

人間の多様な振る舞いを考慮した音声 UI の必要性

岡本淳[†] 庄境誠[†]

音声認識が広く普及しない理由は認識率だけが原因ではない。UI のデザインにも多くの問題点が潜んでいる。誰でも使える音声 UI をデザインするにはユーザの行動を観察して、問題点を洗い出す必要がある。しかも、使い続けたいと感じる音声 UI を提供するにはユーザの行動をリアルタイムでモニタして、音声 UI がユーザに適応する必要がある。

Needs of Voice UI on which a variety of human behavior is taken into account

Jun Okamoto and Makoto Shozakai

The reason why ASR is not widely used is not only for accuracy but also for UI design aspect. In order to design an easy-to-use Voice UI (VUI), it is necessary to observe user's behavior and clarify the problems. Moreover, to provide the VUI everyone wants to keep using, it should monitor the user's behavior in real time and the VUI should adapt to the user.

1. はじめに

我々は会話で意志の疎通を行っている。それは人間にとって最も自然で、簡単で、早い方法であるからだ。機械との意思疎通も音声で行いたいと思うのは当然なことだろう。そんな想いから音声認識は発達してきた。しかし、音声認識は広く普及しているとは言いがたい。条件を整えば、80~90%以上の認識率が出るにも関わらずである。これは、機械への音声入力を実は人間にとって難しいからなのだろうか？ここでは、2006年度から2008年度にかけて行われた、経済産業省研究開発プロジェクト「情報家電センサー・ヒューマンインターフェイスデバイス活用技術・音声認識基盤技術」（以下、国プロ）の成果を踏まえて、今後必要とされる音声 UI について考察する[1]。

2. 何が問題なのか？

我々は、国プロに取り組む準備として市販されている音声認識機能付きのカーナビゲーションシステム（以下、カーナビ）を2台購入し、社内のスタッフ10名を被験者として簡単なユーザビリティテスト（以下、UT）を行った。納品されたばかりでこれから試しに使うという状況を想定した。このため、事前に説明したことは音声認識の開始方法、タッチキー、リモコンによる操作方法だけである。尚、被験者にはカーナビのマニュアルを渡し、使い方がわからない場合は参考できるようにした。被験者には、観光地のパンフレットを渡し、好きなところを選んで目的地の設定をしてもらうというタスクを与えた。但し、この時なるべく音声認識を使用し、音声認識で達成できないと判断したら、タッチキーやリモコンにより実行するようにしてもらった。

UTを開始してみると、皆いきなり操作を開始することが観察された。しかし、途中で操作方法がわからなくなり、10人中9人の被験者が途中でマニュアルを調べていた。なぜ最初にマニュアルを読まなかったかと質問してみると、「自分はマニュアルを読まなくても大丈夫だと思った」、「分厚いマニュアルを見ただけで読む気がしない」などと回答していた。

また、UT中の被験者の行動を観察していると、被験者は画面に表示されている言葉、シンボルや音声ガイダンスに誘導されやすいことがわかった。また、音声コマンドとリモコン、タッチキーによる操作を自由な順序で組み合わせて処理できると思いついでいるために思うように操作できないということもわかった。

これらの結果は、認識率より音声 UI そのものに解決しなければならない問題が潜んでいることを示唆していると考えられる。

国プロの UT の際、被験者に音声認識機能付きのカーナビを使ったことがあるかど

[†] 旭化成株式会社 新事業本部 情報技術研究所
Information Technology Laboratory, New Business Development, Asahi Kasei Corporation

うかをインタビューした。使ったことがあると回答した被験者に「どうでしたか」と質問すると、ほとんどの被験者は「認識してくれない」と回答していた。しかし、被験者は適切なタイミングで、適切な音声コマンドを発話したのだろうかという疑問が残った。我々が社内のスタッフを対象にして行った UT では、ほとんどの被験者は最初にマニュアルを読まなかった。国プロの被験者もマニュアルを読まずに適当に操作を開始した人がほとんどだと考えてよいのではないか。もしそうであれば、発話のタイミングや発話コマンドは、カーナビの開発者が想定していたものと異なっていると考えられる。そのために、被験者は「認識してくれない」と考えた場合も多かったと思われる。認識率の向上も大事であるが、「使える」ということをユーザにアピールできる音声 UI をデザインすることが重要である。

