

共同作業タスクにおける対話の分析

田本 真詞 川端 豪

NTT基礎研究所

〒243-01 神奈川県厚木市森の里若宮 3-1

tamoto@idea.ntt.jp

あらまし 音声対話システムと人間との快適なコミュニケーションを実現させるには、音声対話システムに対話を協調的に進めるための機構が必要と考えられる。そこで人間同士の対話記録の分析から対話の協調的機構に関わる知識を獲得し、対話システムへ応用することが検討されている。本研究では、実際の音声対話における間投詞、終助詞などの発話の開始・終了符号、あいづち・復唱などの応答などの対話の調整やそのふるまいなどを観察し、対話の協調的機構のための知識の獲得を検討する。特に、対話の調整の観察のためのタスクとして目的指向型協調作業に着目し、タスクと同時発話や言い差し、間投詞的応答など発話権の移動に関わる対話の調整との関連を分析する。

キーワード 音声理解、自然言語処理、音声対話データベース、JUNO

Investigation of Cooperative Task Dialogue

Masafumi TAMOTO Takeshi KAWABATA

NTT Basic Research Laboratories

3-1 Morinosato Wakamiya, Atsugi-city, Kanagawa 243-01 Japan

tamoto@idea.ntt.jp

Abstract This report describes some feature of dialogue coordination that convey intentional and operational information of utterances. For constructing effortless speech conversation system, it is necessary to implement the coordination mechanism for dialogues. To analyze the dialogue coordination, we collect two kinds of task-oriented cooperative speech dialogues named Client-Manipulator task and Client-Advisor task. We then analyze the correlation of the phenomenon concerning turn-taking with dialogue coordination in various dialogue tasks and conditions.

key words Speech Understanding, Natural Language Processing, Speech Dialogue Database, JUNO

1 まえがき

音声を用いた会話が人間同士のコミュニケーションの手段として日常的な習慣となっていることから、人間と情報を交換するシステムのインターフェースのひとつとして、音声会話の利便性は高いものと考えられる。そこで、インターフェースとしての音声対話システムに人間にとって快適なコミュニケーションを実現させるには、対話を協調的に進めるための機構が必要となる。そのために人間同士の対話記録を分析し、得られる知識を協調的機構の構築に利用することが検討されている。

日常の音声対話では、発話のプランニング [1] や実時間下での生成 [2]、伝達の不確かさなどの制約から対話の動機となる情報伝達とは別に話題の継続や転換、発話権の確保、維持、放棄に関する情報や、伝達された情報や発話生成の内的な処理状況に関する情報などが重畳されている [3]。

これらの情報は、発話の意図や目的を伝達する協調的機構として、談話構造や発話構造のパターン系列 [4] や発話速度、韻律などの変化 [5, 6]、また、対話の進行中に情報伝達そのものを確実なものにしていく機構として、問投詞、終助詞などの発話の開始・終了符号、あいづち・復唱などの応答などの対話の調整 [7, 8] によって表出されると考えられる。

これまでに談話構造、発話プランニングなどの機構に関してさまざまな検討が行なわれているが、対話の調整に関する表出や言い換え、言い淀み、言い直し表現や言い射しなど非文や不適格文の要素を扱う枠組は、人間と音声対話システムの間には違和感のない対話を実現するために不可欠のものと考えられる [9, 6]。

調整のための表出が対話の目的や進展状況、コミュニケーションの環境に応じてどのようにふるまうかを観察すれば、これらの表出の担う情報と情報のやりとりによって達成される対話の協調に関する知識を得ることが出来る。

よって、タスクや環境を様々に設定した音声対話実験による試行を通して、対話の調整のための表出の観察が重要であると考えられる。本研究では、対話の協調的機構に関する知識を獲得するために実際の音声対話を通して、対話の調整やその表出のふるまいなどの観察を行なう。同時に、対話の調整やその表出のふるまいのモデルを検討する。

