

対話における相槌の時間構造の解析

杉原 史郎*¹ 三宅 美博*¹

Analysis of Time Structure of Aizuchi in Spoken Dialogue

Shiro Sugihara*¹ and Yoshihiro Miyake*¹

Abstract - In this study, we aimed to analyze the Aizuchi act in conversation to construct the system which can simulate the Aizuchi between the system and user. By using machine which sends the message, we analyzed quantitatively the Aizuchi act of listener who receives the message. To be concrete, we analyzed listener's Aizuchi duration and pause duration in experimental spoken dialogue, in which the machine speaks a sentence and the listener chimes in. As a result, in regular spoken dialogue that doesn't change the speech motion, machine's pause duration and speech duration influenced listener's pause duration. In the transition process of spoken dialogue, the change of machine's pause duration caused the slow change of listener's pause duration, and the change of machine's speech duration caused the sudden change of listener's pause duration. These results show that Aizuchi act has hierarchy in time structure, and suggest that the hierarchy is associated with duality in communication.

Keywords : Aizuchi Act, Spoken Dialogue, Time Structure, Hierarchy, Duality

1. はじめに

近年、音声を通じて人間と機械がコミュニケーションを行う機会が増加している。それに伴い、人間-機械系の音声コミュニケーションの場において、人間が何ら応答を示さない機械に対して一方的に働きかける機会も増大する傾向にある。具体例としては、音声によるテキスト入力ソフト、ボイスメール、留守番電話などがあげられる。しかし、このような人間から機械への一方的な働きかけは人間に違和感などを抱かせ、我々がその人工物と共生する上での障害となっている。そこで、本研究では対話中に見られる相槌行為を用い、メッセージの円滑な入力を支援するシステムの構築を目指す。

人間同士の対話を考える際にコミュニケーションの持つ二重性という性質が重要になると考えている[1-3]。二重性とは、コミュニケーションにおける身体的なほたらきと、意識的なほたらきの相補的な関係を指している。そして、人間のコミュニケーションは意識の及ばない身体相互作用と、そこに現れる意識とが相互に拘束し合う構造を持つと考えられる[4-5]。しかし、従来の音声対話研究において、そのほとんどが情報を伝えるという意識的な側面ばかりに注目していたため、情報が伝わってしまうような無意識的な側面、あるいは、これら両側面について、あまり考慮されてこなかったように思われる。

そこで本報告では意識的なほたらきとしての発話と、呼吸機構のような身体的なほたらきと関係する間(ま)に注目し[6-8]、相槌行為の構造をコミュニケーションの二重性を切

り口とし分析する。また、相槌行為をメッセージの送り手と受け手の関係が固定される対話の一形式と捉え、実験中に送り手と受け手の関係が交替しない実験的な対話を設定する。すなわち、話者は文を発話し、メッセージを送り続け、聴者はそのメッセージを受け取り、相槌を打ち続ける。さらに、話者の発話動作を機械的に構成し、話者の発話の時間長と間(ま)の時間長を制御することで、聴者の相槌の時間長と間(ま)の時間長への影響を調べる。

2. 実験設定

2.1 実験の対話内容

対話の内容は話者である機械が1文のメッセージを発話し、それに対して聴者である被験者が相槌「うん」を入れ、6秒間繰り返す。文は単語(3モーラ) + 「ですね」の計6モーラで構成され(例えば「あぐらですね」と発話される)、単語の内容は「あぐら」「やなぎ」「みなみ」「ゆびわ」「ゆのみ」「でんわ」の6種類を採用した(提示順序はランダムに行われた)。これは文と文の間の意味的なつながりを極力排除し、コンテキストに左右されない相槌行為の基本構造の解析を狙ったものである。また、単語直後に付加された「ですね」は文の語尾のアクセントパターンを一定にし、終助詞「ね」の機能によって被験者の相槌を促す働きをもつ。

2.2 対話特徴量の定義

本研究で計測する聴者側の特徴量、及び、機械によって制御する話者側の特徴量について説明する(Fig. 1)。

<発話長>

話者である機械が文を発話し、その発話が終了するまでの時間長を「発話長(SP)」と定義する。すなわち、機械による

*1: 東京工業大学大学院 総合理工学研究科

*1: Tokyo Institute of Technology, Interdisciplinary Graduate School of Science and Engineering

