

音声 UI 設計とその評価 -普及促進に向けて，システム構築の立場から-

石川 泰[†]

音声認識技術は，統計的手法による認識方式など長年の研究の蓄積により，高い精度の認識方式が確立し，近年の MPU の高速化，メモリの低価格化により，組み込みシステムにおいても大語彙音声認識が実現できるようになった。特に，運転を阻害しないことに対する期待から，カーナビゲーションシステムでの実用化が進められている。しかし，開発者の期待に対して，現状，十分な普及が進んでいるとはいえない。これらの理由については，種々の優れた技術的な検討が行われているが，実際の製品開発について考えると，UI 構築の技術的方法論に加え，開発コスト，エンジニアの知識，さらには，開発・評価・改良のサイクルの形成などにも大きな問題がある。本稿では，これらの問題を整理し，解決のための一方法として，連続キーワード入力からのユーザ意図推定方式を提案する。有効性を論じ，開発環境としてのメリットと音声 UI 普及への貢献の可能性を示す。

Design of Speech UI and Its Evaluation - Towards the Wide Spread of Speech UI from Developers' Point of View -

Yasushi Ishikawa[†]

Speech recognition has been progressed by the accumulation of studies such as statistical acoustic and language models. Now the method which provides practical good performance is established. This method of large vocabulary speech recognition is available on embedded systems with progress of MPUs and memories. The practical speech interface has been realized especially on automotive systems, because of expectations for an interface which does not distract use's attention from driving. However, contrary to developers' expectation, still the spread of speech interfaces is not sufficient. The reasons of block of usage have been discussed from technical points of view. But when we consider development of products, there are many reasons not only technical but also problems of costs, knowledge of developers, process of development, evaluation and improvement. In this paper, a method of speech UI in which user's goal is extracted from connected key words is proposed as a solution to these problems, then its advantages and possibilities are discussed

1. はじめに

音声認識技術は，基本的な統計的認識手法，騒音下音声認識方式の研究の蓄積により，認識率の観点では十分に精度の高い方式が確立されている。特に，近年の MPU の高速化，メモリの大容量化・低価格化により，組み込み機器においても大語彙連続音声認識が十分に動作するようになってきている。特に，運転動作を阻害しないという特徴から，カーナビゲーションシステムをはじめとする車載機器での音声認識の実用化は高い期待を受け，1990年代中盤から実用化が進められてきている[1,2]。

しかし，この間，技術的には認識機能の大語彙化，高騒音下での認識精度の向上などがあったものの，開発者の期待にくらべて，普及が十分に進んでいるとはいえず，いまだに車載機器の必須機能として大多数のユーザが利用する状況とはなっていない。

音声認識がなぜ利用されないのかについては，これまでも多くの優れた研究・議論がなされてきている[3]。しかし，これらのうちの多くの議論は，ユーザ視点，あるいは実環境下での音声インタフェースの性能に焦点をあてた技術的な問題点，あるいは，マンマシンインタラクションの視点でのユーザビリティ上の問題点を中心に議論されている。これらの議論は音声インタフェースのユーザビリティ向上のために，重要な示唆と与えるものであり重要ではあるが，開発者視点の問題点，また多くの議論がなされながら改善が進まない理由，さらには，今後の車載情報機器のアプリケーションの変化，それに伴うシステム構築上の問題点を検討する必要がある。

本稿では，2章において，開発者視点での問題点を述べ，3章において，従来の音声 UI (User Interface) の代表的な設計手法を示し，4章において筆者らが提案するフレキシブルショートカットの概要を示す。その上で，5章において将来の車載機器への要求について述べ，開発者視点で，本方法が従来の問題点を解決できる可能性を示す。

2. 音声 UI 設計における開発者視点での問題点

現在の車載機器は，ナビゲーション機能については，3D 表示などの高機能化・POI (Point of Interest, 施設名，名所などの目的地となりうる情報) の大規模化など機能向上が著しい。さらにこれに加え，CD のリッピングによる大量の楽曲の蓄積が可能なオーディオ機能，欧米では重要な機能として位置づけられているハンズフリー電話機能とそのための電話帳の処理機能に加え，路車間通信に基づく交通情報の提供などその機能は複雑になってきている。音声によるインタフェースはこのような複雑な処理に対して有効なインタフェースと考えられているが，実際の音声 UI の設計開発においては，以下の問題が生じている。

[†] 三菱電機株式会社 情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation.

