

これらから、Ab (効率) はどの被験者も重視するが、Bb (透過性) や Cb (リズム) については重視しない被験者も存在すると言える。

なお、いずれの指標も評価者による対話への参加の影響を受けないと仮定して、総合的評価 Xb を目的変数、指標 Ab、Bb、Cb を説明変数として重回帰分析を全評価結果を用いて行なったところ、1%あるいは5%の有意水準で重視される傾向が確認された指標は各被験者ごとに表 3.28 の通りであった。この表からも、上の考察が導かれる。

表 3.14: 相手についての総合指標:重回帰分析 (対話聴取後)
(**は 1%、*は 5%の有意水準で重視されることを示す。)

被験者	A	B	C	D	E
Ab (効率)	**		*	**	*
Bb (透過性)				*	*
Cb (リズム)				*	

(8) 相手についての詳細指標:Xb との相関係数 (対話聴取後)

評価時期：対話後 (HE)

評価対象：相手の参加者についての詳細指標 (b)

相手の参加者の話し方についての総合的評価と詳細な指標との相関について、被験者によらず相関係数を調べた結果および各被験者毎に相関係数を調べた結果を表 3.15 に示す。なお、3.3.1 節に示した様に、対話聴取後の評価指標の多くは評価者自身の対話への参加の影響を受けるため評価者自身が参加していない対話への評価結果を分析の対象とした。

C2b (スピード) E1b (表現の多様性) E2b (話し方の多様性) E3b (リズムの多様性) は評価者自身の対話への参加の影響を受けないため全対話への評価結果を、それ以外の詳細指標は評価者自身の対話への参加の影響を受けるため評価者自身が参加した対話への評価結果のみを分析に用いた。

この表において、B1b (処理状態)、B2b (意思伝達)、C1b (間) についてはいずれの被験者の評価結果も総合的評価 Xb との相関が強い。従って、これらは被験者によらず重視される指標と言える。

一方、A2b (無駄時間) については被験者 A の評価結果において、A3b (発話量) については被験者 A および C の評価結果において低い相関を示していることから、これらの

表 3.15: 相手についての総合指標: Xb との相関係数 (対話聴取後)

被験者	全体	A	B	C	D	E
A2b (無駄時間)	0.5984	0.0909	0.4915	0.6766	0.8307	0.5023
A3b (発話量)	0.4716	0.0909	0.5497	0.2557	0.5384	0.4961
B1b (処理状態)	0.5561	0.3223	—	0.3105	0.8941	0.6501
B2b (意思伝達)	0.4107	0.3656	0.2651	0.0491	0.8410	0.3133
B3b (安心感)	0.7000	0.6364	0.5513	0.6295	0.8627	0.8008
C1b (間)	0.5343	0.2132	0.4207	0.5704	0.8353	0.1414
C2b (スピード)	0.2155	0.2548	-0.2447	-0.0331	0.4256	0.6000
E1b (表現の多様性)	0.3423	0.1741	0.3052	0.4436	0.4274	0.3333
E2b (話し方の多様性)	0.3295	0.1612	0.3874	0.1299	0.4274	0.3333
E3b (リズムの多様性)	0.3771	0.3015	0.5709	-0.0776	0.4869	0.4648

指標は被験者によって重視される場合とそうでない場合があると言える。

B3b (安心感) については、被験者 E の評価結果が総合的評価 Xb と負の相関を示していることから、被験者によって重視される場合とそうでない場合があると言える。

また、C2b (スピード)、E1b (表現の多様性)、E2b (話し方の多様性)、E3b (リズムの多様性) については、いずれの被験者の評価結果も総合的評価 Xb との相関が弱く、これらは被験者によらず重視されない指標と言える。

以上の考察をまとめると、対話聴取評価時に相手の話し方についての指標に対して評価者が重視する指標の個人差に関して以下の結論が導ける。

- 効率はどの被験者も重視するが、透過性やリズムについては重視しない被験者も存在する。
- ただし効率の詳細指標である発話量や無駄時間については、重視しない被験者も存在する。
- 処理状態の透過性、意思伝達の透過性、間といった指標については、被験者によらず重視される指標と言える。
- 安心感については、被験者によって重視される場合とそうでない場合があると言える。
- スピード、表現の多様性、話し方の多様性、リズムの多様性については、被験者によらず重視されない指標と言える。

なお、対話後および対話聴取後の評価に際して、被験者にそれぞれ重視する指標を選択させた（複数選択可）。その結果を表 3.16 に示す。この表から、多くの被験者は論文検索の効率（タスク効率）を重視し、被験者によってはその他の指標も重視するといった傾向があると言える。