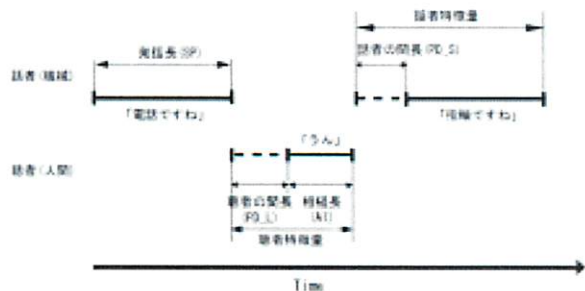


図1 特徴量の定義

Fig.1 Definition of parameter

音声提示時間を「発話長(SP)」とする。

<相槌長>

聴者である人間が相槌「うん」を発話し、その発話が終了するまでの時間長を「相槌長(AI)」と定義する。すなわち、人間の発話時間を「相槌長(AI)」とする。

<間長>

一方の発話が終了してから他方の発話が開始されるまでの時間長を「間長」と定義する。具体的には、話者である機械が発話を終了してから、聴者である人間が発話を開始するまでの時間長を「聴者の間長(PD_L)」と呼び、聴者である人間が発話を終了してから、話者である機械が発話を開始するまでの時間長を「話者の間長(PD_S)」と呼ぶ。

<話者特徴量>

話者である機械によって決定される特徴量を「話者特徴量」と定義する。具体的には「発話長(SP)」と「話者の間長(PD_S)」を指す。

<聴者特徴量>

聴者である人間によって決定される特徴量を「聴者特徴量」と定義する。具体的には「相槌長(AI)」と「聴者の間長(PD_L)」を指す。

2.3 実験システム

本研究で使用した計測システムの概要図をFig. 2に示す。ノートパソコン(IBM ThinkPad570, MMX Pentium2, 333MHz, メモリ:128MB, OS:Windows98)にはミニミキサー (FOSTEX, MN06), A/Dコンバータ(ADTEK, AXPAD02, PCMCIA 準拠)が接続されている。マイクロフォン付きヘッドフォン(audio-technica, ATC-H7COM)は機械の音声を提示し、人間の音声を

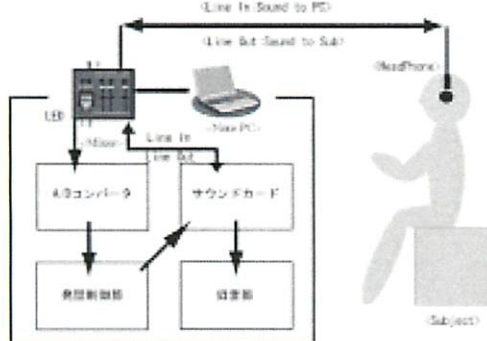


図2 システム構成

Fig.2 Structure of System

センシングする。ミニミキサーではステレオ音声による入力となされ、人間の音声はR側、機械の音声はL側のラインに分離される。また、ミニミキサーに搭載されているLEDは音声入力が行われると点灯し、ミキサー上部に取り付けられた光センサによって検出される。検出された信号はON-OFFの2値化信号に変換され、ノートパソコンに格納される。またノートパソコンでは、独自に開発したアプリケーション(Inspire社製 BorlandC++ Builder ver. 3.0で開発)が人間および機械音声の録音、機械の発話タイミングの制御、機械音声の再生を行う。

2.4 データ解析方法

実験終了後、独自に開発した解析ソフト(BorlandC++ Builder ver. 3.0で開発)を用い、ノートパソコンに格納されたデータから発話長、相槌長、間長を算出した。

3. 実験

3.1 話者が定常的に発話する場合

本実験では対話中に発話の長さ、間の長さに変化を生じない定常的な対話において、聴者の相槌行為の特徴を明らかにすることを目的とする。具体的には話者特徴量を固定し、聴者特徴量の特徴を調べる。話者特徴量の設定条件は(SP, PD_S)= (900, 100), (900, 150), (900, 200), (900, 250), (900, 300), (850, 200), (950, 200)の条件が用いられた。被験者は20代の男子学生5人で、普段の対話と同じように相槌を入れるよう指示された。

Fig. 3は発話長(SP)を900ms、話者の間長(PD_S)を200msに固定した時の聴者特徴量(AI, PD_L)の一例である。相槌長(AI)は60秒間の測定時間で定常となり、聴者がほぼ固定長の相槌を入れていることがわかる。一方、聴者の間長(PD_L)は実験開始直後、100ms程度のゆらぎを持ちながら遷移するが、次第に話者特徴量に適応し、定常な状態になることが明らかになった。また、これらの結果は他の設定条件、被験者においても確認された。

