

表 3.36: タスク効率の主観評価と対話時間

(時間の単位は [sec])

	対話時間	一発話当たりの対話時間	無音時間	一発話当たりの無音時間
Correl	-0.63	-0.44	-0.52	-0.42

3.4.2 サブタスク毎の客観量

ここでは、前述の客観量について、検索能力の影響を取り除くために各対話におけるサブタスク数（検索回数）によって正規化したものと、各効率の主観評価との相関分析の結果を示す。

(1) 発話数

まず、対話効率およびタスク効率の主観評価とサブタスク数で正規化した発話数との相関を調べた結果を表 3.37 に示す。

表 3.37: 効率の主観評価とサブタスク数で正規化した発話数

対話番号	N	S 発話数	S 発話数/N	U 発話数	U 発話数/N	総発話数	総発話数/N
1	2	10	5.0	6	3.0	16	8.0
2	5	20	4.0	12	2.4	32	6.4
3	1	6	6.0	5	5.0	11	11.0
4	5	14	2.8	8	1.6	22	4.4
5	4	15	3.8	8	2.0	23	5.8
6	6	23	3.8	14	2.3	37	6.2
7	3	13	4.3	7	2.3	20	6.7
8	5	18	3.6	9	1.8	27	5.4
9	3	9	3.0	3	1.0	12	4.0
10	7	22	3.1	11	1.6	33	4.7
11	6	20	3.3	11	1.8	31	5.2
12	5	23	4.6	17	3.4	40	8.0
13	5	22	4.4	9	1.8	31	6.2
14	4	13	3.3	9	2.3	22	5.5
15	4	20	5.0	10	2.5	30	7.5
Average	4.3	16.53	4.0	9.27	2.3	25.80	6.3
Correl(Ab)	-0.48	-0.54	0.22	-0.48	0.26	-0.54	0.25
Correl(A1a)	-0.52	-0.61	0.25	-0.61	0.24	-0.63	0.26

(注) N はサブタスク数を表す。

発話数に関して、サブタスク数で正規化することにより、いずれの効率の主観評価との相関も、負の強い相関から正の弱い相関に変わっている。これは、サブタスク数自体、効率との相関が強いという結果に起因している。定義の上では、本来タスク効率とサブタスク数の相関は強く、対話効率とサブタスク数の相関は弱いはずであるが、上の表でいずれも相関が強くなっているのは、先に述べた様に対話効率の主観評価においてタスク効率と独立して評価することが難しいためと考えられる。

対話制御において対話効率の主観評価を向上する場合には、これらの客観量を基準とするのは適切ではないといえる。

(2) 発話時間

次に、効率の主観評価とサブタスク数で正規化した発話時間との相関を調べた結果を表 3.38 に示す。

表 3.38: 効率の主観評価とサブタスク数で正規化した発話時間

対話番号	N	S 発話時間	S 発話時間/N	U 発話時間	U 発話時間/N	総発話時間	総発話時間/N
1	2	28.57	14.29	13.38	6.68	41.95	20.98
2	5	87.27	17.45	37.62	7.52	124.90	24.98
3	1	18.21	18.21	9.59	9.59	27.80	27.80
4	5	72.45	14.49	18.03	3.61	90.48	18.09
5	4	74.15	18.54	18.86	4.72	93.01	23.25
6	6	93.57	15.60	62.97	10.50	156.50	26.08
7	3	75.07	25.02	15.30	5.10	90.37	30.12
8	5	79.02	15.80	45.91	9.18	124.90	24.98
9	3	52.45	17.48	19.24	6.41	71.69	23.89
10	7	83.84	11.98	33.84	4.83	117.70	16.81
11	6	69.75	11.63	40.71	6.79	110.50	18.42
12	5	93.52	18.70	41.01	8.20	134.50	26.90
13	5	126.50	25.30	20.24	4.05	146.80	29.36
14	4	61.34	15.34	18.89	4.72	80.23	20.06
15	4	92.11	23.03	33.08	8.27	125.20	31.30
Average	4.3	73.86	17.52	28.58	6.68	102.40	23.81
Correl(Ab)	-0.48	-0.54	-0.09	0.48	-0.08	-0.59	-0.67
Correl(A1a)	-0.52	-0.66	-0.20	-0.55	-0.07	-0.70	-0.77

(注) N はサブタスク数を表す。

発話時間に関しても、サブタスク数で正規化することにより、いずれの効率の主観評価

との相関も、負の強い相関から弱い相関に変わっている。

対話制御において対話効率の主観評価を向上する場合には、これらの客観量を基準とするのは適切ではない。

(3) 対話時間

次に、効率の主観評価とサブタスク数で正規化した対話時間との相関を調べた結果を表 3.39 に示す。

表 3.39: 効率の主観評価とサブタスク数で正規化した対話時間

対話番号	N	対話時間	対話時間/N	無音時間	無音時間/N
1	2	58.40	29.20	16.45	8.22
2	5	177.56	35.51	52.67	10.53
3	1	35.69	35.69	7.89	7.89
4	5	126.96	25.39	36.48	7.30
5	4	143.63	35.91	50.62	12.66
6	6	215.03	35.84	58.49	9.75
7	3	142.84	47.61	52.47	17.49
8	5	240.20	48.04	115.27	23.05
9	3	113.95	37.98	42.26	14.09
10	7	278.56	39.79	160.88	22.98
11	6	320.22	53.37	209.76	34.96
12	5	279.73	55.95	145.20	29.04
13	5	258.15	53.63	121.40	24.28
14	4	169.62	42.41	89.39	22.35
15	4	228.10	57.03	102.91	25.73
Average	4.3	186.58	42.22	84.14	18.02
Correl(Ab)	-0.48	-0.62	-0.52	-0.52	-0.53
Correl(A1a)	-0.52	-0.63	-0.55	-0.48	-0.52

