HMM 音声認識ソフトを構音障害評価に使用する可能性

ニュー天草病院リハビリテーション科
言語聴覚士　山口信
理学療法士　新井淑子・高橋慎太郎・福田裕二
・片岡千奈美・竹森浩
MD　古賀由紀

2002.7.1 作成

1.はじめに

構音障害の評価は大別して発声発語器官の検査と構音の聴覚印象による評価の二つがある。前者は一定の客観性が期待できるが、後者は専門教育を受けた言語聴覚士（以下 ST）の耳に頼っているのが現状である。特にスピーチ全体に対する評価は、伝統的に 5 段階の発話明瞭度により主観的になされてきた。発話明瞭度は次のようなものである。1 ° よくわかる、2 ° ときどきわからないことばがある、3 ° 話題を知っていれば見当がつく、4 ° ときどきわかることばがある、5 ° まったくわからない 1)。

この研究では HMM 音声認識ソフトの認識率を各人のスピーチの客観的評価に使用しうる可能性について考察した。

2.HMM 音声認識方式の特徴

得られた特徴パラメータ時系列を用いてパターン認識を行う。近年は以下のような利点のため、隠れマルコフモデル（HMM）による音声認識が主流になりつつある。

- スペクトル時系列の統計的変動をモデルのパラメータに反映できる。
- 比較的簡単なモデルのパラメータ推定法が存在する。
- 確率モデルでは二つのモデルを結合する際、モデル間の結合も確率で表すことができるため、滑らかに結合した新たなモデルを得ることができる。
- 認識時の計算量は比較的少ない。

一方、統計的計算が基本となるため、次に挙げるような欠点も存在する。

- モデルのパラメータを決定するための学習処理が、やや複雑で計算量が多い。
- 不特定話者で音素単位の認識を行うためには、学習に音素のバランスの取れた大量の音声を必要とする 2)。

特に、最後の項に挙げた欠点が本研究のような用途に使用する場合のネックになると思われる。

3.研究の方法

① 当院の患者・家族・職員から構音障害の有無を問わず便宜抽出した 50 人の被験者に同一の文章を読んでもらい、そのスピーチを音声認識ソフトに認識させる。

② 認識率について、構音障害の有無、性別、年齢、脳血管障害の既往の有無などから比較する（サンプルの採取方法が便宜抽出でサンプル数も少ないため統計的に厳密なものではない）。

③ 認識率については認識された文章を音素に直し、音読に用いた文章と比較し、誤認識された音素の数から判定する。
（全体の音素数－誤認識された音素数／全体の音素数）×100（%）＝認識率

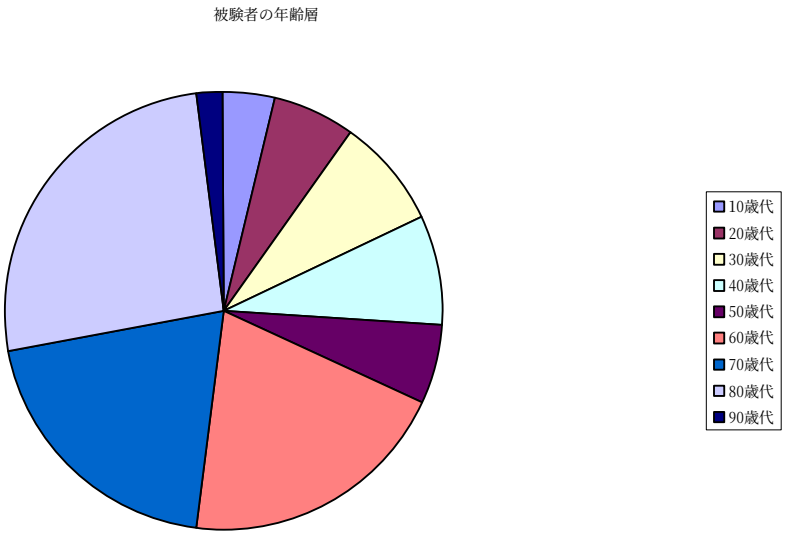
④ 構音障害の有無については理学療法士（以下 PT）５名に匿名のスピーチを聞いて判断してもらい、3 名以上が構音障害ありと判断したものを構音障害群とする。

