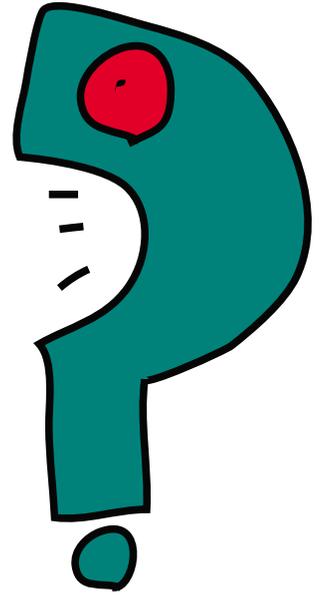


解説

# 音声補完： 言い淀むと助けてくれる 音声インタフェース



後藤 真孝 m.goto@aist.go.jp  
産業技術総合研究所 情報処理研究部門

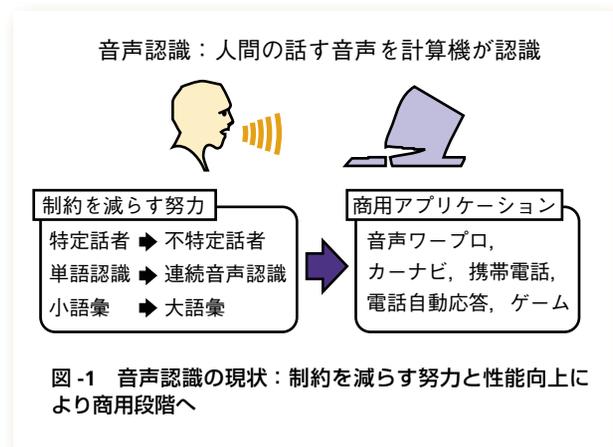
ユーザがある単語を一部しか思い出せずに断片だけを言って言い淀むと、計算機側がその残りを補って入力することを可能にする「音声補完」という音声インタフェース機能を紹介します。従来の音声認識は言語情報しか利用しておらず、非言語情報の1つである言い淀みは、単に誤認識を招く一因と考えられていた。音声補完では、その言い淀みを逆に活用し、ユーザが単語発声途中で有声休止（母音の引き延ばし）によって言い淀むと、それを含む補完候補の一覧を見ることができる機能を提供します。これにより、入力中に困って言い淀めば手助けが受けられる使いやすい音声入力の実現ができた。音声補完の研究は、今後他の非言語情報も活用していくことで、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。

## ■音声認識の現状：あなたはアナウンサーのように話せますか

実は、現在の音声認識は、かなりすごい。アナウンサーが新聞記事を読み上げているような音声に対しては、すでに、95%以上の高い認識率が達成されている。実際に、音声認識技術は身近に使われ始め、実用化の段階に入っている。たとえば、音声で文章を入力できる音声ワープロ（ディクテーション）ソフトウェア<sup>1)</sup>、音声で行き先を指定できるカーナビゲーションシステム、音声で接続先を決められる携帯電話、音声で問合せ可能な自動電話応答システム、音声入力で遊べるゲームなどが各種商用段階になっている（図-1）。

とはいえ、音声認識はまだまだインタフェースとしては使いにくい。なぜか。人間同士がコミュニケーションするときには、音声は非常に便利で快適な情報伝達手段なのに、計算機に音声で話しかけようとする、まだ到底その便利さには到達できていないことが分かる。それには、そもそも話している意味を理解するのが困難であるという人工知能に関連した問題もあるが、それ以前に、現在の音声認識が、人間が口にするさまざまな音

声の、ごく一部の現象しか扱えていないという制約が問題となっている。音声認識の歴史は、特定話者から誰が話しても入力できる不特定話者へ、単語認識から文章が入力できる連続音声認識へ、小語彙から数千語以上の大語彙へ、制約を少しでも減らそうという努力の積み重ねであった<sup>2)</sup>。その結果、アナウンサーのように、文法的に正しい書き言葉を、静かな部屋で淀みなく読み上げたような音声ならば、高い精度で認識できるようになった。しかし、親しい人同士の気ままな日常会話のような、発声があいまいな話し言葉は、音響的にも言語的にも大き



