

# 3 使い勝手の良い 音声インタフェースの実現

庄境 誠

旭化成（株）新事業本部情報技術研究所

## 音声インタフェースの現状と課題

エレクトロニクス技術、情報通信技術の進展は、情報家電機器や車載情報機器の複雑化をもたらしている。音声認識技術、音声合成技術およびこれらの管理制御技術から構成される音声インタフェース技術は、それらの複雑な機能を初心者や高齢者を含む「誰でも」が「簡単に」活用できるための解決手段として期待が大きい。たとえば、代表的な車載情報機器である、カーナビゲーションシステム（以下、カーナビと略す）は、過去、ボタン、タッチスクリーンと画面表示からなるグラフィカルユーザインタフェース（以下、GUIと略す）による方式が一般的であった。しかし、近年は、運転中の安全なハンズフリー操作を実現したいという需要の拡大により、音声インタフェースが付加される製品が増えてきている。しかしながら、GUIと音声インタフェースの間にミスマッチが多々あり、カーナビの音声インタフェースは必ずしもユーザにとって使い勝手が良いとは言えなかった。そのほかに、カーナビの音声インタフェースが、必ずしも利用されていない実情の原因として、以下の大きな課題を指摘することができる。

**課題1：**自動車運転中の走行雑音に対する音声認識性能が十分高くないため、あるいは、さまざまな声質や発話様式を有するユーザ音声のバリエーションに対する音声認識性能が十分高くないため、初めて音声インタフェース機能を使った際に誤認識が発生し、音声インタフェースに対する印象が急速に悪化してしまい、二度と音声インタフェースを使わない場合が多発している。

**課題2：**音声インタフェースの利便性に対する世の中の共通理解がなく、ユーザもどの機能が音声で入力・制御可能なのか、どのような言い方が音声

インタフェースにより受理されるのかなど、音声インタフェースの基本的な使い方が分からないので、音声インタフェースを試そうとしない場合が多発している。

そのような現状を打開するために、初心者や高齢者を含む「誰でも」が「簡単に」情報家電機器や車載情報機器を活用できるユニバーサルな音声インタフェースの実用化をターゲットとして、2006年度からの3年間（先立って行われた先導研究の期間を含めると4年間）にわたって、経済産業省のプロジェクトとして「音声認識基盤技術の開発」が実施された。

この「音声認識基盤技術の開発」プロジェクトの中で、我々は、上記の課題1と課題2の解決に資する、以下の研究開発を行ったので、その成果を解説する。

まず、課題1の本質は、実環境における誤認識要因が十分に解明されていないことにある。音声認識性能を客観的に評価し、平均音声認識性能だけではなく、音声認識性能のばらつきを考慮した音声認識性能分布の形態で定量的に評価する必要がある。そこで、まず最初に、音声認識性能の劣化要因を、実環境の発話コーパスを用いて物理量の観点から分析する<sup>1)</sup>。

一方、課題2の本質は、使い勝手の良い音声インタフェースの設計方法論が確立されていないことにある。そこで、我々は、GUIと親和性が高く、使い勝手の良い車載情報機器用音声インタフェースを実現することを目指して、ユーザビリティ評価テストを繰り返し、車載情報機器実証システムを徐々に改良した。ユーザを適切にナビゲートできた結果、最終的には95%のユーザに対して95%のタスク達成率を達成した。普及のためには使い続けたいと感じる音声インタフェースであることが必要である。



## 特集 音声認識技術の実用化への取り組み

被験者種別	被験者詳細	人数	目的
一般被験者	一般人	220名	日本語発声の話者による広がり把握
プロ被験者	運転のプロ	40名	高速走行中の発話様式の変化を把握

表-1 被験者種別

そこで、タスク達成率だけでなく、音声認識機能が付いたカーナビの使用意欲に関しても調査を行ったので、この成果に関して解説する<sup>2)</sup>。

これらの知見や教訓を踏まえて、最後に、次世代音声インタフェースのあるべき姿について論じる。

### 性能評価技術

#### ●自動車運転行動中の音声データの収集

##### 《自動車運転行動中発話コーパスの概要》

自動車運転行動中の発話における対雑音性能、対話者性能、対語彙性能を明らかにするためには、それらの分析が可能な発話コーパスの確保が必要である。自動車運転行動中の発話コーパスとしては、CIAIR コーパス<sup>3)</sup>が知られているが、被験者の発話内容が完全に制御されておらず、対雑音性能、対話者性能、対語彙性能を客観的に分析するには好適ではなかった。そこで、以下の条件を満足する自動車運転行動中の発話コーパス（260名、延べ170時間）を新たに収集した。

