

相づちのリアルタイム性の分析と 自動生成アルゴリズムの設計

香曾我部 多門¹ 山田 節夫² 西野 哲朗¹

概要: 音声対話システムにおいて円滑なコミュニケーションを実現するためには、システム側が適切なタイミングで相づちを打つことでユーザの発話を促すことが、有効な手段の一つであると考えられる。本論文では、まず、人が相づちを打つタイミングの分析を行い、その分析結果を基に設計した相づちのリアルタイム自動生成アルゴリズムについて述べる。さらに、その相づち自動生成アルゴリズムを組み込んだ音声対話システムを構築し、相づちの生成タイミングや頻度に関する評価実験を行うことで、提案手法の有効性を検証する。

キーワード: 相づち自動生成, 雑談型対話, 音声対話

TAMON KOHSOKABE¹ SETSUO YAMADA² NISHINO TETSURO¹

1. はじめに

近年、音声対話システムの形態のひとつとして、明確な目的を持たない対話によってユーザーの欲求を満たすことを目的とした、非タスク指向型対話システムが検討されるようになってきている。こうしたシステムは、ユーザーの退屈や寂しさを紛らわす効果があり、話し相手のいない高齢者のケアなど介護分野への利用も期待されている。このような話し相手の役割を果たすシステムを実現するために、傾聴という機能を搭載することが提案されている [1]。傾聴とは相手の話に耳を傾け、同意や共感をしめす反応をすることで、積極的な会話を促す行為のことである。

傾聴における同意や共感をしめす反応の一つに相づちや復唱があり、これを自動的に生成するための研究が既に行われている。生成タイミングを推定するために、先行発話の韻律情報や言語情報から、相づちが打たれやすいパターンを推定し相づち生成タイミングを決定する手法や [2], [3], [4], それらの情報を素性として機械学習を用いる手法 [5], [6], [7] が提案されている。しかし、音響的手法による相づちは、発話の意味的な要素を考慮しないため、相手の発話に理解を示すという相づちの役割を果たさず、必要以上に相づちを

生成したり、人間の打つ相づちの特徴と異なることがある。また、韻律情報と言語情報の両方の特徴を訓練データとした決定木による手法は、発話の意味的内容を考慮したタイミングで相づちを生成することを可能としているが、ユーザの話し方や対話の内容によっては、訓練データとの特徴と異なることが原因で自然なタイミングで相づちを生成することができないことが小林らによって示されている [6]。

本研究では、これらの問題を解決するため、人間の相づちの特徴を考慮し、ユーザの自由な発話に対して、より自然なタイミングで相づちを生成するシステムの構築を目的とする。相づち自動生成システムがユーザに与える人間らしさの印象は、相づちのタイミングが影響すると指摘されており [6]、目的とする相づちの自然なタイミングとは、ユーザがシステムとの対話を人間らしいと感じるようなものである。本論文では、人間同士による対話から、相づちが打たれる際の言語的特徴と、相づちの生成頻度の特徴を分析し、得られた知見から相づちの自動生成アルゴリズムを構築した。提案するアルゴリズムでは、表層的な言語情報と生成頻度に関する調節により、人間の打つ相づちの特徴を再現するように生成タイミングを決定する。このアルゴリズムを評価するために、音声対話システムを構築し、被験者 12 名による主観評価を行ったところ、提案手法は音響的手法よりもより自然なタイミングで相づちを生成しており、より人間らしい対話であると評価された。以下本論文では、2 節

¹ 電気通信大学
University of Electro Communications

² 日本電信電話株式会社
Nippon Telegraph and Telephone Corporation

で人間の打つ相づちの分析方法と分析結果について述べ、3節で提案する手法の詳細、4節で評価実験について述べる。5節では評価実験の結果と結果に対する考察を述べる。

2. 人間の打つ相づちの分析

人間による相づちのタイミングには、先行発話の音声ピッチや発話終端からの無音区間長などの音響的、あるいは直前の単語の特徴などの言語的な情報からなる複合的な要因が関係していると考えられている [3], [4], [5], [7]. そこで本論文では、相づち自動生成アルゴリズムの設計のために、このような人間の打つ相づちの特徴を分析した。

