

音声対話の発話交代に関わる現象の分析

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

Analysis of Speaker Transition in Cooperative Task Dialogues

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 an effortless speech conversation system, it is necessary to implement the coordination mechanism in spoken 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 analyze the relations of turn-taking behaviors to the dialogue coordination in various tasks and dialogue 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]、伝達の不確実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、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 [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できる。

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、そのために、あいづちや間投詞的表現などの他の対話の調整の発話系列の時間的構造の間の関連を検討する。さらに、発話片末尾の韻律的構造と、発話片の統語・談話構造的な分類を施すことで発話権の獲得と維持、譲渡を統御する発話間の構造について知見を得る。

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になため、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を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) の二つの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に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]。地図タスクでは、二人の被験者が地図を持つ 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の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。

3 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 [12] 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。

4 発話系列の時間的構造の分類

同一被験者ペアによる二つの協調作業タスク [13]、タングラムタスク（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

次頁に示す図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 (Figure 1.)。

操作タスクであるタングラムタスクでは、指示者側 (Client) の発話権維持を示す発話接続の数が操作者 (Manipulator) 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ている。異なり被験者ペアの対話記録でも同じ傾向が観察されている。助言タスクである地図タスクにおいても情報提供者側 (Giver) の発話権維持が情報追随者側 (Follower) より多い傾向が観察されている。

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 (Figure 2.)。

Client のグラフは指示者 (Client) から操作者 (Manipulator) への発話権の移動、逆に Manipulator の欄は操作者 (Manipulator) から指示者 (Client) への移動の際のポーズ長の分布を表している。

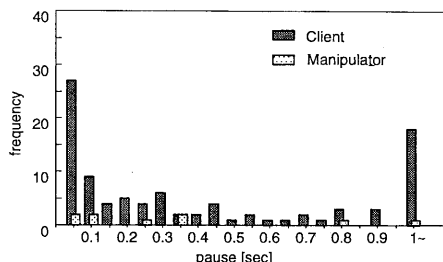


Figure 1. 発話権の維持 (タングラムタスク)

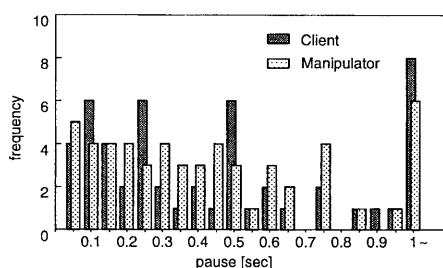


Figure 2. 発話権の譲渡 (タングラムタスク)

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 (Client) の発話に対して操作者 (Manipulator) 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 (Manipulator) 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考えられる。

一方、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) が大半をしめる結果が見られた。

発話権の獲得

最後に、相手の発話の途中に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示す (Figure 3, 4.)。

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、あるいは生じている。

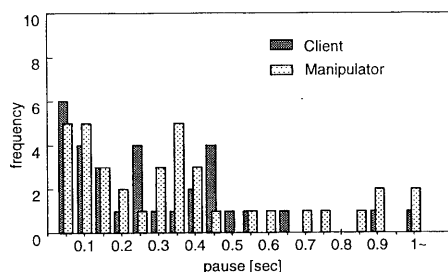


Figure 3. 発話権の獲得 (タングラムタスク)

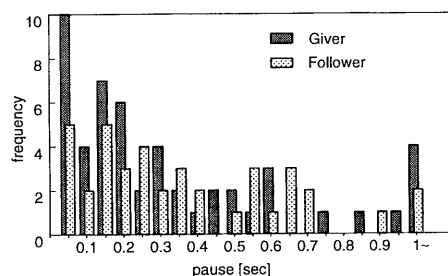


Figure 4. 発話権の獲得 (地図タスク)

5 発話末尾の韻律的情報

発話系列の時間的構造に着目した分析では、タスクや発話権管理の機能の違いによるポーズ長やオーバーラップの分布の差が観察された。しかし、ポーズ長やオーバーラップなどの時間的構造だけでは発話権の管理における機能、特に発話権の維持・譲渡・獲得における対話の調整としての働きを十分に説明することができない。