な変動や不規則な現象を含むため、ほとんど認識できない。計算機に向って、アナウンサーのように淀みなく丁寧に話しかけようとしても、何を言えればいいかを口を開ける前にきちんと考えなければならず、だんだん話す気がしなくなってくる。人間同士では快適なはずの音声、計算機が相手になると使いにくく感じるのは、こういった現在の音声認識技術の持つ制約が一因となっている。

では、音声認識を便利で快適な入力手段にするには、どうすればよいのであろうか。それには、**インタフェースの観点から音声認識を捉え直す**ことが重要である。従来の音声認識研究の多くは、いかに認識率を上げるかを競ってきたが、それをインタフェースとしてどのように使うかは十分には検討されてきていない。1つの方向性として、人間同士の対話のように計算機と対話することを目指す音声対話技術も研究されているが、人間は対話だとくだけた発話をしがちであり、音声認識が困難な場面も多い。そこで以下では、現在の音声認識技術を活かしながら、どうすれば音声インタフェースをより使いやすくできるかを考えていきたい。

## ■音声の持つ豊かな情報：言語情報と非言語情報

音声は豊かな情報を含んでおり、インタフェースの観点から見ると、大きな潜在能力を持っている情報伝達手段である。音声の持つ情報は、**言語情報** (verbal information) と**非言語情報** (nonverbal information) <sup>☆1</sup> に大別できる (図-2)。言語情報は、音韻や単語のような書き言葉の持つ情報で、文字で容易に表記される。一方、非言語情報は、韻律 (イントネーション、アクセント、発話リズム、声の高さ、話す速さなど)、言い淀み、声質、話者の感情・態度のような情報で、文字としては表記されないか、正確に書き起こしにくい。たとえば、同じ「はい」という言語情報でも、普通に語尾を下げる「はい」や、語尾を上げて聞き返す「はい?」、最初の音を延ばす「はいー」、しぶしぶ言う「はい」などのように、非言語情報によって、多様な情報が伝達される。このように人間は、言語情報と非言語情報の両者を活用しながら、自然に音声から豊かな情報を得てコミュニケーションしている。

しかし、現在の音声インタフェースは、この音声の持つ潜在能力を引き出していない。これまでの音声認識研究は、言語情報をいかに音声から得るかに焦点を当ててきた。非言語情報の中で韻律に関しては、主に音声認識率を向上させる目的で利用を試みられたこともあったが、「現在実用に供されている音声認識技術ではピッチ

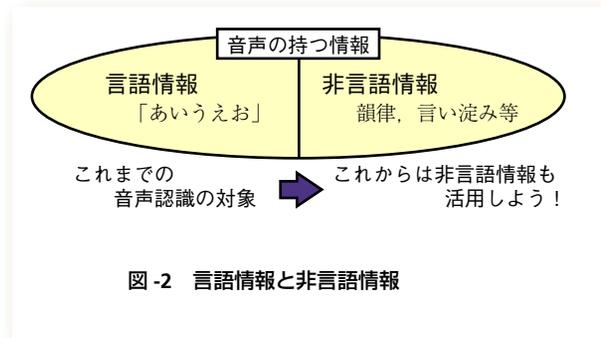


図-2 言語情報と非言語情報

情報はまったくといってよいほど利用されておらず<sup>4)</sup>、「韻律情報は対話処理に関しては重要だと思われるが音声認識にはその有効性があまり示されてない」<sup>5)</sup>と指摘される状況である。このように、非言語情報を用いていない現在の音声認識は、いわばキーの認識誤りを起こすようなキーボード (文字を入力しようとしてもなかなか完全には入らないキーボード) にしか過ぎず、いくら音声認識率を高くしたとしても、キーボードを超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。

