

音声対話における発話タイミングの 影響に関する知覚実験

伊藤 敏彦† 北岡 教英‡ 西村 良太††

†北海道大学大学院 情報科学研究科
〒060-0814 札幌市北区北14条西9丁目

‡名古屋大学大学院 情報科学研究科
〒464-8603 名古屋市千種区不老町

††豊橋技術科学大学 情報工学系
〒441-8580 豊橋市天伯町雲雀ヶ丘1の1

E-mail: t-itoh@media.eng.hokudai.ac.jp, kitaoka@nagoya-u.jp,
nishimura@slp.ics.tut.ac.jp

あらまし先行研究で行ってきた対話リズムに関する分析結果の妥当性を検証するために、発話タイミングに関する分析結果に基づいた音声対話を作成し、対話の自然性や合成音声の違和感、発話の聞き取りやすさなどを調査する知覚実験を行った。知覚実験1では合成音声による4ターンの短いタスク指向対話を、知覚実験2では実音声と合成音声の1分程度の長めの雑談対話を用いて実験を行った。その結果、発話における自然な発話タイミングが存在すること、発話内容に即した発話タイミングに自然性を感じるなど、対話リズムに関する分析結果の妥当性を示すことができた。

キーワード 知覚実験, 音声対話, 対話リズム, 発話タイミング, 自然性

Subjective Experiments on Influence of Response Timing in Speech Dialogues

Toshihiko ITOH† Norihide KITAOKA‡ Ryota NISHIMURA††

†Graduate School of Information Science and Technology, Hokkaido University,
N14, W9, Kita-ku, Sapporo, 060-0814, JAPAN

‡Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-0843, JAPAN

††Department of Information and Computer Sciences,
Toyohashi University of Technology,

1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, JAPAN

Abstract In order to examine the validity of the findings on the dialogue rhythm which we previously pointed out, we made some dialogue samples with various rhythm/utterance timing and evaluated them subjectively from the view points of naturalness of whole the dialogues, unnaturalness of synthesized speech, and intelligibility of the dialogues. We used short task-oriented four-turns dialogues using speech synthesizer in Experiment No. 1, and approx. one-minute chat-like dialogues in No. 2 using natural human utterances and synthesized voices. The results of these experiments supported our previous analysis that the utterance timing is important for natural dialogue and the timing of each utterance mainly depends on the contents of the utterance.

keyword Subjective Experiment, Speech Dialogue, Dialogue Rhythm, Response Timing, Naturalness

1. はじめに

近年、音声対話システムの性能は飛躍的に向上してきており、実用化を目指し、多くの研究がなされている [1]。しかしながら、現在のところ人対機械のコミュニケーションが人間同士のものと同等であるとは言えない。従来、人対機械におけるコミュニケーションにおいては、音声認識や言語理解の精度、合成音声の質などが重要であると考えられてきたが、我々の先行研究 [2] により、発話タイミング、ピッチ、発話速度などの対話リズムや感情などのシステ

ム発話の韻律情報が十分でないとは機械と人間らしい自然な対話が行えないことを示した。

我々はこの対話リズムに着目し、対話リズムを考慮したタスク指向型や雑談型の音声対話システムの開発を行っている [3, 4]。これらのシステムは、人間同士の協調的な対話においては発話者の音響的特徴（発話タイミング・F0・発話速度）が対話が進むにつれ同調する傾向がある [5] ことを反映するように、ユーザ発話の音響的特徴にシステム発話の音響的特徴を同調させる機能も有している。

我々のシステムは、対話リズムや同調傾向を考慮しない場合に比べて、対話の自然性に関しては若干の改善が見られたものの人間同士の対話の滑らかさや自然さには達していない。この原因として、システムの発話タイミングが対話相手（ユーザ）の音響的特徴やキーワードのみに依存し、システム側の状態を一切考慮していないことや、同調機能に関しても常に相手の発話の音響的特徴に無条件にシステム発話の音響的特徴を同調させているためであると考えている。発話タイミング、F0、発話速度などの対話リズムや音響的な同調の度合いなどは、発話者の状態（対話の盛り上がり、対話意図、感情など）により大きく変化すると考えられる。また、我々は発話者の状態の中でも、特に対話意図が対話リズムに与える影響が強いと考えているが、対話意図と対話リズムの関係を分析した研究はほとんど存在せず、またタスク指向対話における同調傾向についても研究されていない。

