

音声補完の評価

後藤 真孝 伊藤 克亘 速水 悟

産業技術総合研究所
m.goto@aist.go.jp

あらまし 本稿では、ユーザがある単語を思い出せずに断片だけを発声しても、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」という音声インタフェース機能とその評価について述べる。既にテキストインタフェースでは、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いるが、音声では効果的な補完機能は提案されていなかった。我々は、ユーザが単語発声途中に有声休止(母音の引き延ばし)によって言い淀むと、それを含む補完候補の一覧を見ることが出来るインタフェースを構築し、労力をかけずに自発的に補完機能呼び出ししながら音声入力することを可能にした。45名の被験者による評価実験の結果から、音声補完は使いやすく便利な機能であり、その有効性が高いことが確認された。

Evaluation of Speech Completion

Masataka Goto Katunobu Itou Satoru Hayamizu

National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan

Abstract This paper describes a speech interface function, called *speech completion*, which helps a user enter a word or phrase by *completing* (filling in the rest of) a phrase fragment uttered by the user. Although the concept of completion has been widely accepted in text-based interfaces, effective completion for speech has not been proposed. We enable a user to invoke the speech-completion function intentionally and effortlessly by building an interface that displays completion candidates when a filled pause is uttered (a vowel is lengthened) during a phrase. Experimental results with forty-five subjects confirmed the intuitiveness, usefulness, and effectiveness of speech completion.

1 はじめに

現在の音声入力インタフェースは、音声の持つ潜在能力を引き出していない。音声は、音韻や単語のような言語情報だけでなく、韻律や言い淀みのような非言語情報も含んでいるが、これまでの音声認識は、主に言語情報のモダリティーしか利用していなかった。そのため、いわば認識誤りを起こすようなキーボードにしか過ぎず、いくら音声認識率を高くしたとしても、キーボードを越えるような使いやすいインタフェースは構築困難であった。音声ならではのメリットを引き出すためには、音声の持つ非言語情報が、人間同士のコミュニケーションでどのような役割を果たしているのかを問い直し、その役割を積極的に活用したインタフェースを構築する必要がある。

そこで我々は、非言語情報の中でも特に話者の思考状態が現れやすい有声休止 (filled pause) に着目し、ユーザが言い淀むとシステム側が手助けをしてくれる「音声補完 (speech completion)」という新しい音

声インタフェース機能を提唱した^{1)~5)}。有声休止は言い淀み現象の一つであり⁶⁾、発話したい内容が断片的にしか思い出せないときや、何を発話しているのか判断に迷うときに、発声されることが多い。音響的には持続した有声音 (母音の引き延ばし) として現れ、例えば、話者が「音声補完」という単語を最後まで思い出せないときには、「おんせいー」と言い淀んだりする(「いー」が有声休止)。このとき、対話相手はしばしば話者の言いたいことを推測し「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¹⁾。これは、発話された単語の断片の残りを補うことで、話者が述べようとしている単語全体の候補を提示している、つまり、単語を補完していると見なすことができる。そこで、この対話相手の役割を計算機に担わせることで、ユーザが音

¹⁾ このように、本来音声を使う場合には、いい加減で断片的な情報を伝えても、対話相手が様々な形で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となっている一つの理由であると考えられる。

声入力中に有声休止すると、計算機が補完候補を提示してくれる機能「音声補完」を実現した。

本稿では、これまで文献1)~5)で提案してきた音声補完機能についてまとめると共に、未報告であった音声補完の有効性を確認する評価実験の結果を中心に述べる。以下の章では、まず2章で音声補完の利点と三つの補完方式について述べ、3章で提案した音声補完機能を持つ音声入力インタフェースを紹介する。次に、4章で音声補完の有効性を確認する評価実験の結果を述べる。そして、5章で関連研究やマルチモーダルインタフェースに関する議論をおこない、6章で音声補完機能が効果的に活用される応用例を紹介する。最後に、7章でまとめを述べる。

2 音声補完

「音声補完」とは、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、一部だけ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。

