社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－NL－155 (17)
2003－SLP－ 46 (6)
2003／5／27

対話音声中の言い直し発話の検出と認識

北岡 教英† 角谷 直子† 中川 聖一†

† 豊橋技術科学大学 〒441-8580 愛知県豊橋市天伯町雲雀ケ丘 1-1
E-mail: †{kitaoka,naoko,nakagawa}@slp.ics.tut.ac.jp

あらまし コンピュータと人間が音声を通じてコミュニケーションを図る場合，誤認識は避けられない．ここで，誤認識の訂正のために，ユーザが誤認識された部分を言い直した場合に，それが言い直しの音声であるとシステムが判定できれば，誤認識からの回復が容易になると考えられる．本稿では，自然な対話音声中に含まれる繰り返し部分の存在を検出することにより言い直し発話を検出することを試みる．システムとの対話を収集した際に実際に生じた誤認識時の言い直し発話について，DP マッチングの最適パスの線形性判定と DP 距離の閾値処理を用いた方法と，音声認識を行った結果得られる N-best 候補中に含まれる認識単語の重なり度を用いた方法とを組み合わせることで，閾値などの設定に用いないオープンタスクにおいて再現率 91.9%，適合率 87.7%の検出性能を得た．さらに，この検出結果を音声認識の際の語彙・文法の制約に用いることによって対話音声の認識性能を向上させることができた．

キーワード 言い直し発話, 言い直し発話検出, DP マッチング, 認識候補の重なり度

Detection and Recognition of Correction Utterances on Misrecognition of Dialog System

Norihide KITAOKA†, Naoko KAKUTANI†, and Seiichi NAKAGAWA†

† Toyohashi University of Technology, Aichi, 440-8580 Japan
E-mail: †{kitaoka,naoko,nakagawa}@slp.ics.tut.ac.jp

Abstract Misrecognition is inevitable, when we communicate with computers through speech interface. The recovery becomes easy, if the system can detect the user's repetition of misrecognized part. In this report we proposed to detect correction utterances by detecting repetition parts in spontaneously spoken dialog. We used the combination of DTW-based method and N-best list overlapping measure and achieved 91.9% recall rate and 87.7% precision on a test set collected using a real dialog system. We also improved the recognition performances of the utterances by constraining the lexicon and the grammer to be used to recognize the utterances.

Key words correction utterances, detection of correction utterances, DTW, recognition candidate overlapping measure

1. はじめに

コンピュータと人間が音声を通じてコミュニケーションをはかる場合，誤認識は避けられない．しかし，現在は音声インターフェース技術が未熟であるために，誤認識からの回復が困難である．一般に，ユーザはシステムの誤認識に対して，同じ内容の言い直し (繰り返し) で対処しようとすることが多い．すなわち，システムがユーザの言い直しを検出できれば，誤認識からの回復が容易になると考えられる．

現在も言い直し (繰り返し) 音声に関する研究はいくつかなされている．言い直し発話の韻律的な特徴についていくつか研究がされており，繰り返し発話は元のユーザ発話に比べてピッチ・継続時間長が大きく発話速度の低下が見られる [1] [2]，訂正発話では継続時間が長くなるがパワーやピッチについては大きな差は見られない [3]，認識誤りの訂正と棄却誤りの訂正では認識誤り訂正の方が継続時間がより長くなる [4]，訂正連鎖においてエラーからより遠い訂正は近い訂正よりピッチ・パワーが大きく継続

System：目的地を設定してください
User：えーと，静岡県の浜松インターに行きたい
System：静岡県浜松駅に設定します
User(言い直し)：東名高速道路の浜松インターです

図 1　言い直しの例（文発声）

長が長くゆっくりで先行ポーズが長い [5] などの報告がある．また，F0 とパワーを用いて訂正発話の特徴を分析したところ，誤認識された発話と初回訂正発話の変化は有意ではなかったが，変化のタイプによって被験者を分類し再分析したところ，有意水準 1%で変化が有意であったと報告されている [6]．我々の以前の研究 [7] においても，言い直し発話は声の大きさ，声の高さによる抑揚が大きくなっていることが分かっている．