(注) N はサブタスク数を表す。

対話時間および無音時間ともに、サブタスク数で正規化しても効率の主観評価と負の強い相関が見られた。

また、無音時間については、他の全ての客観量がタスク効率との相関の方が対話効率との相関よりも強いのに対して、正規化しない場合と同様にその逆の傾向が見られた。これは、わかりやすく言えば、相手の話し方の効率を主観的に評価する際、検索回数に関わらず平均的に無駄な時間が多いか少ないかが影響しているということである。これらのことから、サブタスク数で正規化した無音時間は対話制御において対話効率の主観評価を向上

する場合に有効な客観量と言える。

次節にはこの無音時間について、対話制御における客観量として利用することを目的に、さらに詳細に分析する。

3.4.3 無音時間と対話戦略について

上述の分析結果より、システムの対話制御によって対話効率の主観評価を向上するには、無音時間を極力減らすことが有効であると結論付けられる。ここでは、収録対話において、システム役がどのようにして無音時間を減らしているかを分析する。

無音時間を減らす（間を埋める）対話戦略としては、フィラー（繋語）やメタ発話（「ちょっと待って下さい」など）の他に、確認発話、申し出發話などがある。

表 3.40 に各対話における各種類の発話の出現頻度を示す。

表 3.40: 無音時間と対話戦略の関係

対話番号	無音時間	フィラー発話	メタ発話	確認発話	申し出發話	合計	総発話数	割合 (%)
1	8.23	7	0	5	2	14	22	63.6
2	10.53	11	1	8	0	20	64	31.3
3	7.89	4	0	1	1	6	13	46.2
4	7.30	8	0	6	2	16	42	38.1
5	12.66	7	0	4	2	13	46	28.3
6	9.75	9	0	9	2	20	61	32.8
7	17.49	10	1	6	3	20	32	62.5
8	23.05	5	0	6	0	11	55	20.0
9	14.09	6	0	5	1	12	22	54.5
10	22.98	18	2	7	0	27	79	34.2
11	34.96	12	0	5	0	17	72	23.6
12	29.04	17	0	8	2	27	101	26.7
13	24.28	24	3	7	0	34	72	47.2
14	22.35	10	0	4	3	17	51	33.3
15	25.73	10	1	10	1	22	65	33.8
Total	—	158	8	91	19	276	797	34.6

この表より、無音時間を減らす対話戦略として、場繋ぎ的に用いられるフィラーや、無音時間を生かして問題解決を円滑に進行させようとする確認発話、メタ発話などが頻繁に用いられていることがわかる。

また、これらの客観量と、対話効率（表 3.2 の効率 (b)）の主観評価結果との相関分析の結果を表 3.41 に示す。なお、相関分析では、被験者に関わらず対話毎の主観評価結果の平均値と客観量の数値との相関係数を求めた。

表 3.41: 対話効率の主観評価と客観量の相関関係

客観量	相関係数
平均発話回数（検索回数で正規化）	-0.24
平均発話時間（一発話当たり）	0.15
対話時間に対する無音時間の割合	-0.45

この結果から、効率評価の向上には、無音時間の削減が発話回数の削減以上に有効であることがわかった。

つまり、従来音声対話制御においては確認発話によって音声認識の信頼性の低さを補う際に、直接的な確認と間接的な確認をバランス良く用いることによって全体の発話数を極力抑えることが効率向上に重要とされてきた [30]。しかしながら、この原則を単純に適用することは不適切であり、検索処理の間などの無音時間を減らすような確認発話は回数が増えても問題がないと言える。同様に、音声認識などの処理中に間を埋める発話の出力も効率向上に有効であると言える。

これらの『間』を繋ぐ機能を持つ発話は、それぞれが予測される『間』に対応していると考えられる。従って対話制御においては、音声認識や問題解決などの処理にかかると予測される時間の大小に応じて、間を繋ぐ発話の種類を選択すれば良い [31]。今後はこうした対話制御の方法について具体的に検討する。

次節では、こうした人間同士の対話の分析結果に基づいて、効率を重視した対話制御のモデル化を試みる。

3.5 効率を重視した対話制御のモデル化

音声対話システムの動作中、システムは音声認識や問題解決の処理中であることや入力を受理されていないことなどのシステムの状態を伝達することが望ましい [32][33]。対面対話であれば視覚的出力によって、非対面対話であれば発話によってシステムの状態を的確に提示することで、誤解を最小限に抑え円滑な対話を促すことができる。

