人工知能学会研究会資料
SIG-SLUD-B403-02

音声対話システムにおける発話の誤分割修復要否判定のユーザ適応

User Adaptive Classification of Restoration Necessity for Incorrectly-Segmented Utterances in Spoken Dialogue System

堀田尚希[1*] 駒谷和範[2] 佐藤理史[1] 中野幹生[3]
Naoki Hotta[1] Kazunori Komatani[2] Satoshi Sato[1] Mikio Nakano[3]

[1] 名古屋大学大学院
[1] Graduate School of Engineering, Nagoya University
[2] 大阪大学産業科学研究所
[2] The Institute of Scientific and Industrial Research, Osaka University
[3] ホンダ・リサーチ・インスティチュート・ジャパン
[3] Honda Research Institute Japan Co., Ltd.

Abstract: A spoken dialogue system should respond quickly after a user finishes speaking, but this often causes incorrect segmentation of user utterances by erroneous voice activity detection. We previously developed a method that performs a posteriori restoration for the incorrectly segmented utterances. A crucial part of the method is to classify whether the restoration is required or not. In this paper, we improve the accuracy by adapting the classification to each user. We focus on speaking tempo of each user, which can be obtained during dialogues. We reveal a correlation between each user's tempos and their appropriate thresholds used in the classification. We then derive a linear regression function that converts the tempos into the thresholds. We adapt two classifiers: that simply using a threshold and decision tree learning. Experimental results showed the proposed user adaptation for the two classifiers improved the classification accuracies by 3.3% and 2.1%.

1 はじめに

音声対話システムでは “何を話すか” という応答内容の正しさとともに，“いつ話すか” という応答タイミングの適切さを考慮する必要がある．一般的に音声対話システムでは，ユーザの発話に対してシステムは素早く応答することが好ましい．Ward らは，音声対話システムにおいてユーザがストレスを感じる原因として，音声認識・理解の誤りとともに，応答までに要する時間を挙げている [8]．つまり，応答が遅い音声対話システムは，ユーザにストレスを与える可能性がある．

一方で音声対話システムで素早い応答を実現しようとする場合，ユーザの発話中にシステムが話し始めることがある．これはシステムが，ユーザ発話中の短い無音区間を発話終了であると誤って認定するからである．このときシステム内部では，元来一発話であったユーザ発話が，無音区間により複数の発話区間に分割されている．本研究ではこの現象を発話の誤分割と呼ぶ．

これまで我々は，この誤分割により生じる 2 つの問題を，事後的に修復するシステムを提案してきた [3]．

1. ユーザの発話中にシステムが話し始める問題
   → そのシステム発話を停止するルールを追加
2. 誤った発話区間に対して音声認識がなされる問題
   → 発話断片を結合して再度音声認識を実行

本稿では誤分割の修復が必要か否かの判定精度を向上させるため，修復要否判定におけるパラメータを，ユーザに応じて適応させることを考える．修復要否判定では，発話断片間の時間間隔に対する閾値が，重要なパラメータである [3]．この閾値を，それまでのユーザの振る舞いに応じて変更する．

本稿ではユーザが発話をするテンポに着目し，これを利用して修復要否判定のユーザ適応を行う．まず，本研究で特徴として使用する発話テンポを定義する．次

*連絡先：名古屋大学大学院工学研究科
〒 464-8603 愛知県名古屋市千種区不老町 C3-1(631) 名古屋大学大学院工学研究科電子情報システム専攻 佐藤・松崎研究室
E-mail: n_hotta@nuee.nagoya-u.ac.jp

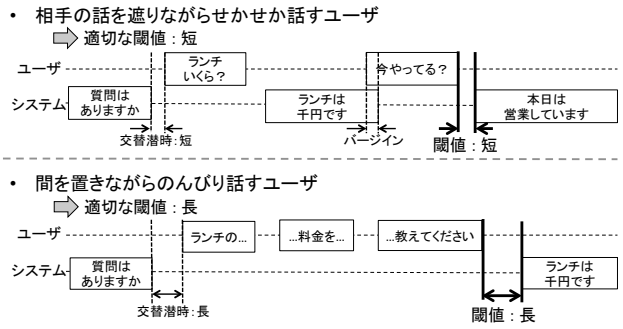

図 1: ユーザの発話交替潜時と適切な閾値

に，発話テンポと修復要否判定における適切な発話間間隔の閾値との相関を調べ，線形回帰式を得る．その後，この線形回帰式を用いた閾値処理を行う場合と，線形回帰式を用いてパラメータを変更して決定木学習を行う場合の，2 つの方法で修復要否判定を行い，その精度を調べる．

