ギブス分布に基づく拡張ベイズ推定によるテキスト指定形話者照合

九州工業大学　○植木琢也　溝部祐太　西田健　黒木秀一

Text-Prompted Multistep Speaker Verification Using Gibbs-Distribution-Based Extended Bayesian Inference

Takuya Ueki, Yuta Mizobe, Takeshi Nishida, and Shuichi Kurogi
Kyushu Institute of technology

Abstract: This paper presents a method of text-prompted multistep speaker verification for reducing verification errors. The method is developed for our speech processing system which utilizes competitive associative nets (CAN2s) for learning piecewise linear approximation of nonlinear speech signal to extract feature vectors of pole distribution from piecewise linear coefficients reflecting nonlinear and time-varying vocal tract of the speaker. This paper focuses on reducing verification errors by means of multistep verification using Gibbs-distribution-based extended Bayesian inference (GEBI) in text-prompted speaker verification. The effectiveness of GEBI and the comparison to BI (Bayesian inference) is shown and analyzed by means of experiments using real speech signals.

1. まえがき

本論文ではテキスト指定形話者照合の方法を提示する. テキスト指定形話者照合は詐称者や数字列のなりすましに対応するために開発された[1]. 本論文では照合誤差の低減に焦点を当て，多段ベイズ(BI)推定を用いた方法とギブス分布に基づく拡張ベイズ(GEBI)推定を用いた方法の比較実験を行い，GEBI推定の有効性を示す.

さらに, 近年音声信号が非線形であるという多くの報告があり, 非線形を扱う手法により高い認識性能を実現できるのではないかと考え, 本論文では非線形関数を区分的に線形近似する能力を持つニューラルネットの1つとして提案されている競合連想ネット(CAN2:Competitive Associative Net 2)を使用する．これまでにCAN2はアンサンブル学習の1つであるバギング法をCAN2に適用したbagging CAN2を用いて，音声時系列を性能よく再現し，また, 再現する際に抽出される音声時系列の極の分布の配置パターンを特徴パターンとして用いた話者認識は，高い認識率を得ている[4,5].

2. 多段ベイズ(BI)推定とギブス分布に基づく拡張ベイズ(GEBI)推定

多段ベイズ(BI)推定の問題を簡単に提示し(詳細については[5]を参照)，ギブス分布に基づく拡張ベイズ(GEBI)推定を導入する．まず，話者の集合を $S=\{s_i \mid i\in I_S\}$, 数字の集合を $D=\{d_i \mid i\in I_D\}$ とする．さらに，学習機械として実数関数を学習する回帰学習機械(RLM: Regression Learning Machine; 詳細は後述)を用いることとし，その集合を $\mathrm{RLM}^{[M]}=\{\mathrm{RLM}^{[m]} \mid m\in M\}$ と表す．話者照合の場合，$\mathrm{RLM}^{[m]}$ は話者 $m\in M=S$ の音声信号を学習し，数字照合の場合，$\mathrm{RLM}^{[m]}$ は数字 $m\in M=D$ の音声信号を学習するものとする．$\mathrm{RLM}^{[m_i]}(m_i\in M)$ の出力を $v^{[m_i]}\ (=1 \text{ or } -1)$ とし，$\mathrm{RLM}^{[M]}$ の出力を $\boldsymbol{v}^{[M]}=\left(v^{[m_1]},\cdots,v^{[|M|]}\right)$ と表す．さらに，第 t 段(ステップ)での $\mathrm{RLM}^{[M]}$ の出力を $\boldsymbol{v}_t^{[M]}$ とし，$t=1,2,\cdots T$ に対する出力時系列を $\boldsymbol{v}_{1:T}^{[M]}=\boldsymbol{v}_1^{[M]}\boldsymbol{v}_2^{[M]}\cdots\boldsymbol{v}_T^{[M]}$ と表す．すると多段BIによる $m\left(\in M=S \text{ or } D\right)$ に対する事後確率は次式で与えられる.

$$p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right)=\frac{1}{Z_t}\,p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)p\left(\boldsymbol{v}_t^{[M]} \mid m\right) \tag{1}$$

ここで Z_t は正規化定数であり，$\sum_{m\in M}p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right)=1$ をみたす．さらに $t=1,2,\cdots$, とし次式を得る.

