

## 共通部分単語の提示を併用した単語音声入力インタフェース

張 用起<sup>†1</sup> 甲斐 充彦<sup>†1</sup> 王 龍標<sup>†1</sup>

大語彙の単語(キーワード)入力タスクに対し、認識精度の影響による入力効率の低下を抑えることを目標とした GUI 併用の音声入力インタフェースを提案する。この音声入力インタフェースは、発話の認識結果をもとに生成される複数候補リストを利用者に提示し、そこから利用者が選択する GUI(Graphical User Interface) の仕組みを併用することを前提とする。この様なユーザインタフェースシステムの有効性は認識結果の N-best 候補のみを提示する場合、その有効性は音声認識システムの利用者や環境による認識精度の違いに大きく影響される。我々の方法は、認識性能が高くない悪環境下での入力効率を向上させるため、GUI として提示される複数候補の中に、一般的な認識結果の N-best 候補だけでなく検索(絞り込み)候補を含めるアプローチに基づく。検索候補は、事前に定義された共通部分単語と N-best 出力結果から動的に作成される。本稿では、共通部分単語の定義や、表示する候補リストの構成法の違いによる性能比較について報告する。雑音環境を想定した約 13000 語の単語入力タスクに対して評価実験を行った結果、提案するインタフェースは認識結果の N-best 候補のみを提示するベースライン手法を上回る性能を示し、入力効率改善された。一方、共通部分単語や構成法の違いはインタフェースの性能に影響を与えるが、いずれもベースライン手法より高い性能を示した。

### Speech Interface on Combination of Search Candidates from the Common Word Parts

YONGGEE JANG,<sup>†1</sup> ATSUHIKO KAI<sup>†1</sup>  
and LONGBIAO WANG<sup>†1</sup>

In this paper, we propose a new user interface based on a combination of speech input and graphical user interface(GUI) to restrain decrease of input efficiency by influence of recognition accuracy. A new speech interface is based on combination using GUI which can make users select candidates from multiple candidates list made by recognition results. But, an effectiveness of these user interfaces is strongly influenced by a difference of recognition accuracy from a user or an environment with typical speech recognition systems if they simply display N-best candidates of recognition results only. To improve input efficiency under adverse conditions (where the recognition accuracy is not very

high), our method uses not only N-best candidates from the recognition result but search candidates which can narrow candidates. The search candidates are dynamically made from N-best output result and common word parts defined in advance. In this paper, we compare different methods of extracting common word parts and two methods of generating candidates list. We carried out an experiment of word input task for about 13,000 vocabulary words in several adverse conditions, and evaluated the performance of our GUI-assisted speech interface system by simulating a GUI-assisted (multimodal) isolated-word input task. As a result, the proposed systems indicated improvement for input efficiency and performance over the baseline system. Although the difference of a definition of common word parts and a way to make candidates list influenced to performance of interface system, all of proposed systems still show better performance than the baseline's.

#### 1. はじめに

音声認識システムを用いるアプリケーションにおいて、誤認識の扱いは避けられないことである。画面表示を備えた情報機器での機器操作アプリケーションのように単語の音声入力タスクを想定したとき、複数の認識結果を表示する GUI 併用の音声入力インタフェースはこの問題に対する簡単かつ効率的な改善策といえる<sup>1)2)5)</sup>。GUI を併用した音声入力インタフェースを実現する最もシンプルなアプローチとして、認識尤度が高い順に複数の認識候補を提示する方法がある<sup>4)5)</sup>。このような GUI 併用によるインタフェースは、必ずしもコマンド操作のような単純な単語入力タスクのためだけでなく、システム主導のマルチモーダル対話インタフェースのシンプルな実現方法としても有効と考えられる。しかし、大語彙のタスクにおいては、発話した単語と類似した発音の単語との誤認識のため、入力した単語が候補リストの中に現れにくくなりやすい。一般的に、このようなインタフェースは、認識精度が下がるとその効果も小さくなり、その原因は表示する候補の数と発話単語の含有率が比例しないためである(2章を参照)。

