社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 話者認識技術を応用した知覚的年齢分布の自動推定

山内　景太†　　峯松　信明††　　広瀬　啓吉†††

† 東京大学工学部
†† 東京大学大学院 情報理工学系研究科
††† 東京大学大学院 新領域創成科学研究科
〒 113-0033 東京都文京区本郷 7-3-1
E-mail: †{kta-yama,mine,hirose}@gavo.t.u-tokyo.ac.jp

**あらまし**　本研究ではまず，広い年代の話者を含む音声データベースに対する聴取実験によって，各話者の知覚的年齢を推定した．次にデータベースの各話者を混合ガウス分布モデルを用いてモデル化した．未知入力話者の知覚的年齢を，各話者モデルに対する尤度を重みとして用い，聴取実験を通して定義された各話者の知覚的年齢の期待値操作によって推定することを試みた．ここで，各話者の知覚的年齢は，聴取者間で平均値をとることにより一つの値として表現することも可能であるが，聴取実験によりその分布が得られている．そこで各話者の知覚的年齢を，ラベルとして与える場合と，分布として与える場合について検討した．実験の結果，いずれの方法でも知覚的年齢と自動推定年齢間の相関値は約 0.9 となったが，分布として与えた場合により高い精度で推定される様子が観測された．
**キーワード**　知覚的年齢，聴取実験，年齢分布，話者認識，期待値操作，知覚的インターフェイス

# Estimation of perceptual age distributions using speaker recognition techniques

Keita YAMAUCHI†, Nobuaki MINEMATSU††, and Keikichi HIROSE†††

† Faculty of Engineering, University of Tokyo
†† Graduate School of Information Science and Technology, University of Tokyo
††† Graduate School of Frontier Sciences, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan
E-mail: †{kta-yama,mine,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract**　This paper proposes a new technique to estimate perceptual age of a speaker by using speaker recognition techniques. Firstly, listening experiments were carried out to define perceptual age of speakers in a database, who covered a wide range of real age. Next, every speaker in the database was modeled as GMM. Estimation of the perceptual age of an unknown speaker was done by expecting the perceptual age of all the speakers in the database, where likelihood scores calculated by matching the GMMs with the input speaker were used as weights. Perceptual age of every speaker in the database can be defined either as a value of age or a distribution of age. In this work, both definitions were compared experimentally. Correlation between the perceptual age defined by listening and that estimated by machines were approximately 0.9 in both cases. But some benefits were found in the experiments when the perceptual age was modeled as a distribution.
**Key words**　perceptual age, listening experiment, age distributions, speaker recognition, expectation, PUI

## 1. はじめに

近年，計算機技術の飛躍的発展により，音声情報処理技術を用いたシステム開発が広く行われるようになってきた．その多くは，人間による音声入力を認識技術を用いて「音声文字変換」し，文字化された情報のみに基づいてシステム動作を制御している．しかし，我々が音声から得ている情報は言語情報のみではなく，文字言語には陽に含まれない話し手の意図や態度といったパラ言語情報，更には，個人性，性別，年齢，感情などの非言語情報まで含まれる．特に，個人性，性別，年齢などの情報は話し手が意図的に音声を通して伝搬している情報ではないが，聞き手はこれらの「意図されていない」情報に敏感に察知し，それに基づいてマナーや発話スタイルを調整する「気の利いた」対応を可能にしている．特に日本のような単一民族国家では「阿吽の呼吸」という言葉で表現される，その場の雰囲気を察知し，適切な判断を下す能力のある無しが「気の利いた」相手であるかどうかの判断基準となっている．

これらを考慮すると，今後，より高度なインタラクションを

実現しようとする場合，文字情報のみでなく，相手の状態を知覚するインターフェイス (PUI: Perceptual User Interface) が不可欠であろう．そこで本研究では，PUI が持つべき機能の一つとして，話し手が無意識的に発している情報である「年齢」の情報を音声から抽出する技術構築について検討する．年齢という場合，話し手の生物学的年齢と，話し手の音声を聞いた時に聞き手が感じる知覚的年齢の二つが考えられるが，本研究では後者を対象とする．これは，聞き手が年齢の情報を元に行動を律する場合，常に知覚的年齢に基づいているからである．

本研究ではまず，子供から高齢者までの音声データベース (以下，DB と略す) 内の話者に対して，30 名の被験者による聴取実験を通して知覚的年齢を推定する．と同時に，DB の全話者を GMM を用いてモデル化する．その後，知覚的年齢を，各 DB 話者との GMM スコアを重みとして用い，DB 話者毎に定義されている知覚的年齢の期待値操作を行なうことで推定することを試みる．ここで，知覚的年齢の与え方として 2 種類の方法を検討する．即ち，各 DB 話者に対して知覚的年齢を 1) ラベルとして与えた場合と，2) 分布として与えた場合を検討する．