音声補完を通常の音声入力に効果的に導入するには、ユーザが候補を見たいと思うタイミングで、自発的に補完機能呼び出せることが重要となる。なぜなら、ユーザの望まないタイミングで次々と補完候補を出すような自動補完は、認識時に曖昧性が大きい音声では、煩わしく不適切な機能となりやすいからである。そこで、言い淀み現象の一つである有声休止で呼び出すことにより、ユーザが候補を見たいと思うタイミングで、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、単語のみを取り上げて議論する。つまり、単語補完機能に論点を絞る。ただし、ここでの単語は、音声認識システムの単語辞書上(言語モデル上)の1単語とする。したがって、例えば姓名が一つの単語として登録されている

ときには、姓だけのよう部分発声から残りが補完される。

ユーザが単語のどの部位を発声して入力したいか(補完したい方向)に応じて、以下の三種類の音声補完方式を用意した。ここでは、「宇多田ヒカル」や「Blankey jet city」という一単語を補完する場合を例に説明する。

1. 音声フォワード補完(単語の頭がわかる場合)

単語の前半(頭)がわかっているときに、その最後の音節で有声休止をおこなうことで、それに続く後半を補完する方式である。例えば、「うただー」と「だ」の音で有声休止をおこなって入力すると、「宇多田ヒカル」が補完候補の一つとして得られる。仮に「宇多田」という単語も単語辞書に登録されているときには、ユーザが、単に「宇多田」と入力したいのか、それとも補完候補を得たいのかをシステムが識別する上でも、有声休止で補完機能呼び出すことが重要となる。

2. 音声バックワード補完(単語の末尾がわかる場合)

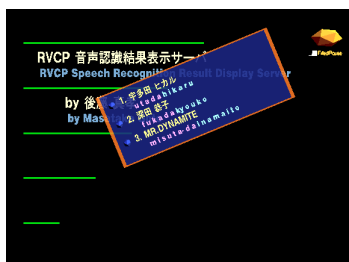
単語の後半(末尾)がわかっているときに、事前に定めたキーワード(以下、ワイルドカードキーワードと呼ぶ)を言いながらその最後の音節で有声休止をおこない、続いて後半を発話することで、その前につながる前半を補完する方式である。例えば、「なんとか」をキーワードと定めた場合、「なんかーひかる」と入力すると、「宇多田ヒカル」が補完候補の一つとして得られる。この「なんかー」は、任意の文字列にマッチするワイルドカードに相当する。入力したい単語辞書の中に、キーワードを部分文字列として含むような単語が仮にあったとしても、有声休止によってキーワードは識別可能であり、意図した箇所でのみ音声バックワード補完呼び出すことができる。

3. 音声ツーウェイ補完(単語の中央部分がわかる場合)

単語の中央部分がわかっているときに、上の二つの補完方式を応用し、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呼び出すことで、補完入力が可能となる。例えば、「Blankey jet city」を「jet」に基づいて補完したいときに「なんかー じゅっとー」と入力すると、中央部分に「じゅっと」の音を含む「Blankey jet city」が補完候補の一つとして得られる。



(1) 「うただー」と入力



(2) 「だー」の有声休止中



(3) 補完候補ウィンドウが回転しながら登場



(4) 「1番」と入力した直後



(5) 1番の候補が輝いて跳躍

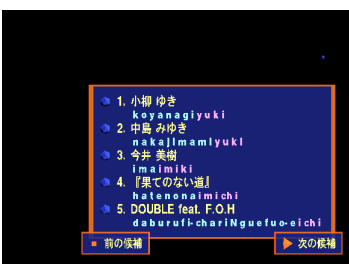


(6) 1番の候補「宇多田ヒカル」が認識結果として確定

図 1: 音声フォワード補完中の画面表示例



(a) 「なんとかー」と入力した直後



(b) 「ゆき」と入力した直後に補完候補ウィンドウが登場



(c) 「1番」と入力して「小柳ゆき」が認識結果として確定

図 2: 音声バックワード補完中の画面表示例

なお、上記以外の場合として、単語の頭と末尾が共にわかる場合(単語の中央部分だけがわからない場合)が考えられるが、その場合には、頭か末尾のいずれか一方を用いて、音声フォワード補完か音声バックワード補完をすればよい。

