音声認識システムの統合を目的とした識別学習の枠組み*

○太刀岡勇気 (三菱電機・情報総研), 渡部晋治, ルルージョナトン, ハーシージョン (MERL)

1 はじめに

異なる音声認識システムの仮説を統合することで、Recognizer Output Voting Error Reduction (ROVER) [1] のように、たとえ補助システムの性能が元のシステムの性能よりも低くとも、音声認識の性能改善を図ることができる。効率的にシステムを統合するためには、異なる傾向を持つ仮説を統合することが重要であり、補助システムの仮説が元のシステムの仮説と似通っていたり、誤りが過分の場合には、統合により性能が向上しないこともあり得る。ただし古典的なシステム統合手法は、理論的な背景に乏しく、補助システムの構築には試行錯誤を必要とする。

この問題に対処するため、我々は、正解ラベルと元のシステムと補助システムの仮説の傾向の関連性が明らかな、相互情報量最大化 (MMI) 基準に基づく、音響モデルの学習法を提案した [2]。本報では、すでに提案した学習法を拡張し、システム統合のための系列識別学習の一般的な枠組みを提案する。提案法は識別的音響モデルや特徴量変換といった幅広いモデル学習に対応できる。ここでは、音響モデル (すでに提案したガウス混合モデル (GMM) に加え、深層神経回路網 (DNN)) と識別的特徴量変換へ適用した。我々は識別学習の目的関数を一般化し、正解ラベルに関する目的関数と、元のシステムの仮説に関する目的関数の調整ができるようにした。提案法は、従来のラティスに基づく識別学習の単純な拡張であることから、識別学習と明確な類似性を持つという利点がある。これに加え、提案法はマージンを考慮した識別学習になっており、補助システムの出力を元のシステムの出力からどの程度離すかを調整できる。

2 節において、補助システム構築のための一般的な識別学習の枠組みについて述べる。音響モデル (DNN) の系列の識別学習 (3 節) や識別的特徴量変換 (4 節) に適用し、5 節で、提案法の有効性を実験的に検証する。

2 補助システムの識別学習

提案法では、補助システムのモデルは、ある初期モデルから識別的に学習を進めることで構築する。提案の補助システムの識別学習法は、一般的な識別学習の原理を拡張したものになっている。Q 個の既存のシステムに対し、提案の目的関数 $\mathcal{F}^c$ は、通常の識別学習の目的関数 $\mathcal{F}$(正解ラベル ω_r と関連) から、元のシステムにより生成された 1 位の仮説 $\omega_{q,1}$ $(q = 1, \ldots, Q)$ に関連する項を引き去ったものである。

$$\mathcal{F}^c_\varphi(\omega_r, \omega_{q,1}) = (1+\alpha)\mathcal{F}_\varphi(\omega_r) - \frac{\alpha}{Q}\sum_{q=1}^{Q}\mathcal{F}_\varphi(\omega_{q,1}) \tag{1}$$

φ は最適化される補助システムのモデルパラメータの組であり、α はスケーリング係数である。もし α が零の時には、目的関数は旧来の識別学習のそれに一致する。式 (1) の第 1 項は識別学習の基準に従って、当該システムの性能を向上させる一方で、第 2 項は当該システムを元のシステムの出力結果から遠ざける役割を持っている。α は両者のバランスを取っている。次節以降、式 (1) における目的関数とモデルパラメータの具体的な形を検討する。

3 音響モデルの識別学習

本節では、MMI 基準を上述の枠組みに適用する。MMI 学習では、正解ラベル列 ω_r と初期モデル (例えば ML モデル) により生成されたラティス上の仮説 ω に対する、以下に示す目的関数を最大化する。

$$\mathcal{F}^{\mathrm{MMI}}_\lambda(\omega_r) = \ln \frac{P_\lambda(\omega_r, \boldsymbol{X})}{\sum_\omega P_\lambda(\omega, \boldsymbol{X})} \tag{2}$$

$$= \ln \frac{\sum_{s_r \in \mathcal{S}_{\omega_r}} p_\lambda\left(s_r, \boldsymbol{X}\right)^\kappa p_L(\omega_r)}{\sum_\omega \sum_{s \in \mathcal{S}_\omega} p_\lambda\left(s, \boldsymbol{X}\right)^\kappa p_L(\omega)} \tag{3}$$

