

人生検索エンジン「ライフリサイクル」

ー音声ライフログの柔軟な検索システムー

1. 背景

日常生活のデータを蓄積し、検索やデータマイニングに利用する「ライフログ」は近年大きな注目を集めている。しかし一般に利用されているライフログシステムの多くは位置情報や画像、動画を利用したものであり、会話データを対象とするものは広く普及するに至っていなかった。その理由として、日常会話に対する音声認識の精度が依然として低く、正確な検索やマイニングのための単語抽出やインデキシングが困難であったことが挙げられる。これは多くの音声認識システムで使用される音響モデル／言語モデルが講演音声をもとに学習したものであり、日常会話のように発話速度が一定しない場合、精度が低くなってしまうことが原因である。現在、日常会話をもとにした音響モデル／言語モデルの構築が進められているが、それによって精度がどれだけ向上するかは未知数である。

2. 目的

本プロジェクトでは音声認識技術をベースとして用い、日常会話音声の検索を可能にするシステム「ライフリサイクル」の開発を行った。また、音声認識に新たな拡張を付与することによって、学術的にも新しい手法の開発を目指した。各要素技術はモジュールとして実装し、類似する目的を持ったシステムの基盤となるように努めた。

「ライフリサイクル」の開発は以下の要素の実現を目的として進められた。

1. 日常会話の認識に適した音声認識モジュール
2. 高速な検索を可能にさせるインデキシングモジュール
3. 柔軟で使いやすい検索を実現する検索インタフェース

最終的にこれらの機能に加え、発話速度の揺らぎに対応する手法、ならびに音声データの自動コピーモジュールの実装を行った。

3. 開発内容

本プロジェクトで開発された「ライフリサイクル」の詳細は以下の通りである。

3. 1 システムの全体構成

ライフリサイクルではICレコーダによって録音された日常会話データに対して音声認識を適用し、得られた単語列ならびに音素列をデータベースに蓄積する。これを後日、Web検索エンジンを模した検索インタフェースで検索して利用する、という流れを基本としている。

ユーザはシステムを以下の手順に従って使用する。

1. ICレコーダを用いて発話を記録し、PCに接続する。

2. 音声認識が自動で実行され、単語列にフィルタリングがおこなわれる。
3. 単語列と音素列がデータベースに蓄積される。
4. 検索インタフェースで検索を行う。
5. 検索結果が検索インタフェース上で表示される。