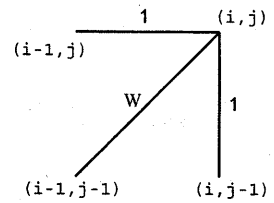
また，言い直し発話の検出についても研究されており，井ノ上，今井らは，未知語処理のための孤立単語の繰り返し音声検出の手法として，(1) 認識候補の重なり度による識別手法，(2) 認識尤度差による識別手法，(3) パワーの時系列ベクトル間の距離による識別手法，の 3 通りの手法を提案しており，手法 1 と手法 3 を組合せることによって，再現率 95.8%，適合率 95.9%の検出精度を得ている [8] [9]．さらに我々は，カーナビゲーションシステムの地名入力タスクにおいて，ユーザが誤認識，特に地名の下位階層のみの誤認識に対してその箇所のみを言い直すという傾向から，その部分的な言い直しを DP マッチングと認識候補の重なり度を用いて検出する手法を実現し，さらにその結果を認識辞書の制約として反映することで言い直し発話，言い直しでない発話の両方において認識性能を向上させた [10] [11]．

本報告では，さらに言い回しの自由度を認めた対話性の高い対話システムにおけるユーザの言い直し発話を検出することを考える．[10] の手法をベースに，より自然な文発声における言い直しに対応できる手法を提案する．より自然な文発声における言い直し発話は，図 1 のようにその前後に別の単語，間投詞，否定語等を含む場合が多く，言い直し発話中のどこに言い直し部分（図 1 では「浜松インター」）が存在するのか分からない．また，その言い直し部分が複数存在する場合も十分に考えられる．よって，言い直し発話とその直前の発話間に共通部分が存在するか否かでその発話が言い直しであるのかどうかの判定を行う．また，ユーザの言い直し発話を検出することにより，認識性能を向上させる方法についても検討する．

2. 言い直し検出方法

2.1 DP マッチングによる検出

文発声における言い直しは 2 発話文中のどこに言い直し (共通) 部分が存在するかが分からない．従って，2 発話間の任意区間に存在する共通部分を抽出する必要がある．そこで，直前の発話と現発話に対して図 2 のような傾斜制限のない DP パスを用いて最適照合パスを求め，線形



for $i = 1, 2, ..., I; j = 1, 2, ..., J$

$$g(i,j) = \min \left\{ \begin{array}{l} g(i-1,j) + d(i,j) \\ g(i-1,j-1) + Wd(i,j) \\ g(i,j-1) + d(i,j) \end{array} \right\}$$

図 2　DP パスと漸化式

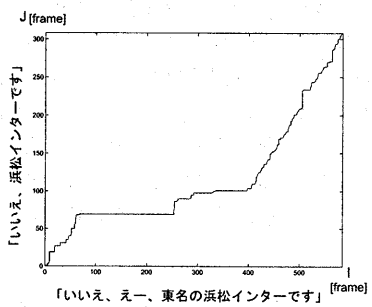


図 3　DP マッチングによる判定の例（文発声の場合）

パス (斜めに進むパス) の区間を共通部分として抽出する．

また，抽出された共通部分の平均距離・最大距離，最小フレーム数についても判定条件とする [13]．

パスの線形性の判定基準は DP パスが i 方向，j 方向に連続して 5 フレーム (40ms) 以上進むならば共通部分でないとする．線形と判定された区間について平均距離，最大距離のいずれかがそれぞれ閾値 (MEAN=8,MAX=20) よりも大きいならば共通部分でないとする．さらに，助詞などの短い区間のマッチングを回避するため，20 フレーム以上継続する区間のみを抽出する．また，パワーによる閾値を与え，無音区間のマッチングを回避する．以上の条件を全てクリアしたものを共通部分と判定し，最終的に共通部分が抽出されれば現発話は言い直し発話であるとする．また，母音などの定常的な音に対して，斜めに進むパスを優先するために，斜めに進むパスの重み W を 1.9 とした．

