

探索的検索のための音声入力インタフェースの検討

西本 卓也^{†1} 岩田 英三郎^{†2}
櫻井 実^{†2} 廣瀬 治人^{†2}

音声認識技術は普及しつつあるが、音声インタフェースがより有効に利用されるためには、音声入力の利点や欠点を踏まえ、マルチモーダルインタフェースのトータルなデザインが重要である。このような観点から、我々は、オンラインショッピングなどにおいて候補の検索や比較検討などの要素を含む探索的検索を、音声インクリメンタル検索とドラッグ&ドロップ操作によって直感的に行うインタフェースを提案・試作したので報告する。

A Voice Interface System for Exploratory Search

TAKUYA NISHIMOTO,^{†1} EIZABURO IWATA,^{†2} MINORU SAKURAI^{†2}
and HARUTO HIROSE ^{†2}

A multimodal interface system is developed to verify the effectiveness of speech in exploratory search. The prototype system is designed in consideration of the basic principles and organization principles of interface that are required for comfortable input systems. Desired organization of interface using speech and graphical user interface is discussed from the viewpoints of less physical demand, easiness to learn, and transparency of the system.

1. はじめに

我々がコンピュータや情報機器によって行う作業の多くは探索的検索 (exploratory search)¹⁾ の要素を含んでいる。つまり選択肢の中から既知の項目を効率よく指定するのではなく、複数の候補を比較検討しながら自分が望む項目を選んでいく場合が多くなっている。例えばウェブでショッピングを行う場合には、条件を指定していくつかの候補を絞り込み、それぞれの候補の価格や仕様や評判情報などを閲覧し、場合によっては条件を修正しながら最終的な候補を選びとっていく作業となる。また、情報機器や携帯電話の操作においても、何かを行いたい場合に階層メニューをたどるだけでは所望の操作を見つけ出すことができなったり、所望の機能がどのような名前で登録されているのかが分からなかったりするため、探索的検索を必要とする場合が多い。

キーボード入力によるキーワード検索は探索的検索の有用な手段と考えられる。例えばPCの操作においてもデスクトップ検索は必須の機能となりつつある。しかし、情報家電やモバイル機器などではキーボード

を用いることが困難な場合もあり、このような場面では音声入力がある有力な手段になりうる。

近年の音声認識技術の進歩は目ざましく、種々の優れたシステムが実現されている。しかし、代替の入力手段を考慮してなお音声入力の必然性を感じさせるアプリケーションを開発することは容易ではなく、音声認識の利用は必ずしもヒューマンインタフェースの改善につながっているとは言えない。

音声インタフェースの研究は、要素技術の個々の改良の積み重ねによって洗練されたトータルシステムを作るという、いわば還元主義的な思想で進められる場合が多かった。これに対して小林・西本ら^{2),3)} は、望ましいインタフェースシステムとはどのようなものであるか、といった全体像を描いた上で要素技術たる音声入力のあり方を検討し、グローバルな意識からマルチモーダル環境における音声入力の有効利用を提案した。

本稿では、キーボードによる文字入力は不可能であるが、画面表示とポインティング操作 (タッチパネルなど) が使用できる、という前提で、探索的検索のための音声入力インタフェースの設計について、ヒューマンインタフェースの原則³⁾ に基づいて議論し、試作システムの実装および評価について報告する。

^{†1} 東京大学 大学院情報理工学系研究科

The University of Tokyo

^{†2} ユニバーサルロボット株式会社

Universal Robot Inc.

2. インタフェースの原則と音声利用

2.1 視覚障害者支援技術からの知見

我々は視覚障害者と対面朗読者の音声対話を分析し、弁当屋のメニューから食べたいものを選ぶ対話が「対象の全体的な把握」「分類や検索による大まかな選択」「お気に入りリストへの追加」「お気に入りリストからの最終決定」によって構成されていることなどを確認し、この知見を活かして、音声合成とキーボードで操作可能なウェブシステムを試作した⁴⁾。