2　発話テンポに基づくユーザ適応

本稿では，修復必要性の判定において重要な特徴である発話間間隔に着目し，この閾値をユーザに応じて変更することで，修復要否判定精度の向上を図る．音声対話システムがユーザに適応して発話を行うことは，システムの性能向上に寄与する [5] だけではなく，対話の初期段階からユーザとの同調傾向が見られやすい [6] という利点もある．

音声対話システムでは，様々な話し方をするユーザの利用が想定される．相手の話を遮りながら急いで話すユーザの対話例を図 1 の上部に示す．このようなユーザはシステムの発話終了後，素早く話し始めることが多い．さらに，相手の発話中に話し始めること (バージイン) も，相手の話を遮りながら急いで話すユーザによく見られる特徴である．一方で，間を置きながらのんびり話すユーザの対話例を図 1 の下段に示す．このようなユーザは，システムの発話終了後，少し待ってから話を始める傾向がある．

我々は，適切な閾値は，それぞれのユーザの話し方に依存すると考える．相手の話を遮りながら急いで話すユーザは，発話中の言い淀みが少なく，発話中のポーズも短いと予想される．このため発話間間隔の閾値を短く設定し，不要な修復処理やそれに伴う応答の遅延を避ける．またこのようなユーザは，システムが素早く応答しない場合，発話が受理されなかったと思い，同じ内容を再度発話をする可能性がある．これは発話の衝突を防ぐという面やユーザインタフェースの観点から，避けられるのが望ましい [1]．一方で間を置きながらのんびり話すユーザは，発話中に長いポーズがある可能性がある．そのため，発話間間隔の閾値を長く設定し，発話断片間の間隔が長い場合でも発話断片対を修復するのが望ましい．

3　発話テンポと適切な閾値の関係

本章では，ユーザの話し方を定量的に表すパラメータとして，発話テンポを定義し，修復が必要とみなす適切な閾値の関係性を調査する．発話テンポと適切な閾値に関連があるならば，ユーザとの対話から得られる発話テンポから，修復が必要とみなすための適切な閾値を推測することができる．

3.1　発話テンポの定義

ユーザの話し方を定量的に表すパラメータとして，本研究では発話テンポを定義する．本稿ではあるユーザの発話テンポを，当該ユーザの対話データにおける，システムの発話終了からユーザの発話開始までの時間間隔 (本研究ではユーザの発話交替潜時と呼ぶ) の平均とする．これに加えて，システムの発話中にユーザが話し始めた場合 (バージイン) 時に対しても，交替潜時を定義する．ここでは交替潜時は負の値とし，その絶対値はユーザの発話開始からシステムの発話終了までの時間間隔 (システムとユーザが同時に話していた時間) とする．

3.2　修復要否判定の閾値

発話テンポと適切な発話間間隔の閾値の関係性を調査するため，ユーザ毎に適切な閾値を決定する．本稿では適切な発話間間隔の閾値を，修復が必要か否かを高精度に判定可能な閾値とする．すなわち，ユーザ毎に修復が必要が否かを，発話断片対と正解ラベルを入力とし，発話間間隔の特徴のみを用いて，サポートベクタマシン (Support Vector Machine (SVM)) で判定したときに，閾値となる値を適切な発話間間隔の閾値とする．この判定には機械学習ソフトウェア Weka(version 3.6.9) の SMO を使用した．SVM を用いる理由は，マージン最大化を行う機械学習手法であるからである．SVM ではクラス間での距離が最大となる位置に識別境界を設定する．これは適切な閾値を決定する上で好ましい性質である．

学習データに十分な正例または負例がなく，適切な発話間間隔の閾値が導出できないユーザは，固定の閾値を使用した．本稿では，全ての発話が “修復が必要である” 場合は，適切な発話間間隔の閾値は 2.00[秒] と

