

## 音声インタフェースにおける認知的負荷測定法とその評価

西本 卓也 † 高山 元希 ‡ 荒木 雅弘 ‡

東京大学大学院 情報理工学系研究科 † 京都工芸繊維大学 工芸学部 ‡

〒113-8656 東京都文京区本郷 7-3-1 E-mail : nishi@hil.t.u-tokyo.ac.jp †

〒606-8585 京都市左京区松ヶ崎御所海道町 E-mail : araki@dj.kit.ac.jp ‡

あらまし 音声インタフェースを利用するユーザの認知的負荷を測定することによって、音声応用システムの使いやすさや利便性を向上させる手がかりが得られる。本報告では、音声の聴取と応答を要求する主課題を、画面表示とキーボード入力による副課題と並行して行う、二重課題法の実験を提案する。本手法を用いることにより、音声応用システムを実際に使用しながら、ユーザが高い認知的負荷を要求された箇所を知ることができる。負荷の大小が異なる単語復唱課題を用いた実験の結果、提案手法は課題の認知負荷を反映しやすく、被験者の慣れの効果などの影響を受けにくいなど、提案手法の有効性が確認された。

キーワード 音声インタフェース, 認知的負荷, 二重課題法, 音声応用システム, 単語復唱課題

### An Evaluation Method of Cognitive Load for the Voice Interface Systems

Takuya NISHIMOTO †, Motoki TAKAYAMA ‡, and Masahiro ARAKI ‡

Graduate School of Information Science and Technology, The University of Tokyo †

Department of Electronics and Information Science, Kyoto Institute of Technology ‡

7-3-1 Hongo, Bunkyo, Tokyo, 113-8656 JAPAN E-mail : nishi@hil.t.u-tokyo.ac.jp †

Matsugasaki, Sakyo, Kyoto, 606-8585 JAPAN E-mail : araki@dj.kit.ac.jp ‡

**Abstract** The measurement of the cognitive load of using the voice user interface systems is helpful for improving the usability of such systems. We propose an evaluation method which performs the main task of listening and answering via voice and the subtask of pressing the key-board responding to the signs on the visual display simultaneously. Using this dual-task method, we can be aware of the part in the voice interface systems that the user paid much attention to, where the developer of the system must improve. Our preliminary experiments using the shadowing message tasks showed that the results of the method reflected the difference of the task difficulties significantly, and the results are less subject to the effects of the practice.

**Key words** Voice user interface, Cognitive load, Dual-task method, Voice application systems, Shadowing message task

## 1 はじめに

電話による音声ポータルや車載用音声対話システムなどにおいては、音声認識と音声合成を用いた音声インタフェースをいかに用いるべきか、あるいは、いかに評価するべきか、といったことが重要である。

音声インタフェース設計のガイドラインについては各種の提案がある。また、システム開発のできるだけ早い段階において、実際のユーザによる反復的なユーザビリティ・テストと改良が重要であるとされる（例えば [1]）。ユーザビリティ・テストにおいてはユーザ発話の認識率、タスク達成率、各操作手段の利用度、アンケートによる主観評価などが検討対象となり、検討結果に基づいてシステム発話（プロンプト）、受理可能なユーザ発話（文法）、状態遷移モデル、エンジンの設定などのチューニングが行われる。

音声インタフェースには利便性の高さが求められる場合と、認知的負荷の低さが求められる場合がある。利便性に関する評価は従来のユーザビリティ・テストで行われてきたが、車載情報機器の安全性などに影響する認知的負荷に関しては、その評価手法は十分に検討されてこなかった。

認知的負荷の低い音声インタフェースを実現するためには、実際に音声インタフェースを利用しているユーザの認知的負荷を測定し、認知的負荷の高い箇所を知ることが重要である。そこで我々は、音声インタフェースに対して二重課題法をいかに適用すれば、簡単に実験ができて信頼性の高い結果が得られるかを検討した。そして、画面表示とキーボード入力による早押しゲームを作成し、主課題である音声課題と副課題である早押しゲームを被験者に同時に行わせることにより、比較的長時間にわたる音声課題において、音声課題の認知的負荷を反映しやすく、被験者の慣れの効果などの影響を受けにくい評価ができることを確認した。本手法を用いることで、複数の音声インタフェース設計の良し悪しを比較したり、特定のシステムにおいて認知的負荷の高い部分を検出することが可能になる。

