

インタフェースシステムの導入原則に関する一考察

東京大学大学院情報理工学系研究科 西本 卓也

[概 要]

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、システムを与えられた状況に適合させ、評価や改良を行っていくプロセスが重要である。本稿では音声インタフェースを念頭に、筆者が過去に論じたインタフェース原則を補完する「インタフェースの導入原則」を提案し、「有用性」「適合性」「妥当性」の3つの観点から概要を述べる。

1 はじめに

電話応答システム、カーナビゲーション、人間型ロボットにはじまり、最近ではスマートフォンの操作においても音声認識の利用が広まりつつある。

音声というメディアがコミュニケーションに使われ続ける限り、機械と人間の音声インタフェースは必要不可欠である。そう考えてきたが、この十数年のインターネット技術の普及は「人と人のリアルなコミュニケーションの場」を大幅に減らしてしまった。電子メールやウェブといった「機械の都合に人間が歩み寄った技術」は「未成熟なインタフェース技術」であり、いずれ技術が成熟すれば「人間に歩み寄った快適なインタフェース」に置き換わっていく。そのように楽観的に考えることができた。しかし「電子メールやウェブが普及したこの十数年」は残念ながら「雇用崩壊の時代」でもあった。今後は「要素技術としては未成熟なままの音声対話システム」が「経済的な事情」に後押しされて、生身の人間の職場を奪っていく、そんな事例が増えるのかも知れない。

視覚障害を持つ人々は、音声合成を「ディスプレイの代用」として使いこなしている。音質が悪くても、読みやアクセントの誤りが多くても、彼らは「他に手段がない」から仕方なく（不完全な技術に歩み寄って）使っている。あらゆる人々が「仕方なく音声対話システムを使う」時代が永遠に来ないと言えるだろうか？

本稿では、音声インタフェースを念頭に「インタフェースシステムの導入原則」について論じる。これはインタフェースの原則 [1, 2] を補完するものである。以下の「有用性」「適合性」「妥当性」の3項目から構成される。

2 有用性の原則

主張：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。

2.1 必然性の考慮

音声認識技術の利用が必然となる「現場」を探すことが重要である。初期の実験システムでは「電話応答システムにおいて数字を入力する代わりに、音声で読み上げた数字を認識できる」といった提案が行われた。これに対して「数字は電話機の数字ボタンで入力すればよいので、音声入力の方が優位性につながらない」という意見があった。一方で「数字は任意のリソースを示すユニバーサルな手段」として重宝される場面もある。

文献 [1] では、作図システムにおいてマウスを描画エリアのポインティングに専念させるために、その他の操作を音声入力で行う、という提案を行った。「慣れた被験者は左手でキーボードショートカットを操作し、右手でマウス操作を行うので、熟練すると音声の優位性は小さくなる」という見方もあったが、「しばらく利用をやめるとショートカットの利用は思い出すことが難しく、そういった場合に音声は再度有効性を発揮する」ということもあわせて示された。

「数字は電話機の数字ボタンで入力すればよい」という考え方は、固定電話については成り立っていた。しかし携帯電話においては「ボタンを押すためにはいったん顔から電話機を離す必要がある」ので、数字の音声入力にも必然性がある。このように「必然性の定義」は、時代の変化や技術の動向に応じて変化する。

音声インタフェースを用いることに必然性がある応用として、ロボット、そしてモバイル機器を挙げることができる。ロボットにキーボードやスイッチをつけることは不適切である。モバイル機器におけるボタン入力は万能ではない。タッチ操作もなにかを選択するためには有効だが、いわゆる「探索的検索」には向いていない。

音声技術をマルチモーダル・インタフェースの要素技術として捉えなくてはならない。「音声と非音声の手段を適切に組み合わせる」ことが有効な場合もある。さまざまな立場の設計者との協力も必要である。

2.2 動機付け

音声認識に熟練しているユーザは少ない。音声を使ってみたいと思わせる「動機付け」が重要である。ノーマンは「エモーショナル・デザイン」[3] の議論において、美しさ・楽しさなどを与えてくれるインタフェースシステムは、ユーザのタスク達成率が有意に高くなる、というエピソードを紹介している。この理由についてノーマンは、ユーザが楽しいと感じるほど、操作がわかりにくい場合や予期しない反応があったときに、積極的に試行錯誤を行って問題を解決するようになるため、という考えを示している。

チクセントミハイの「フロー理論」[4] は社会心理学の立場から「仕事」と「遊び」は作業の目的によって区別されるべきではないこと、あらゆる作業には自己目的的な要素がありうるこ