音声対話は、複数話者による発話の系列で構成されている。発話と発話の間には、息継ぎや小休止によって無音区間（ポーズ）が挿入されることがある。これらの無音区間のふるまいは、生理的必要性だけではなく、統語・意味・語用論的な制約を受けていると考えられる。「間」や「呼吸」など対話の様子の時間的側面を表す慣用語は、これらが対話を進める上でなんらかの調整機能をはたすことを示している。

本報告では、ポーズに加え、言い淀みや、言い差し・

割り込みなどの現象を発話権管理の観点からとらえ、発話権管理のために、発話系列の時間的構造に反映される対話の調整機能を分析する、また、あいづちや問投詞の表現などの他の対話の調整の表出と発話系列の時間的構造の間の関連を検討する。

2 音声対話データ

音声対話データには、音声対話データベース構築のために設定した目的指向型協調作業の収録結果を用いる。目的指向型協調作業では、話者の間で立場や目的に関する知識が明らかでないため、対話の展開やその終了が明確かつ予測できる。また、異なる設定で実験を反復することで対話の変化をとらえやすく、タスクによって語彙のある程度制限することができる。

2.1 対話タスクの設定

本報告では、対話参加者の役割付けの異なる操作タスク (Client-Manipulator task) と、助言タスク (Client-Advisor task) 各課題の対話記録を用いて、課題の違いによる対話構造の変化とそれともなう対話の調整のふるまいの変化を比較した。

タングラムタスク (操作タスク)

三角形、四角形、五角形からなる数片のタイルを用いてゴールとなる図形を組み立てる。client に提示されるパズルの完成図は、全体または一部がシルエットで描かれており、タイルの形状を考えながら配置を推定しなければならない。manipulator がタイルを扱うため、タイルの形状・配置について情報の交換が必要となる。

タングラム課題では、双方の表情の映像に加え、操作者の手元にあるタイルを指示者の側に映像で提示している。

地図タスク (助言タスク)

地図タスク (Map Task) [10] は、Edinburgh 大学で種々の対話現象の観察のために設定されたタスクで、すでに多くの分析結果が提示されている。また、同様の実験が日本語で試みられている [11, 12]。地図タスクでは、二人の被験者が地図を持ち、あらかじめ経路の記入された地図を持つ information giver が経路の記入されていない地図を持つ information follower に経路情報を再現させるものである。information giver の持つ地図には、出発・目標点を結ぶ経路と名前をついた特徴物が提示される。information follower の地図には、出発・目標点と特徴物のみ提示される。

2.2 対話収録システム

話者はそれぞれが別の防音室に収容され、外界やお互いの部屋とは隔離されるため、音響的・視覚的な情報を直接伝達することはできない。

音声の伝達は話者の襟に取り付けられたラベリア (仕込み) マイクロフォンとイヤレシーバを介して行なわれる。話者の正面にはディスプレイを設置し、相手側正面に設置されたカメラで撮影した映像を表示する。これにより話者はお互いの上半身及び表情の動きを観察できる。話者の前面の卓上には共同作業に使うオブジェクトを配置し、作業の様相を上方からカメラで撮影する。作業の映像はタスク設定に応じて相手側に提示する。

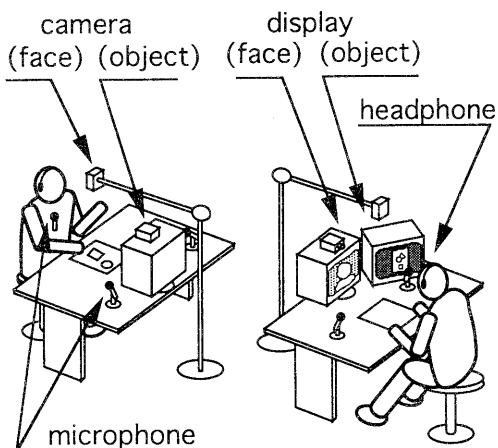


Figure 1. Layout for dialogue data collection.