そこで、時間的構造に韻律的な情報を加え、発話権の維持・譲渡・獲得をモデル化する試みが行なわれている [14, 15, 16, 17]。

発話系列の時間的構造を分析した対話データから、さらに発話終端の2モーラのピッチを抽出し、時間的構造と韻律的な情報による分類が発話権の維持・譲渡・獲得にどのように関わるかを調べた。

5.1 韻律的構造レベルの分類

各発話を「自立語+付属語(列)」または、ポーズで区切られるなどして時系列的に孤立した「自立語」または「付属語」に断片化し、発話片ごとに終端の2モーラのピッチを抽出し、次の表に示される通り分類した (Table 2.)。

また、これらの情報と談話・統語レベルの知識との関連を調べるため対話に出現する語彙に対して次のような分類を行なった。

Table 2. 韻律的構造レベルの分類

ラベル	継続	交替
U 上昇	15	6
UU 上昇持続	3	14
F 平坦	7	3
D 下降	75	31
DD 下降持続	35	20
DF 下降平坦	7	6
FD 平坦下降	16	11
FU 平坦上昇	3	6
UD 上昇下降	45	30
DU 下降上昇	52	28

5.2 統語レベルの分類

先に示した通り、各発話を「自立語+付属語(列)」または、「自立語」、「付属語」に断片化し、発話片ごとに次の表に示される通り統語レベルの分類を行なった (Table 3.)。

Table 3. 統語レベルの分類

ラベル	例
名詞(句)	北の・図書館の・右側に
動詞(句)	行って・通りながら・まわってー
孤立した助詞	てー・をー・でー・んでー
接続詞	それで・だから・ほんで
間投詞	はい・うん・えーと・んー
副詞	そう・ずーっと・ちょっと・いま

5.3 談話標識レベルの分類

統語レベルで「副詞」、「接続詞」、「間投詞」と分類されていた語を、談話の構造化を中心とした意味的・語用論的機能に着目して次の表に示される通り再分類する [18] (Table 4.)。

Table 4. 談話標識レベルの分類

ラベル	例
フィラー的	あのー・えーと・んー
応答的	はい・うん・いいよー
接続的	じゃ・そしたら・それだから
様相的	ね・よ・さ・じゃん・てか

5.4 意味レベルの分類

さらに、時間的構造と意味的な情報に着目した分類として、言い直し、言いさし、復唱、唱和などの現象にラベルを与える (Table 5.)。

Table 5. 意味レベルの分類

ラベル	例
言い直し	みぎ、(んー)、ひだり
言いさし	まちの、(相手発話)、北をー
復唱	したにね、(相手発話)、した
唱和	りんごのはやし、(同時に)、りんごばやし

地図タスク (308 秒間、471 発話) の対話データの統語ラベルの頻度を表に示す (Table 6.)。

Table 6.

発話断片における統語／談話標識レベルの分類

統語ラベル	-	フィラー	応答	継続	様相
名詞	46				
名詞句	81				3
動詞	14				
動詞句	32				31
孤立した助詞	8				
接続詞				17	
間投詞	36	32	33		2
副詞	62	2	5		6

6 発話末尾の時間・韻律的構造と発話交代

発話片相互の時間的構造と韻律的な情報に加え、統語・談話レベルの分類を施して、発話権の維持・譲渡・獲得における関連を分析した結果、次のような特徴が見い出された。

上昇型ピッチで終了する発話

聴覚上は F_0 周波数が上昇していると思われる発話片末の多くは、上昇型ピッチに分類されなかった。

発話片の聴覚実験から聴覚上の F_0 周波数の上昇とピッチ情報が必ずしも一致していないことがわかっていて、これらの場合のピッチ変化は平坦か、やや下降していることが多い。聴覚上の F_0 周波数の上昇を表す条件をみいだすことで、疑問文末のピッチ上昇を抽出できると考えられる。

