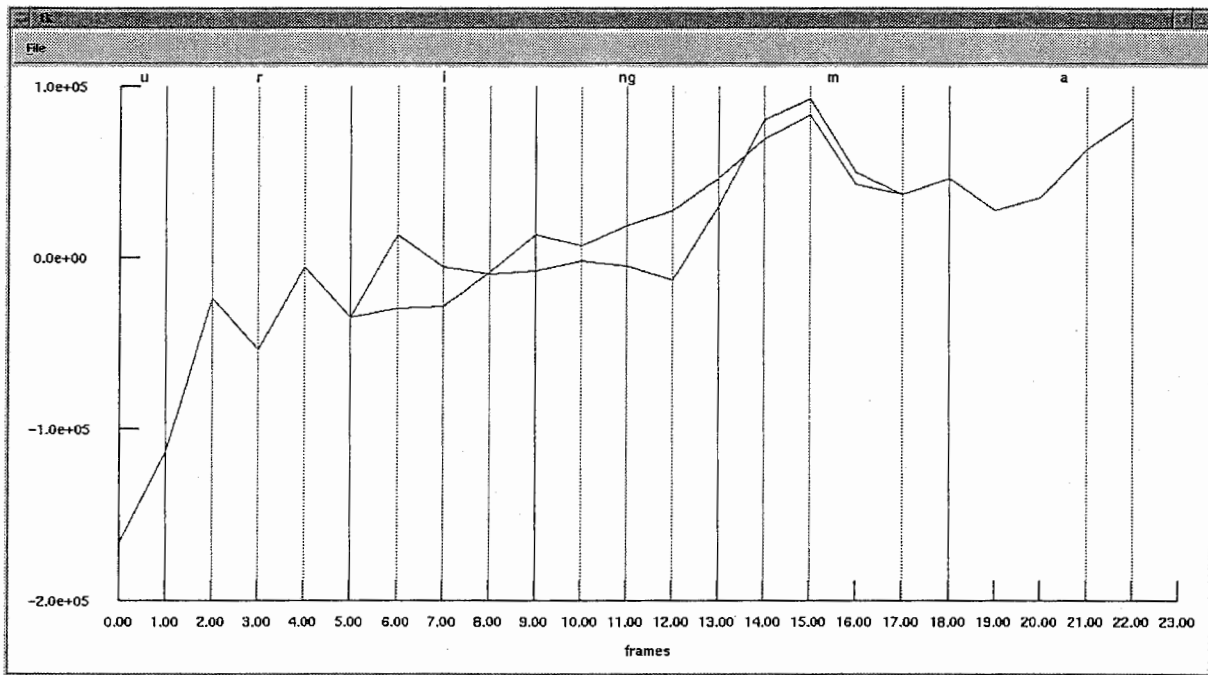


図 1: 先行音素への ng の吸収

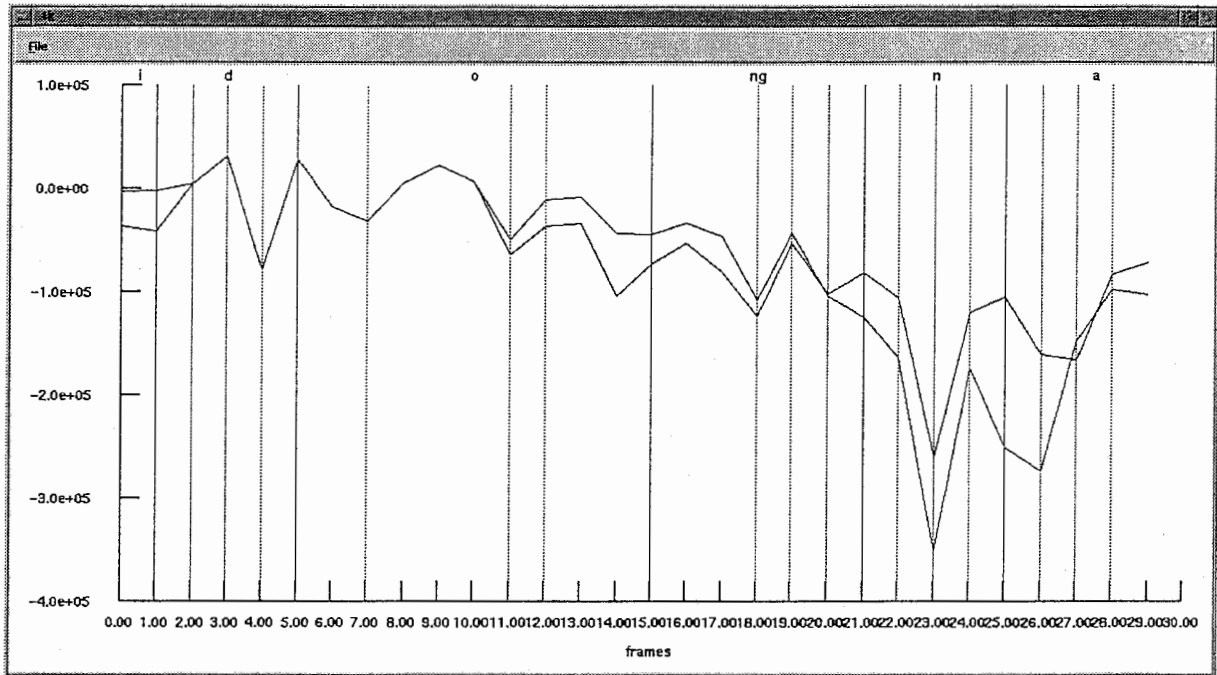


図 2: 後続音素への ng の吸収

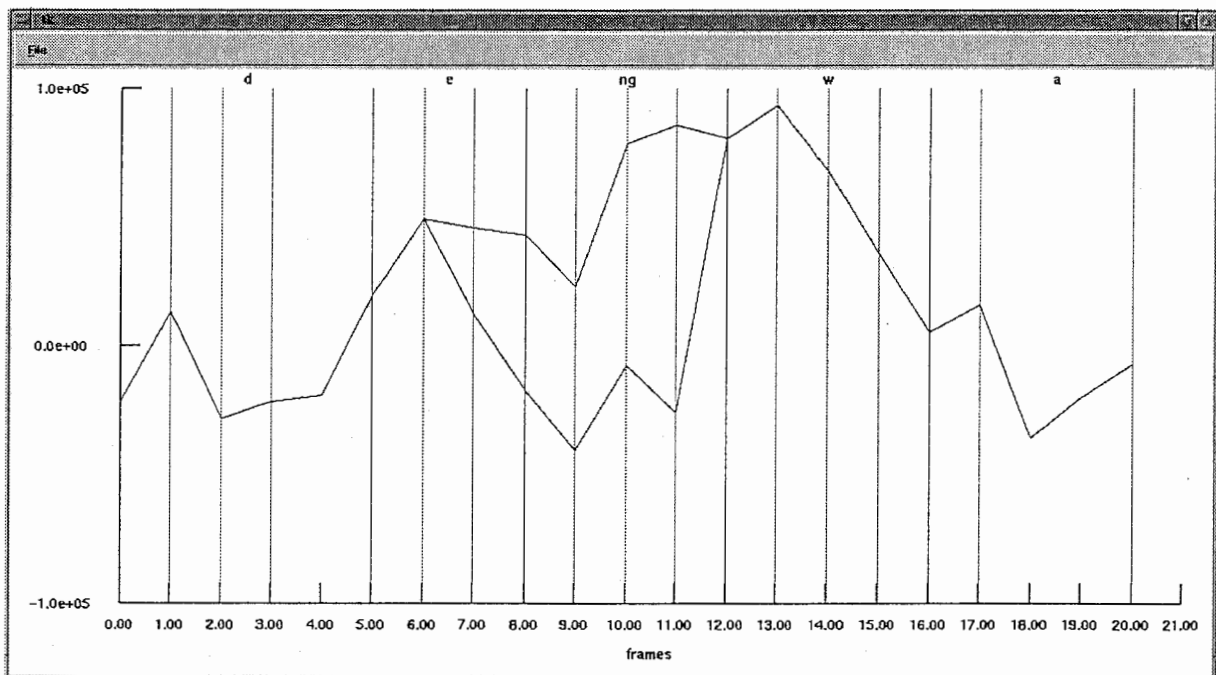


図 3: ng の脱落

表11からは、先行音素による ng の吸収の顕著な傾向は見受けられないが、i と o は ng を吸収する傾向があると思われる。表12より、音響的な特徴が似ている n や m といった鼻音に吸収されることが多いことがわかる。図1と図2は、正解系列と認識系列のグラフの形状が似ており、両者の尤度差が大きいとはいえない。しかし、図3では正解系列として ng に割りふられたフレームにおいて、尤度間に大きな差が出ており、ng の音響的な特徴量が現れていないと考えられる。

### 3 音素の連鎖からの解析

前章では、尤度値やフレーム数といったHMnetの構造に関する視点、言い替えれば音響的な視点から脱落の解析を行なった。本章では、テキストデータより得られる言語的な情報からngの脱落の解析を試みる。

#### 3.1 二項分布による脱落数の評価

表4の子音の中で、(bigramの)学習サンプル数や評価データ中での出現回数が多い子音は、統計的な解析を行なった際に、十分信頼できる値になると考えられる。そこで、本章では二項分布を用いて統計的にngの脱落の解析を行なう。

二項分布は以下の式により得られる。