例えば、ユーザの発話内容を一部繰り返す確認発話、発話の間を「えと」「あの」などの間投詞表現で埋める自己調整発話、「ただいま検索中です」のように直接的にシステム

の状態を言い現すメタ発話などは、システムの内部状態を提示する役割を持つ発話と言える。従って、効率を重視した対話制御において無音時間を埋める発話を行なう際、このような状態提示発話のいずれかを選択して適切なタイミングで出力することが有効になる。

一般に、最適な対話戦略およびその出力タイミングはシステムの処理性能に依存する。これまでに、問題解決に必要な情報を入手するための対話戦略を音声認識精度に応じて学習する方法が提案されている [34][35]。また、音声認識精度あるいは音声認識結果の尤度に応じて確認発話戦略を決定するモデルが提案されている [30][36]。

以下には、自己調整発話やメタ発話などを含めて、より一般的にシステムの内部状態を提示する役割をもつ発話について、システムの処理精度および処理時間から決定される内部状態の曖昧さに応じて出力タイミングを決定するモデルを提案する。

ここで提案する対話制御モデルでは図 3.3 に示す様に、発話を受理する毎にシステムの応答に要する内部の処理時間を予測し、予測された処理時間や処理精度に応じて対話戦略を選択し、さらにその出力タイミングを適切に決定する。以下には本モデルについて、論文検索対話を例に具体的に説明する。

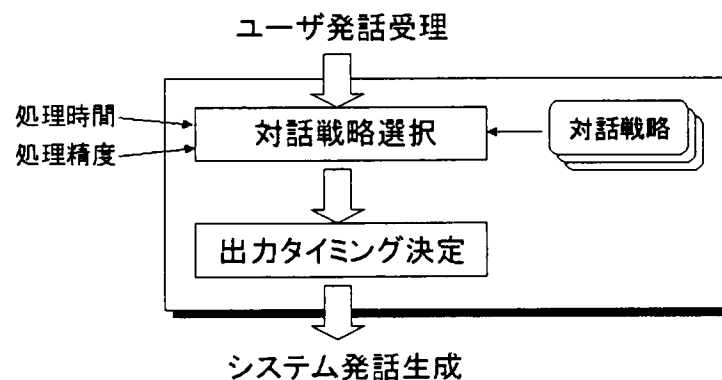


図 3.3: 効率を重視した対話制御モデルの概要

3.5.1 対話戦略の選択

対話戦略は、音声認識や論文検索などの処理についての、予測される処理時間や、処理精度に応じて決定される。処理時間の予測については、システムの動作環境に加え、入力

発話長や検索要求の複雑さなどを考慮することで経験的に精度よく予測することができると思われる。ここでは詳細な説明を省く。

以下に、ここで扱う対話戦略として、状態提示発話の種類ごとに、その具体的な役割と選択の基準を示す。

メタ発話 一部の処理が終ったことや処理時間が極端に長いと予測されることなどを発話によって伝えれば、ユーザにとってシステム状態予測が容易になる。また、それほど処理時間が長くなくても、結果としてユーザを待たせた場合には、正常な処理である旨を発話によって伝えることで、事後の状態の曖昧さを削減でき、さらに以後の状態予測を容易にさせる。

発話受理後、音声認識や論文検索の処理などに比較的長い処理時間が予測された場合に選択される。

確認発話 確認を行なえば、「相手の発話を聞いている」ことを示すことにつながり、その結果、音声処理が終了したことを提示でき、ユーザにとっての状態の曖昧さを大きく削減できる。

発話受理後、検索処理に比較的長い処理時間が予測された場合、および、受理した発話の音声認識結果の尤度の大小に応じて選択が決定される。

自己調整発話 処理時間が極端には長くないが若干長くなることが予測される場合には、自己調整的な発話を行なうことにより、発話権を継続するだけでなく、システムが処理中である旨を伝えることにつながる。

発話受理後、音声認識処理に比較的短い処理時間が予測された場合に選択される。

図 3.4 に、論文検索対話における戦略選択の例を示す。例えば、ユーザ発話「西川さんの論文を探してるんですが」を受理した後、最初の区間検出処理 p_1 が終了した時点で、区間長の長さから比較的短い音声認識処理時間が予測できた場合、システムは自己調整発話「ええと」を出力して、システムが区間検出処理を終了した旨を伝える。さらに、音声認識処理 p_2 終了後、その認識結果の尤度が比較的低く、なおかつ比較的長い検索処理時間が予測される場合、確認発話「西川さんですね」を出力しながら、検索処理 p_3 を並列に