⑤ 研究材料：音声認識ソフト‐IBM Via Voice version9。
録音ソフト‐Microsoft IME2000。
録音マイク‐Via Voice 付属のマイクロフォン。

⑥ HMM 音声認識ソフトの起動には特定人物の声によるエンロールが必要だが、これには検者である 40 歳男子（脳血管障害の既往なし、構音障害を含むスピーチの障害なし）の声を使用し、エンロールは最低限に止めた。

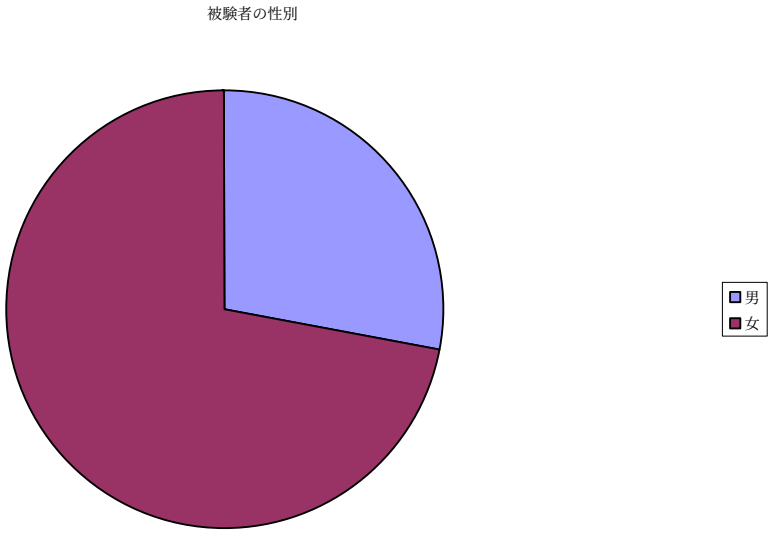
4.被験者の年齢・性別・脳血管障害の既往・構音障害の有無

① 被験者の年齢（グラフ 1）
被験者の年齢は 10 歳代 2 人、20 歳代 3 人、30 歳代 4 人、40 歳代 4 人、50 歳代 3 人、60 歳代 10 人、70 歳代 10 人、80 歳代 13 人、90 歳代 1 人であった。