では、音声の持つ豊かな可能性を引き出すために、認識率向上とは視点を変えて、インタフェースの観点から非言語情報を活用してみると、現在の音声認識技術でも、もっと使いやすい音声インタフェースを構築できるのではないだろうか。以下では、このような発想で生み出された新たな音声インタフェース機能「音声補完」を紹介する。

## ■音声補完：言い淀むと助けてくれる新たな音声インタフェース

「**音声補完**」 (speech completion) <sup>6), 7)</sup> は、非言語情報の1つである言い淀み現象を積極的に活用した音声インタフェース機能である。言い淀み現象には、有声休止 (音節の引き延ばし)、無声休止 (次の発話内容を考えているときの無音区間)、言い直し (例:「イタ、イタリア料理店は」) などがあるが<sup>☆2</sup>、ここでは、特に話者の思考状態が現れやすい**有声休止** (filled pause) を活用する。有声休止は、発話したい内容が断片的にしか思い出せないときや、何を発話していい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、たとえば、話者が「音声補完」

<sup>☆1</sup> ここでは便宜上、パラ言語情報<sup>3)</sup> は非言語情報に含まれるものとする。

<sup>☆2</sup> このような言い淀みは、書き言葉には通常現れない、話し言葉特有の現象といえる。人間が考えながら話す場合には高い頻度で出現するが、書き言葉を読み上げたような音声を前提とする音声認識では、誤認識を招く一因であった。

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（図-3）。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、対話相手がさまざま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である1つの理由となっている。

このような対話相手による「音声補完？」という手助けは、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。この**補完** (completion) の概念は、テキストインタフェースですでに広く受け入れられている。たとえば、tsh や bash などの UNIX シェルや、GNU 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、**補完トリガキー**と呼ぶ）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。また近年、WWW ブラウザの URL 入力や、携帯電話の文字入力、ペン入力においても、自動補完機能（文字を入れるごとに、補完候補が次々と提示されていく機能）を持ったインタフェースが実用化されている。しかし、音声入力では、自然に補完機能を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提案されていない状況であった。音声補完では、このような補完による手助けという概念を初めて音声認識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に成功している。

#### □音声補完で何ができるか

「音声補完」は、音声入力中に、ユーザが補完機能を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することによって、以下の3つの利点が得られる。

- **記憶補助** 入力したい内容がうろ覚えでも、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、すべての音を最初から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



図-3 音声補完：全部を思い出せなくて言い淀むと計算機側が残りを補って手助けをしてくれる

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に対しては、煩わしく不適切な機能となりやすいからである<sup>☆3</sup>。そこで音声補完では、有声休止に補完トリガキーの役割を担わせることで、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を呼び出すことを可能にしている。

音声補完の対象には、単語や文節、文章などさまざまなレベルが考えられるが、現段階では単語補完機能のみが実装されている。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、たとえば姓名が1つの単語として登録されているときには、姓だけの部分発声から残りが補完される。

ユーザは、図-3の「おんせいー」の場合のように必ずしも単語の頭が分かるわけではなく、末尾や中央部分に基づいて補完したい場合もある。そこで、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、図-4の3種類の音声補完方式が提案されている<sup>☆4</sup>。ここでは、「宇多田ヒカル」、「小柳ゆき」、「Blankey jet city」という単語辞書上の1単語を補完する場合を例に説明する。

<sup>☆3</sup> キーボードでは各キーを押したことを認識する段階で曖昧性がなく、文字を入力することに補完候補を出せるが、音声では文字の境界を決定することすら困難なため、そもそもいつ補完候補を出すがが一意に決まらない。また、仮に一定間隔で出しても、適切な候補を精度よく出し続けることは難しく、煩わしくなりやすい。

<sup>☆4</sup> 実行画面を収録したデモンストレーションムービーを、<http://staff.aist.go.jp/m.goto/SpeechCompletion/index-j.html> で見ることができる。