1. 録音ゲインを全収録で固定
  2. 発話内容が複数の語彙種を含んでおり、全話者で共通
  3. 発話の際の自動車運転行動が、全話者で共通
- これにより、雑音性、話者性、語彙性を切り分けた分析が可能となった。

##### 《運転行動状態》

自動車内でカーナビを操作する状況を考慮し、アイドリング中 (Idle)、市街地走行中 (City)、高速走行中 (High) の3状態で収録を行った。ただし、安全上の理由から、被験者を一般被験者とプロ被験者に分け、一般被験者は公道での収録は行わず、運転教習所のコース内で市街地走行を模擬し、高速走行は行わなかった。被験者種別の詳細を表-1に示す。

サンプリング周波数	48kHz
ビット数	24bit
チャンネル数	4チャンネル
使用マイク	SONY ECM-77B
ハイパスフィルタ	SHURE A15HP
ゲイン	全話者全状態同一ゲイン
発声語彙の指示	TTSによる合成音声

表-2 収録条件

音圧	Spectral Subtraction 後の 10ms フレーム最大平均パワー
SNNR	プリエンファシス後の発話/雑音区間の音声+雑音と雑音のパワー比
発話速度	単位時間当たりのモーラ数
滑舌度	話者依存モデルの5母音間距離
話者固有度	話者依存モデルと話者非依存モデル間の距離

表-3 分析方法

##### 《収録条件》

収録条件を表-2に示す。なお、ハイパスフィルタは、高速走行時に路面の段差等による波形振幅のオーバーフロー(クリッピング)を防ぐために用いた。

#### ●音声データの物理量計測と統計的解析

収集した音声データの分析を行った。なお、分析には、自動車メーカーが音声認識用およびハンズフリー通話用にマイクを設置する場所として主流であるマップランプに設置した2つのマイクのうち、運転席側のマイクで収録した音声を使用した。分析を行った変動要因(物理量)は、音圧、SNNR (Signal with Noise to Noise Ratio)、発話速度、滑舌度、話者固有度の5つである。表-3に各物理量の分析方法を示す。

##### 《一般被験者のデータ分析》

一般被験者の全データを用いて、物理量の分析を行った。結果を表-4に示す。音圧、SNNRに関してはアイドリング時の男女差が大きく、走行中は差が小さくなることが分かる。これは、走行雑音の影響で音圧が相対的に大きくなる傾向(ロンバード効果)と走行雑音の含有比率が高くなることによると思われる。また、女性の標準偏差が大きく、話者による声量の差が大きいことが予測できる。発話速度は走行状態にかかわらず、女性の方がややゆっくり

### 3 使い勝手の良い音声インタフェースの実現

		Idle		City	
		Male	Female	Male	Female
音圧 [dB]	Mean	74.1	69.9	75.2	71.9
	SD	3.4	4.7	3.4	4.3
SNNR [dB]	Mean	21.1	18.4	16.5	15.5
	SD	4.6	5.1	5.0	5.2
発話速度 [more/sec]	Mean	7.7	7.1	7.8	7.2
	SD	1.2	1.0	1.3	1.0
滑舌度	Mean	3.0	3.0	2.8	2.9
	SD	0.5	0.4	0.4	0.4
話者固有度	Mean	0.68	0.65	0.57	0.56
	SD	0.14	0.15	0.12	0.15

Mean：平均 SD：標準偏差

表-4 一般被験者の物理量分布

		Idle		City		High	
		Male	Female	Male	Female	Male	Female
音圧 [dB]	Mean	74.0	68.0	76.1	71.8	79.1	77.3
	SD	2.8	4.3	3.0	4.5	2.7	3.6
SNNR [dB]	Mean	19.0	16.7	13.1	12.0	6.5	5.9
	SD	3.9	4.0	4.9	4.6	2.4	2.6
発話速度 [more/sec]	Mean	8.1	7.5	8.2	7.5	8.3	7.4
	SD	1.1	0.9	1.1	1.0	1.1	1.0
滑舌度	Mean	3.4	3.1	3.1	2.9	2.6	2.4
	SD	0.5	0.3	0.4	0.4	0.3	0.5
話者固有度	Mean	0.66	0.66	0.58	0.63	0.62	0.66
	SD	0.14	0.14	0.10	0.16	0.13	0.19