3 音声補完機能付き音声入力インタフェース

構築した音声補完機能付き音声入力インタフェースの機能を説明する。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。

1. [音声フォワード補完の場合]

単語の発声途中で母音を引き延ばすと、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. 図 1のように、

「うただー」と入力すると、「1. 宇多田ヒカル, 2. 深田恭子, 3. MR.DYNAMITE」といった補完候補が表示される²⁾。)

[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最後の母音を引き延ばし、続いて単語の後半を発声すると、その発声された断片が末尾に付く補完候補の一覧が、同様に番号付きで表示される。(ex. 図 2のように、「なんとかーゆき」と入力すると、「1. 小柳ゆき, 2. 中島みゆき」等のような補完候補が表示される。)

[音声ツーウェイ補完の場合]

ワイルドカードキーワードの最後の母音を引き

²⁾ テキストの補完とは異なり、たとえ「うただ」から始まる単語が辞書中に一つしかなくても、音声の曖昧性から候補を一つに絞り込めないことが多い。上記の例では、/utadahikaru/, /fukadakyouko/, /misuta-dainamaito/ が音響的な類似度から補完候補として得られた。

延ばし，続いて単語の中央部分の最後の音節の母音を引き延ばすと，その発声された断片が中央に含まれる補完候補の一覧が，番号付きで表示される．

- 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る．その場合，そのマークを読めば他候補が見られる．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．
- ユーザは3通りの方法で補完候補を選択できる．
 - 候補の番号を言う．(ex.「1番」か「1」と言う．)
 - 単語の続きや単語の頭を読み上げる．(ex.「ひかる」「こやなぎ」と言う．)
 - 単語全体を頭から読み上げる．(ex.「うただひかる」「こやなぎゆき」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．

以上の機能を持つ音声入力インタフェースは，有声休止の区間を得る有声休止検出部，音声認識中に補完候補を作成する音声認識部，インタフェース全体の状態管理をするインタフェース管理部，補完候補一覧や認識結果を提示する画面表示部の4つの構成要素で実現できる．これらは，効果的に負荷分散が可能で，拡張性が高くなるように，ネットワークプロトコル RVCP (Remote Voice Control Protocol)^{1),3)} を用いて複数のプロセス群として実装された．具体的な実装方法は，文献1),3)に委ねて省略する．

4 評価実験

音声補完の有効性を確認するために，3章で述べたインタフェースを用いて，以下の点を調査する実験をおこなった．

- 音声補完の使用に慣れた後，音声補完を使用するか使用しないかを自由に選んで入力するときに，どのように入力するか．
 - うろ覚えの単語を入力するときに，音声補完を使用するか使用しないか．
 - 音声補完の使用後にどのような印象を受けたか．
- なお，実験時間の関係で，本実験では，音声フォワード補完と音声バックワード補完の二つを対象に評価した．

本実験では，音楽検索等の現実的な応用を想定し，日本のポピュラー音楽のヒットチャート(2000年度の

すべての週間ランキングのシングル上位20曲)から，曲名(342語)とアーティスト名(179語)のデータベースを作成し，音声補完対象の単語辞書(計521語)とした．実験には，20～50代の45名の被験者(男性24名，女性21名)が参加した．

4.1 実験方法

被験者に，音声フォワード補完と音声バックワード補完の入力方法を説明した後³⁾，紙面に記載された単語を異なる条件で音声入力させた．被験者は，まず練習として，ある1単語(「ボクの背中には羽根がある」)について「ぼくのー」や「なんとかー，はねがある」のように指示された通り読み上げて発声し，音声補完を体験した．次に，あらかじめ音声認識システムの単語辞書の中から決められた5単語(曲名もしくはアーティスト名)