図 3 に共通部分を含む 2 発話間の DP マッチングの例を示す．共通部分である「いいえ」と「浜松インターです」に対する DP パスがほぼ線形になっていることが分かる．

2.1.1 認識候補の重なり度による方法

現発話 A と直前の発話 B を音声認識した結果得られる各 N-best リストで forced-alignment をとり，DP マッチ

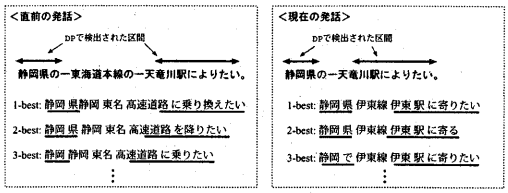


図 4　認識候補の重なり度による判定

ングにより言い直しとして検出された区間 (実際には少しでも重複する区間) に対応する単語を比較し (図 4)，同一の単語が含まれる度合 (重なり度) を求める．DP マッチングにより検出された言い直し区間で N-best リストに存在する単語を W_n とし，重なり度を以下の式で定義する．

$$\text{重なり度} = \frac{\sum_{k=1}^{K} 2 \times \sqrt{C_k^A \times C_k^B}}{N^A + N^B} \tag{1}$$

ここで C_k^A, C_k^B はそれぞれ発話 A, B 中で比較対象となった単語 k の頻度，N^A, N^B はそれぞれ発話 A, B 中に現れた単語の総数である．

言い直し発話の場合は重なり度が大きく，言い直しでない発話の場合は重なり度が小さいと予測できるため，これに閾値処理することにより言い直しか否かを判定する．ここでは最大 200-best(平均 150-best) を用いた．

2.2　2 つの手法の組み合わせ

まず，言い直しである確率を定義する．発話 W が言い直しである場合を $C(W) = 1$，言い直しでない場合を $C(W) = 0$，W に対するある特徴量 $\mathbf{x}$ の値を $\mathbf{x}_W$ とする．このとき，W の特徴量 $\mathbf{x}$ の値が $\mathbf{x}_W$ であった場合に W が言い直しである確率，すなわち $P(C(W) = 1|\mathbf{x} = \mathbf{x}_W)$ を考える．

ここで，言い直しである確率は $\mathbf{x}$ の関数 $f(\mathbf{x})$ に従っているとする．

$$f(\mathbf{x}) = \frac{1}{1 + \exp(g(\mathbf{x}))} \tag{2}$$

本稿では，DP における平均局所距離と重なり度を用いて $\mathbf{x} = (x_1, x_2, 1)^T$ として，3.1 節で説明する実際のデータによって言い直し確率を推定する．$g(\mathbf{x})$ としては線形結合 $g(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ $(\mathbf{a} = (a_1, a_2, a_3)^T)$ を用いた．$f(\mathbf{x})$ とサンプルの言い直し/言い直しでない (1/0) の 2 乗誤差を最小化するようにパラメータ $\mathbf{a}$ を推定することにより，言い直しである確率の関数を推定する [12]．実際に推定した関数を図 5 に示す．各サンプルに対するこの関数の値を閾値処理することによって言い直しか否かを判定する．

3.　評価実験

3.1　評価データ

各手法の閾値や組合せパラメータなどの設定のためのデータとして，図 1 に示したような目的地設定タスクの音声対話システム (認識語彙数は約 15,000 語) を使用して収集した男性話者 10 名による言い直し発話 192 発話，言い直しでない発話 320 発話を用いた．また，評価には，電話番号認識などを含む面会受付タスクの音声対話システム (認識語彙数 864 語) によって収集した言い直し発話 124 発話，言い直しでない発話 136 発話を用いた．ともに文脈自由文法駆動の連続音声認識を用いている．

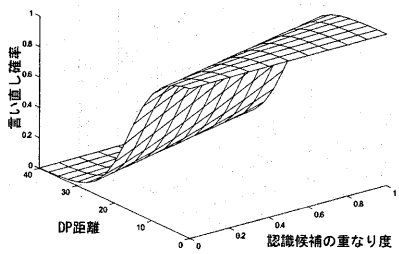


図 5　言い直しである確率の表現

表 1　音声分析条件など

音声特徴量	10 次元 LPC ケプストラム ＋ 10 次元 Δ
サンプリング周波数	12 kHz
分析窓	ハミング 窓
フレーム幅	21.33 ms
フレームシフト	8 ms
音響モデル	114 音節 HMM (4 状態 全共分散 4 混合分布)

