E-022

聞き役雑談対話システムのための非言語音響特徴量に関する検討
A Study on non-verbal acoustic features for listening-oriented dialogue system

栂井 良太[†1]　伊東 伸泰[†2]　綱川 隆司[†1]　西田 昌史[†1]　西村 雅史[†1]
Ryota Togai[†1]　Nobuyasu Ito[†2]　Takashi Tsunakawa[†1]　Masafumi Nishida[†1]　Masafumi Nishimura[†1]

1. はじめに

近年，高齢者に生きがいを与えるための聞き役雑談対話システムが注目を集めているが，その多くはユーザー入力として完璧なテキスト入力を前提としている[1][2]．一方，聞き役雑談対話システムにおいてはユーザー入力として言語情報以外の情報も有用であると考えられている[3]．我々は先に，ユーザー入力として簡易的な非言語音響情報だけを利用した統計的対話システムを提案し，その選択応答の自然さについての評価を行った[4]．結果として，人間が選択した応答に近い評価を得ることができ，対話システムにおける非言語音響情報の可能性を示すことができた．

しかし，先の報告は学習対象者が 1 名の場合の結果に過ぎなかった．本報告では統計的対話システムの学習者を 5 名に増やして評価結果の一般性を高めるとともに，対話システムの学習対象話者が，当該システムを使った場合の対話の評価に加え，学習対象話者外の話者との対話の自然性についても，第 3 者に評価させた．また，ユーザー入力として基本周波数や MFCC などのさらに複雑な非言語音響情報の利用についての検討も行う．

2. 非言語音響情報を利用した統計的対話システム

2.1 システム入出力と対話モデル

本研究ではユーザーの発話音声データから発話区間中の平均発話音量，発話長，有音率の 3 種類の簡易的な音響特徴を抽出し，これを対話システムの入力とする．言語情報は利用しない．有音率とは発話区間中の無音時間に対する有音時間の比率である．システムでは，これら 3 種類の非言語音響情報を測定し，学習データから推定した平均値と標準偏差に基づいて閾値を設定し，大，中，小の 3 段階に分類する．

ユーザー発話の入力が完了した後，システムは現状態から次状態へと状態遷移を行う．システムが遷移できる状態としては表 1 の 5 状態である．状態遷移が完了したとき，システムは遷移先の状態に関連した発話を生成する．各状態における発話例を表 1 に示す．システムがどの状態に遷移するかは文献[5]に示される Utility rule によって決定する．

表 1 システム状態と各状態における発話例

システム状態	発話例
話題転換	話題を変えましょう。最近はまっている趣味は何ですか？
質問	その趣味はいつごろ好きになりましたか？
傾聴	うんうん、それで？
自己開示	私の趣味は自作パソコンです
共感	やっぱりそうですよね

2.2 学習方法

Utility rule は対話コーパスから最尤推定することによって次状態の選択規則を対話コーパスに近づけるように学習することができる．本研究ではシステムとの対話のターンごとにユーザーが応答の自然性を評価し，その評価値が高い対話パターンがより多く出現するような対話コーパスを生成し，それを対話モデルの学習データとすることでユーザーが自然と感じる応答選択規則を学習させる．

学習の流れとしては，まずターンごとにシステム応答の自然性に対する 5 段階評価値(1～5)を記録する対話実験を学習者ごとに 30 ターン×3 回行う (ステップ 1)，その後，複数の学習者による評価値の総和が高い対話パターンがより高い確率となるような対話パターンの確率分布を推定する(ステップ 2)．最後に，この確率分布に従って対話コーパスを生成し，単一の対話モデルを再学習する(ステップ 3)．ステップ 3 が終了するとステップ 1 に戻り，学習済みの対話モデルを用いて再度対話実験を行い，学習用対話コーパスを更新する．

この一連の流れを 1 つの学習サイクルとし，最終的に複数のユーザーの総合評価値の高い対話がより多く出現する対話モデルの作成が可能となる．本研究では学習用話者を 20 代の男子学生 5 名とし，ランダムな応答を返す初期対話モデルから評価値が十分に高まったと判断できた第 4 サイクルまで学習を行った．

3. 評価実験

3.1 実験条件

本研究では提案手法の有用性を示すために①その都度人間が判断した最適応答を返す Wizard 手法，②非言語音響情報から統計的対話制御を行った本研究の提案手法，③ランダムに応答を返すランダム手法，それぞれの手法に基づく 3 種類の対話システムを用意し，比較評価を行う．具体的には，それぞれのシステムに対し 6 人の対話者が 30 ターンの対話を行い，その様子を録音した(3 システム×対話者 6 人×30 ターン)．なお，対話者の中の 3 人は 2.2 で述べた対話モデルの学習を行った対話者の中から無作為に抽出した 3 人とし，後の 3 人はその学習者とは別人である．

評価時には，客観的に対話の自然さを評価するために対話者と重複しない 3 名の評価者に対話の録音データを聞いてもらうことで評価を行った．まず 1 名分の対話データ(3 システム分)に対し，3 名の評価者に対話の自然さについて 1～5 段階で評価をつけてもらった(3 システム×評価者 3 名=9 個の評価値)．その後システムごとに評価値の平均値をとり，それをその対話者におけるシステムごとの評価値と

†1 静岡大学大学院 総合科学技術研究科 情報学専攻
†2 日本アイ・ビー・エム（株） 東京基礎研究所

した．さらにこの評価方法を残りの 5 名分の対話録音データに対して行うことで，6 人の対話者それぞれにおける 3 つのシステムの評価値を算出した(対話者 6 人×3 システム=18 個の評価値)．なお，評価者は 3 システムでそれぞれ異なる．