し，全ての発話が “修復が必要でない” 場合は，閾値は 0.00[秒] とした．

3.3 対象データ

本研究では世界遺産検索システム [7] により収集されたデータを調査対象とした．世界遺産検索システムでは 35 名のユーザの音声データが，対話ログとともに保存されたほか，4 名のユーザの音声データが保存された (対話ログは保存されなかった)．各ユーザは約 8 分の対話を 4 回ずつ行った．対話の仕方に関しての教示はなく，ユーザに自由に対話をしてもらうことで，音声データは収集された [7]．

本稿では 26 名のユーザのみを調査対象とした．ユーザ適応を行うためには，十分な数の発話断片の対が保存されている必要があるためである．具体的には，発話間間隔が近接しており (2.00 秒未満)，雑音ではない発話断片対 (各発話断片が 0.80 秒以上) が 6 対以上あるユーザのみを調査対象とした．我々は，元来一発話である可能性がある (修復が必要である可能性がある) 発話断片の対を調査対象としており，発話間間隔が 2.00 秒以上である発話は，元来一発話である可能性はないと判断したからである [3]．また，対話ログが保存されていない 4 名のユーザは，調査対象から除外した．

調査対象とする 26 名のユーザからは，全体で 3099 個のユーザ発話，390 対の発話断片対が得られた．これら 390 対の発話断片対には，文献 [4] と同様に，元来一発話であるか否かのラベルを付与した．元来一発話である発話断片対は修復の必要があり，それ以外の発話断片対は修復の必要はない．3099 個のユーザ発話は，発話テンポの導出に使用した．ただし，ユーザの発話交替潜時が –3.5 秒以上，6 秒未満の発話のみを調査対象とした．交替潜時の正の絶対値が大きい場合は，ユーザが長考したためと考えられ，一定以上の値に大きな意味はない．交替潜時の負の絶対値が大きい場合は，システムの発話停止が遅れたというシステムの遅延によるものであると考えられる．この値は実験的に決定した．

3.4 相関の調査

調査対象とする 26 名のユーザそれぞれに対して，発話テンポと適切な発話間間隔の閾値を調査した．これらの関係をプロットしたものを図 2 に示す．縦軸は適切な発話間間隔の閾値 [秒] を示し，横軸はシステムの発話終了からユーザの発話開始までの時間間隔の平均 [秒]，すなわち発話テンポを示す．

図 2 においてユーザの発話テンポと適切な発話間間隔の閾値の相関係数を求めたところ，0.63 であった．これは，Guilford の規則 [2] に基づけば，中程度の相関があることが示される．さらに，図 2 からユーザの発話テンポと適切な発話間間隔の閾値の線形回帰式として，式 1 を得た．

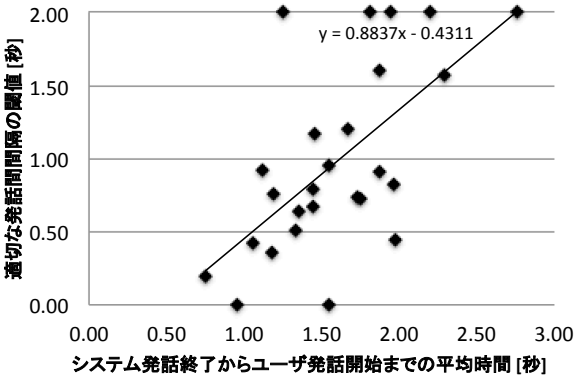

図 2: ユーザの発話テンポと適切な発話間間隔の相関

$$\text{閾値 [秒]} = 0.884 * \text{発話テンポ [秒]} - 0.431 \qquad (1)$$

この線形回帰式を用いることで，対話から得られる特徴である発話テンポから，未知の特徴である適切な閾値を推測することができる．

4 修復要否判定の適応

3 章で示した発話テンポと閾値の相関が，これを用いたユーザ適応により，修復要否判定の精度向上に寄与するかどうかを検証する．ユーザ適応の概要を図 3 に示す．本手法の目的は，ある発話断片の対を，2 つの発話として別々に解釈する（修復不要）か，または 1 つの発話として統合して解釈する（修復必要）かを判定することである．ここではまず，3 章で得られた線形回帰式を用いて，ユーザの発話テンポから，発話間間隔の閾値を適応させる．これによりユーザに適応した修復要否判定を行う．本章では，2 種類の修復要否判定手法におけるユーザ適応の方法を示す．