本論文では、上述したような複数候補提示のユーザインタフェースシステムにおける入力効率の問題を改善するため、選択肢の絞り込みを可能とする新しい GUI 併用の音声インタフェースシステムを提案する。提案するシステムでは、複数候補提示において絞り込みを可能とするため、検索候補を新たに導入する。また、想定するタスク語彙の中の任意の単語間で発音および表記が共通する単語断片のことを共通部分単語と呼び、これらを検索候補とし

<sup>†1</sup> 静岡大学  
Shizuoka University

て用いる。

関連研究<sup>1)</sup>では、組織名入力タスクのための複数候補表示のインタフェースシステムにおいて、組織名に共通する単語断片を基本単語単位と定義し、音声認識の単位として用いることで組織名のカバー率向上の効果を示した。しかし、限られた候補数の中で単語断片の候補をどのように選択して提示するかについては検討されていない。

提案する音声インタフェースは、共通部分単語から得られる検索候補を用いて、単語（キーワード）候補の絞り込みを可能とする。共通部分単語を抽出する方法や、検索候補を提示する候補リストに挿入する方法については様々な方法が挙げられる。例えば、共通部分単語の抽出基準は、共通部分単語の長さ、共起頻度などが考えられ、候補リストの構成法では候補リストの信頼度スコアや表示回数の期待値を用いる方法が考えられる。

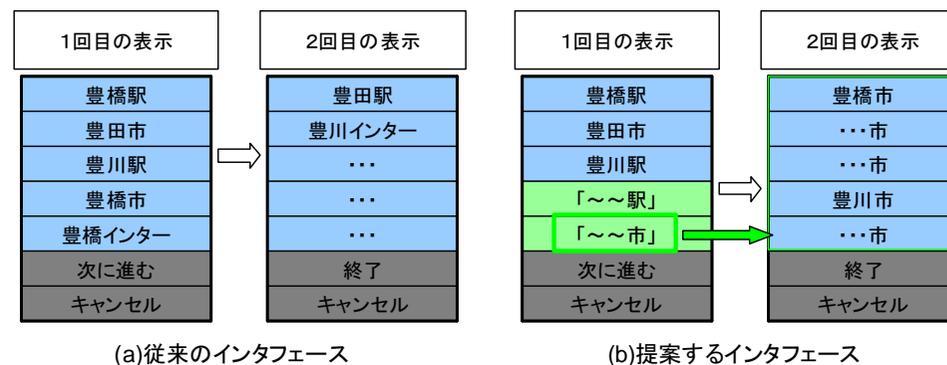
提案手法の評価実験としては、あらかじめ録音された音声の認識結果を用い、比較するシステムが実際に生成する候補リスト出力に基づいて、実際のユーザの行動を想定したシミュレーションの評価実験を行った。その結果、提案法は共通部分単語の規模によらず効果を示した。また、提案するインタフェースにより、ユーザの満足度を向上させる可能性が示され、さらに音声認識システムの認識精度が異なる様々な条件において頑健性を持つことが分かった。

2章では、複数候補を提示する典型的な音声入力インタフェースを紹介する。そして、孤立単語の音声入力タスクに対するベースラインシステムの性能も示す。共通部分単語の抽出と検索候補の生成については3章で述べ、共通部分単語の定義や検索候補を含む候補リストの構成法について説明する。4章では評価実験によるベースラインシステムとの性能比較結果を示す。

## 2. 複数候補を提示する音声入力インタフェース

3章で述べる提案法は、音声認識システムの出力として尤度スコアの順で順位付けられたN-best 認識候補を利用する。複数候補を提示する最もシンプルな方法は、図1(a)に示すように複数の候補がN-best 認識尤度の順に表示され、GUIを通してその候補が選択できるというものである<sup>4)5)</sup>。表1は4章で述べる評価実験のタスクにおいて、提示する単語候補数と正解の発話単語が候補リストの中に含まれる割合の関係を示している。

この結果から分かるように、高い雑音レベル（低いSNR）の環境下では100候補を提示しても正解の発話単語が含まれる割合が高くない。つまり、大語彙のタスクにおいて認識精度が十分でない場合は、複数候補を提示するユーザインタフェースが必ずしも効果的に機能しない。