- yaen front 4 men feat. saki
(ヤエン フロント フォーメン フィーチャリング サキ)
- 水・陸・そら，無限大
(みずりくそら，むげんだい)
- 恋はスリル，ショック，サスペンス
(こいはスリル，ショック，サスペンス)
- 神経がワレレ暑い夜
(しんけいがワレレあつийよる)
- 関東裸会 三羽鳥
(かんとうはだかかいさんばからす)

を1セットとして，そのセットを以下の3つの条件で入力した．

- 「音声補完」を使用しないで入力
- 「音声補完」を使用して入力
- 「音声補完」を使用するか使用しないかを自由に選んで入力

(2)の条件で，音声フォワード補完，音声バックワード補完のどちらを使用するか(単語中のどの箇所を発声するか)は，単語ごとに自由とした．5単語は全被験者を通じて共通だが，順番をほぼランダムに変えたものを5通り用意して被験者に均等に割り当てた．被験者は紙面に記載された単語セット(読みも記載)を見ながら，まず一番上の条件で5単語を入力し，次に一つ下の条件で同じ5単語を入力していく．ただし，(1)と(2)の条件の順番は，被験者ごとに交互に変わるようにした．

³⁾ 我々は，音声補完を初めて使用するユーザには，新しい音声インタフェース機能として具体的な入力方法を教示することを想定している．そのような教示をしない場合にユーザがどのように振る舞うかの調査は，今後の課題である．

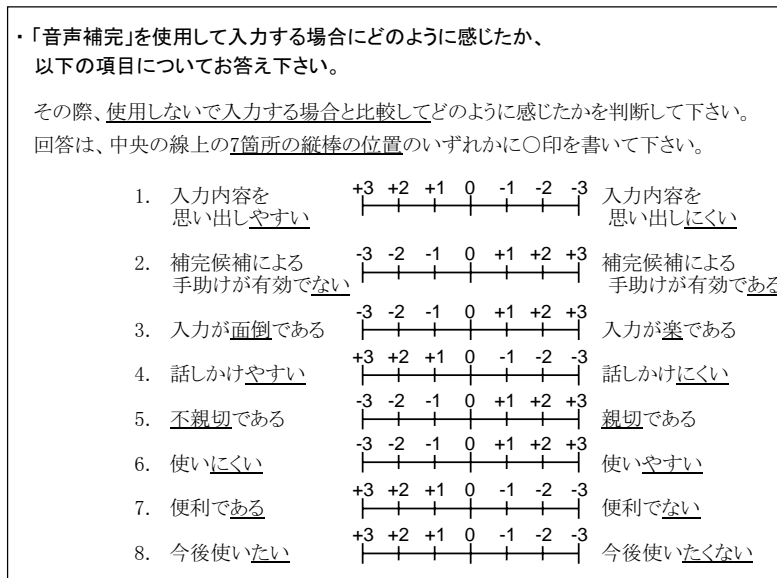


図 3: 被験者に対する音声補完使用後のアンケート (-3 ~ +3 の数字は実際の用紙には記載されていない)

それから、単語セットの紙が取り去られ、被験者は何も単語を見ない状態で、再び同じ 5 単語をそれ以上思い出せなくなるまで入力した。その際、単語の入力順序は自由とした。これを入力条件 (4) とする。

(4) 何も単語を見ない状態で思い出しながら自由に入力

単語を意図的に記憶しないように、被験者は、入力条件 (1) ~ (3) の実験中には、後で思い出しながら入力する実験があることは知らされていない。

最後に、被験者にはアンケートとして、図 3 の 8 項目について 7 段階で評価させた。

4.2 実験結果

実験の結果、被験者全員がすべての入力を完了することができ、音声補完を使用するのに特別な訓練は不要であった。被験者一人が練習も含めて全条件で入力するのに平均で約 7 分間を要した。

入力条件 (3) で、音声補完を使用するか使用しないかを自由に選べるときに、どのように入力したかを調査した結果を表 1 に示す。この結果から、入力条件 (1), (2) で音声入力に慣れた被験者が、74.2% の割合で音声補完を使用して単語を入力したことがわかる。そのうち、音声フォワード補完と音声バックワード補完の使用割合は、前者が 61.7%、後者が 38.3% であった。また、39 名の被験者が 5 回中 3 回以上音声補完を使用しており、1 回も使用しなかった被験者はいなかった。