4.1 閾値処理の場合

発話間間隔のみを用いた最も単純な手法として，閾値処理による修復要否判定を考える．まず最も単純な場合で，ユーザ適応の有効性を確認する．

閾値のユーザ適応の処理の流れを図 4 に示す．ユーザ発話の断片対（その発話間間隔）を入力とし，3 章で得られた線形回帰式を利用して，修復要否判定を行う．具体的にはそれまでの対話履歴から発話テンポを

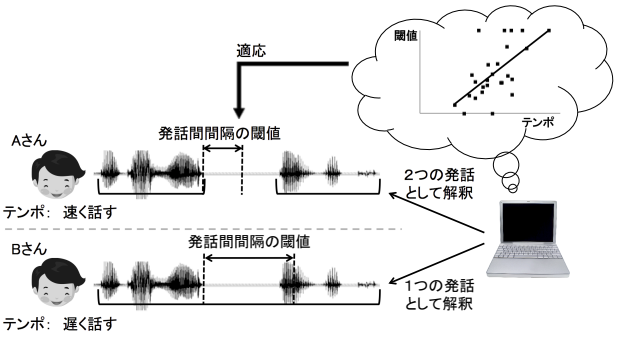

図 3: 発話テンポに基づく修復要否判定のユーザ適応の概要

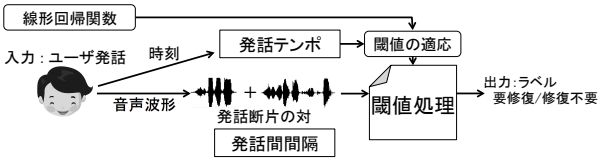

図 4: 閾値処理におけるユーザ適応

計算し，線形回帰式を用いて，それに対応する閾値を得る．続いて，この閾値を用いた修復要否判定により，修復が必要であるか否かのラベルを出力する．具体的には，入力された発話断片対の発話間間隔が，線形回帰式により求めた適応後の閾値よりも小さい場合には修復が必要であるとし，大きい場合には不要であると判定する．

4.2 決定木学習の場合

閾値処理よりも高精度な修復要否判定を行う手法として，決定木学習を考える．この場合においても，ユーザ適応が有効に作用することを示す．

決定木に対するユーザ適応の処理の概要を図 5 に示す．ここでは，文献 [4] で有効とされた，発話間間隔と，4 つの特徴（前半断片の平均の音声認識信頼度，GMM を用いた雑音判別結果，前半断片全体の F0 レンジ，前半断片の最大音量）を用いる．つまり入力は，発話断片の対（その発話間間隔）とそれらから得られる 4 つの特徴量であり，これらを決定木に入力することにより，修復の要否を判定する．

本稿でのユーザ適応は，これらの特徴のうち，発話間間隔のみを変換することで実施する．前節の閾値処理の場合と同様に，対話のその時点までに得られたユーザの発話テンポに基づき，3 章で得られた線形回帰式を用いて，特徴量を線形変換する．

この変換は，学習時と実行時の両方で実施する．決定木学習は全てのユーザのデータを用いて行うため，決

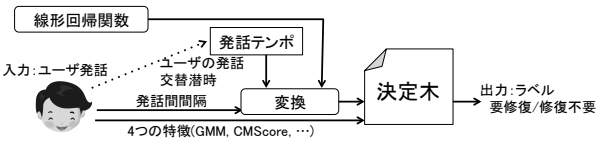

図 5: 決定木の特徴量のユーザ適応

定木の中で設定される発話間間隔の閾値は全ユーザで共通である．このため，閾値を変化させるのではなく，学習時と実行時の両方で，特徴量自体にユーザごとに同じ変換を加えることで，ユーザに応じた判定を行う．

この線形変換には，当該ユーザの発話テンポから得られる閾値と，全ユーザの発話間間隔の閾値の平均との比を用いる．これにより，ユーザごとに偏りがある発話間間隔の特徴量を平均化し，単一の閾値で，ユーザに適応した判定の実現を狙う．具体的には，ユーザの発話断片から得られた発話間間隔の特徴量を I_0，適応後の特徴量を I として，式 2 で線形変換を行う．

$$I = I_0 \times \frac{T_0}{T} \tag{2}$$

T はユーザの発話テンポに対応した発話間間隔の閾値である．T_0 は定数であり，調査対象とする 390 対の発話断片対の発話間間隔の平均値である，0.519[秒] とした．

