音声補完: 音声ワイルドカード補完機能の実現*

◎ 後藤 真孝　伊藤 克亘　秋葉 友良　速水 悟 (電総研)

1. はじめに

音声の持つ潜在能力を引き出した音声インタフェースを実現するには，言い淀みや韻律のような非言語情報が人間同士のコミュニケーションで果たす役割を，積極的に活用する必要がある．そこで我々は，ユーザが音声入力中に言い淀むと，計算機が補完候補を提示して手助けをしてくれる「**音声補完**」という新しい音声インタフェース機能を提唱し，プロトタイプシステムの構築を進めてきた[1),2)]．そして，ユーザが入力したい単語(フレーズ)の後半を思い出せずに，途中で言い淀む(有声休止をおこなう)と，システム側がその後半を補完してくれるインタフェースを実現した．この単語の頭から末尾へ向けて補完する方式を，本稿では「**音声フォワード補完**」と呼ぶ．

実際に，人間同士の音声対話でも，話者が「音声補完」という単語を最後まで思い出せず,「おんせ**いー**」と言い淀むと (「**いー**」が有声休止)，対話相手は「音声補完？」のように候補を提示することで，話者が思い出すのを手助けしてくれる．このように，音声では断片的な情報を伝えても，対話相手が自分の発話や思考の手助けをしてくれることが期待でき，それが快適で優れた情報交換手段である一つの理由となっている．文献1),2)の音声フォワード補完は，有声休止を補完のトリガーとすることで，これを音声入力インタフェースに自然な形で取り入れることに成功した．しかし，補完箇所は単語の後半に限定されていた．

そこで本稿では，入力したい単語の前半を思い出せずに言い淀みながら後半だけを発声しても，システム側が前半を補って入力することを可能にする「**音声バックワード補完** (**音声ワイルドカード補完**)」という新たな補完方式を提案し，その実現方法を述べる．

2. 音声バックワード補完 (音声ワイルドカード補完)

音声バックワード補完とは，事前に定めた**ワイルドカードキーワード**を言いながらその最後の音節で有声休止をおこない(母音を引き延ばし)，続いて後半を発声することで，その前につながる前半を補完する方式である．例えば,「なんとか」をキーワードと定めた場合,「宇多田ヒカル」のような姓名の姓が思い出せずに「**なんとかー**ひかる」と入力すると,「宇多田ヒカル」が補完候補の一つとして得られる．この「**なんとかー**」はワイルドカードに相当するため，これを音声ワイルドカード補完とも呼ぶ．入力したい単語辞書の中に，キーワードを部分文字列として含むような単語が仮にあったとしても，有声休止によってキーワードは識別可能であり，意図した箇所でのみ音声バックワード補完を呼び出すことができる．

「音声フォワード補完」と「音声バックワード補完」が共に可能な音声入力インタフェースの操作の流れを図 1 に示す．音声バックワード補完の場合，ワイルドカードキーワードの後に発声された断片が末尾に付く補完候補の一覧が，番号付きで表示される．ユーザは，3通りの方法(候補の番号を言う，単語の前半を読み上げる，単語全体を頭から読み上げる)で選択でき，選択された候補が強調表示されて確定する．候補が多くて画面に入りきらないときには,「前の候補」「次の候補」というマークが表示され,「前(の候補)」や「次(の候補)」と言えば他候補を見ることができる．

以上のように，単語の頭がわかるときは「音声フォワード補完」，末尾がわかるときは「音声バックワード補完」を用いれば良いが，両者がわからず中央部分だけがわかるときは問題が生じる．これは，両補完方式を組み合わせることで実現できる．逆に，中央部分がわからず頭と末尾が共にわかる場合には，いずれか一方の補完方式を適用すればよく，問題がない．

3. 二つの補完方式を組み合わせた補完

単語の頭と末尾がわからず，中央部分だけがわかるときは，まず既知の中央部分までを音声バックワード補完で入力しつつ，その中央部分の最後の音節で有声休止をして音声フォワード補完を呼び出すことで，補完入力が可能となる．例えば,「Dragon Ash feat. ラッパ我リヤ」を入力したいときに「feat. (フィーチャリング)」しかわからない場合,「**なんとかー** ふぃーちゃりん**ぐー**」と入力すると，中央部分に「ふぃーちゃりんぐ」の音がある候補一覧を見て選択できる．

4. 実現方法: 音声認識部 (補完候補作成) の拡張

文献1),2)では，音声補完インタフェースが，有声休止検出部，音声認識部，インタフェース管理部，画面表示部の4つの構成要素で実現可能なことを述べたが，以下では，音声認識部の音声バックワード補完への対応方法を中心に説明する．

有声休止検出部で文献3),4)の手法によって有声休止が検出されると，音声認識部は，補完方式に応じて候補一覧を作成する．ここでは，単語発声の補完を説明するが，連続音声中の単語の補完も同じ枠組で実現できる．入力対象の単語辞書(人名等)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．有声休止が検出されると，その時点で最も尤度の高い仮説がワイルドカードキーワードかどうかを判定し，音声フォワード補完と音声バックワード補完のどちらを実行するかを決定する．

音声フォワード補完の場合[1),2)]，尤度の高い順に上位 N_{seed} (= 5) 個の仮説から葉の方向へたどって補完候補を生成し，上位 N_{choice} (= 20) 個を番号付けして，インタフェース管理部へ送る．それらの仮説に対

* "Speech Completion: Supporting Speech Wildcard Completion" by Masataka Goto, Katunobu Itou, Tomoyosi Akiba, and Satoru Hayamizu (ETL)