表 3.16: 被験者が重視するとした指標

評価指標	RE	HE
効率 (a)	2	1
効率 (b)	0	0
論文検索の効率 (a)	3	4
無駄時間 (a)	0	1
無駄時間 (b)	1	1
発話量 (a)	0	0
発話量 (b)	0	0
透過性 (b)	1	1
処理状態 (b)	1	1
意思伝達 (b)	1	0
安心感 (b)	0	2
リズム (a)	0	0
リズム (b)	1	0
間 (a)	2	1
間 (b)	0	1
スピード (b)	1	0
バランス (a)	0	0
主体性 (主導権) (a)	0	1
表現の多様性 (b)	0	1
話し方の多様性 (b)	0	0
リズムの多様性 (b)	0	0

3.3.3 仮説 H_3 の検証：指標間の関係

(1) 対話についての効率の指標:Aa との相関係数

評価時期：対話後 (RE)

評価対象：対話についての効率 (Aa)

まず、対話についての指標のうち、効率の総合指標 Aa と詳細指標 A1a（論文検索効率）、A2a（無駄時間）、A3a（発話量）の相関関係を調べた結果を、表 3.17 に示す。

この表から、効率 Aa の詳細指標とした A1a、A2a、A3a のうち、A1a（論文検索効率）

表 3.17: 対話についての効率の指標:Aa との相関係数 (対話後)

被験者	全体	A	B	C	D	E
A1a (論文検索効率)	0.9075	1.0000	0.9820	0.8660	0.9449	—
A2a (無駄時間)	0.8207	0.9449	0.9820	0.8660	0.8660	0.9820
A3a (発話量)	0.7522	0.5000	0.9820	-0.8660	0.9449	—

と A2a (無駄時間) はいずれの被験者においても Aa との相関が強く、下位分類として機能し得るが、A3a (発話量) に関しては Aa との相関が強い被験者が 5 人中 2 人おり、下位分類として相応しくない可能性があると言える。

(2) 対話についてのリズムの指標:Ca との相関係数

評価時期：対話後 (RE)

評価対象：対話についてのリズム (Ca)

次に、対話についての指標のうち、リズムの総合指標 Ca と詳細指標 C1a (間)、C3a (バランス) の相関関係を調べた結果を、表 3.18 に示す。

表 3.18: 対話についてのリズムの指標:Ca との相関係数 (対話後)

被験者	全体	A	B	C	D	E
C1a (間)	0.8985	1.0000	0.9449	0.5000	0.8660	1.0000
C3a (バランス)	0.3385	0.5000	0.7559	0.5000	0.8660	-0.5000

この表から、リズム Ca の詳細指標とした C1a、C3a のうち、C1a (間) はいずれの被験者においても Ca との相関が強く、下位分類として機能し得るが、C3a (バランス) に関しては Ca との相関が強い (相関係数が 0.5 以下) 被験者が 5 人中 3 人おり、下位分類として相応しくない可能性があると言える。この考察は、3.3.2 で示した予測と一致しており、以降、C3a (バランス) についてはその分類について再考する必要がある。

(3) 相手についての効率の指標:Ab との相関係数

評価時期：対話後 (RE)

評価対象：対話についての効率 (Ab)

次に、相手の話し方についての指標のうち、効率の総合指標 Ab と詳細指標 A2b (無駄

時間)、A3b (発話量) の相関関係を調べた結果を、表 3.19 に示す。

表 3.19: 相手についての効率の指標:Ab との相関係数 (対話後)

被験者	全体	A	B	C	D	E
A2b (無駄時間)	0.9087	0.5000	0.9449	—	1.0000	—
A3b (発話量)	0.7750	0.5000	0.8660	—	1.0000	—

この表において相関係数の算出ができないサンプルが多く存在するため、信頼性の低い結果ではあるが、効率 Ab の詳細指標とした A2b、A3b のどちらもいずれの被験者においても Ab との相関が強く、下位分類として機能し得ると考えられる。

(4) 相手についての透過性の指標:Bb との相関係数

評価時期：対話後 (RE)

評価対象：対話についての透過性 (Bb)

次に、相手の話し方についての指標のうち、透過性の総合指標 Bb と詳細指標 B1b (処理状態)、B2b (意思伝達)、B3b (安心感) の相関関係を調べた結果を、表 3.20 に示す。

表 3.20: 相手についての透過性の指標:Bb との相関係数 (対話後)

被験者	全体	A	B	C	D	E
B1b (処理状態)	0.6938	—	1.0000	-0.5000	0.8660	—
B2b (意思伝達)	0.6898	1.0000	—	0.0000	—	—
B3b (安心感)	0.4660	0.8660	-0.5000	-0.5000	0.8660	—

この表において、B1b (処理状態) および B2b (意思伝達) については相関係数の算出ができないサンプルが多く、比較が困難である。B3b (安心感) については相関係数が算出できる 4 名の被験者の評価結果のうち、2 名が正の高い相関、2 名が負の相関を示している。これは 3.3.2 項に示した予測とあわせて、分類の不適切さによるものと考えられる。従って、以降では B3b (安心感) の分類を再考する必要がある。

(5) 相手についての透過性の指標:Cb との相関係数

評価時期：対話後 (RE)

