

J-006

# 時空を越えたコミュニケーションを実現する音声インタフェースの提案

## Proposal of Speech Communication Interface Covering Time and Space Differences

田島 大嗣†  
Hirotsugu Tabata

久保 陽太郎†  
Yotaro Kubo

樽松 明†  
Akira Kurematsu

白井 克彦†  
Katsuhiko Shirai

### 1. はじめに

音声は、我々が人同士のコミュニケーションに用いる身近な存在であり、最も手軽な情報伝達手段であるため、誰にでも使いやすいインタフェースへの利用として注目を集めている。音声を利用したアプリケーションの従来研究としては音声対話システムがあげられる。例えば、交通案内システムや天気案内システムなどである。これらの音声対話システムでは、ユーザとシステム間でのやり取りを対話で行っている。また、前提として正確な音声認識が求められていた。

一方、本研究では、より人と人とのコミュニケーションに近いような状況を想定し、ユーザとユーザとの間にシステムが介在し、楽しくコミュニケーションをとる新たな音声インタフェースの提案を行う。さらに、多少の誤認識を容認させることや音声インタフェースに空間情報や時間情報を組み合わせることで音声インタフェースに新たな付加価値をつけ、時空を越えたコミュニケーションを実現することを目指す。

空間情報の取得にはGPS[1], [2] やICタグ[3]などがあげられる。しかしICタグの場合、空間情報に関しては正確に得ることができる反面、対象地域全体に敷設する必要があるといった問題点がある。それに対してGPSは大規模な設備を必要とせず空間情報の取得を行うことができるため、GPSによって行う方式を採用する。

### 2. 提案システム

ユーザがGPSを搭載したモバイルPCを利用しながら空間を楽しんでもらう。対話空間で発生した対話内容を連続音声認識し、その結果を分析することでキーワードの抽出を行う。このキーワードとGPSから得た位置情報との関連付けた情報を空間情報とし、サーバ上で他の対話空間と共有できるようにする。さらに、提案システムを通して時間軸的に遠い対話空間のユーザがそれらの空間情報を知ることにより対話をさらに促進させることができる。

従って、システムは現在利用しているユーザの声に対して直接フィードバックを行うのではなく、過去のユーザが同一地点で話した内容をフィードバックする。つまり、ユーザの感覚としては、時間軸的には遠いが、同一場所で多くの人が感じた自然な声を聞くことにより、多くの人と同時にコミュニケーションをとっているような感覚を味わうことができる。

システムを利用した際の対話情報の流れを図1に示す。

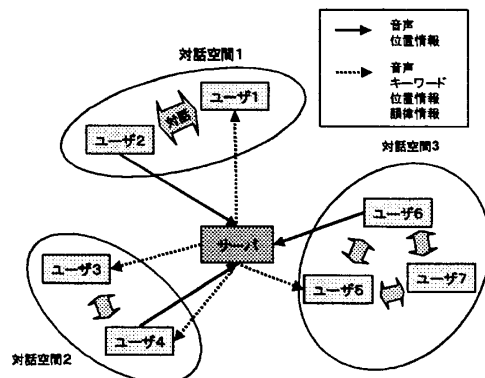


図1 システムを利用した際の対話情報の流れ

#### 2.1 提案システムにおける情報提示の方法

抽出したキーワードは、文字情報をブラウザ上に出力すると共に音声データを出力することにより、ユーザに情報提示を行う[4]。さらに、現在地以外でも気になる地点を指定すると、現在地の際と同様に空間情報を提示するとともに音声出力も行う。

#### 2.2 キーワード抽出

音声認識の結果をそのまま情報提示すると、誤認識などにより文法的に破綻していたり、意味が通じない場合が考えられる。従って、下記手法によりキーワード抽出を行う。音声認識により得た認識結果を単語毎にわけ、式(1)のように各単語の全体に対する出現頻度と各発話地点を中心とした一辺4秒四方(縦約120m×横約100m)の範囲に対する出現頻度を計算する。

$$\text{出現頻度} = \frac{\text{各単語のカウンタ数}}{\text{単語数の合計}} \quad \dots (1)$$

さらに、それらの値を式(2)に代入し、品詞が名詞もしくは形容詞である単語の中で、有効度の上位5個をキーワードとして提示する。つまり、全体でよく出現する単語をキーワードとするのではなく、各地点特有の単語をキーワードとする。

$$\text{有効度} = \frac{\text{範囲内の出現頻度}}{\text{全体の出現頻度}} \quad \dots (2)$$

ただし、単語のカウンタ数が3回以下の単語に関しては、誤認識の可能性が極めて高いため有効度が高いとしてもキーワードから除外する。また、認識尤度の上位3個を音声認識結果として利用することでキーワードの取りこぼしを軽減させる。

このような手法を用い情報提示を行うことで、場所の特徴(臭い, 広い, 高い, うるさいなど)を知ることができ、音声認識の問題点である誤認識やGPSの問題点である場所による位置情報の誤差などにも対応できる。

†早稲田大学理工学術院, Waseda Univ.