分析対象として、2人による対話を収録したものを文字データに書き起こし、相づちタイミングと前回打った相づちからの経過時間をタグとして付与した対話コーパスを作成した。対話は、大学生及び大学院生による2組4名が、対話ドメインとして話しやすい趣味の一つ決めてもらい、それに関する内容を互いに語り合うという条件で収録した。対話では自然に話者交代が行われているが、一方の話者が話し手に回っている部分を切り出してその話者1人分のデータとし、計4人分のデータを作成した。このデータをそれぞれ a, b, c, d とする。対話時間はそれぞれ20分程度であり、4人の合計で487箇所相づちが打たれていた。

2.1 表層的言語情報に関する分析

文節が発話された直後に相づちが打たれる場合が多く、また文節の最終形態素の品詞によって相づちの打たれる割合が異なることが分かっている [7]. 上記の4人のデータ全体対して、すべて文節の最終形態素の品詞を調べ、品詞ごとの直後に相づちが打たれた割合を表1に示す。この表の見方について、例えば助動詞の値は、コーパス中の全ての文節の最終形態素において助動詞が出現した回数が389回、そのうち直後に相づちが打たれた回数が110回で、その割合が $110/389 = 28.3\%$ であることを表している。ここで、文節境界および形態素情報の解析には日本語係り受け解析器 CaboCha[9] を使用した。また、品詞の分類は、CaboChaがデフォルトで採用している IPA 品詞体系によるものである。

表1より、文節の最終形態素の品詞が、助動詞、名詞、助詞のときに共通して高い割合で相づちが挿入されており、次いで動詞、人により形容詞のときに相づちが打たれやすいことが分かった。

また、4つのデータ全体に対して、文節の最終形態素の品詞が助詞のとき、助詞の分類ごとの相づちの挿入割合を調査した結果、助詞の種類が終助詞のときに 59.1% ($94/159$) の割合で相づちが挿入されており、助詞全体での割合である 16.1% を大きく上回っていたことが分かった。終助詞は文の末尾に付き、その文を完結させるため、音声対話においても意味的な区切りを表していると考えられる。同様に、

助動詞においても、丁寧な断定の助動詞の「です」、特殊活用の「ます」など、いわゆる“ですます調”が文末に現れるときに相づちが打たれやすく、4つのデータ全体においても、それぞれ「です」が $24/34 = 70.5\%$ 、「ます」が $10/14 = 71.4\%$ の割合で相づちが打たれている。

表1 文節の最終形態素の品詞と相づち挿入割合

品詞	相づち挿入割合
助動詞	28.3% (110/389)
名詞	18.1% (31/171)
助詞	16.5% (262/1584)
動詞	13.2% (27/204)
形容詞	7.7% (9/117)
副詞	1.1% (3/264)
接続詞	0.6% (1/167)
連体詞	0.0% (1/110)
感動詞	0.0% (0/11)

以上の分析より、先行発話の文節の形態などの表層的な言語的特徴によって相づちの打たれやすさが変化し、とりわけ終助詞や断定の助動詞「ます」などの文末表現で発話が終了しているときは、比較的強い意味的区切りを表しており、相づちが打たれやすいことが分かった。