実行する。このような処理により、予測される処理時間や精度に応じた適切な対話戦略を選択できるようになる。

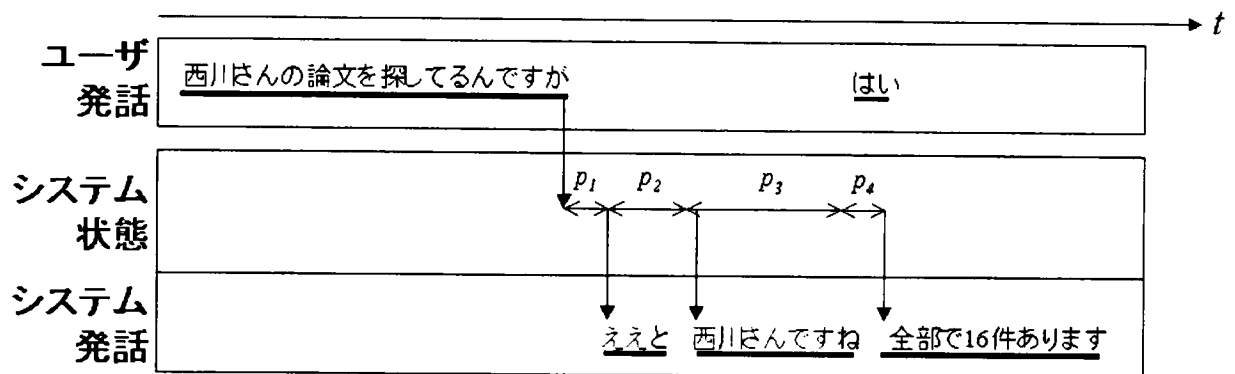


図 3.4: 対話における戦略の選択の例

3.5.2 出力タイミングの決定

上述の図 3.4 に示した例において、自己調整発話や確認発話の出力タイミングは区間検出処理や音声認識処理が終了した時点となっている。しかし、処理時間の実測値とユーザの予測とは異なるため、これは必ずしも適切とは言えない。

以下には、ユーザの予測を考慮した、システムの状態提示発話の出力タイミング決定のモデル化を試みる。

(1) 処理過程についての状態

一般に、ユーザはシステムを利用するうちに、比較的正確に実際のシステムの状態を予測することができる。音声対話システムについて知識を持つユーザであれば一層正確な予測が可能となる。なぜなら、システムを構成する各処理の処理時間は処理性能とユーザの入力によって決定されるある分布に準じており、その分布を経験的に知るからである。したがって、システムから何の反応もなくともユーザはある程度処理過程を予測していると考えることができる。ただし、システムについての知識の少ないユーザにおいては、個々に持つシステムイメージから期待する処理時間の長さが異なるため、それぞれのユーザに

とっての状態の曖昧さも異なる。

これは、システムの状態提示度とユーザの状態把握度の違いに起因する問題であり、ここでは処理時間の性能から求めるシステムの状態提示度をもって、ユーザの期待の要素を省いた客観的基準とする。

次項には、任意の時刻における状態の曖昧さを各処理の処理時間長の確率密度からエントロピーで表わす方法について以下に説明する。

なお、音声対話システムにおける各処理は、上位の処理によって必要に応じて呼び出されるサブプロセスとして位置付けられる。各処理は、開始と終了の時刻が明確に計測できるものであるとして、処理時間長はガウス分布に従うものと仮定する。各処理の分布の平均および分散の値は、各処理毎の性能評価において実測して得ることが可能である。厳密には、各処理のアルゴリズムとしての性能の他、計算機性能や通信環境などに依存する。

(2) 処理過程のモデル

システムの内部状態の集合

$$S = \{S_1, S_2, \dots, S_n\} \quad (3.1)$$

において、状態 S_i の処理内容を p_i と表わす。システムの動きは逐次的で、状態の一つの例としてシステムの動作が以下のように記述されることとする。

$$\sum(\alpha) = S_{\alpha_1} S_{\alpha_2} \cdots S_{\alpha_n}, S_{\alpha_j} \in S \quad (3.2)$$

状態遷移列 $\sum(\alpha)$ はユーザ発話やシステムの処理結果によって変化するが、ここでは単純なマルコフモデルで近似して、 S_{α_j} が S のうち S_k 、 $S_{\alpha_{j+1}}$ を S_l とすると、 S_k から S_l の状態遷移確率 P_{kl} で状態遷移が表現されるとする。この状態遷移確率 P_{kl} の値は、入力発話の処理結果の信頼性や処理 p_k の精度に応じて決定される。例えば、ユーザの検索要求を示す発話に対して、認識精度が充分高ければ検索処理への遷移は確実であるが、認識精度が低い場合には別の処理への遷移が起こる可能性が高まる。こうした傾向についてユーザも経験的に知ることができると考えられ、従って厳密にはこの状態遷移確率の値を考慮すべきである。しかしながら、一般に単純なタスクでは、 P_{kl} は 0 か 1 に近い値を取るものと考えられ、ユーザにとってシステムの動きは予測しやすいものとなっている。そ

ここで、簡単の為、状態遷移列 $[S_1 S_2 \dots S_n]$ で表わされる一定の状態遷移を仮定し、状態遷移確率の影響を除く。

システムが処理を開始してからの時刻 t において状態 S_i にある確率を $P_i(t)$ とする。また、状態 S_i の継続時間長 D_i が正規分布 $N(\mu_i, \sigma_i^2)$ に従い、その確率密度関数が $f_i(t)$ で表されるとする。これらの関係を図 3.5 に示す。