そのため、我々はタスク指向型の対話を収集し、対話意図と対話リズム（発話タイミング・F0・発話速度）の関係について分析を行った[6]。さらに、発話タイミングの分布の傾向、個人差による影響、詳細なレベルの対話意図（3.1.1節の対話例参照）と対話リズムの関係、先行発話の影響についても調査を行った[7]。これらの分析結果から、対話リズムに関するいくつかの重要な知見を得た。特に、発話タイミングに関する知見は、音声対話システムとの人間らしい自然な音声対話を実現する上で重要であると考えている。

そこで、本稿ではこれらの知見の妥当性を検証するために、発話タイミングに関する知見に基づいた音声対話を作成し、その自然性などを調査する知覚実験を行った。本稿の最初に先行研究で得られた分析結果に関して簡単に述べ、その後今回行った知覚実験について述べる。

2. 自然でスムーズな人-機械間の対話実現を目指して

我々が自然でスムーズな人-機械間の対話を実現するために、対話意図と対話リズムに関して行った分析[6, 7]について簡単に述べる。

まず、分析対象とした音声対話コーパスは、我々が独自に収録したホテルの検索・予約に関する客役と店員役の二者による非対面のタスク指向音声対話（人数17名、対話数50対話、平均対話時間4分57秒）である。このコーパスに対して、発話位置情報（総発話数 客役:3,510, 店員役:3,844）や、対話意図タグ[8]（総対話意図数 客役:2,520, 店員役:

2,215）などのタグを付与し、そのタグ情報を利用して各対話意図における発話タイミング、F0、発話速度などを分析した。これらの分析結果から、対話意図による対話リズム（発話タイミング・F0・発話速度）の違いの分析、対話意図や間投詞の発話タイミングの分布傾向、対話リズムにおける個人差の影響、詳細なレベルの対話意図と対話リズムの関係、対話相手の先行発話による発話者の対話リズムへの影響などについて調査した。これらの調査結果から、我々は2話者によるタスク指向対話において対話リズムを以下のように考えている。

まず、対話意図と対話リズムの関係には、個人による影響（個人差）は存在するが、各対話意図間の関係を大きく崩すものではなく、一般性がある。

発話タイミングに関しては、対話意図毎に発話タイミングの分布は異なる。対話意図よりさらに詳細な単位である発話内容で分析してみると、より分散は小さくなり、発話タイミングがある程度固定化する。しかし、発話内容の場合でも発話タイミングの範囲が存在し、その範囲の大きさは発話内容ごとにかなり異なるが、範囲の開始位置（時間）は全ての発話内容においてほぼ共通である。この範囲の大きさは、発話内容（exchange構造内の種類を含む）や対話構造の状態（話者交替/話者継続）、話題の切り替わりの有無（新しい話題を始める発話であるかどうか）により決定される。基本的に働き掛けの場合より応答の場合の方が、発話タイミングの範囲は狭まる。また、話題が切り替わる位置では、タスクに関する整理や対話の流れのプランニングをするために発話タイミングの範囲が広がり、発話タイミングが変動しやすい。そして、この発話タイミングの範囲終了までが発話者が発話内容決定にかけられる主な時間となる。発話者は、発話タイミングの範囲内に発話内容を決定できない場合や、決定できないと予想した場合に間投詞を発話する。実際の最終的な発話タイミングは、「発話タイミングの範囲」と「発話内容、発話内容の重要度、発話内容の決定度（発話内容がどの程度決まっているか）」の関係により決定されると考えられる。そのため、発話タイミングは、「対話相手の先行発話の予想しやすさ」、「対話相手の発話内容の理解しやすさ」、「発話内容」、「発話内容の重要度」、「対話相手が予想（期待）している発話内容と決定した発話内容との差」、「対話相手の発話持続時間」「情報検索の時間」などが強く影響すると思われる。