視覚障害者に対する支援を想定した対話として、弁当屋のメニューから食べたい商品を選ぶという課題を選んだ。メニューには多くの商品が記載されており、商品名だけでなく、価格、商品の説明、写真など、さまざまな情報が含まれる。気分や体調などによって利用者の食べたいものは変化する。また、最初に食べたいと思ったものが、その店のメニューに含まれているとは限らないため、妥協して別の候補を探す場合もある。いわば、ユーザの漠然とした要求をだんだん具体化していく探索的検索タスクである。

試作されたシステムは被験者におおむね好意的に評価された。一方で、同じような対話を音声認識によって実現するべきであるか、という質問に対しては「人間(対面朗読者)との対話よりもキーボードの上下キーなどを用いたスクリーンリーダ操作の方が快適である」といった意見が得られた。我々は「対面朗読の熟練者はタスクを効率的に達成するための対話を行う」と期待してこの実験を行った。しかし例えば「直前に読み上げられた情報を聞き返す」という場合に人間同士の対話では「もう一度読んでください」といった依頼が必要であるが、スクリーンリーダでウェブサイトを読む場合にはカーソルキーの操作だけで聞き返し操作が可能である。

音声インタフェースシステムの設計においては単に人間同士の対話を模倣するだけではなく、既存の入力手段の利点や欠点を踏まえて、快適なインタフェースのあり方を合理的に検討する必要がある。

2.2 インクリメンタル検索の有効性

ヒューマンインタフェースの原則³⁾では、インタフェースの基本原則として「操作労力」「システムの透過性」「頑健性」の3つを挙げている(付録 A.1)。マウスとキーボードを用いたインタフェースにおいては近年「操作労力」「システムの透過性」に関する配慮が多くなされるようになった。いわゆるインクリメンタル検索は、GNU Emacs などのテキストエディタだけでなく Apple iTunes や Google Suggest などのシステムで用いられている。Raskin⁵⁾ はインクリメンタル検索の有効性を特に強く主張しており、音声入力によるインクリメンタル検索の実現可能性も示唆している。

インクリメンタル検索は

- 操作労力の原則：ユーザの操作が妥当であることを、絞り込まれた結果や候補数の提示などによってフィードバックする

- システムの透過性の原則：所望する候補が選択可能になれば入力を打ち切ることができるため、ユーザに操作労力削減の機会を与える

などによりインタフェースの基本原則に適合している。音声によるインクリメンタル検索もインタフェースの基本原則に適合すると考えられるが、そのためには

- 発話中に候補やその個数を逐次表示する
- 発話の途中で内容が確定したら、発話を中断しても許容される

という要件を満たす必要がある。

2.3 音声入力のリアルタイム性

従来の音声認識システムは性能を重視する一方で認識結果を出力する際の遅延はやむを得ないものと見なす場合が多い。多くの大語彙連続音声認識システムが、trigram 言語モデルを適用したり N-best 候補を得るために2パスの探索を行っており、入力同期の第1パス処理だけでは最終結果を得られない。しかし、音声認識システムに対するしゃべり方や発話内容が妥当であるか否かを発話終了まで知ることができないのは、音声入力システムの「透過性」においては不十分である。これに対して、人間同士の対話において、人間は相手の表情から反応を読むことができる。人間は一方が話している間も頷いたり首をかしげたりするだろうし、相手の発話が聞き取りにくければ直ちに「え？」などと聞き返すことができる。自分の発話を理解しているのか理解できないのかを示してくれる相手とは、会話がしやすいのではなからうか⁶⁾。

音声認識技術におけるリアルタイム性の検討は過去にも行われてきた。例えば放送字幕作成のための音声認識システム⁷⁾ においてはニュース音声の認識結果を2秒以内に得ることを目標とし、2パスデコーダを改修して一定フレームごとに結果を確定する、といった手法が提案されている。また音声認識エンジンからの情報に基づいてユーザの発話中に頷いたり相槌を打つ音声対話システムが NTT や早稲田大学などで試作されており、従来の HMM による音声認識に限らず WFST などの利用も提案されている。これらはリアルタイム音声インタフェースの可能性を示すデモであり、自然な音声対話を実現したとされているが、音声インタフェースの汎用的な枠組みの提案には至っておらず、操作の効率性に貢献する提案とは言えない。

2.4 音声入力と効率性

音声認識は効率的な入力手段として期待されるが、スイッチやボタンなど何らかの代替手段が利用できる場面では、入力モダリティの適切な組み合わせを考慮する必要がある。音声入力においても「長い文章やコマンドは発話することに多くの労力を要する」と考え

