音声言語情報処理 17－3
（1997. 7. 18）

音声対話における属性情報・認識誤りフィードバックの効果

北井 幹雄，相川 清明
ＮＴＴヒューマンインタフェース研究所 〒239 横須賀市光の丘 1-1

あらまし 情報検索など，認識結果の正誤をユーザに確認しながらタスクを達成する電話音声対話システムにおいて，システムが間違った候補を確認した場合に，ユーザに入力単語の属性情報あるいは確認候補単語の誤り程度を指定させて認識候補の順位を再評価し，ユーザに正解が確認されるまでの対話回数を低減する方法について提案する．国名，会社名を認識候補としたシミュレーションによる評価の結果，１位候補が誤りであった場合の２位候補の正解率を５％以上改善できることを確認した．

和文キーワード 音声対話，対話システム，認識応答システム

On the Use of Feedback of the Spoken Word Category or the Incorrectness Level of a Confirmed Word Candidate

Mikio KITAI, Kiyoaki Aikawa
NTT Human Interface Labolatories
1-1 Hikari-no-oka Yokosuka-Shi Kanagawa 239 Japan

Abstract Confirmation of word candidate correctness by yes/no for speech recognition result is inevitable to get accurate input word, especially in an telephone-based application. However this type of confirmation may be awkward to the caller, speedy and natural confirmation interface is expected. In this paper, we propose two types of estimation methods for the order of word candidates by indicating input word category or incorrectness level of a confirmed word candidate. Then we evaluate the effects of the methods by simulation using recognition results.

英文 key words speech dialogue, dialogue system, voice activated system

1 はじめに

一般に、認識システムに直接マイクで音声を入力する場合に比べ、電話回線を経由した音声の認識性能は、帯域制限や回線を始めとする各種電話系のノイズにより劣化する。そこで、我々がこれまで検討して来た電話音声認識応答システムでは入力項目に対する認識結果は復唱して正誤を確認するとしてきた[1]. このシステムは，ユーザはシステムから確認された認識結果が間違いであった場合、例えば「やり直し」と発声して再発声を行なってシステムに再認識させるか、あるいは「いいえ」や「次候補」などと発声して、システムが復唱した候補が誤りであることを指示する必要がある。前者の場合、ユーザがシステムにとってより協力的な、すなわちより認識し易い発声を行えば、システムが正しく認識する可能性は高くなるが、システムに習熟していないユーザの場合，どのように発声すれば良いか分からず，区切った発声を行なうなど逆に認識しずらい発声を行なう可能性も高い．後者の場合、システムは次候補の正誤を確認するが，この方法では上位の認識候補に正解がない場合、正解にたどり着くまで多くの対話を要することになる。

本論文では，ユーザに正解候補が確認されるまでの対話回数を減らすことを目的として，誤った認識結果が確認された場合に，(i) 入力単語に関する属性情報を指定する方法，(ii) 否定語により全否定するのではなく誤りの程度を指定する方法，についての検討結果を報告する．

2 属性情報のフィードバックによる認識結果の再評価

ユーザが入力単語の属性情報をシステムに指示することにより、入力単語の認識結果の優先順位を再評価する方法について検討した。入力単語の属性情報とは、例えば国名がシステムに入力する単語である場合、地域名は一種の属性となる．

入力単語の属性情報の指定は、入力単語の認識結果が間違いであった場合にのみ行なうのが最も効果的である。そこで、今回は、1位認識候補の正誤は従来通り「～ですか」と確認し、この1位候補が例えば「いいえ」により誤りと指定された場合に、ユーザに入力単語の属性情報を入力して貰い、その属性情報の認識結果により、2位以下の認識候補の優先順位を再評価する方法について検討した。

2.1 属性情報の指定方法

入力単語の認識結果の再評価に使用する属性情報の総数と，その知名度に応じて，属性情報の指定方法を変えることを想定し，以下の2つの方法を評価した．

【方法1】すべての属性情報を認識させ、その認識結果に応じて、入力単語の認識結果の順位を再評価する。

【方法2】入力単語の認識結果の上位候補の属性情報N個に、それ以外という単語、例えば「その他」を、属性情報として認識させ、その認識結果に応じて、入力単語の認識結果の順位を再評価する。