また発話全体のF0と発話速度の決定には、「発話内容」、「発話内容の重要度」、「対話相手が予想（期待）している発話内容と決定した発話内容との差」

などが強く影響する。

さらに聞き手も各対話意図（発話内容）の平均的な発話タイミング、F0、発話速度、さらに可能ならば対話相手別の発話タイミング、F0、発話速度をモデル化しており、そのモデル化された対話リズムと実際の発話の対話リズムの差から、聞き手は様々なパラ言語情報を含む非言語情報を取得していると考えられる。そのため、スムーズで自然なタスク指向型の音声対話システム実現には、平均的な対話意図と発話タイミング、F0、発話速度との関係について人間同士の対話を用いてモデル化し、音声対話システムがそのモデルを用いてユーザの状態とシステムの状態を考慮した対話リズムを実現することと、さらに実際のユーザ発話とモデルの違いから非言語情報の取得ができる必要がある。

3. 対話リズムに関する知覚実験

前節では、対話リズムに関する分析で得られた知見について述べ、人間らしい自然な音声対話を行う音声対話システムを作るために必要な要素に関する妥当性を検証するために、被験者を用いて知覚実験を行った。

今回の知覚実験で検証を行う分析結果からの仮説は以下の2点である。

- 聞き手は、対話意図や発話内容における一般的な発話タイミングをモデル化している。
- 聞き手は、発話内容に即した人間の発話タイミングを最も自然に感じる。

上記の検証のために、知覚実験として2種類の実験を行った。知覚実験1では合成音声による4ターンの短いタスク指向対話における発話タイミングの違いによる影響を調査し、知覚実験2では1分程度の長めの雑談対話における発話タイミングの違いによる影響について調査した。また、知覚実験2においては、実音声又は合成音声により構成された音声対話試料を用いることにより、声の質、感情の欠落、抑揚の欠落などの要素と発話タイミングの関係についても調査した。

3.1 知覚実験1

本実験では、4ターンの非常に短いタスク指向対話を合成音声で作成し、対話中の各話者の発話タイミングを様々に変化させ、対話全体の自然性、合成音声の声質への違和感、各話者の印象などについて調査した。

3.1.1 音声試料

本実験で作成した音声試料の対話内容と対話意図（発話内容の下段左）、詳細レベルの対話意図（発話内容の下段右）を以下に示す。

交代潜時	話者	発話内容
α	A:	お名前の方よろしいですか？ [未知情報要求] [名前要求]
	B:	鈴木太郎です。 [未知情報応答] [名前応答]
β	A:	鈴木太郎様？ [確認] [名前確認]
	B:	はい。 [肯定・受諾] [確認肯定]

対話例中の α 、 β 、 γ は前発話の終了から当該の発話を発話するまでの時間（交代潜時）を示している。この発話タイミングを様々に変化させ、実験用音声試料とした。実験に用いた発話タイミングパターンは以下のような7種類である。括弧内の値は実験に用いた発話タイミングの実際の時間（秒）である。

- 1-1 各対話意図の発話タイミング平均値 ($\alpha:0.65, \beta:0.75, \gamma:0.11$)
- 1-2 各発話内容の発話タイミング平均値 ($\alpha:0.21, \beta:0.47, \gamma:0.09$)
- 1-3 各発話内容の発話タイミング平均値のランダム適用 ($\alpha:0.47, \beta:0.09, \gamma:0.21$)
- 1-4 各発話内容の発話タイミング平均値のランダム適用2 ($\alpha:0.09, \beta:0.21, \gamma:0.47$)
- 1-5 発話タイミングを1秒に固定 ($\alpha:1.00, \beta:1.00, \gamma:1.00$)
- 1-6 発話タイミングを0.5秒に固定 ($\alpha:0.50, \beta:0.50, \gamma:0.50$)
- 1-7 発話タイミングを-0.3秒に固定 ($\alpha:-0.30, \beta:-0.30, \gamma:-0.30$)