一連の流れを図 1 に示した。

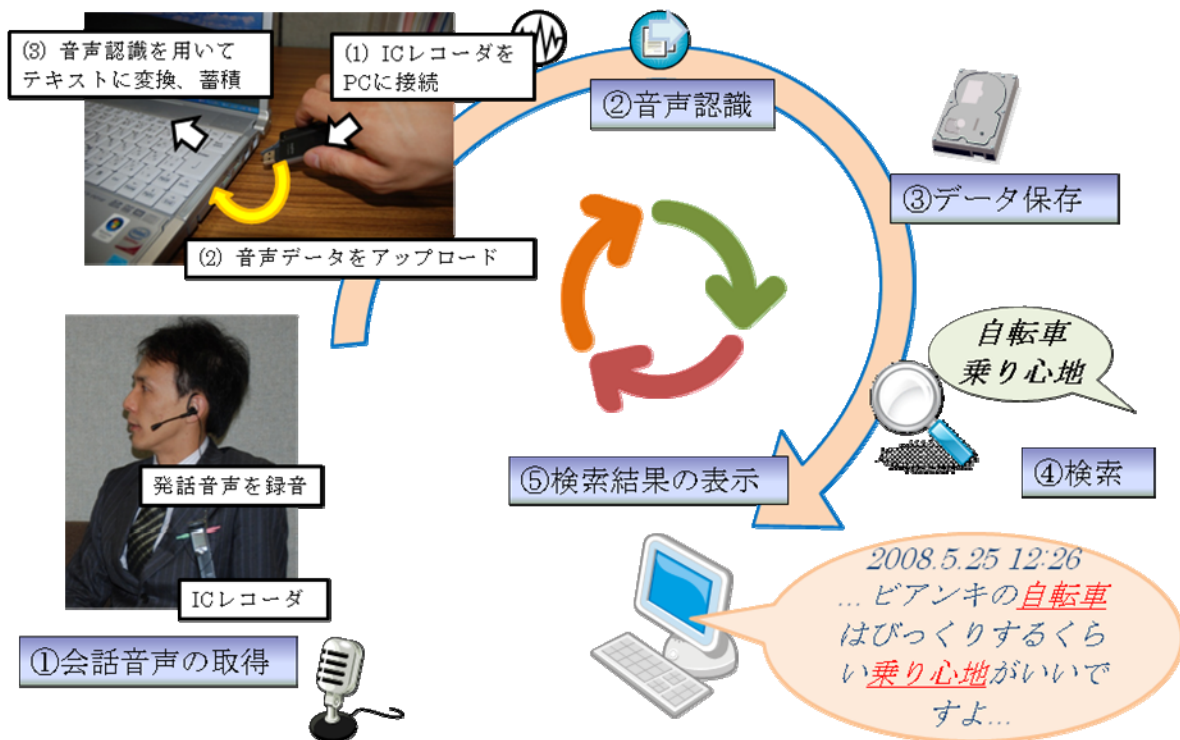


図 1： ライフサイクルの全体構成

本システムにおいて中心的な位置を占め、開発においても主要な部分を占めた音声認識／インデキシング機能の構成を図 2 に示す。

まず、音声データを認識し、テキスト化を行うため、音声認識エンジン Julius が使用される。この際、広く使われている日本語話し言葉コーパス（CSJ コーパス）の音響モデル・言語モデルが利用される。

認識結果の単語列は意味空間モデルでフィルタリングが行われた上で、リレーショナルデータベースに蓄積される。同様に音素列もリレーショナルデータベースに蓄積される。

以下の節ではこれらの機能の詳細を述べる。

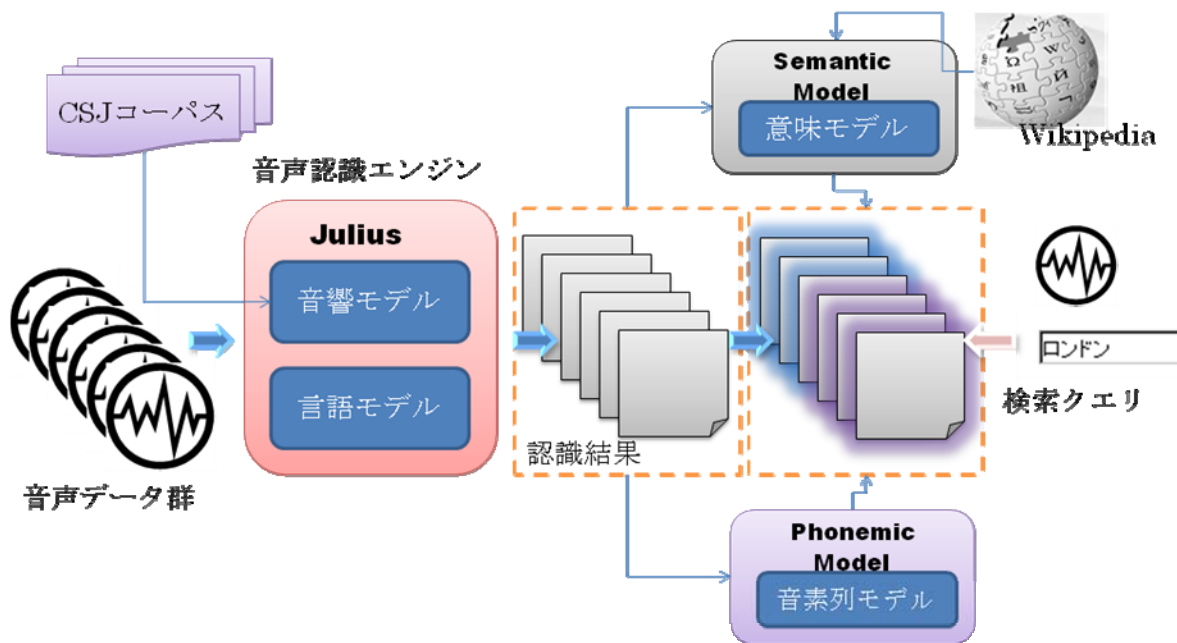


図 2： ライフサイクルの音声認識／インデキシング機能の構成

3. 2 自動コピー機能

会話音声进行認識／インデキシングする際、IC レコーダ中の音声データを手動で PC にコピーする形ではユーザにおける負荷が大きい。そこで利便性を向上させるため、自動コピー機能を実装した。

IC レコーダを PC に接続した際、コピープログラムが自動で起動し、ファイルの転送が行われる。また、Julius が起動され、インデキシングまでの一連の流れが実行される（図 3）。

IC レコーダと PC のそれぞれに設定ファイルを置くことで、すでにコピーされたファイルを改めてコピーすることを防ぎ、音声認識／インデキシングの際のオプション設定を変えられるようにしている。