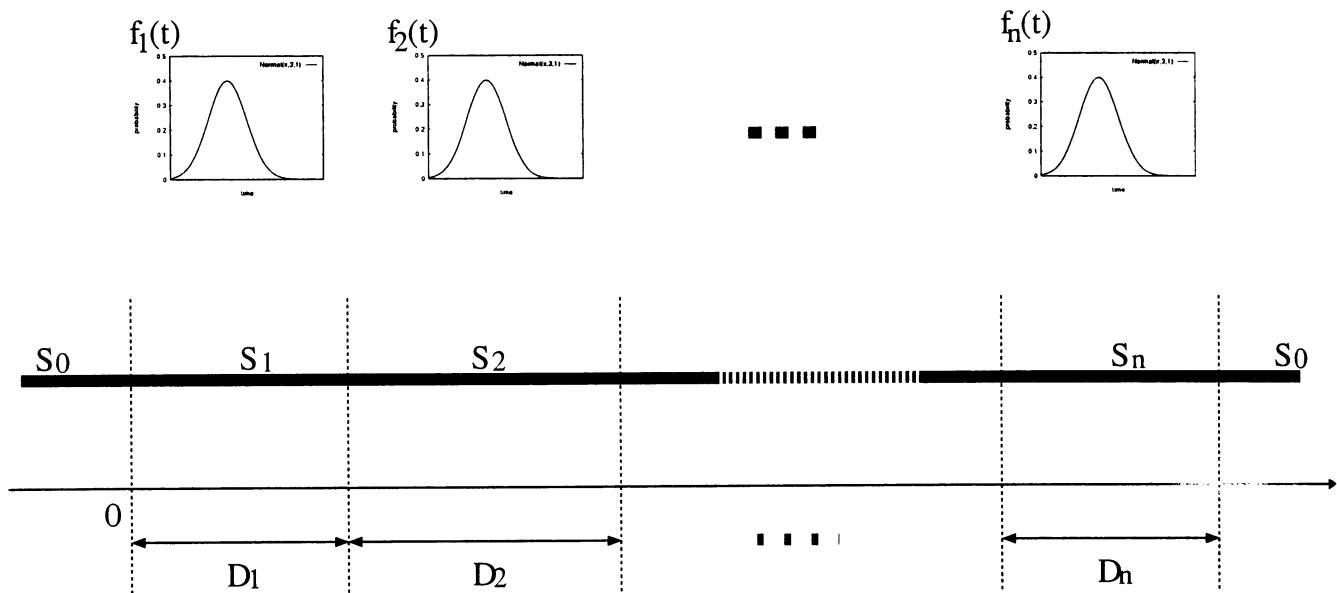


図 3.5: 処理過程と状態の遷移

ここで、状態 $S_i (i = 1, 2, \dots, n)$ のどれでもない、つまり処理を行っていない状態を S_0 とすると、以下の式が成り立つ。

$$\sum_{i=0}^n P_i(t) = 1 \tag{3.3}$$

まず、 $P_1(t)$ は、時刻 t に状態 S_1 にある確率であり、すなわち状態 S_1 の継続時間長 D_1 が t 以上となる確率に等しいから、

$$P_1(t) = \int_t^{\infty} f_1(x) dx \tag{3.4}$$

と現せる。標準化変換を施して以下のようなになる ($g(z)$ は標準正規分布 $N(0, 1)$ の確率密

度関数)。

$$\begin{aligned} P_1(t) &= \int_{\frac{t-\mu_1}{\sigma_1}}^{\infty} g(z) dz \\ &= \phi\left(\frac{t-\mu_1}{\sigma_1}\right) \end{aligned} \quad (3.5)$$

次に、時刻 t において、既に状態 S_1 が終了し現在状態 S_2 である確率 $P_2(t)$ は、すなわち状態 S_1 の継続時間長 D_1 と状態 S_2 の継続時間長 D_2 の和が t 以上となる確率から状態 S_1 にある確率を引いたものに等しい。 D_1 と D_2 の和はガウス分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ に従うから、 $D_1 + D_2$ の確率密度関数を $f_{1+2}(x)$ とすれば、

$$P_2(t) = \int_t^{\infty} f_{1+2}(x) dx - P_1(t) \quad (3.6)$$

と表せる。 $P_1(t)$ と同様に標準化変換を施せば、

$$\begin{aligned} P_2(t) &= \int_{\frac{t-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}}^{\infty} g(z) dz - P_1(t) \\ &= \phi\left(\frac{t-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) - P_1(t) \end{aligned} \quad (3.7)$$

同様にして、時刻 t における状態 S_i にある確率 $P_i(t)$ ($i = 1, 2, \dots, n$) を以下の式で求めることができる。

$$P_i(t) = \phi\left(\frac{t - \sum_{j=1}^i \mu_j}{\sqrt{\sum_{j=1}^i \sigma_j^2}}\right) - \sum_{j=1}^{i-1} P_j(t) \quad (3.8)$$

なお、システム内の全ての状態が終り、 S のいずれの状態でもない場合を S_0 とすると、(3.3) 式より、時刻 t において状態 S_0 にある確率は、以下の式で表される。

$$P_0(t) = 1 - \sum_{i=1}^n P_i(t) \quad (3.9)$$

以上からそれぞれの状態にある確率を用いて、時刻 t における状態のエントロピー $H(t)$ は以下の式で表せる。

$$H(t) = - \sum_{i=0}^n P_i(t) \log P_i(t) \quad (3.10)$$

(3) 状態エントロピーのモデル例

図 3.6 に、システムを構成する 3 つの処理の継続時間長 D_1, D_2, D_3 の分布が与えられた 2 つのケースの状態エントロピーの推移を示す。それぞれの平均状態エントロピーは 0.55, 0.51 と計算できる。なお、図の凡例のカッコ内の値は、3 つの与えられた処理時間長分布の平均と標準偏差である。