本報告ではまず、二重課題法を用いた実験手法について概説し、関連研究について述べる。次に、提案手法の詳細について述べる。さらに、提案手法の有効性

を調べるために行った実験とその結果について述べ、考察を行う。最後にまとめと今後の課題について述べる。

## 2 認知的負荷と二重課題法

本章では認知的負荷の理論と二重課題法の関連研究について概説する。

### 2.1 キャパシティー一定の法則

ある対象に心理的にどれだけ注意をしていたかを測定するためには二重課題法が用いられる。速く歩いているときに難しい質問を受けると、歩く速さを保つことが難しくなるように、人が一つの課題を行っている間に別の課題が加わると、心理的な折り合いをつける必要が生じる。人は心的負荷を複数の課題に配分することができても、費やすエネルギーを一定以上に増やすことができない（「キャパシティー一定の法則」）、心的エネルギーの限界を超えた場合は、課題に対する応答が遅れたり誤りが増加したりする。

例えば Reeves ら [2] は、映像に対する注意の度合いを測定するために、第一課題として被験者にビデオ映像を注視させながら第二課題として音を聞かせる実験を行っている。音はランダムなタイミングで発せられ、被験者は音が聞こえたらボタンをなるべく早く押すように教示される。音に対する反応時間は、被験者がビデオ映像に注意を傾けるほど遅くなる。

「キャパシティー一定の法則」が唱えられている一方で、人には心的チャンネルがいくつもあり、同時並行的に課題を遂行できるとする「マルチチャンネル理論」も提唱されており、現在もその是非は議論されている [3][4]。また、複数の課題を同時に解く場合には、課題の類似性、課題の困難性、経験値、の3つの要素の影響を受ける。しかも、二重課題法の影響は一般的に双方向であるため、第一課題が第二課題に影響を与えるだけでなく、第二課題が第一課題に影響を与えてしまうこともある。

本研究では、認知心理学的な知見そのものを議論の対象とはせず、これらの知見の工学的な応用という立場を取る。つまり、個々のモデルのいずれが正しいかを議論するのではなく、それぞれのモデルをある条件下での現象を説明するための手段と考える。本研究

においては、音声インタフェースの認知的負荷測定において、「キャパシティー一定の法則」を成立させるためにどのような条件を設定すればよいか、実験を通じて明らかにすることが重要である。

## 2.2 関連研究

清水ら [5] は車載用音声対話システムについて、利便性としてタスク達成時間に注目し、安全性として LED 刺激への反応遅れ割合に注目した評価を行っている。タスク達成時間の推定に新美ら [6] のモデルが有効であることを示したが、安全性については有意な実験結果が得られていない。

この実験では、巡回コースを運転しながら LED 刺激反応タスクと交通情報検索タスクを同時に繰り返し行っている。しかし、LED 刺激反応タスクの成績が音声対話システムの負荷を測定する規準となり得ることが実証されていない（関連研究 [7] では記憶課題の有無が LED 刺激反応タスクの成績に関連することのみが示されている）。また、この実験では、被験者に与える課題が、運転、LED 刺激反応タスク、音声対話、の三重課題になっている。被験者によって音声対話タスクの負荷が運転と LED 刺激反応タスクのどちらに影響するかわからないので、予測される結果が出ていないと考えられる。

Strayer ら [8] は二重課題法を用いて携帯電話での会話が運転に与える影響を調べている。被験者は、模擬運転課題（ハンドル操作）を行いながら信号反応課題（赤信号が点灯したらできるだけ速くブレーキを踏むこと）を行い、これらと並行して音声による課題を行わせている。模擬運転課題では 3 種類の正弦波を加算したコースを用いている。結果として、携帯電話での会話は、携帯電話の形状（ハンズフリー型、ハンドヘルド型）とは無関係に模擬運転課題や信号反応課題に悪影響を与えること、単語復唱課題は運転に影響を与えないが、単語生成課題（聴取した単語に基づいて新たな単語を考える課題）は影響を与えること、などが確認されている。この研究は、視覚と手の操作を伴う第一課題と音声による第二課題の間で二重課題法が利用可能であることを示している。しかし、一部の実験がハンドル操作と信号反応を含む三重課題になっている。また音声課題がない場合と単語復唱課

題の間に有意な差がないことから、負荷測定の分解能が十分でない可能性がある。

これらの先行研究は、いずれも自動車運転という具体的な状況を想定したものであり、一般道や試験コースにおける実車の運転や、ドライビングシミュレーターなどの設備を必要とする。運転課題は負荷が一定であったり反復的であったりしては課題として成立しないため、コース中に高負荷の状況とそうでない状況が含まれる。従って、運転課題が他の課題から受ける影響の大小はコースの状況によって異なる。