(a)従来のインタフェース (b)提案するインタフェース  
図1 複数の候補を表示する GUI 併用の音声入力インタフェース

Fig.1 Example of GUI-assisted speech interfaces displaying multiple candidates

表1 異なる雑音レベルにおける表示候補数と正解含有率の関係  
Table 1 The performance of baseline system in different noise level

表示候補数	SNR=+10[dB]		SNR=+15[dB]		CLEAN	
	入力成功率 (平均所要表示回数)	入力成功率 (平均所要表示回数)	入力成功率 (平均所要表示回数)	入力成功率 (平均所要表示回数)	入力成功率	入力成功率
5	51.87% (1.00)		90.52% (1.00)		99.75%	
10	53.87% (1.04)		91.77% (1.01)		100.00%	
15	56.11% (1.12)		93.02% (1.04)		100.00%	
20	56.86% (1.15)		93.77% (1.06)		100.00%	
100	63.34% (2.09)		96.51% (1.29)		100.00%	

## 3. 検索候補を導入した複数候補の提示手法

### 3.1 新しいユーザインタフェースの概要

本章では、2章で述べたような複数候補の提示において、検索候補を導入した新しいインタフェースシステムについて説明する。拡張したインタフェースシステムは、ユーザが入力した発話の単語<sup>\*1</sup>を選択できるのに加え、発話した単語の部分のみが検索候補として提示されたとき、その検索候補を選択することによって候補の絞り込みを可能とする。例えば、ユーザが「～～市」という検索候補を選択すると、提案するインタフェースは次の画面に「市」で終わる単語候補だけを提示する（図1の(b)）。

\*1 本稿で「単語」とは、想定するタスクにおいて一発話単位として定義される単位のことであり、複合名詞としてのキーワードを含む。

提案するシステムが利用する検索候補は、想定するタスクの語彙から事前に作成した共通部分単語の集合から、N-best の信頼度スコアに基づき選択される。共通部分単語はシステムの単語辞書において複数の単語が共通に持つ単語の断片を意味し、単語辞書が用意されれば事前且つ自動的に作成される。本研究の評価実験(4章)で用いる施設検索タスクの場合、共通部分単語として市、駅、町のようなランドマーク名に共通する語尾部分などが含まれている。

複数候補リストの一部として表示される検索候補は、発話した単語とは一対一対応ではないが、認識候補の中から検索候補と同じ共通部分単語を持つ単語候補だけに絞り込む役割を持つ。前述した例では、「~~市」という検索候補を選択すると、次の画面からは、「~~市」を共通部分単語として持っている単語候補のみが出力される。

想定するタスク語彙から共通部分単語を抽出する方法については次節で述べる。一方、検索候補をどのように候補リストの中に挿入するか、つまり候補リストの構成法については3.3節で説明する。

### 3.2 共通部分単語の抽出法

3.1節で述べたように、共通部分単語は複数の単語が共通に持っている単語断片を意味し、システムの単語辞書から事前に抽出される。ユーザ側から見れば、共通部分単語は検索候補として提示されるので、ユーザインタフェースとしてはそれ自体が分かりやすい単位であることが望ましい。そこで、形態素を共通部分単語の基本単位として、以下に示す手順で単語辞書から自動的に共通部分単語を抽出する。

- 単語辞書の各単語(キーワード)に対して形態素解析を行い、全ての単語を形態素単位に分割する
- 形態素の出現頻度を求め、単語の先頭部分や末尾部分、もしくは中間部分における任意の形態素列の出現頻度が3回以上で文字列の長さが2以上であるものだけを抽出する
- ある共通部分単語が他の共通部分単語の一部であり、出現頻度が同じであれば削除する

図2は単語辞書から共通部分単語を抽出するオフライン処理の流れである。上述したアルゴリズムで抽出した共通部分単語は次節で述べるように検索候補として複数候補を提示するインタフェースに用いられる。

### 3.3 検索候補を含む候補リストの構成法

提案するインタフェースシステムは、図1(b)に示したような複数候補を提示するGUIを想定し、音声認識器のN-best出力から得られる単語(キーワード)候補と共通部分単語の集合から得られる検索候補とで構成される。この提案法の有効性は、一度に表示される限られた長さの候補リストとしてどのような単語候補及び検索候補を提示するかによって大きく変動する。本論文では、一回で表示できる候補リストのサイズLが与えられたとき、 $N_K$