この変換の狙いを，具体的な傾向を用いて示す．図 2 で示されている傾向として，発話テンポが速いユーザに対しては閾値を小さくするのが望ましい．ここでの変換は，このようなユーザに対する発話間間隔の特徴量を，デフォルトの値よりも相対的に大きくするものである．つまり，学習する決定木の閾値が全ユーザで共通であるため，閾値を小さく設定する代わりに，相対的に特徴量を大きくしている．実際に，発話テンポに対応する閾値 T が小さいユーザは，この比 $\frac{T_0}{T}$ が大きくなり，適応後の特徴量 I も大きくなるように変換が行われる．

5 実験的検証

4 章で示した閾値処理と決定木学習のそれぞれに対して，修復要否判定の精度を調査する．これにより，3 章で示した発話テンポと適切な閾値の相関が，判定精度向上に実際に寄与するかどうかを検証する．さらに，逐次的に適応を行うのではなく，ユーザのデータが一括で得られるとした場合（バッチ適応）との性能比較を行うことで，適応による性能の上限や，適応の収束速度について考察する．

表 1: 交差検定における閾値処理の判定精度

	closed	交差検定
ユーザ適応前	281/390 (72.1%)	281/390 (72.1%)
ユーザ適応後	294/390 (75.4%)	293/390 (75.1%)

表 2: 交差検定における決定木の判定精度

	closed	交差検定
ユーザ適応前	312/390 (80.0%)	271/390 (69.5%)
ユーザ適応後	320/390 (82.1%)	300/390 (76.9%)

5.1 ユーザ適応による判定精度

提案するユーザ適応が判定精度の向上に寄与することを検証する．検証は，閾値処理と決定木学習のそれぞれについて行い，また closed test と交差検定のそれぞれの場合によって行う．closed test では，適応と評価を同一のデータを用いて行った．交差検定では，ユーザ単位の 1 個抜き交差検定 (leave-one-out cross-validation) を実施した．具体的には，閾値処理では，26 名のユーザのうち 1 名のユーザを除いた 25 名のデータを用いて，線形回帰式を導出し，残りの 1 名のユーザのデータで評価を行った．決定木学習では，線形回帰式（式 1）を用いた変換式は既知であるものとし，全てのユーザの発話間間隔の特徴量を変換した後，26 名のユーザのうち 25 名のデータを用いて決定木の構築を行い，残りの 1 名のデータで評価を行った．決定木学習で線形回帰式を既知としたのは，実験の簡便さのためであるが，後述するように，この回帰式のパラメータ数は 2 と少ないため，交差検定を実施した場合でも値はほぼ変わらないことを確認しており，問題はない．

閾値処理における判定精度を表 1 に示す．ユーザ適応前は，全ユーザに対して単一の閾値 (0.822[秒]) により判定した場合である．この閾値は，全てのユーザに対して，元来一発話か否かを，発話間間隔の特徴のみを用いて，Weka の SMO により判定した場合の閾値である．ユーザ適応により，closed の場合では 3.3 ポイント，交差検定の場合では 3.0 ポイントの判定精度の向上を得た．これにより，閾値処理ではユーザ適応が判定精度の向上に寄与したといえる．

次に，決定木学習における判定精度を表 2 に示す．ユーザ適応前は，4.2 節の式 2 で示す線形変換を行わずに決定木を構築した場合の結果を示す．この場合もユーザ適応により，closed の場合では 2.1 ポイント，交差検定の場合では 7.4 ポイントの判定精度の向上を得た．これにより，決定木学習の場合でも，ユーザ適応が判定精度の向上に寄与したといえる．

閾値処理と決定木学習の双方に対して，closed test と交差検定の結果を比較することで，過学習の有無について考察する．閾値処理では，ユーザ適応前と後の両方で，closed の場合と交差検定の場合とで，精度はほぼ変わらない．これは判定器が学習データには依存しておらず，つまり過学習は起こっておらず，未知のユーザにも対応できる可能性が高いことを示している．その要因として，閾値処理手法では，学習すべきパラメータの数が少ないことが考えられる．閾値処理において学習すべきパラメータは，線形回帰式 $y = ax + b$ における a と b のみである．交差検定の場合のパラメータ a と b の値を表 3 に示す．ここでは a と b の両方で，平均値は 3.4 節における式 1 の係数とほぼ一致し，大きく変化することはなかった．その結果，closed の場合とほぼ同一の結果が得られたと考えられる．