1. 音声フォワード補完   
「うただー」 ➡ 「宇多田ヒカル」
2. 音声バックワード補完   
「なんとかーゆき」 ➡ 「小柳ゆき」
3. 音声ツーウェイ補完   
「なんとかーじえっとー」 ➡ 「Blanky jet city」

図-4 3種類の音声補完方式

1. 音声フォワード補完 (単語の頭が分かる場合)  
単語の前半 (頭) が分かっているときに、その最後の音節で有声休止を行うことで、それに続く後半を補完する方式である (画面表示例: 図-5).
2. 音声バックワード補完 (単語の末尾が分かる場合)  
単語の後半 (末尾) が分かっているときに、「なんとか」などの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行い、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 (画面表示例: 図-6).
3. 音声ツーウェイ補完 (単語の中央部分が分かる場合)  
単語の中央部分が分かっているときに、上の2つの



(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

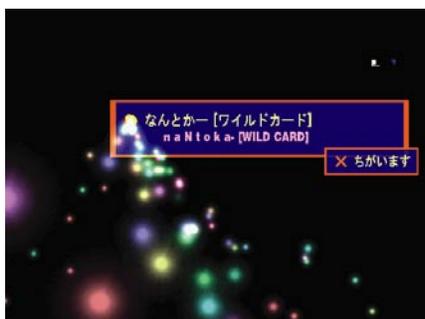


(5) 1番の候補が輝いて跳躍



(6) 1番の候補「宇多田ヒカル」が認識結果として確定

図-5 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-6 音声バックワード補完中の画面表示例

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を呼び出すことで、頭と末尾を補完する方式である。

上記のような入力を行うと、補完候補（単語）の一覧が、番号付きで即座に表示される（図-5 (3)、図-6 (b)）。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（図-6 (b)）、「前（の候補）」や「次（の候補）」と言えば他候補が見られる。こうしてユーザは候補の中から入力したいものを見つけ、3通りの方法（候補の番号を言う、単語の残りを読み上げる、単語全体を頭から読み上げる）で選択する。すると、選択された候補が強調表示されて、音声認識結果として確定する。

#### □音声補完の仕組み

このような音声補完の機能を持つインタフェースを構築するには、どんな技術が必要かを考えよう。まず、任意の単語中の有声休止を、リアルタイムに検出するための技術がなければならない。次に、検出した有声休止をトリガとして、補完のための候補を作成・提示する技術も必要となる。ここで注意しなければならないのは、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（「宇多田ヒカル」）が辞書中に1つしかなくても、音声の曖昧性から候補を1つに絞り込めないことが多い、ということである。そこで、発声と音声認識の単語辞書との音響的な類似度に基づいて、常に複数の補完候補を作成する必要がある。たとえば、図-5の例では、音響的な類似度から、「1. 宇多田ヒカル (*Utadahikaru*)、2. 深田 恭子 (*Fukadakyouko*)、3. MR.DYNAMITE (*Misutadainamaito*)」のような補完候補が得られている。以下では、このような処理を実現する技術について簡単に紹介する。

#### —有声休止をどうやって検出するか

音声補完では、有声休止を高い精度でリアルタイムに検出する必要がある。また、ユーザがどの単語のどの母音で有声休止を行うかは事前に分からないので、言語に依存せずに（言語情報を一切使わずに）検出しなければならない。