Mean：平均 SD：標準偏差

表-5 プロ被験者の物理量分布

発声することが分かる。滑舌度、話者固有度ともに、走行中は小さくなる傾向があるが、これは、滑舌度、話者固有度の差が走行雑音によりマスクされたためと考えられる。

#### 《プロ被験者のデータ分析》

プロ被験者の全データを用いて、物理量の分析を行った。結果を表-5に示す。アイドリング、市街地走行では、一般被験者と同様の結果である。高速走行においては、SNNR がかなり低下していることが分かる。

#### ●認識性能と物理量の関係

##### 《認識性能分布曲線》

収集したコーパスのプロ被験者の高速走行中のカーナビコマンド発声について認識性能を評価した。

音響モデル	monophone, 43音素, 単一正規分布
認識対象語彙	カーナビ用コマンドを想定した語彙
待ち受け語彙数	約 1,000 語
特徴ベクトル	MFCC (12) DMFCC (12) Dlog power (1)

表-6 評価条件

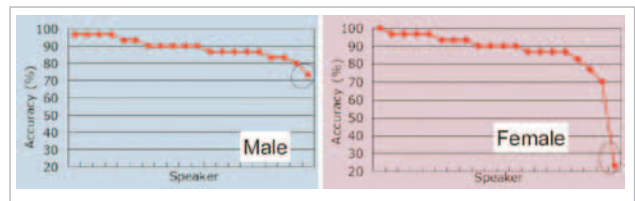


図-1 認識性能分布

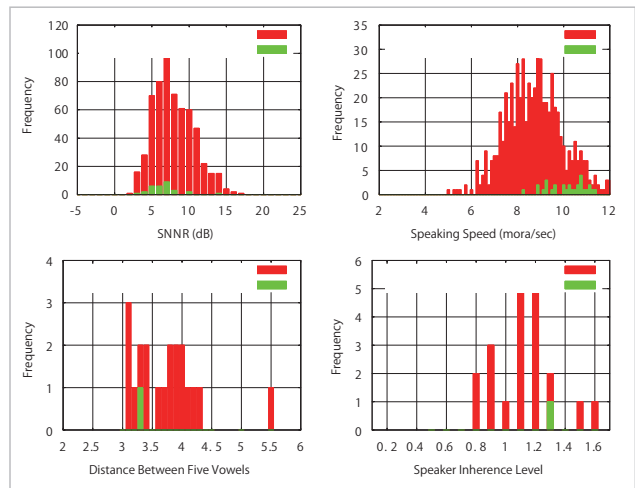


図-2 音声認識性能が低い話者(男性)の物理量ヒストグラム

評価条件を表-6に示し、認識性能分布曲線を図-1に示す。

#### 《音声認識性能が低い話者の物理量分析》

音声認識性能が低い話者の物理量の特徴を分析するために、図-1の○で囲んだ話者に着目し、分析を行った。

##### 男性

図-2に示す各物理量ヒストグラムにおいて、緑で示した部分が当該話者のヒストグラムである。当該話者は、発話速度が速いこと、滑舌が悪いこと、話者固有度が比較的大きいことが分かる。

##### 女性

図-3に示す各物理量ヒストグラムにおいて、緑で示した部分が当該話者のヒストグラムである。当



該話者は、SNNRが非常に悪いこと、発話速度の発話による変動が大きいこと、滑舌が悪いこと、話者固有度が大きいことが分かる。

以上の分析から、性能劣化に大きくかかわる要因は、滑舌度と話者固有度であり、元々SNNRが低い傾向にある女性話者に関しては、SNNRも音声認識性能劣化の大きな要因となることが分かった。

《複数の物理量が認識性能に与える影響》

前記から、音声認識性能が低い話者は、複数の物理量の分布が全体の分布に対して偏っていることが分かった。そこで、滑舌度と話者固有度に着目し、物理量の分布が偏っている話者の認識性能との関係を調査した。なお、本調査は、女性話者についてのみ行った。