評価対象：対話についてのリズム (Cb)

次に、相手の話し方についての指標のうち、リズムの総合指標 Cb と詳細指標 C1b (間)、C2b (スピード) の相関関係を調べた結果を、表 3.21 に示す。

表 3.21: 対話についてのリズムの指標:Cb との相関係数 (対話後)

被験者	全体	A	B	C	D	E
C1b (間)	0.7533	1.0000	0.7559	—	1.0000	—
C2b (スピード)	0.2558	—	0.7559	—	1.0000	—

この表から、リズム Cb の詳細指標とした C1b、C2b のうち、C1b (間) は相関係数が算出できるいずれの被験者の評価結果においても Cb との相関が強く、下位分類として機能し得ると言える。C2b (スピード) に関しては相関係数の算出ができないサンプルが多く、比較が困難であるが、算出できたサンプルに関しては Cb との相関が高く、個人差も存在しないことが予測される。

以上、対話後の評価における指標間の関係の分析結果から、以下のようにまとめられる。

- 対話についての指標のうち、論文検索効率と無駄時間はいずれの被験者においても効率の詳細指標として機能する。
- 対話についての指標のうち、発話量に関しては効率の下位分類として相応しくない可能性がある。
- 対話についての指標のうち、間はいずれの被験者においてもリズムの下位分類として機能する。
- 対話についての指標のうち、バランスに関してはリズムの下位分類として相応しくない可能性がある。
- 相手の話し方についての指標のうち、無駄時間、発話量はいずれの被験者においても効率の下位分類として機能し得る。
- 相手の話し方についての指標のうち、間はいずれの被験者においてもリズムの下位分類として機能し得る。
- 相手の話し方についての指標のうち、スピードに関してもいずれの被験者においてもリズムの下位分類として機能し得ると予測される。

次に、対話聴取後の評価結果について指標間の関係を分析する。なお、3.3.1 節に示した様に、対話聴取後の評価指標の多くは評価者自身の対話への参加の影響を受けるため評価者自身が参加していない対話への評価結果を分析の対象とした。

(6) 対話についての効率の指標:Aa との相関係数 (対話聴取後)

評価時期：対話聴取後 (HE)

評価対象：対話についての効率 (Aa)

まず、対話についての指標のうち、効率の総合指標 Aa と詳細指標 A1a (論文検索効率)、A2a (無駄時間)、A3a (発話量) の相関関係を調べた結果を、表 3.22 に示す。

表 3.22: 対話についての効率の指標:Aa との相関係数 (対話後)

被験者	全体	A	B	C	D	E
A1a (論文検索効率)	0.8192	0.8939	0.8380	0.7725	0.8707	0.9577
A2a (無駄時間)	0.7787	0.5202	0.7948	0.8793	0.8252	0.9350
A3a (発話量)	0.7175	0.7634	0.8078	0.7929	0.3490	0.8446

この表から、効率 Aa の詳細指標とした A1a (論文検索効率)、A2a (無駄時間)、A3a (発話量) は、いずれの被験者においても Aa との相関が強く、下位分類として機能し得る。ただし、A3a (発話量) のみ被験者 D の評価結果で弱い相関を示しており、この場合は分析に十分なサンプル数があるために、この点において仮説 H_3 が正しくない可能性もある。

(7) 対話についてのリズムの指標:Ca との相関係数 (対話聴取後)

評価時期：対話聴取後 (HE)

評価対象：対話についてのリズム (Ca)

次に、対話についての指標のうち、リズムの総合指標 Ca と詳細指標 C1a (間)、C3a (バランス) の相関関係を調べた結果を、表 3.23 に示す。

この表から、リズム Ca の詳細指標とした C1a、C3a のうち、C1a (間) はいずれの被験者においても Ca との相関が強く、下位分類として機能し得るが、C3a (バランス) に関しては Ca との相関が強い (相関係数が 0.5 以下) 被験者が 5 人中 2 人おり、下位分

表 3.23: 対話についてのリズムの指標:Ca との相関係数 (対話聴取後)

被験者	全体	A	B	C	D	E
C1a (間)	0.9156	0.7385	0.9325	0.9800	0.9459	0.9207
C3a (バランス)	0.5932	0.0000	0.2375	0.7736	0.8307	0.9536

類として相応しくない可能性があると言える。対話後の評価についての分析結果からの考察と同様に、以降、C3a (バランス) についてはその分類について再考する必要がある。

(8) 相手についての効率の指標:Ab との相関係数 (対話聴取後)

評価時期：対話聴取後 (HE)

評価対象：対話についての効率 (Ab)

次に、相手の話し方についての指標のうち、効率の総合指標 Ab と詳細指標 A2b (無駄時間)、A3b (発話量) の相関関係を調べた結果を、表 3.24 に示す。

表 3.24: 相手についての効率の指標:Ab との相関係数 (対話聴取後)

