

論文 / 著書情報  
Article / Book Information

論題(和文)	クロス言語検索を用いた中国語音声認識による乗換案内システム
Title(English)	
著者(和文)	張 琦, 大西 翼, 篠崎 隆宏, 古井 貞熙
Authors(English)	Oonishi Tasuku, Takahiro Shinozaki, SADAOKI FURUI
出典(和文)	日本音響学会2011年春季講演論文集, , No. 2-5-7, pp. 61-62
Citation(English)	, , No. 2-5-7, pp. 61-62
発行日 / Pub. date	2011, 3

# クロス言語検索を用いた中国語音声認識による乗換案内システム\*

張 琦, 大西 翼, 篠崎 隆宏, 古井 貞熙 (東工大)

## 1 はじめに

近年中国から訪日する旅行者が大幅に増えているが、日本語を入力できないために、日本での情報検索が容易でない。本研究では、日本語がわからない中国人ユーザーが、手持ちのモバイル端末 (iPhone) で利用できるサーバークライアント型中国語音声認識による情報検索システムを構築した。現在日本で使われている漢字のほとんどは中国語にもあるため、日本語がわからない中国人ユーザーでも中国語読みで日本語にある漢字が読める。駅名の中で漢字が占める割合が高いことと、旅行者のニーズが高いことから、システムを乗換案内に特化した。従来の音声認識システムでは、通常、入力言語と出力言語が同じであるが、本研究においては、音声入力と異なる言語での検索を実現した。認識結果の N-best 候補を提示することにより検索精度を高めるだけではなく、漢字とかなが混在する駅名の入力も可能とした。

## 2 システム概要

### 2.1 サーバー、言語モデルおよび音響モデル

サーバー側に通信プログラムと WFST を利用した音声認識システム (T<sup>3</sup> デコーダ) を置く。言語モデルはネットワーク文法を用いる。音響モデルの訓練用コーパスは "King-ASR-018" で、広い年齢層の 850 人 (男性 430 人、女性 420 人) の話者、計 150 時間の音声データが収録されている。音響モデルは声調なしの 3 状態、25 次元 (MFCC とその とエネルギー) のトライフォン HMM である。

### 2.2 単語発音辞書

認識対象語彙数は 1060 である。中国語 日本語クロス言語検索を可能とするため、以下の三点を実現した。

- 1) 複数の中国語読みを持っている漢字に関する対応：複数の中国語読みを持つ漢字は、中国語のコンテキストがあれば発音がほぼ確定できるが、日本語環境では確定できないため、発音辞書に可能な発音をすべて記載した。
- 2) 中国語にはない日本語の漢字に関する対応：駅名に出てくる中国語にはない日本語の漢字に対して、予備実験で 20 名の被験者の読み方を調べ、発音辞書に追加した。

- 3) ひらがな、カタカナに関する対応：仮名は発音しない。従って品川と品川シーサイドが同じ評価スコアとなるので、音声入力後ユーザーに選択してもらうことにした。

### 2.3 クライアント

ユーザーが発声した「東京駅から吉祥寺駅まで」などに相当する中国語音声、16kHz、16bit の Linear PCM で録音し、ストリームでサーバーとリアルタイムで通信する。録音終了後、サーバーから認識結果の N-best を受け取り、最も尤度が高い結果を表示すると同時に、リスト形式ですべての候補を表示する。リストにある候補をタッチすることによって、認識結果が修正できる。発音した駅名が表示されないときは再度音声入力するか、キーボードで入力する。目的語と一致する認識結果が表示されたら、検索ボタンで経路を検索することができる。

### 2.4 インタフェース

発声に関する制約と、録音区間制御法の異なる UI A, B, C の三通りのインタフェースを検討した。UI A: 「東京駅から吉祥寺駅まで」のような文発声のみ受理する。UI B: 文音声の他、「東京駅」のような孤立単語も受理可能で、個別に駅名を修正できる。UI A と UI B はユーザーがメインボタンを押している間のみ音声を録音する。UI C: 加速度センサーでマイクを耳に近づける動作を検出し、録音開始と終了を制御する。ディスプレイ方向への 0.3G の加速度で録音を開始し、カメラ方向への 0.3G の加速度で録音を終了する (ハイパスフィルターを使い、重力の影響を除いた)。文発声のみ受理する。