滑舌度

滑舌度の値が小さい4名の話者に関して、話者固有度とSNNR、音声認識性能を表-7に示す。表の話者C、Dから、滑舌度が小さい話者でも、話者固有度、SNNRのどちらかが良い値を示した場合は高い音声認識性能を示すことが示唆されている。

話者固有度

話者固有度が大きい4名の話者に関して、滑舌度とSNNR、音声認識性能を表-8に示す。表の話者c、dから、話者固有度が大きい話者でも、滑舌度やSNNRが良い値を示した場合は高い音声認識性能を示すことが分かる。

以上の結果から、音声認識性能劣化要因のうち、複数の物理量の値が偏っていることで、音声認識性能が著しく劣化することが確認できた。

ユーザビリティ評価技術

●車載情報機器実証システム

GUIと親和性が高く、使い勝手の良い音声インタフェースを備えた車載情報機器実証システム（以下、実証システムと略す）を開発することを目指した。今回我々が実装した、音声で操作できる機能は以下の5つである。

- 電話番号発信

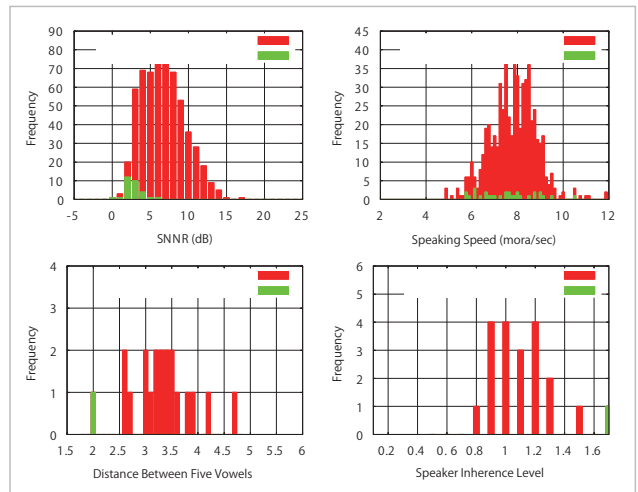


図-3 音声認識性能が低い話者(女性)の物理量ヒストグラム

話者	滑舌度	話者固有度	SNNR	Accuracy (%)
A	1.92	1.65	2.17	23.33
B	2.56	1.42	3.34	70
C	2.55	0.89	3.44	96.67
D	2.64	1.14	6.43	86.67

表-7 滑舌度の小さい話者の他の物理量と認識性能

話者	話者固有度	滑舌度	SNNR	Accuracy (%)
a	1.65	1.92	2.17	23.33
b	1.42	2.56	3.34	70
c	1.23	3.72	7.92	86.67
d	1.22	3.05	6.25	96.67

表-8 話者固有度の大きい話者の他の物理量と認識性能

- 電話帳発信
- 住所検索
- 施設検索
- 楽曲検索

実証システムの開発にあたり、市販のカーナビを利用して社内のスタッフを被験者とした簡単なテストを行った。このテストでは音声操作の開始方法だけを事前に説明し、その他の使い方に関しては説明を行わなかった。ただし、使い方で困ったときのためにカーナビに付属している使用説明書を渡しておいた。このような条件で、音声で目的地を設定するというタスクを被験者に実施してもらったところ、このテストに参加した被験者は使用説明書に目を通すことなくカーナビを使い始めるということが観察

### 3 使い勝手の良い音声インタフェースの実現

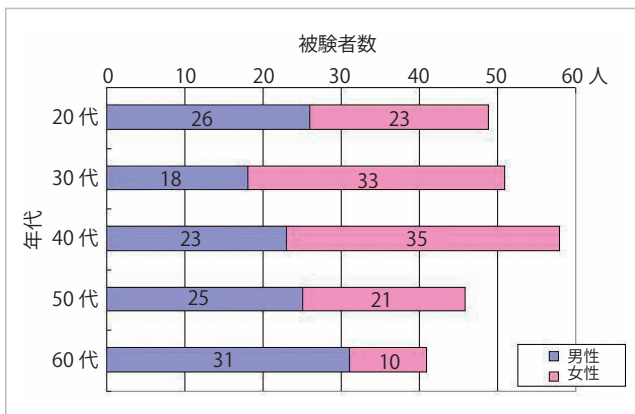


図-4 被験者の内訳

された。また、画面上に表示されているメッセージや音声ガイダンスを頼りに操作するという行動が観察された。これらの結果から、説明書を読まなくても操作できるようにGUIの表示ラベルと音声コマンドを一致させて、GUIとの親和性を向上させることとした。

