

音声インタフェース評価における慣れの影響の分析

菊池 英明 北村 泰司 本島 嘉朗 山田 哲史 宮澤 幸希 白勢 彩子

早稲田大学人間科学学術院

〒359-1192 埼玉県所沢市三ヶ島 2-579-15

kikuchi@waseda.jp

あらまし 音声認識技術実用化に向けて、実用的な音声インタフェースの設計指針を明確にするためのインタフェース評価方法の確立を目指している。評価実験においては多くの要因を同時に制御する必要が生じるため、統制が容易な要因についてはその影響を明らかにして事前に統制することが望ましい。本稿では、事前訓練や経験による習熟（本研究では「慣れ」と呼ぶ）が被験者の操作や評価に与える影響を調べ、幾つかの知見を得た。具体的には、(1)慣れを考慮する場合には利用回数と利用間隔の両方を考慮する必要があること、(2)トークスイッチ操作の認知的負荷にも慣れの影響があり得ること、(3)生体信号計測は被験者の負荷状況の推測に利用できる可能性があること、などである。

キーワード 音声インタフェース、ユーザビリティ、習熟、認知的負荷

Influences of experience on evaluation of speech interface

KIKUCHI Hideaki KITAMURA Taiji MOTOJIMA Yoshirou
YAMADA Tetsushi MIYAZAWA Kouki and SHIROSE Ayako

Faculty of Human Sciences, Waseda University

2-579-15, Mikajima, Tokorozawa-shi, Saitama, 359-1192, JAPAN

kikuchi@waseda.jp

Abstract We aim to establish a methodology of speech interface evaluation in order to pursue real use of speech recognition technologies. We analyzed influences of experience on operation and evaluation of speech interface through experiments. Twelve subjects are divided into three groups which differ with interval of use. Experimental results show that response time of subjects and objective evaluation of cognitive workload decrease with experience of use of the speech interface. Moreover, the results show that interval of use have influences on speech interface evaluation.

Keywords Speech interface, Usability, Experience, Cognitive workload

1. はじめに

音声メディアには他のメディアにない数多くの利点がある[1]ことや、近年の認識・合成技術の発展に対して、音声インタフェースの普及は十分とはいえない。石川ら[2]が指摘するように、要素技術の性能向上とは別に、音声インタフェースの使用性(ユーザビリティ)を正しく評価し向上させることが今後重要になってくる。

一般的なインタフェースのユーザビリティ評価方法としては、標準規格が作成され広く利用されている。カーナビゲーションシステム用音声入力の性能評価のためのガイドライン案[3]が作成されるなど、音声インタフェースのユーザビリティ評価の基本的な方法はある程度整いつつあるといえる。

そうした中で、これまで質問紙などによる主観評価に依存していた認知的側面の評価を、より詳細かつ客観的に行なう方法の導入が試みられ始めている。西本ら[4]は対話に伴われる利用者の負荷量を二重課題法によって測定することを試み、負荷が大きくなる対話状態の発見に成功している。野田ら[5]は音声対話における負荷量を二重課題法やNASA-TLXによって分析し、対話戦略と負荷量の関係を明確にする見通しを得ている。他方、自動車運転に音声インタフェース操作が与える負荷については[6][7][8]など多くの検討が進んでいる。音声インタフェースのより詳細かつ実際のユーザビリティ評価を実現するために、これらの先行研究によって得られた知見を含めて評価方法を確立することが必要である。

音声インタフェースのユーザビリティ評価方法を確立するために、評価実験を繰り返し、音声インタフェースに対する評価と各要素技術の性能や利用者特性との関係を認知科学的な側面から分析するアプローチをとる。その際、音声認識性能や利用者特性のバランスをとるうえで、被験者の習熟の程度を考慮して統制することが必要になる。そこで、本研究ではまず音声インタフェース評価に習熟が与える影響を明らかにすることを目指す。

音声インタフェース利用における習熟の影響を調べた研究として原ら[9]の楽曲検索システムを対象とした事例がある。被験者12名が1日1時間のセッションを5日間にわたって行

なう実験を通して、被験者が音声インタフェースの利用回数を重ねるごとに音声認識性能が向上し、それに伴ってタスク成功率と満足度が向上することが報告されている。習熟による認知的負荷の変化は調査されていない。一方、内田ら[10]はカーナビゲーション操作タスクを設定して実験を行い、ビデオ解析による視認回数・時間測定を行った。実験は4名の被験者を対象に行われ、1名の被験者につき5回の走行(20分)を1~3日間隔で実施した。その結果、操作の習熟過程の段階と各段階における特徴を見出した。音声操作は対象とされていない。