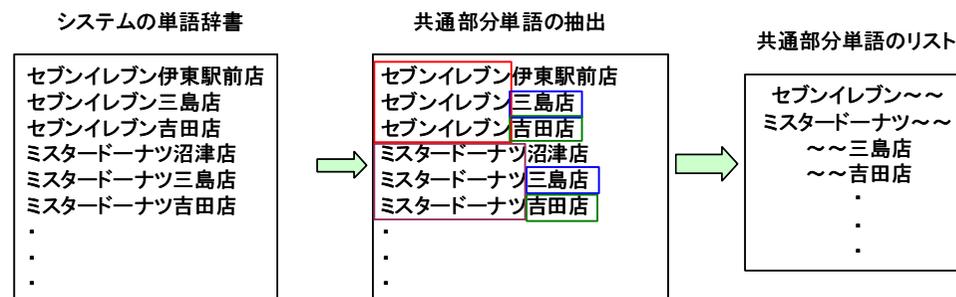


図2 共通部分単語の抽出過程の流れ  
Fig.2 Flow of extracting common word parts

個の単語候補と  $N_S (= L - N_K)$  個の検索候補からなる候補リストを決定する方法として、2通りの評価基準に基づくアルゴリズムを考える。一つは候補リストにおける信頼度スコアの合計を最大化するアルゴリズムであり、もう一つは想定されるユーザの操作回数の期待値を最小化するアルゴリズムである。

まず、共通部分単語に対する信頼度スコアを定義する。音声認識器の出力である順位付けられた単語候補(N-best)集合  $\mathbf{W}_R = W_1, W_2, \dots, W_N$  が与えられたとき、ある共通部分単語  $w$  の信頼度スコアを次式のように定義する<sup>3)</sup>。

$$C(w) = \frac{\sum_{i=1}^N P_A(X|W_i)\delta(W_i, w)}{\sum_{i=1}^N P_A(X|W_i)} \quad (1)$$

ここで、 $P_A(X|W_i)$  は発話  $X$  に対する単語  $W_i$  の音響尤度であり、共通部分単語  $w$  が単語  $W_i$  の部分である場合は  $\delta(W_i, w) = 1$ 、そうでなければ  $\delta(W_j, w) = 0$  とする。

語彙単語や共通部分単語の信頼度スコアの計算が行われると、後述のアルゴリズムでは信頼度スコアが高い順に単語候補や検索候補として選択する。そこで、式(1)で順位付けられた単語候補と検索候補からなる候補リストの構成法としては様々な方法が考えられるが、本論文では以下の2つのアルゴリズムを考える。

#### (1) 信頼度最大化基準による構成法

候補リストにおける候補の信頼度スコアの合計を最大化するアルゴリズムである。以下に示すアルゴリズムは、単語候補を  $N_K$  個含む候補リストにおいて次式で示される信頼度スコアの合計  $S_{conf}(N_K)$  を最大化するような単語候補集合  $\hat{\mathbf{W}} = \hat{W}_1, \hat{W}_2, \dots, \hat{W}_{N_S}$  を近似的に求める手順に基づいている。