表 3: 交差検定時のパラメータ

パラメータ	a	b
平均値	0.883	−0.431
標準偏差	0.034	0.057

一方で決定木学習では，ユーザ適応前において，closed の場合では閾値処理の精度を上回ったのに対し，交差検定では閾値処理の精度を下回った．決定木学習では学習すべきパラメータが多く，学習データ中の個々のユーザのふるまいに過度に適応した，つまり過学習した決定木が構築されたためと考えられる．このため，closed の場合における精度は不当に高かったとみなせる．

決定木学習でユーザ適応を行った場合，closed と交差検定の両方の場合において，決定木学習の判定精度は閾値処理を上回った．これは，ユーザ適応を行うことにより，閾値処理から決定木学習へとモデルを複雑にしても，より一般的な判定器が学習され，過学習の発生が抑制されていることを示唆している．

学習により得られた決定木のうち，高さが 4 以下の部分を図 6 に示す．この決定木では高さ 1 の部分に，ユーザ適応後の発話間間隔が使用されていた．このことから，ユーザ適応後の発話間間隔の特徴は，確かに有用であったことが示されている．

5.2 バッチ適応との比較

前節までの実験結果は全て，対話開始から判定時点の発話までの，ユーザごとの発話テンポを使用している．本稿ではこれをオンライン適応と呼ぶ．

本節ではこれに対し，各ユーザの，対話全体における発話テンポを用いて，適応を行う場合を考える．本稿ではこれをバッチ適応と呼ぶ．バッチ適応では，対象ユーザの対話データが全て事前に得られていることを仮定しており，ユーザの特性が既に十分に得られて

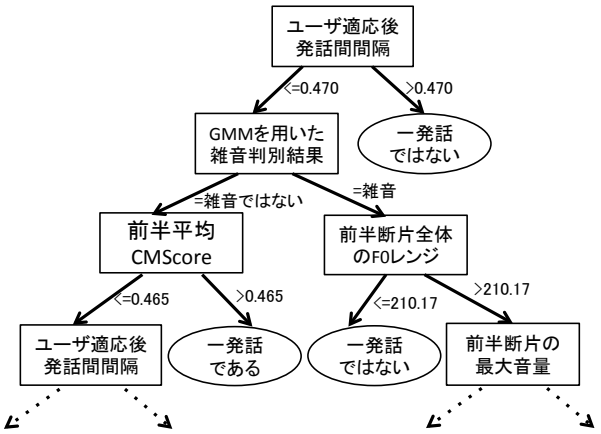

図 6: 構築された決定木（高さ 4 以下の部分）

表 4: オンライン適応とバッチ適応による判定精度

	閾値処理	決定木学習
ユーザ適応前	281/390 (72.1%)	312/390 (80.0%)
オンライン適応	294/390 (75.4%)	320/390 (82.1%)
バッチ適応	306/390 (78.5%)	331/390 (84.9%)

いる場合に相当する．これを用いて，ユーザ適応による性能向上の上限について考察する．

バッチ適応とオンライン適応の場合の判定精度を，表 4 に示す．閾値処理の場合と，決定木学習の場合のそれぞれで結果を示している．ただし，決定木学習におけるオンライン適応は，バッチ適応のデータを用いて決定木を構築し，オンライン適応のデータを用いてテストを行った場合である．閾値処理・決定木学習ともに，バッチ適応では，オンライン適応に比べ，それぞれ 3.1 ポイント，2.8 ポイント，高い判定精度が得られた．つまり，オンライン適応では，各ユーザのデータが十分に得られていない状況において，判定精度が不安定である可能性が示唆されている．

5.3 適応の収束速度の分析

オンライン適応とバッチ適応の判定精度の違いに関する詳細を，閾値処理の場合において調査する．バッチ適応は，全ての発話が事前に得られていると仮定しているため，利用可能な発話数が増えると，オンライン適応はバッチ適応に収束すると予想できる．利用可能な発話数の増加に応じたオンライン適応の精度について分析し，考察を行う．

