

E-073

適応的言語処理を用いた音声認識利用による楽曲検索システムの検討 A Design of Retrieval System Using Speech Recognition with Adaptive Processing

大山 博延† 樽松 明† 白井 克彦†
Hironobu Oyama Kurematu Akira Shirai Katuhiko

1. はじめに

近年、ネットワーク上に大量のデータを保存することが可能になってきたことと同時に、HDDを搭載したカーナビゲーションやiPodに代表される大容量ミュージックプレイヤーの普及、携帯電話のような端末での音楽配信サービスの開始により、その大量の楽曲の中から必要な情報を検索する機会が多くなってきた。このように大量の楽曲の中から好きな曲が選択可能な状況になると、自分の目的の曲を検索し選択することに手間がかかるようになる。さらに、これらの楽曲を扱う端末の小型化やハンズフリー化が利便性の高まりによって進んでいる。その結果、従来のキーボードのような入力インタフェースに変わり、音声を入力インタフェースに利用する必要性も高まってきた。このような現状のなか様々な音声入力インタフェースを利用したシステムが検討され、研究が進められている。例えば、音声による語彙フリー多言語マルチメディア検索システム[1]や、車の中の楽曲検索を想定した対話システム[2]、音声による楽曲検索システム[3]などが挙げられる。これらの研究では、音声認識技術に着目し音声入力インタフェースのみでの入力を想定してハンズフリー化の実現を検討している。しかし、実用化までには音声認識の精度が低いという問題や、ユーザビリティが不十分であるため、効率的な検索が行えないといった問題がある。現在実用化されている様々なシステムでも同様の問題があるため、需要があるにもかかわらずいまだに普及しているとは言えないのが現状である。

本研究ではこれらの問題を改善することで、音声によって円滑に楽曲検索を行えるようなシステムの実現を目指す。

2. 提案手法

2.1 適応的言語処理

音声認識精度が低いといった問題は、認識する対象が楽曲特有の単語である事が挙げられる。楽曲検索をするような場面では大量のデータの中から楽曲を検索する事になるため、似たような曲名であったり短い曲名であったりする場合、正確に楽曲を特定する事が難しい。そこで、対象とする言語モデルをシステムの実行途中で限定する適応的な言語処理を行うことにより認識精度の向上を図る。

2.2 システム側のフィードバックによるユーザビリティの向上

ユーザビリティが不十分であるため、効率的な検索が行えないといった問題は、ユーザがシステムの認識状況を把握できないということに起因している。そこで、システム側の状態をユーザが把握できるような工夫をすることで、次にどんな入力をすれば自分の目的の楽曲が効率的に検索できるかをユーザに分かりやすくすることで改善を図る。

3. システム

3.1 システム概要

システムは、「音声入力部」「音声認識部」「リスト作成部」「結果出力部」「曲の選択」の5つの部分で構成されている。

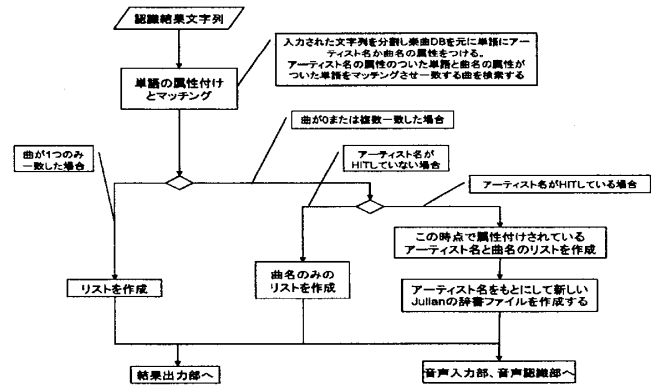


図1 リスト作成部概要

この中で、「音声入力部」と「曲の選択」はユーザが手動で行うことにし、その他の三つの手順をプログラムによって動作するようシステムを構成した。

3.2 音声認識部

音声認識ソフトとしてJulian[4]を使用した。Julianは有限状態文法に基づく連続音声認識を行うことのできるソフトで、タスクの決まった音声認識に対して高精度な認識を行うことができる。認識結果は尤度が高いものから5つ候補が上がったものを次の「リスト作成部」に送る。尤度は、言語モデルと音響モデルを統合して算出した数値である。