被験者	全体	A	B	C	D	E
A2b (無駄時間)	0.6831	0.4545	0.6473	0.3722	0.7163	0.8372
A3b (発話量)	0.4994	0.4545	0.7249	0.0553	0.6560	0.7667

この表において、効率 Ab の詳細指標とした A2b、A3b のどちらも被験者 C のみ相関が低くなっている。この点において仮説 H_3 が正しくない可能性もある。

(9) 相手についての透過性の指標:Bb との相関係数 (対話聴取後)

評価時期：対話聴取後 (HE)

評価対象：対話についての透過性 (Bb)

次に、相手の話し方についての指標のうち、透過性の総合指標 Bb と詳細指標 B1b (処理状態)、B2b (意思伝達)、B3b (安心感) の相関関係を調べた結果を、表 3.25 に示す。

この表において、B1b (処理状態) についてはいずれの被験者の評価結果も Bb との相関が強い。B2b (意思伝達) および B3b (安心感) については、いずれの被験者も Bb との相関が弱い、被験者 D の結果のみ若干強い相関を示している。この傾向は、B3b に

表 3.25: 相手についての透過性の指標:Bb との相関係数 (対話聴取後)

被験者	全体	A	B	C	D	E
B1b (処理状態)	0.8991	0.9258	—	0.7878	1.0000	0.8461
B2b (意思伝達)	0.5671	0.2521	-0.1267	0.4152	0.8464	0.6035
B3b (安心感)	0.5833	0.3133	0.4097	0.4686	0.8699	0.7884

については前節に示したように分類の不適切さによるものと考えられるが、B2b については特に説明する理由がない。この点において仮説 H_3 が正しくない可能性もある。

(10) 相手についての透過性の指標:Cb との相関係数 (対話聴取後)

評価時期：対話聴取後 (HE)

評価対象：対話についてのリズム (Cb)

次に、相手の話し方についての指標のうち、リズムの総合指標 Cb と詳細指標 C1b (間)、C2b (スピード) の相関関係を調べた結果を、表 3.26 に示す。

表 3.26: 対話についてのリズムの指標:Cb との相関係数 (対話聴取後)

被験者	全体	A	B	C	D	E
C1b (間)	0.7622	0.0000	0.8863	0.7025	0.8926	0.2689
C2b (スピード)	0.4171	0.5071	0.3780	-0.1001	0.8312	0.5704

この表から、リズム Cb の詳細指標とした C1b、C2b のうち、C1b (間) は、5 人中 3 人の結果では相関が強いが残る 2 人の結果では極端に相関が弱くなっている。また、C2b についても、被験者 D の結果のみ強い相関を示し、他の被験者の結果では弱い相関を示している。以上からも仮説 H_3 が正しくない可能性もある。

以上の結果から、仮説 H_3 が必ずしも正しいとは言えない指標があり、指標の分類を再考する必要があると言える。

3.3.4 仮説 H_4 の検証：対話後と対話聴取後の比較

ここでは同一被験者の対話後および対話聴取後の評価結果の比較を行なう。

(1) リアルタイム評価結果と聴取評価結果の一致率

まず同一被験者の同一対話に対しての評価結果について一致率 [28] を調べた結果を表 3.27 に示す。

表 3.27: リアルタイム評価結果と聴取評価結果の相関および一致率

評価指標	相関係数	κ
効率	0.5596	0.5
論文検索の効率	0.68277	0.2
無駄時間	0.55385	0.2
発話量	0.52972	0.3
透過性	0.20069	0.1
処理状態	0.20719	0.2
意思伝達	0.67135	0.5
リズム	0.78654	0.4
間	0.40459	0.2
スピード	0.50463	0.2
バランス	0.71898	0.3
主体性 (主導権)	0.78267	0.1
多様性		
表現の多様性	0.08392	0
話し方の多様性	-0.1829	0
リズムの多様性	-0.2729	0.1

この表において、 $.70 < \kappa$ であることが信頼性の目安であるが [29]、いずれの指標においてもその基準を下回っている。しかしこれは主観評価自体が必ずしも絶対的に行なわれなため当然の結果である。ここでは相関係数から以下の知見を提示するにとどめる。

- 効率、リズム、主体性に関しては聴取評価による代替がある程度可能である。
- 透過性、多様性に関しては聴取評価による代替は不可能。

(2) 仮説 H_2 の検証結果についての比較

また、3.3.2 節の仮説 H_2 の検証結果について、対話後および聴取後の比較して整理する。まず、対話についての指標に関して以下の様にまとめる。

1. 効率およびタスク効率の重要性は、対話後評価では個人差があるのに対して対話聴取後評価では個人差が少ない。

2. 無駄時間および発話量の重要性は、いずれの評価においても個人差が見られるが、重要としない被験者は対話後評価と対話聴取後評価で異なる。
3. リズムの重要性は、対話後評価では個人差があるのに対して対話聴取後評価では個人差が少ない。
4. 間の重要性は、いずれの評価においても個人差が少ない。
5. バランスの重要性は、いずれの評価においても個人差が見られるが、重要としない被験者は対話後評価と対話聴取後評価で異なる。
6. 主体性の重要性は、いずれの評価においても個人差が見られるが、重要としない被験者は対話後評価と対話聴取後評価で異なる。

