

音声蓄積サービスにおける音声インタフェースの評価

3S-8

角田 進 加藤 泰久

NTT情報通信網研究所

1 はじめに

現在、電話回線を通してボイスメールや情報提供サービスなどの音声蓄積サービスが提供されており、最近では機能の多様化によりPB（プッシュボタン）インタフェースでの操作がかなり複雑になっている。一方、音声認識技術の進歩により音声によるサービスの提供が可能となっており[1]、よりユーザフレンドリーなインタフェースを実現するために、ユーザからみた使いやすさという観点からの評価が行われるようになってきている[2][3]。さらに音声インタフェースは、利用者の経験、対話制御、認識性能などの様々な要因により使いやすさがかなり異なると考えられる。そこで今回、音声掲示板サービスを対象に利用者の経験、認識結果の確認方法、認識性能がインタフェースの使いやすさにどの程度影響するかを調べることを目的として主観評価実験を行なった。本稿ではその評価実験概要と結果について報告する。

2 音声掲示板サービス

今回評価の対象とする音声掲示板サービスの音声インタフェースによる典型的な操作手順の例を図1に示す。ログイン・ユーザ認証はPB入力で行い、その他のタスクにおいて音声入力を適用する。

3 主観評価実験システム概要

音声認識を適用したサービスの操作を行なう場合、認識率は被験者の発声方法などに大きく左右される。そこで本評価実験システムでは音声認識装置による発話・終話検出後、認識結果に関わらず設定したフローに従った次のガイダンスを出力し、フロー通りの動作を行う。認識結果を用いない疑似的対話システムなので誤認識、認識不能の割合をシステム側で設定できる。試作した実験システムによる操作フロ

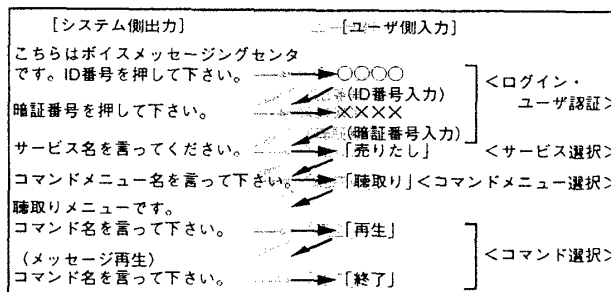


図1 音声掲示板サービス操作手順

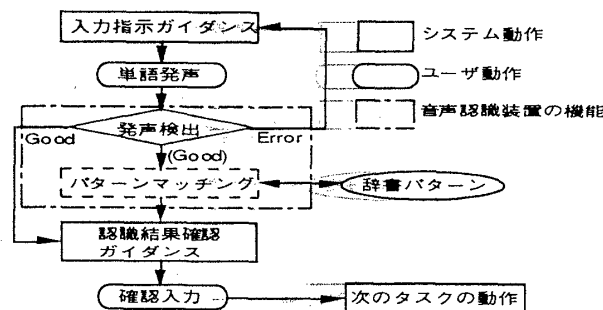


図2 実験システムの操作フロー

ーは図2の通りである。本実験システムにより被験者は予め設定した操作フローに従い操作を行う。

4 設定条件

評価するパラメータとして、認識結果の確認方法、誤認識・認識不能の回数を用いて、11種類の条件(1~11)を設定する(表1)。さらに条件12(表1)として従来のPBインタフェースを用いる。音声インタフェース(条件1~11)では離散単語発声、ガイダンス中の割込み入力は不可能とし、PBインタフェースでは認識結果の確認はないものとする。

- (1)確認方法：音声つまり「はい」「いいえ」を発声することにより確認する方法、またはPBで確認する方法を用いる。確認なしの条件では入力した時点で次のタスクの入力指示のガイダンスが出力される。
- (2)認識性能：誤認識は第1候補として間違えた単語を提示し、第2候補または第3候補の提示で認識が成功する。それに対して認識不能は候補単語の提示はせず再発声を促すガイダンスを出力し、被験者は同じ単語を発声する。表1中の認識失敗の回数は

1つの入力で起こる連続の認識失敗の回数を示しており、その割合は4回の入力に対して1回とする。また条件4, 8は通常音声で確認し、4回の入力に対して1回だけ複数候補を提示してPBで選択する。