3.3 リスト作成部

リスト作成部の概要を図に示す(図1)。

このリスト作成部で本研究の提案手法の一つである「適応的言語処理の利用」を行っている。

まず、音声認識結果から形態素解析された単語に対し、楽曲データベースとマッチングさせ、マッチしたものにアーティスト名または曲名の属性を付ける。この属性付けが終わった後、属性の付いた単語の組み合わせを実際のアーティスト・楽曲情報と照らし合わせ、それぞれが一致するものを検索する。その後条件分岐によりシステムの振る舞いを決める。

3.4 結果出力部

「結果出力部」で本研究の提案手法のもう一つの「システム側のフィードバックによるユーザビリティの向上」を行っている。

「リスト作成部」から送られてきたリストを画面に表示する。条件に応じて「曲の選択」、または「音声入力部、音声認識部」に移行する。ここで曲が一つに絞れていない場合でもリストを表示する。そうすることによってシステム側の認識状況を伝え、ユーザに次の動作をうながしている。

4. 実験

4.1 実験概要と目的

初期状態での音響モデルと言語モデルが同じで、本研究で作成したシステムを使用する場合と、初期の音響モデルと言語モデルのみを用いて適応的処理を使用しない場合に対して、楽曲の検索を行い、結果を比較することによって、作成したシステムの評価を行った。

† 早稲田大学理工学術院

評価は「単語認識率」「タスク達成率」「ユーザビリティ」の三点である。「単語認識率」の評価実験は、提案手法の一つである適応的言語処理の有効性を評価するために行った。「タスク達成率とユーザビリティ」の評価実験は、作成したシステムによって楽曲検索をした場合のタスク達成率、その楽曲検索のタスクを行ったときのシステムのユーザビリティを調べることで、提案手法の適応的言語処理とシステム側のフィードバックによるユーザビリティがどの程度実現できたかを評価するために行った。さらに、対象とする楽曲数を489曲から1000曲に増やし、それぞれの実験を行い楽曲数を増加させたときの評価を行った。

4.2 単語認識率

楽曲検索システムのために「アーティスト名」「曲名」を認識できるように文法と辞書を定義した実験データは、事前に録音した音声認識させた。アーティスト名と曲名が認識できた場合を「認識成功」、どちらかが成功した場合を「～の認識成功」、両方認識失敗した場合を「認識失敗」とし、全体に対するそれぞれの割合を各認識率の基準の値とした。ここからさらに認識結果を認識が失敗したものだけに限り、本研究で作成したシステムを使用しもう一度認識を行い、上記の方法と同じように認識率を出し基準の値と比較した。

4.3 タスク達成率とユーザビリティ

タスク達成率は1タスクにつき最大3回楽曲検索を行い、目的の曲が「何回目」で検索できたか、または3回以内にできなかった場合を「認識失敗」で評価する。さらに、この時のシステムを使った感想を「有用性」「効率性」「満足度」の3つの観点を、3を標準とし5段階評点により、ユーザビリティの評価を行った。

4 実験結果

「単語認識率」の結果を図に示す(図2)。

図からわかるように、「認識成功率」「曲名の認識成功率」がシステム不使用時と使用時では、使用時のほうが高くなっていることが分かる。さらに、楽曲数が489曲と1000曲を比べると、不使用時の認識率は低くなっているが、使用時にはほとんど変わらない認識率になった。このことから、楽曲数が増加しても適応的言語処理の有用であると考えられる。「タスク達成率」「ユーザビリティ」の結果を図3と図4に示す。

「単語認識率」の評価実験と同様に、システム使用時には、楽曲数に関わらず高いタスク達成率と、高いユーザビリティ評価が得られた。しかし、ユーザビリティの「効率性」の値が他の「有用性」「満足度」に比べて低くなっており、なお改善の余地がある。