2.3 被験者の組合せ

同じ2名の被験者ペアを用いて各タスクについて収録を行なった。また、被験者の違いによる変動をみるため、被験者ペアの双方と第3の被験者による比較対象用の収録をあわせて行なった。被験者同士の親近性は高い。

3 音声データの書き起こし

話者ごとに収録された音声を、文節単位の発話に区切り、話者番号、開始時刻、終了時刻とともに列挙する。

時刻で表される発話の順序関係に加え、割り込みや、挿入の生じた時点を明確にするため、表記に[番号]で示される同期ラベルを付加する。さらに、文節内のポーズや、言い淀みを記述するために、ポーズの開始・終了時刻の情報を別の行に表し、表記に<番号>で表されるポーズラベルを付加する。これらのラベルによって、発話の順序関係や発話内外のポーズの区別を明記できる。

発話終端などの長母音化を「ー」、補助的な情報として、書き起こし作業者の聴取によるピッチ上昇を「↑」で表す。これらの規則によって記述された書き起こし結果を次に示す。

-
- 1 126.28 127.06 みぎて [1] の一、
 - 2 126.91 127.10 [1] うん。
 - 1 127.58 128.92 ひとさしゆび <#1> の、
 - 1 128.36 128.72 <#1>
 - 1 128.92 130.30 そのへん <#2> [1] を一、
 - 1 129.50 129.94 <#2>
 - 2 129.88 130.06 [1] ん。
 - 1 130.82 131.96 [2] うん、みずいろの一、
 - 2 130.79 131.25 [2] こっち↑
 - 2 131.84 132.08 うん。
 - 1 133.26 134.22 したのへんに一、
 - 1 135.50 136.03 くっつけて。
 - 1 136.30 136.56 うん、
 - 1 136.56 137.41 したがわのへん。
-

4 発話権管理からみた発話系列の時間的構造

発話権の管理における機能 [13] として、次表の項目を考える。

Table 1. 発話権の管理における機能

開始:	発話権の開始
維持:	自己発話権の維持
確認:	自己発話権の確認
譲渡:	相手話者への発話権の譲渡
獲得:	相手発話者からの発話権の獲得
承認:	他者発話権の維持

また、発話権の管理における機能の発話系列の時間的構造への分類を次のように考える。

発話権の維持

一方の話者による発話が無音区間を通して接続する。

発話権の譲渡

一方の話者の発話終了直後、ないし無音区間を通して他方の話者の発話が接続する。

発話権の獲得

一方の話者の発話終了直後、ないし発話区間の重複(オーバーラップ)を生じて他方の話者の発話が接続する。

5 発話系列の時間的構造の分類

同一話者ペアによる二つの協調作業タスク、タンگرامタスク(操作タスク)地図タスク(助言タスク)の発話間のポーズ、オーバーラップの分布を分析した。

発話権の維持

以下に示す表に同一話者による発話接続間のポーズ継続時間長と出現頻度を示す。書き起こし作業者による精度のゆらぎを考慮して50msec以下のポーズを省いた値を示す。50msec以下ないしポーズ無しで接続する発話接続は、タングラムタスクで話者1が56回、話者2が17回出現し、地図タスクで話者1が132回、話者2が14回出現している。また、1sec以上のポーズは、まとめて表示している。

Table 2. 発話権の維持 (タングラムタスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	27	28.4	2	22.2
0.10 ~	9	9.5	2	22.2
0.15 ~	4	4.2		
0.20 ~	5	5.3		
0.25 ~	4	4.2	1	11.1
0.30 ~	6	6.3		
0.35 ~	2	2.1	2	22.2
0.40 ~	2	2.1		
0.45 ~	4	4.2		
0.50 ~	1	1.1		
0.55 ~	2	2.1		
0.60 ~	1	1.1		
0.65 ~	1	1.1		
0.70 ~	2	2.1		
0.75 ~	1	1.1		
0.80 ~	3	3.2	1	11.1
0.90 ~	3	3.2		
1.00 ~	18	19.0	1	11.1
合計	95	-	9	-

