

UI 設計とユーザビリティ —音声インタフェースの課題—

石川 泰

三菱電機株式会社 情報技術総合研究所

Ishikawa.yasushi@eb.MitsubishiElectric.co.jp

あらまし 音声認識技術の発展により、音声インタフェースの実用化が進んでいるが、その利用分野の拡大、普及は十分に進んでいるとは言えない。本稿では、音声インタフェースをタスク達成のインタラクションの視点で、GUI と比較しながら、ユーザビリティ視点で整理し、その問題点を検討することで、今後の音声インタフェース設計の課題を議論する。

キーワード 音声認識 音声インタフェース 音声対話 GUI UI 設計 ユーザビリティ

UI Design and Usability - Subjects of Speech Interface -

Yasushi Ishikawa

Information Technology R&D Center, MITSUBISHI Electric Corporation

Abstract In association with technical progress of speech recognition, products with speech interface have been developed. However, the spread of application areas and the diffusion are still inadequate. In this paper, the usability of speech interface is described from the viewpoint of interaction for task completion comparing GUI. And from investigation of problems, the subjects of speech interface design are discussed.

Keywords Speech Recognition, Speech Interface, Speech Dialog, GUI, UI design, Usability

1. はじめに

音声認識技術は、1990年代に入り、統計的手法の導入により不特定話者音声認識が十分に高い性能を示すようになり、プロセッサ技術の進歩により、製品開発が本格化した。電話系音声認識技術による CTI (Computer Telephony Integration) システム、コールセンター応用が注目され、ディクテーション S/W が販売され、携帯電話などへの利用が図られてきた。しかし、現在、カーナビゲーションでの利用がハンズフリー、アイズフリーという利点か

ら実用システムとして継続的に注目され、利用されてはいるものの、音声認識技術者の期待からは、その利用分野も、普及も十分とは言えない。

一方、ユーザインタフェースの重要性は、機器・システムの機能の複雑化、扱う情報量の飛躍的拡大、携帯電話やデジタル音楽プレーヤなど小型機器の普及を背景に、製品価値を決定する大きな競争軸としてますます高く認識されており、高度なグラフィック処理による GUI が開発されている。音声認識の利用を推進す

るためには、単に音声人間にとって最も自然な情報伝達の手段であるということに基づいた音声認識の有効性の主張や、ハンズフリー・アイズフリーといった制約からの利点の訴求ではなく、音声認識の特長、制約・問題点を考慮した UI 設計技術の検討が必須である。

もちろん、これまでも、音声認識の実用化を阻害する要因の分析や、課題の検討が行われ、すぐれた知見が示されてきた[1][2][3]。音声認識技術の問題、音声対話の制御技術・対話記述の技術の問題、さらには、人間機械間のインタラクションがどうあるべきか、これら議論はすぐれた先行研究にゆだねることとし、本稿では、GUI との比較により、音声インターフェースがどのような特徴を有するかを検討する。

2 章では、ユーザインターフェースにおけるユーザビリティについて概説し、3 章では、従来の音声インターフェース設計例を GUI と比較することで、ユーザビリティ上の問題の明確化を試み、4 章において、これらの検討から音声インターフェースの課題を議論する。

2. UI とユーザビリティ

ユーザビリティは Nielsen によれば、「学習容易性」「効率性」「記憶性」「エラー」「満足度」により決まるシステムの有用性と定義される[4]。この定義は、人間工学、認知工学、感性工学の観点から考察されたものであり、UI (ユーザインターフェース) の問題点抽出の視点で重要な着眼点である。しかし、評価の観点では、相互の関係性も強いいため、1998 年に制定された ISO9241-11 では、「指定された利用者によって、指定された利用の状況下で、指定された目標を達成するために用いられる際の有効さ、効率および、満足度の度合い」として定義されている。ここでは、この定義に基づき、主に有効さと効率について検討する。ここで、有効さ

は、「目標を達成する上での正確さと完全性」を、効率とは「ユーザが目標を達成する際に、正確さと完全性に費やした資源」と定義される。ただし、これらを一般的に論じることは極めて困難であり、また音声認識における誤認識の問題を含めると、複雑になるため、ここでは、タスクを単純化した場合についての GUI との比較により検討する。