表 1: 発話タイミング比較実験結果

	第1グループ				第2グループ	
	第1セット		第2セット		第3セット	
	自然	声質	自然	声質	自然	声質
1-1	3.5	3.5	4.6	4.5	4.1	4.0
1-2	4.7	4.4	4.8	4.4	5.4	4.5
1-3	3.0	3.8	4.4	4.7	3.5	3.7
1-4	3.8	4.0	3.9	4.5	4.2	4.2
1-5	4.5	2.8	3.3	4.7	3.7	3.8
1-6	5.3	4.2	5.5	4.7	5.9	4.5
1-7	1.6	3.6	1.6	3.7	1.2	3.3

音声試料中の発話音声は音声合成システム「Fine-Speech SDK V2.1」を用いて作成した。話者の違いを明確にするために話者 A は女性、話者 B は男性の声として音声合成を行った。声の高さ、感情、抑揚、発話速度などはすべてデフォルトの値を使用し、特別な後処理なども行っていない。その後、自作のプログラムを用いて各発話音声を発話タイミングまでの無音データを挟み込みながら接続することで各音声試料を作成した。なお、無音データは完全な無音であり、環境雑音などは一切考慮していない。

3.1.2 実験手順

本実験は2つの被験者グループに対して行った。被験者第1グループはN大学情報系大学院生 男女混合39名、被験者第2グループはH大学情報系学部学生 男女混合58名である。どちらのグループも音声認識、音声言語処理、音声合成などに関する専門的な知識はない。実験を行った場所は、どちらも70名から80名程度が入れる通常の教室であり、雑音や反響に関する特別な処置は一切行っていない。教室の中央前方にスピーカーを設置し、教室全体に音声が届くように方向、音量調整などを行った。

その後、2つの被験者グループに対して以下のような流れで知覚実験を行った。

1. 被験者に実験内容、音声試料の対話内容、印象リスト、アンケートが書かれた紙を配布・熟読させ、さらに口頭でも簡単な説明を行った。
2. 7パターンの音声試料を評価させずに一通り再生。
3. 音声試料をランダムに選び各3回ずつ再生し、自然性、合成音声への違和感を7段階で評価してもらった。さらに可能ならば各話者に対する印象を印象リストから選択するか、自由記入で回答してもらった。

なお、被験者第1グループには評価してもらう対話の自然性について特別な説明は一切行わず実験を行い、被験者第2グループには発話内容と発話タイミングに着目して対話の自然性を判断するように指示して実験を行った。さらに被験者第1グループに対してのみ、評価の揺らぎを調べるために同じ7パターンの音声試料を再度ランダムに再生し、もう一度別に評価してもらった。ただし、この時に被験者には直前に評価してもらった音声試料と同じ音声試料であることは伝えていない。

3.1.3 実験結果

知覚実験1の結果を表1に示す。表中の「自然」は対話の自然性に関する7段階評価（7＝自然）の平

均、「声質」は合成音声への違和感に関する7段階評価（7＝違和感なし）である。

非常に興味深いことに、どの評価セット、どの評価項目においても、評価の第1位、第2位、第7位の高得点、低得点を獲得した音声試料はそれぞれ同一の発話タイミングパターンのものであり、発話タイミングなどの対話リズムが対話の自然性へ強い影響を与えていることや、適切な（自然な）発話タイミングが存在することを示している。また、発話タイミングの自然さと合成音声への違和感には強い相関があり、自然な発話タイミングであればあるほど、合成音声への違和感は薄れていく。しかしながら、第1位の評価結果である音声試料1-6は発話タイミングを0.5秒に固定したものであり、我々がこれまでの対話分析結果から提案している発話内容に基づいた発話タイミングを使用した音声試料1-2は高得点ながら第2位の評価結果となった。最も評価の低い発話タイミングパターンは、常にオーバーラップしている音声試料であり、人間の発話タイミングの範囲内だとしてもオーバーラップの使いどころには注意が必要であることが分かる。その他の発話タイミングパターンは、評価セットによって順位が変動しているため、全体の平均点で調べてみると、音声試料1-1 > 1-4 > 1-5 > 1-3の順となり、対話意図での発話タイミングもランダムな発話タイミングよりは自然であるとの評価を得た。