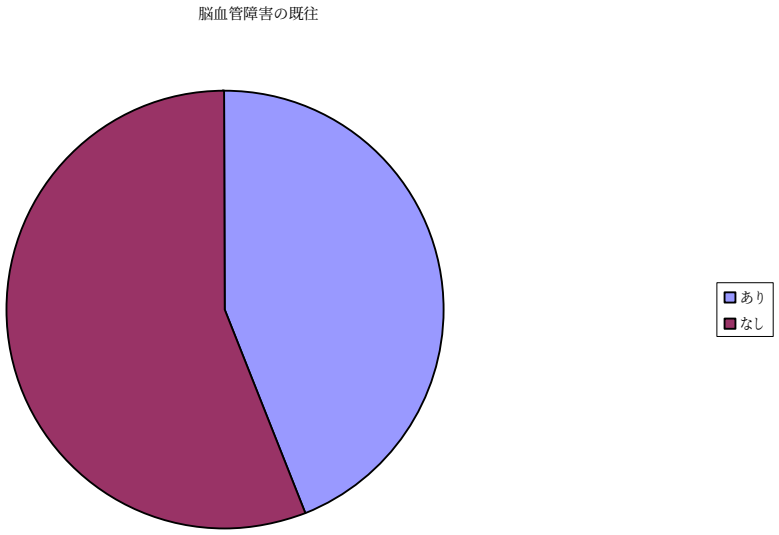
② 被験者の性別（グラフ 2）

被験者の性別は、男性 14 人、女性 36 人であった。



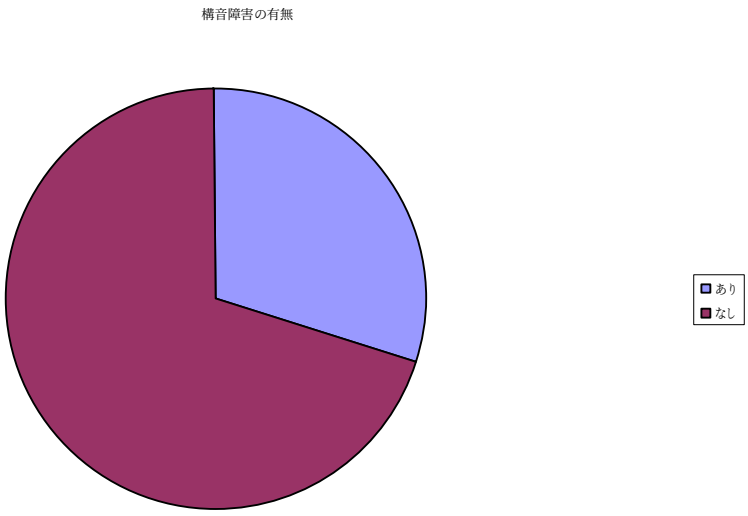
③ 脳血管障害の既往（グラフ 3）

脳血管障害の既往は、有りが 22 人、無しが 28 人であった。



④ 構音障害の有無（グラフ 4）

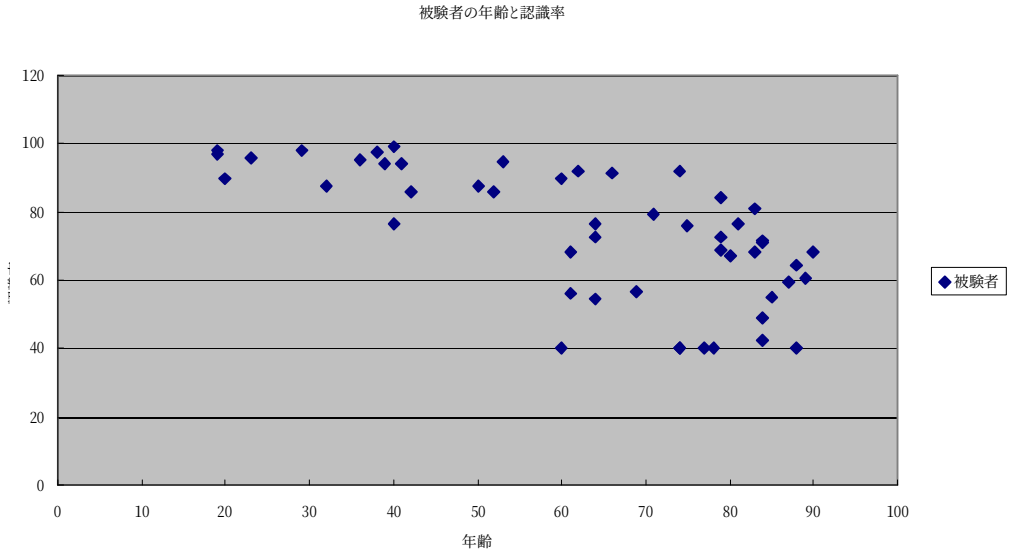
構音障害の有無は、有りが 15 人、無しが 35 人であった。



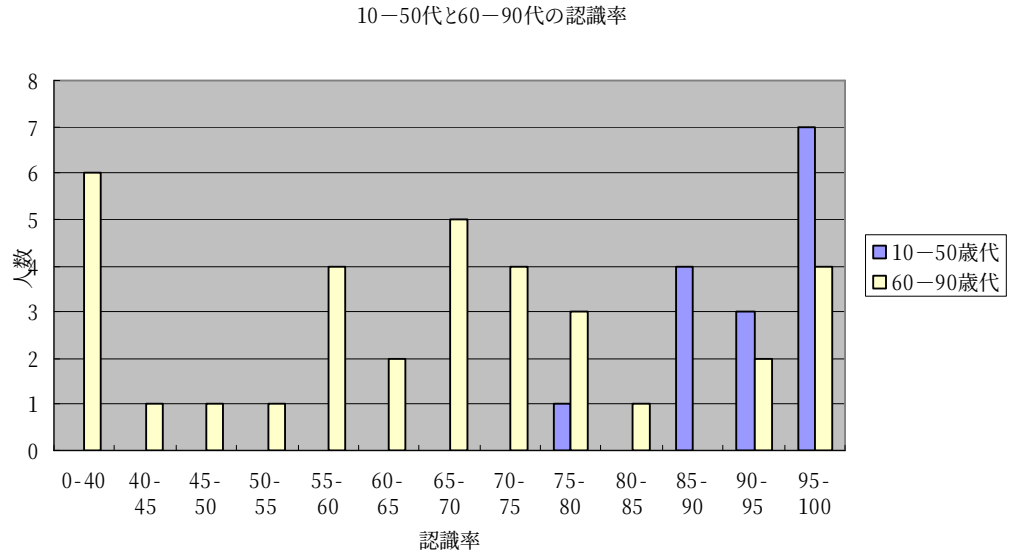
5.結果

① 年齢と認識率