2.2 頻度に関する分析

聞き手が傾聴し、話し手に対して相づちを打つことにより同調を示しているとき、相づちを打つ頻度や間隔にどのような特徴が現れるかを調べるため、直前に打った相づちからの経過時間に対する相づちの挿入頻度について分析した。一方が傾聴する場面を対象とするため、対話コーパスのうち、一方の話者が30秒以上連続で発話し、かつ聞き手がその発話区間に対して2回以上相づちを打っている部分を分析対象とした。図1に、4人それぞれの経過時間1秒ごとの相づち挿入頻度を示す。これより、相づちの回数は経過時間がある秒数をピークとした釣鐘型の度数分布を描き、ピークとなる経過時間が個人で異なることが分かった。また、対話全体の単語数に対する相づち回数も、最も多い人で約14単語あたり1回、最も少ない人で約26単語あたり1回と、個人間で大きな差がみられた。

頻度に関する特徴には個人差があり、これは聞く立場になったときに、相手の相づちの頻度によって、受ける印象に大きな変化があることを示唆している。相づちの自動生成システムにおいては、このような個人差に柔軟に対応できるような設計が求められる。

3. 相づち自動生成アルゴリズム

2節の分析結果を踏まえ、相づちを自動生成するためのアルゴリズムを設計し、音声入力に対してリアルタイムに相づちを応答可能な音声対話システムを構築する。

2.1節において述べたように、相づちは意味的な区切りで

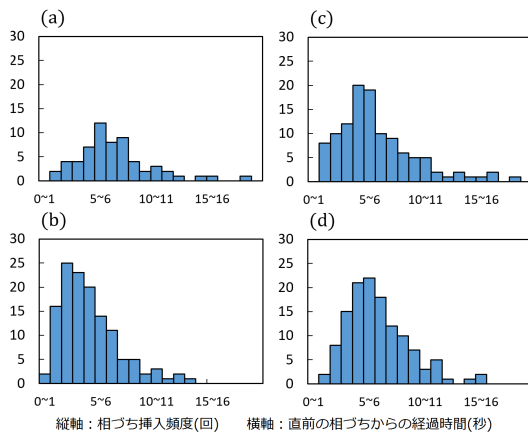


図 1 直前に打った相づちからの経過時間に対する相づちの挿入頻度生成されることが多い。しかし、品詞の形態や前回打った相づちからの経過時間によっては、明確な意味的区切りでなくとも相づちが打たれる場合も多く、こうしたタイミングでは同じ条件でも必ずしも打たれるわけではない。つまり、人間による相づちのタイミングは、品詞や頻度の特徴を捕らえつつも、ランダム性を持つものだと考えられる。このことを踏まえ、提案する相づち自動生成アルゴリズムでは、相づち生成タイミングを、先行話者の発話の意味的区切りで生成される場合と、ある条件のもとに確率的に生成される場合とに分け、それぞれ別な方法で相づち生成タイミングを決定する。前者の場合には、先行発話が文末表現を含むかどうかを基準にルールベースで相づち生成の適否を判定する。後者の場合には、2 節で分析した文節の最終形態素の品詞の特徴と、直前に打った相づちからの経過時間に対する頻度の特徴から、相づち生成確率を求め、確率的に相づちを生成するか否かを決定する。

構築した音声対話システムの構成を図 2 に示す。以下では、このシステムの各部の詳細について説明する。

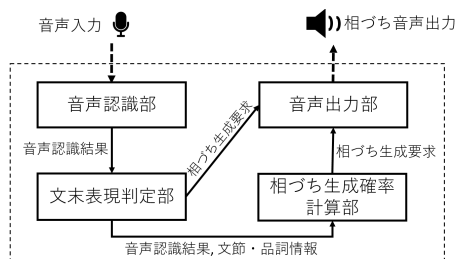


図 2 構築した音声対話システムの構成

3.1 音声認識部

音声入力の言語情報をテキストとして得るために、音声入力に音声認識エンジンの Julius[10] を適用する。本システムでは 200 ミリ秒の無音区間が検出されるごとに音声認識結果を出力するように設定した。音声認識部は Julius の音声認識結果を受け取るたびに、記号や空白文字を除去したテキストを文末表現判定部に送り、相づち生成の判定を