$$S_{conf}(N_K) = \sum_{k=1}^{N_K} C(W_k) + \sum_{s=1}^{N_S} C(w_s) \quad (2)$$

ここで、信頼度最大化基準による構成法（以下、信頼度最大化法と呼ぶ）は以下のようなアルゴリズムで実行される。

- $t$  回目 ( $t = 1$ ) の候補リストを生成するとき、以下の内容を実行する。
1.  $m=1$  if  $t = 1$ ,  $m=0$  if  $t > 1$ .
  2.  $k=m \sim L$  に対して、ステップ 3 を実行する。
  3.  $CandList(k) \leftarrow \emptyset$ .
    - (a)  $CandList(k) \leftarrow CandList(k) \cup W_i$  for  $i = 1, 2, \dots, k$ : 単語候補を追加。
    - (b)  $W_R \leftarrow W_R - W_i$ : 選択された単語候補を除く。
    - (c) 以下の過程を  $N_S = (L - k)$  回実行する。
      - (c1)  $W_R$  に対応する全ての共通部分単語  $w$  に対して、信頼度  $C(w)$  を計算する。
      - (c2)  $CandList(k) \leftarrow CandList(k) \cup \hat{w}$ , ここで  $\hat{w} = \underset{w}{\operatorname{argmax}}(C(w))$ 。
      - (c3)  $W_R \leftarrow W_R - W_i (\forall W_i \supset \hat{w})$ : 検索候補を部分として持つ単語候補を削除。
    - (d) 式 (2) により、信頼度スコアの合計  $S_{conf}(k)$  を求める。
  4.  $\hat{k} = \underset{k}{\operatorname{argmax}} S_{conf}(k) (m \leq k \leq L)$  を求め、  
単語候補が  $\hat{k}$  個挿入された候補リスト  $CandList(\hat{k})$  を採用する。

上記のアルゴリズムの (c3) のステップは、選択された検索候補  $w$  によって絞り込み可能となる単語を、この後に生成する候補リストに単語候補として含めないようにするためのものである。

## (2) 表示回数最小化基準による構成法

候補の信頼度スコアを正解確率とみなして推定する候補リストの表示回数の期待値を最小化する最適アルゴリズムである。例えば、0.5 の信頼度スコアを持つ単語候補は、0.5 の確率で画面表示は 1 回で終わると推定される。もし、表示された単語候補の中に正解が含まれなかった場合は、検索候補が該当無しの状態となり、より多い表示回数が必要となる。そこで、単語候補が  $N_K$  個挿入された候補リストにおいて、次式で推定する表示回数の期待値を最小化するような単語候補及び検索候補を近似的に求めるアルゴリズムを考える。

$$S_{count}(N_K) = C_1 \sum_{k=1}^{N_K} C(W_k) + C_2 \sum_{s=1}^{N_S} C(w_s) + C_3 \left( 1 - \sum_{k=1}^{N_K} C(W_k) - \sum_{s=1}^{N_S} C(w_s) \right)$$

ここで、各変数の意味は「信頼度最大化法」と同様であり、 $C_1, C_2, C_3$  は定数で、それぞれ「単語候補選択」、「検索候補選択」、「該当なし」のケースで予想される操作回数である。この定数は  $C_1 = 1$  となり、 $C_2$  と  $C_3$  は本研究は経験的にそれぞれ 2, 4 に設定した。その時、表示回数最小化基準による方法（以下、表示回数最小化法と呼ぶ）は以下のようなアルゴリズムで実行される。

- $t$  回目 ( $t = 1$ ) の候補リストを生成するとき、以下の内容を実行する。
1.  $m=1$  if  $t = 1$ ,  $m=0$  if  $t > 1$ .
  2.  $k=m \sim L$  に対して、ステップ 3 を実行する。
  3.  $CandList(k) \leftarrow \emptyset$ .
    - (a)  $CandList(k) \leftarrow CandList(k) \cup W_i$  for  $i = 1, 2, \dots, k$ : 単語候補を追加。
    - (b)  $W_R \leftarrow W_R - W_i$ : 選択された単語候補を除く。
    - (c) 以下の過程を  $N_S = (L - k)$  回実行する。
      - (c1)  $W_R$  に対応する全ての共通部分単語  $w$  に対して、信頼度  $C(w)$  を計算する。
      - (c2)  $CandList(k) \leftarrow CandList(k) \cup \hat{w}$ , ここで  $\hat{w} = \underset{w}{\operatorname{argmax}}(C(w))$ 。
      - (c3)  $W_R \leftarrow W_R - W_i (\forall W_i \supset \hat{w})$ : 検索候補を部分として持つ単語候補を削除。
    - (d) 式 (2) により、信頼度スコアの合計  $S_{count}(k)$  を求める。
  4.  $\hat{k} = \underset{k}{\operatorname{argmax}} S_{count}(k) (m \leq k \leq L)$  を求め、  
単語候補が  $\hat{k}$  個挿入された候補リスト  $CandList(\hat{k})$  を採用する。