## 3 評価実験

### 3.1 実験手順

周囲に人の声やキーボードを打つ音がある研究室で、中国人留学生男女各 5 名を被験者としてシステムの評価を行った。最初に、簡単なガイダンスの後、好きな駅名を使い二回練習してもらった。評価実験では、iPhone を操作しながら、プログラムがランダムに抽出した二つの駅名を使い「x 駅から y 駅まで」のように発話してもらった。インタフェースごとに 10 セット、計一人 30 セットのテストを行った。終了

\* A transfer guide system based on cross-language search using Chinese speech recognition . by Qi Zhang, Tasuku Oonishi, Takahiro Shinozaki, Sadaoki Furui (Tokyo Institute of Technology)



Fig. 1 UI B の画面の例

後、主観評価の各項目に対して5段階評価をしてもらった。

使い方がわからないときはヘルプを参照してもらい、それでもわからないときは適切に助言した(主に、iPhoneの操作になれない人と加速度センサーの仕組みがわからない被験者で、全体の2割を占めた)。

### 3.2 実験結果

#### 3.2.1 客観評価

以下の各項目 [1, 2] について評価を行い、結果を Table 1 に示す。

タスク時間：プログラムが駅名を提示した時から被験者が乗り換え検索のボタンを押すまでの時間。初回認識率：最初の発話に対する正解率。二回以上発話する確率：10 best に表示されても、ユーザーが見つけれない場合は再度発話する可能性がある。平均ターン：ユーザーの発話回数の平均。タスク達成率：音声入力でタスクを達成する確率(数回の音声入力希望の結果が出ない場合、音声入力をあきらめ、キーボードで入力するユーザーがいる。試みる回数はユーザーの判断に任せ、制限しなかった。)

タスク時間に関しては、B方式がA,Cより2秒くらい短い。Aは音声による片方の駅名修正ができないので、修正に使う発音時間が長くなるだけでなく、誤認識によって、元々正解だった片方の駅名を誤認識で上書きしてしまう可能性がある。Cは加速度センサーで開始と終了を示すので、加速度の検出(0.2秒ごとに加速度データをアップデート)に時間がかかる上、ゆっくり動かし場合に、特に終了の検出がうまくいかないことが数回あった。同じタスクを音声を使わずに両手を使ってキーボードで入力する場合は、平均時間が45.6秒で長くなり、音声入力との差が歴

Table 1 それぞれのUIに対する客観評価

	UI A	UI B	UI C
タスク時間 (s)	12.4	10.5	12.7
初回認識率	0.78	0.76	0.78
二回以上発話する確率	0.15	0.19	0.16
平均ターン	1.18	1.23	1.19
タスク達成率	0.97	1	0.99

Table 2 それぞれのUIに対する主観評価

	UI A	UI B	UI C
総合満足度	4.1	<u>4.7</u>	4.1
使いやすさ	4.3	<u>4.8</u>	4.3
レイアウト	4.4	4.4	4.5
レスポンス	4.8	4.7	<u>4.2</u>
修正のしやすさ	4	<u>5</u>	4.1

然である。

二回以上発話する確率と平均ターンの項目で、Bの値はA,Cより大きい。個別の駅名修正ができるので、リストで目的語を探すより再度発話することで目的語を入力するを好むユーザーがいた。候補リストが長い場合にこの傾向が強くなる。

#### 3.2.2 主観評価

使いやすさなどに関する主観評価 [2, 3] の結果を Table 2 に示す。有意水準5%で両側検定を行い、最大値もしくは最小値であり、かつ他のインタフェースと有意差が認められるものを下線で示した。総合満足度、使いやすさ、修正のしやすさともB方式がA,Cより高い評価を受けた。客観評価と合わせると、誤認識を簡単に修正できるか否かが、タスク時間、タスク達成率、総合満足度に大きく影響することがわかる。Cのレスポンスが低い評価を受けているのは、発声終了の検出ミスが原因である。

## 4 まとめ

中国語音声認識による日本の乗換案内システムを構築し、入力言語と出力言語が違う音声認識システムの実用性を示した。3種類のインタフェースの評価実験を通して、修正しやすいインタフェースが重要であることが確認された。

## 参考文献

- [1] 石川他, 音響学会誌 61-2, pp.79-84, 2005.
- [2] J.Novak *et al.*, 日本音響学会 2010 年春季講演論文集, 1-Q-23, pp.209-210, 2010.
- [3] J. Larson, Ten Criteria for Measuring Effective Voice User Interfaces, <http://www.larson-tech.com/Writings/10UICriteria.htm>, 2005.