5. まとめ

本稿では、実用的な音声による楽曲検索システムを実現するために、適応的な言語処理による音声認識による音声認識精度の向上と、システム側からのフィードバックによるユーザビリティの向上を組み込んだシステムの設計と評価を行った。「単語認識率」の評価実験により、適応的言語モデルを用いた音声認識は、有効性が示された。さらに、「タスク達成率とユーザビリティ」の評価実験では、ユーザインターフェースを含むシステムの有用性が得られた。しかし、一回目の認識率の低さが目立ち、ユーザビリティを下げる原因になっていることが分かった。今後は、音声認識精度の向上の手法を検討していくとともに、検索対象の楽曲データベースの曲数を増やし、評価実験を行う予定である。

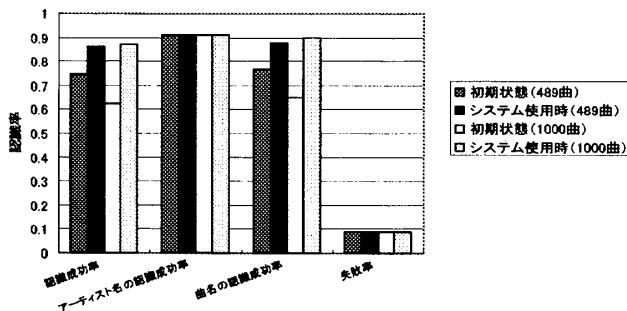


図2 単語認識率の評価実験結果

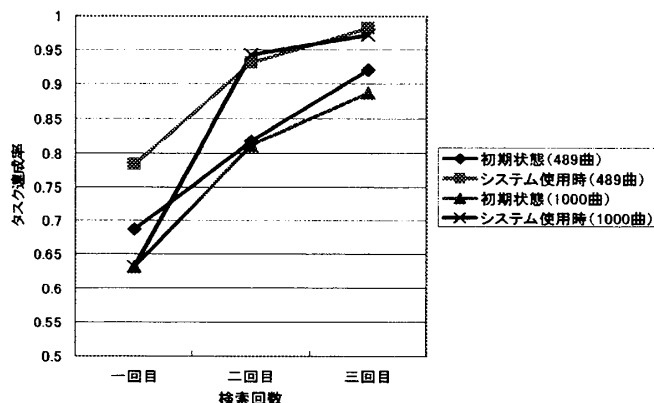


図3 タスク達成率の評価実験結果

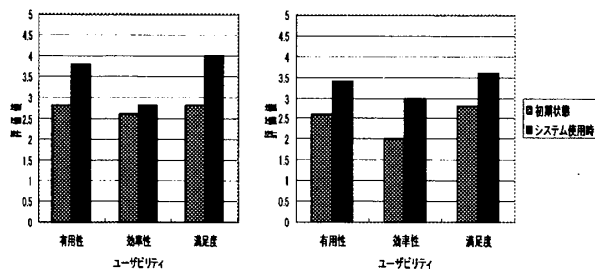


図4 ユーザビリティの評価実験結果

謝辞

本研究の一部は、早稲田大学理工学研究所の研究課題「自発的コミュニケーション機構を有するマルチモーダルヒューマンインタフェースの研究」、平成19年度科学研究費基盤研究(B)課題番号17300066「対話状況に応じた自発的コミュニケーション機構の研究」によるものである。

参考文献

[1]Stefan Schulz,Hilko Donker
 "An User-Centered Development of an Intuitive Dialog Control for Speech-Controlled Music Selection in Cars" INTERSPEECH2006 ICSLP, 2006

[2]名古屋大学情報科学研究科音言語行動信号処理研究室,
 "楽曲検索システム MusicNavi"
<http://moscowmule.sp.m.is.nagoya-u.ac.jp/index.html>

[3]李時旭, 児島宏明, 田中和世"音声による語彙フリー多言語マルチメディア検索システム"知能システム研究部門 研究成果展示会 オープンハウス 2004.

[4]李晃伸, 河原達也, "連続音声認識ソフトウェア Julius,"人工知能学会誌, vol.20, no.1, pp.41-49, 2005.