表 2 は、入力条件 (4) で、被験者が何も単語を見ない状態で入力するときに (うろ覚えの単語を入力する

表 1: 単語を見ながら入力する際に

音声補完を使用するか使用しないかを調査した結果

	使用	不使用
一人当たりの平均回数	3.71 回	1.29 回
割合	74.2%	25.8%

表 2: うろ覚えの単語を思い出しながら入力する際に

音声補完を使用するか使用しないかを調査した結果

	使用	不使用
一人当たりの平均回数	2.56 回	0.62 回
割合	80.4%	19.6%

ときに)、音声補完を使用するか使用しないかを調査した結果である。被験者は平均 3.18 単語を思い出しながら入力できたが、その 80.4% で音声補完を使用していた。そのうち、音声フォワード補完と音声バックワード補完の使用割合は、前者が 66.1%、後者が 33.9% であった。また、被験者が事前に入力する単語を知っていたかどうかをアンケート時に報告させたところ、平均 1.67 単語を既に知っていた (ただし、事前に知っていた単語を必ずしも入力できていなかった)。一つも事前に知らなかった被験者は 11 名いたが、その全員が思い出して入力する際には音声補完を必ず使用していた。

図 3 のアンケートを集計した結果、8 項目のそれぞれについて -3 ~ +3 の 7 段階尺度で評定した値の平均は表 3 となった。評定値の比率を帯グラフで図 4 に示す。最も評定値の平均が高かったのは 2. と 7. の項目で、実際に、補完候補による手助けが有効であり、音声補完が便利であったことがわかる。次に、1.,

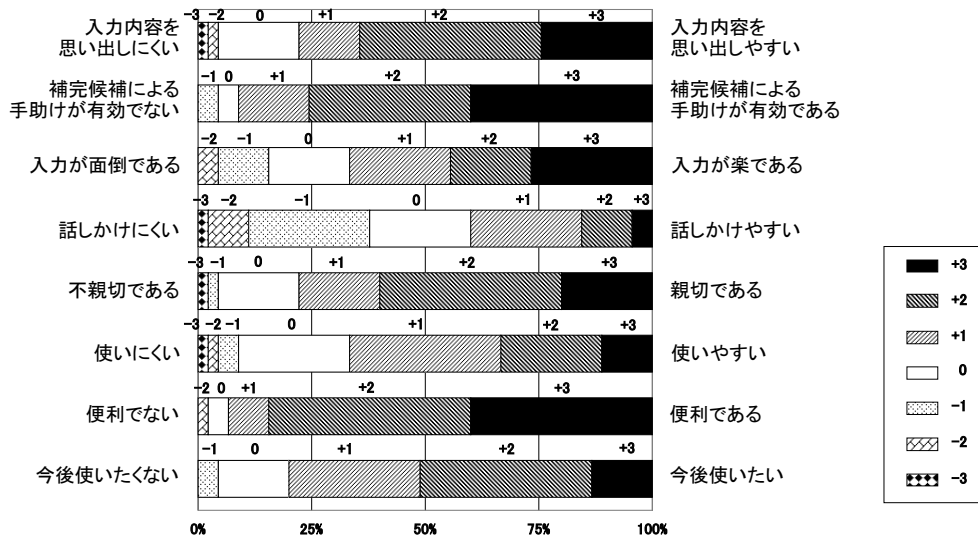


図 4: 音声補完使用後のアンケートの集計結果 (帯グラフ)

表 3: 音声補完使用後のアンケートの集計結果 (評定値の平均)

項目	評定値の平均	+1 ~ +3の占める比率
1. 入力内容を思い出しやすい	+1.56	77.8%
2. 補完候補による手助けが有効である	+2.02	91.1%
3. 入力が楽である	+1.18	66.7%
4. 話しかけやすい	+0.09	40.0%
5. 親切である	+1.49	77.8%
6. 使いやすい	+0.96	66.7%
7. 便利である	+2.13	93.3%
8. 今後使いたい	+1.40	80.0%