音声分析条件と音響モデルは表 1 の通りである．

3.2　評価尺度

各発話に対する言い直しか否かの判定を，再現率，適合率，F 値によって評価した．

$$\text{再現率} = \frac{\text{正しく言い直しと判定された発話数}}{\text{言い直し発話数}} \tag{3}$$

$$\text{適合率} = \frac{\text{正しく言い直しと判定された発話数}}{\text{言い直しと判定された発話数}} \tag{4}$$

$$F\text{ 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \tag{5}$$

3.3　実験結果

図 6 に各手法による再現率–適合率曲線を示す．図が右上に近付くほど性能が良いといえる．また，F 値最大となる場合の値を表 2 に示す．組み合わせた方法が DP マッチングや認識候補の重なり度による方法よりも良い性能を示しており，F 値最大の場合，再現率 92.7%，適合率 89.1%を得た．

また，表 3，4 に，パラメータ学習データおよびテストデータにおけるクローズドテストの結果を示す．最終的に得られる結果はオープンでもクローズドでもほぼ同等となり，頑健なパラメータ設定であることを示している．

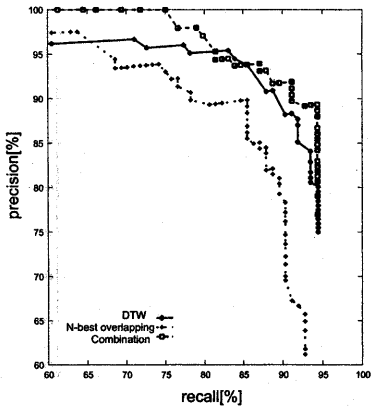


図 6　言い直し判定の Recall-precision 曲線

表 2　F 値最大における各手法の判定性能 (オープンテスト)

手法	再現率	適合率	F 値
DP	91.9%	87.7%	89.8
重なり度	92.7%	62.5%	74.9
組合せ	92.7%	89.1%	90.9

表 3　F 値最大における各手法の判定性能 (クローズドテスト)

手法	再現率	適合率	F 値
DP	95.3%	87.6%	91.3
重なり度	51.0%	93.3%	66.0
組合せ	94.8%	89.2%	91.9

表 4　F 値最大における各手法の判定性能 (テストセットでパラメータ推定しテストセットでテスト)

手法	再現率	適合率	F 値
DP	92.2%	88.1%	90.1
重なり度	85.5%	89.8%	87.6
組合せ	94.4%	89.3%	91.8

ただし，重なり度のみによる結果は大きく異なっている．これは，二つのタスクで語彙数が大きく異なるために，認識結果の重なり度の絶対値に差が生じたためである．従って重なり度のパラメータ設定に用いるデータの選択はやや慎重に行う必要がある．

4.　言い直し判定による効果

4.1　対話システムへの導入

ここでは，システム主導型対話システムにおいて入力ステップの度に確認を求めない対話戦略を考える．ステップ毎の確認はユーザにとって回りくどい印象を与えるため行わない方が良いが，システムの誤認識の際の訂正手段を別途必要とする．先にも述べた通り，人間は同内容を繰り返すこと (言い直し) で訂正を行おうとする．しかしシステムは次の質問に移行しており，訂正発話を受理するためには現在の質問と直前の質問の両方に対する回答発話を認識できるように語彙・文法を設定しておく必要があるが，語彙・文法を増加させることは認識率の低下につながる．ここで言い直し発話が検出できれば語彙・文法を限定して認識性能を向上できる可能性がある．

表 5　パープレキシティによる評価

受付項目	判定なし	判定あり
面会者の名前	121（名前・時間）	58（名前）
面会時間	46（名前・時間・住所）	17（時間）
住所	183（時間・住所・電話）	126（住所）
電話番号	38（住所・電話）	14（電話）
平均	97	54