$$p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right)=\frac{1}{Z_t}\exp\left(-t\left(\tilde{L}_{1:t}^{[m]}-\frac{1}{t}\log p_0(m)\right)\right) \tag{2}$$

ここで $p_0(m)=p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:0}^{[M]}\right)$ は事前確率を表し，$\tilde{L}_{1:t}^{[m]}\equiv-\frac{1}{t}\left(\sum_{k=1}^{t}\log p\left(\boldsymbol{v}_k^{[M]} \mid m\right)\right)$ は正規化対数尤度である．また，$\sum_{m\in M}p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right)=1$ より $m\in M$ と $m_\nu=\arg\max_{m_i\in M}p_{\mathrm{B}}\left(m_i \mid \boldsymbol{v}_{1:t}^{[M]}\right)$ の確率の比率は $t\to\infty$ のとき

$$\begin{aligned} r_{\mathrm{B},i,\nu} &\equiv \frac{p_{\mathrm{B}}\left(m_i \mid \boldsymbol{v}_{1:t}^{[M]}\right)}{p_{\mathrm{B}}\left(m_\nu \mid \boldsymbol{v}_{1:t}^{[M]}\right)} \\ &= \frac{p_{\mathrm{B}}(m_i)}{p_{\mathrm{B}}(m_\nu)}\exp\left(-t\left(\tilde{L}_{1:t}^{[m_i]}-\tilde{L}_{1:t}^{[m_\nu]}\right)\right) \\ &\to \begin{cases}1, m_i=m_\nu\\ 0, m_i\neq m_\nu\end{cases} \end{aligned} \tag{3}$$

となる. したがって, 入力音声が未登録の場合でも,

登録話者または登録数字 $m = m_\nu$ に対する学習機械の事後確率 $p_{\mathrm{B}}\left(m \mid \boldsymbol{v}_{1:0}^{[M]}\right)$ が非常に大きくなる．この問題を回避するために以下に示すギブス分布を導入する．

$$p_{\mathrm{G}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right) \equiv \frac{1}{Z_t} \exp\left(-\beta\left(\tilde{L}_{1:t}^{[m]} - \frac{1}{t}\log p_0(m)\right)\right) \quad (4)$$

ここで $p_0(m) = p_{\mathrm{G}}\left(m \mid \boldsymbol{v}_{1:0}^{[M]}\right)$ は事前確率を表し，β は逆温度である．以下に示すように $m \in M$ と $m_\nu = \arg\max_{m_i \in M} p_{\mathrm{G}}\left(m_i \mid \boldsymbol{v}_{1:t}^{[M]}\right)$ の確率の比率は t の増加に伴い，1より小さい値に収束する．

$$r_{\mathrm{G},i,\nu} \equiv \frac{p_{\mathrm{G}}\left(m_i \mid \boldsymbol{v}_{1:t}^{[m]}\right)}{p_{\mathrm{G}}\left(m_\nu \mid \boldsymbol{v}_{1:t}^{[m]}\right)} \to \exp\left(-\beta\left(\tilde{L}_{1:t}^{[m_i]} - \tilde{L}_{1:t}^{[m_\nu]}\right)\right) \to c_i^{\beta} < 1 \quad (5)$$

このことは上記のBI推定の問題を回避できる可能性を示唆する．ここで式(4)より，GEBI確率の漸化式を以下のように導出する．

$$p_{\mathrm{G}}\left(m \mid \boldsymbol{v}_{1:t}^{[M]}\right) \equiv \frac{1}{Z_t} p_{\mathrm{G}}\left(m \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)^{\beta_t/\beta_{t-1}} p\left(\boldsymbol{v}_t^{[M]} \mid m\right)^{\beta_t} \quad (6)$$

式(6)において，$\beta_t = \beta/t, (t \geq 1), \beta_0 = 1$ とする．なお，従来のBI推定法では $\beta_t = 1, (t \geq 0)$ で与えられていた．

