音声対話インタフェースにおける発話タイミング制御とその評価

武藤　ゆみ子[*1]　高野　弘二[*1]　大良　宏樹[*1]
小林　洋平[*2]　山本　知仁[*3]　三宅　美博[*1]

Utterance Timing Control on Speech Dialog Interface and its Evaluation

Yumiko Muto[*1], Kouji Takano[*1], Hiroki Ora[*1],
Yohei Kobayashi[*2], Tomohito Yamamoto[*3] and Yoshihiro Miyake[*1]

Abstract - In this study, to develop the speech dialogue interface for the smooth dialogue with human, we discussed the structure of psychological effects of the utterance timing to the user. Concretely developing the simple speech dialogue avatar which never presents the non-verbal information such as facial expression, we evaluated the relationship between the avatar's speech timing and the user's subjective impression to the avatar. As the result, we found that the appropriateness of the avatar's speech timing is necessary to the user's good impression.

Keywords: dialogue, human-robot interaction, avatar, switching pause, timing control

1.　はじめに

「間（ま）」を合わせることは人間同士の協調作業において不可欠である．特に，音声対話において「間（ま）」の長さ（交替潜時）は人間の発話タイミングにより決定される．このことから，音声対話は人間が生来持っているタイミング機構と関係があることが推測される．本研究では，このような対話の時間的側面に注目し，人間と自然な対話を実現できるような対話インタフェースの設計を目標としている．

これまでわれわれの研究グループでは，コミュニケーションでの協調プロセスにおけるタイミング共有の重要性に注目し研究を進めてきた[1][2]．特に，音刺激に合わせて指でタップを行うタッピング実験を用いて，外部刺激に対し人間がどのように応答しているか調べることにより人間のタイミング機構を明らかにしてきた[3][4]．タイミングを合わせることは，他者との共同作業や円滑なコミュニケーションの達成において重要である．さらに，言語の獲得に先行して幼児の動作と母親の発話タイミングが同調することも知られている[5]．このようなことから，タイミング共有は人間のコミュニケーションの根底であることが推測される．

近年，コンピュータ技術の発展に伴い人間とのコミュニケーションを目的とした人工物が多く提案されてきている．そのなかで，人間と円滑で自然なコミュニケーションを実現する人工物の設計のために，人間同士の対話における頷き[6]や発話タイミング[7]が注目されている．これらは，話し手が頷きを行う相手が好意的に感じるという知見[8]や，「間」の長さ（交替潜時）や発話長などが同調する傾向があるという知見[9]に基づいており，コミュニケーションにおける対人的な共感行動の一型であると示唆されている．さらに，人間同士の音声対話において，発話タイミングにより決定される「間（ま）」は極めて豊かな感性情報を有し，この交替潜時の長さが話者の印象を形成する要因のひとつとなっていることも示されている[10][11]．このようなことから，ユーザー心理を考慮した音声対話インタフェースの構築には，発話などのタイミングを考慮することが重要であると考えられる．

しかしながら，人間同士の対話と人間-人工物の対話は，使用できるコミュニケーションチャンネルの制限から，その情報伝達のしくみが異なることが想定される．たとえば，人間同士の言語・非言語チャンネルを使用可能な「対面場面」と言語チャンネルのみ使用可能な「非対面場面」とでは，発話内容や発話頻度に差があることが示されている[12]．対面場面では，表情や身振りなどの非言語チャンネルから感性情報も含めた情報をやりとりすることができるのに対し，非対面場面では言語チャンネルに頼るほかないからである．このようなことから，人間同士の対話を解析する一方で，コミュニケーションチャンネルの制限をつけた人工物と人間の対話を解析し，その情報伝達のしくみを明らかにすることが対話を目的とした人工物設計のために必要である．

そこで本研究では，表情や身振りなどの非言語的チャンネルを持たないアバタを構築し，人間と人工物のやり取りで多く観察される，人間側がアバタに指示を与えア

*1: 東京工業大学　知能システム科学専攻
*2: 京都産業大学　*3: 金沢工業大学
*1: Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology
*2: Kyoto Sangyo University　*3: Kanazawa Institute of Technology