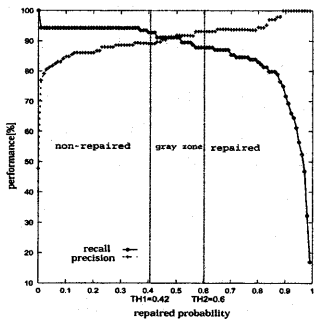


図 7　グレーゾーンの設定

以降では，3.1 節でテストデータの収録時に用いた面会受付システムを想定して評価を行う．テスト文 260 文の内訳は，面会者の名前 58 文，面会時間 78 文，住所 40 文，電話番号 84 文である．各文は例えば面会者名だけではなく，「○○さんに会いたい」のような文になっているものが多い．

4.2　パープレキシティによる評価

言い直し判定のない場合（常に可能性のある語彙・文法を設定しておく）と言い直し判定のある場合（随時適切な語彙・文法に限定することができる）とで，パープレキシティを求めた．

各受付項目毎の判定なし／ありの場合において，文法 (CFG) から求めたテストセットのパープレキシティを表 5 に示す．また，表 5 の括弧内に，パープレキシティを求める際に用いた文法を示す．判定なしの場合は，可能性のある複数の文法（即ち，各項目の文法とその前後の項目の文法）を用いた場合の結果である．判定ありの場合は，各項目の文法に限定した場合の結果である．判定のない場合に比べて，受付項目毎，全項目，両者において，パープレキシティは約 1/2 に減少している．このことからも，言い直し判定を行うことにより，認識率の向上が期待できる．

表 6　言い直し判定による認識率向上の効果

方法	言い直し発話	言い直しでない発話
言い直し判定なし	42.7%	70.5%
100%判定が成功	51.6%	79.4%
実際の判定性能		
グレーゾーンなし	49.2%	77.9%
グレーゾーン 1	49.2%	78.7%
グレーゾーン 2	50.0%	77.9%

4.3　認識率による評価

4.3.1　グレーゾーンの導入

言い直し判定により，正しく判定された発話に対しては適切な語彙・文法に限定することができる．しかし，その反面，誤って判定された発話に対しては誤った語彙・文法が設定されてしまうため，必ず誤認識されてしまうという問題がある．

そこで，言い直しかどうか曖昧な場合には言い直し判定の失敗による誤認識を軽減するため，可能性のある複数の語彙・文法を設定しておくグレーゾーンを導入する．図 7 に閾値による再現率と適合率の変化を示す．唯一の閾値で言い直しか言い直しでないかを 2 値的に判定するのに対して，言い直し／どちらともいえない／言い直しでない，の 3 値に判定する．ここでは，グレーゾーン 1：0.42〜0.6，グレーゾーン 2：0.3〜0.7 の 2 種の設定を行った．

4.3.2　認識結果

ここでは，(1) 言い直し判定しない場合，(2)100%判定が成功したと仮定した場合，(3) 実際の判定性能 (再現率 92.7%・適合率 89.1%) の場合（グレーゾーンあり／なし），の 3 通りで認識実験を行った．

認識結果を表 6 に示す．言い直し判定なしの場合の認識率がベースライン，100%判定が成功したと仮定した場合の認識率が上限である．言い直し発話，言い直しでない発話，両者に対して認識性能向上の効果があることが確認できた．また，グレーゾーンの導入による効果も見られた．

5.　音声対話システム

5.1　構　　成

言い直し検出を用いた面会受付システムを構築した．システムの構成を図 8 に示す．本システムは，面会者の名前，面会時間，来訪者の住所，電話番号を順次尋ねていき，各入力を認識するたびに復唱はするが確認は求めない．誤認識の場合には，次の質問をされていてもユーザは前の質問に対する回答を再度行うことができ，システムはこれを言い直し検出によって検出し，辞書を一つ前の状態に戻して認識を行うことで対処する．

5.2　発声方向検出

一般に音声認識では，発話者はトークスイッチを押すなどして発話の開始を認識システムに教示する．しかし音声入力する度となると煩わしいものである．そこで，発話者の発声方向を検出する手法を音声対話システムに導入した．