10～50 歳代では最高 99.2%、最低 86%、平均 92.3%、標準偏差 5.95、分散 35.5 と、比較的高くかつバラツキの少ない認識率を示した。一方 60～90 歳代では最高 92.2%、最低では判定できた範囲で 42.6､%、判定不能（40%より低い可能性が高い）も 6 人で、判定不能を除いた平均 70.2%、標準偏差 12.9、分散 165.6 と、認識率に大きなバラツキがあった（グラフ 5）。年齢と認識率の相関係数は-0.72 であった。

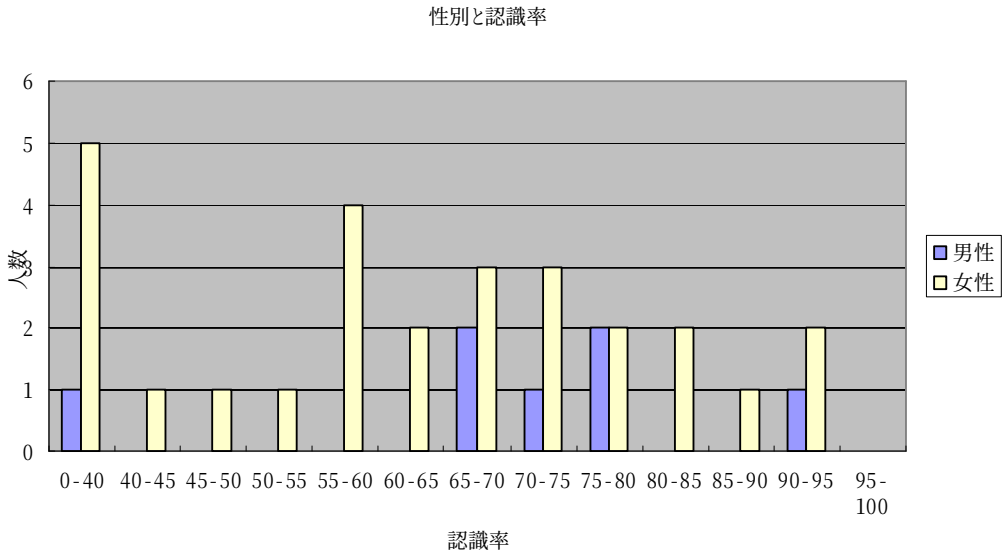


被験者を 10～50 歳代と 60～90 歳代の 2 群に分けてヒストグラムにするとこの傾向はよりはっきりする（グラフ 6）。2 群間の平均差は 12.8（t=2.6E-9）であった。