だが、上記の結果をそのまま受け取ると、対話における発話タイミングは、人間が発話する発話タイミングの範囲の中間的な値である0.5秒程度のタイミングに常に一定であることが最も自然で好ましいという結論になる。そこでもう少し別の観点からの分析を行った。まず音声試料1-6と1-2においてどちらに高得点を与えたかを調査したところ、第1セットの被験者39人のうち音声試料1-6側に15人、音声試料1-2に7人、その他は同点であった。第2セットにおいては、39人中、音声試料1-6側に17人、音声試料1-2側に7人であった。しかしながら、対話内容と発話タイミング着目して評価してもらった第3セットにおいては、58人中、音声試料1-6側に26名、音声試料1-2側に18名であり、その差は縮まっていた。また、各音声試料の話者についての印象や全体的なコメントなどを調べてみると、音声試料1-6に対しては「落ち着いている」、「理性的」、「冷静」、「のんびり」などの印象が多く、音声試料1-2に対しては、「人間的」、「誠実」、「せっかち」、「緊張」などの意見が多かった。また、「合成音声は淡々とゆっくりしゃべるので、間があいてるほうが自然に感じる」との全体コメントもあった。確かに、本

実験で用いた音声試料に関しては感情や抑揚、発話速度などはすべてデフォルト値であり、実際にその内容を人間が発話をするときの感情や抑揚、発話速度とは異なっている。そのため、人間的な発話タイミングと機械的な感情、抑揚、発話速度の合成音声を組み合わせたために、今回のような結果が出た可能性もある。そこで、知覚実験2においては、これらの点を考慮して知覚実験を行った。

3.2 知覚実験2

本実験では、知覚実験1において対話の自然性に関して高得点を得た「発話内容を考慮した発話タイミング」と「0.5秒固定の発話タイミング」にのみ着目し、実音声における対話の自然性での発話タイミングの重要性、および対話音声の感情情報、抑揚、ピッチ、発話速度などの発話タイミング以外の対話リズム要素と発話タイミングの関係について調査した。

3.2.1 音声試料

本実験で用いた音声試料は、国立国語研究所から提供されている「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese; CSJ) [9] から用いた。対象とした音声対話は、CSJ コーパスの2者雑談対話である「D02F0025」の中から比較的盛り上がっている部分 (79.214sec~143.860sec; 合計 64.646sec) を抜き出して実験に使用した。切り出した対話はL話者22発話、R話者18発話の合計40発話である。この音声対話データを基に以下の4種類の音声試料を作成した。

- 2-1 対話中の発話は実音声であり、実音声の発話タイミングをそのまま利用 (原音声対話)
- 2-2 対話中の発話は実音声であり、すべての発話タイミングを0.5秒に固定
- 3-1 対話中の発話は合成音声であり、実音声の発話タイミングをそのまま利用
- 3-2 対話中の発話は合成音声であり、すべての発話タイミングを0.5秒に固定

音声合成システムには、galatea ツールキットに含まれる galatea talk [10] を用い、コーパスに付属の書き起こしデータを用いて発話音声を作成した。なお、元音声は女性同士の対話であるが、合成音声では話者の聞き分けが難しいため、女性対男性 (R話者を男声に変更) の対話として発話音声を作成している。