従来の車載情報機器で使用されてきた施設検索はジャンル検索が主流であったが、開発者が想定するジャンルの分類とユーザが想定するジャンルの分類が必ずしも一致するわけではないため、うまく使いこなせないという問題があった。一方、インターネットの世界ではキーワードによる検索が主流であり、近い将来、車載情報機器はインターネットに常時接続されるようになると予想される。その際には、施設検索の方法はインターネットで利用されているキーワード検索の方法に置き換わるものと予想される。このため、実証システムの施設検索には、キーワード検索方式を採用することにした。

旭化成（株）製組込み用音声認識エンジンVORERO<sup>4)</sup>を用いて、PC上で動作する実装システムを試作した。さらに、早稲田大学で開発されたプロキシエージェント技術<sup>5)</sup>も実装し、被験者の行動をモニタリングできるようにした。操作中のログデータをすべて蓄積できるので、これを後から分析することにより、音声インタフェースの継続的な改良を行うことができた。ロギングおよびユーザビリティ評価は、使い勝手の良い音声インタフェースを実現するために大変重要な技術である。

#### ●被験者評価実験

被験者によるユーザビリティテスト（以下、UTと略す）の実施と実証システムの改良を交互に何度も繰り返した。UTに参加した被験者は男性123名、女性122名の合計245名であった。被験者の内訳を図-4に示す。

被験者の資格条件として、次の5つを設定した。

- 運転免許証を持っている
- カーナビを所有している、または使用したことがある
- 携帯電話を所有している
- PCの使用経験がある
- 年齢は20代から60代まで

UTは実験室または実車でを行い、両環境でタスク達成率に差がないことを確認した後は、すべてのUTを実験室で行った。UTの実施にあたっては、車載情報機器を購入した直後の状況を想定した。どのような機能が音声で使えるのか、どうやって音声認識機能をスタートさせるのかという初歩的な事項の説明を自動車販売店で受けている、またはあらかじめパンフレットを読んで知っていると想定した。そこで、UTの事前教示では以下の5つの説明を行った。

- 音声で操作ができること
- ハンズフリー機能を利用して、電話番号や電話帳に登録した相手の名前を音声で入力して電話をかけることができること
- 音楽をCDなどからハードディスクに保存し、再生することができること
- 発話ボタンを押すことで、音声認識が開始されること
- ビデオを用いた音声認識の利用方法

以上から分かるように、実証システムの具体的な操作方法に関しては一切説明を行っていない。UTで実施したタスクは、以下の8種類である。

- 電話番号で電話をかける
- 電話帳で電話をかける



## 特集 音声認識技術の実用化への取り組み

タスク	1回目	2回目	3回目
電話番号で電話をかける	1	1	1
電話帳で電話をかける	2	2	2
住所検索	3	4	3
施設検索	4	3	4
楽曲検索	5	5	5
電話番号で電話をかける (画面なし)	6		
電話帳で電話をかける(画面なし)	7	6	6
楽曲検索(画面なし)	8		

表-9 タスクの実施順序

- 住所検索
- 施設検索
- 楽曲検索
- 電話番号で電話をかける(画面なし)
- 電話帳で電話をかける(画面なし)
- 楽曲検索(画面なし)

ここで、画面なしとは運転中の場面を想定しており、画面には地図しか表示されない状態でのタスクとなっている。各タスクの内容を実施の前に被験者に説明した。この説明では、あらかじめ用意したタスク指示書を被験者に読み上げてもらい、内容を理解してもらった。

実証システムを操作している最中の被験者の考えを確認するために、操作の途中で音声認識が一時停止したタイミングで被験者にインタビューを行い、何を考えていたか、また、どういう点に戸惑ったかということの聞き取りを行って問題点の抽出を行った。また、実証システムではタッチパネルも使用できるが、音声インタフェースの問題点を抽出するために、被験者には音声だけで操作するように指示した。中には、つつい手で操作してしまう被験者も観察されたために、タッチパネルで操作することができないように設定して UT を実施した。

音声インタフェースを広く普及させるには、「誰でも操作できる」というだけでは不十分で、「ずっと使い続けたい」と感じることでなければならない。そこで、全体のタスク達成率が90%に達してからは、UTの前後で音声認識機能が付い