こうした条件を満たす手法として、文献8)の有声休止箇所の実時間検出手法が採用された。この手法は、有声休止が持つ次の2つの音響的特徴

- 声の高さがほぼ一定のままである（基本周波数の変動が小さい）。

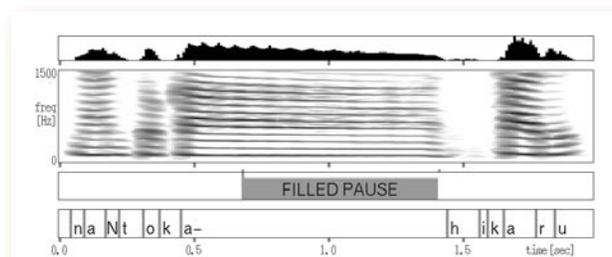


図-7 「なんとかーひかる」に対する有声休止区間の検出例：パワーとスペクトルの時間変化（上2段）とシステムが検出した有声休止区間（中段）、有声休止区間での音素遷移を抑制した音声認識システムが出力した音素系列（下段）

- 母音が途中で変わらずに声が持続する（スペクトル包絡の変形が小さい）。

を信号処理によって検出することで、任意の母音の引き延ばしを検出できる。検出結果の例を図-7に示す。検出性能に関して、再現率と適合率のトレードオフは調整でき、音声補完の目的に合うように適合率を高く（誤検出を少なく）設定することが可能である。自由発声音声の対話コーパスを用いて、音声補完で言い淀むような400ms以上の継続時間を持つ有声休止を対象に評価すると、再現率0.91、適合率0.93の性能が得られている<sup>8)</sup>。ここで決定した有声休止開始点は、次に述べる拡張された音声認識システムへ送られる。

#### —補完候補をどうやって作成するか

補完候補の作成は、隠れマルコフモデルに基づく典型的な音声認識システムを拡張して実現する（ここでは、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みで可能である）。通常の音声認識システムでは、音声の各発話区間に対して、それが事前に用意された単語辞書上のどの単語に最も似ているかを求める。単語辞書は、図-8のように各経路が異なる単語に対応する木構造で保持され、入力された音声がこの木構造上のどの経路に近いかを評価する。その際、音声には曖昧性があるため、さまざまな経路をたどる仮説を考え、適切なものを選ぶ必要がある。この認識処理の最中の仮説を、図-8ではくさび形のマークで表す。発話が開始されると、刻一刻と、これら多数のくさび形マークが根 (root) から葉 (leaf) の方向へノード (node) をたどって移動していき、発話終了時点で、葉にたどり着いた仮説の中で最ももっともらしい（尤度の高い）仮説が、認識結果と見なされる。

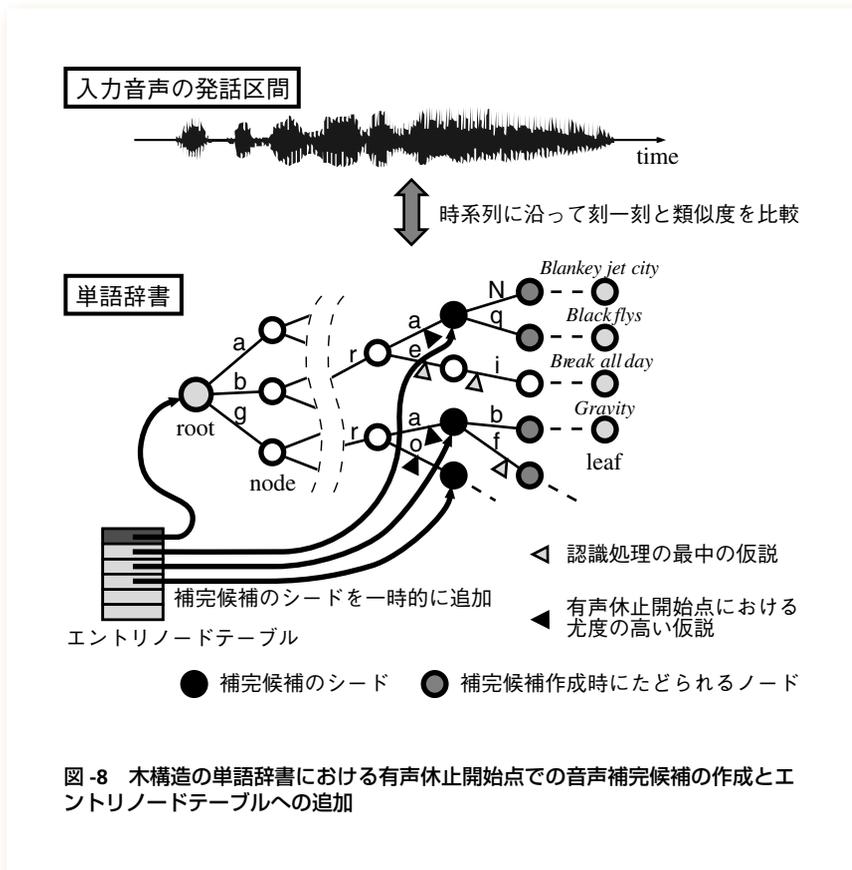
音声補完ではこの仕組みを流用して、入力された断片と音響的に類似した部分を持つ単語を補完候補として求