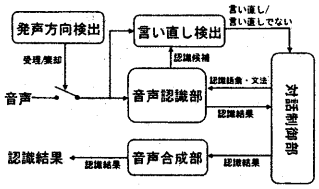


図 8　音声対話システムの構成

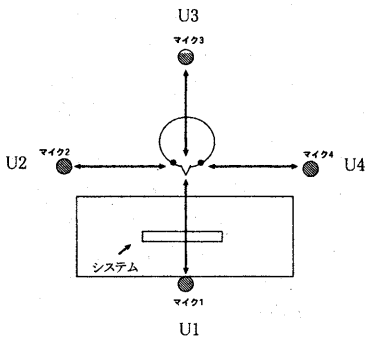


図 9　音声パワーの比を用いた発声方向検出

人間の発話時の音圧分布は，話者の顔面の正中面からの角度や周波数に応じて変化することが音響管モデルやレプリカによる測定で報告されている [14]．唇や頭部の形状が放射される音声の指向性に影響し，頭部周囲の音圧は，水平面および頭部正中・鉛直面の分布とも正面に比べ，正面からの角度が大きくなるにつれて後背部で数〜十数 dB 減衰し，特に高域成分ほど減衰が大きい [15]．

そこで，図 9 に示すように発話者の周囲にマイクロホンを設置し，また，マイク 1，2，3，4 の方向に向かった発話をそれぞれ U_1，U_2，U_3，U_4 として現発声が 4 方向のうちどの方向に向かったものなのかをを判定する．

あらかじめマイク 1，2，3，4 に向かって収録した学習データから，ある一定の音声区間 (16kHz サンプリングで 256 点窓長．128 点シフト) 毎の音声パワーを求め，2 個のマイクロホン間の有声区間のパワーの比の対数 x を計算し，それらが正規分布 $P(x|U_i) = N(\mu_i, \Sigma_i)(i = 1, 2, 3, 4)$ をなすと仮定して各方向毎に平均・(共) 分散を推定しておく．音声の入力があった場合には，各フレームから尤度を求め，尤度最大となる方向を推定発声方向とする．厳密には壁などによる反射音，残響，周波数による影響が考えられるが，ここでは特に考慮していない．

特徴量としてマイク 1 と 2，1 と 3，1 と 4 の間のパワー比の対数を用い，また各フレーム毎の判定結果を 50 フレーム (0.4sec) の区間で集計して最もその方向と判定された回数の多い方向を，その区間の発声方向とした．

話者 10 名 (男性 7 名，女性 3 名) に各方向に，「あ」「い」

表 7　発声方向の判定結果

	U_1	U_2	U_3	U_4
U_1	0.998	0.001	0	0
U_2	0.068	0.851	0.080	0
U_3	0	0.113	0.886	0
U_4	0.088	0.136	0.119	0.655

「う」「え」「お」の母音もしくは 3〜4 秒程度の短い文章 3 種類を発声させ，収録した音声 10 人分のうち 5 人分を学習データとし，残り 5 人分で評価を行った．その結果をコンフュージョンマトリックス（行：正解，列：入力）として表 7 に示す．ほぼ正しく方向を判定できているのがわかる．

実際のシステムにおいては，音声区間検出を併用して連続する音声区間に対して発声方向を判定する．従って発声によっては 50 フレームに満たない場合もあるが，予備実験により，20〜30 フレーム程度あればほぼ安定して判定できることを確認した．

5.3　システム動作例

図 10 に，システムが動作している様子を示す．ディスプレイに受付係が表示され，先に示した内容を順次質問する．システムとの対話中に同伴者と横を向いて会話してもシステムは動作せず，システム方向の発話のみによって対話が進行する．また，言い直し判定を用いた対話も良好に動作する．

6. ま と め

本稿では自然な対話音声中に含まれる繰り返し部分の検出によって言い直し発話を検出することを試みた．DP マッチングをベースとした手法，および音声認識の結果の N-best 候補中の単語の重なり度を用いた方法とを組み合わせることによってオープンテストで再現率 91.9%，適合率 87.7%の検出性能を得た．さらに，この検出結果を用いて音声認識の際の語彙・文法の制約に用いることで対話音声の認識性能を向上できることを確認した．

