音声認識における事前教示・訓練の影響*

○網田泰裕，大橋浩輝，宮澤幸希，山田哲史，菊池英明（早大・人科）

1　はじめに

音声インタフェースの実用化に向けて使用性(ユーザビリティ)を正しく評価することが重要であり，著者らはこれまでに音声インタフェース評価における習熟の影響について分析を行った[1]．本稿では，ユーザへの事前教示や訓練が音声認識に与える影響に関して行った調査について報告する．

音声認識技術の性能は年々向上しているが，音声認識について特別な知識を持たないユーザは，現状の音声認識技術にとって対応困難な振舞いをとることが多い．このことは音声認識技術を用いた音声インタフェースの評価においても考慮しなければならず，事前に教示や訓練を行うことによって，音声認識技術の使用方法についての事前知識をどの程度統制できるかを知っておく必要がある．そこで我々は，事前の教示や訓練によってユーザ発話の発話速度，明瞭度，音声認識の成功率がどのように変化するかを実験により観察した．

2　実験

2.1 被験者

被験者は日本語を母語とする男女 20 代 18 名と 40 代 18 名の計 36 名とし，教示・訓練の内容が異なる群を 6 つ設定し，それぞれ 20 代 3 名 40 代 3 名とした．群ごとの教示・訓練方法の内訳をTable 1に示す（訓練(1),(2)の内容については後述）．なお，本稿で述べる分析の範囲では，一部データの不備からこのうち 26 名分を対象にして分析を行っている．

Table 1 音声認識実験群

群	教示	訓練
1	○	(1)
2	×	(1)
3	○	(2)
4	×	(2)
5	○	×
6	×	×

2.2 手順

被験者には，教示や訓練の前後に，事前に選定した地名･施設名 40 語の発声を音声認識向けに行ってもらい，その音声を収録した．なお，その際フィードバックは行わない．

例えば 1 群の被験者は，始めに初期状態での音声収録を行い，次に事前教示を見せて，再度音声収録を行う．さらに訓練を行った後，再び音声収録を行った．

なお，全被験者とも実験の約 1 週間後にあらためて音声収録を行った．

教示内容は音声認識をする上で注意すべき基本的なポイントとデモンストレーションである．主なポイントは，声の大きさ，発話速度である．また，音声認識における良い例，悪い例のデモンストレーションを 2 単語について行った．なお，教示内容を一定にするために，教示ムービーを事前に収録し，用いた．

訓練(1)では，測定時に使用する 40 語とは異なる，ランダムに選んだ地名・施設名 20 語を読み上げてもらい，その際，被験者に認識結果，音声のパワー，発話速度の値を提示する．また，訓練(2)では，訓練(1)に加えて，被験者からの質問に対して，実験者がアドバイスを行った．

2.3 分析

収録音声に対し，音声認識率，発話速度，明瞭度を求める．音声認識率の測定には，Julian[2]を用いた．音響モデルは Julian に付属の性別非依存 PTM・triphone モデルを利用した．受理可能単語数は約 500 語で，孤立単語認識のみとした．発話速度は毎秒モーラ数とする．また，明瞭度の 1 つの指標として 5 母音間のマハラノビス距離[3]を求める．マハラノビス距離は文献[3]の手順通りに算出する．文献[3]においては 31 種の音素を対象にマハラノビス距離を算出しているが，本実験ではモデル学習量が不十分であるため，5 母音を対象にマハラノビス距離を算出する．なお，

* Effects of prior teaching and training in speech recognition, by AMITA, Yasuhiro, OHASHI, Hiroki, MIYAZAWA, Kouki, YAMADA, Tetsushi and KIKUCHI, Hideaki (Waseda University).