単語の属性は，入力単語の認識結果を絞り込むための追加情報として入力するため，基本的に認識誤り率が可能な限り低いことが望まれる．よって，【方法1】は，属性数が少なく(例えば10未満)，且つ各属性間の類似度が小さい場合に適していると考えられる．【方法2】は，属性数が多い場合に適していると考えられ，例えば入力単語が会社名で属性が業種の場合，システムが「業種は，○，△，□…のいずれですか」などとユーザに入力可能なメニューを提示することを想定している．

2.2 評価方法

評価用のデータとしては，国の名前，会社名の音声をそれぞれ認識させた結果(1位正解率が93％程度の場合の結果)のうち，1位候補が間違いである認識結果(10位まで)を用いた．認識には当研究所のHMM-LR方式の不特定話者電話音声認識ソフトウェアを用いた[2].

提案した方法による入力単語の認識結果の再評価は以下のように行なった．【方法1】では，1位候補が間違いであった場合に，ユーザが入力単語の属性(国名の場合は地域名，会社名の場合は業種)を入力し，それが1位で正しく認識されたと仮定して，入力単語の2位以降の認識結果の中からその属性を持つ候補を，そ

の属性を持たない候補より上位とした．【方法2】では，選択したN個の属性の中に正解の属性がある場合は方法1と同じで，正解の属性がない場合は，「その他」が入力されたと仮定(しかもそれが1位で正しく認識されたと仮定)して，選択した属性を持たない単語候補を選択した属性を持つ単語候補より上位とした．なお，選択した属性を持つが候補にないものについては，今回は候補の最後から最大10位まで追加することにした(順位は表記順)．

2.3 結果と考察

表1，2にそれぞれ国名，会社名の結果を示す．今回の実験では，国名(正解の国の名前および国際機関名，1国で複数ありで218単語)の場合の属性としては地域名(アジア，アフリカ，オセアニア，北アメリカ，国際機関，南アメリカ，ヨーロッパの7つ)を，会社名(有名な会社，200単語)の場合の属性としては業種名(35種)を使用した．

表1: 属性指定の場合の国名の累積認識率

	なし	全属性対象	上位4属性
2位	50.8	60.8	60.0
3位	60.0	68.8	65.0
4位	61.7	70.0	65.8
5位	63.3	70.4	66.7
10位	66.7	73.8	68.3

表1より，全属性を認識対象として指定した場合の方が上位4個の属性(プラス「その他」)を認識対象とした場合よりも良い結果が得られた．双方とも従来法より，2位候補の正解率で10%程度，3位以下でも累積認識率が改善した．国名の場合は，属性として使用した地域名は7個と少ないので，全属性を認識対象とする方法の採用が適当と考えられる．但し，各国がどの地域に属しているかの正確な知識を持つことは通常は困難であるため，隣接地域の国では双方の地域でも検索できるようにしておく必要がある．また，地域名が分からない場合の対策として例えば「分からない」というような指定が可能なようにしておく必要がある．

表2の会社名の場合も，国名の場合と同様に性能が改善されているが，こちらの方が改善率が高く，全属性を認識対象とした場合の10位までの累積認識については40%近く改善している．これは，属性が国名の場合に比べ細かく分類されていることが最大の理由である．しかし，属性が35個では属性自体の認識性能の影響が実際の場面では問題となってくると考えられるので，上位N個(今回は4)の属性を指定する方法が現実的である．

表2: 属性指定の場合の会社名の累積認識率

	なし	全属性対象	上位4属性
2位	33.6	54.7	45.3
3位	38.3	61.7	51.6
4位	41.4	68.8	55.5
5位	43.8	72.7	57.0
10位	54.7	92.2	66.4

ところで，提案法では，属性の入力を伴うので，認識結果をそのまま確認する従来の方法に比べ入力回数が1回増えていることになる．よって，発話回数の比較と言う意味では，従来法のn+1位までの累積認識率と，提案法でのn位までの累積認識率の比較の方が適切である．この意味では，従来法では2度目の確認(3度目の発声)で約51%が正解と同定されるのに比べ，属性指定では3度目の発声では何も同定されず，従来法より1度多い4度目の発声で60%が正解と同定されるので，4度目の入力以降の改善率が良くても，それだけで単純に従来法より良いとは評価できない．評価は，候補の提示手段(音声and/or画面)，候補の確認方法(複数候補の逐次確認／一括確認)により変わる．