バタが応答するという場面を設定する．具体的には，アバタと被験者である人間との間にターゲットである積み木を用意して，人間が「それをとってください」と指示したのに対し，アバタが「はい」と応答する状況を用意する．そして，このアバタの「はい」の発話タイミングと発話長を制御することにより，人間側がアバタに抱く印象がどのように変化するのかを調査する．

2. 実験方法

2.1 被験者

被験者は，23 歳から 51 歳までの 12 人（男性 5 人，女性 7 人）であった．

2.2 実験装置

アバタは 40 型透過型スクリーンにプロジェクタによって表示され，アバタの音声はスクリーンの両サイドに置かれたスピーカーから提示された．アバタは表情や身振りを一切持たないものであった(図 1)．被験者はスクリーンから 1.5m の位置にアバタと対面して座り，被験者の音声は被験者の前に固定して置かれたマイクに向かって発せられた．アバタの表示と発話制御は，いずれも C++ 言語で構築されたプログラムを実行することにより行われた．実験中，被験者の手は膝の上に固定するように指示され，体を動かすことも禁止された．ターゲットとして，底辺 5cm，高さ 3cm，幅 3cm の赤い三角形の積み木がアバタの前に用意された． 被験者の斜め左前に，被験者の発話開始を指示するためにノート PC が配置され，それによって発話の開始まで 3〜10 秒程度の時間がランダムに与えられた．この開始の合図は，「それをとってください」「はい」のリズム化を防ぐことを目的とし power point を用いて行われた．本実験の映像と音声はビデオカメラによって記録された．

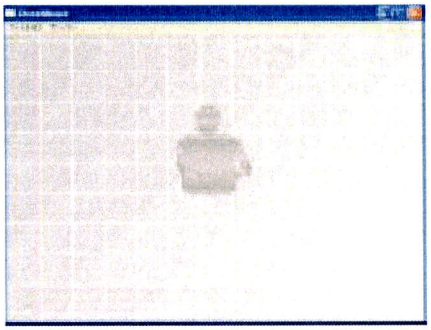

図 1. 本研究で使用したアバタ

2.3 実験課題

実験では，アバタと被験者の間にターゲットとして積み木を用意して，被験者が「それをとってください」と指示したのに対し，アバタが「はい」と応答する課題を用いた．図 2 に本課題における人間の指示発話長・交替潜時・アバタの応答発話長の関係を示した．アバタの応答発話長は「長(270ms)・中(200ms)・短(170ms)」の 3 種類が用意され，さらにそれぞれの発話長に対し 6 種類(300ms,600ms,900ms,1200ms,1500ms,1800ms)の交替潜時がランダムに提示された（表 1）．その際，被験者はできるだけ自然に言いやすい発話速度で「それをとってください」と指示を行い，その発話速度を意識的に変化させることは禁止されていた．1 種類の交替潜時につき，「それをとってください」「はい」を 10 セット行い，10 セット終わった直後に被験者はアバタの印象を評定し質問紙に回答を記入した．質問紙への回答が終わるごとに 2~3 分の休憩が与えられた．

表 1. 制御パラメータの種類と試行回数

〈アバタの応答発話長〉	〈交替潜時〉
3種類	×（6種類×各10セット）

質問紙には，アバタの印象について 10 種類の形容詞(話しやすい，友好的な，感じのよい，親切な，親しみやすい，肯定的な，落ち着いている，信頼できる，丁寧な)，アバタとの距離がどのくらいに感じるかを調べるための 1 種類の形容詞（近い），さらにアバタの発話タイミングが早いか遅いかについて調べるための 1 種類の形容詞（早い）の合計 12 種類の形容詞に関し，「全くあてはまらない:1」から「非常によくあてはまる：5」の 5 段階評価を行った．

3. 結果

3.1 アバタの発話タイミングに対する評定結果

アバタの発話タイミングが被験者にとって早いか遅いかを調査し，得られた被験者の平均評定値を図 3 に示す．この結果から，アバタの応答発話長の長さに関わらず交替潜時が 900ms であるとき，被験者はアバタの発話タイミングが適切であると評定していることがわかる．さらに，交替潜時が短いときほど被験者は「非常に早いと言える」と回答しており，交替潜時が長くなればなるほど「全く早いと言えない」と回答していることがわかる．