オンライン適応に利用可能なデータ数を変化させたときの，修復要否判定の正解数を図 7 に示す．ここでは，閾値処理において，発話テンポを，対話開始から x

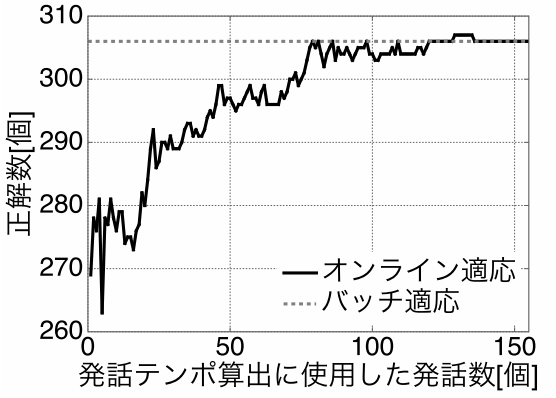

図 7: 閾値処理における既知の発話数と判定精度

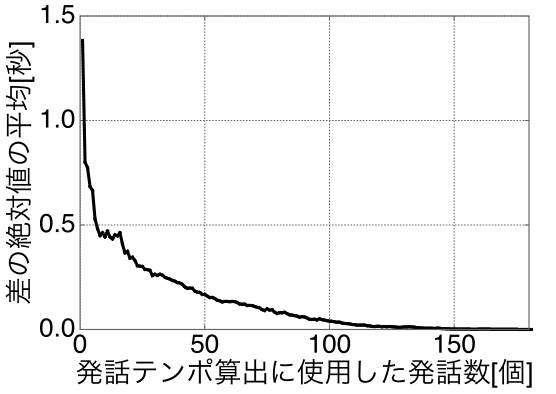

図 8: オンライン適応とバッチ適応の閾値の差

発話目までを使用して計算した場合の正解数を示している．縦軸は正解数，横軸は使用した発話数である．図上部の点線は，バッチ適応の場合の正解数（$y = 306$）を示している．オンライン適応の場合，既知の発話数が 10 発話未満と少ない時には，正解数は大きく変動し，また正解数自体も少ない（275 発話程度）．一方，既知の発話が増えるにつれて正解数は増加し，既知の発話数が 80 発話程度になると，バッチ適応とほぼ同等の正解数となっている（305 発話程度）．このことから，既知の発話が多くなるほど，オンライン適応の判定精度はバッチ適応に近くなることと，既知の発話数が 80 発話程度でバッチ適応と同等になったことを確認した．

オンライン適応の判定精度がバッチ適応に近づくことは，既知の発話が多くなるほど，発話テンポがバッチ適応における値に近づくことからも確認できる．これを図 8 に示す．図 8 は，対話開始から x 話目までを用いて発話テンポを計算した場合に得られる閾値と，バッチ適応の場合の閾値との差の絶対値について，26 名のユーザの平均を取ったものである．すなわち，ユーザ i について，対話開始から x 発話目までを用いて発話

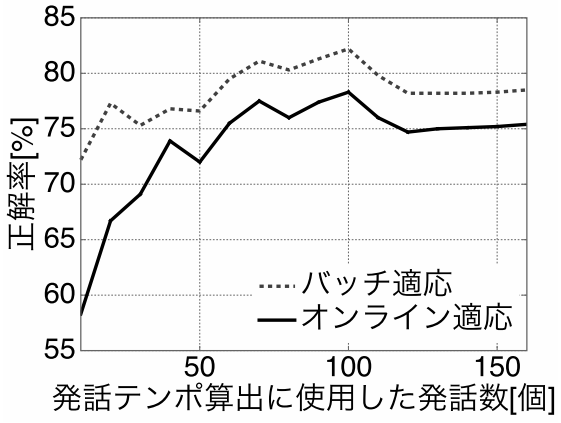

図 9: オンライン適応における既知の発話数と正解数

テンポを計算した場合の閾値を T_x^i，ユーザ i におけるバッチ適応の閾値を T_0^i とすると，x 発話目のとき，以下の式 3 で求められる．

$$\text{バッチ適応との差の絶対値の平均} = \frac{\sum_{i=1}^{26} |T_x^i - T_0^i|}{26} \quad (3)$$

既知の発話数が少ないときは，バッチ適応との閾値の差が，平均で 1 秒以上存在していた．既知の発話が増加すると，閾値の差は大きく減少し，100 発話程度では 0.05 秒程度まで減少した．このことから，既知の発話が多くなるほど，閾値がバッチ適応における値に近づくことが示されている．閾値がバッチ適応における値に近づけば，その修復要否判定の結果もバッチ適応に近づく．