## 2. 聴取実験による知覚的年齢の推定

### 2.1 使用した音声データと被験者

被験者は 20 歳前後の成人 30 名である．知覚的年齢推定に使用した音声データは男性話者のみで，JNAS-DB(20〜60 歳)153 名 [1]，S-JNAS-DB(60〜90 歳)202 名 [2]，子供音声 DB(6〜12 歳) 141 名 [3] の合計 496 名である．聴取実験は USB ヘッドホンを使用し，Web 上で行った．高齢者の特徴に発声音量が小さいことが考えられるが，システム実装時にマイク距離によって推定年齢が変化してしまうのは好ましくないため，音量等化処理を施した後，提示音声として使用した．

### 2.2 実 験 手 順

被験者は各 DB 話者につき 1 文の音声を聴取し，何歳に感じるかを 0〜110 歳，1 歳刻みの中から選び，さらに雑音レベルを高，中，低から選ぶようにした．音声はランダムに並べた．なお，推定知覚年齢が文章の内容に引きずられる (特に子供音声は童話を読み上げている) ことが考えられるので，発話内容に依存した年齢推定をしないよう指示し，また，同一発声者の文音声の内容についても被験者間で異なるようにした．更に，方言話者 (主に関西) が多く混ざっているため，方言であることを理由にした年齢推定も避けるよう教示した．本実験に先立って，聴取者本人の知覚的年齢イメージを明確にするため，各年代の音声を実年齢を伝えずに計 7 文聴取させた．

### 2.3 実験結果と考察

全被験者 (30 人) の知覚的年齢の平均と標準偏差を，横軸に各 DB 話者 (計 496 名) をとってプロットしたものを図 1，図 2 に示す．これを見ると，知覚的年齢が 0 歳から 80 歳までの間に広く分布しているが，中にはデータ量の少ない年齢も観測されている．また，子供データのばらつきはたかだか ±2 歳程度だが，成人以上のデータは ±4〜15 歳程度という大きな幅を持っていることがわかった．しかし，知覚的年齢に比例してばらつきが大きくなると考えるとこれは妥当な結果であるといえる．

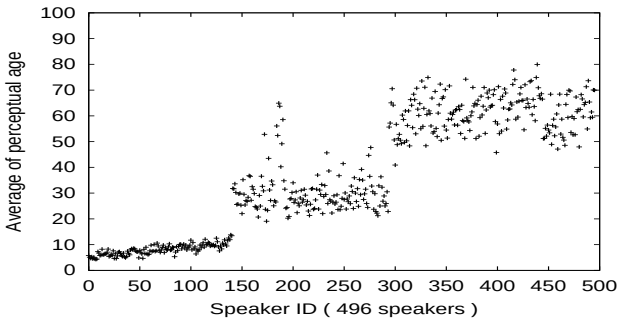


図 1 各 DB 話者ごとの推定年齢の平均

Fig. 1 Averaged perceptual age of DB speakers over the listeners.

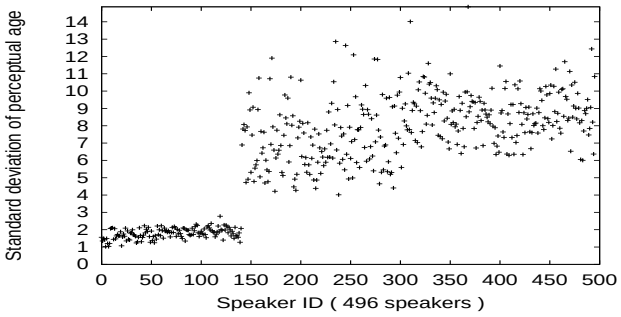


図 2 各 DB 話者ごとの推定年齢の標準偏差

Fig. 2 Standard deviation of perceptual age over the listeners.