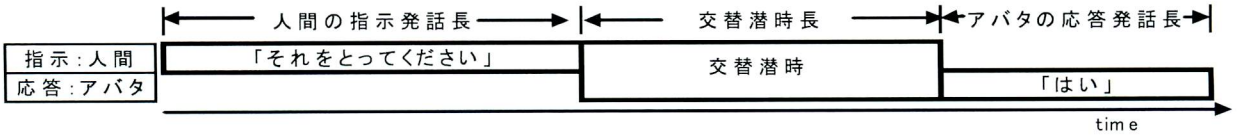

図 2. 人間の指示発話長，交替潜時，アバタの応答発話長の関係

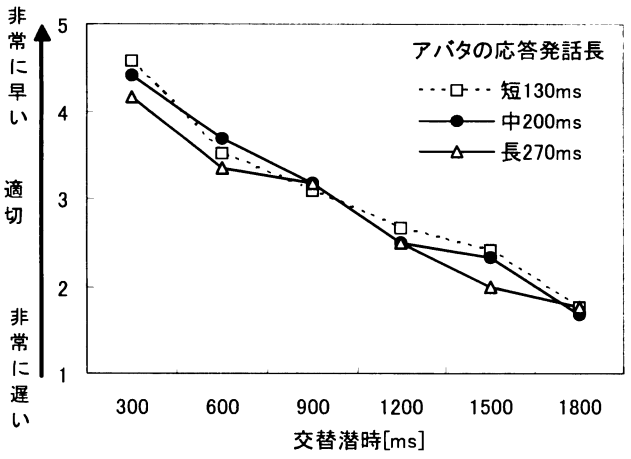

図3. アバタの発話タイミングに対する評価

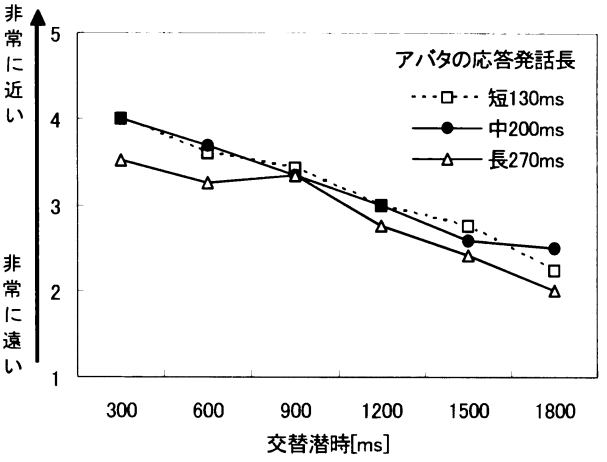

図5. アバタとの距離感の評価

3.2 アバタの発話長が印象評定値に与える影響 (交替潜時：900ms)

被験者がアバタの発話タイミングが適切であると評価した交替潜時900msにおいて，アバタの応答発話長（短130 ms・中200 ms・長270 ms）の違いが印象評定値に与える影響を調べた．交替潜時900msにおけるアバタの平均印象評定値を図4に示す．アバタの応答発話長が270msであるとき，他の条件（130 ms,200ms）に比べて評定値が有意に低い値を示していた（$p<0.05$）．