1の効率については、比較的客観性の高い指標であるために本来個人差が少ないことが予測できるが、対話聴取後評価よりも対話後評価で個人差が多く見られるのは、対話後評価の方が記憶に頼る割合が大きいため評価が安定しないことが原因として考えられる。

2の無駄時間および発話量については、どちらの分析結果がその被験者の特性を示しているかはデータから明らかにできないが、先の考察とあわせて対話聴取後評価の方がより信頼性が高いと予測される。

3のリズムについては、先の効率についての考察と同様、対話聴取後評価よりも対話後評価で個人差が多く見られるのは、対話後評価の方が記憶に頼る割合が大きいため評価が安定しないことが原因として考えられる。

5のバランスについては、2の無駄時間と同様のことが言える。6の主体性についても同様である。

次に、相手の話し方についての指標に関して、以下の様にまとめる。

1. 効率の重要性は、対話後評価では個人差があるのに対して対話聴取後評価では個人差が少ない。
2. 透過性の重要性は、いずれの評価においても個人差が見られる。
3. リズムの重要性は、いずれの評価においても個人差が少ない。

1の効率については、対話についての指標における考察と同様のことが言える。

3のリズムについては、対話後評価では総合的評価との相関がどの被験者も高いのに対

して、対話聴取後評価では一人を除いてどの被験者も低い。これはリズムが経験的あるいは感覚的な指標であり、対話を実際に行なう方がより実感できるためと考えられる。

なお、詳細指標については、対話後の評価結果のサンプル数不足により比較分析ができないためここでは結論を控えることにする。

(3) 仮説 H_3 の検証結果についての比較

また、3.3.3 節の仮説 H_3 の検証結果について、対話後および聴取後の比較して以下のようによまとめられる。

まず、対話についての指標に関して以下の様にまとめる。

1. 効率の指標について、いずれの評価においても論文検索効率と無駄時間は総合指標との相関が強く、発話量は相関が弱い被験者もいる。
2. リズムの指標について、いずれの評価においても間は総合指標との相関が強く、バランスは相関が弱い被験者もいる。

次に、相手の話し方についての指標に関して、以下の様にまとめる。

1. 効率の指標について、対話後評価では無駄時間も発話量もどの被験者においても総合指標との相関が強いが、対話聴取後評価では相関が低い被験者がいる。
2. リズムの指標について、間もスピードも、対話後評価ではどの被験者においても総合指標との相関が強いが、対話聴取後評価ではそうでない被験者もいる。

処理状態の透過性および意思伝達の透過性については対話後評価結果のサンプルが不足しているため、比較が困難である。また、安心感については先の分析によって分類の再考を要すると考えられる。

2 のリズムについては先に述べた様に、リズムが経験的あるいは感覚的な指標であり、対話を実際に行なう方がより実感できるためと考えられる。

3.3.5 評価指標についての知見

これまでの分析により得た知見を以下にまとめる。

(1) 効率

- 効率およびその詳細指標については、聴取評価による代替が可能である。
- 効率およびその詳細指標については、聴取評価の方がより安定した結果が得られる。
- 効率およびその詳細指標の対話聴取後評価には、評価者本人の対話への参加・不参加が影響しない。
- ただし、相手の話し方における発話量の評価に限ってはその影響が見られるため、評価者の対話への参加・不参加を考慮する必要がある。
- 効率およびその詳細指標の重要性は評価者に関わらず高い（ただし対話後評価ではその様に現れない場合もある）。
- 論文検索効率は、意識的にも重要視される傾向にある。

(2) 透過性

- 透過性の評価には、評価者本人の対話への参加・不参加が影響する。
- 処理状態の透過性の評価には評価者本人の対話への参加・不参加が影響しないが、意思伝達の透過性および安心感の評価には影響する。
- なお、安心感は透過性の下位分類として相応しくない。

(3) リズム

- リズムの評価には、評価者本人の対話への参加・不参加が影響する。

3.3.6 論文検索タスクにおける機能目標の分析

25 種類の指標のうちの相手の話し方についての指標で、総合指標（表 3.2 の総合評価 (b)）に対して大きな影響を与えている指標を調べた結果を表 3.28 に示す（総合評価 (b) を目的変数、その他の指標を説明変数とした重回帰分析による結果である）。

この結果から、相手の話し方の主観評価において、効率指標が総合指標に高い影響力を持つことが5人中4人の被験者において5%の有意水準で示されている。したがって、論文検索タスクで今回設定した状況における対話では、効率が最も重視される指標であると言える。

表 3.28: 総合指標とその他の指標の重回帰分析
(**は1%、*は5%の有意水準で優位性が認められたことを示す。
また、A~Eは被験者を示す。)