## 3. 知覚的年齢ラベルを用いた未知入力話者の知覚的年齢推定

スペクトル情報に高齢者らしさがあらわれることから，知覚的高齢者と非高齢者の識別を話者認識技術を用いて 91 %程度の精度で行うことができる [4]．また，一般に高齢者の音声認識率は下がるが，それに対して高齢者データでモデルを再学習/適応することで認識率が改善される [5]．また，子供の音声認識率も音響的特徴が成人と大きく異なるために非常に悪いが，同様に子供の音響的性質ごとにモデルを再学習/適応することで認識率が改善される [6]．これより，本研究においても話者認識技術を採用することとし，それによって聴取実験被験者 30 名それぞれの知覚的年齢の推定の様子を再現することを試みる．

### 3.1 推 定 方 法

推定方法は次のようである．1. 各 DB 話者の GMM を作成．2.(未知) 入力音声と GMM との尤度スコアを計算．3. 各 DB 話者の知覚的年齢ラベルと尤度スコアの期待値操作によって知覚的年齢を計算．以下で計算法などについて詳しく説明する．

**実験条件** 実験条件を表 1 に示した．2.2 節において主観的雑音レベルが高いと判定された 89 名を DB 話者から除き，その結果残った 407 名を使用した．除かれた DB 話者の内訳は，子供音声 DB59 名，JNAS-DB18 名，S-JNAS-DB12 名である．ここで子供音声が多く除かれたのは子供音声の収録が非常に難しいためである [6]．また，無音部分には話者情報は含まれていないので，音声は無音部分を除いて使用した．GMM は混合数 16 とし，学習データは各 DB 話者につき 60 秒とした [7]．

**尤度スコアの計算** 入力音声の GMM に対する対数尤度を，フレーム長で正規化し，尤度に戻した値を重みとして使用した．

表 1　実 験 条 件
Table 1　Conditions of experiments.

| | |
|---|---|
| 学習データ (各 60s) | 子供 (82 名)，JNAS(135 名)，S-JNAS(190 名) |
| 評価データ (各 5s) | 同上 |
| サンプリング周波数 | 16kHz |
| フレーム周期, 分析窓 | 10ms, ハミング窓 25ms |
| プリエンファシス | $1-0.97z^{-1}$ |
| 特徴パラメータ | $12MFCC+12\Delta MFCC+\Delta$パワー |
| 特徴パラメータ抽出の前処理 | スペクトル等化処理<br>無音区間除去 |
| GMM | 対角分散共分散行列・混合数 16 |

**期待値操作による知覚的年齢の計算**　本研究では，以下のような計算式で知覚的年齢 (PA) 推定を検討している．

$$PA=\frac{\sum P(x|o)\times x}{\sum P(x|o)} \tag{1}$$

ここで $o$ は音響観測量，$x$ が (聴取実験より定義される) 知覚的年齢である．即ち $o$ に対する事後確率を用いた期待値操作で知覚的年齢を推定する．ここで $P(x|o)$ はベイズ則により

$$P(x|o)=\frac{P(o|x)P(x)}{P(o)} \tag{2}$$

と変形され，$P(o)$ を定数項と考えれば期待値操作の重みは $P(o|x)P(x)$ となる．$P(x)$ は知覚的年齢に対する事前確率であり，例えば，最終的に実装されたシステム利用者の知覚的年齢分布がそれに相当するが，現時点でそれを仮定することは不可能であり，本研究では $P(x)$ として一様な分布をもつ確率密度関数を考える．以上の結果，$P(o|x)$ を重みとした期待値操作へと帰着される．但し本研究で利用する DB 話者の知覚的年齢分布は明らかに偏りがあり，それを是正する処理を必要に応じて導入することになる．例えば，図 1 が示すように各知覚年齢のデータ数に差があるため，尤度スコアを重みにして (話者単位で) そのまま期待値を求めると，データ数の多い年齢に重みをかけた操作となってしまう (つまり $P(x)$ が一様ではない)．そこで DB 話者を年齢毎に分類し，各年齢毎に尤度スコア ($P(o|x)$) を算出することとした．ここでは話者モデル尤度を用いて $P(o|x)$ を以下の 2 通りで近似した．DB 話者中，知覚的年齢が $x$ 歳である $i$ 番目の話者のモデルを $M_x(i)$ とする．

$$P(o|x) = \overline{P(o|M_x(i))} \tag{3}$$

$$P(o|x) = \max_i P(o|M_x(i)) \tag{4}$$

上式のように定義された $x$ 歳としての尤度スコアを用いた期待値操作によって知覚的年齢を推定する．しかし，年齢幅は有限であるために，最終的に得られる期待値の (年齢幅中の) 位置によっては，期待値より低い年代幅，期待値より高い年代幅に差が生じ，これが原因となって推定値が偏りを持つことがある．これを解消するために $P(o|x)$ の上位 $N$ 年齢について期待値操作を行ない，最終的な知覚的年齢とした．もし尤度スコアが完全に年齢のみに起因する場合は $N$=1 の時，即ち最大尤度スコアの知覚的年齢ラベルを推定結果とすればよい．しかし，話者認識技術を用いた年齢推定を検討しているため $N$>1 として期待値操作を施している．$N$ は実験的に随時決定した．