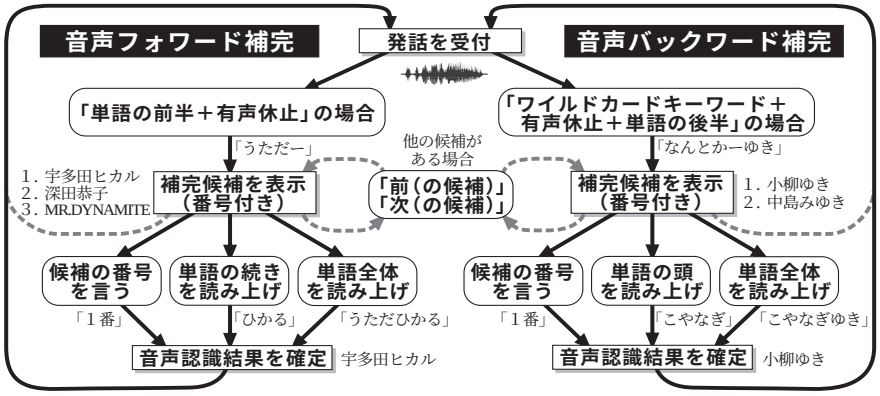


図 1: 音声補完の操作の流れ

(a) 「なんとかー」と入力した直後

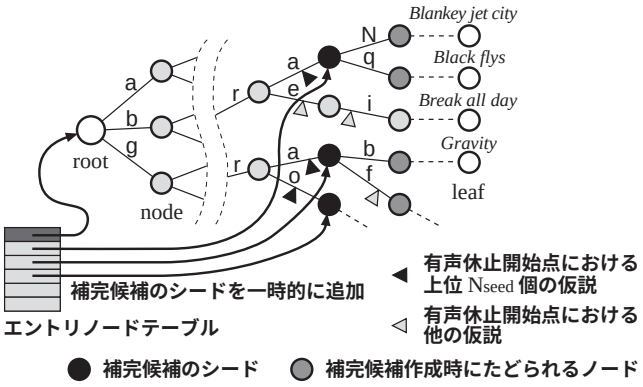
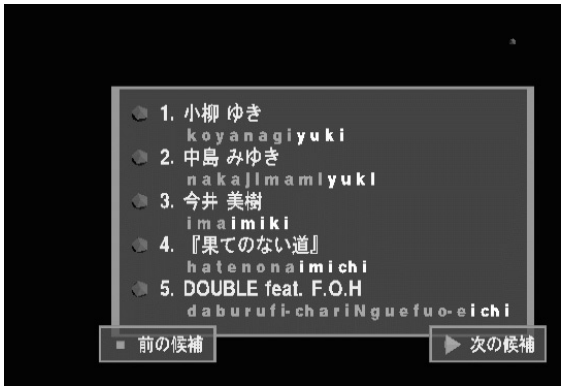


図 2: 木構造の単語辞書における有声休止開始点での音声補完候補の作成とエントリノードテーブルへの追加



(b) 「ゆき」と入力した直後に補完候補ウィンドウが登場

図 3: 音声バックワード補完中の画面表示例

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，ここ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．

一方，音声バックワード補完の場合，有声休止終了時点以降に発声された単語の後半部分を認識し，補完候補を生成する必要がある．この単語途中からの認識は，辞書中の全単語の途中の音節を，エントリーノードテーブルに一時的に(ワイルドカードキーワードの直後だけ)追加することで実現する．そして葉に到達した仮説から尤度の高い順に上位 N_{choice} 個を送信する．その後，単語の頭(未発声の前半)を言っても選択できるようにするために，各候補で発声されなかった音素列の終端を葉とする単語を一時的に登録する．例えば,「小柳ゆき」を「**なんとかー**ゆき」で入力した場合，/koyanagi/の末尾を葉とする単語を追加する．

5. 実装と結果

提案した全補完方式を可能にする音声入力インタフェースを実装し，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動作確認をした．実際に運用した結果，入力内容がうろ覚えなときや長くて複雑なときにも，音声補完機能を呼び出しながら，容易に単語入力できることを確認した．特に，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，音声補完では思いついた断片だけを発声すればよく，心理的抵抗が少なく使いやすい点が優れていた．

6. おわりに

本稿では,「音声バックワード補完」という新たな補完方式を導入し，任意の箇所に対する補完を可能にした．今後，音声補完を発端として，他の非言語情報も導入していくことで，さらに使いやすい音声インタフェースが構築できる可能性がある[5]．キーボードとの対比で考えれば，従来の音声認識が扱ってきたのは，通常キー(文字キー)の一部に過ぎない．それに対して，本研究での有声休止の位置付けは，いわば特殊キーの Tab (UNIX シェルや Emacs エディタの補完トリガーキー)に相当する．これを第一歩として，音高等の他の非言語情報を特殊キーとして活用するような研究が，今後発展していく余地は大きい．

参 考 文 献

[1] 後藤 他: 音声補完: “TAB” on Speech, 情処研報2000-SLP-32-16, pp. 81–86 (2000).
[2] 後藤 他: 音声補完: 単語補完ができる新たな音声入力インタフェース, 音講論集 秋季 2-Q-10 (2000).
[3] Goto *et al.*: A Real-time Filled Pause Detection System for Spontaneous Speech Recognition, *Eurospeech '99*, pp. 227–230 (1999).
[4] 後藤 他: 自然発話中の有声休止箇所のリアルタイム検出システム, 信学論(D-II), **J83-D-II**, 11, pp. 2330–2340 (2000).
[5] 後藤 他: 音声補完: 音声入力インタフェースへの新しいモダリティの導入, *WISS2000*, 近代科学社, pp. 153–162 (2000).