3. テキスト指定形話者認識システム

3.1 話者と数字の一段認識

CAN2を用いたテキスト指定形話者認識システムのブロック図を図1に示す．一般に話者認識は音声をアナログ信号からディジタル信号へ変換をする前処理，特徴量抽出，パターンマッチング，話者の決定の4工程で構成される．話者認識は話者照合と話者識別に分類され，本論文ではある者が本人の主張している通りの人物であるかを判断する話者照合を行う．また，テキストとして数字を使用しテキスト指定形話者照合を実現する．話者 s または数字 d に対する回帰学習機械 $\mathrm{RLM}^{[m]}$ $(m = s \text{ or } d)$ の目的関数は次式で表される．

$$y^{[m]}(\boldsymbol{q}) = \begin{cases} 1, \text{ if } \boldsymbol{q} \in \mathrm{Q}^{[m]} \\ -1, \text{ otherwise} \end{cases} \quad (7)$$

ここで $\mathrm{Q}^{[m]}$ は $m\,(= s \text{ or } d)$ に対する特徴ベクトル $\boldsymbol{q}$ の集合を表す．この目的関数の学習後，$\mathrm{RLM}^{[m]}$ は連続関数 $\hat{y}^{[m]} = \hat{f}^{[m]}\left(\boldsymbol{q}^{[m]}\right)$ を出力するが，これを次式より2値化して一段照合あるいは2クラス分類を行う．

$$v^{[m]} = \begin{cases} 1, \text{ if } \hat{y}^{[m]} \geq y_\theta^{[m]} \\ -1, \text{ otherwise} \end{cases} \quad (8)$$

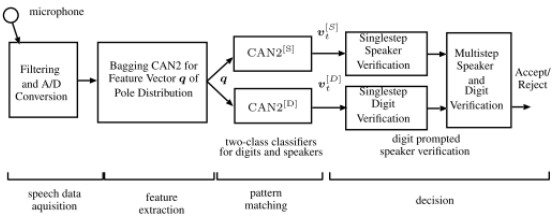

Fig.1. Diagram of text-prompted speaker verification system using CAN2s

すなわち、$v^{[m]} = 1$ の場合 m を受理し，それ以外の場合は棄却する．ここで閾値 y_θ は2.3に示すように安定した照合結果が得られるように調整する．

3.2 数字と話者の多段照合

テキスト指定形話者照合を実行するため，数字列と話者の多段照合を並列に行い、その結果を組み合わせ，数字列が受理された場合のみ話者照合を行う．話者または数字の多段照合の学習機械 $m \in M\,(= S \text{ or } D)$ の特徴ベクトル $\boldsymbol{q} \in Q$ から得た $\mathrm{RLM}^{[M]}$ の二値化出力ベクトル $\boldsymbol{v}^{[M]} = \left(v^{[m_1]}, \cdots v^{[\|M\|]}\right)$ の条件付き確率を式(9)に示す．

$$p(\boldsymbol{v}^{[M]} \mid s) = \prod_{m_i \in M} p(v^{[m_i]} \mid m) \quad (9)$$

さらに，話者または数字の入力時系列 $m_{1:T} = m_1 m_2 \cdots m_T$ に対する出力時系列 $\boldsymbol{v}^{[M]}$ を $\boldsymbol{v}_{1:T}^{[M]} = \boldsymbol{v}_1^{[M]} \boldsymbol{v}_2^{[M]} \cdots \boldsymbol{v}_T^{[M]}$ とする．出力 $\boldsymbol{v}_{1:T}^{[M]}$ が参照ト入力(正解とすべき入力) $m_{1:T}^{[r]}$ に対する応答であるかを判定するために，GEBI推定を使用し，式(10)(11)に示すように $t = 1,2,\cdots,T$ のときの2つの再帰的な事後確率を算出する．

$$p_{\mathrm{G}}\left(m_{1:t}^{[r]} \mid \boldsymbol{v}_{1:t}^{[M]}\right) = \frac{1}{Z_t} p_{\mathrm{G}}\left(m_{1:t-1}^{[r]} \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)^{\beta_t/\beta_{t-1}} p\left(\boldsymbol{v}_t^{[M]} \mid m_t^{[M]}\right)^{\beta_t} \quad (10)$$

$$p_{\mathrm{G}}\left(\overline{m_{1:t}^{[r]}} \mid \boldsymbol{v}_{1:t}^{[M]}\right) = \frac{1}{Z_t} p_{\mathrm{G}}\left(\overline{m_{1:t-1}^{[r]}} \mid \boldsymbol{v}_{1:t-1}^{[M]}\right)^{\beta_t/\beta_{t-1}} p\left(\boldsymbol{v}_t^{[M]} \mid \overline{m_t^{[M]}}\right)^{\beta_t} \quad (11)$$