本研究では、これらの先行研究を踏まえて、例えば利用回数だけでなく利用間隔の影響や、習熟に伴う操作や評価の詳細な変化、認知的負荷の変化を明らかにすることを目的として実験を行なう。具体的には、利用間隔の異なる3つの群に被験者を分けて、音声インタフェース利用時の行動や評価の変化の違いを比較する。2章で実験方法を示し、3章には実験の結果、4章にはまとめを述べる。

なお、[9]では事前訓練の影響も調べており、事前にフィードバックを与えて訓練を行なうことでタスク成功率が向上することがわかっている。音声対話システムにおいて4分間の事前訓練がユーザのパフォーマンスや満足度の継続的な向上に貢献することも報告されている[11]。このように事前訓練も、利用経験と同様に習熟して慣れた状況に達する重要な手段であるが、本稿では利用経験の影響に焦点を絞り、事前訓練は行なわない。また、本稿では事前訓練や経験による習熟を以降、慣れと呼ぶ。

2. 実験方法

本章では音声インタフェース利用経験が被験者の操作や評価に与える影響を調べるための実験について説明する。

2.1 実験設定

被験者は12名とし、音声インタフェース評価の経時変化を見るために、各被験者は時間をおいた計3日、同一の試行を行った。慣れの状況による違いを見るために、12名を4名ずつの3群に分け、それぞれ3日の試行を実施する間隔を(1)毎日、(2)2、3日おき、(3)一週間お

きとした。

タスクはカーナビゲーションシステムにおける目的地設定とし、特に施設名による目的地検索を対象とした。実験に先立ち、市販されている3つのカーナビゲーションシステムの製品マニュアルを参考にして施設名による目的地検索の実現方法を調べた。既存の製品の實現方法にならって、目的地設定や誤認識時の訂正などの音声操作が Wizard of OZ 方式で實現できるようにシミュレータを実装した。

また、将来的に生体信号計測やプロトコル分析などの認知実験手法を用いて、被験者の認知的な状況を綿密に分析して評価に用いることを計画している。本実験においてもそれらの手法と慣れの関係を見るために予備的に計測を行なった。実験における計測対象は以下のとおりである。

音声：

ヘッドセットマイクからの被験者の音声とシステムからの出力音(声)

生体信号：

皮膚伝導性(発汗量)、心拍、血液容積脈(いずれも緊張状態、集中状態、ストレスに関係している。)

プロトコル分析：

音声インタフェース利用時の発話プロトコル、機器操作記録、システム状態、行動を記録した動画像

質問紙：

音声インタフェースの評価項目に対する主観的評価として NASA-TLX を用いる。音声インタフェース評価に適した形で日本語化した[5]を参考にした。

実験の環境を図1に示す。被験者の音声インタフェース利用の様子として音声、ビデオカメラ2台による映像、トークスイッチ操作、生体信号(心拍、発汗量、血液容積脈)が記録される。

2.2 実験手順

一人の被験者は、一日の実験で12個の目的地設定を行う。一回の実験のうち任意の4回の発声に対して Wizard of OZ 方式によって誤認識を発生させる。なお、一人の被験者が行なう3日にわたる実験ではそれぞれ目的地群を変えた。

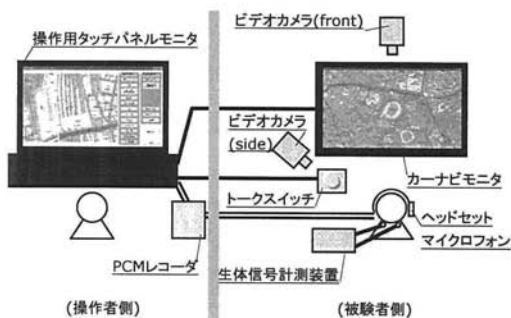


図1: 実験環境

発声に際してはトークスイッチを押してプロンプト音を聞いた後に発声するよう、あらかじめインストラクションを与えた。

以上の設定で行った実験の結果から、特に慣れがユーザの振る舞いに与える影響や、インタフェース評価に与える影響を調べた。次章に実験の分析結果を述べる。

3. 実験結果

3.1 慣れによる振る舞いの変化

転記テキストを作成し、被験者応答時間やシステム発話時間などの情報から、慣れによって被験者の振る舞いがどのように変化したかを分析した。分析の対象とするポイントを図2の模式図に示す。発話タイミング(1)はユーザの応答時間に、発話タイミング(2)はトークスイッチから発話に至るリズムに関係することが想像される。また、発話速度は音声の音響的特徴に影響を与えると考えられる。いずれも、音声インタフェース利用に際しての認知的負荷に関係すると考えられる。