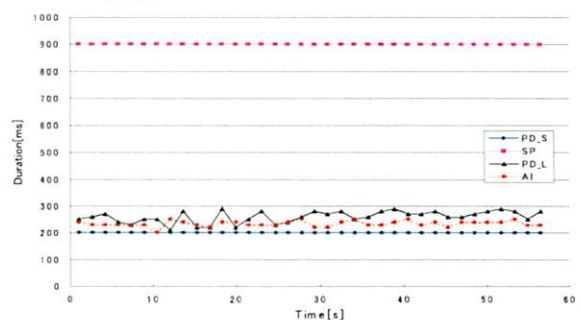


図3 聴者特徴量の時間発展

Fig.3 Temporal development of listener parameter

3.1.1 話者の間長から聴者特徴量への影響

Fig. 4は発話長(SP)を900msに固定した条件下で、各被験

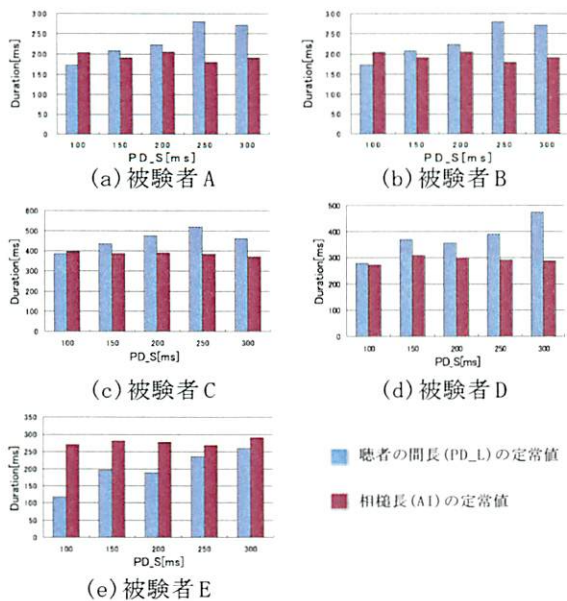


図4 話者の間長と聴者特徴量の定常値の関係
Fig.4 Relation between stationary value of PD_S and listener parameter

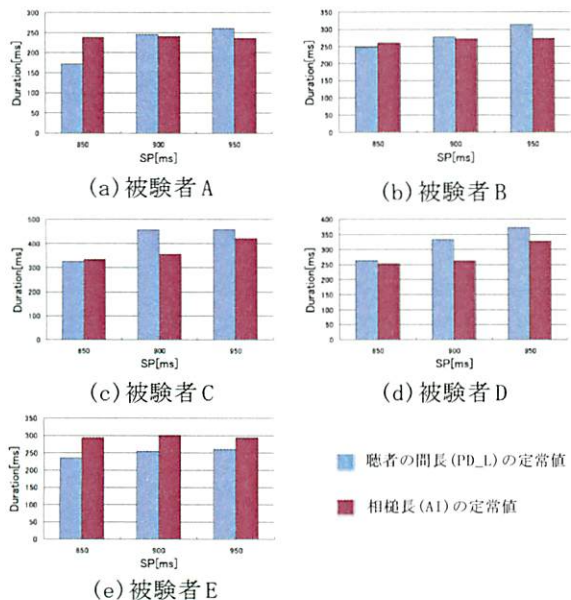


図5 発話長と聴者特徴量の定常値の関係
Fig.5 Relation between stationary value of SP and listener parameter

者毎に話者の間長 (PD_S) と聴者特徴量の定常値の関係を示したものである (設定条件は (SP, PD_S) = (900, 100), (900, 150), (900, 200), (900, 250), (900, 300) の5種類). 聴者特徴量の定常値は各条件下の試行において開始20秒以降に観測された聴者特徴量の平均値を示している. 図が示すように話者の間長 (PD_S) が増加すると, 聴者の間長 (PD_L) の定常値は増加し, 相槌長 (AI) の定常値はあまり変化しない. すなわち, 話者の間長が聴者の間長に影響するが, 相槌長にはあまり影響しないことがわかる. ここで, 統計的手法により定量的に確認するために, PD_S = 100ms時の聴者特徴量を基準とし, 図

		PD_S					PD_S					
		100	150	200	250	300	100	150	200	250	300	
sub.a	1	1.202	1.289	1.618	1.572		sub.a	1	0.936	1.005	0.877	0.936
sub.b	1	1.405	1.769	1.887	1.903		sub.b	1	0.948	1.04	1.072	0.964
sub.c	1	1.133	1.237	1.352	1.2		sub.c	1	0.974	0.982	0.967	0.932
sub.d	1	1.332	1.289	1.412	1.708		sub.d	1	1.136	1.103	1.07	1.059
sub.e	1	1.675	1.607	2.009	2.214		sub.e	1	1.041	1.026	0.993	1.078