Table 3. 発話権の維持 (地図タスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	39	24.4	7	31.8
0.10 ~	20	12.5	3	13.6
0.15 ~	8	5.0	2	9.1
0.20 ~	6	3.8	4	18.2
0.25 ~	9	5.6	2	9.1
0.30 ~	14	8.8		
0.35 ~	9	5.6		
0.40 ~	4	2.5		
0.45 ~	6	3.8		
0.50 ~	7	4.4		
0.55 ~	8	5.0	1	4.5
0.60 ~	3	1.9	1	4.5
0.65 ~	3	1.9		
0.75 ~	3	1.9		
0.80 ~	2	1.2		
0.85 ~	3	1.9		
0.95 ~	1	0.6		
1.00 ~	15	10.0	2	9.1
合計	160	-	22	-

操作タスクであるタングラムタスクでは、指示者側(話者1)の発話権維持を示す発話接続の数が操作者(話者2)の数と際立った対比をみせている。これは、発話権の管理が指示者側に存在することをしめしている。異なり話者ペアによる対話記録でも同じ傾向が観察されている。

1sec以上のポーズが挿入されても話者1側で発話権が維持されている会話例を次に示す。

1 145.65 146.27 すとつぶって
 1 146.27 146.68 いうか [1] ら。
 2 146.57 146.82 [1] うん。
 2 146.82 147.56 ゆっくりね。
 1 147.49 147.75 うん。
 1 150.06 150.28 あー！
 1 150.28 150.55 すとつぶ。
 1 152.87 153.26 それでー、
 1 153.26 153.94 あの一、
 2 153.74 154.07 うん。

ここでは、操作者が指示者の監視のもとで、オブジェクト(タイル)を操作し、次の指示に移るまで2秒以上待機している。

助言タスクである地図タスクでの、情報提供者側(話者1)の発話権維持が情報追随者側(話者2)より多い傾向が示されている。

発話権の譲渡

次に発話権がポーズをはさんで移動する場合のポーズ長と頻度の分布を示す。話者1の欄は話者1から話者2への発話権の移動、逆に話者2の欄は話者2から話者1への移動の際のポーズ長の分布を表している。

発話権の維持ではポーズ長の分布が短い(~100msec)側にかたよる傾向が見られたが、発話権の譲渡ではポーズ長と頻度の間に強い関連が見られない。これは、発話権譲渡の要因が発話系列の時間的構造のみに依存しないためと考えられる。

発話の接続を観察すると、タングラムタスクでは、指示者側(話者1)の発話に対して操作者(話者2)があいづちを返す(21例、平均ポーズ長0.25sec)場合と指示に対して操作者(話者2)が指示代名詞などで質問を返す(12例、平均ポーズ長0.88sec)のふたつの分布が存在している。タングラムタスクでは、指示者側に操作者の手元にあるタングラムのタイル映像が提示されるので、指示表現が多くなっているものと思われる。

地図課題では、映像の提示がないため、指示代名詞による問いかけは表れず、あいづち(30例)が大半をしめている。

Table 4. 発話権の譲渡 (タングラムタスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	4	7.8	5	9.1
0.10 ~	6	11.8	4	7.3
0.15 ~	4	7.8	4	7.3
0.20 ~	2	3.9	4	7.3
0.25 ~	6	11.8	3	5.5
0.30 ~	2	3.9	4	7.3
0.35 ~	1	2.0	3	5.5
0.40 ~	2	3.9	3	5.5
0.45 ~	1	2.0	4	7.3
0.50 ~	6	11.8	3	5.5
0.55 ~	1	2.0	1	1.8
0.60 ~	2	3.9	3	5.5
0.65 ~	1	2.0	2	3.6
0.75 ~	2	3.9	4	7.3
0.85 ~	1	2.0	1	1.8
0.90 ~	1	2.0		
0.95 ~	1	2.0	1	1.8
1.00 ~	8	15.7	6	10.9
合計	95	-	9	-