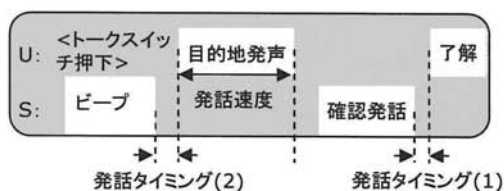


図2: 分析の対象とするポイント (Uは被験者、Sはシステム)

3.1.1 発話タイミング(1) 応答時間

システム確認発話終了後、被験者が応答を開

始(音声誤認識の場合はトークスイッチ押下)するまでの時間を計測したところ、図3の結果が得られた。正しく認識した場合(“Correct-Recognition”)、誤認識の場合(“Incorrect-Recognition”)、およびそれらを区別しない場合のいずれも、利用回数が進むにつれて応答時間が短縮されている。平均短縮率は正しく認識した場合で 12.7%、誤認識の場合で 20.7%であった。

ここでこの結果を詳細に観察するために、2~3日おきに実験を行う群(2)に所属する1名をとりあげて試行ごとの変化の違いを示した図が図4である。1回目は平均870.3[ms]であったのが2回目は316.6[ms]と大幅に(64%)減少した。また、誤認識に限って見た時、1回目が2004.0[ms]であったのが2回目には694.0[ms]となり、これも65%減少した。この図に見られるとおり、誤認識時を除けば各回の中でのばらつきが小さいのに対して回数の違いによる変化は顕著である。

回数による応答時間の違いを、正しく認識した場合に限定して、利用間隔を変えた被験者群ごとに見たのが図5である。

慣れの影響によって応答時間が短縮するのは、2~3日おきに利用した場合までで、間隔

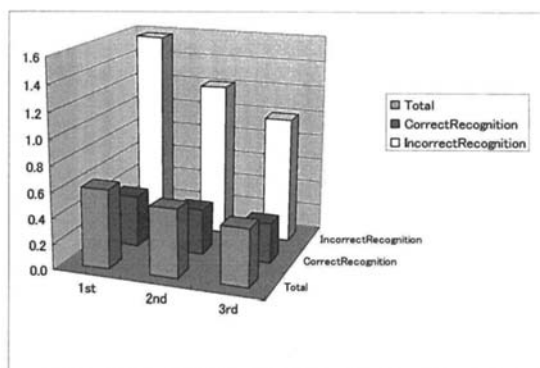


図3: 被験者全体の応答時間の経時変化(単位は[s])

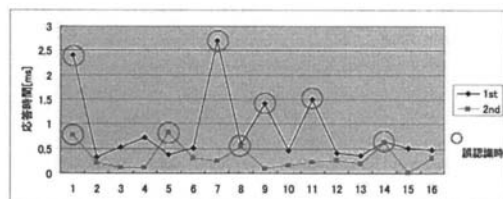


図4: 被験者一名の応答時間の経時変化

が一週間あくと慣れの効果が見られない。

これらのことから、被験者の操作において利用経験の影響を考慮する際には、利用時間だけでなく利用回数やその間隔を考慮する必要があるがえる。

3.1.2 発話タイミング(2) トークスイッチ

次に、トークスイッチ押下時のビープ音後、発話開始までの時間(発話開始時間)を計測した結果を図6に示し、正しく認識した場合に限定した被験者群ごとの違いを図7に示す。

図6では、図3の応答時間と同様に利用回数とともに時間が短縮されている。全体の平均短縮率は14%程度であった。慣れによってスイッチ押下から発声までのリズムの習得が進んだものと考えられる。図7において被験者群による傾向の違いは見られなかった。

3.1.1の結果とあわせて、回数を追うごとの変化の割合は3回の試行を終えても収束していないため、4回目以降の変化の程度をあらためて調査する必要がある。

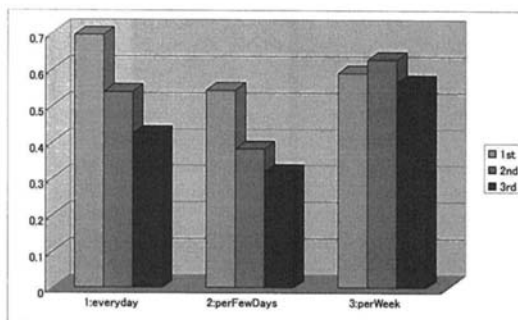


図5: 利用間隔を変えた被験者群ごとの応答時間の変化(単位は[s])