$$P^i(1-P)^{(D-i)} \times \frac{D!}{i!(D-i)!} \quad (1)$$

上式で、 $P$ は学習サンプル中での対象音素のbigramの数をunigram数で割った値、 $D$ は認識系列における対象音素のunigramの脱落数、 $i$ は認識系列における対象音素のbigramの脱落数である。上式を $0 \rightarrow D$ 、または $D \rightarrow 0$ までの積分(今回は総和)をとることにより、対象音素のbigramの二項分布を得ることができる。

この時、あるbigramにおけるngの脱落数より求められる分布が<sup>s</sup>、二項分布の両端にあてはまるものであれば、平均的でないことになる。つまり、脱落数が0からの積分値が0.025未満の分布にあてはまったならば、そのbigramやngの脱落は統計的に極めて少ないことになり、逆に0.975を越える分布にあてはまった時は脱落が統計的には極めて多いことが示される。

表13に、ngのbigram、認識系列において対象のbigramが出現した時のngの脱落数( $d$ )、脱落数 $d$ の時の二項分布値、 $0 \rightarrow D$ の総和、 $D \rightarrow 0$ の総和とbigramの学習サンプル数を示す。表中で、二重線の上部が<sup>s</sup>ngの先行音素に着目した時、下部が<sup>s</sup>ngの後続音素に着目した時の値である。また、後続音素に関して、 $0 \rightarrow D$ の総和が0.975を境界として実線をひいた。