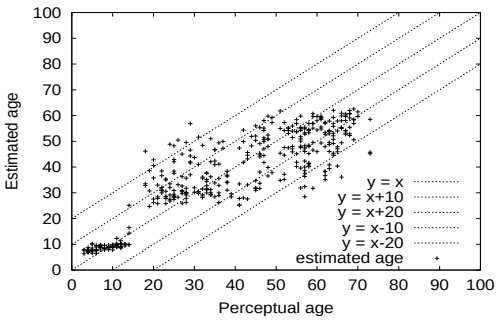


図 3　被験者 28 番の推定結果
Fig. 3　Result of estimation of M28 subject.

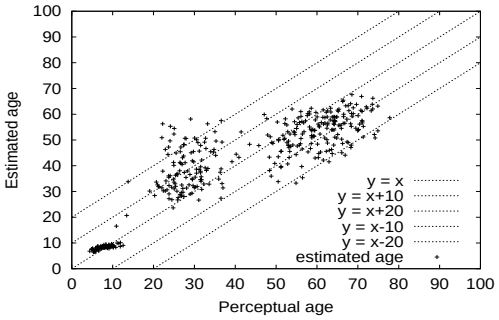


図 4　全被験者の平均ラベルを用いた推定の結果
Fig. 4　Result of estimation using all subject.

### 3.2　評 価 実 験

雑音を含むものを除いた DB 話者 (計 407 名) 各々の知覚的年齢を，その話者自身以外の全 DB 話者を用いて上記方法にて推定し，結果をその話者の知覚的年齢ラベルと比較することで評価した．比較には相関係数を用いた．評価データ長が 5 秒程度あれば結果が収束する [4] ので，評価音声は 5 秒のものを用意した．聴取実験被験者のうち 1 名のラベルを用いて推定した結果を図 3 に，全被験者の知覚的年齢ラベルの平均を用いて推定した結果を図 4 に示す．なお式 (4) を用いている．相関係数は図 4 のときが最大で 0.89，平均が 0.85 であり，比較的高い相関が得られたと．いずれも式 (4) 利用時のスコアである．

### 3.3　検討と考察

**尤度スコアの代表方法の比較**　尤度スコアは各年齢ごとに平均値で代表させる方法と最大値で代表させる方法の 2 通りを採用している．各被験者において相関最大となった方法を集計すると，前者が 9 名，後者が 22 名 (平均ラベルによる推定を含む) となった．一見，最大値法のほうが有効であるようにみえるが，最大値法は，正解と大きく異なる年齢の中に一人でも個人性の似ている話者が存在した場合にその影響を大きく受けるため，推定性能の安定性という意味では問題のある近似方法である．実際のシステム構築時には実験的な検討が必要であろう．

**相関図に対する考察**　各被験者において相関係数最大となった $N$ の値は平均 12，最小で 7 であった．これは期待値操作の必要性を直接支持している．また，図 3，4 を見ると，知覚的年齢ラベルと自動推定結果の関係は全体として $y=x$(正解直線) 周辺に分布しているが，分布全体を直線近似するとその傾きは 1 未満となる．これは，期待値操作時の年齢数を $N$ 固定としており，特に高齢者の場合は正解年齢よりも高い年齢が存在しない

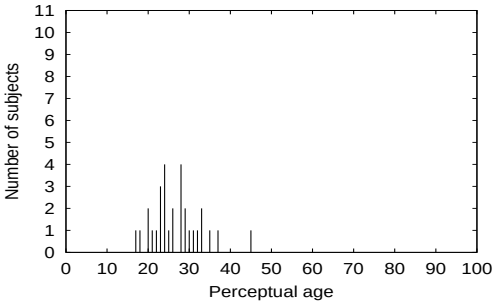


図 5　JNAS-DB 話者の知覚的年齢分布の例

Fig. 5　An example of perceptual age distribution of a speaker in JNAS-DB.

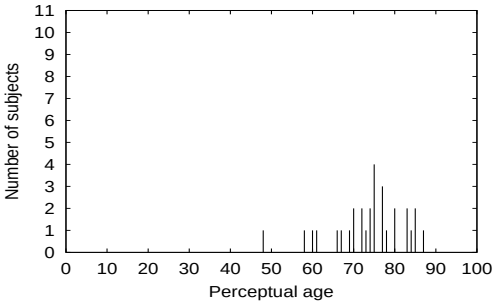


図 6　S-JNAS-DB 話者の知覚的年齢分布の例

Fig. 6　An example of perceptual age distribution of a speaker in S-JNAS-DB.