める。ここでは簡単のため、音声フォワード補完をどう実現するかだけを考える。まず、任意の時点で有声休止が検出できるように、上記の通常の音声認識システムに加えて、有声休止の検出処理を常に並列に動作させておく。そして、発声中に有声休止が検出されると、音声認識システムに割り込みがかかった状態となり、その時点で尤度の高い上位数個の仮説（図-8の黒いくさび形マークが表し、入力された断片と音響的に類似した部分の末尾に位置する）から葉の方向へたどって補完候補を作成し、尤度の高い順に番号付けて、上位数十個を候補とする。つまり、有声休止が検出された時点でもっともらしい仮説は、もしその続きが普通に発声されたとすると、その先にある葉へ高い尤度で到達していくことが期待されるので、その葉に対応する単語を補完候補にしようというアイデアである。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。

こうして、「うただー」と言えば「宇多田ヒカル」のような候補が得られるようになったが、今度は、次の発語で「ひかる」と言って候補を選択できなければならない。ただし、通常の音声認識では、辞書登録されているものしか認識できないため、そのままでは「ひかる」は認識できない。単語中のあらゆる音で言い淀む可能性があるため、そのような途中からの語句をすべて辞書に登録するのも現実的でない。そこで、単語の続きを言って選択するための工夫として、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-8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。これにより、前述した例で「ひかる」と言ったときに、エントリノードテーブルから追加エントリ「ひかる」へ飛んで認識可能となる。音声バックワード補完、音声ツーウェイ補完も、このエントリノードテーブルを活用して実現するが、誌面の制約から、詳細は文献7)に譲って省略する。

#### □音声補完は実際に有効だったか

以上述べてきた音声補完機能を持つ音声インタフェースシステム（現在はノートパソコン上でも動作可能）に



より、45名の被験者（男性24名、女性21名）に対して、音声補完の有効性を確認する実験がなされた。音声補完対象の単語辞書（計521語）には、日本のポピュラー音楽のヒットチャートから得た曲名（342語）とアーティスト名（179語）のデータベースが用いられた。その結果、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入力しても、使用されることが多いことが分かった。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、実用的な機能であったことを意味する。特に、うろ覚えの内容を入力する際に効果的であることも分かった。被験者に対するアンケートの結果からも、音声補完の有効性が確認されている<sup>7)</sup>。

さらに、音声補完の応用例として、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・再生システム）も実現されている。音声補完の有用性は曲名とアーティスト名の入力で確認されているが、これは住所入力や各種固有名詞の入力等のさまざまな局面にもすぐに適用できる基本的なアイデアである。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、音声インタフェースを構築する上で、不可欠な機能の1つになることが予想される。

## ■今後の展望：音声インタフェース研究に新しい流れを

今後の音声インタフェースはどのような方向へ向って行くのだろうか。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、今後の音声インタフェースに関していくつかの方向性を示唆している。