3. どうやって開発者の考えをユーザに伝えるか？

よく音声 UI と GUI は比較される。GUI はわかりやすいが、音声 UI はわかりにくいと結論付けられることが多い。しかし、本当にそうであろうか？ Web を閲覧していると、時々どうしたら良いかわからなくなる場合がある。GUI でもデザインが悪ければ使い勝手の悪いものになってしまう。しかし、GUI の場合は目で見ることができる。したがって、ユーザから見るといくつ選択肢があるのかということがわかりやすい。音声 UI の場合はどうであろうか。システムに「コマンドをどうぞ」と言われてもどんなコマンドがあるのかはわからない。コマンドを読み上げられて、その中から選択するような音声 UI の場合、コマンドの数が多いと全てを覚えることができない。このように考えると、音声 UI の場合は如何にしてユーザに手掛かりを呈示するかという事が大事だと言うことがわかってくる。そして、その手がかりが紛らわしい場合、ユーザは誤った行動をとることになる。

我々は、国プロで音声 UI が具備されたカーナビゲーションシステムの実証システムを開発し、20代から60代の被験者を対象に UT を行った。マニュアルを読まなくても操作方法がわかるようにするために、画面上のボタンに書いてある言葉を音声コマンドに対応させるようにした。何もヒントがなければ画面上にある言葉を話すのではないかと想像し、音声認識が開始された時には「コマンドをどうぞ」という単純なガイダンスを流すだけにした。

最初のプロトタイプを開発し、6人の被験者で UT を行った。最初のタスクは電話をかけるという内容である。このタスクでは、2人の被験者は自力でタスクを達成することができず、残りの4人は自力でタスクを達成できたが時間がかかってしまった。このような結果になった原因は、以下の2つである。

- 「これぐらいのことはわかってくれるだろう」という安易な考えに基づいて音声

UI をデザインした

- エラーが起きた際に回復させる方法をユーザに呈示していなかった

2番目の原因のためにユーザはエラーが起きてもすぐに回復することができなかった。エラーが起きても自力で回復できるようにしておくことは非常に大事である。開発者はシステムの仕組みを知っているため、エラーが起きてもそれがなぜ起きたかを推測することができる。しかし、ユーザはシステムの仕組みを知らない。このためエラーが起きてもその原因を正しく推測することは困難である。正常な状態で操作する方法だけでなく、異常が起きた状態での操作方法もユーザに伝わるように、UI をデザインすることが大事である。

しかし、そうは言っても開発者の考えをユーザに伝えることはなかなか難しい。難しい原因として次の2つが考えられる。

- 音声ガイダンスをあまり長くできない
- ある動作を言い表すのに誰もが同じ表現を使うとは限らない

誤解を与えないように音声ガイダンスを懇切丁寧にするという方法が効果的なように思われる。しかし、そのようにすると音声ガイダンスを聞いているうちに最初に出てきた内容を忘れてしまうようになる。そのようなことを避けるためには、音声ガイダンスはあまり長くしないほうが良い。

電話をかけることは他に何と言いつけることができるだろうか？ ちょっと考えただけでも「電話をつなぐ」、「ダイアルする」、「発信する」という表現が思い浮かぶ。普段自分が使っている表現と違い過ぎると、とっさにその意味を理解できなくなってしまう。特に、エラーが起きた場合は、ユーザは多少なりとも動揺しているので余計に意味を理解することは難しい。ではどうするかというと、音声ガイダンスをなるべく具体的な表現をすることが大事である。具体例を示すとさらに効果が上がる。

しかし、これだけでは十分ではない。国プロでの経験では、約10%の被験者は音声ガイダンスを注意深く聞いてくれなかった[2]。エラーが起きた後で具体的な操作方法を音声ガイダンスで説明しても、同じ操作を繰り返した被験者が多かった。インタビューをしてみても、「意味がわからなかった」、「何が起きているのかわからなかった」という回答が返ってくるだけであった。

最終的に、我々は、エラーが起きた時に被験者の注意は他の事柄に向けられているのではないかと考えた。そのような状態の時にいくら音声ガイダンスを流しても聞いてもらえないのではないかと考えたわけである。人と人とのコミュニケーションでは、相手が自分の話を聞いていないと感じた時には、「ちょっといいですか」とか「聞いて下さい」という言葉を相手に投げかける。機械と人がコミュニケーションを行う場