対話指標	A	B	C	D	E
効率 (b)	*		*	**	*
透過性 (b)				*	*
リズム (b)				*	
主体性 (主導権) (b)					
話し方の多様性 (b)					

3.4 効率の主観評価と客観量の相関分析

次に、データベース検索タスクにおける機能目標としての効率評価向上のための対話制御を実現するために、効率の主観評価と客観量の相関関係を明らかにする。

まず始めに、分析の対象とした対話データの概要を表 3.29 に、それに対して対話参加者自身が行なった主観的評価結果を表 3.30 に示す (時間の単位は [sec])。†

表 3.29 において、発話単位は 400msec のポーズで区切られる音声区間をさらに発語内行為 [29] によって分割統合したものである。なお、あいづちやフィラーなど、発語内行為を与えられない発話は対象外とした (対話時間には含まれている。)

実験では各被験者が3対話ずつ対話を行ない、独立に分析・評価した。つまり、これらの表に示した対話番号のうち、1、2、3は一人の被験者によるもの、という様になっている。

以下には、これらのデータを元に効率について分析した結果を示す。

†なお、対話収録環境は 3.2 の表 3.1 に主観的評価に用いた質問文は表 3.2 に示してある。

表 3.29: 対話データ 99d01 の概要

対話番号	発話数 S	発話数 U	発話時間 S	発話時間 U	対話時間
1	10	6	28.57	13.38	58.40
2	20	12	87.27	37.62	177.56
3	6	5	18.21	9.59	35.685
4	14	8	72.45	18.03	126.96
5	15	8	74.15	18.86	143.63
6	23	14	93.57	62.97	215.03
7	13	7	75.07	15.30	142.84
8	18	9	79.02	165.9	240.20
9	9	3	52.45	19.24	113.95
10	22	11	83.84	33.84	278.56
11	20	11	69.75	40.71	320.22
12	23	17	93.52	41.01	279.73
13	22	9	126.50	19.24	268.15
14	13	9	61.34	18.89	169.62
15	20	10	92.11	33.08	228.10
Total	248	139	1107.83	547.70	2798.63
Average	16.53	9.27	73.86	36.51	186.57

3.4.1 全体的な傾向

まず「効率」について、タスク効率と対話効率に分けて定義する。つまり、音声対話システムによってタスクを遂行する際、音声対話はあくまでユーザの要求を達成するためのサブタスクの一つであり、タスクそのものではないとする考え方である。タスク効率は「いかにタスクを効率良く達成できるか」を示し、対話効率は「いかに対話を効率良く達成できるか」を示す。

ここで、論文検索タスクは、図 3.2 に示す様に、「ユーザの検索要求 (Demand) を迅速かつ正確に推定し、時には曖昧な検索要求を具体化 (embodiment) し、得られた検索要求 (Condition) に対して検索 (search) を繰り返しその結果 (Result) を提示することで、検索要求を満足させるタスク」であると言える。このタスクでは、タスク効率は検索要求の具体化能力や検索能力（あるいはデータベースの性質）に依存し、対話効率はこうした能力とは独立した対話の進行能力に依存する。音声対話検索システムの「効率」を向上するには、これらのいずれも向上する必要があると考える。

しかしながら、これらの効率を定量的に評価することは容易であるのに対して主観的に独立に評価することは難しい。実際に 3.3 の人間同士の対話評価においては、表 3.17 で示

表 3.30: アンケートによる主観評価結果 (5段階評価)

評価指標	記号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
総合評価 (a)	Xa	5	3	5	5	5	1	4	2	4	4	3	5	3	3	1
総合評価 (b)	Xb	5	4	5	5	5	3	4	4	4	4	3	4	4	3	4
効率 (a)	Aa	5	4	5	3	5	1	3	2	4	4	2	2	2	3	1
効率 (b)	Ab	5	4	5	3	5	3	4	4	4	5	2	3	3	4	3
論文検索の効率 (a)	A1a	5	4	5	2	4	1	2	2	4	4	2	1	2	2	2
無駄時間 (a)	A2a	4	2	5	2	4	1	4	2	4	3	1	2	2	4	1
無駄時間 (b)	A2b	4	4	5	3	5	2	4	4	4	5	2	3	3	3	3
発話量 (a)	A3a	4	4	5	3	5	2	4	4	3	5	2	3	3	3	3
発話量 (b)	A3b	4	4	5	3	5	1	2	4	3	5	2	3	3	3	3
透過性 (b)	Bb	5	4	5	5	4	5	4	2	4	5	3	4	2	2	2
処理状態 (b)	B1b	5	5	5	5	3	5	3	4	4	5	3	5	3	2	4
意思伝達 (b)	B2b	5	4	5	5	5	5	2	3	4	4	4	5	4	2	2
安心感 (b)	B3b	4	3	5	5	5	1	2	4	4	5	2	5	2	2	2
リズム (a)	Ca	4	2	4	4	5	2	3	4	4	3	2	4	3	2	2
リズム (b)	Cb	4	3	4	3	5	2	4	4	4	4	2	4	3	3	3
間 (a)	C1a	4	2	4	4	5	2	3	4	4	3	2	4	3	2	2
間 (b)	C1b	5	4	5	5	5	1	3	4	4	4	2	4	3	3	3
スピード (b)	C2b	5	5	5	5	5	5	4	4	4	5	4	5	3	3	4
バランス (a)	C3a	4	4	5	3	5	3	2	2	4	4	3	4	3	3	1
主体性 (主導権) (a)	Da	4	4	4	3	5	1	2	2	4	4	3	4	3	3	1
表現の多様性 (b)	E1b	4	5	3	3	5	4	4	4	2	3	3	3	2	2	3
話し方の多様性 (b)	E2b	4	5	4	5	3	4	5	3	2	3	3	3	3	3	3
リズム (b)	E3b	3	4	3	5	5	4	4	2	2	3	2	3	3	3	3