また、被験者が走行コースなどに慣れてしまうことにより、運転課題の負荷が下がり、第二課題に与える影響が小さくなってしまう恐れもある。

音声課題と併用する第二課題を作成するにあたっては、周期的に一定の条件で結果を測定可能で、慣れの効果が少ないように工夫する必要がある。このような課題を音声課題と並行して行うことによって、第二課題の成績から音声課題の認知的負荷の高低を知ることができる。

本研究と目的は異なるが、自動車の運転状況から運転者の余裕度を推定し、運転余裕のあるときに音声対話を用いた情報提示を行う運転状況適応型音声インタフェースの提案があり、内山ら [9] は二重課題法を用いた運転余裕度の定量化を試みている。人による音声言語の処理は記憶負荷を伴うことから、合成音声をも 5 単語ずつ聴取して回答する記憶課題が用いられ、アクセルオフの状況において正答率が下がることが確認されている。この実験は、運転操作の認知的負荷を測定するものであるが、運転課題と音声課題の間で二重課題法が利用可能であることを示している。

本報告は、内山らの研究と類似の状況を想定しつつ、音声課題を第一課題に見立てて「対話余裕度」を得る手法の提案である。

## 3 提案手法

本章では、我々が提案する実験手法について述べる。

### 3.1 早押しゲーム

上述の要件を備える副課題として、早押しゲームを作成した。実装には Microsoft Visual Basic 6.0 を用いており、動作環境は Microsoft Windows 2000 で

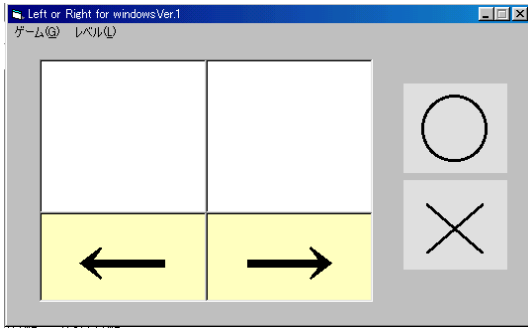


図 1: 早押しゲームの画面構成

ある。早押しゲームの画面構成を図 1 に示す。

画面左には 2 つの空欄および 2 つの矢印が並んでおり、ゲームが始まると 2 つの空欄の一方に左右いずれかの矢印 ( ← , → ) がランダムに現れる。左の空欄に左向きの矢印が表示された場合と、右の空欄に右向きの矢印が表示された場合には、 ← に対応する上矢印 ( ↑ ) キーを押すこととする。それ以外の場合には、 × に対応する下矢印 ( ↓ ) キーを押すこととする。画面右には上下に × の記号が並んでいるが、これは および × と上下矢印キーの対応関係を示している。キーを押すと、正解不正解に関わらず 1 秒後に次の矢印が現れる。ゲーム実行中は、1ms 単位でのユーザの応答時間と出題や応答などの内容を記録している。

このゲームは慣れるとほぼ反射的に操作できるようになり、約 2 秒ごとに試行が行われる。同時に音声によって他の課題を行うことも可能である。しかし、意識の集中の程度によって応答時間に差が出ることが予想される。

### 3.2 二重課題法の実験手法

提案手法は、まず、被験者が十分慣れるまで早押しゲームを練習させる。続いて、被験者に第二課題として早押しゲームをさせながら、同時に第一課題である音声課題を行わせる。被験者には事前に、二つの課題を同時に行う場合には音声課題を優先すること、早押しゲームについては正確さを優先しつつ、できるだけ速く応答すること、を指示しておく。

被験者実験の後、早押しゲームにおける誤回答の試行は欠損値とし、同時に行われた音声課題の内容と対応付けて各試行の結果を集計する。ある音声課題における早押しゲームの応答速度が、他の課題と比較し

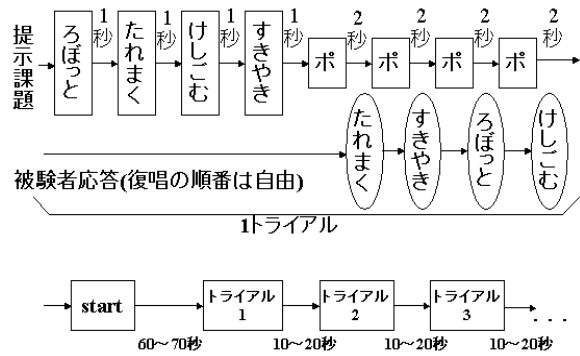


図 2: 4 単語の単語復唱課題

て有意に遅れていれば、その音声課題は有意に認知的負荷が高いとみなすことができる。

## 4 提案手法の評価

本章では、提案手法の有効性を検証するために、認知的負荷の違いが自明であるような課題を被験者に与え、有意な差が得られることを確認する。内山ら [9] と同様、我々も、音声インタフェースを用いた作業においては、聴取したプロンプトを記憶しておくことの負荷が大きいと考え、単語復唱課題を用いる。