- 1) 音声 UI についての設計の基準，設計手法が確立されておらず，設計においては経験や予想に基づいた設計が行われる．また，その設計者が必ずしも音声認識，UI デザインについての専門家であるとはかぎらない．
- 2) GUI や H/W のスイッチキー設計では，視認性，GUI 制約下での理解しやすさなどのみを考える専門家によりなされる．音声 UI 設計者との共同作業が行われず，音声 UI 設計が GUI デザインありきで検討されることも多い．
- 3) 音声 UI を含んだ UI 評価には多大な労力がかかるだけでなく，評価方法が確立していないため，事前の専門家によるインスペクション評価の確認しか行われないケースがある．
- 4) 対象機能，アプリケーションが複雑化していることもあり，音声 UI の構築には，多大な時間と費用がかかるため，プロトタイプ作成と改良をひとつの対象製品の開発に対して繰り返し行うことが不可能に近い．
- 5)

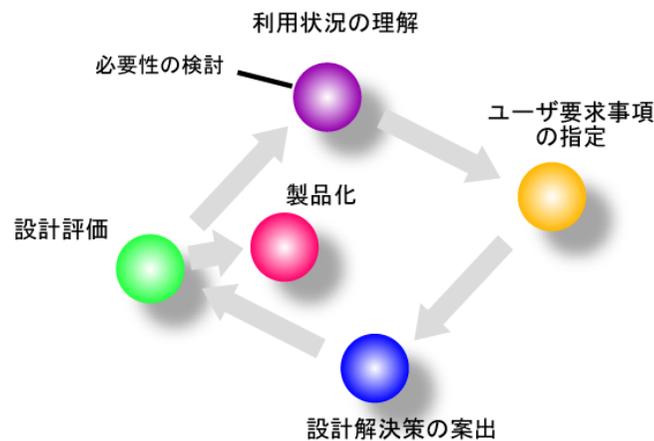


図 1. インタラクティブシステムにおける人間中心のデザインプロセス (ISO-13407)

たとえば，ISO-13407[4]には，対話型のシステムにおける人間中心のデザインプロセスが規定されている．しかし，音声 UI デザインにおいては，このプロセスを踏むことが上記理由から困難であり，プロセスを十分に経ないで製品化が進む場合もある．具体的には，

- 1) 必要性について，漠然とした音声の有効性，運動動作への妨げないことを理由とし，具体的な根拠が明確でない場合がある．

- 2) 利用状況の理解においては，運転中であるという環境の想定は容易にできるものの，ユーザが音声 UI に対してどのような理解をもっているかなど，ユーザの状況理解が困難である．また，他の入力手段の特長欠点などの総合的な UI 環境の想定も容易ではなく，音声 UI のみが検討対象となる場合がある．
- 3) 設計・評価を行うには，試作システムを構築する必要があるが，多大なコストが必要であり，また評価手法が確立されていない．
ことが問題である．

したがって，音声 UI 設計においては，単に音認識・音声 UI 技術の専門化が，設計解決策の一事例や概念だけを示すのみではなく，実際の対象のシステム，その具体的な GUI など音声以外のモダリティによる UI 設計，さらには，対象ユーザを明確にし，有効性を評価し，その問題点を把握，ユーザの要求を抽出，改善策を考えるプロセスを，開発現場で実現しなくてはならない．

3. 従来の音声 UI デザイン

3.1 音声 UI の対象タスク

音声 UI の設計については，対話の主導権や，音声入力に対する制約などをさまざまな視点からの分類が成り立つが，ここでは，階層化が可能なシステムの機能やデータを選択を対象とし，ユーザが現在の対象アプリケーションの状況からなんらかのシステムとの対話を行い，目的の機能・データまで到達するタスクを考える．このような対象タスクにおける階層化は通常 GUI あるいは H/W キー入力デザインの中で行われることが多い．これは，GUI や H/W キーが面積や，キーの数に対する制約があるため階層化せざるを得ないという理由も考えられるが，一方では，対象の機能やデータがユーザにとって理解されやすいものとするための有効な方法であるとも理由として考えられる．対象業務を分類し，上位の概念からユーザに「提示する」ことにより，対象が理解されやすく，またその分類方法に一貫性があれば，ユーザは対象の機能を効率的に把握することが可能であると予想できる．これは，インタラクションが若干多くても，GUI が広く一般ユーザに受け入れられてきた理由と考えられる．