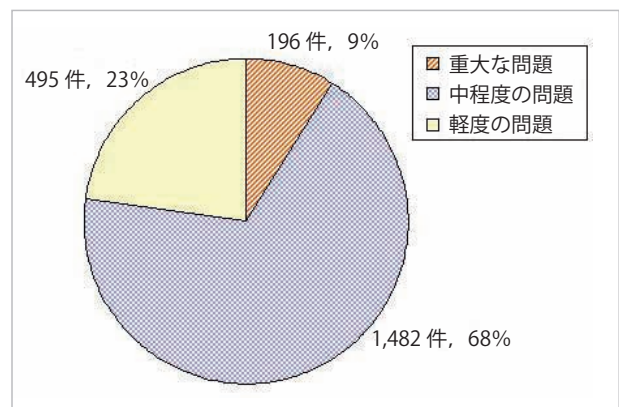


図-5 重要度別問題発生件数

重要度	定義
軽度の問題	操作は可能だが、気になる問題
中程度の問題	時間はかかるが、自力でタスクを達成
重大な問題	自力でタスクを達成できない

表-10 重要度の定義

たカーナビ(以下、音声認識カーナビと略す)の使用意欲に関する調査を行った。具体的には、実証システムの機能を説明した後に、「音声認識カーナビを使いたいと思いますか?」と質問し、そう思うならば5点、どちらとも言えない場合は3点、そう思わない場合は1点の5段階評価で点数をつけてもらった。また、UTが終了した時点で、「今日使った音声認識カーナビをずっと使い続けたいと思いますか?」と質問し、同様に5段階評価で点数をつけてもらった。

たった1回のタスクの実施で、被験者が本当に「ずっと使い続けたい」と判断することは難しいと考えられる。なぜなら、タスクを1回実施した直後の被験者はまだ操作に慣れていない状況にあり、何度か使ってみて操作に慣れた状況を再現できていないからである。そこで、表-9に示すように一部のタスクを何度か繰り返し実施した上で、使用意欲の調査を行った。

### ●設計指針

UTの結果、総計で2,173件の問題の発生を観察した。これらの問題を重要度別に分類すると、図-5のようになった。重要度の定義は、表-10に示す通りである。

### 3 使い勝手の良い音声インタフェースの実現

音声インタフェースの問題点として誤認識の問題が上位にくると想像しがちであるが、UTの結果を見るとアプリケーションの構造が分かりにくいという問題や、音声ガイダンスが曖昧なため被験者に誤解を与えてしまい、その結果としてタスクが達成できなかったという問題が多く発生したことが分かる。したがって、このような問題の発生を防ぐには、次の2つを設計指針として音声インタフェースを設計するとよいことが分かる。

- ①アプリケーションの仕組みを知らせる
- ②具体的な対応方法をユーザに知らせる

観察された重大な問題をさらにその内容で分類した結果を図-6に示す。

技術者はなるべく正確な情報をユーザに与えようと考えがちであるが、そのようにするとユーザには逆に分かりにくい表現になりがちである。ユーザは説明書を隅から隅まで読んでくれるとは限らないので、どんなユーザでも使えるような音声インタフェースを実現するには、アプリケーションをなるべくシンプルな構造にして、次に具体的にどんな操作をしたらよいかということをメッセージ表示や音声ガイダンスを活用して知らせることが大事である。

しかし、中にはメッセージ表示を読んでもくれない、音声ガイダンスを聞いてくれないというユーザも現実には存在し、年齢が上がるにつれて多くなる傾向が見られ、およそ10%程度を占めていた。そのようなユーザでも使えるような音声インタフェースにするには、

- ③情報を呈示する際は、その前にユーザの注意を引く

という設計指針に従って音声インタフェースを設計すると効果的であった。以上の3つの設計指針に従って音声インタフェースを設計すると、音声認識のタスク達成率を大幅に改善することができた。