Table 5. 発話権の譲渡 (地図タスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	3	5.3	7	15.2
0.10 ~	5	8.8	3	6.5
0.15 ~	7	12.3	7	15.2
0.20 ~	3	5.3	2	4.3
0.25 ~	2	3.5	1	2.2
0.30 ~	7	12.3	4	8.7
0.35 ~	3	5.3	3	6.5
0.40 ~	1	1.8	4	8.7
0.45 ~	3	5.3		
0.50 ~	3	5.3	2	4.3
0.55 ~	1	1.8	4	8.7
0.60 ~	2	3.5	1	2.2
0.65 ~	3	5.3	1	2.2
0.70 ~			1	2.2
0.75 ~	1	1.8	1	2.2
0.80 ~	2	3.5		
0.85 ~	1	1.8		
0.90 ~	2	3.5		
0.95 ~	3	5.3	1	2.2
1.00 ~	5	8.8	7	15.2
合計	57	-	46	-

発話権の獲得

最後に、相手の発話の途中で割り込んで発話権を獲得する場合のオーバーラップの間隔と頻度の関係を観察する。

オーバーラップは、短い時間 (~200msec) に集中している。多くは、最初の話者の発話終端が長母音化したときに次の発話が始まったり、あいづちが挿入され

ることで生じている。比較的長いオーバーラップは、発話が同時に始まる時、相手の発話を復唱するときに生じている。

Table 6. 発話権の獲得 (タングラムタスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	6	19.4	5	13.5
0.10 ~	4	12.9	5	13.5
0.15 ~	3	9.7	3	8.1
0.20 ~	1	3.2	2	5.4
0.25 ~	4	12.9	1	2.7
0.30 ~	1	3.2	3	8.1
0.35 ~	1	3.2	5	13.5
0.40 ~	2	6.5	3	8.1
0.45 ~	4	12.9	1	2.7
0.50 ~	1	3.2		
0.55 ~	1	3.2	1	2.7
0.60 ~			1	2.7
0.65 ~	1	3.2		
0.70 ~			1	2.7
0.75 ~			1	2.7
0.85 ~			1	2.7
0.90 ~	1	3.2	2	5.4
1.00 ~	1	3.2	2	5.4
合計	37	-	31	-

Table 7. 発話権の獲得 (地図タスク)

ポーズ [秒]	話者 1		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	10	19.6	5	12.2
0.10 ~	4	7.8	2	4.9
0.15 ~	7	13.7	5	12.2
0.20 ~	6	11.8	3	7.3
0.25 ~	2	3.9	4	9.8
0.30 ~	4	7.8	2	4.9
0.35 ~	2	3.9	3	7.3
0.40 ~	1	2.0	2	4.9
0.45 ~	2	3.9		
0.50 ~	2	3.9	1	2.4
0.55 ~	1	2.0	3	7.3
0.60 ~	3	5.9	1	2.4
0.65 ~			3	7.3
0.70 ~			2	4.9
0.75 ~	1	2.0		
0.85 ~	1	2.0		
0.90 ~			1	2.4
0.95 ~	1	2.0		
1.00 ~	4	7.8	2	4.9
合計	51	-	41	-

6 課題の設定・話者性による変動

タングラム課題や地図課題のように役割付けの明確なタスクでは、発話権維持の観察にみられるような発話権の管理の偏りが生じやすい。そこで、このような

役割付けによる発話権管理の偏りの傾向を調べるためにこれまでの被験者ペアとは異なる被験者との組み合わせの対話データを分析した。

被験者の入れ換えによって、地図課題の発話権維持・獲得におけるポーズ長の分布に変化を生じた例を示す。

Table 8. 発話権の維持 (地図タスク)

ポーズ [秒]	話者 3		話者 2	
	頻度	[%]	頻度	[%]
0.05 ~	23	35.9	28	45.9
0.10 ~	7	10.9	6	9.8
0.15 ~	1	1.6	8	13.1
0.20 ~	2	3.1	2	3.3
0.25 ~	7	10.9	2	3.3
0.30 ~	3	4.7	2	3.3
0.35 ~	7	10.9	3	4.9
0.45 ~	2	3.1	2	3.3
0.50 ~	6	9.4	2	3.3
0.60 ~	4	6.2	1	1.6
0.75 ~	1	1.6	1	1.6
0.80 ~	1	1.6	2	3.3
合計	64	-	61	-