また、ユーザビリティの評価においては、しばしば、「目標の達成」、すなわち評価タスクを規定した評価が行われる。しかし、携帯電話やカーナビゲーション、デジタルテレビなど、今日の情報機器では、その機能の高度化・多様化により、ユーザがシステムに対して使用時に抱く目標が、対象システムで実現できるかどうか分からない、あるいは逆に、システムが保有する機能がユーザに利用されない、つまり目標とされないことが頻発する。すなわち、実際のシステムとユーザ期待とのミスマッチである。UI デザインでは、このようなミスマッチを防ぎ、すべての機能をユーザに提供することが重要となる。広くは、「有効さ」「効率」にかかわる問題であるが、具体的なミスマッチがどのように起きるかを含め検討することとする。

3. GUI と音声インターフェース

対象システムの機能、UI を一般化することは困難であるが、ここでは、単純化し、機能選択型のタスクと、データ入力型のタスクに分類する。さらに、単純な音声のみのインターフェースと、GUI の典型的手法について比較検討する。

3.1. 機能選択タスク

3.1.1. GUI

GUI に着目すると、ユーザが希望する機能を多くの機能から選択する UI は、図 1 のように、木構造で表現される階層化されたメニューの選択で表すことができる。すなわち、ある状

況からのユーザの目標に相当するノードへの到達問題である。

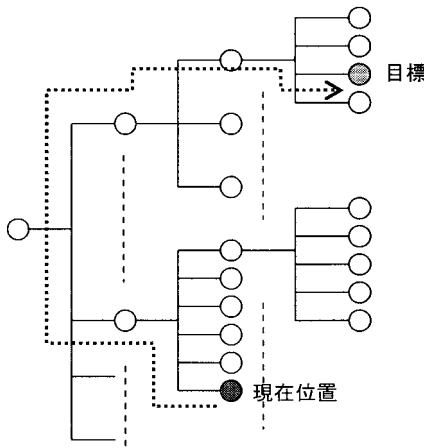


図 1. GUIによる機能選択タスク

したがって、効率は、現在の位置から、目標のノードまでの、操作回数(階層の深さ)と、各階層でノードを選択するために要する資源(労力)によって決まる。一般に、このノードの選択のためには、言語表現によるメニュー、あるいは、グラフィカルアイコンによる選択、またはその組み合わせが用いられる。労力は、それらのメニューの理解、選択のための操作に費やされる。

一方、「有効さ」は、提示される用語や、グラフィカルアイコンがユーザの目標、あるいはその上位概念をあいまい性なく、表現しているかによって、正しく目標に到達できるかという問題として考えられる。

3.1.2. 音声インタフェース

機能選択タスクの音声による実現手段は種々考えられる。ひとつには、GUIと完全に一致した処理体系での実現であり、ポインティングデバイスあるいは、H/W スイッチの代わりに音声を利用しようとするものである。このような

設計思想であれば、基本的に比較は、音声入力の労力と GUI、H/W 操作の労力との差異、および、グラフィカルアイコンあるいは、表示文字と、音声コマンドの想起の比較となる。しかし、一般には、音声認識の語彙の大きさの活用が図られる。その典型例がすべての状態から、目標操作に到着するための音声コマンドを定義する手法である。

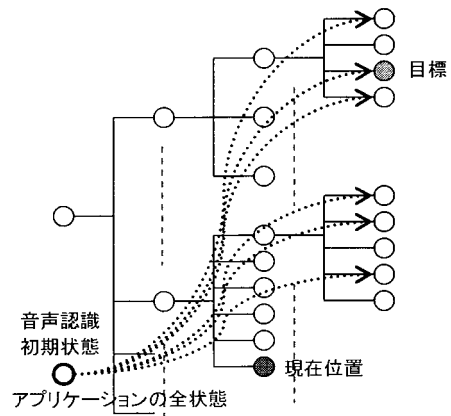


図 2. 音声コマンドによる機能選択タスク

このような方法をとることにより、操作回数が減少し、効率向上が期待できる。しかし、「有効さ」について考えると、目標機能を表現する、設計者の発想による一意な「音声コマンド」をユーザが選択できるかどうかという問題が生じる。音声認識では、音声コマンドが「日常使用している言葉」であるという、人間同士のコミュニケーションのアナロジーから、容易にユーザが記憶できると誤解される場合がある。機器の機能が複雑になればなるほど、すなわち、GUI 操作のインタラクションが増え、1 発話による入力の有効性が高い場合ほど、このコマンドの一意性の保持と、ユーザの期待との合致は困難になり、音声インタフェースの有効さを著しく低下させる要因となる。

連続音声認識、自然言語理解技術は、ユーザの目標に対する音声操作の自由度を高めるものではあるが、自由度を高めることは、ユーザ発話が、目標のノードだけではなく、GUI におけるその上位ノードの表現である場合や、それが現在の状態に依存した表現になることをも許容しなくてはならず、困難な課題であることは言うまでもない。