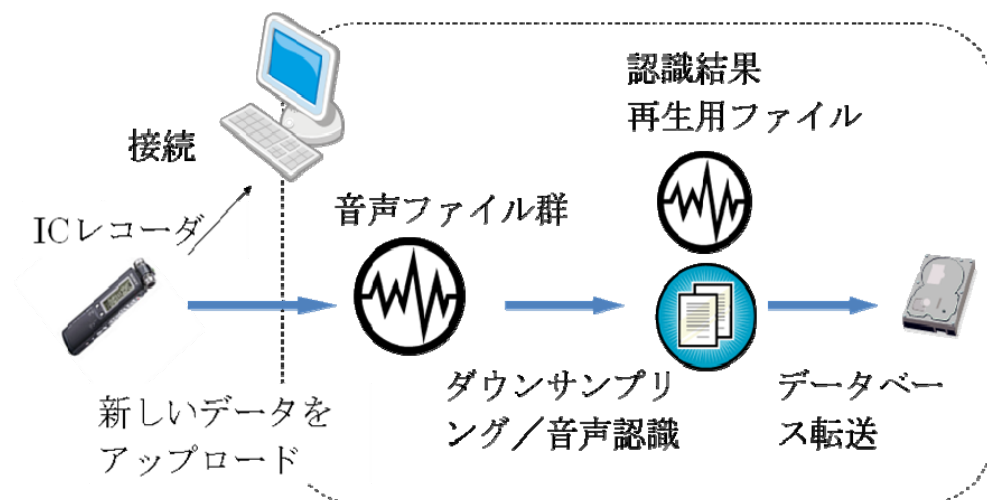


図 3： 自動コピー機能

3. 3 音声認識モジュール

音声認識モジュールではフリーの音声認識エンジンである **Julius** を使用した。

従来の音声認識エンジンにおいて主に対象とされてきた講演音声と異なり、日常会話では発話速度に大きな揺らぎが存在するため、認識精度が低くなってしまうという問題があった。

この問題を解決するため、認識モデルの学習時とは異なるフレームシフトと窓長を複数用いて認識を行い、もっとも尤度の高い単語列／音素列を選択することで、発話速度の揺らぎに対応させた。

3. 4 意味空間を用いたフィルタリング

音声認識エンジンの出力結果には誤認識が多く含まれる。誤認識が生じた場合、結果は実際の発生と音声的に似ているだけであり、意味的にはまったく異なる単語であることが多い。すなわち、その会話に現れる他の単語とは意味的に大きく異なる単語であることが多い。

そこで本システムでは意味空間モデルを用いることで、会話中の他の語から意味的に大きく離れた単語をフィルタリングする機能を考案し、実装した。これによって正しく認識された単語がフィルタリングされてしまう可能性もあるが、単語の登録から漏れた場合も、音素列はデータベースに登録されるため、音素列検索によってデータの欠損を補えると考えられる。

実際の所、人間もまた意味的な知識を用いてフィルタリングや補正を行っている可能性がある。そのため未知の音素列は会話中でも聞き取りにくいという傾向がある。

具体的な実装では **World Wide Web** 上の大規模な百科事典である **Wikipedia** のデータを用いて **pLSI** (**probabilistic Latent Semantic Analysis**) を実行し、意味空間モデルを構築した。この意味空間での距離を用いて単語のクラスタリングを行い、最大クラスタの重心を会話の意味的な中心とみなした。この重心から閾値以上離れた単語はデータベースに登録しないことで、フィルタリングを実現した。

具体的には以下のステップに従って処理を行っている (図 4)。

1. 音声認識エンジンが出力する単語認識の尤度を利用し、もっとも尤度の高い名詞の集合を取得する。
2. pLSI から得られた単語-意味空間の変換行列を用いて意味空間への変換を行い、クラスタリングを行う。
3. 最大クラスタに含まれる単語集合を意味空間に変換し、重心を求める。会話中の各単語と重心との距離を求め、閾値以内に入る単語のみをデータベースに登録する。