単語復唱課題の語彙サイズは 166 単語である。単語の頭文字が濁音を除く 50 音としてバランスよく出現するように 2~4 モーラの普通名詞を無作為に選び、聞き取りにくいものを除く。読み上げには合成音声 ( 東芝 LaLaVoice2001 ) を使用してランダムな順序で提示する。

### 4.1 実験 1

まず実験 1 として、早押しゲーム課題と音声課題を同時に行う場合に「キャパシティー一定の法則」を満たす結果が得られることを確認する。

実験 1 で用いる単語復唱課題の概要を図 2 に示す。4 つの単語が 1 秒間隔で提示された後、2 秒間隔で短いピープ音が提示される。この音に続いて、被験者は記憶した単語を 1 つずつ答える。復唱する単語の順番は自由である。これを単語復唱課題の 1 試行 ( 約 20 秒 ) とし、各試行は 10~20 秒のランダムな間隔で繰り返される。

被験者は 20~25 歳の男性 5 名である。各被験者は約 8 分間、早押しゲームのみ ( 単一課題 ) の状態と単

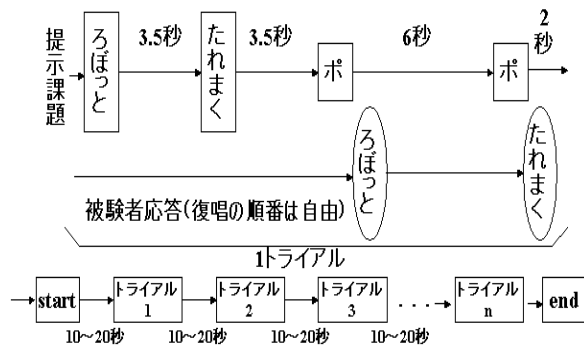


図 3: 2 単語の単語復唱課題

語復唱課題との同時実行（二重課題）の状態を繰り返しながら実験を行った。

## 4.2 実験 2

実験 2 では、認知的負荷が自明に異なる二種類の音声課題を設定した場合に、第一課題の負荷の違いが第二課題の成績に影響することを検証する。単語復唱課題において同時に記憶する単語の数の多少が、課題の認知的負荷の違いをもたらすものとして実験を行う。

単語復唱課題には、2 単語課題と 4 単語課題を用意する。4 単語課題は実験 1 と同一条件である。2 単語課題は図 3 に示すように、単語提示から被験者の復唱に至るまでの 1 試行の時間を、4 単語課題とほぼ同じにするために、単語の提示間隔を 3.5 秒に、回答を促すピープ音の間隔を 6 秒にしてある。これにより、単位時間当たりの負荷をほぼ一定にすることができる。

被験者は 20 ~ 25 歳の男性 9 名女性 1 名である。被験者を 5 名ずつ A, B の 2 群に分け、A 群は 4 単語課題の後に 2 単語課題を行うこととし、B 群は 2 単語課題の後に 4 単語課題を行うこととする。A 群の具体的な実験手順は以下の通りである（B 群の場合は 3-4 と 6-7 が入れ替わる）。

- 1 ゲームの説明と練習
- 2 ゲームだけを実行 (8 分)
- 3 4 単語課題を十分慣れるまで練習
- 4 ゲームと 4 単語課題を同時に実行 (8 分)
- 5 休憩
- 6 2 単語課題を十分慣れるまで練習
- 7 ゲームと 2 単語課題を同時に実行 (8 分)

表 1: 実験 1 における応答時間の平均および有意差

被験者	二重課題	単一課題	有意差
	[msec]	[msec]	
1	596	540	**
2	617	531	**
3	1001	603	**
4	586	496	**
5	460	419	**

表 2: 実験 2 における応答時間の平均および有意差 (A 群)

被験者	2 単語課題	4 単語課題	有意差
	[msec]	[msec]	
1	542	1191	**
2	458	589	**
3	486	675	**
4	496	683	**
5	426	562	**