場合があり，その場合はより低い年齢として推定されてしまう．期待値操作時の年齢幅 (数) の適応的な決定が必要であろう．

## 4. 知覚的年齢分布を用いた未知入力話者の知覚的年齢推定

一般的な年齢の知覚を考える場合，2.3 節の聴取実験結果において，人間の年齢の知覚には幅があることが分っており，一つの値によって代表させることが必ずしも適当とはいえない．そこで，知覚的年齢に確率密度分布を仮定し，それに基づいて入力話者の知覚的年齢分布を自動推定することを検討する．

### 4.1 各 DB 話者の知覚的年齢分布のモデル化

DB 話者毎に，横軸を知覚的年齢，縦軸をその年齢と推定した聴取実験被験者の人数として図 5，6 にその一例を示す．聴取実験の被験者は 30 名と比較的少数であるが，被験者を増やしていくことでこの分布は正規分布に帰着すると推測できる．そこで以降では，各 DB 話者の知覚的年齢分布として被験者全 30 名の推定結果を正規分布でモデル化したものを用いる．

### 4.2 知覚的年齢確率密度分布の自動推定

仮定した正規分布の分散は各 DB 話者の知覚的年齢分布の広がり (年齢不詳の度合い) を表すため，DB 話者の知覚的年齢の様子をより的確に表すことができる．各 DB 話者の知覚的年齢分布を GMM 尤度スコアを重みとして足し合わせることで，入力音声の確率密度分布を推定した．即ち式 (1) における知覚的年齢 $x$ を分布化する訳である．以下で詳説する．

### 4.3 推 定 方 法

3.1 節と同様，期待値操作によって入力話者の知覚的年齢確率密度分布を求めるが，各年齢ごとのデータ量に偏りがあるのでそれをキャンセルする関数を作成する．この関数としては，

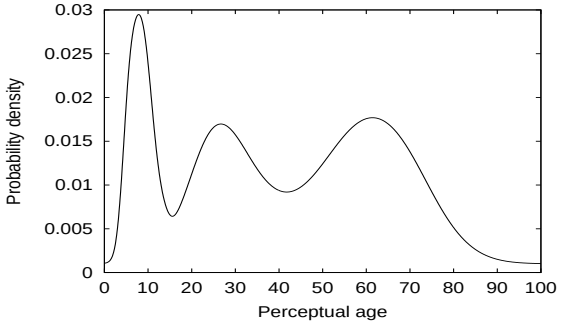


図 7　キャンセル用確率密度関数

Fig. 7　Probability density function to cancel the biased distribution of perceptual age.

全 DB 話者の知覚的年齢分布をそのまま加算し，全区間における積分値が 1 となるよう正規化した関数を使用し，最終的な推定分布は期待値操作で得られる分布をこのキャンセル関数で割ることで取得する．その結果，全 DB 話者から等しく離れた入力話者は確率密度一定，すなわち完全な「年齢不詳者」と扱われる．ただし，キャンセル関数の値が 0 付近の値をとる場合は除算時に問題となるので，キャンセル関数に常に $\varepsilon(\varepsilon \ll 1)$ を加えることによって回避した．ここでは実験的に $\varepsilon = 0.001$ とした．このキャンセル関数を図 7 に示した．

### 4.4 評 価 実 験

評価話者としては，3.2 節の評価実験と同じように DB 話者を自身を除いて推定したものを用いる．キャンセル関数も入力話者自身を除いて作成した．本推定方法関する評価方法としては 2 通り検討した．まず，評価は聴取実験の結果から得られた分布と，4.2 節で推定された分布を比較した．次に，前者の分布の期待値と後者の分布の期待値の相関を調べた．

### 4.5 検討と考察

**推定された確率密度分布の比較による評価**

子供，JNAS，S-JNAS の代表的なものを図 8，図 9，図 10 として示した．結果を見るとピークは仮定した正規分布に近い位置にできている．ただし，年齢幅は有限 (5〜90 歳程度) なので，3.1 節で述べたように，ピークより高い年齢と低い年齢の分布にアンバランスができてしまう．

**推定された知覚的年齢分布の期待値による評価**

推定された知覚的年齢確率密度分布を全区間において期待値操作することで知覚的年齢を推定した結果を図 11 に示した．この相関係数は 0.84 である．この図を見ると全体の傾きが 1 未満になっており，推定された確率密度分布のアンバランスが大きく影響していることが分る．しかし，上述したように，分布のピークは近いところに出ていることから，ピーク年齢を求め，そこを中心に区間を限定して期待値操作を行うことによって推定結果は向上すると思われる．以下では区間限定による知覚的年齢の自動推定について検討する．