さらには、GUI の操作が、操作の範囲をメニュー、アイコンなどでユーザに提示しているのに対して、音声インタフェースでは、一般に操作の範囲が明示されない。したがって、上述のように、ユーザがシステム機能について理解していない場合、実際にはない機能の要求がなされ、また、GUI 操作により、機能の学習ができる、すなわち、提示されることにより、ユーザが知らない機能を学習できるのに対し、音声インタフェースのみからは、学習が進まない可能性も高い。

3.2. データ入力型タスク

一般にスロットフィリングで表現されるタスクを考える。スロットに入る値の種類が少ない場合はメニューの選択としての実現ができるが、ここでは、スロットに入力する値の種類が多い、あるいは、制限がない場合を考える。

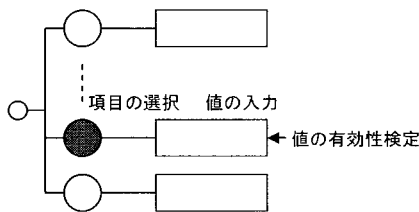


図 3. データ入力型タスク

このタスクを一般化すれば、図 3 に示すように、複数項目についての値入力のタスクとなる。

したがって、ユーザは、なんらかの方法で、項目を選択し、値を入力、さらには、その値がシステムにとって意味のある値かどうかを検定されるとともに、用意された項目の中からタスクを完了するために必要な項目をすべて入力することで、タスクが完了する。

3.2.1. GUI

GUI における効率は、いかに容易に項目が選択できるか、そして値の入力の労力がどのようなものかにより決まる。項目の選択では、ポインティングデバイスなどを用いてユーザが選択するか、システムが順番を規定するかいずれかの方法により実現され、その効率をタスクの特異性にあわせて検討すればよい。値の入力については、値の種類が比較的少ない場合には、GUI のリストセレクションボックスや、ラジオボタンによる選択が有効であり、この場合、値の有効性の検定も同時に可能となるというメリットがある。値の種類が大きい場合は、キーボードなどの入力手段を用いて入力することになり、労力が大きいことが予想できる。

また、「有効さ」については、項目の選択のために、表示されたテキストあるいはグラフィックアイコンがユーザの期待といかに合致し、誤らずに選択できるか、システムが項目の入力順序を規定しているとき、その順序がユーザの期待に合致し、判断誤りを生じさせないかが検討項目である。また、ユーザの入力がシステムの入力値として有効なものとして正しく受け入れられるか、さらには、ユーザが入力したい項目と、システムのタスク完了のために必要な項目とに不一致が生じないかどうか重要な要素となる。

3.2.2. 音声インタフェース

音声インタフェースの実現では、項目選択の実現手段として、GUI 同様システムが入力項

目の順序を規定する場合と、ユーザが項目を選択できる場合とが考えられる。これは、基本的に上記のとおり音声インタフェース特有の問題ではない。しかし、以下の点でそのユーザビリティには大きな相違を生じる可能性がある。すなわち、GUI が順番を規定していても、入力のタイミングまでは規定しないのに対し、音声対話においては、システムが入力を要求する場合、入力のタイミングまでを規定し、かつ、その場合に、ユーザに要求どおり発話する以外のことができない状況を強いる場合が多い。このことは、ユーザに大きな労力を要求することになり、とくに、その順番がユーザの期待と一致しないときは大きな問題になる。また、ユーザが選べる場合には、項目をあらゆる語と設計者の期待との一致が必要であり、これが満たされず、ユーザの期待と異なる項目が選択された場合、著しい効率の低下、あるいは、有効さの低下につながる。

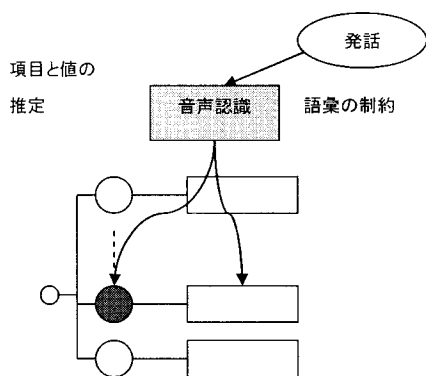


図 4. 音声インタフェースによるデータ入力

また、値の入力においては、ユーザの発話が、正しく認識できれば、効率の上で、他の入力手段に比べ有効である可能性は高い。しかし、「有効さ」について考えると、システムの期待する入力値として有効かどうか、入力時点では判断できず、また、リジェクトが正しく行わ