各発話音声の発話タイミングの変更や無音の挿入には音声波形ソフト GoldWave¹ を用いて音声波形編集することで行った。作成方法は、実音声の発話

¹<http://www.goldwave.com/>

タイミング変更に関しては波形を表示し、ポーズが長い場合には余分な部分を切り、ポーズが短い場合には、無音を挿入することで、発話タイミングをすべて0.5秒に変更した。無音を挿入する場合、環境雑音などは一切考慮していない。話者交替時にオーバーラップが発生していた場合には、重ならないようにオーバーラップ時間長に0.5秒プラスした分だけ時間をずらした。相槌やフィラーなどの発話内容 (情報) を有していない発話に関しては、オーバーラップを含めすべてそのままのタイミングを利用した。合成音声に関して、実音声のタイミングを利用する場合は、各発話のポーズ長を元音声と同じになるように無音を挿入したり、オーバーラップしている場合には発話をオーバーラップするように重ね合わせた。一定の応答タイミングにする場合には、相槌・フィラーなどの発話内容を有していない発話を除くすべての発話に0.5秒の無音を挿入し各発話を接続した。

3.2.2 実験手順

本実験の実験環境は基本的に知覚実験1と同一である。被験者は知覚実験1における被験者第2グループのH大学情報系学部学生 男女混合58名である。

この被験者グループに対して以下のような流れで知覚実験を行った。

1. 被験者に実験内容、印象リスト、アンケートが書かれた紙を配布・熟読させ、さらに口頭でも簡単な説明を行った。
2. 音声試料2-1、2-2の順に2回再生し、自然性、対話の違和感、感情や対話意図の理解しやすさを7段階で評価してもらった。さらに可能ならば各話者に対する印象を印象リストから選択するか、自由記入で回答してもらった。
3. 次に音声試料3-1、3-2の順に2回再生し、自然性、対話の違和感、感情や対話意図の理解しやすさ、合成音声への違和感、発話の聞きやすさを7段階で評価してもらった。さらに可能ならば各話者に対する印象を印象リストから選択するか、自由記入で回答してもらった。

3.2.3 実験結果

知覚実験2の結果を表2、表3に示す。表中の「自然」は対話の自然性 (7=自然)、「違和感」は対話音声への違和感 (7=違和感なし)、「意図」は発話の意図や感情の理解しやすさ (7=理解しやすい)、「声質」は合成音声への違和感 (7=違和感なし)、「聞き取り」は発話の聞き取りやすさ (7=聞き取りやすい) に関する7段階評価である。

表 2: 実音声の比較実験結果

	自然	違和感	意図
2-1	6.0	6.1	5.8
2-2	5.6	5.6	5.3

表 3: 合成音声の比較実験結果

	自然	違和感	意図	声質	聞き取り
3-1	2.4	1.5	1.5	1.7	1.8
3-2	3.4	1.8	2.0	2.1	2.6

実音声での比較実験から発話音声に感情、抑揚、発話速度などの非言語情報が発話内容に即した状態である場合、発話タイミングも0.5秒固定よりも発話内容に基づいた発話タイミングの方がすべての評価に関して良かった。逆に同一の発話内容であるが合成音声の発話音声の場合、発話タイミングが0.5秒固定の対話音声のほうがすべての評価に関して良かった。このことから知覚実験1において0.5秒固定の発話タイミングが最も評価が高かった理由は、合成音声の非言語情報の感情、抑揚、ピッチ、発話速度などが一定の音声と最もバランスが取れた結果であると考えられる。つまり、自然な発話タイミングだけで他の非言語情報を超えて、対話全体の自然性を支配できるほどの重要度や優先度を発話タイミングは持っていない。そのため、発話タイミングの決定には、発話内容だけでもある程度は自然なタイミングを決定できるが、本当に人間らしい自然な対話を実現するには感情などを考慮した発話タイミングの決定と、その感情などに即したピッチ、抑揚や発話速度を持った音声も同時に必要であり、このバランスが重要となる。逆に、どんなに自然な発話タイミングでも音声の非言語情報が十分でない、そこに違和感や不自然さを感じるようになる。今後、音声対話システムなどで発話タイミングを考慮する場合には、音声合成の性能なども考慮した発話タイミング決定が必要になると思われる。また、発話タイミングのみで感情や対話意図を完全に推定することは難しくそうであるが、素性の一つとして利用するには十分な情報量をもっていていると考えられる。