ここで，Z_t は $p_{\mathrm{G}}\left(m^{[r]} \mid \boldsymbol{v}_{1:T}^{[M]}\right) + p_{\mathrm{G}}\left(\overline{m^{[r]}} \mid \boldsymbol{v}_{1:T}^{[M]}\right) = 1$ を満たす正規化定数である．また，

$$p\left(\boldsymbol{v}_t^{[M]} \mid \overline{m^{[r]}}\right) = \sum_{m \in M \setminus \{m^{[r]}\}} p\left(\boldsymbol{v}_t^{[M]} \mid m\right) / \left(|M| - 1\right) \quad (12)$$

とする．

$t = T$ の時の数字の照合を式(13)により行う．

$$V_{1:T}^{[D]} = \begin{cases} 1, \text{ if } p_{\mathrm{G}}\left(d_{1:T}^{[r]} \mid \boldsymbol{v}_{1:T}^{[D]}\right) \geq p_\theta^{[D]} \\ -1, \text{ otherwise} \end{cases} \quad (13)$$

さらに T ステップにおける話者と数字の照合，すな

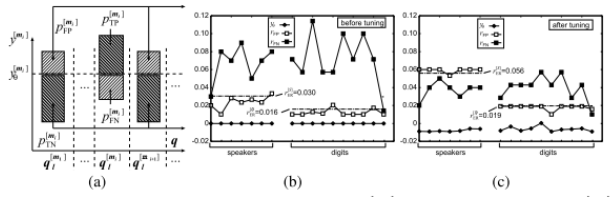

Fig.2. (a) Schematic relationship between the threshold $y_\theta^{[m_i]}$ and the performance ratios $r_{\mathrm{FP}}^{[m_i]}$, $r_{\mathrm{FN}}^{[m_i]}$, $r_{\mathrm{TP}}^{[m_i]}$, $r_{\mathrm{TN}}^{[m_i]}$. The horizontal axis indicates all training feature vectors $\boldsymbol{q}=\boldsymbol{q}_l^{[m_i]}(l=1,2,\cdots)$ obtained from speakers $m=s\in S$ or digits $m=d\in D$. The vertical axis indicates the corresponding output $y^{[m_i]}$ of $\mathrm{RLM}^{[m_i]}$. Experimental results (see 3) of $y_\theta, r_{\mathrm{FP}}, r_{\mathrm{FN}}$ for the original $y_\theta=0$ and tuned y_θ are shown in (b) and (c), respectively.

わちテキスト指定形話者照合の照合は

$$V_{1:T}^{[SD]}=\begin{cases}1, \text{ if } \left(V_{1:T}^{[D]}=1\right)\wedge\left(p_{\mathrm{G}}\left(s_{1:T}^{[r]}\mid \boldsymbol{v}_{1:T}^{[S]}\right)\geq p_\theta^{[S]}\right)\\ -1, \text{ otherwise}\end{cases} \tag{14}$$

より行う．$p_\theta^{[D]}$ と $p_\theta^{[S]}$ は閾値である．以上より，$V_{1:T}^{[SD]}=1$ のとき入力を受理し，それ以外の時は棄却する．

3.3 回帰学習機械の閾値 y_θ の調整

全ての話者と全ての数字に同じ閾値を使用するために，以下の手順により話者 $m=s\in S$ と数字 $m=d\in D$ それぞれについての式(8)における閾値 $y_\theta=y_\theta^{[m]}$ を調整する．まず，入力 m の特徴ベクトル $\boldsymbol{q}$ は $\mathrm{RLM}^{[m_i]}\left(m_i\in S\right)$ によってFP(偽陽性)とFN(偽陰性)に分類される．ランダムに選択されたデータに対するそれらの割合は次のように推定される．

$$\begin{aligned}r_{\mathrm{FP}}^{[m_i]}&=\frac{1}{|M|-1}\sum_{m\in M\setminus m_i}p\left(v^{[m_i]}=1\mid m\right)\\ r_{\mathrm{FN}}^{[m_i]}&=p\left(v^{[m_i]}=-1\mid m\right)\end{aligned} \tag{15}$$