行う。

3.2 文末表現判定部

文末表現判定部では、入力の文末の表現から相づちを生成するか判定する。まず、音声認識部から受け取った認識結果の文字列に対して、CaboCha を用いて文節区切り、および形態素ごとの品詞分類を行う。解析結果から、入力の最後の文節に、特殊型の助詞「ます」、断定の助動詞「だ」「です」を含んでいた、または終助詞で終わっていた場合、相づちを生成できると判定する。相づちを生成できると判定した場合、品詞情報に基づく確率計算部と頻度調節部の処理をスキップし、音声出力部に相づち生成を要求する。文末表現判定部で相づちを生成しない場合、CaboCha の解析結果を品詞情報に基づく確率計算部に送る。

3.3 相づち生成確率計算部

相づち生成確率計算部では、品詞情報と前回打った相づちからの経過時間から、相づち生成確率を計算する。まず、CaboCha の解析結果から品詞情報に基づく相づち生成確率を計算し、次にこの確率に前回打った相づちからの経過時間に応じてバイアスを加えることで最終的な相づち生成確率を求める。この相づち生成確率に従って確率的に相づちを生成するか決定する。以下に具体的な計算方法を説明する。

3.3.1 品詞情報に基づく確率計算

まず、文末表現判定部から受け取った CaboCha の解析結果から、品詞情報を基に相づち生成確率を求める。ここで求める相づち生成確率とは、2.1 節で分析した表 1 に示すような、文節の最終形態素の品詞別の相づちの打たれやすさを考慮し、システムに入力された対話全体で、この品詞別の相づちの挿入割合を再現するように計算される。

品詞情報に基づく相づち生成確率は、あらかじめパラメータとして設定された品詞ごとの基本相づち生成確率と、入力テキストの文節の最終品詞の出現回数から計算する。以下に、パラメータと変数の初期設定と、具体的な計算手順を説明する。

パラメータと変数の初期設定

- (初期設定 1) 品詞ごとの基本相づち生成確率を設定し、 F_{pos} と表す (pos には各品詞名が入る)。この基本相づち生成確率は 2.1 で示したような、システムとして再現したい品詞ごとの相づちの挿入割合を設定する。
- (初期設定 2) 品詞ごとの出現回数を記録する変数を N_{pos} と表し、0 に初期化する。

計算手順

- (手順 1) 文末表現判定部から入力テキストの文節の区切りの情報のすべての形態素の品詞分類の情報を受けとる
- (手順 2) 文節ごとに最終形態素の品詞を調べ、その品

詞に対応する変数 N_{pos} に 1 を足す.

- (手順3) 入力最終文節の最終形態素の品詞を fin_pos とするとき、確率 P_{fin_pos} の試行を N_{fin_pos} 回繰り返したとき少なくとも 1 回成功する確率を、この入力に対する“品詞情報に基づいた相づち生成確率” P_1 とする.
- (手順4) fin_pos の出現回数 N_{fin_pos} を 0 にする.
- 次の入力は手順3から繰り返す. (fin_pos 以外の品詞の出現回数 N_{pos} は次の入力に持ち越される)

以下、入力テキスト「私はおいしいリングを一つ食べた」を例に、相づち生成確率計算部の動作について、順を追って説明する.