一方、上昇型と分類された発話片は、疑問の様相を表す助詞が付属しており、応答的発話による発話交代が生じるか、1sec 前後の長いポーズが後続して次の発話に遷移する。

下降型ピッチで終了する発話

「名詞＋助詞」で構成される発話片の多くは、下降型ピッチで終了する。これらの発話片は、オーバーラップで発話交代を生じる。

下降型ピッチで終了する発話に 200ms 以上のポーズが後続する場合、その発話が応答的でないし、継

続的談話指標を持つ発話でなければ発話交代が生じやすい。

下降型ピッチでは観測されるオーバーラップが、下降持続型ピッチではみられない。また、下降持続型ピッチの発話に後続するのは、応答的またはフィラー的な発話である。

平坦型ピッチで終了する発話

平坦型ないし平坦下降型のピッチは、発話片末の助詞の長母音化や間投詞に多く見られる。ただし、名詞の体言止めや動詞の終止形には見られない。平坦上昇型のピッチで終了する発話片は、上昇型ピッチの場合と同じように疑問の様相を表す助詞が付加するか、名詞などの自立語末のピッチが上昇する。

平坦上昇型のピッチも下降型のピッチと同様に 200ms 以上のポーズが後続する場合は発話交代が生じやすい。

上昇下降型ピッチで終了する発話

下降型ピッチで終了する発話と同様に「自立語＋助詞」で構成される発話片に 200ms 以上のポーズが後続する場合、応答的またはフィラー的な発話で交代が生じる。

下降上昇型ピッチで終了する発話

後続ポーズが 100ms 未満の場合、下降型ピッチで終了する発話と同様に「自立語＋助詞」で構成される発話片は交代を生じない。また、200ms 以上のポーズが後続する場合は、応答的またはフィラー的な発話で交代が生じる。

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、目的指向型協調作業のように、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調作業全体の主導権と密接に関連することがわかった。

さらに、発話片末尾の韻律的構造と、発話片の統語・談話構造の分類を施すことで発話権の獲得と維持、譲渡を統御する発話接続の構造について知見を得た。

今後は、発話接続の時間的・韻律的構造を中心にあ

いづちや問投詞的表現などの対話の調整の表出とを関連づける、詳細な対話の調整モデルの検討を進めた。

謝辞

音声対話理解の研究機会を与えてくださった、NTT基礎研究所情報科学研究部、石井健一郎部長、菅田研究グループリーダーに謝意を表します。また、日頃活発に御討論いただく菅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グループ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問投詞・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—問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 *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.
- [12] 白井克彦. 音声対話における認知モデル—発話権管理による割り込みへの対処—. 1994.
- [13] 田本真詞, 川端豪. 音声対話を通じた共同作業タスクの検討. 電子情報通信学会技術報告, Vol. SP95-36, pp. 89-94, 1995.
- [14] 小坂直敏. あいづちを中心とした会話音声の呼応関係の分析. 電子情報通信学会技術報告, Vol. NLC87-23, pp. 69-74, 1987.
- [15] 川森雅仁, 島津明. 対話における発話交代の分析. 電子情報通信学会技術報告, Vol. NLC95-73, pp. 31-38, 1996.
- [16] 小磯花絵, 堀内靖雄, 土屋俊, 市川熹. 先行発話断片終端部分に存在する次発話者に関する言語的・韻律的要素について. 電子情報通信学会技術報告, Vol. NLC95-72, pp. 25-30, 1996.
- [17] 堀内靖雄, 小磯花絵, 土屋俊, 市川熹. 自発音声対話における話者交代の制御に関わる発話末の統語的・韻律的特徴. 情報処理研究会報告, Vol. 96-SLP-10-9, pp. 45-50, 2 1996.
- [18] 川森雅仁, 川端豪, 島津明. 談話標識語の音韻的研究. 電子情報通信学会技術報告, Vol. NLC95-26, pp. 53-60, 1995.