上記のアルゴリズムにより選択された単語候補と検索候補は、正解の発話（もしくはその一部）を含む可能性が高くなると期待される。一般的に検索候補の導入は最初の画面表示において発話の単語全体を探す効率を低下させるが、検索候補の提示によって結果的にはシステムの操作回数に関する入力効率を向上させ、ユーザビリティの改善も期待される。

## 4. 評価実験

この章では、3 章で提案した音声インタフェースシステムとベースラインシステムの性能を比較する評価実験を行う。また、実験条件の変化による提案法の性能変化についても述べる。

### 4.1 タスクおよび実験条件

本研究では、カーナビゲーションシステムでのユーザインタフェースを想定し、共通部分単語の多様性が得られる施設検索タスクで評価実験を行う。施設検索タスクは 12,845 個の施設名の単語を持ち、静岡県におけるランドマークや様々な施設、店舗などが含まれている。

また、3章で述べたアルゴリズムによって単語辞書から事前に抽出された共通部分単語の種類は、それぞれ先頭部分 430 種類、末尾部分 130 種類、中間部分が 1220 種類であった。

実験に用いた評価用音声データは 4 話者が異なる単語を発話した 401 発話であり、全てクリーン環境下で録音された孤立単語である。しかし、実環境の条件を模擬し、認識精度の差をもたらす音響環境での効果の検証を示すため、雑音を重量して SNR+10[dB], SNR+15[dB] に加工し、さらに各テスト発話に対してスペクトルサブトラクションを行った。一方、本論文における全ての実験には、HMM ベースの連続音声認識器である SPOJUS を利用している<sup>5)</sup>。音響モデルの HMM は、124 種類の音節カテゴリからなる。

#### 4.2 実験・評価方法

評価方法として、オフラインでの音声認識結果を用い、ユーザが音声インタフェースを操作すると想定したシミュレーション実験を行う。模擬する状況設定は次の通りである。まず、ユーザは音声を入力し、発話した単語を候補リストから探そうとする。そして候補リストの中に発話した単語（単語候補）もしくは単語の一部（検索候補）が現れたら、ユーザはその候補を選択する。もし、どの候補も該当しなければ、ユーザはシステムに次の画面を表示させ次の候補リストを参照する。

本論文の評価基準としては、1 回分の提示する候補リストの長さを 5 個に固定し、ある回数までの候補リストの表示に制限した場合の入力成功率と平均所要表示回数を考える。入力成功率とは、シミュレーション実験においてユーザが音声を入力し、その発話した単語を限られた表示回数の候補リストの中から発見できた割合を意味する。最初から単語候補を直接選択しても、検索候補で絞り込みその後の候補リストで単語候補を選択しても同じ入力成功と見なされる。一方、平均所要表示回数とは、シミュレーション実験において、ユーザが入力に成功した場合の候補リストの表示回数の期待値を意味する。シミュレーション実験では、発話単語自身が選択されたら候補リストの表示を終了し、発話した単語自身が現われるまでユーザがシステムの操作を繰り返すと想定する。よって、平均所要表示回数の期待値は入力成功率に影響され、入力効率が同程度の場合、この期待値が小さいほどユーザが正解の発話を早い段階で発見できることを意味する。つまり、平均表示回数が少ないシステムが良い性能のシステムだと言える。

評価するシステムは 3 種類であり、単語候補のみを提示する従来の方法、検索候補を併用する提案法（信頼度最大化法と表示回数最小化法）の性能を比較する。また、共通部分単語の抽出における中間部分の導入による影響や、画面に表示する候補リストの候補数の変化による性能変化も比較する。