- 初期設定 1 品詞毎の基本相づち生成確率として、例えば、 $F_{助動詞} = 30\%$, $F_{助詞} = 25\%$, その他の品詞には全て 0% と設定する.
- 初期設定 2 品詞ごとの出現回数を記録するための変数, $N_{助動詞} = 0, N_{助詞} = 0, \dots$ のように 0 を初期値として設定する.
- 手順 1 入力テキストの文節の区切りの情報とすべての形態素の品詞分類の情報を受けとる. 入力テキストを文節に分割すると、「私は / おいしい / リングを / 一つ / 食べた」となり、さらに品詞に変換すると「名詞 助詞 / 形容詞 / 名詞 助詞 / 名詞 / 動詞 助動詞」となる.
- 手順 2 各分節の最終形態素の品詞は助詞が 2 回で助動詞、名詞、形容詞が各 1 回なので、 $N_{助詞} = 2, N_{助動詞} = N_{名詞} = N_{形容詞} = 1, \dots$ となる.
- 手順 3 このとき、相づち生成の適否は直前の品詞、つまり入力最終品詞によって決めるため、最後の品詞である助動詞に関する相づち生成確率を計算する. $N_{助動詞} = 1$ なので、この入力に対する、“品詞に基づく相づち生成確率” P_1 は $F_{助動詞} = 30\%$ の試行を 1 回行って成功する確率 $P_1 = 30\%$ となる.
- 手順 4 最後に、 $N_{助動詞}$ を 0 に初期化する.

以上が最初の入力が行われたときの計算手順である. また、この入力に対する相づち生成判定の処理の終了後、次の入力として「このリングは」というテキストが入力された場合、文節は「連体詞 / 名詞 助詞」となり(手順1)、各品詞の出現回数は、新たに連体詞が $N_{連体詞} = 1$, 助詞は 1 増えて $N_{助詞} = 3$ となる(手順2). このとき入力最終品詞は助詞なので、この入力に対する、“品詞に基づく相づち生成確率” P_1 は、 $F_{助詞} = 25\%$ の試行を $N_{助詞} = 3$ 回反復したとき少なくとも 1 回成功する確率 $P_1 = 1 - 0.75 * 0.75 * 0.75 = 0.58$ となり、最後に $N_{助詞}$ を 0 に初期化する.

3.3.2 前回打った相づちからの経過時間に応じた調節

2.2 節で示したように、人間の打つ相づちは直前に打った相づちからの経過時間に対して釣鐘型の分布を持つ. よっ

て、システムとの対話全体としてこのような分布を模倣するように、求めた品詞に基づく相づち生成確率 P_1 に対してバイアスかける. 前回打った相づちからの経過時間 x 秒に応じて、相づちのうたれやすさを調節するために、品詞に基づく確率計算部から受け取った相づち生成確率 P_1 に対して、以下の式で表される計算を行い、最終的な相づち生成確率 P_2 を求める.

$$P_2' = P_1 - (1 - 2Q(x))(1 - P_1)$$

としたとき、

$$P_2 = \begin{cases} P_2' & (P_2' \geq 0) \\ 0 & (P_2' < 0) \end{cases} \quad (1)$$

ここで、 $Q(x)$ は模倣する分布の $x + 1$ 秒までの累積相対度数であり、パラメータとして事前に設定する. たとえば図3のような分布をパラメータとする場合、 $Q(0) = 0, Q(1) = 0.01, \dots, Q(6) = 0.55, \dots, Q(15) = 1.00$ となる. このようにパラメータを設定したとき、受け取った P_1 と経過時間 x に応じて、最終的な相づち生成確率 P_2 は図4のように変化する.

こうして求めた最終的な相づち生成確率 P_2 から、確率的に相づちを生成する. 具体的には、システムは 0 以上 1 以下の乱数を生成し、乱数が生成確率 P_2 よりも小さい値だったとき、音声出力部に相づち生成を要求する.