この様に、処理性能の違いによって、システムが潜在的に持つ状態の曖昧さが異なることが、提案した状態エントロピーで表現でき、異なる対話システムの性能を比較する上で有効であると同時に、各システムの対話制御において適切な対話戦略の採用にも有効な手段になると考える。

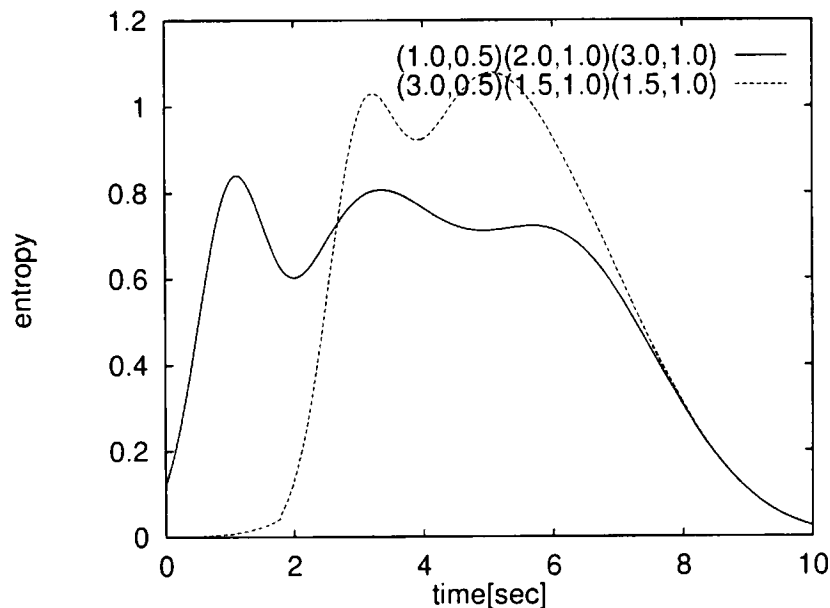


図 3.6: 処理時間長の分布の違いと状態エントロピー

以下、前述の対話戦略のうち、メタ発話を用いる場合の処理過程についての状態エントロピーの変化について説明する。まず、メタ発話によって状態の曖昧さを削減する様子を状態エントロピーで表わすと図 3.7 の様になる。なお、この図はシステムを構成する 3 つの処理の処理時間長の分布が $N(3.0, 0.5^2)$, $N(1.5, 1.0^2)$, $N(1.5, 1.0^2)$ で与えられた場合に、処理を開始してから 3.0 秒後に既に最初の処理 p_1 が終了している旨を発話した時のものである。

この様に、システムが最終的に応答を返す前に、処理状態を提示するような何らかの発

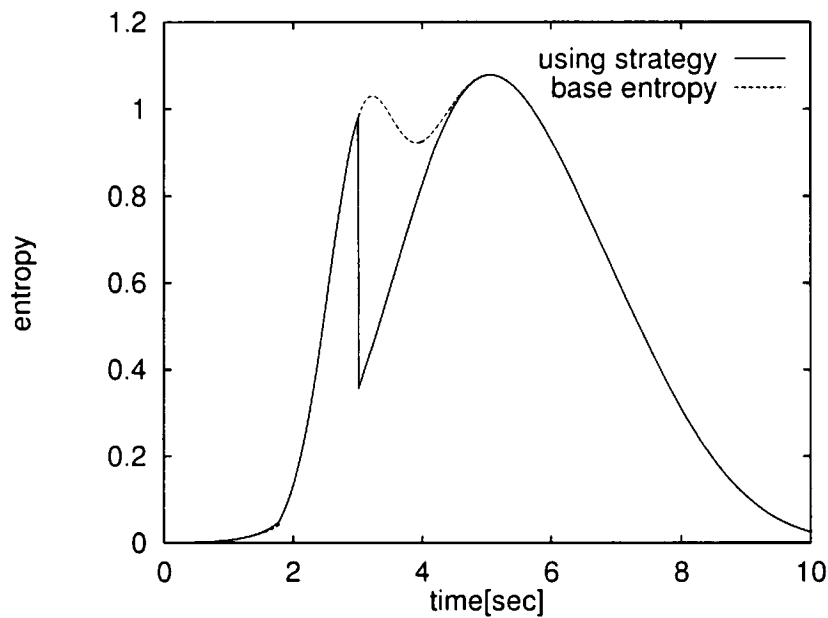


図 3.7: メタ発話による状態の曖昧さ削減

話を入力することにより、ユーザにとっての状態の曖昧さが削減できると考えられる。図 3.7 と同じ条件において、メタ発話出力時刻と平均状態エントロピーの関係を図 3.8 に示す。この図より、処理開始後 2 秒程度ならどの時刻にメタ発話を入力しても、平均状態エントロピーの変化は同程度であることがわかる。この様にして、処理時間によって現した処理過程についての状態エントロピーを計算することで、システムの処理過程についての状態を示す発話を適切なタイミングで出力することが可能になる。