表 13: ng の bigram の二項分布

bigram	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
o ng	22	0.000572	0.999613	0.000959	261
- ng	2	0.017310	0.998764	0.018546	5
u ng	18	0.006752	0.993945	0.012807	247
e ng	17	0.095958	0.397752	0.698206	450
i ng	26	0.015286	0.039555	0.975731	844
a ng	25	0.009052	0.021581	0.987471	850
ng n	22	0.000178	0.999897	0.000281	239
ng m	10	0.002915	0.998715	0.004200	90
ng h	5	0.032237	0.986392	0.045846	47
ng w	7	0.024793	0.986265	0.038528	76
ng t	9	0.020977	0.985885	0.035093	108
ng j	2	0.158406	0.943209	0.215197	21
ng i	2	0.223689	0.872047	0.351642	30
ng sh	2	0.223689	0.872047	0.351642	30
ng b	4	0.155157	0.855444	0.299712	67
ng e	1	0.363185	0.799832	0.563353	20
ng s	4	0.179845	0.786641	0.393204	77
ng o	3	0.226228	0.681696	0.544532	69
ng z	1	0.317580	0.507046	0.810534	40
ng r	1	0.301391	0.468463	0.832929	43
ng p	1	0.290237	0.443857	0.846380	45
ng g	9	0.120175	0.380023	0.740152	256
ng k	5	0.149365	0.371546	0.777820	156
ng -	3	0.144840	0.271283	0.873557	119
ng zh	1	0.070360	0.087294	0.983066	97
ng d	18	0.000002	0.000002	0.999999	978

表 13より、先行音素に関しては u と o の後の ng 脱落数が、後続音素に関しては n と m の前の ng 脱落数が、統計的に見ても極めて多いことが確認できる。

なお、巻末の付録の表 23～表 29に、unigram の学習サンプルが 2000 以上で脱落率が 10 % 以上の子音 (r, g, h, j, n, d, t) について、同様の二項分布表を掲載した。

### 3.2 音素列からの解析

前節では学習サンプルから得られる統計値を基に脱落の解析を行なった。本節では、認識実験に使用される評価データにおいて、どのような音素列が発声された時に ng の脱落が生じるかという、言語的な視点から解析を行なう。表 14 に、脱落の少ない音素列について脱落数 (D)、出現回数 (T)、該音素列が含まれている単語の例を、表 15 に脱落の多い音素列に関して同じデータを示す。

表 14: ng の脱落の少ない音素列

音素列	脱落数	出現回数	例
a i n g d e	0	69	…ないんです
t a n g d e	1	10	…もらったんです
k i n g d e	1	10	現金で
n a n g d e	2	30	…なのですが
r o n g t o	2	13	フロント
g e n g k i	2	9	現金で
u i n g	0	20	チェックインできる
s e n g	1	34	すみません…
s h i n g	2	16	シングルで
j o n g	2	11	4号室
s a n g	4	37	3
n a n g	5	46	…なのですが

表 15: ng の脱落の多い音素列

音素列	脱落数	出現回数	例
d e n g w a	6	6	電話
d o n g n a	3	3	どんな…
h e n g m o	3	3	大変申し訳…
t o n g	7	7	ワシントン, シェラトン
b u n g	4	4	ふたり分
r a n g	4	4	レストラン
r i n g	7	8	ブルックリンまで
n i n g	6	8	…人の, 確認させて…
h u n g	4	5	5分ぐらい…
z h i n g	3	4	友人の, と

発声の際に ng が (「の」からの転化である場合のように) 本来 1 音節としてとらえられる音素列においては、ng の脱落が少なく、ng が前後の母音と融合して鼻母音化し、ng 単独のスペクトルが現れない音素列においては、ng の脱落が多いことがわかる。単語の観点からは、口語体で頻繁に使用される表現や数詞内に現れる ng の脱落は少ないといえる。しかし、今回の実験では、「ワシントン」「ブルックリン」といった、タスクに依存する固有名詞内の ng の脱落が多くみられた。また、口語体独特の表現では、「大変申し訳…」内の ng を認識することが、自由発話の音声認識を行なう上で必要になってくると思われる。

次に、ng と ng の後続音素との関連を調査するために、子音を表 16 のように分類した。

表 16: 子音の分類

クラス	音素
有声破裂音	b, d, g
無声破裂音	p, t, k
有声摩擦音	z, zh
無声摩擦音	s, sh, h
破擦音	ts, ch
鼻音	m, n
弾音	r
促音	q
半母音	w, j

表 16 の分類による各クラス毎、ng の先行母音と ng に後続する子音との ng の脱落数 (D)、出現回数 (T) を表 17 に示す。ただし、表 16 中、脱落数が 1 以下であった破擦音、弾音、促音は省略した。

表 17: 後続子音と ng の脱落数

後続子音	先行母音									
	a		i		u		e		o	
	D	T	D	T	D	T	D	T	D	T
有声破裂音	8	73	8	115	8	23	2	9	5	11
無声破裂音	2	14	4	16	0	0	4	28	4	21
有声摩擦音	0	13	1	4	0	0	0	0	1	4
無声摩擦音	2	3	2	10	0	0	1	3	6	9
鼻音	9	21	9	16	4	5	3	4	6	7
半母音	1	3	1	3	1	1	6	7	0	0

表 17 より、u、e、o を先行母音とし、鼻音を後続母音とした時の ng の脱落が多いことがわかる。また、e を先行母音、半母音を後続子音とした時や、o を先行母音、無声摩擦音を後続子音とした時も ng が脱落しやすい傾向にあることがわかる。

#### 4 話者からの解析

一般に、母音は話者によってフォルマントの分布が大きく異なるため、話者の性質がとらえやすい。音響的な性質が母音に近い ng においても、話者の違いにより、ng の脱落数が増減すると思われる。

そこで、表 18 に、表 1 であげた話者の会話 ID、ng の脱落数 (D) と ng の出現回数 (T) を、表 19 に、表 2 であげた話者の会話 ID、ng の脱落数 (D) と ng の出現回数 (T) を示す。

表 18: 話者毎の ng 脱落数 (badspeakers)

male			female		
会話 ID	D	T	会話 ID	D	T
TCC70109.A	2	11	TCS70004.B	13	28
TCC70201.A	2	4	TCS70010.A	3	9
TCC70212.A	5	18	TCS70020.A	1	8
TCC71008.A	7	20	TCS70047.A	0	7
TCS70034.A	5	14	TSC71013.A	7	18