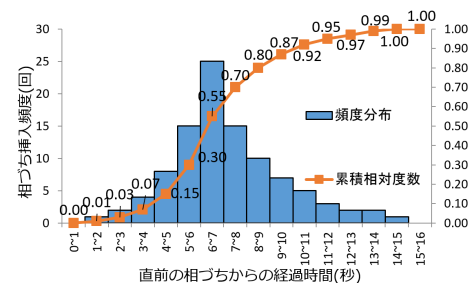


図3 パラメータとして設定する頻度分布

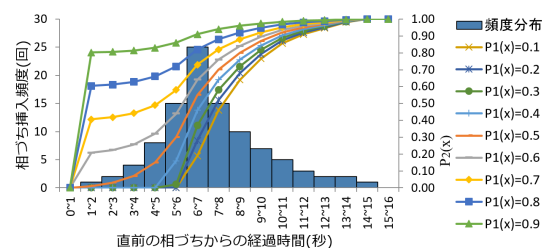


図4 P_1 と経過時間に応じて変化する P_2

3.4 音声出力部

相づちを生成すると判定されたとき、音声出力部より相づち音声を出力する. 出力する相づち音声として、一般的な相づちの発話である「うん」を、音声合成ソフトにより作成したものをを用いた. なお、音声合成ソフトウェアとし

て、CeVIO プロジェクトの CeVIO Creative Studio を使用した。

4. 評価実験

3 節で説明した相づち自動生成アルゴリズムが、音響的手法と比べてより自然なタイミングで相づちを生成し、ユーザに人間らしい印象を与えられるかを確認するため、また、頻度調節による効果を確認するために行った評価実験について説明する。

4.1 評価対象の音声対話システム

音響的手法、提案手法 (頻度調節なし)、提案手法 (頻度調節あり) の 3 通りの相づち自動生成手法を比較するために、それぞれの手法を実装した 3 つの音声対話システムを作成した。

音響的手法は、ユーザの発話長が 200 ミリ秒以上の発話の後に、200 ミリ秒以上の無音区間を検出したときに相づちを生成する。提案手法 (頻度調節なし) は 3 節で説明した手法だが、相づち生成確率に時間経過に応じたバイアスをかけないものである。つまり、3.3 節における品詞情報に基づく相づち生成確率 P_1 をそのまま最終的な相づち生成確率として用いる。ここで、品詞情報に基づく確率計算におけるパラメータである、各品詞の基本相づち生成確率は、助動詞が 0.30、名詞と助詞が 0.20、動詞が 0.15、形容詞が 0.10、その他の品詞が 0.00 と設定した。提案手法 (頻度調節あり) は 3 節で説明した手法そのものであり、品詞情報に基づく確率計算におけるパラメータは案手法 (頻度調節なし) と同様の値を、前回打った相づちからの経過時間に応じた調節におけるパラメータとして表 2 の通りに、それぞれ設定した。なお、このパラメータは 2 節で分析した結果の平均的な値から暫定的に決定した。

4.2 評価用の対話音声の作成

前節で説明した 3 つの音声対話システムの評価を行うために、音声入力に対してシステムが応答する相づちが合成された対話音声を作成する。この対話音声を第 3 者の被験者に聴取してもらい、アンケートによる主観評価を実施した。本節では、対話音声の作成方法について述べ、主観評価については次節で述べる。

本実験では、音声認識による認識誤りを最小限にするために、あらかじめ正しく音声認識されるように収録した音声をシステムの入力として、相づち生成を行った。入力する音声は、被験者 1 名により収録した。また、対話タスクとして、被験者が友人と自分の趣味や思い出について話している雑談対話を想定し、その対話の中で自分が話し続けている場面を切り出した設定で収録し、対話シナリオの異なる 3 つの音声データを作成した。この音声データをそれぞれ音声データ A, B, C と呼ぶ。

次に、音声データ A, B, C のそれぞれに対して、3 つの手法で相づち生成タイミングを検出し、相づち音声を合成する。音声データとシステムによる相づちの合成方法は以下の通りである。音声データを音声対話システムを構築した PC に対してライン入力し、マイクでの音声入力と同様の環境でリアルタイムに相づちが生成されるようにして、入力音声とそれに対する相づちの音声をシステムを構築した PC 上で同時に合成した。このように音声データ A に対して、音響的手法、提案手法 (頻度調節なし)、提案手法 (頻度調節あり) の 3 通りの手法で相づちを合成した 3 つの音声データを作成した。同様の処理を音声データ B, C に対しても同様に行い、計 9 つの相づちが合成された対話音声データを作成した。図 5 に、音声データ A に対して、各手法が生成した相づちの箇所を示す。