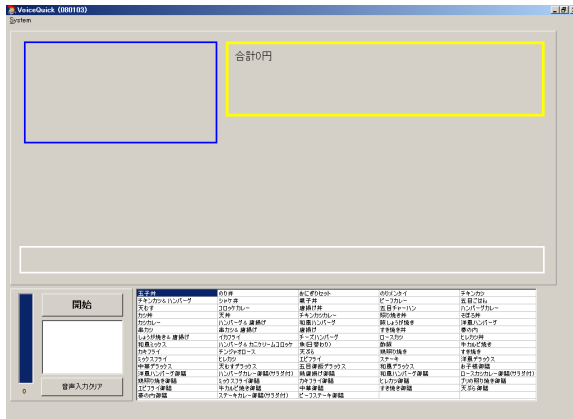


図 1 システムの全体画面（初期状態）

なくてはならない。ユーザに無駄な入力や発話をさせないために、音声インクリメンタル検索をマルチモーダルインターフェースと効果的に組み合わせることが望まれる。

2.5 直接操作型インターフェース

我々はかつて、ラジオ番組制作支援システム「オラビー」の開発⁸⁾ *1 において、心的負荷の高い作業を行っている生放送中のラジオディレクターが、どのような操作系を用いればリスナーからの音声投稿を安全かつ効率的にオンエアできるか、という課題に取り組んだ。その結果、投稿された音声素材をオブジェクトとして画面に配置し、これらを直接ドラッグ&ドロップ操作の対象として、キューシートに見立てた矩形領域内に配置する、というインターフェースの有効性を確認している。

音声入力を操作手段とするシステムは一般的に「対話」あるいは「対話エージェント」をモデル化する「秘書型」のインターフェースである。しかしユーザの興味が操作対象そのものである場合には間接的操作を強いられることは望ましくない⁹⁾。これに対して、操作対象（オラビーの場合は投稿素材）を直接の操作対象にする「道具型」のインターフェースは、アプリケーション全体としてシステム透過性の向上に貢献しており、2.1 で考察した視覚障害者の要求にも近い。

3. プロトタイプシステムの設計

我々はお弁当購入タスク⁴⁾ を対象として、音声インクリメンタル検索と直接操作インターフェース⁸⁾ の要素を盛り込んだプロトタイプシステムの設計を行った。

画面構成を図 1 に示す。画面右下には商品名（73 個）の候補を示すリストボックスがある。画面左下にはタスクを初期化する「開始」ボタン、音声で入力された文字列を示すリストボックス、その内容を初期化

*1 2005 年上期 IPA 未踏ソフトウェア創造事業の成果。http://orange.com/

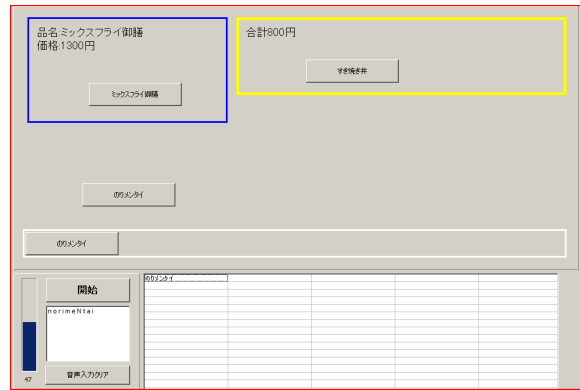


図 2 システムの全体画面（インスペクタおよびカートに商品に乗せた状態）

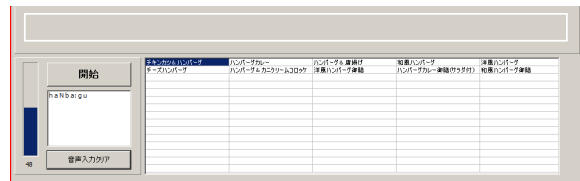


図 3 システムの画面（「ハンバーグ」と発話した直後）

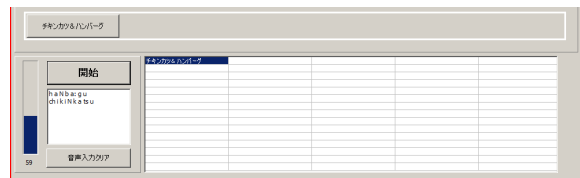


図 4 システムの画面（「ハンバーグ」「チキン」と発話した状態）

する「音声入力クリア」のボタンがある。その左のプログレスバーは音声発話が終了してからの経過時間を示しており、10 秒以内に次の操作を行わないと音声入力クリアが実行される。