とを示した。また「挑戦」と「技能の知覚」のバランスに基づく「フロー体験」のモデルを提案した。フロー理論はゲームデザイン、教育、経営など、多くの分野に影響を与えた。

ヒューマンインターフェースにおいてもフロー理論はしばしば考慮される。例えば Web デザインについては、リンクの少なすぎるページは、ユーザに与える選択肢が少なすぎる（挑戦が低すぎる）ので「退屈」に感じる。リンクの多すぎるページは、ユーザに与える選択肢が多すぎる（挑戦が高すぎる）ので「不安」になる、といったフロー体験のモデルが考えられる。

筆者の検討 [5, 6] から、音声入力インターフェースは、それ自身が自己目的的な楽しさを持つこと、「音声認識の性能が 100 % ではないこと」に起因する「楽しさ」があり得ること、などが示唆された。しかし「楽しさ」の客観的な測定は簡単ではない。また「タスクに起因する楽しさ」と「インターフェースに起因する楽しさ」を区別することは難しい。

「楽しさ（フロー体験）」をもたらすもの、特に「一人ひとりの人間が主役になるメディア環境時代」の新たなアプリケーションも、音声技術を使いこなすための視点である。現在流行中の「ニコニコ動画」「ツイッター」などは、筆者が過去に提案してきた音声応用システム [7, 8, 9] を連想させる。

2.3 有用性の発見を促進する

ある技術が有用であることを発見するためには、特殊な現場や立場からの要求、日常の些細な不便、などを幅広く検討する必要がある。ちょっとした思いつきを、音声技術の専門でない技術者が、気軽に実装して試せる、ということが重要である。そのために必要なのはツールキット、API、Web サービス等の整備と戦略的な無償化である。ノウハウを簡便化して共有するための標準化、そして国際化も必要である。

もはや「音声認識は完成した技術」と見なされることもある。しかし現在の技術では、例えば統計モデルのチューニングのコストがかかりすぎるため、実環境利用においてビジネスモデルが成立しない、という指摘もある。こうした技術の限界が、現在の技術の延長で打破できるのか、シャノンの情報理論に由来する枠組みの限界なのか、定かではない。

「オープンソース」あるいは「フリー戦略」は、インターフェースの指針を個別の問題に適用するための試行錯誤を加速しつつある。

3 適合性の原則

主張：あらゆる年齢や能力の人々に対して使いやすさを提供する。

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。

3.1 ユニバーサルデザイン

筆者の最初の論文 [1] は「インタフェースシステムの構築において、要素還元主義によって構成技術の改良を積み重ねる」従来手法を見直すための「音声インタフェース構築の方法論」であった。これには「技術ではなく人間に着目した要素還元主義」に繋がる。アクセシビリティの検討、すなわち視覚や聴覚などの障害を支援する技術は、人間のモダリティを要素還元主義的に検討できる視点である。

音声認識は聴覚障害の支援に有効なメディアである。具体例は、放送の字幕作成支援技術などである。また音声合成は（健聴者については）視覚障害の支援に有効なメディアである。具体例はスクリーンリーダ（画面読み上げソフトウェア）である。

障害支援という状況では、しばしば特別な技術要求がある。例えば視覚障害の支援では「早口音声合成」が求められる。早口音声合成の聴取には慣れの効果があり、また単語親密度や記憶の効果が関わってくるので、評価には慎重を要する [10] 。

3.2 状況と環境の考慮

カーナビゲーションシステムの目的地設定においては音声入力がある。これは自動車の運転において「ハンズフリー」「アイズフリー」が要求されるからである。これに加えて「マインドフリー」の考慮も重要である。例えば右左折の直前など「運転そのものが大きな負荷」という状況では、音声対話システムがユーザに話しかけたり、操作を求めたりすべきではない。

筆者が関わった一連の実証実験 [11] においては、車載音声対話インタフェースにおいて、自動車の走行状況を考慮して、対話の中断や再開を行う機能を実現し、高い評価を得た。このように、ユーザが行っているタスクに影響を与えない、というのは、重要な配慮である。

また別の着眼点としては、「音声インタフェースを使うことが周囲の人の邪魔にならない」という配慮が挙げられる。このような目的で「ささやき声」など、周囲の人に聴き取りにくい音声を入力・認識できる技術の研究がなされている。