表 19: 話者毎の ng 脱落数 (goodspeakers)

male			female		
会話 ID	D	T	会話 ID	D	T
TAC70017.A	1	6	TAC70019.A	0	5
TAC70021.A	2	9	TAC70102.A	0	11
TAC70023.A	2	8	TAC70303.A	1	6
TAC70103.A	2	8	TCC71007.A	3	11
TCS70074.A	2	7	TCS70028.A	0	9

表 18 と表 19 より、全音素の認識率の悪い話者は、ng の脱落に関しても、全体的に性能が悪くなっているといえる。

次に、2.2 節のようにフレーム数からの解析を行なった結果を示す。表 20 に、全話者 (42 名) の中で ng の脱落数が 5 以上の話者 (男性 5 名, 女性 3 名) の会話 ID と脱落数 (D)、出現回数 (T)、正解系列での HMnet の状態の位置毎の継続フレーム数と総フレーム数の平均を示す。また、各話者の 2 行目に脱落が生じなかった時の数と、継続フレーム数、総フレーム数の平均を示す。

表 20: 位置毎のフレーム数 (脱落数の多い話者)

会話 ID	D	T	1	2	3	4	total
TAC70022.A	7	19	1.43	1.29	1.29	1.14	5.14
	12	19	1.92	3.00	2.83	1.25	9.00
TCC71008.A	7	20	1.29	1.29	1.14	1.14	4.86
	13	20	4.85	1.00	1.08	1.31	8.23
TCC70307.A	5	13	1.40	1.00	1.60	1.80	5.80
	8	13	1.38	1.50	1.62	2.00	6.50
TCS70034.A	5	14	2.00	1.00	1.00	1.60	5.60
	9	14	2.22	3.00	2.89	1.89	10.00
TCC70212.A	5	18	2.00	1.00	2.00	1.60	6.60
	13	18	1.62	1.23	1.31	1.38	5.54
TCS70004.B	13	28	1.46	1.77	1.31	1.38	5.92
	15	28	1.13	1.67	1.20	1.13	5.13
TSC71013.A	7	18	1.86	5.00	2.57	1.57	11.00
	11	18	1.09	1.55	1.18	1.36	5.18
TCC71001.B	5	17	1.60	1.80	1.00	1.40	5.80
	12	17	6.42	3.33	1.42	1.25	12.42

各話者の2行目に示した、ngの脱落が生じなかった時の継続フレーム数と総フレーム数は、ngを発声する際の各話者の特徴を表しているといえる。脱落を起こしている場合は、ある状態が占めるフレーム数の分布が大きくなっている。

巻末の付録の表32と表33に、全42話者のHMnetの状態毎のフレーム数(正解と脱落)を掲載する。



## 5 まとめ

本稿では、自然発話音声認識のための音響モデルの高性能化を目的とし、その前段階として音素タイプライタを用いた音素認識実験の結果を基に、ng の脱落に特化した誤りの解析を行なった。ng の脱落の解析の視点として、本稿では、

1. HMnet の構造
2. 音素の連鎖
3. 話者の相違

に着目した。以上の視点からの解析の結果、

1. ng を構成している HMnet の状態からは、
  - ng と同一音素クラスの n や m といった鼻音の状態との誤りが多く、
  - 状態のフレーム占有度より、ng の音響的な特徴はとれていても後続する鼻音に吸収される傾向が強いことが確認された。
2. 音素の連鎖からは、
  - 実際の発声の際に ng が前後の母音と共に鼻母音化するかどうか、ng の脱落の現象に影響することが確認された。
3. 話者の相違からは、
  - 性別や話者によって、ng を構成する HMnet の状態のフレーム数の分布が変化することが確認された。

本稿の主目的である音響モデルの高性能化のために、

- ng に対応する HMnet の各状態の継続フレーム数を全体的に増やす
- 話者の特徴を表現できるように HMnet の各状態の継続フレーム数を変化させる

というように、継続時間の問題に対して state skip のみではなく他の手法による対応を行なう必要があると思われる。

また、対象音素が1音節となり得るかを調査した上で、

- 対照音素周辺の音素連鎖により HMnet の各状態の継続フレーム数を変化させる

必要があると思われる。

## 参考文献

- [1] 山本博史, 匂坂芳典. 接続の方向性を考慮した多重クラス N-gram モデル. 音講論, pp. 75-76, September 1998.
- [2] 山本博史. 多重クラス N-gram による効率的言語モデル表現. 音講論, pp. 77-78, September 1998.
- [3] 柘植覚, 内藤正樹, H. Singer, 深田俊明, 高野優. 連続音声認識用音響モデル (ResearchJ V5). Technical Report TR-IT-0266, ATR, July 1998.
- [4] A. Nakamura, S. Matsunaga, T. Shimizu, M. Tonomura, and Y. Sagisaka. Japanese speech databases for robust speech recognition. In *Proc. ICSLP*, pp. 2199-2202, Philadelphia, 1996.
- [5] F. H. Liu, R. M. Stern, X. Huang, and A. Acero. Efficient cepstral normalization for robust speech recognition. In *Proc. DARPA Workshop*, pp. 69-74, March 1993.
- [6] M. Ostendorf and H. Singer. HMM topology design using maximum likelihood successive state splitting. *Computer Speech and Language*, Vol. 11, No. 1, pp. 17-41, 1997.
- [7] 政瀧浩和, 松永昭一, 匂坂芳典. 品詞および可変長単語列の複合 N-gram の自動生成. 信学論 (D-II), Vol. J81-D-II, No. 9, pp. 1929-1936, September 1998.
- [8] 清水徹, 山本博史, 政瀧浩和, 松永昭一, 匂坂芳典. 大語い連続音声認識のための単語仮説数削減. 信学論 (D-II), Vol. J79-D-II, No. 12, pp. 2117-2124, December 1996.