3.2 ショートカット型音声 UI

このような階層型の機能に対して，音声コマンドを独立して設計し，ひとつの機能をひとつの音声コマンドで実現しようとする方法を，ここではショートカット型音声 UI と呼ぶことにする (図 2)．このような方法は，音声認識の語彙数が，GUI の階層的なメニューのエントリー個数に比べはるかに大きく設計できること，また音声認識が「自然言語」を扱えるため，語彙が大きくなってもユーザが特別にコマンドを記憶するなどの労力は必要としないだろうという考え方に基づく方法といえる．

しかし，実際には，機能の数だけのユニークな音声コマンドを設計する必要があり，

多少の言い換え語を追加しても、ユーザの期待とシステム設計のミスマッチが生じる(図3)。ミスマッチは、語彙外発話を生じ、誤認識によるシステムの動作は、ユーザに理解されず、混乱のみを招くという問題点がある。

3.3 対話型音声 UI

一方、複雑な機能・データ選択のための処理を、1 発話ではなく、対話的に処理をすることで、上記問題点を解決しようという方法を対話型音声 UI と呼ぶこととする。このとき、音声のみのインタフェースを考えると、プロンプトの設計や、対話設計の自由度の設計、対話の主導権などを着目し、種々の方法が考えられるが、ここでは、GUI・キー入力がある UI を対象として考えるため、有効な方式としては、音声対話の設計が、GUI・キー入力と一貫性がある設計となる。この場合は、GUI の特長である対象の理解のしやすさに加え、入力可能な語彙が限定され、さらに文脈に依存した用語を用いることができるため、音声の操作方法も理解は容易になる(図2)。

しかし、これは、キー入力やタッチパネルの入力を音声で実施しているにすぎず、音声認識に期待される大語彙認識に基づく利便性を捨てることを意味する。さらに、誤認識などを考えると GUI・キー入力に対する優位性には疑問が生じる。

3.4 自由発話理解

音声 UI 設計の視点で、自由発話理解を考えると、図3に示すように、音声 UI において生じ得るユーザの発話行為と、システム設計における語彙・文法設計のミスマッチを、待ちうけ語彙・文法を拡張することで解決しようとする方法と考えることができる。しかしこの場合、ユーザの発話行為をどのように想定するのが問題になるだけでなく、受理したユーザ発話から、ユーザが目的とする機能・データを推論する方法が問題となる。自由発話では、ユーザ発話の文脈・状況依存性をどのように解決するかが極めて困難な課題であり、いくつかの試みはあるものの[5]、実用的なシステムを実現することは容易ではない。特に複雑で多種多様な機能を扱わなくてはならない車載機器においては、現実的ではない。

3.5 データ入力型タスク

階層化が困難、あるいは関連のない複数データの入力タスクでは、入力データの順番の制御を音声対話の主たる課題として扱い、主導権制御や、その実現手段としてのスロットフィリングが音声 UI の問題として検討されることがあるが、入力項目の選択について考えると、ショートカット型、対話型の UI での実現もとして考えることができるため、ここでは陽に扱わない。