5., 8. の項目で評定値が高く, 入力内容を思い出やすく親切であり, 被験者は今後も使いたいと思えていることがわかる. 3. と 6. の項目も, 約 3 分の 2 の被験者が, 入力が楽で使いやすいと支持していた. 一方, 4. については他と異なり, 被験者の反応の大半は -1 ~ +1 にあった. その理由として, 本実験では何を入力するかが明確に指示されていたため, 従来の音声認識に対する話しかけにくさを感じたり, 音声補完が話しかけやすいかどうかを判断できる状況とはならず, 適切に評価されなかった可能性がある. これについては, 今後検討の余地が残された.

以上から, 45 名の全被験者が音声補完を使いこなすことができ, 音声補完を使用するか使用しないかが自由な条件でも, 使用されることが多かったことがわかる. これは, 音声補完が, 使用するのが容易で訓練は不要であり, 実用的な機能であることを意味する. 特に, 思い出しながら入力する際には, 入力単語を一つも事前に知らなかった被験者は必ず音声補完を用いていたことから, うろ覚えの内容を入力する際に効果的であったといえる. また, アンケートの結果からも, 音声補完の有効性が確認できた.

5 議論

音声補完は, 音声認識をインタフェースとして使いやすくするにはどうすべきかという観点から生まれた研究であり, 今後様々な方向への発展が考えられる. 以下では, そのような方向性も含めて議論する.

5.1 関連研究

音声入力インタフェースでは, 効果的な補完機能はこれまで提案されていなかったが, インタフェースにおける補完 (completion) という概念自体は, キーボード入力等で既に広く受け入れられてきた. 例えば, tcsh などの UNIX シェルや Emacs などのテキストエディタは, ファイル名やコマンド名の補完機能を提供している. こうした補完機能では, ユーザが補完機能を呼び出すキー (以下, 補完トリガーキーと呼ぶ. 慣例として “Tab” キーが用いられることが多い.) を押したときに, 途中までタイプされた単語の断片の続きが補われる. このような手動補完以外にも, WWW ブラウザの URL 入力や, Reactive Keyboard⁷⁾ では, 自動補完機能が導入されている. これは, ユーザが

タイプしている最中に、システム側が補完候補一覧を次々と提示していく機能である。また、ペン入力に関しても、自動補完機能を持った予測ペン入力インタフェース⁸⁾やPOBox^{9),10)}が提案されている。これらは、予測インタフェース¹¹⁾とも呼ばれ、有効性が示されてきた。

しかし、音声入力の場合には、上記のような自動補完は不適切な機能となりやすい。キーボード入力や、ソフトキーボードを用いたペン入力では、各キーを押したことを認識する段階で曖昧性はなく、各文字の境界は明らか(明確に分節可能)である。手書き文字認識を用いたペン入力でも、文字の認識時に曖昧性はあるものの1文字の単位は音素より大きく、各文字は分節可能な条件で入力される。一方、音声入力では、音素の認識時に曖昧性が高い上に、音素の境界を決定することも難しい(分節が困難である)。そのため、そもそもどの時点で補完候補を提示するかが一意に決まらず、仮に一定間隔で提示したとしても、キーボード入力等のように高精度で適切な候補を提示し続けることは困難である。「予測を使わない場合に比べて少しでも不都合がある場合には予測インタフェースは使われない傾向がある」¹¹⁾ことから考えても、音声の自動補完は煩わしく、実用的でない可能性が高い。それに対して本研究では、有声休止によってユーザが意図した箇所でも明示的に補完機能呼び出せるようにしたことで、補完して欲しくないときには一切干渉することがない実用的なインタフェースが実現できた。