また，TP（真陽性）とTN（真陰性）の比率は $r_{\mathrm{TP}}^{[m_i]}=1-r_{\mathrm{FN}}^{[m_i]}$ と $r_{\mathrm{TN}}^{[m_i]}=1-r_{\mathrm{FP}}^{[m_i]}$ である．ただし，各 $\mathrm{RLM}^{[m_i]}$ は誤差率の平均二乗誤差 $\left\langle\left(v^{[m_i]}-y^{[m_i]}\right)^2\right\rangle$ を最小化するように学習してある．

$$r_{\mathrm{ER}}^{[m_i]}=\frac{1}{|M|}=\left(r_{\mathrm{FN}}^{[m_i]}+\left(|M|-1\right)r_{\mathrm{FP}}^{[m_i]}\right) \tag{16}$$

図2(a)より，$y_\theta^{[m_i]}$ の値を大きくすると $r_{\mathrm{FP}}^{[m_i]}$ が減少し，$r_{\mathrm{FN}}^{[m_i]}$ が増加していることがわかる．この関係より，三つの要素 $\left(y_{\theta,n}^{[m_i]}, r_{\mathrm{FP},n}^{[m_i]}, r_{\mathrm{FN},n}^{[m_i]}\right)$ からなる $\delta^{[m_i]}$ を得る．

試用閾値は

$$y_{\theta,n}^{[m_i]}=\left(n/n_y\right)\left(\bar{y}_P^{[m_i]}-\bar{y}_N^{[m_i]}\right)+\bar{y}_N^{[m_i]} \tag{17}$$

である．ただし，$n\in I_y=\{0,1,2,\cdots n_y\}$ とする．また $\bar{y}_P^{[m_i]}, \bar{y}_N^{[m_i]}$ は全ての学習データの特徴ベクトル $\boldsymbol{q}$ の出力 $y^{[m_i]}$ の真と偽の平均であり，n_y は分割数である．$\delta^{[m]}\left(m\in M\right)$ のデータを式(18)により分割する．

$$\delta_l^{[m]}=\left\{\left(y_\theta, r_{\mathrm{FP}}, r_{\mathrm{FN}}\right)\in\delta^{[m]}\mid l=\arg\min_{l'\in I_y}\left|r_{\mathrm{FP}}-r_{l'}\right|\right\} \tag{18}$$

ただし，

$$\begin{aligned}r_l&=\left(l/n_y\right)\left(r_1-r_0\right)+r_0, \left(l\in I_y\right)\\ r_0&=\min\left\{r_{\mathrm{FP},n}^{[m]}\mid m\in M, n\in I_y\right\}\\ r_1&=\max\left\{r_{\mathrm{FP},n}^{[m]}\mid m\in M, n\in I_y\right\}\end{aligned} \tag{19}$$

とする．その後 $\left(y_\theta, r_{\mathrm{FP}}, r_{\mathrm{FN}}\right)\in\delta^{[m]}$ の平均と差異を計算する．ここで $\left(\mathrm{E}_{\delta_l^{[m]}}(y_\theta), \mathrm{E}_{\delta_l^{[m]}}(r_{\mathrm{FP}}), \mathrm{E}_{\delta_l^{[m]}}(r_{\mathrm{FN}})\right)$ と $\left(\mathrm{V}_{\delta_l^{[m]}}(y_\theta), \mathrm{V}_{\delta_l^{[m]}}(r_{\mathrm{FP}}), \mathrm{V}_{\delta_l^{[m]}}(r_{\mathrm{FN}})\right)$ はそれぞれ $m\in M, l\in I_y$ とする．$l\in I_y$ を r_{FP} および r_{FN} の全ての分散の和を最小化するものとし，式(20)のように示す．

$$\tilde{l}=\arg\min_{l\in I_y}\sum_{m\in M}\left(\mathrm{V}_{\delta_l^{[m]}}(r_{\mathrm{FP}})+\mathrm{V}_{\delta_l^{[m]}}(r_{\mathrm{FN}})\right) \tag{20}$$