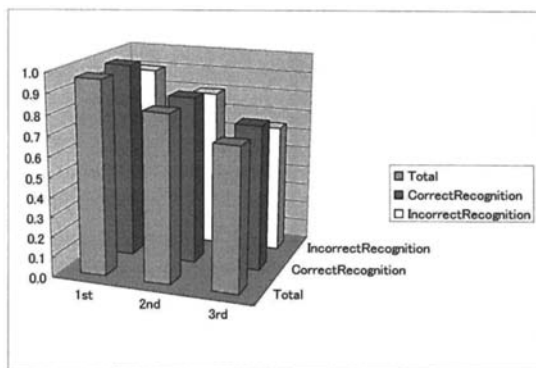


図6: 被験者全体の発話開始時間の経時変化(単位は[s])

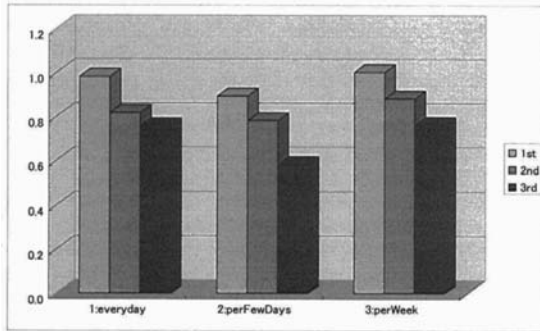


図 7: 利用間隔を変えた被験者群ごとの発話開始時間の変化(単位は[s])

3.1.3 発話速度

目的地設定発話の発話速度を毎秒モーラ数で計測したところ、被験者全体で1回目から3回目にかけてそれぞれ平均 6.95、7.30、7.25であった。被験者群を区別して見ても1回目から2回目にかけて発話速度が速まる傾向は同様だが他に明確な傾向は見られなかった。

一部の被験者において、利用回数が進むにつれて誤認識後の発話速度がより遅くなるケースがあった。慣れによって誤認識後は丁寧に発声する可能性が考えられるが、音響的な変化について言及するにはデータ量が少ないため、今後データ量を増やしてあらためて調査する必要がある。

3.2 アンケート

先に述べたとおり、NASA-TLXに基づくアンケート調査を行った。図8にその結果を示す。図に示したのは2回目と3回目の平均評価点であるが、いずれの項目においても2回目の方が3回目よりも小さい値の評価点となっており、これにより心的負荷状況が低くなったと見ることができる。

さらに被験者群毎のアンケート結果の違いを見たところ、2回目から3回目にかけての総合的な評価点の変化量は群(1)が-0.61、群(2)が-0.19、群(3)が-0.11となっていた。このことから、被験者の評価において利用経験の影響を考慮する際には、実験回数だけでなく間隔に注意する必要があることが示唆される。

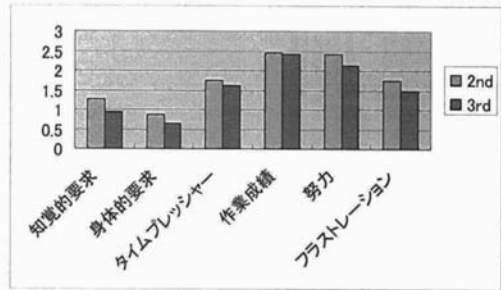


図 8: 主観評価結果

3.3 生体信号計測

最後に、予備的に計測した生体信号とユーザの振る舞いの関係を考察する。

図9に、群(3)に属する被験者1名について、1回の実験中の全ての音声波形と計測した生体信号の推移を示す。音声波形のうち上側が被験者、下側がシステムの音声であり、生体信号のHRは心拍、SCは皮膚コンダクタンス、Tempは皮膚温を表す。また、最下段には参考のために応答時間の推移を示した。

図中、実線で囲んだ箇所は誤認識後のシステム発話を聴取中であることを示したが、このいずれのケースもHR指標に変化が見られる。この被験者は3.1.1の応答時間においても誤認識後には著しく遅れが見られていることから、システム誤認識によって負荷が増したことが生理指標の変化にもあらわれたと考えることができる。

また、点線で示した箇所は読み方の難しい地名を発声している場面であり、他の手段ではとらえにくいそうした負荷状況を生体信号計測によってとらえられる可能性が示唆される。