5.2 音声からテキストへ変換後に補完候補を作成するアプローチに関する考察

文献1),3)で述べたように、本研究では連続音声認識システムを拡張して補完候補を作成する処理を実現した。それに対して、仮に「うただー」と音声入力した発話の断片を、そのまま/utada/のようなテキスト文字列に変換すること(いわゆる音素タイプライタ)ができれば、あとは辞書との文字列間の比較によって候補を生成するようなアプローチも考えられる。しかし、現在の技術水準では音素認識の精度は高くなく^{12),13)}、辞書や文法などの言語的な制約がない状況では多くの誤りを含む。さらに、音素の境界を決定することも難しい。そのため、補完候補を生成するために、認識した音素系列が辞書上の単語の音素の部分列であるかを判定するには、音素間の類似度を考慮しながら時間伸縮や誤りも許容して比較する必要がある。これはまさに連続音声認識システムがおこ

なっていることであり、音素の認識と辞書上の単語との比較をシームレスに実現しているといえる。

5.3 音声中の複数のモダリティーを活用したマルチモーダルインタフェース

従来の音声インタフェースでは、1章でも述べたように、音声認識を中心とした言語情報のモダリティーが主に利用されてきた。それに対して本研究の音声補完では、有声休止のような、音声中に含まれる非言語情報のモダリティーを積極的に利用して、より使いやすいインタフェースを実現した。これは、音響信号が持つ複数のモダリティーを活用した、一種のマルチモーダルインタフェースであると我々は捉えている。

さらに、本研究は、今後他の非言語情報のモダリティー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の構築を目指していこうというメッセージも持っている。キーボード⁴⁾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、音声の音高や話速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。実際に、本研究に続く提案として、高い声をShiftキーを押している状態と見なし、声の高さで音声認識時の入力モードを切替える「音声シフト」^{14)~16)}という機能も研究されている。

しかも、キーボードの機能の範囲に留まる必要はない。音声フォワード補完の有声休止が音韻情報を同時に伝えていたことからわかるように、多くの非言語情報のモダリティーは、言語情報と同時に伝えられるようなメタな情報伝達手段である。このように高い潜在能力を持つ音声のメリットをさらに引き出せば、従来から議論されている音声の利点(ハンズフリー、速い入力速度、等)とも相まって、優れたインタフェースを生み出していけるはずである。

6 応用例: ジュークボックス

提案した音声補完機能付き音声入力インタフェースの応用例として、曲名等を音声認識して、その楽曲を再生するジュークボックスシステム(楽曲検索・再

⁴⁾ 本稿はキーボードが、通常キー(英数字等の文字がそのまま入るキー)と特殊キー(TabやShift等の特別な働きを持つキー)の二種類で構成されると考える。

生システム)を構築した。現在の実装では、4章で述べた179アーティストによる342曲のデータベースを検索できる。基本的には、3章と同様のインタフェースで音声入力するが、音声認識結果が確定すると、それに応じて以下の機能が提供される。

- 曲名が確定したとき
その楽曲のアーティスト名が表示されて、サビが一定時間再生される。
- アーティスト名が確定したとき
そのアーティストの曲名一覧が表示され、ユーザが番号や曲名を読み上げると、上記の曲名が確定したときの動作をする。

本システムは、RWC2001最終成果展示発表会(2001年10月3日～5日)及び産業技術総合研究所情報処理研究部門オープンハウス(2001年10月18日)にて実機デモンストレーションがおこなわれ、多数の来場者が体験した。

上記の機能を実装するために、曲名辞書とアーティスト名辞書を分けておき、辞書の種類を認識結果に付与することとした。また、3章で述べたRVCPによって音声認識結果を受信し、データベースを参照しながら楽曲を再生するプロセスを追加した。実機デモンストレーションでは、説明と実演を繰り返す必要があるため、ボタン(キー)を押すことによって、音声認識部の休止・再開をトグルで切り替える機能も追加した。これは単なるマイクのミュートよりも高機能であり、発話開始後に休止に切り替えても、その発話全体をキャンセルできる。さらに、楽曲の再生中は、音楽がマイクから入力されて誤認識することを防ぐために、自動的に音声認識部が休止するようにした。