した結果から、「対話についての効率」すなわち対話効率と「論文検索効率」すなわちタスク効率の主観評価の相関係数が0.91と非常に高いことがわかっている。従ってここでは対話制御への利用という目的を考慮して、対話効率を中心に分析を行ない、タスク効率の分析は補足的に行なうこととする。なお主観評価においては表3.30で示したA1aがタスク効率にあたり、Abが対話効率にあたる。

以下では、まず対話効率について種々の客観量との関係を分析し、次にタスク効率について同様の分析を行なう。

(1) 対話効率との相関

対話効率と発話数

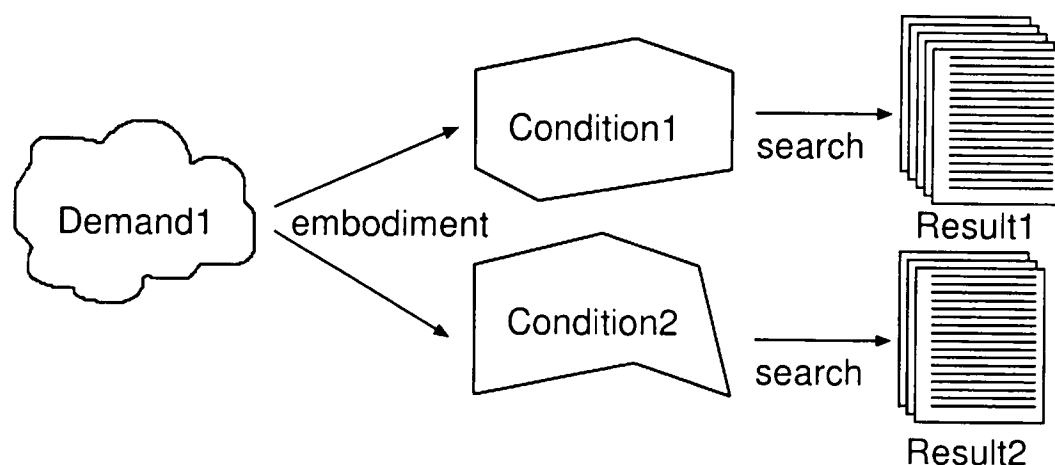


図 3.2: 論文検索タスク

まず、対話効率の主観評価 (Ab) と発話数との相関を調べた結果を表 3.31 に示す。

発話数に関してはいずれも対話効率との強い負の相関が見られているが、全発話数に対するシステム発話数の割合は相関が低い。対話における発話数の大小が対話効率の主観的評価に影響を与えるのは直感的にも予測がつくが、システム発話の割合が効率評価に影響を与えないという結果は意外と言える。

システム発話数の方がユーザ発話数より負の相関が強いのは、システムの対話方法の評価との相関だけに妥当な結果と言える。

対話効率と発話時間

次に、対話効率の主観評価 (Ab) と発話時間との相関を調べた結果を表 3.32 に示す。

この表においては、発話数の場合と同様の傾向があらわれている。総発話時間が最も負の相関が強く、ついでシステム発話時間、ユーザ発話時間の順に強い負の相関が見られ、システム発話の割合にはほとんど相関が見られない。

対話効率と対話時間

次に、対話効率の主観評価 (Ab) と対話時間との相関を調べた結果を表 3.33 に示す。

対話時間とは、対話における最初の話者の発話開始時から最後の話者の発話終了時までの時間のことであり、前項の発話時間における総発話時間に対して、ポーズ部分を含む分

表 3.31: 対話効率の主観評価と発話数

対話番号	Ab	S 発話数	U 発話数	総発話数	S 発話数の割合 (%)
1	5	10	6	16	62.5
2	4	20	12	32	62.5
3	5	6	5	11	54.5
4	3	14	8	22	63.6
5	5	15	8	23	65.2
6	3	23	14	37	62.2
7	4	13	7	20	65.0
8	4	18	9	27	66.7
9	4	9	3	12	75.0
10	5	22	11	33	66.7
11	2	20	11	31	64.5
12	3	23	17	40	57.5
13	3	22	9	31	71.0
14	4	13	9	22	59.1
15	3	20	10	30	66.7
Total	—	248	139	387	—
Average	3.8	16.53	9.27	25.8	64.2
Correl(Ab)	—	-0.54	-0.48	-0.54	-0.13