最後にオーバーラップに関してだが、知覚実験2においてもオーバーラップのある音声対話を用いている。しかし、実音声の音声試料では全く違和感を感じないが、合成音声の音声試料ではオーバーラップしている箇所に関する自然性、違和感、聞きやすさなどの評価は特に悪い。人間らしい自然なオーバーラップを実現するためには、人間らしい発話内容と発話タイミングだけでは不十分であり、その他の非言語情報との組み合わせやバランスが重要であると思われる。今後、それらに関する要素をもう少し詳しく分析をする必要がある。

4. まとめ

本稿では、これまでに先行研究で行ってきた対話リズムに関する分析結果の妥当性を検証するために、発話タイミングに関する分析結果に基づいた音声対話を作成し、対話の自然性や合成音声の違和感、発話の聞き取りやすさなどを調査する知覚実験を行った。その結果、発話における自然な発話タイミングが存在することや、発話内容に即した発話タイミングに自然性を感じるなど、対話リズムに関する分析結果の妥当性を示すことができた。しかしながら、対話の自然性は発話タイミングだけが自然でも不十分で、発話の音響的特徴と発話タイミングのバランスが重要であることも分かった。

今後は、ボイスチェンジャーなどのツールを使い、声の質以外の音響的特徴を残したままの音声や、一部の音響的特徴を除いた音声を用いて同様な知覚実験を行い、自然性を感じるために必要な要素について分析していくことを考えている。また、合成音声を用いても自然であると感じるオーバーラップを実現するために必要な音響的特徴についても分析していきたいと考えている。

参考文献

- [1] 藤原 敬記, 伊藤 敏彦, 荒木 健治, 甲斐 充彦, 小西 達裕, 伊東 幸宏, “認識信頼度と対話履歴を用いた音声言語理解手法”, 電子情報通信学会論文誌, Vol.J89-D, No.7, pp.1493-1503.
- [2] 伊藤 敏彦, 山田 真也, 荒木 健治, “音声認識・言語理解性能や状況の違いによるタスク指向音声対話の言語的・音響的特徴の比較”, 日本音響学会誌, Vol.63, No.5, 2007.
- [3] 東海林 圭輔, 高橋 美佳, 井原 誠也, 伊藤 敏彦, 荒木 健治, “対話に関するリズムや同調作用を考慮した音声対話システム”, 情報処理学会研究報告, 2006-SLP-61, pp.43-48(2006).
- [4] Kitaoka, N., Takeuchi, M., Nishimura, R., and Nakagawa, S., “Response Timing Detection Using Prosodic and Linguistic Information for Human-friendly Spoken Dialog Systems”, Journal of JSAI, Vol.20, No.3 SP-E, pp. 220-228(2005).
- [5] 長岡 千賀, 小森 政嗣, 中村 敏枝, “対話における交替潜時の対話者間影響”, 日本人間工学会誌, Vol.38, No.6, pp.316-323(2002).
- [6] 藤原 敬記, 伊藤 敏彦, 荒木 健治, “タスク指向対話における発話意図の対話リズムへの影響”, 情報処理学会研究報告, 2007-SLP-66, pp.37-42 (2007).
- [7] 藤原 敬記, 伊藤 敏彦, 荒木 健治, “タスク指向対話における相互の発話意図を考慮した対話リズムの分析”, 人工知能学会研究会資料, SIG-SLUD-A701-08, pp.45-50 (2007).
- [8] 荒木 雅弘, 伊藤 敏彦, 熊谷 智子, 石崎 雅人, “発話単位タグ標準化案の作成”, 人工知能学会論文誌, Vol.14, No.2, pp.251-260(1999).
- [9] Maekawa, K., “Corpus of Spontaneous Japanese : Its Design and Evaluation”, Proc. ISCA & IEEE Workshop SSPR 2003, pp.7-12, (2003)
- [10] 嵯峨山, 川本, 下平, 新田, 西本, 中村, 伊藤, 森島, 四倉, 甲斐, 李, 山下, 小林, 徳田, 広瀬, 峯松, 山田, 依, 宇津呂, “擬人化音声対話エージェントツールキット Galatea”, 情報処理学会研究報告, SLP-45-10, pp.57-64, (2003)