4. フレキシブルショートカット

図4に、筆者らが提案するフレキシブルショートカットの動作原理を示す[6]。この提案方法では、基本的な構造を対話型 UI とし、この方法の問題である非効率

性を、連続発話入力と、簡単な語彙の倒置や省略に対する遷移先推定と、あいまい性解消対話によって解決しようというものである。

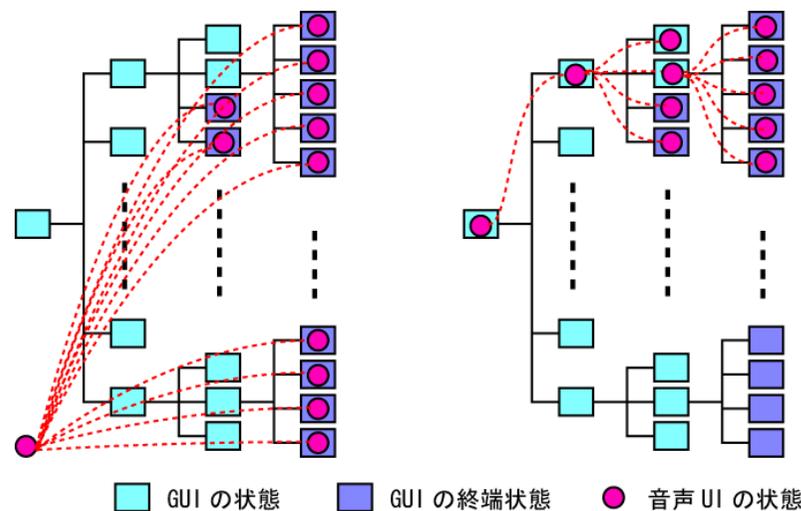


図2. ショートカット型 UI (右) と対話型 UI (左) の原理

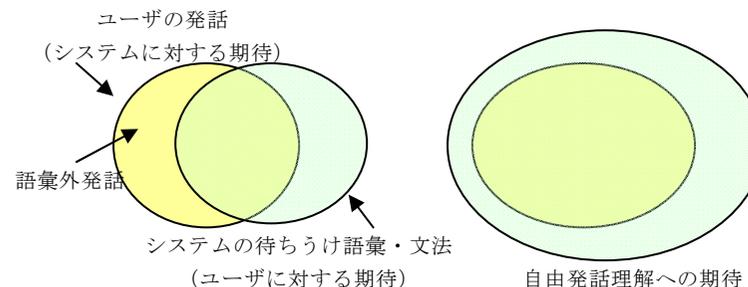


図3. ユーザ発話とシステム設計のミスマッチと自由発話理解

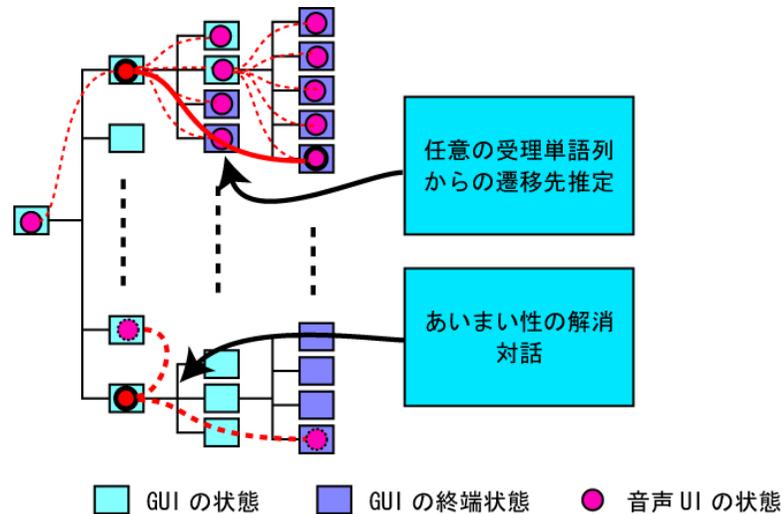


図 4. フレキシブルショートカットの動作原理

具体的には、通常の対話型 UI の音声コマンドに当たる語彙をキーワードとして GUI の状態遷移に対してすべて定義しておく。現在の状態から GUI の状態遷移と同じ状態遷移は、複数のキーワードの連続発声により実行できる。さらに、あいまい性がなければ、キーワードの転置や省略があっても、遷移可能となり効率的なショートカットが実現できる。また、遷移先があいまいであった場合は、ユーザーに複数の遷移先候補と、その遷移に必要なあいまい性のない連続発話入力、すなわちキーワード連鎖を示し、ユーザーの選択により遷移先を決定する。