合も同じようなことを行うのが良いと考え、何度も同じエラーが起きた場合はシステム側に注意を向けるように促すようにした。具体的には「これから使い方の説明を行いますので準備ができたらいと返事をしてください」という内容の音声ガイダンスを流し、さらに画面上にも同じ内容のメッセージを表示するようにした。

4. どうして迷うのか？

国プロでは、合計 245 名の被験者を対象に最終実証評価の UT を行った。この UT で観察された問題点を 3 つのレベル分類した結果が、

である。各レベルの定義は表 1 に示す。全体の問題の内、9%が重大な問題であった。この重大な問題を原因別で分類したのが図 2 である。最も多いのが音声ガイダンスに関連した問題である。この問題に対しては前章で述べた対策が効果的であった。2 番目の問題である Structure Understanding は、システムの構造を被験者が理解できなかったために起きた問題である。具体的には、タスクを達成するにはどのメニューを選択するかわからない、あるいは、画面上にたくさんボタンが並んでいると次に何を選んだらよいかわからなくなるという問題である。重大な問題を年齢層別に分類すると図 3 のようになる。図 3 からわかるように、年齢層が上がるにしたがって、これらの問題の発生件数が増えることがわかる。

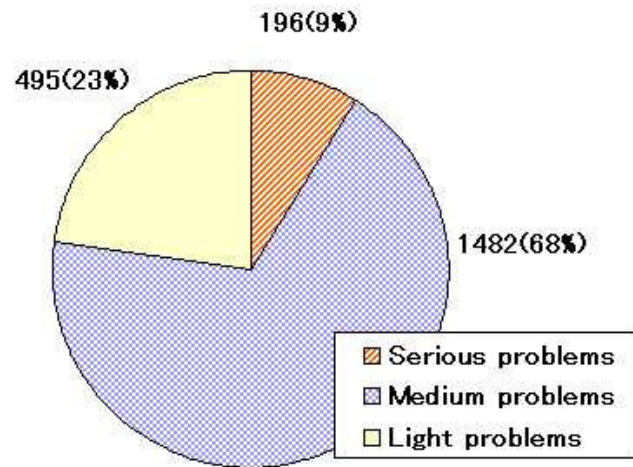


図 1 重要度別問題点の分類

表 1 重要度の定義

Level	Definition
Light	Observed problem has little effect on task completion.
Medium	Observed problem in a task that was completed without any assistance but required a large consumption of time.
Serious	Observed problem in a task which was incomplete or where the subject misunderstand the task.