7 おわりに

本稿では、発話された単語断片の残りを補うことでユーザの音声入力を手助けする「音声補完」という音声インタフェース機能について述べ、評価実験によってその有効性を示した。音声補完のように、コミュニケーションにおいて言い淀み現象が持つ役割を、インタフェース機能として積極的に活用する研究は従来ほとんどなく、新しい試みといえる。本研究では、楽曲再生可能なジュークボックスシステムを応用例として実現し、曲名とアーティスト名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有名詞の入力等の様々な局面にもすぐに適用できる。音声補完による手助けは、ひとたび使用し始めると、手助けがない状態を不便に感じるほど便利なものであり、

今後、音声入力インタフェースを構築する上で、不可欠な機能の一つになることが予想される。

今後は、補完候補の選択操作の自由度を高めたり(タッチパネルとの併用や候補が絞れる場合の自動確定等)、補完対象を単語よりも長い単位に拡張したりしていく予定である。また、音声補完を発端とした新たな音声インタフェース研究の方向性を探求していきたいと考えている。

謝辞

連続音声認識システム niNja の改良や単語辞書作成等でご協力頂いた秋葉 友良 氏、および、本研究に対し有益な議論をして頂いた麻生 英樹 氏に感謝する。

参考文献

- [1] 後藤真孝, 伊藤克亘, 速水悟: 音声補完: “TAB” on Speech, 情処研報 音声言語情報処理 2000-SLP-32-16, pp. 81-86 (2000).
- [2] 後藤真孝, 伊藤克亘, 速水悟: 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] 後藤真孝, 伊藤克亘, 秋葉友良, 速水悟: 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).
- [4] 後藤真孝, 伊藤克亘, 秋葉友良, 速水悟: 音声補完: 音声ワイルドカード補完機能の実現, 音講論集 春季 3-8-2 (2001).
- [5] Goto, M., Itou, K., Akiba, T. and Hayamizu, S.: Speech Completion: New Speech Interface with On-demand Completion Assistance, *Proc. of HCI International 2001*, Vol. 1, pp. 198-202 (2001).
- [6] 田窪行則: 音声言語の言語学的モデルをめざして — 音声対話管理標識を中心に —, 情報処理, Vol. 36, No. 11, pp. 1020-1026 (1995).
- [7] Darragh, J. J., Witten, I. H. and James, M. L.: The Reactive Keyboard: A Predictive Typing Aid, *IEEE Computer*, Vol. 23, No. 11, pp. 41-49 (1990).
- [8] 福島俊一, 山田洋志: 予測ペン入力インタフェースとその手書き操作削減効果, 情処学論, Vol. 37, No. 1, pp. 23-30 (1996).
- [9] 増井俊之: ペンを用いた高速文書入力手法, インタラクティブシステムとソフトウェア IV, 近代科学社, pp. 51-60 (1996).
- [10] Masui, T.: An Efficient Text Input Method for Pen-based Computers, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'98)*, pp. 328-335 (1998).
- [11] 増井俊之: 予測/例示インタフェースの研究動向, コンピュータソフトウェア, Vol. 14, No. 3, pp. 4-19 (1997).
- [12] 堂下修司: 音声対話システムにおける種々の音声理解方式の比較・考察, 音声による人間と機械の対話, オーム社, pp. 332-339 (1998).
- [13] 伊藤克亘: 音声対話システム, 自然言語処理 — 基礎と応用 —, 電子情報通信学会, pp. 302-322 (1999).
- [14] 尾本幸宏, 後藤真孝, 伊藤克亘, 秋葉友良, 小林哲則: 音声シフト: 音高を利用した音声入力インタフェースの提案, 音講論集 秋季 3-1-8 (2001).
- [15] 尾本幸宏, 後藤真孝, 伊藤克亘, 小林哲則: 音声シフト: 音高を利用した新たな音声入力インタフェース, WISS2001, 近代科学社, pp. 17-26 (2001).
- [16] 尾本幸宏, 後藤真孝, 伊藤克亘, 小林哲則: 音声シフト: “SHIFT” on Speech, 情処研報 音声言語情報処理 2002-SLP-40-3 (2002).