4 妥当性の原則

主張：適切なタイミングで、妥当な手法と尺度で評価を行う。

結果を生かして反復的な開発・改良を行う。

妥当性を保証するためには評価手法が重要である。音声対話システム・音声インタフェースがどれだけユーザにとって「楽であるか」を評価できる手法について以下で論じる。

4.1 所要時間に関する評価

同じタスクを与えて複数のインタフェースシステムを被験者に利用させ、所要時間を比較する。複数の被験者で行う場合は、被験者間の平均を比較する。より厳密には、検定や分散分析を行う。複数のタスクを使う場合は、タスクごとに比較をすることで、重要な知見が得られる

こともある．要因（タスクの違い，被験者群の違い，インタフェースの違いなど）を線形モデルで近似し，それぞれの要因の効果を推定する方法もある．

4.2 身体的・物理的な負荷に関する評価

音声入力とポインティングデバイスの比較のために「手や指を動かす距離の累積」「コマンドを入力した回数」を使用できる．文献 [1] では，マウスを使う作図タスクにおいて，モード切替のコマンドを音声入力で行うことにより，すべてをマウス操作で行った場合と比較して，マウスポインタの移動量を削減できた，と主張した．

この議論にはいくつか検討の余地がある．まず「画面上のマウスポインタの移動距離」と「マウスを持つ手の移動距離」が同じであるという仮定を検証しなくてはならない．またマウスの移動量だけを負荷と見なすのは妥当ではなく，「キーボードでショートカット操作を行う場合の指の動き」も検討する必要がある．さらにタブレット操作やタッチ操作における身体的な負荷の評価は，単純ではない．また「音声コマンドを発話することの身体的・物理的な負荷」も無視できない．声の大きさや発声方法も考慮すべきである．

4.3 心的要因に関する評価

二重課題法は心的要因を客観的な尺度で測ることができる一手法である．文献 [12] では，音声認識とテキスト音声合成による対話インタフェースを被験者に使用させながら，画面とキーボードを用いた早押しゲームを行わせ，ゲームの応答速度の変化から，対話の状態によってインタフェースの負荷が変化していることが判断できる，と主張した．

二重課題法を用いる際の注意点としては，モダリティやタスクの干渉を防ぐ必要がある．例えば，主課題が音声対話タスクであれば，副課題は音声を用いないものが妥当である．また，副課題の早押しゲームが「上下左右」を判断させるタスクであれば，主課題として「右折・左折」などの判断をさせるべきではない．副課題は難易度を適切に設定する必要がある．簡単すぎても難しすぎても主課題の影響を受けにくくなる．副課題が短時間で完結するタスクであれば，主課題を評価する時間分解能が高くなる．副課題のパフォーマンスを測る尺度は連続値である方が望ましい．単なる正解・不正解の判定よりも応答時間の変化を比較する方が，主課題の効果に敏感である．

もう一つ，生理的指標を用いた心的負荷測定がある．特にストレスを評価する目的で，呼吸や脈拍などを利用する研究が盛んに行われている．ヒューマンインタフェースの評価としては簡便でなく，結果にノイズが多いため扱いにくいという意見もある．

4.4 主観的評価と NASA-TLX

客観評価では比較できない要因も，主観評価では観測できる可能性がある．

一般的な心理実験の主観評価では，5段階・7段階などの尺度で評定を行う．評価項目の検討は慎重に行わなくてはならない．また「非常に…である」「まったく…でない」など，数量化すべき形容詞・副詞の選び方にも注意すべきである．

NASA-TLX はメンタルワークロードの評価手法である [13]。被験者は 6 つの尺度の重要度を評定し、その後、それぞれの尺度について評定を行う。各尺度の負担の値は重要度で重み付けされる。経験的には妥当な結果が得られることが多い。ただし、重要度による重み付けのプロセスには「本当に被験者が重要度を正しく評定できているから、重み付け評定値が妥当な値を示す」のか、「尺度に関する被験者の理解のばらつきを、結果的にこの評定によって相殺している」のか、疑問が残る。また重み付け評価値は「被験者内でのみ比較可能な値であり、被験者間で比較することは妥当ではない」という立場と「被験者間の比較に使える絶対値である」という立場がある。

筆者は前者の立場で「音声の聞き取りやすさ」の評価に取り組んでおり、被験者ごとに複数の課題を行わせて、平均と分散の正規化を行い、その正規化された値を被験者間で比較する、という手法を提案している [14]。

4.5 反復的な開発プロセス

クーパーはインタフェース・システムの開発プロセスに起因する問題を取り上げ、「まずデザインを作り、それから機能を実装せよ」と主張した [15]。このような目的ではラピッド・プロトタイピング・ツールが有効とされる。しかし本質的に重要なのは、プレゼンテーションとロジックを分離する設計手法である。