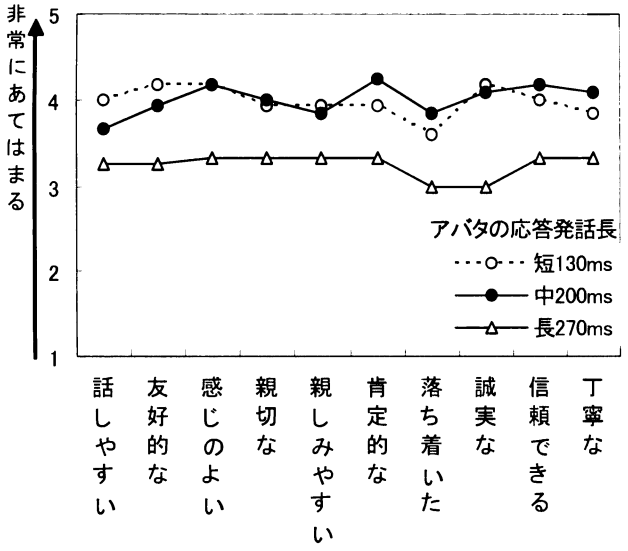

図4. 交替潜時900msにおけるアバタの平均印象評定値

3.3 被験者によるアバタとの距離感の評価

アバタの発話タイミングを制御することにより，被験者が抱くアバタとの主観的な距離感に変化が見られるかを調査し，得られた被験者の平均評定値を図5に示す．この結果から，アバタの発話タイミングが遅くなる（交替潜時が長くなる）に従って，被験者が抱くアバタとの主観的な距離感も遠いと感じられるように変化していることがわかる．

3.4 交替潜時が印象評定値に与える影響 (アバタの応答発話長：中200ms)

アバタの発話タイミング（交替潜時の長さ）を制御することにより，被験者がアバタに対して抱く印象評定に与える影響を調べた．アバタの応答発話長200msにおけるアバタの平均印象評定値を図6に示す．交替潜時が変化することにより，被験者がアバタに抱く印象も変化していることがわかる．また，交替潜時が300msのように短すぎると，落ち着いた印象や丁寧な印象は得られないことがわかる．そして，交替潜時が1800msのように長すぎると，全体的に悪い印象を与えている．さらに，最も評価値が高かったのは交替潜時が900msのときであることがわかる．

4. 考察

以上の結果から，表情や身振りなどの非言語チャンネルを持たないアバタでも，その発話タイミングによって印象が変化することが明らかになった．特に，交替潜時が900msのとき，被験者はアバタの発話タイミングが適切であると評定していた．さらに，その交替潜時900msにおいて，アバタの応答発話長の違いが印象評定値に与える影響を調べた結果，アバタの応答発話長が長いとき(270ms)，他の条件に比べて評定値が下がることが明らかになった．このことは，交替潜時が適切な長さであっても，応答の発話長の長さによっては，印象評定に大きな影響を与えてしまうことを示している．また，アバタは被験者の目の前の同じ位置にあるにもかかわらず，交替潜時が長くなると，被験者はアバタが遠くにいるような感じを抱くことが明らかになった．これは，応答までの時間が長くなることで，アバタが近くにいるという存在感が薄れてしまうことを示している．最後に，交替潜時の長さが変化することにより，被験者がアバタに対して抱く印象も変化していたことが明らかになった．最も評

