

1. はじめに

本稿では、ユーザがある単語を最後まで思い出せずに断片だけを発話しても、システム側がその残りを補って入力することを可能にする「音声補完」という新しい音声インタフェース機能を提唱する。人間同士の音声対話では、話者がある単語の発声途中で言い淀むと、対話相手はしばしば、発話された断片の続きを補完することで、話者が述べようとしている単語全体の候補を提示し、話者が思い出すのを手助けしてくれる。本研究は、このような補完による手助けという概念を音声入力の枠組みに導入することで、音声インタフェースをより使いやすくすることを目的とする。

既にテキストインタフェース (UNIX シェルやエディタ) では、ファイル名の入力等で補完の概念が広く受け入れられており、ユーザが補完機能呼び出すキー (補完トリガーキー) を押せば、続きが補完されて入力できるようになっている。また、近年ペン入力でも、補完機能を持ったインタフェース^{1),2)} が提案されている。しかし音声入力では、自然に補完機能呼び出す手段がなかったこともあって、効果的な補完機能はこれまで提案されていなかった。

2. 音声補完

「音声補完」とは、計算機システムに対する音声入力中に、ユーザが補完機能呼び出すことができるようにするための新たな音声入力インタフェース機能の総称である。ユーザが発話した断片をシステム側が補完してくれることで、以下のような利点が得られる。

- 記憶補助 入力したい内容がうる覚えでも、途中で思い出して発声すれば入力できる。
- 省力化 入力内容が長くて複雑なときに、内容の特定に十分な部分まで発声すれば入力できる。
- 心理的抵抗の低減 従来の音声インタフェースの多くが、すべての音を最後まで丁寧に発声することを強いていたのに対し、音声補完では途中まで発声すればよく、心理的抵抗が少なく使いやすい。

実用的な音声補完を実現するためには、どのように補完機能呼び出すか、つまり音声における補完トリガーキーをどう実現するかが重要である。そこで、言い淀み現象の一つである有声休止 (filled pause) に補完トリガーキーの役割を担わせることを提案する。補完トリガーキーとして音声入力中に有声休止をおこなうことは、人間にとって極めて自然であり、ユーザが候補を見たいと思う任意のタイミングで、労力をかけずに補完機能呼び出すことが可能になる。

音声補完の対象には、単語や文節、文章など様々なレベルが考えられるが、本稿では以下、単語のみを取

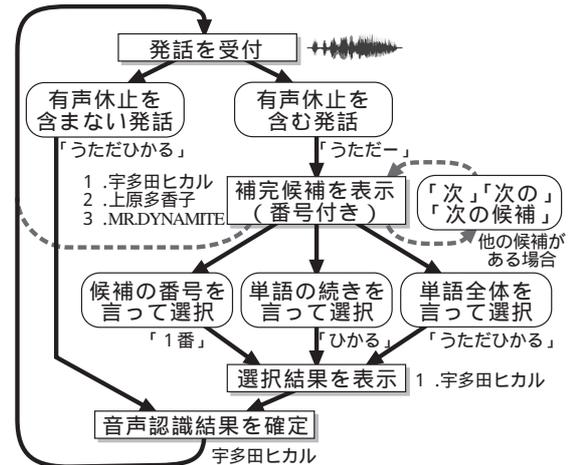


図 1: 音声補完の操作の流れ

り上げて議論する。つまり、単語補完機能に論点を絞り、有声休止としては、単語の途中の母音の引き延ばしだけを対象とする。ただし、ここでの単語は、音声認識システムの単語辞書上の1単語とする。

3. 音声補完機能付き音声入力インタフェース

我々が構築した音声入力インタフェース上で、ユーザは、以下のように有声休止を用いて音声補完しながら、単語を入力することができる (図 1)。

1. 単語の発声途中で母音を引き延ばすと、既に発声された断片から始まる補完候補 (単語) の一覧が、番号付きで即座に表示される。
(ex. 「うただー」と入力すると「1. 宇多田ヒカル, 2. 上原多香子, 3. MR.DYNAMITE」のように補完候補が表示される。)
2. 候補が多くて画面に入りきらないときには、「次の候補」というマークが表示される。その場合「次の候補」等と言えば他候補が見れる。候補が不適切なときや別の単語を入力したいときには、次の3. の選択をせずに別の発話に移ってもよい。
3. ユーザは3通りの方法で補完候補を選択できる。

- (a) 候補の番号を言って選択する。
(ex. 「1番」か「1」と言う。)
- (b) 単語の続きを読み上げて選択する。
(ex. 「ひかる」と言う。)
- (c) 単語全体を頭から読み上げて選択する。
(ex. 「うただひかる」と言う。)

選択すると、その候補は強調表示され、音声認識結果として確定される。

なお、一つの単語を入力中に、音声補完を繰り返し呼び出すこともできる。(ex. 「サザンオールスターズ」を「さざんー (表示)」「おーー (表示)」「すたーず (選択)」で入力。)

4. 実現方法

3章のインタフェースを構築するには、有声休止開始点を決定する有声休止検出部と、単語の途中までの発声を認識して補完候補を作成する音声認識部を実現

* “Speech Completion: New Speech Input Interface Supporting Word Completion” by Masataka Goto, Katunobu Itou and Satoru Hayamizu (ETL)