3 認識誤り情報のフィードバックによる認識結果の再評価

システムが誤った認識候補を確認した場合に，予めシステムが用意した確認候補と正解候補のひらがな表記上の類似性を表す数個の言葉(以後，誤り指摘語と呼ぶ)の中から，ユーザが一つを選択して指定し，システムが指定に応じて認識候補の順位の見直しを行なう方法について述べる．誤り指摘語としては，今回は，一般的な違いを表現する言葉である，全然違う，殆んど違う，半分以上違う，半分位違う，少し違う，の使用を考えた．

ユーザが指定する誤り指摘語により認識結

果の順位を見直すためには，以下の(1)～(3)が必要になる．それぞれ3.1～3.3で説明する．

(1) 確認した第i位の認識候補Ciと第j位の認識候補Cj (i≠j, j＞i) の類似度Sijの定義
(2) 誤り指摘語と上記類似度Sijの関係の定義
(3) 上記(1)と(2)を考慮した認識候補の順位の見直し方法の確立

3.1 類似度の定義

今回は「ひらがな」表記上の類似度を使用した．正誤を確認した第i位の認識候補Ci (iは正の整数．最大値は認識対象候補数から1を引いたもの) とその他の第j位の認識候補Cj (i≠j, j＞i) との類似度は以下のように定義した．

Sij = 100 − 100 × (max{L(Ci),L(Cj)} − CiとCjのひらがな列の一致数) ／L(Ci)

ここで，L(Ci)はCiのひらがな数を，L(Cj)はCjのひらがな数を，max{L(Ci),L(Cj)}はL(Ci)とL(Cj)のうちの大きい数を意味する．CiとCjのひらがな列の一致数は，各ひらがな列の中で互いに一致する部分ひらがな列に対し，以下の条件を満足する部分ひらがな列の長さの総和として与えた．これは，CiとCjのひらがな列に対して，島駆動型検索を行なう際に以下の条件を制約として設けて行なうことに相当する．

【条件1】最長一致候補を含むこと
【条件2】長さ1の縮退，伸長は1回のみ許容
【条件3】部分的な一致は最低長さ2以上

条件1は，今回の検討が候補を音声で確認する場合を対象としているので，人間は短い部分の断続的な一致よりもある程度の長さで連続した一致の方を基準に違いを認識するのではないかと考えて設定した．条件2は，長さ1の縮退や伸長は頻繁に起こるのでこれは許容すべきであるが，これが連続した場合，人間には認識できないと仮定し設定した．条件3も同様で，長さ1では，後に述べる先頭文字と最終文字の場合を除き，殆んど人間は認識できないと考え設定した．

条件3の例外としては以下を設けた．

【例外1】先頭文字と最終文字に限り1文字での一致を許容．
【例外2】一方が他方に濁点，半濁点を付けたものと一致する場合は一致数を0.5とカウント(但し，元の文字数として条件3をクリアすれば問題なしと見なす)．

今回の検討では，異なるひらがなの間の音響的な類似度については上記の例外2以外一切考慮しなかった．

3.2 誤り指摘語と類似度の関係

同じ誤り指摘語を使っても，通常は個人毎に認識する誤りの程度の幅は異なる．よって，本来はその振れ幅自体の検討を行なって，誤り指摘語と類似度の関係を決める必要がある．しかし，今回は，本手法の効果の確認のため，それぞれの誤り指摘語に対して，類似度に対する許容する上限値と下限値，更には代表値を適当に設定して，各認識候補の類似度Sijに対して，許容範囲にあるか否か，更に許容範囲にある場合には代表値にどれだけ近いかに応じ，指摘された誤りの程度に近い単語と判断することにした．同じ近さにある認識候補については認識尤度の高いもの，類似度の高いものをこの順序で優先した．

3.3 認識候補の順位の見直し

認識候補の順位の見直しは，認識尤度と類似度を考慮して以下の手順で行なった．以下では，第i位の認識候補Ciの正誤が確認され，この認識候補Ci(ここでi=1,2,...nで，nは認識候補数) の誤りの程度が誤り指摘語Yにより指定された場合を想定している．