これより，全ての $m\in M$ における r_{FP} と r_{FN} の分散を小さくすることができる．

4. 実験

4.1 実験方法

音声データは研究室の静かな部屋の中でサンプリング周期8[kHz], 分解能16[bit]の下でサンプリングしたを使用した．7人の話者の集合をS={fHS,fMS,mKK,mKO,mMT,mNH,mYM}，10個の数字の集合をD={/zero/,/ichi/,/ni/,/san/,/yon/, /go/,/roku/,/nana/,/hachi/,/kyu/}とする．話者の各数字は2か月間で異なる日時で10サンプルとったものである．数字は $x=x_{s,d,l}\left(s\in S, d\in D, w\in W\right)$ によって発声されたものとし，データセットを $X=\left\{x=x_{s,d,l}\mid s\in S, d\in D, l\in L\right\}$ とする．これらの音声時系列について，out-of-bag(OOB)によって極分布による評価実験を行った．OOBはLOOCV(Leave-one-set-out-cross-validation)より小さい傾きと分散を有すると期待されている[7]. なお，実験結果はテスト話者 s の数字列 $d_{1:T}=d_1d_2\cdots d_T$ とそれに対応する参照話者 $s^{[r]}$ の数字列 $d_{1:T}^{[r]}=d_1^{[r]}d_2^{[r]}\cdots d_T^{[r]}$ について，$r_{\mathrm{CD}}=n_{\mathrm{CD}}/T$ の比率で正しい数字 $d_t=d_t^{[r]}$ を含むという条件の下ランダムに $d_t, s, d_t^{[r]}, s^{[r]}$ を選択した．

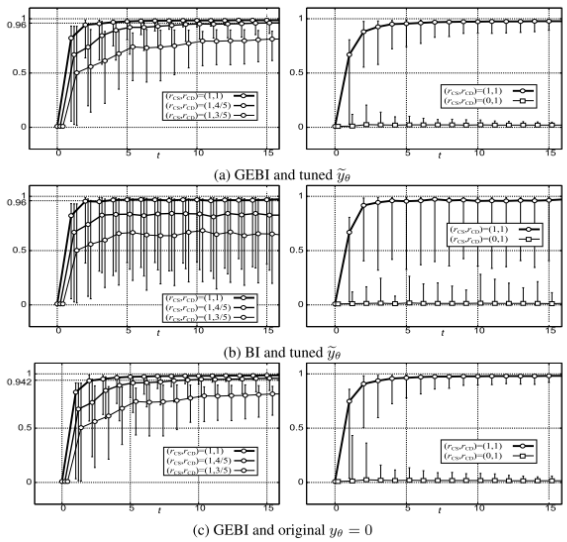

Fig. 3. Experimental result of multistep probability for digits (left) and speakers (right). The plus and minus error bars indicate RMS (root mean square) of positive and negative errors from the mean, respectively. The curves for different datasets are shifted slightly and horizontally to avoid crossovers.

4.2 実験結果と考察

本実験ではテスト数字列は5桁とし，また $T=15=5\times3$ である．表1に実験結果を示す．表1の $(r_{\mathrm{CS}},r_{\mathrm{CD}})=(0,1)$ の行は不正解話者 $(r_{\mathrm{CS}}=0)$ が5桁の数字を全て正しく $(r_{\mathrm{CD}}=1)$ 発声した結果を示し，その下にはそれぞれ正解話者 $(r_{\mathrm{CS}}=1)$ が5桁の数字を $r_{\mathrm{CD}}=1,0,4/5,3/5$ と発声したときの結果である．誤棄却率FRRは $\mathrm{FRR}=100-r_{\mathrm{acc}}^{[SD]}[\%]$ で与えられる．方法(a)では $(r_{\mathrm{CS}},r_{\mathrm{CD}})=(1,1)$ において，ステップ数 t が増加すると誤棄却率は減少する．さらに，$(r_{\mathrm{CS}},r_{\mathrm{CD}})=(1,1/3)$ のおいて，誤受理率FARも減少する．この単調減少の性質は以下の式(4)に示すGEBIの確率 $p_{\mathrm{G}}\left(m\mid \boldsymbol{v}_{1:t}^{[M]}\right)$ がステップ数 t の増加によって単調に収束する性質から得られる．式(2)で与えられるBIの確率 $p_{\mathrm{B}}\left(m\mid \boldsymbol{v}_{1:t}^{[M]}\right)$ もまたステップ数 t の増加に伴い単調に収束するが，GEBI推定に比べて収束が遅い．すなわち，$\tilde{L}_{1:t}^{[m]} >> |\log p_0(m)|/t$ を満たす t において $p_{\mathrm{G}}\left(m\mid \boldsymbol{v}_{1:t}^{[M]}\right)$ は収束に達するが $p_{\mathrm{B}}\left(m\mid \boldsymbol{v}_{1:t}^{[M]}\right)$ は収束に達しないことがわかる．図3(b)では確率曲線の誤差範囲が大きく，しかも変動している．従って，テキスト指定形多段話者照合の照合誤差について，GEBI推定は有効であると考えられる．