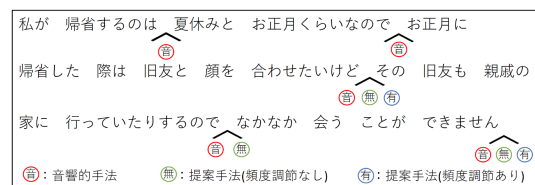


図 5 各手法により生成された相づち箇所の例

4.3 主観評価

システムが生成する相づちについて、ユーザの印象を調査するために、アンケートによる主観評価を実施した。なお、主観評価の方法として、小林らが行った方法を参考にした [6]。被験者 12 名が、前節で作成した対話音声データを聴取し、表 3 に示す項目について、問 1, 問 2, 問 3 には 5 段階で、問 4 には 2 択で評価を行った。各被験者は、9 つの音声データのうち、対話シナリオと相づちの生成手法が重複しないような 3 つを選んで聴取し、評価した。

5. 実験結果と考察

表 4 に、相づち生成手法ごとに評価をまとめたアンケートの結果を示す。なお、表に示す 5 段階評価の評価値は、数字が大きほど良い評価であることを意味する。表より、全てのアンケート項目について、提案手法 (頻度調節あり)、提案手法 (頻度調節なし)、音響的手法の順に高い評価を得ていることが分かる。

小林らは、音響的手法と言語的特徴に基づく決定木による手法を比較する実験を行い、音響的手法の方がより人間らしい相づちを生成できると報告している。本実験で用いた音響的手法は小林らが用いた音響的手法と同一のものであり、これをベースラインとしたとき、提案手法は、より自然なタイミングで相づちを生成し、より人間らしい印象を与えることができたと考えられる。音響的手法による相づちは、他の 2 つの手法による相づちタイミングを 9 割以上網羅したうえで、相づちの総数は 2 倍近くあるため、不必要

表 2 頻度調節部において設定したパラメータ

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$Q(x)$	0.00	0.05	0.15	0.30	0.50	0.65	0.75	0.82	0.87	0.91	0.94	0.96	0.68	1.00

表 3 アンケート項目

記号	質問内容
問 1	相づちのタイミングが自然だった
問 2	相づちの仕方が人間らしい
問 3	相づちの頻度が丁度よい
問 4	相づちが少ない、または多すぎる (問 4 は問 3 にて 1 か 2 を選んだ場合のみ解答)

表 4 アンケートの結果

	音響的手法		提案手法 (頻度調節なし)		提案手法 (頻度調節あり)	
	評価値の平均	評価値の平均	評価値の平均	評価値の平均	評価値の平均	評価値の平均
問 1	3.00		3.42		3.75	
問 2	3.08		3.42		3.50	
問 3	1.92		3.50		4.08	
	選んだ人数		選んだ人数		選んだ人数	
評価値	少ない	多い	少ない	多い	少ない	多い
問 4	0	10	2	1	0	0

なタイミングでの相づちを多く含んでいたと考えられる。実際、音響的手法を頻度が悪い (2 以下と) 評価した全員が問 4 の質問に「多い」と回答していた。

提案手法 (頻度調節あり) と提案手法 (頻度調節なし) の結果から、前回打った相づちからの経過時間に応じた頻度調節により、ユーザにより良い印象を与えることができることが確認できた。頻度調節による相づちの生成のされ方を確認するため、提案手法 (頻度調節あり) と提案手法 (頻度調節なし) による対話音声の前回打った相づちからの経過時間に対する相づち生成頻度の分布を調べ、このグラフを図 6 に示した。この図に示より、相づちの生成総数は、それぞれ 121 回と 107 回であり、大きな違いはないが、提案手法 (頻度調節なし) は提案手法 (頻度調節あり) に比べ、相づちの生成間隔のばらつきが大きくなっていることが分かる。提案手法 (頻度調節なし) による対話では、局所的に相づち頻度が高くなる場面や、逆にしばらく相づちが打たれない場面が見られ、これは人間の打つ相づちの特徴と異なるため、ユーザに悪い印象を与える要因になったと考えられる。