#### 4.3 実験結果

まず、各 SNR 値におけるベースラインシステムの入力成功率及び平均表示回数の期待

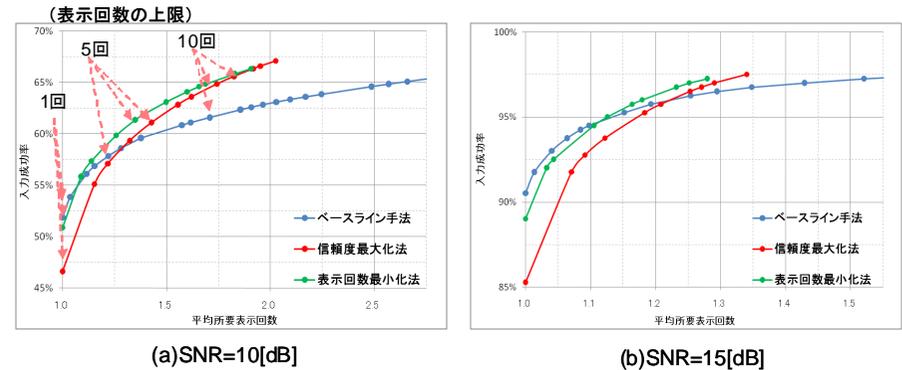


図 3 音声インタフェースシステムの性能比較（ベースライン及び 2 つの提案法）  
Fig. 3 Performance of the speech interface systems (baseline and proposed method)

値を表 1 に示す。たとえば、SNR=10[dB] において表示候補数 20 のとき、入力成功率が 56.86% ということは、候補リストの長さを 5 個、許容される候補リストの最大表示回数を 4 にしたとき、シミュレーション実験によってユーザが入力した単語自身を見つける割合が 56.86% であることを意味する。また、平均所要表示回数の期待値が 1.15 ということは、ユーザが入力単語自身を候補リストから見つけた場合に、画面に候補リストを表示させる回数の平均が 1.15 回であることを意味する。各 SNR 値による結果を見ればわかるように、入力成功率が高いほど、平均所要表示回数は減少し、ユーザにとってはより良いシステムであると判断できる。

次に、提案するインタフェースシステムとベースラインシステムの性能比較を行った結果を示す。図 3 は SNR=10[dB], SNR=15[dB] における性能比較であり、横軸は平均所要表示回数、縦軸は入力成功率である。また、折れ線の各点は、左下から順に候補リストの表示回数の上限を増やしていった場合の結果に対応する。同じ入力成功率であれば所要表示回数が少ないシステムが、同じ表示回数であれば入力成功率が高いシステムが良いシステムだと評価できる。

図 3 の実験結果を見ると、最初の画面表示では提案法の性能はベースラインより落ちている。これは、従来手法と違い提案手法は単語候補の代わりに検索候補を挿入するため、入力成功率がその分低下するからである。しかし、許容される表示回数が増えるにつれ、提案法の性能はベースラインを追い越し、入力成功率も表示回数の期待値もベースラインより良くなっていく。例えば、SNR=10[dB] において 60% の入力成功率を得るため、表示回

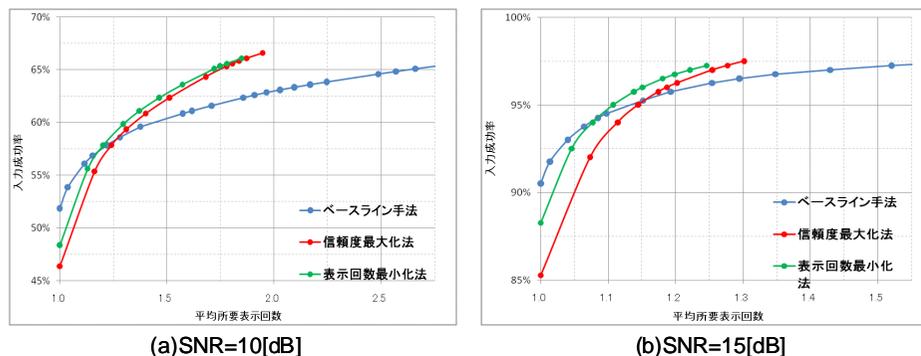


図 4 音声インタフェースシステムの性能比較 (ベースライン及び 2 つの提案法, 中間部分導入)  
Fig. 4 Performance of the speech interface systems (baseline and proposed method, using middle parts of words)