(1) 誤り指摘語Yに対する類似度の代表値，上限値，下限値を取得し，ぞれぞれM{Y}，上限値Max{Y}，下限値Min{Y}とする．
(2) 認識候補Ciに対する認識候補Cj (ここでj＞i) の類似度Sijを求める．なお，認識候補Ciに対する認識尤度はL{Ci}とする．
(3) 認識対象である単語のうち，n位までの認識候補に含まれないが，Ciとの類似度がMax{Y}以下でMin{Y}以上のものがあれば，それを新規に認識候補に加える．これにより認識候補の数はn'(ここでn'≧nとする) に，またn+k (ここでk=0,1,...n'-n) 位の

候補Ｃｋの認識尤度としてはｎ位の候補Ｃｎの認識尤度Ｌ{Ｃｎ}より予め決めた値Ｌ０を引いたものを設定する．
(4)k=i+1 からｎ'に対し(5)を実施．終了後に(6)に進む．
(5) 認識候補Ｃｋに対する評価値Ｐｋを次のように算出する．類似度Ｓikが Max{Ｙ} 以下で Min{Ｙ} 以上であれば，100 からＳikと M{Ｙ} の差の絶対値を引いたものを評価値Ｐｋとする．類似度Ｓikが Max{Ｙ} 以下 Min{Ｙ} 以上の条件を満足しない場合はＰｋを０とする．
(6)k=i+1 とし(7)に進む．
(7)k に１を足し，k がｎ'ならば終了．そうでなければ(8)に進む．
(8) 認識候補Ｃｋに対する認識尤度Ｌ{Ｃk}と１位候補Ｃ１の認識尤度Ｌ{Ｃ1}の差が予め定めたしきい値Ｌ１より大きい場合は終了．
(9) 認識候補Ｃｋの評価値Ｐｋが０であれば(6)に戻る．そうでなければ(10)に進む．
(10)m=k+1 とし(11)に進む．
(11)m から１を引き，m が i+1 ならば終了．そうでなければ(12)に進む．
(12) 認識候補Ｃｍと認識候補Ｃm-1の評価値ＰｍとＰm-1の差の絶対値が予め決めた値Ｔより小さく，且つ認識尤度Ｌ{Ｃｍ}とＬ{Ｃm-1}の差の絶対値が予め決めた値Ｌ２以下でＬ３以上の場合は，候補ＣｍとＣm-1を入れ換え(10)に戻る．そうでない場合は(7)に戻る．

なお，(5)で類似度が条件を満足しない評価値を今回は０としたが，０とせずに計算させた場合の検討は今後の課題である．

3.4 評価方法

評価用のデータは２章と同じものを用いた．誤り指摘語は，表３，４の２通りの場合を考えた．3.3 の処理手順で述べたしきい値は表５のように与えた．なお，１位の認識候補の正解に対する誤り指摘語の選択については，これらの単語間の類似度が表３又は表４のいずれの誤り指摘語に相当するかにより決めた．この誤り指摘語は，本試験では表３，表４に示すように各単語に対応する類似度の許容範囲に重なりを設けなかったため，一意に決めることが出来た．しかし，実際は人間の判断にはバラツキがあるので重なりを設ける必要があるが，この検討は今後の課題とした．よって今回の実験は，提案手法の上限値に近い結果を見ていることになる．今回の実験ではユーザが入力した誤り指摘単語は正しく１位で認識されたものと仮定した．

表 3: 誤り指摘語（４段階）

単語	最小値	最大値	代表値
全然違う	0	20	0
半分以上違う	21	40	30
半分位違う	41	70	55
少し違う	71	99	99

表 4: 誤り指摘語（５段階）

単語	最小値	最大値	代表値
全然違う	0	4	0
殆んど違う	5	20	10
半分以上違う	21	40	30
半分位違う	41	70	55
少し違う	71	99	99

表 5: 候補の順位変更のためのしきい値

Ｌ１	100000 (制限なしと同じ)
Ｌ２	50000 (制限なしと同じ)
Ｌ３	10000 (同じような類似度を持つ候補間での逆転を抑制)
Ｔ	10

3.5 結果と考察

表６，表７にそれぞれ国名，会社名の結果を示す．表６より，国名の場合，誤りの程度を指定することで，２位候補の正解率を５％以上改善することができた．３位以降の累積正解率でも２％から７％改善した．また表７より，会社名の場合も，２位候補の正解率は５％以上改善することができた．一方，３位以降の累積正解率では７％から 12％改善できた．