### 4.6 区間限定をした期待値操作による知覚的年齢自動推定

前節において述べたように，一部例外はあるものの，推定した知覚的年齢確率密度分布のピークは仮定した正規分布のピークに近い位置にできている．ただし，年齢幅が有限なため，ピークより高い年齢と低い年齢の分布がアンバランスになり，その

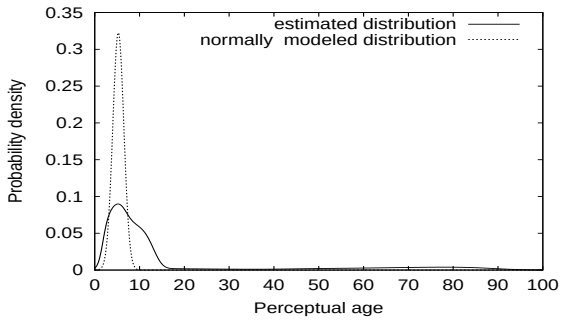


図 8　子供話者の知覚的年齢確率密度分布の例

Fig. 8　An example of estimated distribution of a speaker in children's speech DB.

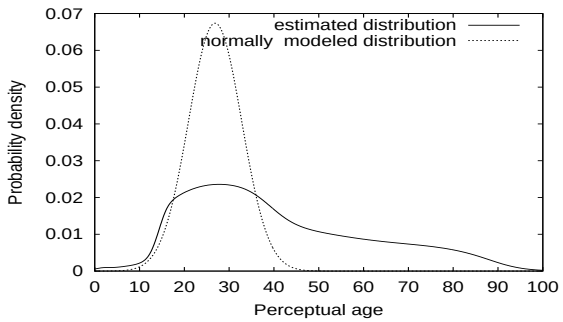


図 9　JNAS 話者の知覚的年齢確率密度分布の例

Fig. 9　An example of estimated distribution of a speaker in JNAS-DB.

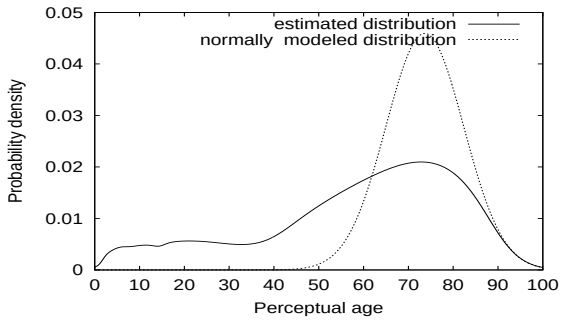


図 10　S-JNAS 話者の知覚的年齢確率密度分布の例

Fig. 10　An example of estimated distribution of a speaker in S-JNAS-DB.

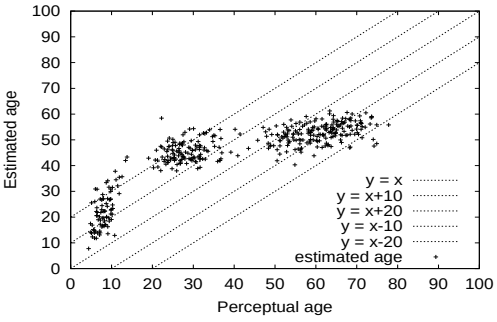


図 11　知覚的年齢分布の全区間を用いた推定結果

Fig. 11　Estimation by expecting the perceptual age distributions.

結果ピーク年齢が少々ずれてしまうことが多い．そこでピーク年齢を基準にして区間を限定した期待値操作を行うことにより，より高い精度で知覚的年齢が推定できると考えられる．以下ではピーク年齢と期待値操作区間の限定範囲の決定方法を示す．

**ピーク年齢検出**　1 歳刻みで 5 点取り，その 5 点の確率密度の

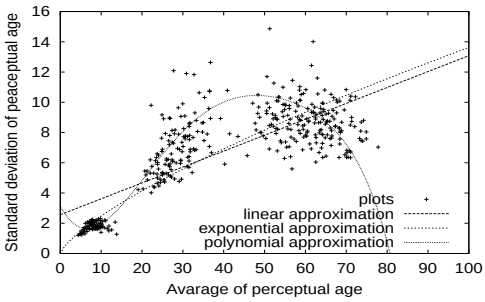


図 12　全被験者の推定年齢ラベルの平均と標準偏差の関係

Fig. 12　Relation between average and standard deviation of the perceptual age.