画面上部には「インスペクタ」「カート」「候補パレット」が並んでいる。音声発話によって候補が一定数（5 個）以下に絞り込まれると、商品が候補パレットに並び、候補パレットの商品はドラッグ&ドロップ操作でパレットの外に取り出すことができる。取り出し操作はパレット外部へのオブジェクト複製の操作となっており、同じ商品を複数回取り出すことができる。前述した 10 秒が過ぎると候補パレットは初期化される。

取り出された商品はインスペクタ内部に配置することで、詳細情報を知ることができる（現在は価格のみ表示）。最終的に商品をカート内部に移動して決定を行うことでタスクは完了する。現在の実装では最終決定操作は実装していない。カート内部の商品の合計価格を動的に表示している。これらの作業の様子を図 2 に示す。

音声インクリメンタル検索の様子を図 3、図 4 に示

す。例えばユーザが「ハンバーグ」と発話しはじめると、発話終了を待つことなく候補リストの更新が始まる。発話が終了した時点では候補リストは「ハンバーグ」を含む商品 10 個に絞り込みが行われている。さらにユーザが (10 秒以内に)「チキン」と発話することで、「チキンカツ&ハンバーグ」のみが候補として残り、候補パレットにこの商品が表示される。

このシステムの構成要素はインタフェース基本原則への適合性において、以下の配慮がなされている。操作労力最小化の原則 部分発話を許容することで、音声入力における労力を最小化する。

システム透過性の原則 従来システムと比較して低遅延で多くの情報を提示することで、より適切なフィードバックを提供する。

頑健性の原則 頑健性の不十分な音声入力を候補選択のみに使用し、他のモダリティで決定操作を行う。これによりシステム全体として頑健性を確保している。

またインタフェース構成原則への適合性においては、以下の配慮がなされている。

初心者保護の原則 直感的な操作体系と画面構成により、マニュアル不要の使いやすさを提供する。

熟練者優遇の原則 発話において省略を許す、というインクリメンタル検索の利点を生かしている。

上級利用移行支援の原則 発話中に逐次画面更新を行うことで、語尾などの発話省略が可能である、ということを直感的に示している。

4. 実装と評価

4.1 実装

本システムの実装について述べる。音声認識エンジンには Julian rev.3.5.2-galatea (fast) ^{*1} Windows 版を使用した。音響モデルは Julius/Julian の音声認識キットとして配布されている性別非依存 PTM 音素モデルである。ネットワーク文法に対応する連続単語音声認識が可能であり、また第 1 パス実行中の内部状態を逐次出力する機能を有する。

アプリケーションの実装には CodeGear C++Builder 2007 を用いている。ドラッグ&ドロップによる GUI は VCL for Win32 を用いて実現されている。

本タスクに合わせて孤立単語のみを認識する文法を作成した。弁当屋の商品名は以下の例を含む 73 種類である。

- 玉子丼、のり弁、おにぎりセット、のりメンタイ、チキンカツ、チキンカツ&ハンバーグ、シャケ弁、親子丼、ビーフカレー、五目ごはん、...

またインクリメンタル検索を実現するために商品名の部分文字列に相当する以下のサブワード候補 (12 種類) を追加した。

- ハンバーグ、カレー、丼、カツ、和風、洋風、中華、フライ、御膳、デラックス、天、鶏
音声認識エンジン Julian の出力する情報は以下の通りである。

発話開始の検出 発話開始検出イベント。入力同期で第 1 パスを実行 (単語対文法を適用)。指定した間隔でフレームごとの 1 位候補を出力。

発話終了の検出 第 1 パス終了イベント。第 1 パス終了時点での 1 位候補を出力。

第 2 パスの実行 ネットワーク文法を適用。N-best 候補の計算。第 2 パス終了イベント。N 位候補の出力。

これらの情報のうち、本実装では発話開始から発話終了までに得られる情報のみを使用して画面更新を行う。具体的には第 1 パス実行中の結果を 50ms (5 フレーム) ごとに取得し、仮説候補の音素列表記が商品名の音素列表記に含まれているか否かという基準で 73 個の全商品から候補リストの絞り込みを行って逐次更新している。音声認識が処理を終了すると次の発話が可能な状態となり、複数の発話が続けて行われた場合は、その全ての音素列表記を含む商品名を絞り込み対象とする。