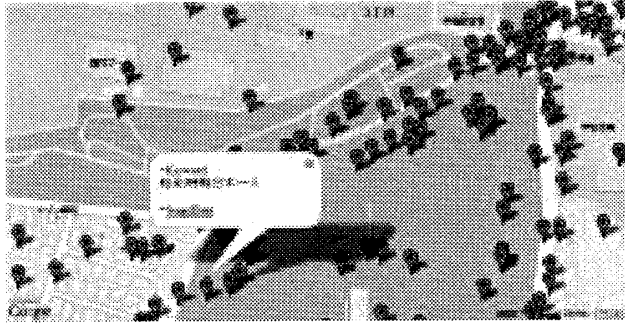


図2 システムの使用イメージ

表1 実験環境

場所	早稲田大学理工学部周辺
被験者	20代男性5名, 20代女性2名 計7名
GPS受信機	VGP-BGU1 (SONY)
通信方式	D02HW (emobile)
マイク	接話マイク
音声認識エンジン	Julius
音響モデル	CSJ 付属の音響モデル (学会+模擬)
言語モデル	CSJ 付属の言語モデル (学会+模擬)
総発話数	1428 文
延べ単語数	15350 個

表2 実験結果

point	keyword				
	構成	頃	高い	場合	ない
箱根山					
カレー屋 (ヤミツキ)	ライブ	病み付き	おいしい	部分	凄
戸山公園内	桜	和	時期	日本	一人

### 3. 予備実験

現状のシステムを用い、データ収集および音声認識やGPSの精度の確認を目的として行った。さらに、抽出されたキーワードの有効性についての確認も行った。

#### 3.1 システムの内容

各地点における提示内容は、5個のキーワードと音声である。また“SoundData”にリンクが張られておりクリックすると、キーワード抽出の際限定した範囲内で録音された音声ランダムに再生される。システムの使用イメージを図2に示す。

#### 3.2 実験環境

被験者は20代男性5名、および20代女性2名の計7名とし、本システムを搭載したモバイルPCを用い、早稲田大学理工学部周辺で予備実験を行った。また、音声認識エンジンはJuliusを使用し、日本語話し言葉コーパス(CSJ) [5]を用いて学習した音響モデルと言語モデルを利用した。総発話数は1428文、延べ単語数15350個である。実験環境を表1に示す。

### 3.3 考察

実験結果の例を表2に示す。この結果からわかるように、地図に載っていないような情報(高い・桜など)をキーワードとして提示できることがわかった。また、同様に感情(おいしい・凄)も抽出できた。

ただし、カレー屋の店名である「ヤミツキ」を「病み付き」となっていたことから、ローカルな固有名詞を言語モデルに学習させる必要があることがわかった。

## 4. 今後の方針

ローカルな固有名詞を言語モデルに学習させた上で、本実験を行う。ただし、実験場所は大きく違うものの助詞などのキーワードとして適正ではないものは同様であると考えられるため、これまで採取したデータは引き続き使用する。

### 4.1 本実験

今回の予備実験では、早稲田大学理工学部周辺で行ったが、本実験では、恩賜上野動物園を実験場所と考えている。理由としては、環境面で周りに高い建物が少ないためGPSの精度が良くなることがあげられる。さらに幅広い世代が楽しめる場所ということもあり、より自然な感情が表れやすいというメリットがある。

評価方法に関しては、下記の4条件のマップを用いて実験し、主観評価のアンケートをとる。評価軸は、「有効性」「面白さ」「便利さ」を10段階評価し、これらのスコアで比較を行う。

- ・ 現在地のみを示すマップ
- ・ 認識結果を加工せず空間情報として提示するマップ
- ・ キーワードを空間情報として提示するマップ
- ・ キーワード+音声を空間情報として提示するマップ

### 4.2 検討課題

音声には感情が文字に表れていなくても、韻律などに表れてくることがある。これらの特徴を分析することで、ユーザが次の目的地を選ぶための誘導を行うことも検討する。また、キーワードを抽出する際に用いている範囲についても適正範囲の検討を行う必要がある。

## 参考文献

- [1] 斉藤和樹, 三上紀子, “観光行動研究における観察データ管理システムの開発,” 地理空間情報フォーラム, 2006.
- [2] 西本昇司, 横尾佳余, 垂水浩幸, “栗林公園モバイルガイドシステムの公開実験,” 電気情報通信学会, 進学技法, CS2006-1.
- [3] 佐々木一郎, 香川考司, 垂水浩幸, “SpaceTag システムの評価実験,” 情報処理学会, 情報処理学会論文誌, Jan. 2004.
- [4] 翠輝久, 河原達也, 正司哲朗, 美濃導彦 “質問応答・情報推薦機能を備えた音声による情報案内システム,” 情報処理学会, インタラクション 2007.
- [5] 日本語話し言葉コーパス: <http://www.kokken.go.jp/>