関係が「小中大中小」となっている場合をピーク年齢候補とした．このピーク年齢候補が複数存在する場合は，このピーク年齢候補の確率密度を前後 $N$ 歳の範囲で積分し，積分値が最も大きいものを最終的なピーク年齢として採用した．ピーク年齢候補の中で最大確率密度を示す点をピーク年齢とする方法も考えられたが，個人性の影響などによって本来推定されるべきでない年齢がピーク年齢となってしまうことをより確実に避けるために上記方法を採用した．積分範囲は予備実験の結果ピーク年齢の相関係数が最大となる $N = 11$ とした．

**期待値操作区間の限定**　区間限定の範囲としては，推定された確率密度分布のピーク周辺の形状が急峻である場合は短く，緩やかである場合は長いというように，分布の大きさによって積分区間を適応的に定めるのが望ましい．図 12 に 2.3 節の聴取実験結果から求めた全被験者の推定年齢ラベルの平均と標準偏差の関係を示した．ピーク周辺の形状はピーク年齢に依存することが図 8〜10 から推測されるので，図 12 に対して引いた標準偏差値に対する近似曲線に基づいて限定範囲を定めることにする．近似方法は線形近似，累乗近似，多項式近似を用いた．

$$y = 0.1052 \times x + 2.5634 \quad \text{(線形近似)} \tag{5}$$

$$y = 0.4503 \times x^{0.7405} \quad \text{(累乗近似)} \tag{6}$$

$$\begin{aligned} y = {} & -4 \times 10^{-10} \times x^6 + 1 \times 10^{-8} \times x^5 \\ & +1 \times 10^{-5} \times x^4 - 0.0012 \times x^3 + 0.0491 \times x^2 \\ & -0.4963 \times x + 3.1186 \quad \text{(多項式近似 : 6 次)} \end{aligned} \tag{7}$$

実際の限定範囲としては，近似式より推定される，ピーク年齢に対して仮定される分布の標準偏差を求め，それを定数倍したものを積分区間とした．定数項の値は実験的に決定した．

### 4.7　推 定 結 果

以上の方法で区間を限定して期待値操作を行った．まず，ピーク年齢推定結果を図 13 に示した．この相関係数は 0.87 である．これよりピーク年齢を取ることで知覚的年齢の推定結果の多くがかなり $y = x$ に近く推定されることが分かる．しかし，中には成人と推定されるべき DB 話者が高齢者と推定されてしまったり，高齢者と推定されるべき DB 話者が成人と推定されてしまっている場合がある．これらについては後で考察する．

以下ではピーク年齢が $y = x$ に近い場合に区間限定を行って期待値操作をすることでさらに $y = x$ に近づけることを試みる．ただし，相関係数などの尺度ではノイズとなるデータの影

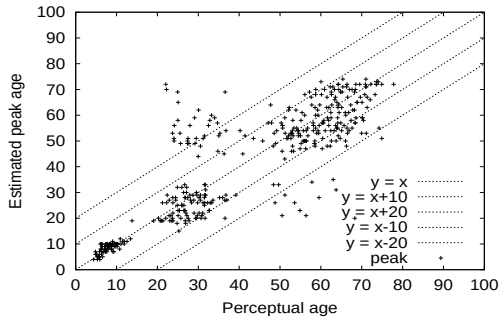


図 13　ピーク年齢のみによる知覚的年齢推定

Fig. 13　Estimation of the perceptual age with age peaks.

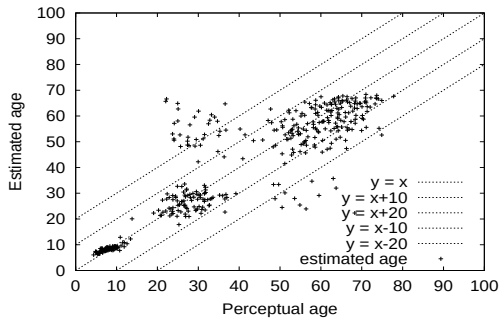


図 14　線形近似を用いた推定結果

Fig. 14　Estimation of the perceptual age with linear approximation.

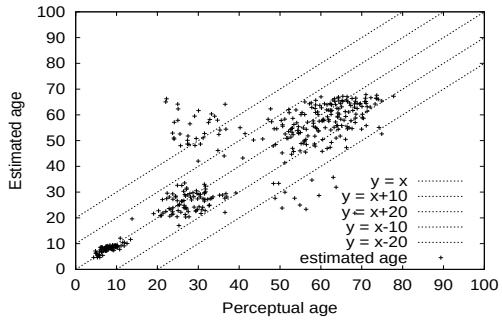


図 15　累乗近似を用いた推定結果

Fig. 15　Estimation of the perceptual age with exponential approximation.