4.2 予備の評価

音声認識システムの内部状況を低遅延でユーザにフィードバックできれば、音声認識システムの性能が完璧でなくてもユーザは許容しやすくなるのではないかと、という仮説を検証するために、音声認識エンジン利用に関する予備の評価を行った。アプリケーションは Visual C++ 2005 および GLUT を用いて、候補リストの逐次表示結果と数字キーによる決定機能を実装した。検索結果をリアルタイムで画面に反映させるシステムと、第 2 パスまでの認識結果が得られるのを待って画面を更新するシステムを、被験者 5 名に評価させ、仮説に対して賛同を得た。特に入力したい商品名が長い場合に有効である、という意見があった。

4.3 国際ロボット展での運用

本システムを 2007 年 11 月から 12 月にかけて東京ビッグサイトで行われた国際ロボット展で運用した。ヘッドセットの接話マイクを使用し、PTT (Push-to-Talk) スイッチは使用せず、マイクアンプのレベル調整のみによって騒音環境に対処した。30 名くらいの利用者がおり、おおむね好意的な評価が得られた。多くの利用者は候補数を可視化する入力プロセスに対して肯定的であり、また、途中まで音声で入力し、最後はマウスで決定する、というマルチモーダルインタフェースは有効であると評価された。

評価者の意見を踏まえて考察すると、「とにかく何か喋ってみる」ということがユーザに次の行動を促すきっかけになったと思われる。何かを発声することでフィードバックが得られ、それによって自分の欲しいものが次第に明確になっていく、ユーザに試行錯

*1 <http://julius.sourceforge.jp>

表 1 語彙外発話と認識結果の対応例

| 認識結果 | 発話 |
|---------|-------|
| のりめんたい | めんたい |
| ちきんかつ | ちきん |
| ちんじゃおーす | ちんじゃお |

誤を促す効果を本システムが備えていたと考えられる。一方で、ドラッグ&ドロップで対象を直接操作する探索的検索インタフェースについては、肯定否定のいずれにおいても目立った意見は得られなかった。この点については、さらなる評価を踏まえて改良が必要と思われる。

4.4 語彙と認識結果の対応

本システムのタスクでは、音声認識の辞書にサブワード候補を手作業で追加した。また一部の商品名は「チキンカツ—チキンカツ&ハンバーグ」「唐揚げ—しょうが焼き&唐揚げ」のように他の商品名のサブワードになっている。これらは本提案の「単語の省略を許容する」という提案において有効に機能したと考えられる。また、実際に画面を見ながら行われた語彙外発話と認識結果の対応例(表1)からは、対立候補の尤度が上がりにくい候補や省略された語尾や語頭の音響尤度が高くない候補が結果として選択されやすい、という傾向が示唆された。

5. ま と め

本報告ではマルチモーダルインタフェースの設計における音声利用の可能性について、インタフェースの原則を踏まえた考察を行い、特にキーボードの使用できない環境において探索的検索を行うためのシステムの設計と試作について述べた。本研究を通じて「人間同士の対話のように素早く理解・応答させる」「エージェントが介在する対話ではなく、タスク対象そのものを視覚化し直接操作を可能にする」という要素の組み合わせの有効性が示唆された。

本研究は音声認識エンジンの実装や利用方法に関する検討は予備的な段階にとどまっており、今後はインタフェース設計における要求を満たすような要素技術の検討を進めていく必要がある。また「上級利用移行支援」に適合する機能として、タスクの情報を積極的に利用してユーザに音声コマンドや絞り込み語彙のヒントを提示することを検討したい。