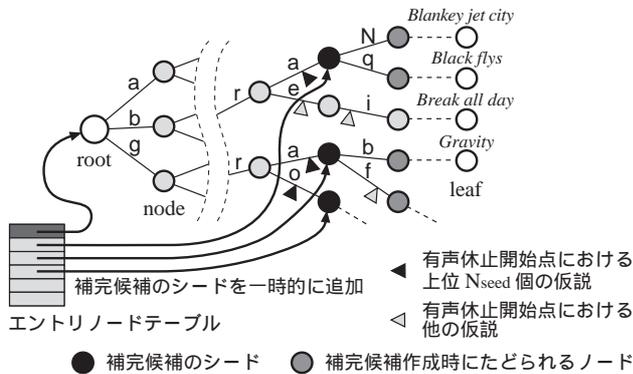


図 2: 木構造の単語辞書における有聲休止開始点での音声補完候補の作成とエントリノードテーブルへの追加

しなければならない。さらに、インターフェース全体の状態管理をするインターフェース管理部と、補完候補一覧や認識結果を提示する画面表示部も必要となる。

4.1 有聲休止検出部

音声補完では、有聲休止を高い精度でリアルタイムに検出することが重要である。しかも、任意の単語中の母音の引き延ばしを検出するため、トップダウン情報を使わない言語非依存な検出をしなければならない。そのような要件を満たす検出手法として、我々が文献3)~6)で提案・性能評価した、有聲休止箇所のリアルタイム検出手法を用いる。ここで決定した有聲休止開始点は、次の音声認識部へ送られる。

4.2 音声認識部 (音声補完候補作成)

音声認識部は、音声入力と有聲休止検出部の結果を受け取り、音声認識結果 (尤度の高い順に上位 N_{result} (= 5) 個) と音声補完候補をインターフェース管理部へと送る。以下、単語発声の補完を説明するが、連続音声の中の単語を補完することも同じ枠組で可能である。

補完候補一覧を作成する処理は、連続音声認識システム $miNja^7)$ を拡張して実現する。単語辞書は、図2のように木構造で保持され、図中のくさび形のマークが認識処理の最中の仮説を表す。有聲休止開始点になると、その時点で有効な仮説 (尤度の高い順に上位 N_{seed} (= 5) 個) から葉の方向へたどって補完候補を生成し、尤度の高い順に番号付けして、上位 N_{choice} (= 20) 個をインターフェース管理部へ送る。それらの仮説に対応するノードを補完候補のシードと呼ぶ。

単語の続きを言っても選択できるように、認識を開始する根を登録するエントリノードテーブルを導入し、単語の途中からの認識を可能にする。通常の単語の頭からの認識では、このテーブルには辞書の根だけが登録されている。単語の途中から認識を開始したい場合には、図2のように補完候補のシードを根として一時的に追加する。これらの追加エントリは、有聲休止を伴う発話の次の発話まで有効とする。

4.3 インターフェース管理部・画面表示部

補完候補の選択等のインターフェース全体としての機能を提供する。まず、有聲休止を含まない発話の場合



図 3: 音声補完中の画面表示例 (補完候補ウィンドウ)

には、単に認識結果を受け取って表示する。一方、有聲休止を含む発話の場合には、音声補完候補を受け取った時点でポップアップ式の補完候補ウィンドウを出現させ、その中に候補一覧を表示する。その際、音素列上で既に発声された部分の色を変えて各候補を表示する。そして、図1の操作の流れに従いながら、次の発話の認識結果に応じて表示・選択・確定等の処理をおこなう。その際、常に1位の認識結果を用いるのではなく、選択操作に該当する結果が上位 $N_{priority}$ (= 3) 個以内であれば、それを優先させて用いる。これは、単語の続きを言って選択する場合等に、その発声に近い単語辞書上の別の単語の尤度が高くなり、適切に選択できない事態を回避するためである。

5. 実装と結果

以上述べてきた音声入力インターフェースのプロトタイプシステムを実装し⁸⁾、曲名とアーティスト名のデータベースを単語辞書として、音声補完の動作を確認した。画面表示例を図3に示す。本システムを運用した結果、提案したインターフェースが実用的に機能し、ユーザが音声補完機能呼び出ししながら、インタラクティブに単語入力できることを確認した。音声補完機能は、使用するのが容易で訓練は不要であり、直感的で使いやすいインターフェースであることがわかった。特に、長い固有名詞を入力する際に、音声補完は非常に有効であった。

6. おわりに

本稿では「音声補完」という新しい音声インターフェース機能を提唱し、補完トリガーキーとして有聲休止を用いることを提案した。実際に、インタラクティブに音声入力可能なシステムを実装し、曲名等の入力で有用性を確認したが、これは住所入力や各種固有名詞の入力といった様々な局面にもすぐに適用できる。今後、音声入力インターフェースを構築する上で、音声補完は不可欠な機能の一つになることが予想される。

参考文献

[1] 福島 他: 予測ペン入力インタ..., 情処学論, 37, 1, 23-30, 1996.
 [2] Masui: An Efficient Text Input ..., CHI'98, 328-335, 1998.
 [3] 後藤 他: 自然発話中の..., 情処研報 99-SLP-27-2, 9-16, 1999.
 [4] Goto et al.: A Real-time ..., Eurospeech'99, 227-230, 1999.
 [5] 後藤 他: 自発的な発話中の言い淀..., 音講論集 秋季3-1-5, 1999.
 [6] 後藤 他: 有聲休止箇所のリアルタ..., 音講論集 春季3-8-8, 2000.
 [7] 伊藤 他: 音素文脈..., 信学論, J75-D-II(6), 1023-1030, 1992.
 [8] 後藤 他: 音声補完: "TAB" on Speech, 情処研報 音声言語情報処理研究会, 2000-SLP-32-16, 81-86, 2000.