なお、ここでは視覚表示を十分に使えず音声のみを用いて対話を行なう状況を想定して説明したが、視覚表示も状態の曖昧さを減らす一戦略として扱うことが可能である。

3.6 対話制御方法の評価実験

前章に提案した対話制御モデルによる状態提示およびそれに伴う無音区間の減少が効率評価向上に有効か否かを調べるために、以下に実験を行なった。

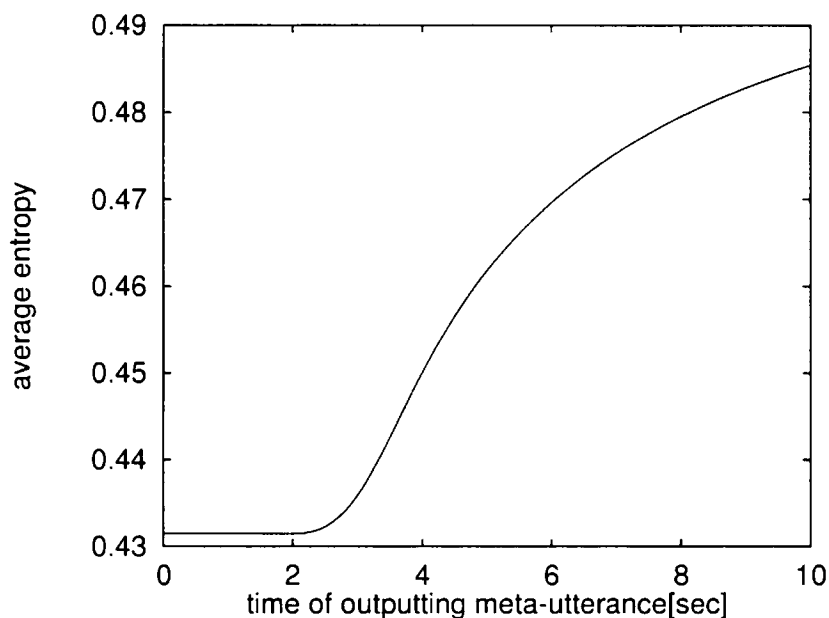


図 3.8: メタ発話出力時刻と平均状態エントロピー

3.6.1 音声対話システムのプロトタイプ

実験においては、学术论文データベース検索をタスクとした音声対話システムのプロトタイプを用いる。システムは図 3.9 に示す様に、音声認識、論文検索、音声合成の主に 3 つの処理によって構成される。学术论文データベースには 500 件の論文が存在し、音声認識ではその書誌情報に対応した約 5000 語の語彙を扱う。音声認識エンジンとしては、モノフォンHMMと単語バイグラムを用いた連続単語認識プログラム [37] を用い、ほぼ実時間での認識処理を実現した。論文検索については、あらかじめ 5000 語に含まれるキーワードと論文とを対応付けておくことにより、高速な検索処理を可能とした。

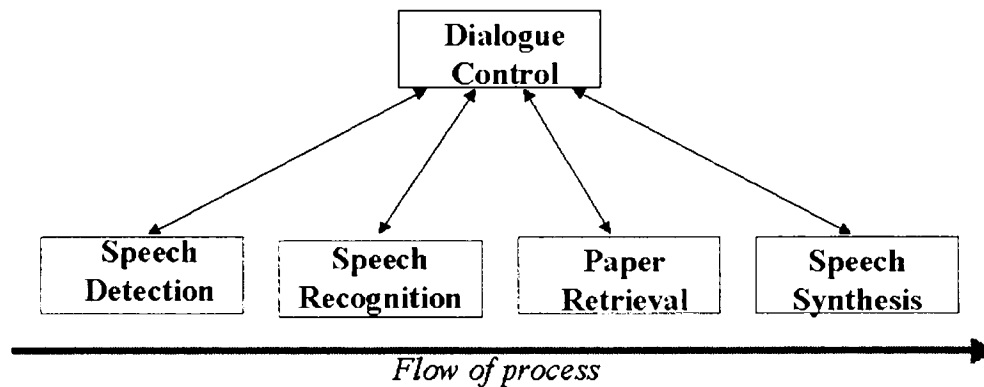


図 3.9: 音声対話システムの構成.

3.6.2 実験手順

実験では、論文検索の処理時間を3種類に設定し、システムがユーザの発話を受理した後に、音声認識処理が終了した旨を伝えるメタ発話を出力するタイミングを変えることで、4種類の対話制御方法を用意した。4種類の対話制御方法はそれぞれ、(a)メタ発話出力なし、(b)音声認識処理終了直後にメタ発話出力、(c)区間検出してから平均論文検索時間後にメタ発話出力、(d)状態エントロピー算出により求めた一定時間後にメタ発話出力するものである。

なお、実験におけるシステムの各処理時間の平均および分散は、予備実験での結果から、音声認識が $(1.84, 0.80^2)$ であり、論文検索については(1) $(2.40, 0.20^2)$ 、(2) $(5.40, 0.50^2)$ 、(3) $(10.40, 1.00^2)$ を設定した(いずれも単位は[sec])。その他の処理時間はいずれも小さい値である為、無視した。これらの値から計算された状態エントロピーは図3.10のようになる。