表1 設定条件

| 条件 | 確認方法 | 認識性能 |
|----|-----------|----------------|
| 1 | 音声 | 誤認識2回 |
| 2 | 音声 | 認識不能2回 |
| 3 | PB | 誤認識2回 |
| 4 | 音声, PB | 3つ候補を提示してPBで確認 |
| 5 | 音声 | 誤認識1回 |
| 6 | 音声 | 認識不能1回 |
| 7 | PB | 誤認識1回 |
| 8 | 音声, PB | 2つ候補を提示してPBで確認 |
| 9 | 音声 | 誤認識なし |
| 10 | PB | 誤認識なし |
| 11 | 確認なし | 誤認識なし |
| 12 | PBインタフェース | |

5 主観評価実験及び結果

被験者はボイスメールサービスの操作経験者5人、未経験者5人とした。被験者には操作方法が記してあるマニュアルと実験指導書を与えた。すべての被験者が指導書の操作フロー通りに「センタへ接続、掲示板選択、メッセージ再生、終了」の操作を条件1~12の順に行ない、操作が終了する毎に使いやすさに関して7段階評価を行なった。それぞれの評点を選択された頻度で荷重平均をとったものをMOS^{注1}値(平均オピニオン評点)とし、使いやすさのパラメータとする。

実験で得られた各条件とMOS値の関係を図3, 4に示す。図3から経験者がPBインタフェースより使いやすいと評価したのは誤認識がなく確認もない条件だけである。つまり確認がある場合には音声入力は有効とはならない。経験者はPBの操作に慣れているため、PBのガイダンス中の割込み入力により操作性が向上できることがPBインタフェースの評価が高い理由と考えられ、音声インタフェースにおいても割込み入力の可否が使いやすさを決める重要な要因であることが推測される。

それに対して図4から初心者は誤認識・認識不能がある条件でも音声インタフェースの方を使いやすいと評価している。また誤認識の回数が同じであれ

ば音声で確認する条件がPBで確認する条件より評価が高くなっている。このことから最終的に認識が成功すれば音声入力が有効であり、確認を含めて音声で操作するインタフェースが適していると考えられる。初心者にはサービス手順やコマンドとPBの対応付けを同時に覚えることがかなり負担になるために、認識失敗がある条件でも音声インタフェースの評価が高くなっていると考えられる。

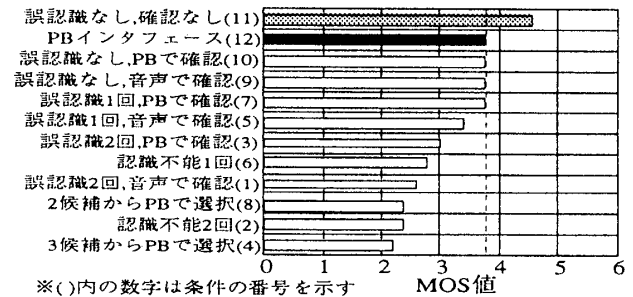


図3 各条件とMOS値の関係(経験者)

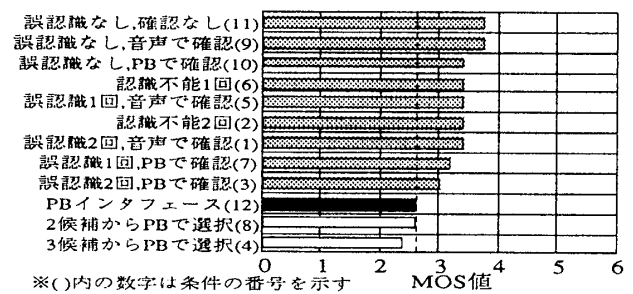


図4 各条件とMOS値の関係(初心者)

6 おわりに

音声掲示板サービスを対象として音声インタフェースの使いやすさを主観評価実験により調べた。離散単語発声、ガイダンス中の割込み入力できない条件において、1)経験者にとっては割込み入力が可能なPBインタフェースの方が使いやすく、2)初心者にとっては認識率が悪くても、掲示板名やコマンド名が分かれば操作できる音声インタフェースが使いやすい、という評価が得られた。今後は今回の評価実験で拘束を与えたガイダンス中の音声入力や付帯語の付加が許容できる条件での評価、さらに複雑な操作についての評価をする予定である。

- 参考文献 [1]菅村昇, 電話網における音声認識の応用と課題, 日本音響学会講演論文集, p. 519, 平成5年3月
 [2] David Fay, "Touch Tone or Automatic Speech Recognition: Which Do People Prefer", p. 339, Human Factors in Telecommunications, 1993
 [3] 新居康彦他, 繰り返し発声を許す簡単な音声インタフェースの利用者満足度モデルの検討, 日本音響学会講演論文集, p. 31, 平成5年10月

^{注1} Mean Opinion Score