(a) 聴者の間長 (PD_L) (b) 相槌長 (AI)

表1 基準値に対する聴者特徴量の比率 (Fig. 4)
Table.1 Ratio of listener parameter

		SP			SP			
		850	900	950	850	900	950	
sub.a	1	1.43	1.517		sub.a	1	1.008	0.991
sub.b	1	1.117	1.266		sub.b	1	1.046	1.05
sub.c	1	1.403	1.406		sub.c	1	1.063	1.254
sub.d	1	1.261	1.409		sub.d	1	1.036	1.298
sub.e	1	1.081	1.106		sub.e	1	1.024	0.997

(a) 聴者の間長 (PD_L) (b) 相槌長 (AI)

表2 基準値に対する聴者特徴量の比率 (Fig. 5)
Table.2 Ratio of listener parameter

4の結果を基準値に対する比率で表す (Table. 1). 話者の間長 (PD_S) と聴者の間長 (PD_L) に関しては $F(4, 24) = 6.26$, $P < 0.01$ であり, 帰無仮説は棄却された. また, 話者の間長 (PD_S) と相槌長 (AI) に関しては $F(4, 24) = 0.29$, $P > 0.2$ であり, 相槌長 (AI) が話者の間長 (PD_S) から有意な影響を受けないことが確認された. しかし, 話者の間長 (PD_S) が十分大きくなると, 話者の間長 (PD_S) から聴者の間長 (PD_L) の定常値への影響が減少する傾向にある (Fig. 4).

3.1.2 発話長から聴者特徴量への影響

Fig. 5は話者の間長 (PD_S) を 200ms に固定した条件下で, 各被験者毎に発話長 (SP) と聴者特徴量の定常値の関係を示したものである (設定条件は (SP, PD_S) = (850, 200), (900, 200), (950, 200) の3種類). 聴者特徴量の定常値は 3.1.1 と同様に各試行において開始20秒以降に観測された聴者特徴量の平均値を示している. 図が示すように発話長 (SP) が増加すると, 聴者の間長 (PD_L) の定常値も増加し, 発話長が聴者の間長に影響することがわかる. 一方, 発話長 (SP) から相槌長 (AI) の定常値への影響については各被験者で異なった. 3.1.1 と同様に, 図5の結果を基準値に対する比率で表し (Table. 2), 統計的手法により定量的に確認する. 発話長 (SP) と聴者の間長 (PD_L) に関しては $F(2, 14) = 9.37$, $P < 0.01$ であり, 帰無仮説は棄却された. また, 発話長 (SP) と相槌長 (AI) に関しては $F(2, 14) = 2.50$ となり, $P < 0.15$ で帰無仮説は棄却されるが, $P < 0.1$ では棄却されない.

3.2 話者の発話に摂動が含まれる場合

本実験では対話中に発話の長さ, 間隔の長さに変化を生じる時に, 聴者の相槌行為の特徴を明らかにすることを目的とする. 3.1の実験では話者特徴量から聴者の間長への影響が明らかになったが, その影響において, 聴者の間長がどのような応答特性を持つかを調べる必要がある. 具体的には話者特徴量に摂動を加え, 聴者特徴量の遷移過程を調べる. なお, 摂動は実験開始から30秒後に加えられ, 発話長への摂

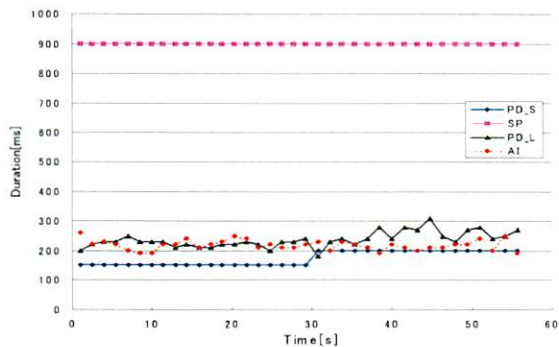


図6 話者の間長変化に対する聴者特微量の時間発展
Fig.6 Temporal development of listener parameter in PD_S shift