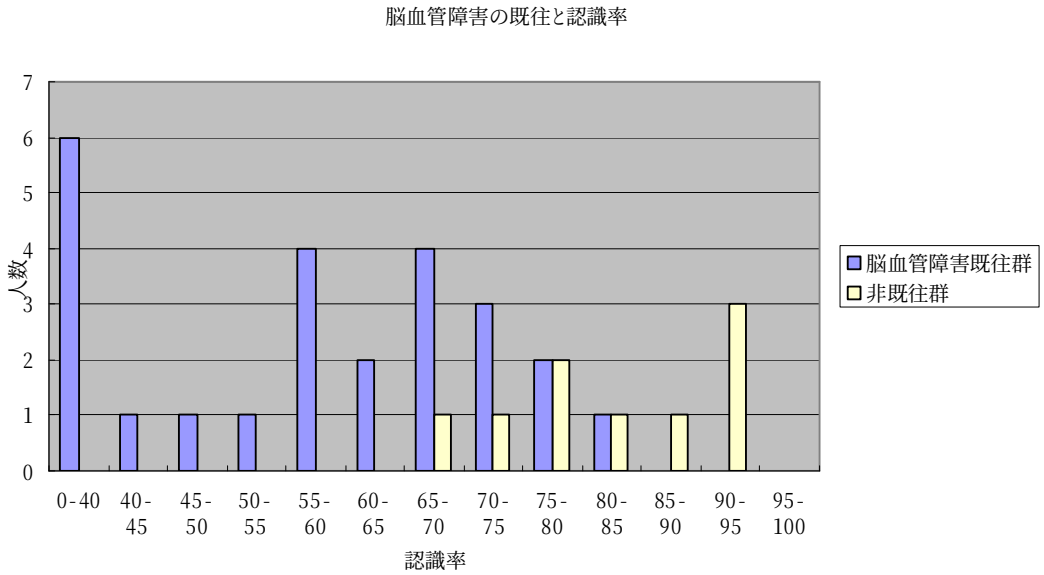
②性別と認識率

男性にやや認識率の高い傾向があった。男女の平均差は 0.4(t=0.01)。60-90 歳代では平均差 3.5(t=0.2)と、認識率に有意差は見られなかった（グラフ 7）。