しかし、「使える」ということと「使い続けたい」ということは別である。ユーザが初めてシステムを使う時には、「うまく使えるか」ということに関心が集

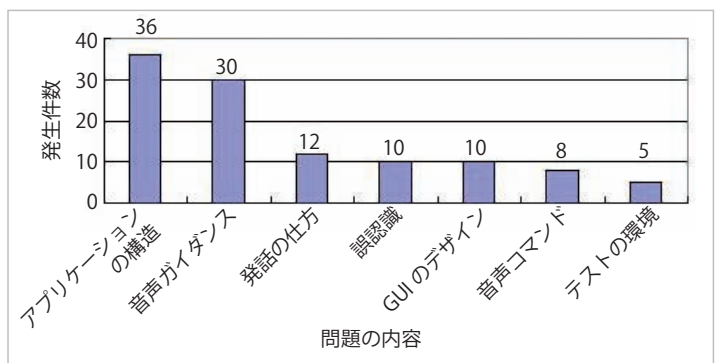


図-6 内容別重大な問題発生件数

中している。しかし、いったんうまく使えるようになると、今度は「使っていて楽しいか」ということに関心が移るようになる。我々は当初、「誰でも使える」音声インタフェースを目指していた。これは、ユーザが初心者であると想定しているということの意味する。このような想定のもとで開発した実証システムの使用意欲を調査した結果、実証システムの使用前には評価を5とした被験者が44名中21名いたのに対し、UT後には14名に減ってしまった。被験者へのインタビューの結果、最も多い原因は使っているうちに音声ガイダンスの冗長性が気になりだしたためであるということが分かった。一度使い方が分かってしまうと、毎回いちいち音声ガイダンスで使い方を説明されることが煩わしいと感じるようになってしまうと推察される。実証システムには音声ガイダンスの途中でもボタン操作で音声ガイダンスを中止させて、音声コマンドを入力させる(以下、手動バージインと呼ぶ)機能を実装していた。しかし、この手動バージインの使い方がユーザには分からなかったためにガイダンスの冗長性に煩わしさを感じるようになったので、手動バージインの方法を積極的に知らせるように工夫した。その結果、実証システムの使用後の使用意欲を5と評価する被験者の人数が使用前の人数よりも上回るようになった。しかし、改良を加えてもガイダンスの冗長性を問題点として挙げる被験者の人数を0にすることはできなかった。そのような被験者は年齢層が高い傾向があった。

このことは、人間は音声インタフェースの使い方



を速やかに学習できる一方で、音声インタフェースは人間の習熟度を計測できず、常にワンパターンの振る舞いしかできないというミスマッチ問題の存在を浮き彫りにしている。音声認識技術を広く普及させるためには、「誰でも使える」というだけでは不十分で、さらなる技術開発が必要であることを示唆していると考えられる。

### 次世代音声インタフェースの可能性

これらの研究成果を踏まえて、情報家電機器や車載情報機器の次世代音声インタフェースが備えることが望ましい機能について考察する。

#### ●習熟度計測

人間にとって音声は生来身に付けた最も自然なコミュニケーション手段である。従来、初心者や高齢者を含む「誰でも」が「簡単に」情報家電機器や車載情報機器を活用できるユニバーサルなインタフェースとして、音声インタフェースに安易に期待する向きが多かった。しかし、現在の音声インタフェース技術の能力はいまだ人間の音声対話能力には及ばない。音声インタフェースがユーザに提供できる機能・性能には制約がある。ユーザはその制約に習熟し、その制約の範囲の中で音声インタフェースを使いこなすことが望まれる。しかし、人間が音声インタフェースの制約に習熟する速度や程度はさまざまである。たとえば、先に述べた被験者評価実験の例では、表-9のタスクの3回目の繰り返しでの施設検索の年代別のタスク達成時間は表-11のようであった。

60代のタスク達成時間の平均値は、20代の約3.3倍である。最大値の1041.0秒を必要とした60代の被験者は、68歳男性であった。この極端な被験者を除いた場合の60代のタスク達成時間の平均値、標準偏差、最大値、最小値は、125.8、38.5、173.0、91.0秒であった。この場合でも、60代のタスク達成時間の平均値は、20代のその約1.5倍である。したがって、年齢層が高くなるにつれて、タスク達成時間は長くなる傾向がある。また、20

年代	平均値	標準偏差	最大値	最小値
20代	84.7	23.0	116.0	51.0
30代	102.1	34.3	152.0	63.0
40代	138.4	54.5	248.0	69.9
50代	117.6	27.5	162.0	73.0
60代	278.3	375.2	1041.0	91.0

表-11 施設検索の年代別のタスク達成時間(秒)