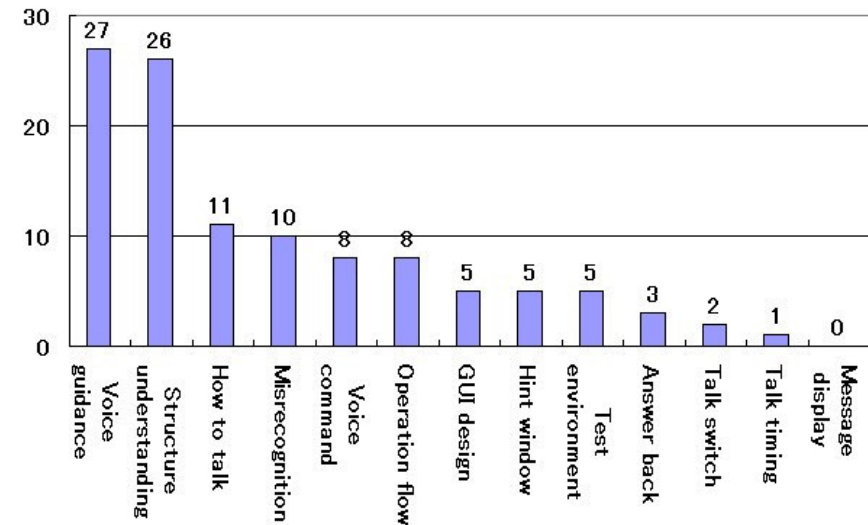


図 2 原因別重大な問題

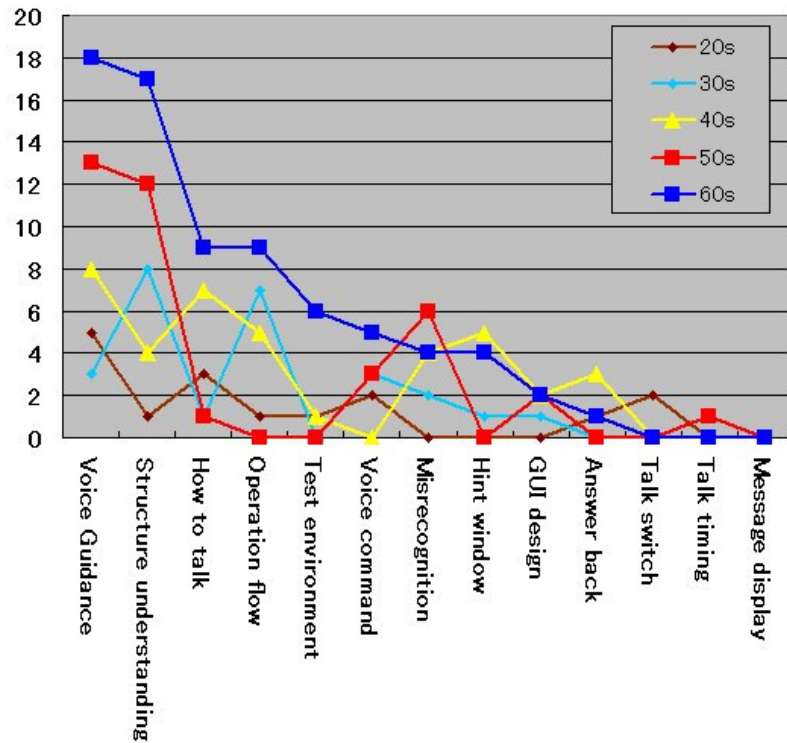


図 3 年齢層別重大な問題

被験者へのインタビューによると、最初のタスクでは、今自分が操作しているシステムはどんな仕組みになっているのだろうかということを主に考えながら操作していることが判明した。最初のタスクを終えると、「だんだん使い方がわかってきました」と回答する被験者が多かった。これらのことから、Structure Understanding の問題を避けるにはシステムの構造をなるべく簡単にすることが大事である。簡単にすることで、システムの仕組みを学習する負荷が低くなる。高い年齢層ほど、この問題の発生件数が多い理由は、システムの仕組みを学習するスピードが年齢層によって異なり、年齢層が上がるほどスピードが遅いからだと考えられる。但し、年齢層が高いほど認知能力や学習能力が低下するためか、あるいは、若い頃にパソコンや携帯電話などのメニュー構造を有する IT 機器に馴染む機会がなかったからかまでは特定できなかった。

表 2 タスク達成率の最終評価結果

Task completion rate	Number of subjects
100%	36
95%	2
85%	1

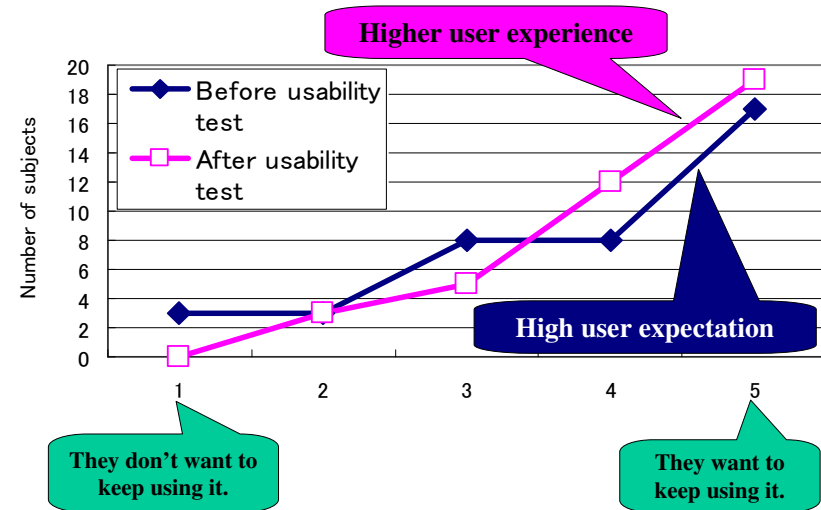


図 4 音声 UI を使い続けたいかどうかの主観評価結果

5. 「使える」ということと「使い続けたい」ということ

音声 UI を広く受け入れてもらうには、まず誰でも使える音声 UI になっている必要がある。前節で示した対策を行うことで、我々は国プロでの最終目標「95%の被験者に対して95%のタスク達成率」をクリアする事ができた。