筆者が視覚障害者のためのタイピング練習ソフト「ウチコミくん」の開発 [16] において取った手法は「まずコンテンツ記述言語とその処理系を実装し、コンテンツの制作と記述言語処理系の開発を並行して行う」というものであった。このような発想は UNIX システムに関するベントリーの考察 [17] 等に遡るが、近年 DSL (Domain Specific Language: ドメイン特化言語) あるいは LOP (Language Oriented Programming: 言語指向プログラミング) として注目されている。

5 おわりに

「原則」とよべるものは網羅的でなくてはならず、また実践的でなくてはならない。本稿では著者の経験や著者の知る事例を踏まえて、考え方の道筋を具体的に示したつもりである。だが、音声技術を生かすツールとしての本稿の完全性を示すのは困難であろう。本論は筆者のサイト [18] にて発展させていく予定である。

これからも「率先して指針を示す」ことを目指しつつ、オープンソースプロジェクトあるいはビジネスの当事者として積極的に関わりながら、工学・考古学・考現学・人間学としての「音声インタフェース論」を考えていきたい。

文 献

- [1] 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.
- [2] 西本 卓也, 岩田 英三郎, 櫻井 実, 廣瀬 治人: "探索的検索のための音声入力インタフェースの検討," 情報処理学会研究報告 2008-HCI-127(2), pp.9-14, Jan 2008.
- [3] ノーマン: 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.
- [4] チクセントミハイ: フロー体験 喜びの現象学, 世界思想社, 1996.
- [5] 西本 卓也, 新美 康永: "音声認識の自己目的的な楽しさ," 人工知能学会研究会資料, SIG-SLUD-9804, pp.13-18, Feb 1999.
- [6] 西本 卓也, 櫻井 晴章, 荒木 雅弘, 新美 康永: "ディクテーション作業における楽しさの分析," 人工知能学会研究会資料, SIG-SLUD-A103-01(3/8), pp.1-6, Mar 2002.
- [7] 西本 卓也, 幸 英浩, 川原 毅彦, 荒木 雅弘, 新美 康永: "非同期型音声会議システム AVM の設計と評価," 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.11, pp.2490-2497, Nov 2000.
- [8] 西本 卓也, 宮川 祥子, 川崎 隆章: "インターネットラジオによる情報発信支援ツールの設計," 電子情報通信学会技術研究報告, WIT 2001-7, pp.35-40, May 2001.
- [9] 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.
- [10] 西本 卓也, 渡辺 隆行: "早口音声の聴取における学習効果と加齢の影響," Human Interface 2007 論文集, 3134, pp.937-942, Sep 2007.
- [11] 西本 卓也, 荒木 雅弘, 小林 哲則: "ディストラクション評価に基づく車内音声対話コンテンツの比較," FIT2003 イベント企画, 「車載情報システムにおける音声インタフェース」予稿集, Sep 2003.
- [12] 西本 卓也, 高山 元希, 櫻井 晴章, 荒木 雅弘: "音声インタフェースのための対話負荷測定法," 電子情報通信学会論文誌 D-II, Vol.J87-D-II, No.2, pp.513-520, Feb 2004.
- [13] 芳賀 繁: メンタルワークロードの理論と測定, 日本出版サービス, 2001.
- [14] 西本 卓也, 狩谷 幸香, 渡辺 隆行: "早口音声聴取における単語親密度とメンタルワークロードの検討," 日本音響学会 2007 年秋季研究発表会講演論文集, 2-Q-23, pp.593-594 (in CD-ROM), Sep 2007.
- [15] アラン・クーパー: コンピュータは、むずかしすぎて使えない!, 翔泳社, 2000.
- [16] 西本 卓也, 高木 治夫: "視覚障害者向けタイピング練習ソフト「ウチコミくん」," 情報処理学会インタラクティブ 2002, pp.191-192, Mar 2002.
- [17] ベントリー: プログラマのうちあけ話 続・プログラム設計の着想, 近代科学社, 1991.
- [18] <http://ja.nishimotz.com/>

著者紹介

1993 年 早稲田大学理工学部電気工学科 (白井・小林研究室) 卒業 . 1995 年 同大学院理工学研究科修士課程修了 . 1996 年 京都工芸繊維大学工芸学部助手 . 2002 年 東京大学大学院情報理工学系研究科助手 . 2007 年 同 助教 .