情報提供者と情報追随者の発話権維持の数がほぼ等しくなっている。さらに、発話権獲得では情報提供者が57例の重複を生じているのに対し、情報追随者が120例の重複を生じている。このばあい、情報追随者が問題解決の主導権を確保しており、発話権の管理に積極的に関わっていることが、書き起こしから観察された。

7 まとめ

対話の中で共有されるべき情報とは独立に、対話を円滑に進める協調的機構として対話の調整ととりいれ、発話系列の時間的構造に対話の調整として作用する機能があるかを検討した。発話系列の時間的側面を発話権の管理における機能にそって分類し、発話権の管理における対話の調整の作用について分析した。分析の結果、発話権の維持と譲渡において、ポーズ継続長と頻度の分布に差が見られた。また、そのような分布を支配する発話接続の構造について知見を得た。

さらに、目的指向型協同作業のように、あらかじめ役割付けの行なわれているタスクでは、役割に合わせて発話権の管理に偏りが生じることが観察された。逆に、発話権の維持を高め、積極的に相手から発話権を獲得する行動が協同作業全体の主導権と密接に関連することがわかった。

今後は、発話接続の時間的構造を中心にあいづちや間投詞的表現などの対話の調整の表出とを関連づける、対話の調整モデルの検討を進めたい。

謝辞

音声対話理解の研究機会を与えてくださった、N T T基礎技術総合研究所、基礎研究所情報科学研究部、石井健一郎部長に謝意を表します。また、日頃活発に御討議いただく誉田研究グループの研究員の皆様、アドバイスをいただいている対話理解研究部の皆様に感謝いたします。

参考文献

- [1] 金水敏. 知識モデルに基づく音声対話の分析. 電子情報通信学会技術報告, Vol. SP93-136, pp. 61-68, 1 1994.
- [2] 堂坂浩二, 島津明. 時間制限下における漸次的発話生成. 電子情報通信学会技術報告, Vol. NLC94-40, pp. 9-16, 12 1994.
- [3] 田窪行則. 音声対話の言語学的モデル: 談話管理標識としての感動詞の分析. 情報処理研究会報告, Vol. 94-SLP-1, pp. 15-21, 5 1994.
- [4] 知野哲朗, 坪井宏之. 談話構造モデルによる電話相談対話の分析. 情報処理研究会報告, Vol. 94-SLP-4, pp. 33-40, 12 1994.
- [5] 高木一幸, 板橋秀一. 音声対話における発話系列の韻律パターン. 日本音響学会講演論文集, pp. 213-214, 3 1994.
- [6] 中川聖一, 小林聡. 自然な音声対話における間投詞・ポーズ・言い直しの出現パターンと音響的性質. 日本音響学会誌, pp. 202-210, 3 1995.
- [7] 島津明, 川森雅二, 小暮潔. 対話の分析—間投詞的応答に着目して—. 情報処理研究会報告, Vol. NL95-9, pp. 65-72, 5 1993.
- [8] 片桐恭弘. 対話調整の分散処理モデル. 情報処理研究会報告, Vol. 94-SLP-2, pp. 1-8, 7 1994.
- [9] 竹沢寿幸, 田代敏久, 森元逞. 音声言語データベースを用いた自然発話の言語現象の調査. 人工知能学会研究会資料, Vol. SIG-SLUD-9403-3, pp. 13-20, 2 1994.
- [10] Henry S. Thompson, Anne Anderson, Ellen Gorman Bard, Gwyneth Doherty-Sneddon, Alison Newlands, and Cathy Sotillo. The HCRC Map Task Corpus: Natural Dialogue for Speech Recognition. In *HLT 93*, pp. 25-30. ARPA, 1993.
- [11] 青野元子. 地図課題コーパス (中間報告). 情報処理研究会報告, Vol. 94-SLP-3, pp. 25-30, 10 1994.
- [12] 青野元子 *et. al.* 千葉大学地図課題コーパス (千葉コーパス). 千葉大学, 1995.