また、最終 UT の前後で、被験者に音声 UI を使い続けたいかどうかに関する 5 段階で主観評価をお願いした。その結果、図 4 に示すように、最終 UI の前は、64%の被験者が音声 UI を「使い続けたい」（主観値 4 または 5）と意思を示し、音声 UI に対する高い期待感を示していたのに対して、最終 UI の後では、79%の被験者が音声 UI を「使い続けたい」（主観値 4 または 5）との高い経験値を示した。

「使える」ということは必ずしも「使い続けたい」ということと一致しない。ユーザが初めてシステムを使う時には、「うまく使えるか」ということに主に関心が集まっている。しかし、一旦使えるようになると、今度は「使って楽しいか」ということに関心が移る。我々が開発した実証システムを被験者は使い続けたいと感じてくれるが、また、何度も使う内に行動に変化が現われないかを観察するために、被験者に同じタスクを3回繰り返してもらった。

どのように被験者の考えが変わっていくかを観察するために、各タスクが終わる毎にインタビューを行った。その結果、1回目のタスク終了時にはわかってきたことや難しいと感じたことを答えていたが、2回目または3回目のタスク終了時には音声ガイドがわずらわしいと感じると回答する被験者が出てきた。それもそのような回答をする被験者は、若年齢層には見られず、高い年齢層の被験者に多かった。

我々が開発した実証システムは、誰でも使えるシステムということを目指した。そのため、ユーザは初心者であると想定していた。ただ、初心者であっても慣れてきたら便利に使えるようにいくつか機能を組み込んでおいた。例えば、音声ガイド再生中でもボタンを押すことにより、音声ガイドをスキップできる手動バージョン機能を用意し、画面表示や音声ガイドでこの機能を被験者に告知した。若年齢層の被験者は注意深く音声 UI を観察しているの、そのような機能に気がついて使うことができた。このため、若年齢層の被験者からは音声ガイドに関するクレームはなかった。しかし、年齢層が上がると基本的な使い方をマスターした段階でそれ以上音声 UI を注意深く観察しようとしなくなる。このため、便利な機能には気がつかない。また、初心者のレベルは超えているにもかかわらず、システムは被験者を初心者扱いするというミスマッチが起きてしまった。このような理由で高い年齢層の被験者から音声ガイドに関するクレームが出たものと考えることができる。

エラーからの回復を対策する際は、被験者へのインタビューから問題点を発見し解決することができた。しかし、使い続けるうちに変化してしまう被験者の考え、振る舞いに対しては同じ方法で対策することは難しい。ところで、機械と人とのやりとりではなく、人と人の場合はどうするだろうか？その場合は相手の様子を観察しながら、慣れてきたなと思ったら「こんな方法がありますよ」とか、「こうしたらもっと便利です」とアドバイスするであろう。機械と人の場合も同様にリアルタイムで振る舞いを観察して問題点を発見し、その対策を立てるようにしないと解決は難しいと考えられる。

6. おわりに

国プロの成果を踏まえて、どのような点に注意したら「使える音声 UI」をデザインできるかについて紹介した。今回紹介した例はカーナビがターゲットであるため、別

のアプリケーションの場合はまた違った観点からデザインする必要があるかもしれない。しかし、「使える音声 UI」を実現するには、認識率を上げるだけでなく、ユーザの行動を観察して、そこに潜在する問題点を洗い出す必要がある点は同じであると思われる。

音声 UI を普及させるには、「使える」という視点だけでは不十分で、「使い続けたい」と感じさせることが不可欠である。そのためには、リアルタイムで音声 UI に対するユーザの振る舞いをモニタリングし、ユーザの習熟度を計測して、それに合わせて音声 UI が変化していくシステムを開発することが必要になるであろう。今後は、ユーザの習熟度を計測する技術の開発が重要になる。

それにより、「サービスとしての」、あるいは、「おもてなしとしての」音声 UI を実現する時代が到来することを期待する。それが、最近、よく語られる User Experience (UE, UX とも略される) の真髄とも通じるであろう。

参考文献

- 1) Jun Okamoto, Tomoyuki Kato, Makoto Shozakai, "Usability Study of VUI consistent with GUI Focusing on Age-Groups", Proc. Interspeech, pp. 1839-1842, Brighton, UK, 2009.
- 2) 岡本淳, 加藤智之, 小柳津聡, 庄境誠. 車載情報機器プロトタイプユーザビリティ評価によるユーザプロファイルの分析. 日本音響学会 2008 秋季研究発表会講演論文集, 3-Q-19