ところで，４段階と５段階の指定の違いは「全然違う」と「半分以上違う」の間に「殆んど違う」を入れたことである．入れた理由は４段階指定の場合に，(1)「全然違う」と言った場合に，候補絞り込みの効果が全くなかったこ

表 6: 誤り指摘を行なった場合の国名の累積認識率 (データ数240)

	なし	4段階指定	5段階指定
2位	50.8	55.8	57.1
3位	60.0	62.1	62.9
4位	61.7	63.3	63.8
5位	63.3	67.1	67.9
10位	66.7	71.7	73.8

表 7: 誤り指摘を行なった場合の会社名の累積認識率 (データ数128)

	なし	4段階指定	5段階指定
2位	33.6	38.3	39.8
3位	38.3	46.1	47.7
4位	41.4	50.0	50.8
5位	43.8	52.3	53.9
10位	54.7	66.4	67.2

と，(2) 効果があったのは「半分以上違う」と「半分位同じ」のみであったこと，による．しかし，表6と表7を見る限り，今回の実験では5段階にした効果は小さかったことが分かる．

今回評価した認識結果が，表3の4段階の誤り程度にどのように分布していたかを示す．国名の場合は，「全然違う」が全体の約51％(2位正解率約36％)，「半分以上違う」が全体の約28％(2位正解率約60％)，「半分位違う」が全体の約13％(2位正解率約63％)，「少し違う」が全体の約8％(2位正解率約100％)，であった．会社名の場合は，「全然違う」が全体の約63％(2位正解率約28％)，「半分以上違う」が全体の約20％(2位正解率約27％)，「半分位違う」が全体の約16％(2位正解率約60％)，「少し違う」が全体の約1％(2位正解率約100％)，であった．よって，「少し違う」場合は，2位正解率が100％であったため，提案法による改善は最初から無理であった．また「全然違う」と指摘された場合には全く結果を改善することができなかった．これは，国名，会社名など同じカテゴリに属する単語間では基本的に音響的に類似性の少ないものが選ばれているためと考えられ，これ自体は避けられない．1位候補が正解と「全く違う」場合にも改善効果を上げるためには，2位候補以下に対しても，正解が提示されるまで，誤り程度の指定を続けて，候補を絞り込む必要がある．

本章の方法と2章の方法の結果をユーザの発話回数の観点から比較する．2章の考察で述べた理由により，2章の提案法のn位の累積正解率と本章の提案法のn＋1位の累積正解率の比較が適切である．更に，属性指定の場合，国名では「全属性指定」の結果を，会社名で「上位N個指定」の結果を比較対象として考える．この場合，本手法の3位候補の累積正解率と2章の2位候補の累積認識率では本章の手法の方が若干良い結果が得られ，それ以降では同等か低くなっている．しかし，本章で提案した方法は，2位以下の候補に対してもその候補が誤りであれば，全く同じ方法で誤りの指示を続けて行くことができ，且つその時点までの評価結果を使うことが出来るので，本章の方法の方が良い可能性が高い．但し，今回は考慮しなかったが，人間の指示の曖昧性を考慮した評価が必要であり，その場合にも今回と同じ結果が得られる保証はなく，今後の検討が必要である．

4 まとめ

音声認識応答システムにおいて，間違った認識結果が確認された場合に，(1) 入力単語の属性を指定，(2) 確認された候補の入力単語からの誤り程度を指定，することで，認識結果の順位を見直して，正解が得られるまでの操作回数を低減する方法を提案した．国名と会社名の認識結果を対象としてシミュレーションによる評価を行なった結果，2位の候補の正解率を5％以上改善することが出来た．今後は，認識対象数やデータを増やした場合の調査，指定する単語の属性数と性能の関係の調査，人間が誤り程度を指定した場合の結果との比較評価などを行なう予定である．

謝辞 日頃御指導を頂き，発表の機会を与えて下さった北脇ＮＴＴヒューマンインタフェース研究所音声情報研究部長に感謝致します．

参考文献

[1] 西宏之，北井幹雄：”尤度と正解率の統計的関係に基づく認識結果確認対話戦略”，信学論A，Vol.J 77-A, No.2, pp.251-258, Feb. 1994.

[2] M.Kitai et al. : ”Experimental Interactive System For Telephone Applications With Speech Recognition and Synthesis Functions”, Proc. of IVTTA'96, pp.25-28, Sep. 1996.