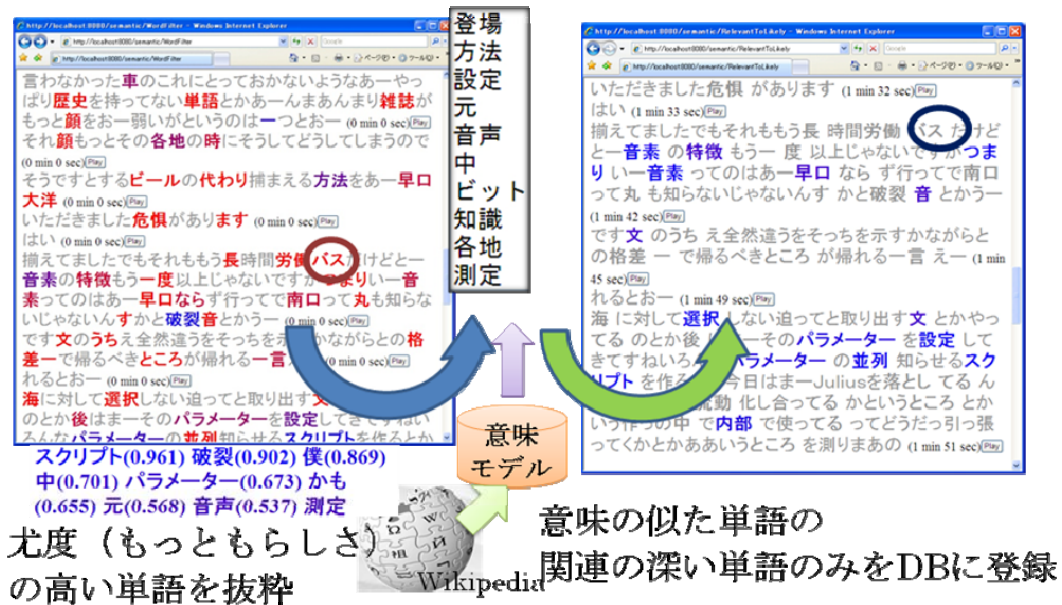


図 4： 意味空間モデルによるフィルタリング機能

3. 5 音素列検索

音素列検索モジュールでは音素列の完全一致ならびに類似に基づく検索を可能にしている。音声認識において誤認識が生じた場合、音素列として見た場合には似ている結果が得られることが多い。そこでユーザが入力したクエリをいったん音素列に変換し、データベース中から類似する音素列を検索することで、検索結果の再現率を向上させる。

漢字をひらがなに変換するツール kakasi を利用し、検索クエリをひらがなに変換した上で、パターンマッチによって音素列に変換を行う。

完全一致したパターンのみだけでなく、レーベンシュタイン距離（編集距離）が一定の距離以下である音素列まで含めて検索する機能を設けた（図 5）。

【認識結果例】

「このようなこと 機能の ことを エコー
ロケーション機能と 呼びます。」
→「こんなこと。昨日のことはこの
て初期の端ます。」

エコーロケーション

e k o: r o k e: s h o N
↓置換 挿入 削除
a k o: r o q k e: s h o

レーベンシュタイン距離 = 3

図 5： 音素列を用いた検索機能

3. 6 検索インタフェース

多くのユーザが使い慣れていると考えられる Web 検索エンジンのインタフェースを踏襲したデザインを選択した（図 6）。

検索フォームにキーワードを入力し、単語列の一致による検索を行う。この方式で検索結果が得られなかった場合には、音素列の一致／類似に基づいた再検索を行う。



図 6：「ライフリサイクル」検索インタフェース

4. 従来の技術との相違

「ライフリサイクル」は日常会話音声を検索するために必要とされる総合的な機能を持っている。各構成要素は独自の技術に基づいており、他システムに組み込んで使用することも可能である。

既存の音声認識エンジンを利用したシステムと比較した場合、日常会話のインデキシングに適しているという特徴がある。具体的には発話速度の揺らぎや発音の不正確さ、語彙の多様性など、日常会話で生じがちな課題への対応を行っている。

さらに、音声認識結果の蓄積を行うシステムとしてすでに市販されているものに「議事録自動作成システム」があるが、これらのシステムを利用する場合、過去の議事録を訓練データとして言語モデルを構築する必要があるなど、大きな手間がかかる。本システムでは手軽に音声認識によるテキスト化と検索を始められるというメリットがある。

5. 期待される効果

本システムはすでに基盤となる機能が実装されているため、今後は企業や研究機関と連携を行い、実用面での完成度を高め、一般への普及に耐えられるシステムを構築していく。

また、開発された手法を論文の形で発表し、学術面での貢献も行っていく予定である。

6. 普及の見通し

開発成果の当初の利用形態としては、会議の議事録や引き継ぎの際の録音等、明確な目的をもった形を念頭においている。特に大学の研究室における学生間の引き継ぎ作業にプロトタイプを使用し、実証実験を行う予定である。

IC レコーダの小型化と低廉化は急速に進んでおり、システムの普及において肯定的な要因となることが予想される。

さらに、携帯電話の録音機能と組み合わせることで日常的な会話を継続して蓄積し、より広い用途で利用されることを目標としている。

7. 開発者名

チーフクリエイター： 手塚 太郎 （立命館大学 情報理工学部）
コクリエイター ： 谷口 忠大 （同上）
 原田 史子 （同上）