まず、音声対話の一場面をインタフェース機能として適切に切り出して提供することで、音声認識をより使いやすくできることを音声補完の研究は示している。一般にインタフェースを実現するには、どのように情報交換するかをユーザとシステムとの間でルール化する必要がある。それがユーザにとって自然で快適であるほど使いやすいインタフェースとなる。音声補完では、人間同士の対話からヒントを得て、言い定めば手助けをしてくれるという、音声対話で人間が経験している自然な情報交換が実現されている。さらに、3種類の音声補完方式としてルール化することで、現在の技術で実装可能な機能となっている。このように、インタフェースの観点から音声対話を見つめ直し、新たな機能としてルール化することで、より快適な音声インタフェースを実現できる可能性がある。

次に、音声補完のどこが優れていたのかを改めて考えてみると、実は、高いリアルタイム性、レスポンスの良さが鍵になっており、これは今後の重要な方向性となると考えられる。多くの音声認識では、発話が終了して初めて認識結果が表示されるというフィードバックが得られる。一方、音声補完では、音声入力中に言い淀むと、発声中に即座に候補が出るような遅延の短いフィードバックが得られ、そこで選択して入力するというインタラクティブな入力手段を提供している。こういったユーザから見た使用感の良さは、音声認識率のような数字には表れないが、インタフェースの観点からは非常に大切な要素といえる。

また、今後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を構築していこうというメッセージも持っている。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（英数字等の文字がそのまま入るキー）の一部に過ぎない。それに対して、音声補完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キー）に相当している。これを第一歩として、声の高さや話す速さなど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、音声補完に続く提案として、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り替える「音声シフト」<sup>9)</sup>という機能も研究されている。また、言語情報を用いずに非言語情報だけで操作するための研究<sup>10)</sup>もなされている。

しかも、非言語情報を活用する際に、インタフェースとしてキーボードの機能の範囲にとどまる必要はない。音声フォワード補完の有声休止（「うただー」の「だー」）が言語情報である音韻（「だ」）を同時に伝えていたことから分かるように、多くの非言語情報は、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。音声は、言語情報と非言語情報を同時に伝達できることが本質であり、そこにこそ音声インタフェースの将来性があると筆者は信じる。今後、音声の持つ大きな潜在能力を引き出す多様な音声インタフェース機能が研究されていくことを期待したい。

**謝辞** 音声インタフェースの在り方について議論し、音声補完を共に研究してきた伊藤克巨氏、速水悟氏、秋葉友良氏に感謝する。また、学会発表等の場でご議論いただいた多くの方々に感謝する。

### 参考文献

- 1) 西村雅史、伊東伸泰：音声ワープロー過去・現在・未来－、情報処理、Vol.40, No.2, pp.164-167 (Feb. 1999).
  - 2) 河原達也：ここまでできた音声認識技術、情報処理、Vol.41, No.4, pp.436-439 (Apr. 2000).
  - 3) 藤崎博也：韻律研究の諸側面とその課題、音講論集 秋季 2-5-11, pp.287-290 (1994).
  - 4) 前川喜久雄：韻律とコミュニケーション、音響誌、Vol.55, No.2, pp.119-125 (1999).
  - 5) 中川聖一：音声認識研究の動向、信学論 (D-II)、Vol.J83-D-II, No.2, pp.433-457 (2000).
  - 6) 後藤真孝、伊藤克巨、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)。(情報処理学会山下記念研究賞受賞)。
  - 7) 後藤真孝、伊藤克巨、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、コンピュータソフトウェア、Vol.19, No.4, pp.10-21 (2002).
  - 8) 後藤真孝、伊藤克巨、速水 悟：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)、Vol.J83-D-II, No.11, pp.2330-2340 (2000).
  - 9) 尾本幸宏、後藤真孝、伊藤克巨、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, pp.13-18 (2002).
  - 10) 五十嵐健夫、Hughes, J. F.: 言語情報を用いない音声による直接操作インタフェース、WISS2001、近代科学社、pp.7-10 (2001).
- (平成 14 年 10 月 4 日受付)