この方法によれば、ユーザー視点では

- 1) 初心者には、従来の対話型 UI 同様、GUI 構造をたどることでシステムの機能や扱うデータの範囲、すなわちシステムの機能構造を確認しながら操作可能。
- 2) 慣れたユーザーで、階層構造を理解していれば、キーワードを連続発声することで効率的なタスク達成が可能。
- 3) さらに、あいまい性解消対話により、あいまい性のない音声入力方法、システムの機能構造を学習することが可能。
- 4) GUI との一貫性により、総合的な学習も容易。また、GUI 操作から音声操作に移行することも容易。

という特長がある。さらに、システム開発者の視点では、以下の特長がある。

- 1) 音声 UI を、状態遷移の手順をプログラムすることなしに簡単に構築可能。
- 2) 理解しやすい GUI デザインが行われていれば、システムの機能構想の表現性がいいことを示し、音声 UI も理解しやすい構造となると予想される。音声 UI と GUI のデザインを同時に実施することを意味し、設計者が異なるための問題が生じにくい。
- 3) 音声 UI 設計者は、GUI デザインを元に、分かりやすく、連続発声が容易なキーワードのみ設計すればよく、設計が容易。ひとつの階層内でキーワードのあいまい性がないように設計すればよく、全体の複雑な機能に対する一意性、妥当性を考える必要がない。
- 4) GUI 設計変更に対して、上記 3 の確認を行えばよく、変更・改良が容易。
- 5) 設計者は音声 UI についての知見、知識を必要としない。

5. 今後の実用システムの特徴と提案手法の有効性

5.1 今後の車載機器の特徴

車載機器ではこれまで、ナビゲーション機能、オーディオ機能、電話のハンズフリー通話機能などを、組み込みソフトウェアとして実現し、その総合的な UI を設計・実装してきた。しかし、車載機器は、今後通信インフラの整備により、車外のデータの利用、車外サーバの機能の活用が進む。また、コンテンツについても、携帯電話や携帯音楽プレーヤ、家庭の情報家電と共有するようになると予想される。

また、個々の機能についても、車のライフサイクルが長いこともあり、ユーザーの満足のため S/W のアップデートや、新たなサービスの追加、ダウンロードなどが行われるようになると考えられる。

5.2 今後の車載機器の UI への要求

このような状況では、開発時のアプリケーションやデータを基に UI を構築することはできず、動的な UI 再構築、また動的データの利用可能性が必須機能となる。

しかし、個別の UI を持つ S/W が、単に車載機器のプラットフォーム上で実行されることは、一貫性のない UI となり、使いにくいばかりではなく、運転行動に対する障害がないことが保障できなくなるため、車載機器としては極めて大きな問題となる。

車載機器の UI は、通常の PC とは異なり、S/W やサービス、データの変更や追加があっても、一貫した操作性をユーザーに提供し、かつ運転障害を生じさせないことが保障されたものでなければならない。

5.3 提案手法の有効性

これに対して、提案手法では、システムの機能構造であるメニューの階層構造と、各遷移に対するキーワードだけが定義されていればいいため、図 5 に示すように、UI

の動的な再構築, 修正が可能である. この再構築はシステム運用中にも可能である.