れない限りは、システムは、入力値を有効な値にマッピングする(誤認識)するという問題が生じる。このとき、何らかの方法で、ユーザが入力した内容が有効ではなかった、つまり語彙外発話であったことが判断できれば、有効さについても、大きな問題を生じることはないが、音声認識はたとえ、待ち受けている語彙を発声しても誤る場合があることを理解しているユーザには、語彙外か誤認識の判断はできない。

このことが、有効さの大きな低下につながる。

3.3. タスクに依存しない問題

一方、GUI と音声インタフェースの比較では、現状では GUI がはるかに入力の多様性がある点を見逃してはいけない。すなわち、タッチパネルに触ること、キーを押すこと、ダイヤルを回すこと、などに加え、動作感覚の制御に基づくハプティックデバイスや、タッチパネルなどの上や、カメラを用いたモーションの認識など、近年急速に多様な入力手段が実現しつつある。これらは、一見発話による明確な意思の伝達に比べ、あいまいな感があるが、実際には、現実のもの操作とのアナロジーにより、直感的に操作の質が理解しやすいメリット、さらには、新たな操作方法についての驚き、楽しさなどこれまでに議論してきた「有効さ」「効率」とどまらず、「満足」への訴求要因となっている。

4. 音声インタフェースの課題

4.1. 「効率」と「有効さ」

音声インタフェースの検討は前述のように、種々の側面からすぐれた検討がなされてはいるが、実際の製品開発の現場では、インタフェースのユーザビリティ視点からの検討が十分には行われず、「有効さ」における問題を、場合によっては「言語表現であること」で解決できているかのように考え、音声認識の「効率」に

についての基本的なポテンシャルを訴求する UI デザインが行われることが多い。

これが、正しくないのは、従来のコンピュータのインタフェースにおける「コマンド」の入力手段が、キーボードから、音声認識に変更されただけの状況を考えればあきらかである。

筆者は、一般的な解決方法を提示はできないが、以下の課題を検討していくことで、知見を蓄積し、音声認識の有効性をそれぞれの対象で効果的に活用する手法が製品として広がり、音声認識技術の広い実用化が果たせるものと考えている。

1) 複雑なインタラクティブシステムの評価手法の検討と標準化

ともすると、タスクの理解があるものとして評価がなされ、現状の組み込みシステムや、デジタル機器のインタフェースに見られるシステムの機能とユーザ理解のミスマッチが評価されない。評価方法を確立し、評価と改良のプロセスが UI デザインの場で行えるようにする必要がある。

2) 音声の「有効さ」の検討

音声インタフェースは、ユーザにとり理解が困難な現時点で操作可能な機能と、その表現方法が設計者理論で規定されてしまっている。このミスマッチを防ぐためには、単なる効率の訴求によるデザインをせず、生じえるミスマッチのインスペクション評価を十分に実施することが必要である。また、機能の自己表現性は、GUI がすぐれているとすれば、GUI との一貫性や、あいまいな入力を設計レベルで許さないのではなく、あいまい性を解消し、学習できる音声インタフェース設計も重要である。

3) GUI で困難な問題の検討

上述のように GUI は自己表現性が高いが、できることの制約や、深い階層の機能選択タスク

ではその構造の理解ができないための「有効さ」の低減が生じる。音声で直ちにこれを解決するのではなく、音声入力によるヘルプ機能など、相補的なインタフェースの設計も重要である。

4) 音声インタフェースの「満足」

GUI は直感的な操作手段を数多く、提供し、提示するコンテンツ・システム状態のグラフィカルにリッチさでユーザの「満足」の向上を図っている。データ入力や機能選択といった明確な目的達成場面ではない、利用方法、他の入力手段の補助手段としてリッチなインタフェースを実現する方法、あるいは、タスクの検討も必要である。

5. まとめ

音声インタフェースと GUI を比較し、インタフェースのユーザビリティを考察することで、課題を検討した。具体的な解決策の提示は筆者には困難であり、検討も稚拙ではあるが、今後、音声開発者の相互議論を進展させ、音声インタフェース実用化を推進したい。

【参考文献】

- [1] 嵯峨山、「音声認識の実用化を阻むもの」信学ソサイエティ大会、Vol.1995 情報・システム、pp.312-313 (1995)
- [2] 畑岡、「音声技術実用化の課題と取り組み」、情処学会、研究報告 SPL-055 (2004)
- [3] 庄境ほか、「音声言語情報処理技術デベロッパーズフォーラム」情処学会、研究報告、SLP107 (2006)
- [4] ニールセン、「ユーザビリティエンジニアリング原論」東京電機大学出版局 (2002)