代でも、タスク達成時間の最大値は最小値の約2.3倍である。

このようにユーザによって、音声インタフェースに対する習熟の速度・程度には大きな差があり、ユーザの習熟度をリアルタイムで計測する技術の確立が待たれるところである。

#### ●パーソナライゼーション

情報家電機器や車載情報機器のユーザは、若年層から高齢者層までさまざまである。さらに、同じ年齢層でも、ユーザの育った環境、性格や嗜好などの複数の要因が組み合わさり、習熟度の違いも相まってユーザのプロファイルは多様性に満ちている。一方で、従来の音声インタフェースは、ユーザの習熟度をリアルタイムでモニタリングする機能を有しておらず、したがって、常に最初に設計・開発されたとおりの画一的な振る舞いしかしない場合がほとんどである。ユーザの習熟度だけではなく、ユーザの声質や発話様式、ユーザの行動パターンを計測できれば、そのユーザにカスタマイズされた音声インタフェースを随時提供できる。このようなユーザの個性やその時点でのユーザの習熟度にマッチした音声インタフェースを提供するパーソナライゼーションの有無が今後の音声インタフェースの競争ポイントになるのではなかろうか。

#### ●アンビエントインタフェース

最近、「いつでも、どこでも、だれとでも」つながるユビキタス情報社会の概念を包含する新しい概念として、「今だけ、ここだけ、あなただけ」を目指すアンビエント情報社会の概念が拡がりを見せている<sup>6)</sup>。これは、その時点でのユーザの環境に依存し



## 3 使い勝手の良い音声インタフェースの実現

たサービスを、ユーザが意識して選択するのではなく、環境側から提示してくれる情報環境の構築により、ユーザにとって快適な生活や社会を実現しようというものである。ICTが埋め込まれた実環境がユーザを包み込み (ambient)、ユーザに意識させることなくその振る舞いを計測し、時々刻々パーソナルな音声インタフェースを提供できるような製品やサービスの実現を目指すべきではないだろうか。

### ●ユーザエクスペリエンス

かつて米国アップルコンピュータ(現アップル)に勤務していた認知心理学者であるドン・ノーマン(Dr. Donald Arthur Norman)は、製品やサービスの使用・消費・所有などを通じて、人間が認知する有意義な体験のことをユーザエクスペリエンスと呼んだ。これは、製品やサービスを利用する過程の品質を重視し、ユーザが真にやりたいことを本人が意識しなくても「楽しく」「面白く」「心地よく」行える点に焦点を当てた、ヒューマンインタフェースやユーザビリティよりもさらに幅広い概念である。ユーザに意識させることなく個性や習熟度を計測し、ア

ンビエントにパーソナライズすることができる音声インタフェースは、まさにユーザエクスペリエンスの概念に合致しており、広く世の中に普及するのではなかろうか。

### 参考文献

- 1) Kato, T., Okamoto, J. and Shozakai, M. : Speech Analyses of Drivers' speech in a Car Environment, Proc. INTERSPEECH2008, pp.1634-1637 (2008).
- 2) Okamoto, J., Kato, T. and Shozakai, M. : Usability Study of VUI Consistent with GUI Focusing on Age-Groups, Proc. INTERSPEECH, pp.1839-1842 (2009).
- 3) 河口, 松原, 岩, 梶田, 武田, 板倉 : 実走行車内における音声データベースの構築, SLP-30-12, 情報処理学会, 音声言語情報処理研究会 (2000).
- 4) <http://www.vorero.com/>
- 5) Nakano, T., Fujie, S. and Kobayashi, T. : Extensible Speech Recognition System Using Proxy-Agent, Proc. ASRU, pp.601-606 (2007).
- 6) 村田 : アンビエント情報社会の実現に向けた取組み, 電子情報通信学会誌, pp.233-238, Vol.93, No.3 (2010).  
(平成 22 年 9 月 1 日受付)

---

庄境 誠 (正会員) [shozakai.mb@om.asahi-kasei.co.jp](mailto:shozakai.mb@om.asahi-kasei.co.jp)

昭和 58 年京大工学研究科修士課程修了。平成 10 年奈良先端大情報科学研究科博士課程修了。工博。現在、旭化成(株)情報技術研究所長、音声ソリューションビジネス推進部長、グループフェロー。音声認識、多次元尺度構成法の研究開発に従事。平成 18 年度本会山下記念研究賞、日本音響学会、電子情報通信学会、人工知能学会各会員。

---