## 付録 A コンフィギュレーションファイル

1.1節の表3の結果を求めるために使用した ATRlattres のコンフィグファイルを付録A.1に、また、ビタビ系列のバックトレースにより正解・認識系列のフレーム毎の HMnet の状態を求めるために使用した atrFrameAlign.py のコンフィグファイルを付録A.2に示す。

## A.1 ATRlattres のコンフィグファイル

```
I/Ocontrol:rpcTable=
I/Ocontrol:rpcNumber=3
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputParamType=
I/Ocontrol:outputParamSize=
I/Ocontrol:outputFormat=Lattice
I/Ocontrol:inputEOFexit=ON
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:inputParamType=float
I/Ocontrol:inputParamSize=26
I/Ocontrol:inputFormat=FrameSync
I/Ocontrol:inputFd=/dept1/work1/ResearchJ/V5/data/CMS_MFCC/(conversation ID).FSYNC
ATRlattice:lexicon=/dept1/work1/ResearchJ/lmodel/19980806/LEX.P
ATRlattice:aname=/dept1/work1/ResearchJ/V5/amodel/AM.M.CMS.bin,/dept1/work1/ResearchJ/V5/amodel/AM.F.CMS.bin
ATRlattice:active_model=1,2
ATRlattice:lmscale=7.000000,11.000000
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=Multi-Class-2,/dept1/work1/ResearchJ/lmodel/19980806/LM.P
ATRlattice:beam=75.000000,75.000000
ATRlattice:work_area=300,50
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=-
ATRlattice:dimension=26
ATRlattice:state_skip=ON,75000
ATRlattice:phone_boundary=OFF
ATRlattice:word_boundary_skip=2
ATRlattice:word_merge=all
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
ATRlattice:amscale=1.000000
# ATRlattice:UTT_END_delay=70

ATRresult:minimum_utt=0
ATRresult:dp_unit=FILE
ATRresult:dp_weight=1.0,1.0,1.0
ATRresult:pause_symbol=-
ATRresult:UTT_END=6
ATRresult:UTT_START=5
ATRresult:re_beam=
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:lattice_out=/dev/null
ATRresult:answer=/dept1/work1/ResearchJ/V5/data/CMS_MFCC/(conversation ID).TRS
```

## A.2 atrFrameAlign.py のコンフィグファイル

```
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=160
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NoHeader
I/Ocontrol:outputParamSize=26
I/Ocontrol:outputParamType=float
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian
```

```
ATRwavecut:pause_symbol=-
          ATRwavecut:PausePeriod=NOT
ATRwavecut:SamplingFrequency=16000.0
```

```
ATRwave2cep:Preemphasis=0.98
ATRwave2cep:FrameLength=20
ATRwave2cep:FrameShift=10
ATRwave2cep:SamplingFrequency=16000
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:LpcOrder=16
ATRwave2cep:CepstrumOrder=12
ATRwave2cep:FrequencyWarping=mel
ATRwave2cep:FilterBankOrder=16
ATRwave2cep:CutoffLowFrequency=0
ATRwave2cep:CutoffHighFrequency=8000
ATRwave2cep:AnalysisType=fft
          ATRwave2cep:DebuggingLevel=0
```

```
ATRcep2para:CepstrumOrder=12
ATRcep2para:LDA=
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:rho=1.0
ATRcep2para:OutputParameter=pow+cep(12)+dpow+dcep(12)
```

```
ATRLattice:lexicon=/dept1/work1/ResearchJ/lmodel/19980806/LEX.P
ATRLattice:amname=/dept1/work1/ResearchJ/V5/amodel/AM.M.CMS.bin,/dept1/work1/ResearchJ/V5/amodel/AM.F.CMS.bin
ATRLattice:active_model=1,2
ATRLattice:lmscale=7.000000,11.000000
ATRLattice:wpenalty=0,0
ATRLattice:ngram=Multi-Class-2,/dept1/work1/ResearchJ/lmodel/19980806/LM.P
ATRLattice:beam=75.000000,75.000000
ATRLattice:work_area=300,50
ATRLattice:frame_shift=10
ATRLattice:pause_symbol=-
ATRLattice:dimension=26
ATRLattice:state_skip=ON,75000
ATRLattice:phone_boundary=OFF
ATRLattice:word_boundary_skip=2
ATRLattice:word_merge=all
ATRLattice:UTT_START=5
ATRLattice:UTT_END=6
ATRLattice:backward_frame=-1
ATRLattice:amscale=1.000000
```

---

```
ATRviterbi:amname=/dept1/work1/ResearchJ/V5/amodel/AM.M.CMS.bin,/dept1/work1/ResearchJ/V5/amodel/AM.F.CMS.bin
ATRviterbi:active_model=all
ATRviterbi:beam=0
ATRviterbi:frame_shift=10
ATRviterbi:dimension=26
ATRviterbi:boundary=FRAME
```

## 付録 B 全話者の音素認識率

1.1節で行なった、Multi-Class Composite N-gram を用いた全 42 話者に対する音素認識実験の結果を表 21 に示す。表中には、会話 ID、音素認識精度、音素正解率、会話中に出現した総音素数、挿入、脱落、置換誤りの数を示してある。