長くなっている[†]。

全対話時間と対話効率の主観評価との相関はきわめて強い。対話時間の長短が、対話自体の評価だけでなく相手の話し方の評価に大きく影響することがわかる。つまり、話し方の効率を評価する際、どの程度話していたかということ以上に、どの程度無駄な時間があつたかということの影響が強いと考えられる。このことを裏付ける様に、対話時間から発話時間を引いた無音時間（厳密には重複の考慮が必要であるが）は、表 3.32 で示した総発話時間よりも対話効率の主観評価との相関が強い。

(2) タスク効率との相関

次に、タスク効率の主観評価 (A1a) と各客観量との相関を分析する。

タスク効率と発話数

まず、タスク効率の主観評価と発話数との相関を調べた結果を表 3.34 に示す。

[†]ただし総発話時間では重複を考慮していないため総発話時間の方が対話時間より長くなる可能性もある。

表 3.32: 対話効率の主観評価と発話時間

対話番号	Ab	S 発話時間	U 発話時間	総発話時間	S 発話の割合 (%)
1	5	28.57	13.38	41.95	68.1
2	4	87.27	37.62	124.90	69.9
3	5	18.21	9.59	27.80	65.5
4	3	72.45	18.03	90.48	80.1
5	5	74.15	18.86	93.01	79.7
6	3	93.57	62.97	156.50	59.8
7	4	75.07	15.30	90.37	83.1
8	4	79.02	45.91	124.90	63.3
9	4	52.45	19.24	71.69	73.2
10	5	105.84	33.84	139.70	75.8
11	2	69.75	40.71	110.50	63.1
12	3	93.52	41.01	134.50	69.5
13	3	126.50	20.24	146.80	86.2
14	4	61.34	18.89	80.23	76.5
15	3	92.11	33.08	125.20	73.6
Total	—	1108.0	428.7	1536.7	—
Average	3.8	73.9	28.6	102.4	72.2
Correl	—	-0.54	-0.48	-0.59	0.05

発話数に関して、対話効率の場合以上に、強い負の相関が見られている。

タスク効率と発話時間

次に、タスク効率の主観評価と発話時間との相関を調べた結果を表 3.35 に示す。

発話時間に関して、発話数の場合と同様に、対話効率の場合以上の強い負の相関が見られている。

タスク効率と対話時間

次に、タスク効率の主観評価と対話時間との相関を調べた結果を表 3.36 に示す。

対話効率の場合と比較して、対話時間は若干相関が強くなり、無音時間は相関が弱くなっている。

(3) 対話効率およびタスク効率に関する結論

この様に、ほとんど全ての客観量が、対話効率以上にタスク効率との相関が強い。これは、先にも述べた様に、これら 2 つのレベルの効率を主観的に独立に評価することは難し

表 3.33: 対話効率の主観評価と対話時間

対話番号	Ab	全対話時間	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
1	5	58.40	3.65	16.45	1.03
2	4	177.56	5.55	52.67	1.65
3	5	35.69	3.24	7.89	0.72
4	3	126.96	5.77	36.48	1.66
5	5	143.63	6.24	50.62	2.20
6	3	215.03	5.81	58.49	1.58
7	4	142.84	7.14	52.47	2.62
8	4	240.20	8.90	115.27	4.62
9	4	113.95	9.50	42.26	3.52
10	5	278.56	8.44	138.88	4.21
11	2	320.22	10.33	209.76	6.77
12	3	279.73	6.99	145.20	3.63
13	3	258.15	8.65	121.4	3.92
14	4	169.62	7.71	89.39	4.06
15	3	228.10	7.60	102.91	3.43
Total	—	2798.6	105.5	1240.1	45.3
Average	3.8	186.6	7.04	82.7	3.02
Correl	—	-0.62	-0.48	-0.58	-0.50

表 3.34: タスク効率の主観評価と発話数

	S 発話数	U 発話数	総発話数	S 発話の割合 (%)
Correl	-0.61	-0.61	-0.63	0.00

い上、これらの客観量が対話の進行能力以上に検索能力などの影響を大きく受けるためと考えられる。ただし、ここで示した客観量のうち唯一、対話効率との相関の方がタスク効率との相関よりも強かった無音時間については、検索能力よりも対話の進行能力により大きく影響を受ける、対話効率を客観的に測定するのに重要な尺度であると考えられる。

表 3.35: 対話効率の主観評価と発話時間

(時間の単位は [sec])

	S 発話時間	U 発話時間	総発話時間	S 発話の割合 (%)
Correl	-0.58	-0.55	-0.65	-0.01