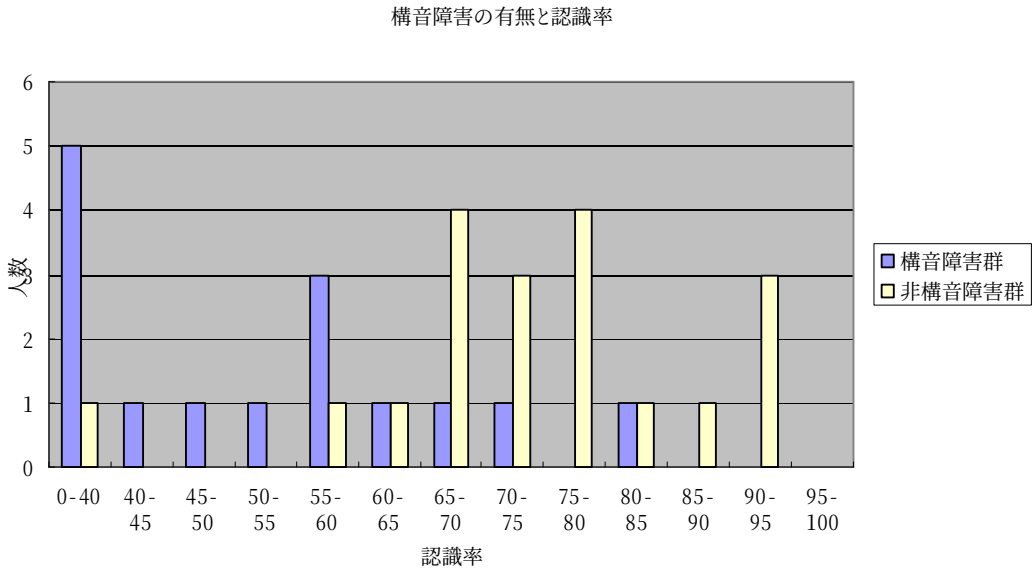
③ 脳血管障害の既往と認識率

脳血管障害の既往がある被験者の方がない被験者より相対的に認識率が低い傾向があった。2 群の平均差は 23.4(t=6.2E-9)。全被験者でより顕著だったが、60〜90 歳代でもこの傾向は変わらなかった（グラフ 8)。2 群の平均差は 18.0(t=0.0002)。特に判定不能なほど認識率の低い被験者は全員脳血管障害の既往群に含まれていた。



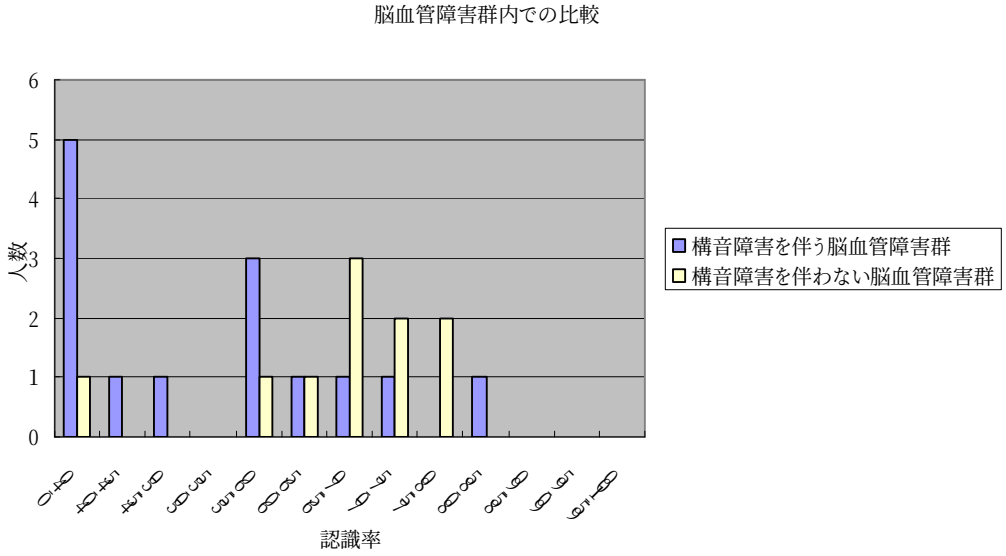
④ 構音障害と認識率

構音障害があると判定された被験者のほうがなしと判定された被験者より認識率が低い傾向にあった。2 群の平均差は 27.4(t=3.6E-5)。全被験者ではより顕著だったが、60～90 歳代でもこの傾向は変わらなかった（グラフ 9）。2 群の平均差は 16.8(t=0.03)。



⑤脳血管障害と構音障害

構音障害を有する脳血管障害既往群はそうでない脳血管障害既往群に比べて認識率が低下した可能性がある。平均差は 8.3(t=0.048)。



6.考察

① 被験者の年齢による認識率には、1.年齢とともに低下する傾向がある、2.年齢が大きくなるほどバラツキがあるという傾向があった。その要因としては 60〜90 歳代では脳血管障害の既往が多い、構音障害を有する被験者が多い、発話明瞭度に影響がなくても歯牙欠損、弓状声帯、発声発語器官の老人性変化などによりボイス・スピーチに軽微な障害を有するものが多い、などが考えられる。したがって、この年齢層に限って、これらの要因について認識率との関係を考える必要がある。