Table 1. Experimental result of acceptance rates $r_{\mathrm{acc}}^{[\mathrm{D}]}$ and $r_{\mathrm{acc}}^{[\mathrm{SD}]}$ achieved by the methods using (a) GEBI and tuned $\tilde{y}_\theta$ (c) GEBI and original $y_\theta=0$, and four datasets with $(r_{\mathrm{CS}},r_{\mathrm{CD}})=(0,1),(1,1),(1,1/4)$ and $(1,1/3)$ The values of $r_{\mathrm{acc}}^{[\mathrm{D}]}$ and $r_{\mathrm{acc}}^{[\mathrm{SD}]}$ are expressed by the rate[%]to the total 1000 test sequences for each case. The thresholds are $p_\theta^{[D]}=0.96$ for (a) and (b),0.942 for(c), $p_\theta^{[S]}=0.5$ and $T=15$.

		(a) GEBI & $\tilde{y}_\theta$		(b) BI & $\tilde{y}_\theta$		(c) GEBI & $y_\theta=0$	
r_{CS}	r_{CD}	$r_{\mathrm{acc}}^{[D]}$	$r_{\mathrm{acc}}^{[\mathrm{SD}]}$	$r_{\mathrm{acc}}^{[D]}$	$r_{\mathrm{acc}}^{[\mathrm{SD}]}$	$r_{\mathrm{acc}}^{[D]}$	$r_{\mathrm{acc}}^{[\mathrm{SD}]}$
0	1	**97.5**	*0.0*	**94.9**	*0.0*	**98.4**	*0.0*
1	1	**98.1**	**98.1**	**96.3**	**94.4**	**98.0**	**98.0**
1	4/5	*76.3*	*76.3*	*72.5*	*70.5*	*84.8*	*84.8*
1	3/5	*0.2*	*0.2*	*44.5*	*43.7*	*0.8*	*0.8*

5. 結論

照合誤差を低減するためにGEBIを用いたテキスト指定形多段話者照合の方法を提案した。実音声信号を用いた確率および実験結果の分析により，GEBIの確率はBIの確率より安定ではるかに誤差率を低減することを示した。本論文ではテスト話者は登録話者のみであったが、今後の追加研究として，テスト話者が未登録話者の場合の方法を検討する．

参考文献

[1] Beigi, H.: Fundamentals of speaker recognition. *Springer-Verlag New York Inc (C)* (2011)
[2] Melin, H., Lindberg, J.: Prompting of Passwords in Speaker Verification Systems, Fonetik-97, Phonum 4, Umera University, Sweden, May 28-30. (1997)
[3] Kurogi, S., Ueno, T. and Sawa, M.: A batch learning method for competitive associative net and its application to function approximation. *Proc. SCI2004*, Vol. V, pp.24–28 (2004)
[4] Kurogi, S., Mineishi, S. and Sato, S.: An analysis of speaker recognition using bagging CAN2 and pole distribution of speech signals. Proc. ICONIP2010, Part I, LNCS 6443, pp.363–370 (2010)
[5] Mizobe, Y., Kurogi, S., Tsukazaki, T., and Nishida, T.: Multistep speaker identification using Gibbs-distribution-based extended Bayesian inference for rejecting unregistered speaker. Proc. ICONIP2012, Part V, LNCS 7667, pp.247–255 (2013)
[6] Campbell, J.P.: Speaker Recognition: A Tutorial. Proc. the IEEE, Vol. 85, No.9, pp.1437–1462 (1997)
[7] Kurogi, S.: Improving generalization performance via out-of-bag estimate using variable size of bags, J. Japanese Neural Network Society, Vol. 16, no. 2, pp.81–92 (2009)