また，この方法を対話システムに導入した．この対話システムは発声方向検出機能も備えており，ユーザはシステムに対して手を用いることなく，また誤認識時にも確認なしに繰り返し入力で訂正することができる，柔軟なシステムとなっている．

図 10　音声対話システムの動作の様子

しかし，人間は繰り返しを含んだ発声をした場合にも必ずしも訂正発話ではない．例えばさらに情報を付加する場合などにも一部を繰り返した発声を行う場合もある．今後は，繰り返しの検出と，人間の繰り返しの発声傾向の分析とを統合して対話戦略・意味解析に利用していくことを考えている．

文　献

[1] 平沢 純一, 宮崎 昇, 相川 清明, “質問-応答連鎖からの音声対話システムの誤解の検出”, 情報処理学会研究報告, 2000-SLP-34-41, pp.239-244, 2000.

[2] 平沢 純一, 宮崎 昇, 相川 清明, “音声対話システムの誤解に対するユーザ応答の分析”, H12 年度春季日本音響学会講演論文集, 3-8-10, pp.85-86, 2000.

[3] S. Oviatt, M.t MacEachern and G.-A. Levow, “Predicting hyperarticulate speech during human-computer error resolution”, Speech-Communication, Vol.24, pp.87-110, 1998.

[4] G.-A. Levow, “Adaptation in spoken corrections: Implications for models of conversational speech”, Speech-Communication, Vol.36, pp.147-163, 2002.

[5] M. Swerts, D. Litman and J. Hirschberg, “Correction in Spoken Dialogue System”, ICSLP2000, Vol.2, pp.615-618, 2000.

[6] 山肩 洋子, 河原 達也, “音声対話システムにおける訂正発話の韻律的特徴の分析”, 人工知能学会研究会資料, SIG-SLUD-A101-3, pp.5-12, 2001.

[7] 角谷 直子, 北岡 教英, 中川 聖一, “カーナビの地名入力における誤認識時の訂正発話の分析と検出”, 情処処理学会研究報告, 2001-SLP-37-11, pp.61-66, 2001.

[8] 井ノ上 直己, 今井 裕志, 橋本 和夫, 米山 正秀, “誤認識訂正のための繰り返し音声検出手法”, 電子情報通信学会論文誌, Vol.J84-D-II, No.9, pp.1950-1959, 2001.

[9] 今井 裕志, 井ノ上 直己, 橋本 和夫, 米山 正秀, “未知語処理のための繰り返し音声検出手法”, 電子情報通信学会技術研究報告, SP99-26, pp.1–6, 1999.

[10] 北岡 教英, 角谷 直子, 中川 聖一, “カーナビの地名入力における誤認識時の言い直し発話の検出と認識”, 電気学会電子情報システム部門誌 (C), Vol.122-C, No.12, pp. 2020–2027, 2002.

[11] N. Kakutani, N. Kitaoka, S. Nakagawa, ”Detection and recognition of repaired speech on misrecognized utterances for speech input of car navigation system”, ICSLP2002, pp. 833–836, 2002.

[12] 北岡 教英, 赤堀 一郎, 中川 聖一, “認識結果の正解確率に基づく信頼度とリジェクション”, 電子情報通信学会論文誌, Vol.J83-D-II, No.11, pp.2160-2170, 2000.

[13] 兵後 裕子, 中川聖一, “連続発声された二発話文間の DP マッチングによる共通部分の抽出”, 電子情報通信学会春季全国大会講演論文集 [分冊 1],A-22,pp.22, 1989.

[14] N. Miki et. al., “Transfer function of 3-d vocal tract model with higher mode”, 1st ESCA Tutorial and Research Workshop on Speech Production Modeling — 4th Speech Production Seminar, pp.211-214. ESCA, 1996.

[15] Jemes L.Flanagan, “Speech Analysis Synthesis and Perception”, Splinger–Verlag, 1972.