本提案の有効性については以下のような観点から評価を行う必要がある。まず、入力に対するフィードバックの遅延の大きさが使いやすさにどのように反映されるか、どのくらい早い応答が必要であるか、といった評価を行いたい。また、低遅延のフィードバックのどのような要素が有効性につながっているのか、「不安を感じにくくなる」「ユーザが発話のコツをつかみやすくなる」などの仮説に基づいて検証していきたい。また、画面更新のタイミングを音声認識エンジンに委ねるの

ではなく、一定フレーム数だけ連続して同じ仮説候補が得られた場合に結果を更新する、ユーザが PTT スイッチを離したら結果を更新する、といった代替手法も比較対象として挙げられる。

最後に、フィードバックの方法についても比較検討を行いたい。我々は、タスクの対象そのもの(選択肢の数がリアルタイムに減っていくこと)の可視化が有効である、という立場を取っているが、入力に対して予測や補完の候補を提示する、あるいは候補数のみを数字やグラフなどで可視化する、などの代替手法も比較対象として挙げられる。

謝辞 本研究の一部は(独)情報通信研究機構・先進技術型研究開発助成金による研究開発として実施された。4.2 節における検討は東京大学工学部・宮城康暢氏の協力によって行われた。

付 録

A.1 インタフェースの原則

インタフェースの原則としてはさまざまな提案がなされている¹⁰⁾¹¹⁾¹²⁾¹³⁾。これらの提案を参考にしながら、音声の特質の検討の指標として利用しうるインタフェースの原則を提案した³⁾。

ここでは、望まれるインタフェース像を、インタフェースが持つべき機能に関する基本的要求と、それらの機能を組み上げてシステムを構成する際に考慮すべき要求との2つの側面から見直し、インタフェースの基本原則および構成原則として整理する²⁾。

A.1.1 インタフェースの基本原則

ここでは、人間は道具を使って作業を行なうにあたっては、楽に仕事を進めることを何より望んでいるものと考え、この観点から問題を整理する。

「楽である」とは、労少なく操作を実行できること、労少なく操作法について知ることができることの2つに分類できる。また、これらの2つにかかわる重要な問題として、システムの頑健性があげられる。

ここではこの3つの観点を満たすためにシステムが従うべき原則をインタフェースの基本原則とよぶ。

I 操作労力に関する原則

a. 位置移動最少の原則: コマンドを入力するために必要となる、手・マウス等の位置移動は少ないほど良い。

b. 指定操作回数最少の原則: 一つのコマンドを指定するための操作回数は少ないほど良い。

c. 指定操作容易性の原則: 一つのコマンド操作は単純で容易なほど良い。

II システムの透過性に関する原則¹⁰⁾

a. 状態理解容易性の原則: システムの状態とそこで可能な命令は容易に把握できることが好ましい。

b. 手順連想容易性の原則: ある命令を実行するための操作は容易に連想できることが好ましい。このた

め、操作の手順はアプリケーションの内外で一貫性を持ち、コマンド名なども連想しやすいことが好ましい。

c. フィードバックの原則：操作がもたらす結果は、常に予測可能であることが好ましい。操作には、常に適切なフィードバックを得られることが好ましい。

III 頑健性に関する原則¹⁰⁾

a. 誤入力防止の原則：システムに対する誤入力はなるべく防止できることが好ましい。

b. 修復容易性の原則：操作はできるかぎり可逆にし、誤操作が致命的な影響を及ぼさないことが好ましい。誤り易い操作に対する修復の操作は特に容易であることが望ましい。

A.1.2 インタフェースシステムの構成原則

上で述べた基本原則の間にはトレードオフが存在する。例えば、キーボードでコマンドを入力する場合、労力最少化の立場からは、コマンド名は短いほどよく、1 回程度の打鍵で一つのコマンドが実行できることが望ましい。しかし、その手順の連想を容易にする立場からは、求める機能に対応する名称をフルネームでコマンド名とする方が望ましい。このように、どれかの基本原則を立てれば、どれかは立たなくなるといった関係が随所にある。