## 5 結果と考察

本章では、前章で述べた実験の結果を提示し考察を行う。

### 5.1 実験 1

実験 1 における各被験者のゲーム応答時間の平均と分散分析の結果を表 1 に示す (\*\*は 1%水準で有意であることを示す)。

全ての被験者において、単語復唱課題を実行している間（二重課題時）のゲームの応答時間は、単語復唱課題の待機中（単一課題時）よりも有意に増加した。この結果より音声課題と提案した早押しゲームに関して「キャパシティー一定の法則」が適用可能であることを確認できた。

### 5.2 実験 2

実験 2 における各被験者のゲーム応答時間の平均と分散分析の結果を各被験者についての応答時間の平均と分散分析の結果 (p 値) を表 2 および表 3 に示す (\*\*は 1%水準で有意であることを示す)。

表 2 および表 3 より、被験者 10 人中 9 人において 1%水準で有意差が確認され、2 単語課題に比べて

表 3: 実験 2 における応答時間の平均および有意差 (B 群)

被験者	2 単語課題 [msec]	4 単語課題 [msec]	有意差
1	695	1060	**
2	401	418	-
3	523	868	**
4	712	941	**
5	478	519	**

4 単語課題を実行中のゲーム応答時間が遅くなることが確認された。

また、被験者 10 人の結果を、1 回目と 2 回目の慣れの差 (要因 1) と単語数による課題の差 (要因 2) に分けて、それぞれ検定を行ったところ、慣れの差については  $p = 0.7752$  (有意差なし)、課題の差については  $p = 0.0165$  (5%水準で有意) という結果が得られ、提案手法が課題の差を反映しやすいだけでなく、慣れの差の影響を受けにくいことが確認された。

## 6 まとめと今後の課題

本報告では、音声インタフェースを使用する課題の認知的負荷を二重課題法によって測定するために、画面表示とキーボード入力による早押しゲームを用いた実験方法を提案し、単語復唱課題を用いてその有効性を検証した。

提案した早押しゲームは 1 回の試行が 2~3 秒であり、反復的に実行されるため、数秒の時間分解能で認知的負荷の高かった箇所を知ることができる。また、被験者の慣れの効果の影響を受けにくいため、所要時間の長い音声課題を評価したり、一人の被験者が繰り返し評価実験を行うような場合にも適している。また、模擬運転課題を用いた Strayer らの先行研究と異なり、比較的認知的負荷が低い単語復唱課題においても、負荷の高低を測定できることが確認された。

これらの結果が得られた理由として、使用した早押しゲームの成績が 1 ミリ秒単位の応答時間で得られるため小さな認知的負荷の差であっても結果に反映されやすいこと、課題の所要時間が短いため統計的有意差をもたらすサンプル数が得やすいこと、課題の負荷が常に一定であるため模擬運転などと比較して安定した結果が得られること、などが挙げられる。

今後の課題としては、音声応用システムの改良に役立つと期待されるさまざまなガイドラインや対話制御手法を取り上げて、本手法を用いてその有効性の検証を行っていききたい。これにより、例えば、車載用音声対話システムにおいて安全性を高める手法を、経験的ではなく実験的に導いていきたい。その際には、本手法で得られる認知的負荷と、自動車の運転など実環境における負荷の関連性についても検討していく必要がある。

## 参考文献

- [1] Bruce Balentine, Devid P. Morgan: How to Build a Speech Recognition Application, 2nd Ed., EIG Press, 2001.
- [2] Byron Reeves, Clifford Nass: The Media Equation, CSLI, 1996. (細馬宏通 訳: 人はなぜコンピューターを人間として扱うか「メディアの等式」の心理学, 翔泳社, 2001.)
- [3] Michael W. Eysenck, Mark Keane: Cognitive Psychology – A Student’s Handbook, 4th Ed., Psychology Press, 2000.
- [4] Nick Lund: Attention and Pattern Recognition, Routledge, 2001.
- [5] 清水司, 小島真一, 脇田敏裕, 本郷武朗: 運転中における音声対話システムの評価, 情処研報, 2000-SLP-32-17, pp. 87-92, 2000.
- [6] 新美康永, 西本卓也, 荒木雅弘: 確認対話の制御方式の効率と音声認識システムの性能との関係, 情処研報, 99-SLP-27-17, pp. 111-118, 1999.
- [7] 小島真一他: 音声対話の運転への影響評価法の開発, 自動車技術会学術講演会前刷集, No. 91-99, pp. 17-20, 1999.
- [8] David L. Strayer, William A. Johnston: Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone, Psychological Science, 12, pp. 462-466, 2001.
- [9] 内山祐司, 小島真一, 本郷武朗, 脇田敏裕: 運転状況適応型音声情報提示システム, シンポジウム「ケータイ・カーナビの利用性と人間工学」, 11, 2001.