既知の発話が多くなるほど，オンライン適応の判定精度はバッチ適応に近くなることは，全体の正解率の推移からも説明できる．図 9 はオンライン適応における，既知の発話数と正解率の関係を示している．縦軸は誤分割修復の判定精度を示し，横軸は既知の発話数の上限を示す．上側の点線はバッチ適応の場合を示しており，下側の実線はオンライン適応の場合を示している．このグラフは例えば，判定対象の発話断片対より前に存在する発話が，50 発話以下であった場合のみに限定すると，オンライン適応における判定精度は約 72%であり，バッチ適応における判定精度は約 77%であったことを示している．既知の発話数が少ない場合は，バッチ適応とオンライン適応の正解率の差は大きいが，既知の発話数が増えるにしたがってその差は小さくなる．このことから，既知の発話数，つまりオンライン適応に使用可能な発話数が少ない間はオンライン適応とバッチ適応の性能差は大きく，既知の発話数が多くなれば，オンライン適応の精度はバッチ適応に近づくことが示されている．

6 結論

本研究では，修復が必要か否かはユーザの話し方の違いに依存することに着目し，ユーザに適応した修復必要性の判定を行った．修復が必要か否かの閾値とする発話間間隔と，ユーザの発話テンポには中程度の相関があることを実験的に示し，それらの線形回帰式を得た．その後，閾値処理の場合と決定木学習を行う場合において，対話開始から現在までに得られた特徴を用いてユーザ適応を行った場合 (オンライン適応) の判定精度を調べた．その結果，閾値処理と決定木学習の両方において，単一の閾値を用いるベースラインに比べ，判定精度が向上することを確認した．

今後，修復が必要か否かの判定精度をさらに向上させる方法として，以下の 2 点が考えられる．

1. ユーザの話し方を表す特徴の追加

   ユーザの話し方を表す特徴は，本研究で使用した発話テンポ以外にも存在する．例えばユーザの発話速度や，言い間違いの回数などが考えられる．これらの特徴を併用することで，精度が向上する可能性がある．

2. 発話間間隔以外の特徴のユーザ適応

   ユーザによってその傾向が異なる特徴は，発話間間隔以外にも存在する．例えば，文献 [4] では発話断片の音量を特徴として用いていたが，ユーザの声の大きさは，ユーザ毎に違うと考えられる．またユーザによっては，対話の最後を上げ調子で話すなど，話し方に癖がある場合もある．このような特徴を用いてユーザ適応を行うことで，精度が向上する可能性がある．

謝辞

本研究の一部は，カシオ科学振興財団の支援を受けた．

参考文献

[1] K. Funakoshi, M. Nakano, K. Kobayashi, T. Komatsu, and S. Yamada. Non-humanlike Spoken Dialogue: A Design Perspective. In *Proc. SIGDIAL*, pp. 1391–1394, 2010.

[2] J.P. Guilford. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1956.

[3] 堀田尚希, 駒谷和範, 佐藤理史. ユーザ発話の誤分割に起因する問題を事後的に修復する音声対話システム. 情報処理学会研究報告 2013-SLP96-5, pp. 1–8, 2013.

[4] N. Hotta, K. Komatani, S. Sato, and M. Nakano. Detecting incorrectly-segmented utterances for posteriori restoration of turn-taking and ASR results. In *Proc. INTERSPEECH*, pp. 313–317, 2014.

[5] K. Komatani, S. Ueno, T. Kawahara, and H. G. Okuno. User modeling in spoken dialogue systems for flexible guidance generation. In *Proc. EUROSPEECH*, pp. 745–748, 2003.

[6] 大石周平, 尾田政臣. 話者間の精神テンポの差がコミュニケーションの円滑化に及ぼす影響. 電子情報通信学会技術報告 2006-HIP105-18, pp. 31–36, 2006.

[7] 佐藤隼, 中野幹生, 駒谷和範, 船越孝太郎, 奥乃博. ドメイン外発話が扱え拡張性が高い対話ドメイン選択フレームワーク. 情報処理学会研究報告 2011-SLP86-12, pp. 1–8, 2011.

[8] N. G. Ward, A. G. Rivera, K. Ward, and D. G. Novick. Root Causes of Lost Time and User Stress in a Simple Dialogue System. In *Proc. INTERSPEECH*, pp. 1565–1568, 2005.