ただし、いずれの考察も今後事例を増やしたうえで再度あらためて検討を行う必要がある。

4. おわりに

音声インタフェースの評価方法を確立し、音声インタフェースの設計指針を明確にすることを目的として、まず評価に慣れが与える影響を調べるための予備的な実験を行なった。実験を通して、慣れの影響を見るための基本的な方法の確認を行い、幾つかの知見を得た。具体的には、(1)慣れを考慮する場合には利用回数と

利用間隔の両方を考慮する必要があること、(2)トークスイッチ操作の認知的負荷にも慣れの影響があり得ること、(3)生体信号計測は被験者の負荷状況の推測に利用できる可能性があること、などである。

今後は、事前訓練や知識と慣れの関係性を明らかにしたうえで、被験者の慣れの程度を統制して音声インタフェースの評価実験を進める予定である。

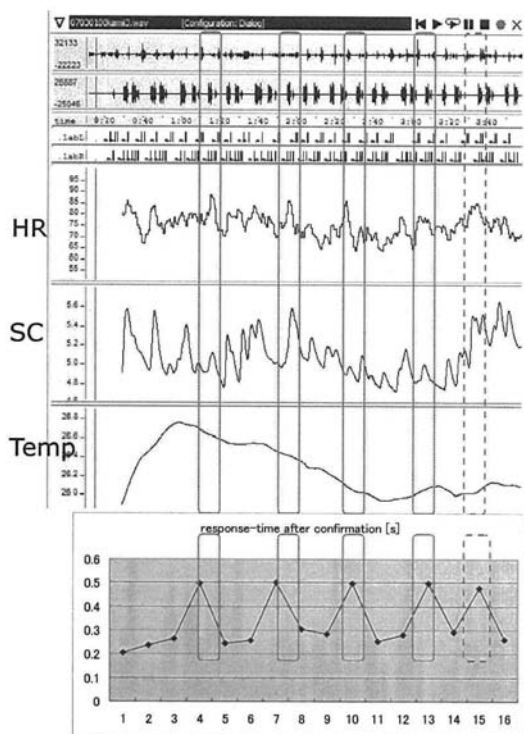


図9: 生体信号計測結果と応答時間の推移 (実線の囲みはシステム誤認識後の発話時、点線の囲みは読み方の難しい目的地名を発声する発話時を示す。)

謝辞

日頃熱心に討論いただく菊池研究室(言語情報科学研究室)の皆さんに感謝致します。

参考文献

- [1] 中川聖一他, "岩波講座 言語の科学 2 音声," 岩波書店, pp.179 (1998)
 [2] 石川泰他, "音声インタフェースの評価," 日本音響学会誌, vol.61, no.2, pp.79-84 (2005)

- [3] 西本卓也他, "音声認識応用に関する学会試行標準," 情報処理学会研究報告, 2005-SLP-55, pp.47-52 (2005)
 [4] 西本卓也他, "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, vol. J87-D-II, no. 2, pp.513-520 (2004)
 [5] 野田幸志他, "心的負荷状況における車載情報機器のための音声対話戦略の分析," 情報処理学会研究報告, 2006-SLP-64, pp.149-154 (2006)
 [6] 伊藤敏行他, "音声読み上げ聴取が運転に及ぼす影響評価," 日本人間工学会, シンポジウム「ケータイ・カーナビの利用性と人間工学」, pp.133-138 (2002)
 [7] 富永茂他, "音声入力カーナビゲーション操作に伴うドライバの作業負荷の検討," 自動車研究, vol.25, no.5, pp.21-24 (2003)
 [8] 北村康宏他, "ドライバーに影響を与える作業負荷の検討 -生理的・行動的变化から-, " ヒューマンインタフェース学会研究報告集, vol.6, no.1, pp.33-37 (2004)
 [9] 原直, 白勢彩子, 宮島千代美, 伊藤克亘, 武田一哉, "音声対話インタフェースの長期利用における学習効果の評価," 情報処理学会研究報告, 2005-SLP-55, pp.17-22 (2005)
 [10] 内田信行, "カーナビ使用における認知的側面の習熟過程:操作手順把握に伴う視認操作パターンの変化," 自動車研究, vol.26, no.9, pp.503-508 (2004)
 [11] C. A. Kamm, et. al, "From novice to expert: the effect of tutorials on user expertise with spoken dialogue systems," proc. of ICSLP '98, pp.1211-1214 (1998)