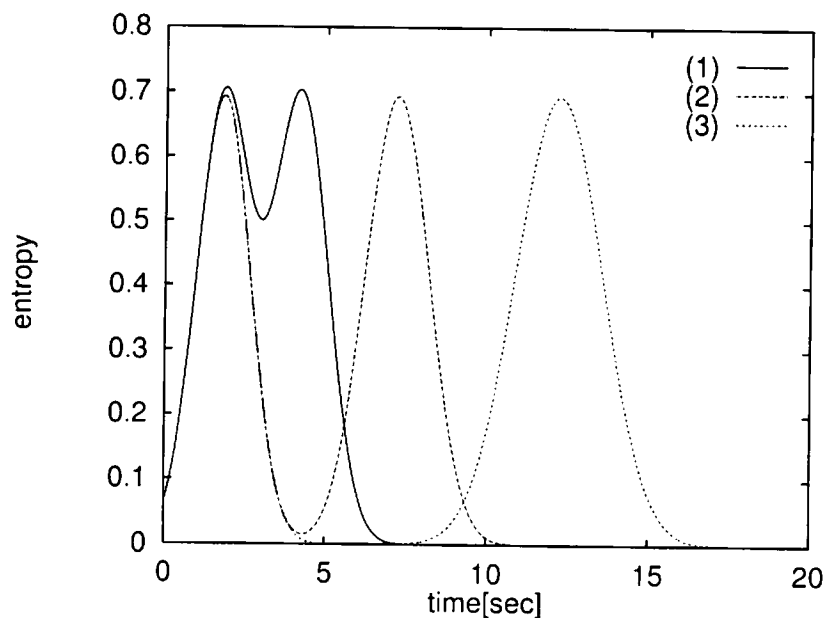


図 3.10: 実験での対話制御における状態エントロピーの推移

状態エントロピーは発話の出力タイミング範囲を得るものであるが、便宜上実験では、論文検索時間の変化に応じて、図3.10のそれぞれの曲線で最初の極小値を与える時刻(1)3.2、(2)4.2、(3)7.3[sec]後にメタ発話を出力した。

なお、5人の被験者はそれぞれの制御を行なうシステムに対して、あらかじめ教示され

た検索要求を達成すべく対話を行ない、さらに、実験終了直後に表 3.2 に示した 25 種類の項目についてのアンケート評価（5 段階）を行なった。

3.6.3 実験結果

アンケート評価結果のうち、相手の話し方の効率（表 3.2 の効率 (b)）についての評価の平均を表 3.42 に示す。まず、(a) 出力なしの場合と、提案モデルによる (d) の場合とで比較すると、平均論文検索時間を変えたいずれの場合も評価が大きく向上している。さらに、状態提示発話の出力タイミングについて、提案モデルによる (d) の場合と (b)(c) の場合とで比較すると、(d) が最も高くその次に (c)、(b) の順となっている（(1) の場合のみ (b)、(c) の順）。

これらの結果から、無音時間を埋める状態提示発話を出力することによって対話の効率評価を向上させ、なおかつ提案した対話制御モデルを用いることによってシステムの処理性能に応じた最適な出力タイミングを決定できることが確認できた。

表 3.42: 各メタ発話制御方法に対する効率の評価結果
((1)(2)(3) はそれぞれ平均論文検索時間を 2.4, 5.4, 10.4[sec] とした場合を示す。)

対話制御方法	(1)	(2)	(3)
a) 出力なし	2.5	2.0	1.8
b) 音声認識処理終了直後	3.3	3.8	3.0
c) 平均論文検索時間後	3.1	4.1	3.3
d) 状態エントロピー	3.6	4.3	4.3

3.7 まとめ

本章では、音声対話インタフェースのコミュニケーション調整機構における音声対話制御の一例として、状態提示行為制御のモデルを提案した。

まず主観的な評価指標の特性と対話制御の機能目標を明らかにすべく、人間同士の模擬的な対話を収集して評価を行った。実験結果に基づいて評価指標の特性と機能目標の分析を行い、そこから導かれる重要な結論として、被験者によらず効率が重視されることから、効率に関する収集対話に対する主観評価結果と客観量の関係性を分析した。その結果、

従来、効率の低下につながるとされていた発話数が、実際には効率の主観評価に影響を与えず、むしろ音声認識や検索処理などの間を埋める状態提示発話を積極的に出力すべきであることを、人間同士の対話分析により明らかにし、その上で処理性能に応じて適切な発話の種類やその出力タイミングを決定するモデルを提案した。提案した対話制御モデルの一部を実際の対話システムに実装、評価した結果、対話効率の向上に有効であることを確認した。

本章では、扱うタスクの性質上、対話における状態としてシステムの内部処理の状態に限定し、さらにその中で一定の状態系列が予測される状況に限定してモデルの説明をした。しかし、一般の対話では多様なレベルの状態が存在するであろうし、状態遷移もより複雑になることが予測される。ここで示した状態エントロピーのモデルにおいて、状態の定義を対話における参加者の心的状態にまで拡大し、なおかつ各状態間の遷移確率を数量化することにより、そうした多様な対話にも対応可能と考える。