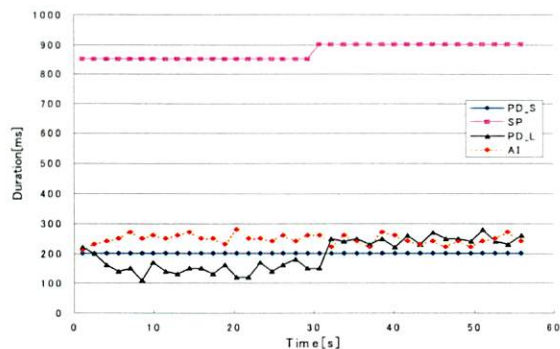


図7 発話長変化に対する聴者特微量の時間発展
Fig.7 Temporal development of listener parameter in SP shift

動付加は話者の発話スピードを変化させることにより実現された。

3.2.1 話者の間長に摂動を加えた場合

Fig. 6は発話長(SP)を900msに固定した条件下で、話者の間長(PD_S)を150msから200msに変化した時の結果である(被験者A)。図が示すように、話者の間長(PD_S)を増やすと聴者の間長(PD_L)だけが増加し、一方、相槌長(AI)は変化しないことがわかる。これは3.1.1で示された話者の間長から聴者の間長への影響はあるが、相槌長には影響しないという結果を支持する。また、聴者の間長(PD_L)は摂動付加以前の定常値から摂動付加後の定常値に、約10秒程度の時間スケールでゆっくりと収束することが明らかになった。

3.2.2 発話長に摂動を加えた場合

Fig. 7は話者の間長(PD_S)を200msに固定した条件下で、発話長(SP)を850msから900msに変化させた時の結果である(被験者A)。図が示すように、発話長(SP)を増やすと聴者の間長(PD_L)が増加し、3.1.2で示された話者の発話長から聴者の間長への影響があるという結果を支持する。さらに、聴者の間長(PD_L)は摂動付加直後すぐに変化し、摂動付加以前の定常値から摂動付加後の定常値へ急速に収束することがわかる。

4. 考察

4.1 話者が定常的に発話する場合

3.1.1では定常的な対話において話者の間長から聴者の間長への作用が強く、相槌長への作用が弱いことが明らかになった。このような間(ま)から間(ま)への作用は自由対話においても間長の同期現象として報告されており[6]、話者の間と聴者の間が自然に合わせられていると考えられる。また、3.1.2では定常的な対話において発話長から聴者の間長への作用が強いことが明らかになった。このとき聴者は話者の発話と合わせるために、意図的にタイミング(ま)を変えて相槌を入れていたと推測される。このように聴者の間には自然に合うといった無意識的な側面と、意図的に合わせるといった意識的な側面が存在すると推測される。そして、この2つの側面が相槌行為という特殊な対話を成立させていると考えられる。

4.2 話者の発話に摂動が含まれる場合

3.2では話者特微量の変化に対する聴者の間長の応答特性を調べるために摂動付加実験が行われた。その結果、話者の間長変化に対する応答は長い時間スケールを、発話長変化に対する応答は短い時間スケールを持ち、時間的な階層性を構成することがわかった。摂動付加実験終了後、被験者に感想を求めると、発話長の変化は知覚できるが、間長の変化は知覚できないという意見を得られた。これは意識化される側面と意識化されない側面が存在し、知覚過程においても2つの側面があることを示唆する。そして、意識化される側面が短い時間スケールで意識的な側面に影響し、意識化されない側面が長い時間スケールで無意識的な側面に影響すると考えられる。この内容はコミュニケーションの二重性の存在を示唆すると同時に、二重性と時間的な階層性との関連を示唆している。

6. 参考文献

- [1]清水, 久米, 三輪, 三宅: 場と共創; NTT出版, 2000
- [2]Y. Tatsumi, and Y. Miyake: Hierarchy and Fluctuation in Time Structure of Spoken Dialogue; pro. of 1999 IEEE International Conference on Systems, Man and Cybernetics (SMC' 99), pp. V-1040-V-1044, 1999
- [3]マイケル・ボラニー: 暗黙知の次元; 紀伊国屋書店, 1980
- [4]三宅: 「生命」における設計; 現代思想, vol. 25, no. 6, pp. 301-317, 1997
- [5]三宅: 「二中心モデル」とインタフェース表現; 日本ファジィ学会誌, vol. 9, no. 5, pp. 637-647, 1997
- [6]長岡, 小森, 中村: 自由対話における話者交替の潜時と呼吸の関連; 第13回ヒューマン・インタフェース・シンポジウム, pp. 311-314, 2000
- [7]渡辺, 大久保, 稲留: 対話コミュニケーションにおける呼吸の引き込み; 第13回ヒューマン・インタフェース・シンポジウム, pp. 271-276, 1997
- [8]杉藤: 日本語音声の研究1 日本人の声, 和泉書院, 1994