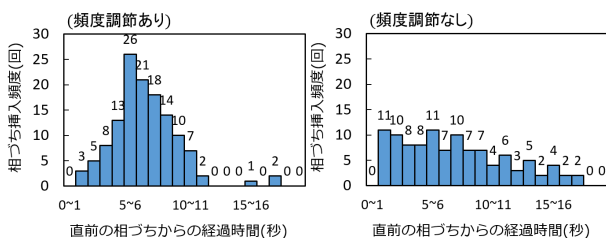


図 6 提案手法の音声対話システムによる直前に打った相づちからの経過時間に対する相づちの挿入頻度

6. おわりに

本研究では、実際の雑談対話から人間の打つ相づちの特徴を分析し、分析結果から得られた特徴を用いて、人が打つ相づちを模倣するシステムの構築を行った。システムはユーザの発話入力に対してリアルタイムに相づちを生成することが可能であり、パラメータを変更することで相づちの生成タイミングを調節することも可能である。本論文ではパラメータには、分析結果から得られた平均的な値を設定した。音響的手法と提案手法を用いて音声対話システムを構築したうえで、ユーザの音声に対してシステムが生成する相づちタイミングについて評価実験を行った。その結果、提案手法が音響的手法に比べより自然なタイミングで相づちを生成でき、ユーザにより人間らしい印象を与えることができることが確認できた。

今後は、介護分野などの実用も視野にいれ、実際に提案手法のシステムを用いた際の効果や問題点を調べるため、傾聴型対話システムにおいて相づちの果たす役割や、ロボット等による相づちの表現方法によるユーザが受ける印象についても調査する予定である。

参考文献

- [1] 横山祥恵, 山本大介, 小林優佳, 土井美和子. 高齢者向け対話インタフェース—雑談継続を目的とした話題提示・傾聴の切替式対話—. 2010-SLP-80(4), pp1-6, 2010
- [2] 平沢純一, 川端豪: 音声対話システム Noddy —ユーザ発話途中でのうなずき・相槌生成—, 情報処理学会研究会報告, 1997-SLP-020, pp.51-52 (1998).
- [3] N. Cathcart, J. Carletta, and E. Klein: A shallow model of backchannel continuers in spoken dialogue, In Proc. of EACL2003, pp.5158, (2003).
- [4] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一: 韻律情報を用いた相槌の挿入, 情報処理学会論文誌, Vol.40, No.3, pp.469-477 (1999).
- [5] 竹内真士, 北岡教英, 中川聖一: 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, 2004-SLP-050, pp.87-92 (2004).
- [6] 小林一樹, 船越孝太郎, 小松孝徳, 山田誠二, 中野幹夫: ASEに基づく相づちによるロボットとの対話体験の向上, 人工知能学会論文誌, v.30, pp.604-612 (2015).
- [7] 神谷優貴, 大野誠寛, 松原茂樹, 柏岡秀紀: 同調的対話システムにおけるあいづち挿入タイミング, 言語処理学会年次大会発表論文集, v.16, pp.365-398 (2010).
- [8] 神谷優貴, 大野誠寛, 松原茂樹, 柏岡秀紀: 音声対話コーパスに基づく相づち生成タイミングの検出とその評価, 言語処理学会年次大会発表論文集, v.17, pp.103-106 (2011).
- [9] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, v.43, pp.1834-1842, 2002.
- [10] 河原達也, 李晃伸: 連続音声認識ソフトウェア Julius, 人工知能学会誌, Vol.20, No.1, pp.41-49, (2005).