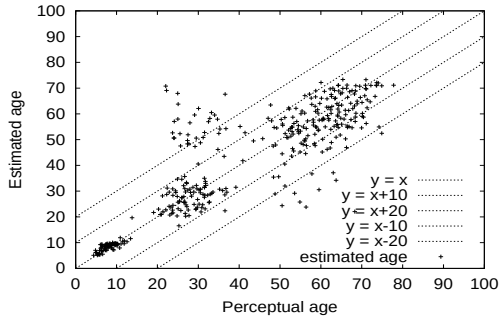


図 16　多項式近似を用いた推定結果

Fig. 16　Estimation of the perceptual age with polynomial approximation.

響で，視覚的に整ったグラフが最適な結果となるわけではない．よってここでは例として 3 つの近似法による区間限定をした期待値操作の結果を挙げるにとどめる．この図を図 14，図 15，図 16 として示した．この 3 つの図の相関係数は全て 0.88 であり，3.2 節の知覚的年齢ラベルを用いた場合の推定結果に比べて相関係数としてはそれほど改善されていない．しかし，図を見る限り $y = x$ に極めて近く推定されているものが明らかに増加しており，聴取実験より得られる知覚的年齢に分布を仮定することの有効性が示されたと言える．以下で正解ラベルから明らかに外れてしまっている場合について考察する．

**正解ラベルから大きく外れたデータについて**　これらの音声を聴取したが，このような極端な間違われ方をする原因は見つからなかった．また，聴取による知覚的年齢分布の分散がその音声の推定し難さを表すと考え，分散の大きさと正解年齢からのずれの関係を調査したが，強い相関関係は見当たらなかった．また，録音サイトの違い，ピーク検出の精度などについても検討したが，原因と思われる現象が観測されなかった．これらの検討より，このノイズといえるデータは話者認識技術を用いた推定の限界であると予想される．実際，先行研究では韻律的特徴に基づく方法 (話速やパワーの局所変動など) についても検討を行い，それを新たなパラメータとして加えることで高齢者非高齢者の識別率の向上を得ており [4]，本手法においても，それらの検討の必要性が示唆される．今後の検討課題の一つである．

## 5. ま と め

本研究では，知覚的年齢をラベルとして与えた場合，正規分布を仮定して分布として与えた場合の 2 種類で知覚的年齢の推定を試み，その結果としてともに約 0.9 の相関値を得た．特に，この知覚的年齢などのように知覚量を認識する場合は，音声認識などの正解か不正解という二者択一的な性質と異なり，期待値操作を行うことができる，分布として扱えるなどの連続的な性質を有するため，それを利用することでより高精度な認識結果を得ることができると思われる．本研究では分布として捉えることの有効性を数字では示せていないが，視覚的には表現されており，今後は分布として捉える方法で推定結果が大きく外れてしまう場合の原因を追求し，それを除去することで高精度な知覚的年齢の推定を実現したい．それには韻律的な議論を入れていく必要があるのではないかと思われる．

### 文　献

[1] 新聞記事読み上げ音声コーパス JNAS
http://www.milab.is.tsukuba.ac.jp/jnas/

[2] 高齢者話者データベース S-JNAS
http://db.ciair.coe.nagoya-u.ac.jp/dbciair/koureisha_files/index.htm/

[3] 子供の声データベース CIAIR-VCV
http://db.ciair.coe.nagoya-u.ac.jp/

[4] 峯松 信明, 広瀬 啓吉, 関口 真理子, “話者認識技術を利用した主観的高齢者の同定とそれに基づく主観的年代の推定,” 情報処理学会論文誌, vol. 43, No. 7, pp. 2186–2196 (2002).

[5] Müller Chiristian and Wasiger Rainer, “Adapting Multimodal Dialog for the Elderly,” Proceedings of the ABIS-Workshop 2002 on Personalization for the Mobile World, Hannover Germany, Oct. 9-11, (2002).

[6] 小川 厚徳, 山口 義和, 松永 昭一, “小学生音声データベースの構築とそれを用いた子供音声認識の一検討,” 信学技報, SP2002-36, pp. 1–6 (2002).

[7] K. P. Markov, “Text-independent speaker recognition based on frame lebel likelihood transformations,” Docter thesis, Toyohashi University of Technology (1999).