表 21: 全話者の音素認識率

会話 ID	accuracy	correct	total	Ins	Del	Sub
TAC70016.A	83.05	85.06	348	7	25	27
TAC70017.A	93.93	94.33	247	1	9	5
TAC70021.A	92.54	93.71	429	5	13	14
TAC70022.A	77.21	80.94	509	19	49	48
TAC70023.A	92.37	94.22	485	9	18	10
TAC70103.A	88.09	89.17	277	3	14	16
TAC70202.A	79.65	80.18	570	3	65	48
TAC70304.A	78.93	83.91	261	13	18	24
TCC70109.A	82.45	85.64	376	12	25	29
TCC70201.A	67.16	70.71	338	12	52	47
TCC70212.A	77.60	80.36	616	17	57	64
TCC70307.A	84.51	87.15	568	15	38	35
TCC71008.A	78.80	80.26	684	10	76	59
TCS70034.A	82.43	86.26	626	24	38	48
TCS70055.A	85.23	86.56	677	9	51	40
TCS70070.A	80.67	83.33	300	8	25	25
TCS70074.A	92.43	94.74	304	7	7	9
TAC70015.A	83.00	85.21	453	10	26	41
TAC70019.A	88.72	90.00	390	5	18	21
TAC70101.A	95.50	96.17	444	3	10	7
TAC70102.A	93.35	96.01	526	14	7	14
TAC70201.A	89.22	92.02	501	14	17	23
TAC70203.A	87.83	90.11	526	12	27	25
TAC70301.A	85.23	87.92	447	12	24	30
TAC70303.A	95.62	96.93	457	6	3	11
TCC70103.A	90.16	92.08	366	7	10	19
TCC71001.B	80.90	82.64	864	15	85	65
TCC71007.A	89.47	91.73	617	14	19	32
TCC71016.A	82.55	87.78	573	30	22	48
TCC71035.A	88.74	90.88	373	8	14	20
TCS70004.B	78.45	80.07	1109	18	117	104
TCS70010.A	78.77	80.82	391	8	41	34
TCS70013.A	90.75	93.00	400	9	12	16
TCS70020.A	78.75	81.29	433	11	33	48
TCS70023.A	88.48	92.03	903	32	37	35
TCS70025.A	90.31	91.63	454	6	16	22
TCS70028.A	94.32	95.17	352	3	8	9
TCS70047.A	87.50	90.25	400	11	16	23
TCS70059.A	89.66	90.50	358	3	18	16
TCS70082.A	88.04	89.60	644	10	25	42
TSC71005.B	86.84	89.85	532	16	15	39
TSC71013.A	77.20	80.93	886	33	78	91
male	82.90	85.19	7615	174	580	548
female	86.25	88.56	13399	310	698	835
both	85.03	87.34	21014	484	1278	1383

## 付録 C HMnet の状態の正解数

表 22 に正解系列において ng がラベリングされているフレーム内で使用されている HMnet の全状態、その状態の正解系列の ng 内での位置 (pos)、正解系列での出現フレーム数 (T)、認識系列において正解系列と同一フレームに出現したフレーム数 (C) と正解率を示す。

表 22: HMnet の状態の正解数

状態番号	C	T	rate
772	165	182	90.66
657	129	148	87.16
17	52	49	85.71
749	101	119	84.87
750	28	33	84.85
523	174	207	84.06
491	158	190	83.16
782	151	183	82.51
526	129	160	80.63
326	45	58	77.59
513	191	248	77.02
548	39	52	75.00
507	133	180	73.89
262	69	97	71.13
188	167	235	71.06
263	74	107	69.16
185	43	63	68.25
490	39	58	67.24
394	113	168	67.26
195	38	57	66.67
376	36	59	61.02
281	42	73	57.53
264	38	84	45.24
395	77	173	44.51
105	29	76	38.16
800	14	44	31.82
705	20	76	26.32
706	10	40	25.00
641	7	32	21.86

## 付録 D 二項分布値

3.1節で行なった、二項分布による脱落数の評価を、1.2節の表4の音素の中で、音素脱落率が10%以上の音素(r,g,h,j,n,d,t)に対する bigram、認識系列において対象の bigram が出現した時の各音素の脱落数(d)、脱落数dの時の二項分布値、 $0 \rightarrow D$ の総和、 $D \rightarrow 0$ の総和と bigram の学習サンプル数を表23~表29示す。

表 23: r の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
- r	6	0.007686	0.997549	0.010138	52
i r	23	0.012324	0.984078	0.028246	435
u r	26	0.045893	0.895249	0.150644	606
a r	50	0.059373	0.739513	0.319859	1347
ng r	1	0.335891	0.558077	0.777814	43
e r	20	0.056616	0.189730	0.866885	704
o r	25	0.001237	0.002575	0.998662	1151
r a	50	0.000016	0.999987	0.000029	830
r j	6	0.041994	0.975195	0.066800	82
r u	35	0.024841	0.947559	0.077282	798
r i	28	0.035342	0.117227	0.918115	999
r e	25	0.014234	0.037607	0.976627	995
r o	7	0.000034	0.000046	0.999988	656

表 24: g の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
- g	12	0.000000	1.000000	0.000000	56
a g	18	0.000097	0.999957	0.000141	204
o g	17	0.012622	0.987602	0.025019	305
i g	14	0.104239	0.732742	0.371497	373
ng g	9	0.137074	0.642630	0.494444	256
u g	16	0.039303	0.101781	0.937523	644
e g	9	0.000000	0.000000	1.000000	989
g e	13	0.000000	1.000000	0.000000	31
g i	7	0.000106	0.999985	0.000121	33
g j	1	0.140444	0.987836	0.152608	5
g u	12	0.041285	0.953952	0.087333	235
g o	19	0.077731	0.278634	0.799097	664
g a	43	0.000012	0.000021	0.999992	1903



表 25: j の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
ij	8	0.008046	0.996429	0.011617	90
nj	2	0.040743	0.995078	0.045665	10
hj	4	0.033331	0.989233	0.044099	39
rj	6	0.038777	0.978259	0.060518	82
uj	9	0.067395	0.924962	0.142433	176
aj	9	0.095484	0.863188	0.232296	199
chj	9	0.102286	0.843550	0.258736	205
ngj	1	0.350906	0.840554	0.510352	21
pj	1	0.365845	0.679738	0.686108	34
zhj	17	0.093636	0.368514	0.725121	558
kj	4	0.151464	0.331329	0.820135	166
shj	15	0.076955	0.244053	0.832902	538
oj	9	0.075592	0.185996	0.889596	368
-j	1	0.044366	0.053618	0.990748	135
ej	4	0.028016	0.043817	0.984199	269
ja	36	0.004626	0.994248	0.010378	761
ju	29	0.031455	0.941263	0.090192	690
jo	34	0.000225	0.000423	0.999802	1542