② 性別については男性に認識率の高い傾向が出たが、これは男性の年齢層が若く、したがって脳血管障害の既往、構音障害ともにない被験者が多かったことが多大に影響している。実際に 60〜90 歳代について男女を比較してみると必ずしも男性が高いとはいえない。いずれにしてもこの年齢層の男性が少なすぎることは、今回のサンプル採取の大きな問題点である。

③ 脳血管障害の既往については 60〜90 歳代の被験者に限っても既往有りの被験者群では認識率が相対的に低い傾向がある。これは脳血管障害に伴って多かれ少なかれ発声発語器官の運動制限が発生することが関係していると思われる。判定不能なほど認識率が低かった 6 名は全員脳血管障害の既往群であり、５名は ST の判定では明瞭度 3°‐話題を知っていれば見当がつく‐以下の明らかな構音障害で、PT 全員が「構音障害あり」と判定した。残り 1 名の被験者は PT1 名のみが「構音障害あり」と判断したが、ST の聴覚的印象では、声が小さいほかは少なくとも重大なスピーチの障害はなかった（発話明瞭度 1:よくわかる）。

④ 構音障害の有無に関しては 60〜90 歳代の被験者に限っても PT5 名中 3 名以上が「あり」と判定した被験者群では認識率が相対的に低い傾向があった。しかし、4 名の PT から構音障害と判定され、かつ実際に言語聴覚療法を施行中の患者（発話明瞭度は 2°‐ときどきわからないことばがある‐）であっても、80%代の認識率を示した被験者も存在する。これは PT に依頼するときに重症度については一切言及しなかったこと、同一院内の患者であるためにスピーチから患者が特定できた場合にはある程度のバイアスがかかったことなどが原因と考えられる。

⑤ 脳血管障害の既往がある被験者群内の比較では、「構音障害あり」と判定された被験者群の方の認識率が低下した可能性がある。発話明瞭度は 1〜4 とさまざまであり、必ずしも全員が重症というわけではない。単なる「CVA スピーチ」と「構音障害（dysarthria）」の間には何らかの質的変化があるのかもしれない。

7. 結論

結論から言えば、HMM 音声認識ソフトを構音障害評価に利用することは、まったく不適当であるわけではないが、現状の技術では時期尚早である。それは、1, 全体的な認識率そのものが低すぎて、人間の聴覚的印象と大きな乖離がある。2，その結果、発話明瞭度でいえば 3 以下の話者では認識率が低すぎ、認識された文章の認識率の判定が不可能になってしまう。3，スピーチに問題がある被験者ではそれが軽度のものであっても認識率の判定が極めて煩雑かつ専門的な知識を必要とする。4，設定段階で特定話者のエンロールを必要とするため、エンロールを行った話者のボイス・スピーチの特徴によっては偏りのあるデータが集積する可能性がある。などの理由からである。

ただ、1，認識率はエンロールした話者の年齢・性別よりボイス・スピーチの特徴との相関が強い可能性が高い。2，スピーチに問題があると判定された被験者では問題なしと判定された被験者より認識率が低い。3，設定次第では最初のエンロール以外には人間の耳でいう「慣れ」に当たるものがないため、「慣れ」を発話明瞭度の向上と錯覚する可能性が薄い。4，現在不特定話者を対象とした音声認識システムが開発されつつある。などの理由で、今後スピーチの評価に利用できるソフト、あるいはスピーチの評価に特化したソフトが登場する可能性も十分に考えられる。今後とも音声認識システムの技術の進歩に注目していきたい。

8.参考文献

1) 医療研修推進財団監修『言語聴覚士指定講習会テキスト』医歯薬出版 P210
2) http//ips9.main.eng.hokudai.ac.jp/research/hata/recognition.html