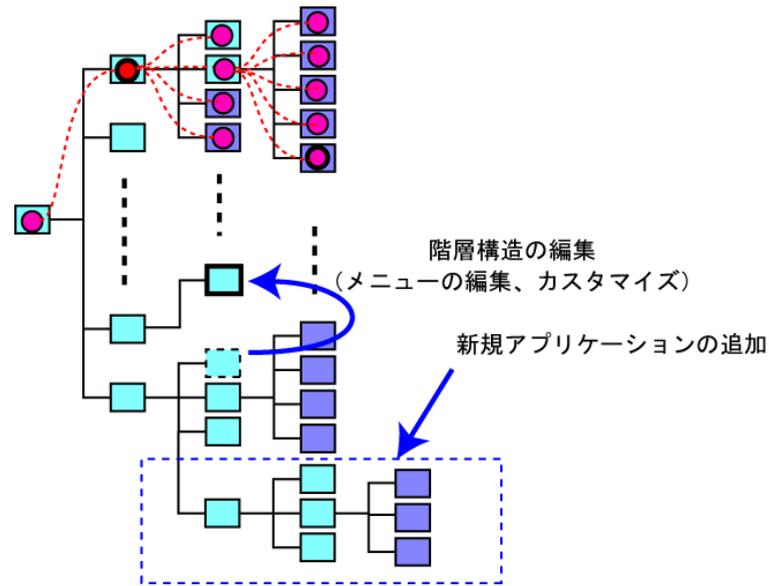


図5 フレキシブルショートカットにおける UI の再構成

このとき, GUI を含めて, 機能構造からの UI 再構築が可能であるため, GUI 実現方法に, 表示アイコンの数や, スクロールなどに対する制約事項, 運転状況に応じた音声入力などの操作制約などを備えておけば, 新たな機能やアプリケーションやデータの追加, またユーザのカスタマイズに対しても, 統一され運転を阻害しない UI を実現することが可能である.

筆者らは, すでに, XML 形式の機能記述方法を検討し, 音声・GUI の自動構築の可能性を実証している.

5.4 開発のサイクル

提案したフレキシブルショートカットは, GUI との親和性の高い音声 UI を, 機能構造表現から, 自動生成する手法であり, 音声 UI の開発を容易にし, 一貫性のある UI を持つ製品を短期間に市場に投入することが可能である. これにより, ユーザの経験と理解が増し, 音声インタフェースの普及促進が可能となると考える.

しかし, システムの機能構造の表現方法の優劣, その GUI 実装方法, あるいは, 音

声 UI のためのキーワード定義, さらには, 機能構造上の音声による状態遷移到達可能領域の定義, キーワード列からの推定方法, などによりそのユーザビリティは大きく影響を受ける可能性がある. したがって図 1 に示す設計プロセスは依然として重要である. これを効率的に実施するには, 本手法に限らず, UI 設計の基本原則を包含した上位 S/W コンポーネント, すなわち UI エンジンを開発するとともに, 音声認識のためのリソースを共有することが重要である. さらに, 評価手法を確立し, その評価結果・知見を音声 UI デザインの設計基準として共有することが重要な課題である[7] 広く音声 UI 設計者が, 知識や経験なしに, 設計・インスペクション評価が可能ない開発環境を構築することが必要である.

6. まとめと今後の展望

音声 UI の設計について, 従来からのユーザ視点に加え, 開発者視点の問題点を示し,これを解決するひとつの音声 UI 手法を提案し,その効果を議論した. 提案手法は, GUI との親和性が高く, ユーザに理解しやすいのみではなく, 効率的操作も両立できる方法である.

しかし, 開発者がさらに, ユーザビリティの高いシステムを実現するには, 評価改良のプロセスを効率的に実行できるよう, UI レベルの S/W ツール可, リソース整備が必須である. さらに, 開発者間でそれら評価結果, あるいは評価から得られた知見を共有することが, 普及促進の鍵となるであろう. 今後は, この実現に向けて, 具体策を検討する予定である.

参考文献

- 1) 小窪, 天野, 畑岡: 車載用音声認識における騒音対策とその評価, 信学会論文誌, J83-D-II(11), pp.2190-2197
- 2) 山崎, 野木, 岩崎: カーナビゲーションの音声認識技術, 三菱電機技報, 74(9) (832) pp.563-566 2000/9
- 3) 嵯峨山: 音声認識はなぜ使われないか・どうすれば使われるか?, 情処学会研究報告, SLP 94 (40) pp. 23-30
- 4) 黒須, 堀部, 平沢, 三樹: ISO13407 がわかる本, オーム社, 2001 年 12 月
- 5) 永井, 石川: 概念素に基づく音声理解への統計的言語制約の導入, 情処学会研究報告 SLP 98(68), pp.99-104
- 6) 熊井, 中野, 小林, 石川: アプリケーションの機能構造に基づく音声インタフェースの提案と評価, 情処学会研究報告, SLP 2007(75), pp.109-114
- 7) 中野, 小林: Proxy-Agent を核とした双方向型音声認識アプリケーション開発支援の実現, 情処学会研究報告, SLP 2009(10), pp.63-68