表 26: h の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
ih	15	0.026650	0.971232	0.055418	191
ah	13	0.087966	0.840092	0.247874	207
ngh	3	0.212295	0.783120	0.429174	47
uh	6	0.161546	0.496679	0.664867	131
eh	4	0.150909	0.325794	0.825115	111
-h	7	0.121000	0.314936	0.806064	176
oh	25	0.059686	0.193502	0.866184	573
hj	8	0.000021	0.999997	0.000024	39
hi	17	0.000097	0.999961	0.000137	210
ho	24	0.000300	0.999819	0.000481	401
he	6	0.167283	0.631964	0.535319	194
hu	6	0.163478	0.516197	0.647281	217
ha	12	0.000000	0.000000	1.000000	1361

表 27: n の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
e n	11	0.019217	0.984897	0.034320	288
- n	5	0.063665	0.962510	0.101156	120
a n	27	0.058079	0.853997	0.204082	1123
ng n	5	0.179557	0.634090	0.545467	239
o n	32	0.075476	0.363547	0.711929	1670
i n	13	0.082514	0.253566	0.828948	781
u n	13	0.027346	0.061935	0.965412	950
n o	49	0.006808	0.989603	0.017204	1872
n j	1	0.165728	0.982175	0.183553	10
n a	28	0.027614	0.950999	0.076615	1061
n i	23	0.091904	0.631078	0.460826	1095
n e	5	0.000000	0.000000	1.000000	1189

表 28: d の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
i d	17	0.002167	0.998572	0.003595	461
- d	8	0.010126	0.995251	0.014875	176
a d	14	0.092469	0.804482	0.287986	649
u d	7	0.155528	0.601401	0.554127	384
o d	16	0.051383	0.140640	0.910743	1141
e d	11	0.046363	0.107388	0.938974	874
ng d	12	0.033428	0.074816	0.958611	978
d a	30	0.000000	1.000000	0.000000	461
d o	20	0.048602	0.915094	0.133508	867
d e	35	0.000000	0.000000	1.000000	3391

表 29: t の bigram の二項分布

biphone	d	値	$\sum$ from 0	$\sum$ from D	学習サンプル数
- t	13	0.000000	1.000000	0.000000	106
a t	14	0.026460	0.972589	0.053871	472
ng t	4	0.091862	0.948066	0.143796	108
q t	9	0.139883	0.599774	0.540110	477
e t	7	0.154669	0.545295	0.609373	393
o t	9	0.138471	0.529213	0.609258	504
u t	4	0.108146	0.207256	0.900890	350
i t	19	0.000216	0.000358	0.999858	1818
t e	21	0.043404	0.927395	0.116009	881
t a	33	0.089516	0.482695	0.606821	1836
t o	25	0.062129	0.210765	0.851364	1577

## 付録 E rの音素系列

1.2節の表4において、ngとともに脱落率が20%をこえているrに関しても、3.2節と同様の調査をおこなった。表30にrの脱落の少ない音素列、脱落数(D)、出現回数(T)、音素列を使用する単語列を、表31にrの脱落の多い音素列に関するデータを示す。

表 30: rの脱落の少ない音素列

音素列	脱落数	出現回数	例
karim	0	44	わかりました
sored	0	25	それでは…
chira	0	19	こちらは…
goro	0	17	10時頃から
ari	0	16	ありがとう
jorosh	1	46	よろしく
kuri	1	15	送りたい
sore	2	48	それでは
maru	2	19	505号室
gurai	2	16	…ぐらい
kered	4	59	けれども

表 31: rの脱落の多い音素列

音素列	脱落数	出現回数	例
nara	7	9	…しなければならなく
mora	7	10	…してもらえませんか
guru	6	6	シングルで
tsure	6	7	失礼ですが…
tara	6	17	あったら持って…
naruru	6	9	…になるんですか
shir	5	7	お知らせ

## 付録 F 全話者の位置毎のフレーム数

全話者(42名)の会話IDと脱落数(D)、出現回数(T)、正解系列でのHMnetの各状態の継続フレーム数と総フレーム数の平均について、ngの脱落が生じなかった場合を表32に、脱落が生じた場合を表33に示す。

表 32: 全話者の位置毎のフレーム数(正解)