どのような割合でどの基本原則を重視してシステムを構成すべきかについての指針をインタフェースの構成原則と呼ぶ。ここでは、構成原則として次の3つを挙げる。

a. 初心者保護の原則¹⁰⁾：入力システムには、当該アプリケーションならびに計算機自体に不慣れなユーザのための、透過性の高い入力手段を用意すべきである。

b. 熟練者優遇の原則¹³⁾：入力システムには、精通したユーザのための、アプリケーションに特化した効率的な入力手段を用意すべきである。

c. 上級利用移行支援の原則¹²⁾：入力システムは汎用手段を用いてシステムを利用しているユーザに対して、特化手段の存在を知らせ、この利用を促す自然な枠組みを用意すべきである。

これらは初心者に対する関を下げながらも、使い込んだときの利便性を重視し、利用者にいち早く熟練者モードへの移行を促すのが良いという主張である。

A.2 インタフェースシステムの導入原則

インタフェースシステムを成功させるためには、アプリケーションそのものの選択や設計により深く関与し、システムをどのような状況に適合させ、どのように評価や改良を行っていくか、というプロセスが重要になっている。付録 A.1 のインタフェース原則を補完するものとしてこれらを暫定的にまとめたものが、以下のインタフェースの導入原則である。

a. 有用性の原則：システムが使用される現場における必然性を考慮して設計と導入を行う。ユーザに動機付けを与える心地よさ・美しさ・楽しさを盛り込む(エモーショナルデザイン¹⁴⁾)。

b. 適合性の原則：あらゆる年齢や能力の人々に対して可能な限り使いやすさを提供する(ユニバーサルデザイン)。システムが使われる状況・環境を考慮する。ユーザ以外の人に悪影響を与えない。ユーザが行っている他のタスクに悪影響を与えない。

c. 妥当性の原則：妥当な時期に妥当な尺度で評価を行う。評価の結果を生かして反復的な開発・改良を行う。

参 考 文 献

- 1) Gary Marchionini: "Exploratory search: from finding to understanding", Communications of the ACM, Vol.49, Issue 4, pp.41-46, 2006.
- 2) 小林 哲則, 竹内 陽児, 白井 克彦: "音声・マウス・キーボードによる統合的入力環境," 電子情報通信学会技術報告, HC92-68, Mar 1993.
- 3) 西本 卓也, 志田 修利, 小林 哲則, 白井 克彦: "マルチモーダル入力環境下における音声の協調的利用-音声作図システム S-tgif の設計と評価-", 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp.2176-2183, Dec 1996.
- 4) 西本 卓也, 嵯峨山 茂樹, 藤原 扶美, 下永 知子, 渡辺 隆行: "対面朗読者と視覚障害者の対話の分析とその応用," 情報処理学会研究報告, 2007-SLP-11, pp.55-60, Feb 2007.
- 5) Jef Raskin, 村上 雅章 (訳): ヒューメイン・インタフェース, ピアソン・エデュケーション, 2001.
- 6) 嵯峨山 茂樹, 西本 卓也, 中沢 正幸: "擬人化音声対話エージェント," 情報処理学会誌, Vol.45, No.10, pp.1044-1049, Oct. 2004.
- 7) 安藤 彰男: リアルタイム音声認識, 電子情報通信学会, 2003.
- 8) 西本 卓也, 川崎 隆章: "ラジオ放送支援システム「オラビー」の開発," 電子情報通信学会技術報告, WIT2006-25, pp.49-54, Jul 2006.
- 9) Brenda K. Laurel: "Interface as Mimesis," User Centered System Design, edited by Donald A. Norman, Stephen W. Draper, pp.67-86, Lawrence Erlbaum Associates, New Jersey, 1986.
- 10) Donald A. Norman, 野島 久雄 (訳): 誰のためのデザイン? - 認知科学者のデザイン原論, 新曜社, 1990.
- 11) B. Shneiderman: Designing the User Interface, Second Edition, Addison-Wesley Publishing Company, 1992.
- 12) 海保 博之, 原田 悦子, 黒須 正明: 認知的インタフェース, 新曜社, 1991.
- 13) 海保 博之, 加藤 隆: 人に優しいコンピュータ画面設計, 日経 BP 社, 1993.
- 14) Donald A. Norman, 岡本 明, 伊賀 聡一郎, 安村 通晃, 上野 晶子 (訳): エモーショナル・デザイン 微笑を誘うモノたちのために, 新曜社, 2004.