3.2 実験結果

対話者 6 人の客観評価結果と 6 人の平均値を図 1 に示す．学習対話者及び学習外対話者それぞれ 3 名に対する平均評価値は多くの対話者において Wizard 手法，提案手法，ランダム手法の順に高い値となった．なお，学習者と同一である対話者と学習者と異なる対話者間の評価結果において，特に有意な差は見られなかった．また，対話者 6 人の平均評価値を見ると，提案手法は人間が選択した Wizard 手法に対し 1.1 の低下にとどまった一方、ランダム手法に対し 0.7 の有意を示すことができた．よって，本研究の提案手法である簡易的な音響情報を用いた統計的対話システムは複数の学習者を用いた場合においても，ある程度自然と感じてもらえる対話制御が可能ということを示すことができた．一方，対話者 3 のように提案手法がランダム手法よりも評価値が低くなってしまうという場合も存在した．

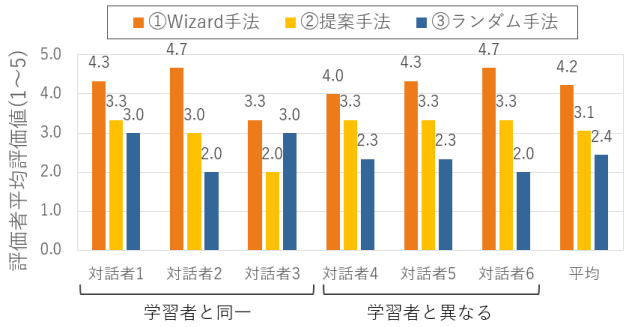

図 1 対話者別平均評価値

4. 他の音響情報の利用

対話者 3 において提案手法がランダム手法よりも評価値が低くなった原因として，ユーザーはシステムが不自然な応答をしたとしてもある程度我慢して発話するという性質が観測できた．本研究で利用している簡易的な音響情報ではそのようなユーザーの心的状態を検知できず，システムはユーザーにとって不自然な発話をし続けてしまうため，自然性に対する評価値が低下してしまったと考えられる．

そのようなユーザーの心的状態を検知して対話戦略を変えるためには，より複雑な音響情報を利用する必要がある．複雑な音響情報の例としては感情認識に用いられる特徴量が考えられる[6]．ここでは対話制御に利用するために INTERSPEECH 2009 Emotion Challenge にて用いられた 5 つの特徴量及びその 12 個の素性値に加え，2.1 で示した 3 つの簡易的音響情報を用いて発話のクラスタリングを行うことを試みる．

クラスタリングに用いた音声データは 3.1 にて収録した聞き役対話のうち，話者 1 人の発話音声に後から収録した疑問文発話を加えた 366 発話を使用し，クラスタリングアルゴリズムは EM アルゴリズムとした．クラスタリング結果を表 2 に示す．

表 2 聞き役対話音声クラスタリング結果

クラスタ番号	データの割合[%]	クラスタ詳細
Cluster1	15	疑問文
Cluster2	25	ポジティブ回答
Cluster3	11	短い共感、同調
Cluster4	1	認識ミスによる再発話
Cluster5	7	長文語り
Cluster6	23	平常発話
Cluster7	18	ネガティブ回答

クラスタ詳細は分類が完了したあとの音声を聞き，どのような発話がそのクラスタに集まっているかを示したものである．クラスタリング結果から直接にユーザーの心的状態は取得できなかったが，同じクラスタに分類される発話が連続で続いたなどの対話中にターンごとにユーザー発話を分類したクラスタの組み合わせによって，より正確にユーザーの状態を検知できる可能性がある．POMDP などの統計的手法の場合，ユーザーの状態は直接観測できなくとも前後の観測結果から正確なユーザー状態把握が可能となるため，今後はこのようなクラスタリング結果を対話マネジメントに適用する予定である．

5. おわりに

本研究では簡易的な音響情報を用いた統計的対話システムにおいて複数の学習者で学習を行ったモデルを適用した場合の評価を行った．また，聞き役雑談対話の発話を感情認識に使用する特徴量を用いてクラスタリングし，ユーザーの心的状態を考慮した対話システムの可能性について検討した．

謝辞

本研究の一部は JSPS 科研費 16K01543 の助成を受けたものである．

参考文献

[1] 中島悠, 梅井良太, 伊東伸泰, 西田昌史, 西村雅史, ”回想法を模擬した高齢者向け対話システムの構築に関する研究”, 情報処理学会第 78 回全国大会, (2016).

[2] 横山 祥恵, 山本大介, 小林優佳, 土井美和子, “高齢者対話インターフェース-雑談継続を目的とした話題提示・傾聴の切り替え式対話法-“, 情報処理学会研究報告書, pp. 1-6, (2010).

[3] Tang Ba Nhat, 目良和也, 黒澤義明, 竹澤寿幸, ”音声に含まれる感情を考慮した自然言語対話システム”, HAI シンポジウム, pp. 87-91 (2014).

[4] 梅井良太, 中島悠, 伊東伸泰, 西田昌史, 西村雅史, ”非言語音響情報を利用した聞き役対話システムに関する検討”, 情報処理学会第 78 回全国大会, (2016)

[5] Pierre Lison, “A hybrid approach to dialogue management based on probabilistic rules”, *Computer Speech and Language* 34, pp. 232-255, (2015).

[6] Bjorn Schuller, Stefan Steidl, Anton Batliner, ”The INTERSPEECH 2009 Emotion Challenge”, pp. 312-315, INTERSPEECH (2009).