会話 ID	C	T	1	2	3	4	total
TAC70016.A	12	13	2.58	4.58	1.75	0.92	9.83
TAC70017.A	5	6	1.80	1.20	1.20	1.00	5.20
TAC70021.A	7	9	1.14	1.14	1.00	1.71	5.00
TAC70022.A	12	19	1.92	3.00	2.83	1.25	9.00
TAC70023.A	6	8	1.33	1.17	1.17	1.50	5.17
TAC70103.A	6	8	1.33	2.50	1.33	2.00	7.17
TAC70202.A	9	11	1.89	5.78	1.11	1.11	9.89
TAC70304.A	8	9	9.00	5.38	6.88	1.75	23.00
TCC70109.A	9	11	1.11	1.33	1.11	1.00	4.56
TCC70201.A	2	4	1.00	1.00	1.00	1.50	4.50
TCC70212.A	13	18	1.62	1.23	1.31	1.38	5.54
TCC70307.A	8	13	1.38	1.50	1.62	2.00	6.50
TCC71008.A	13	20	4.85	1.00	1.08	1.31	8.23
TCS70034.A	9	14	2.22	3.00	2.89	1.89	10.00
TCS70055.A	10	12	1.90	1.70	1.10	1.00	5.70
TCS70070.A	4	5	1.25	1.00	1.00	1.00	4.25
TCS70074.A	5	7	1.20	1.00	1.00	1.40	4.60
TAC70015.A	6	9	1.67	1.33	1.17	1.50	5.67
TAC70019.A	5	5	1.20	1.20	1.20	1.00	4.60
TAC70101.A	8	9	1.38	1.25	1.25	1.62	5.50
TAC70102.A	11	11	1.36	6.55	1.91	1.27	11.09
TAC70201.A	5	6	2.00	1.20	1.20	2.00	6.40
TAC70203.A	13	17	1.46	1.62	1.00	1.31	5.38
TAC70301.A	5	8	2.20	11.60	1.80	1.40	17.00
TAC70303.A	5	6	1.40	1.00	1.20	1.20	4.80
TCC70103.A	7	8	1.57	1.29	1.57	1.43	5.86
TCC71001.B	12	17	6.42	3.33	1.42	1.25	12.42
TCC71007.A	8	11	1.88	2.62	1.50	1.38	7.38
TCC71016.A	10	12	1.50	6.40	1.30	1.30	10.50
TCC71035.A	5	6	3.20	2.20	1.20	2.00	8.60
TCS70004.B	15	28	1.13	1.67	1.20	1.13	5.13
TCS70010.A	6	9	1.83	1.50	1.33	1.00	5.67
TCS70013.A	6	6	3.17	2.00	1.33	1.17	7.67
TCS70020.A	7	8	1.00	1.29	1.14	1.14	4.57
TCS70023.A	11	13	1.64	1.18	1.27	1.18	5.27
TCS70025.A	5	7	1.00	1.40	1.40	2.80	6.60
TCS70028.A	9	9	1.33	1.44	1.00	1.56	5.33
TCS70047.A	7	7	1.71	2.00	1.00	1.71	6.43
TCS70059.A	4	7	1.50	1.25	1.50	1.00	5.25
TCS70082.A	10	13	1.30	1.40	1.10	1.20	5.00
TSC71005.B	7	9	1.29	1.00	1.00	1.14	4.43
TSC71013.A	11	18	1.09	1.55	1.18	1.36	5.18

表 33: 全話者の位置毎のフレーム数 (脱落)

会話 ID	D	T	1	2	3	4	total
TAC70016.A	1	13	1.00	1.00	1.00	1.00	4.00
TAC70017.A	1	6	1.00	10.00	1.00	1.00	13.00
TAC70021.A	2	9	2.50	1.50	2.50	2.00	8.50
TAC70022.A	7	19	1.43	1.29	1.29	1.14	5.14
TAC70023.A	2	8	1.00	1.00	1.00	1.00	4.00
TAC70103.A	2	8	1.00	2.00	1.00	1.00	5.00
TAC70202.A	2	11	3.00	1.00	2.00	1.00	7.00
TAC70304.A	1	9	1.00	2.00	1.00	2.00	6.00
TCC70109.A	2	11	1.00	1.50	1.50	2.00	6.00
TCC70201.A	2	4	1.00	1.00	1.00	1.00	4.00
TCC70212.A	5	18	2.00	1.00	2.00	1.60	6.60
TCC70307.A	5	13	1.40	1.00	1.60	1.80	5.80
TCC71008.A	7	20	1.29	1.29	1.14	1.14	4.86
TCS70034.A	5	14	2.00	1.00	1.00	1.60	5.60
TCS70055.A	2	12	1.00	1.00	1.00	1.00	4.00
TCS70070.A	1	5	1.00	1.00	1.00	1.00	4.00
TCS70074.A	2	7	1.00	1.00	1.00	1.00	4.00
TAC70015.A	3	9	4.00	3.33	1.67	1.00	10.00
TAC70019.A	0	5	0.00	0.00	0.00	0.00	0.00
TAC70101.A	1	9	1.00	1.00	1.00	1.00	4.00
TAC70102.A	0	11	0.00	0.00	0.00	0.00	0.00
TAC70201.A	1	6	4.00	6.00	1.00	4.00	15.00
TAC70203.A	4	17	1.00	1.50	1.00	2.50	6.00
TAC70301.A	3	8	1.67	1.67	1.00	1.00	5.33
TAC70303.A	1	6	1.00	1.00	1.00	1.00	4.00
TCC70103.A	1	8	1.00	5.00	1.00	1.00	8.00
TCC71001.B	5	17	1.60	1.80	1.00	1.40	5.80
TCC71007.A	3	11	1.33	3.00	19.33	1.33	25.00
TCC71016.A	2	12	1.50	1.00	1.00	1.00	4.50
TCC71035.A	1	6	1.00	1.00	1.00	1.00	4.00
TCS70004.B	13	28	1.46	1.77	1.31	1.38	5.92
TCS70010.A	3	9	1.00	1.00	1.00	1.00	4.00
TCS70013.A	0	6	0.00	0.00	0.00	0.00	0.00
TCS70020.A	1	8	1.00	1.00	2.00	3.00	7.00
TCS70023.A	2	13	1.00	1.00	3.50	1.00	6.50
TCS70025.A	2	7	1.00	2.00	1.00	1.00	5.00
TCS70028.A	0	9	0.00	0.00	0.00	0.00	0.00
TCS70047.A	0	7	0.00	0.00	0.00	0.00	0.00
TCS70059.A	3	7	1.00	1.00	1.00	1.67	4.67
TCS70082.A	3	13	2.00	1.00	1.00	0.67	4.67
TSC71005.B	2	9	3.00	7.50	2.00	2.00	14.50
TSC71013.A	7	18	1.86	5.00	2.57	1.57	11.00