本研究では 25 次元の音響パラメータを採用する．

2.4 結果

Fig. 1に各群の平均音声認識率，Fig. 2に 2 群の発話速度の変化，Fig. 3に 3 群の明瞭度の変化，Fig. 4に 3 群の音声認識率と明瞭度の散布図を示す．

音声認識率に注目すると，Fig.1 の 3 群と 4 群の 3 回目に上昇していることから訓練(2)の効果は確認できた一方，訓練(1)の効果は確認できなかった．

また，2 群では，教示後から訓練後にかけて毎秒モーラ数が収束している．3 群では，教示後から訓練後にかけてマハラノビス距離が収束している．さらに，3 群の音声認識率とマハラノビス距離の間に強い正の相関が観察された．

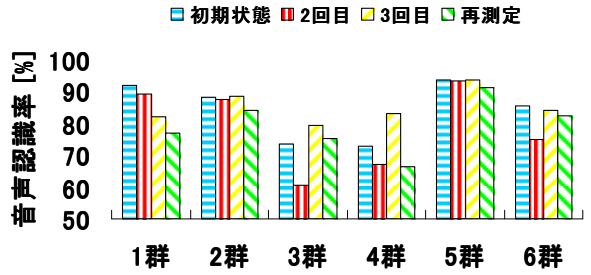

Fig. 1 各群平均音声認識率

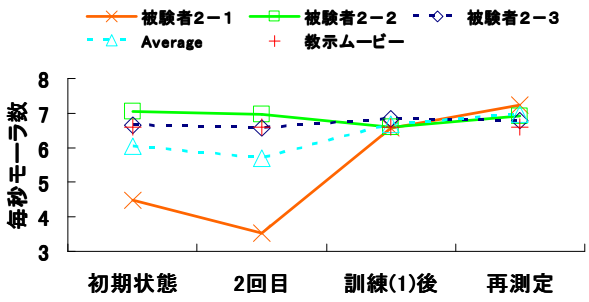

Fig. 2 2 群の発話速度の変化

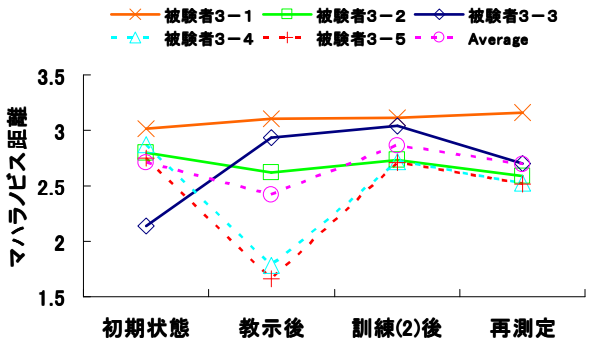

Fig. 3 3 群の明瞭度の変化

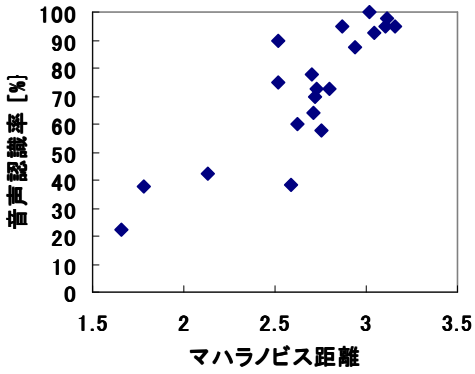

Fig. 4 3 群の音声認識率と明瞭度の関係

3 考察

事前教示の影響は小さいが，訓練の影響は大きいことがわかった．訓練後には発話速度が収束し，ある一定の値に収束しているとも考えられる．しかし，収束している値は群ごとに異なっているので，個人差を考慮する必要があるだろう．また，音声認識率に注目すると，再測定の結果は 3 回目と比べて悪くなっているので，訓練の効果が落ちていると考えられる．訓練によって習得した音声認識技術の利用方法の定着率については検討する必要がある．

4 まとめ

音声認識において，音響的要因に着目し，事前教示や訓練の影響を調査した結果，音声認識率向上にはフィードバックのある訓練が有効だとわかった．また，訓練による発話速度と明瞭度の一定の変化も確認できた．これらの結果を踏まえて，教示・訓練方法と音声認識率の関係を今後より詳細にとらえる必要があると考える．

謝辞

本研究は，経済産業省，平成 19 年度戦略的技術開発委託費「音声認識基盤技術の開発」の一部として実施されたものである．

参考文献

[1] 菊池 他, “音声インタフェース評価における慣れの影響の分析,” 情処研資 SLP-67, pp.97-102, 2007.

[2] 汎用大語彙連続音声認識エンジン Julius/Julian, http://julius.sourceforge.jp

[3] 中村 他, “話し言葉音声の音響的・言語的特徴の分析,” 信学技報, SP-106-78, pp.19-24, 2006.