数最小化法は約 1.25 回, 信頼度最大化法は約 1.35 回, ベースライン手法は約 1.45 回の表示回数が必要となる。全体的に見ると, 提案手法は上限の画面表示回数が 2~4 回に増えたとき, ベースライン手法の性能を上回る。一方, SNR=15[dB] においては, 提案手法の性能がベースライン手法を上回る時点が遅くなり, 5~7 回以上上限の画面表示回数が増えたとき追い越し始める。また, 2 つの提案法間の性能を比較すると, 表示回数最小化法の折れ線が信頼度最大化法より左側に位置し, 表示回数最小化法のほうがより効率が良いと考えられる。

一方, 共通部分単語を抽出するとき, 単語の前方と後方だけでなく, 中間部分も抽出した結果を図 4 に示す。図 3 と図 4 の結果を比較すると, 共通部分単語における中間部分の導入によって, グラフの傾きが少し変化したのが分かる。表示回数の上限が低いときは提案手法の入力効率が低下したが, 表示回数の上限が増えると入力効率が中間部分を利用しなかった場合より良くなっている。特に, SNR=15[dB] において中間部分の導入による有効性が明らかになっており, 上限 5~7 回であった性能の逆転時点が上限 3~5 回に低くなっているのが分かる。

最後に, 画面に表示する候補リストの候補数を 10 個に変更したときの実験を行った結果, 候補リストの候補数が 5 個であった場合の結果と同じ傾向を表し, 同じく提案手法の有効性が示された。この場合も表示回数最小化法のほうが良い入力効率を示した。

## 5. ま と め

本論文では, 複数候補提示のユーザインタフェースシステムにおける入力効率を改善するため, 検索候補と共通部分単語を定義し, 複数候補提示において絞り込みを可能とする検索候補を導入した新しいインタフェースシステムを提案した。また, 評価実験を行い, 複数候補を提示する従来のインタフェースシステムと検索候補を利用する新しいインタフェースシステムの性能を比較した。

その結果, 検索候補および検索候補を含む候補リストの構成法を導入した提案手法は, 実環境に近い雑音条件において入力効率を向上させ, ベースライン手法を上回るシステム性能を示した。また, 検索候補を用いることによって, 認識器が認識に失敗しても部分的に成功した結果をユーザに返し, ユーザにとってはより柔軟な選択が可能になった。また, 共通部分単語における中間部分の導入は, インタフェースシステムの入力効率を向上でき, 表示する候補リストの候補数の変化にも提案手法の有効性が失われなかった。

今後の発展としては, 音声認識器の後処理ではなく, 音声認識システムにおいて共通部分単語を部分的に認識できるようにすることが考えられる。また, 対話インタフェースの一部として応用した場合のユーザビリティ評価や, 未知語の扱いとしての効果を確認することなども挙げられる。

## 参 考 文 献

- 1) 北岡敬英, 押川洋徳, 中川聖一, “ 孤立単語認識と連続基本単語認識の併用に基づく組織名の音声入力インタフェース ”, Proc. of INTERSPEECH 2005, pp.1201-1204, (2005)
- 2) Oshikawa H., Kitaoka N., and Nakagawa S., “ Speech interface for name input based on combination of recognition methods using syllable-based N-gram and word dictionary ”, Proc. ICSLP-2004, pp.177-180, (2004)
- 3) 藤原敬記, 伊藤敏彦, 荒木健治, 甲斐充彦, 小西達裕, 伊東幸宏, “ 認識信頼度と対話履歴を用いた音声言語理解手法 ”, 電子情報通信学会論文誌, Vol. J89-D, No.7, pp.1493-1503, (2006)
- 4) Itoh K., Kai A., Konishi T., Itoh Y., “ An understanding strategy based on plausibility score in recognition history using CSR confidence measure ”, Proc. ICSLP'04, pp.2133-2136, Korea, (2004)
- 5) 趙國, 宮山章子, 山下洋一, “ N-best 音声認識における認識スコアを利用した候補提示数の決定 ”, IEICE Trans., Vol. J88-D-II, pp.1003-1011, (2005)
- 6) 中川聖一, 甲斐充彦, “ 文脈自由文法制御による One Pass 型 HMM 連続音声認識法 ”, 電子情報通信学会論文誌, Vol. J76-D-II, pp.1337-1345, (1993)