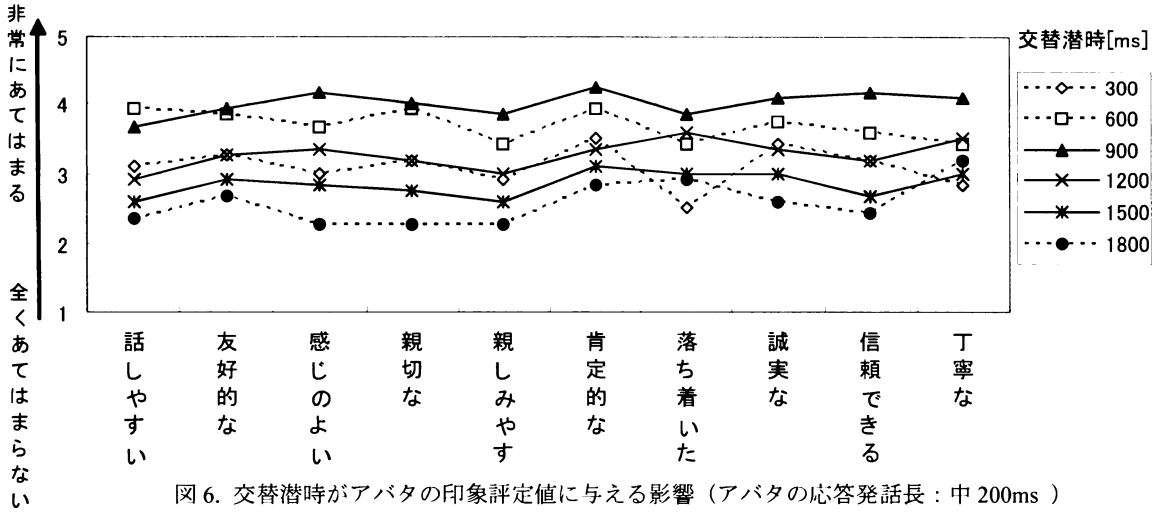

図 6. 交替潜時がアバタの印象評定値に与える影響（アバタの応答発話長：中 200ms ）

価値が高かったのは交替潜時が 900ms のときであり，この値は，被験者が適切な交替潜時であると評定した値と等しかった．

一方，通常の人間同士のやり取りでは，今回の結果よりも交替潜時は短い値を示すと予想される．それは，たとえば指差し動作や，視線の先，表情などで経験的に相手の意図が読めてしまうために，意識して音声情報を聞くことがないからではないかと考えられる．しかし，今回のように存在感はあるが音声情報しか提示されない場合，人間は音声情報に意識を向けることにより，交替潜時や発話長，また今回扱わなかったが音韻情報などから相手の情動と意図を理解しようしていることが伺える．このようなことから，交替潜時が人間同士の場合よりも長くなった可能性が推測される．

さらに，われわれがこれまで行ってきた周期的に外部から提示される音刺激に対しボタンを押す実験では，一番押しやすい刺激周期が 800ms であることが知られている．今後，さらに調べを進め，脳のタイミング機構との関連を明らかにしていく予定である。

5. まとめ

本研究では，表情や身振りなどの非言語チャンネルを持たないアバタを用い，人間と人工物のやりとりで多く観察される「指示-応答」場面を想定し実験を行った．その結果，発話タイミングや発話長を考慮することによりユーザーの人工物に対する印象が変化することを明らかにした．さらに，コミュニケーションチャンネルが制限された状態では，人間は異なる戦略を用いて相手の情動や意図を探ろうとしていることを示唆した．

参考文献

[1] 三宅，辰巳，杉原: 交互発話における発話長と発話間隔の時間的階層性，計測自動制御学会論文集，**Vol.40**,No.6, pp.670-678 (2004)

[2] 今，三宅: 協調タッピングにおける相互同調過程の解析とモデル化，ヒューマンインタフェース学会論文誌，**Vol.7**,No.4, pp.61-70 (2005)

[3] Takano, K., Miyake, Y.: Two types of phase correction mechanism involved in synchronized tapping, Neuroscience Letters,**Vol.417**, pp.196-200(2007)

[4] 武藤，三宅，ペッペル: 外乱を含む同期タッピング課題における認知が運動に与える影響,ヒューマンインタフェースシンポジウム 2006 講演会予稿集，pp.289-294 (2006)

[5] Condon, W.S., Sander L.W.: Neonate movement is synchronized with adult speech, Science,**Vol.83**, pp.99-101(1974)

[6] 渡辺，大久保: コミュニケーションにおける引き込み現象の生理的側面からの分析評価，情報処理学会誌,pp1225-1231(1998)

[7] 竹中，北岡，中川: 韻律・表層的言語情報を発話タイミング制御に用いた雑談対話システム，情報処理学会研究報告,pp87-92(2004)

[8] Matarazzo,J.D,et.al.: Interviewer head nodding and interviewer speech durations. Psychology, Theory, Research and Practice, **Vol.1**, pp.54-63

[9] 大坊，対人行動としてのコミュニケーション 対人行動学研究会(1986)

[10] 長岡，Draguna，小森，河瀬，中村:交替潜時の対話者間影響，ヒューマンインタフェースシンポジウム，pp.311-314(2000)

[11] 長岡，Draguna，小森，中村:音声対話における交替潜じが対人認知に及ぼす影響，ヒューマンインタフェースシンポジウム，pp.171-174(2002)

[12] Welkowitz, J., & Kuc, M.: Interrelationships among warmth, genuiness, empathy and temporal speech patterns in interpersonal interaction. Journal of Social and Clinical Psychology, **Vol.17**, pp.101-102(1973)

[13] Rutter,D.R., & Stephenson, G.M.: The role of visual communication in synchronizing conversation. European Journal of Social Psychology, **Vol.7**. pp.29-37(1977)