λ は最適化される HMM パラメータの組、$\boldsymbol{X} = \{\boldsymbol{x}_t | t = 1, \cdots, T\}$ は T フレームの特徴量ベクトル列である。$P_\lambda(\omega, \boldsymbol{X})$ は、HMM 状態系列 s における、音響モデルスコア p_λ(音響スケール κ) と言語モデルスコア p_L の積である。式 (3) において、音響スコアは正解ラベル s_r および s に関する総和で求められる。$\mathcal{S}_{\omega_r}$ と $\mathcal{S}_\omega$ は各々、正解ラベル ω_r と仮説 ω を出力する HMM 状態系列の組である。以下、単純化のために、Q は 1 とし、インデックス q は省略する。

式 (1) の φ を λ_c に、$\mathcal{F}$ を $\mathcal{F}^{\mathrm{MMI}}$ に置き換えると、補助システムを構築するための目的関数が得られる。

$$\mathcal{F}^{\mathrm{c}}_{\lambda_c}(\omega_r, \omega_1) = \mathcal{F}^{\mathrm{MMI}}_{\lambda_{\mathrm{c}}}(\omega_r) + \alpha \ln \frac{P_{\lambda_c}(\omega_r, \boldsymbol{X})}{P_{\lambda_c}(\omega_1, \boldsymbol{X})} \tag{4}$$

これは MMI 識別学習の枠組み内にあるが、対数尤度比の項が付加されている点が異なる。

---

*A discriminative training framework for speech recognition system combination, by TACHIOKA, Yuuki (Mitsubishi Electric Corp.), WATANABE, Shinji, LE ROUX, Jonathan, HERSHEY, John R.(MERL)

ブーステッド MMI(bMMI) [3] では、MMI の目的関数に、正解率の低い仮説に対応する特徴量を増幅する効果のある係数を導入する。

$$\mathcal{F}_{\lambda}^{\mathrm{bMMI}}(\omega_r) = \ln \frac{\sum_{s_r \in \mathcal{S}_{\omega_r}} p_\lambda\left(s_r, \boldsymbol{X}\right)^\kappa p_L(\omega_r)}{\sum_{\omega} \sum_{s \in \mathcal{S}_\omega} p_\lambda\left(s, \boldsymbol{X}\right)^\kappa p_L(\omega) e^{-bA(s,s_r)}} \tag{5}$$

$A(s, s_r)$ は、HMM 状態系列 s の正解系列 s_r に対する状態/音素/単語正解率で、フレームごとに計算される。式 (4) の単純な拡張により、$\mathcal{F}^{\mathrm{MMI}}$ を $\mathcal{F}^{\mathrm{bMMI}}$ で置き換え、式 (5) と同様に対数尤度比の項に (逆符号の) 増幅係数を加えると、以下の目的関数を得る [1]。

$$\begin{aligned}\mathcal{F}_{\lambda_c}^{\mathrm{c}}(\omega_r, \omega_1) &= \mathcal{F}_{\lambda_c}^{\mathrm{bMMI}}(\omega_r) \\ &+ \alpha \ln \frac{\sum_{s_r \in \mathcal{S}_{\omega_r}} p_\lambda\left(s_r, \boldsymbol{X}\right)^\kappa p_L(\omega_r)}{\sum_{s_1 \in \mathcal{S}_{\omega_1}} p_\lambda\left(s_1, \boldsymbol{X}\right)^\kappa p_L(\omega_1) e^{b_1 A(s_1,s_r)}}\end{aligned} \tag{6}$$

s_1 は元のシステムの 1 位の仮説 ω_1 に対応する HMM の状態系列である。逆符号の増幅係数 b_1 の役割については [2] の議論を参照されたい。この手順は音響モデルの識別学習、識別的特徴量変換のいずれにおいても共通に用いることができる。

DNN-HMM において、MMI 基準 (2) に基づく系列的な識別学習法が提案されている [4]。ここでは、この手法に提案法を適用することを考える [2]。DNN は HMM の状態 j の出力確率を出力する。

$$p_\theta\left(\boldsymbol{x}_t | j\right) = \frac{p_\theta\left(j | \boldsymbol{x}_t\right)}{p_0\left(j\right)} \tag{7}$$

$p_0\left(j\right)$ は学習データから計算される事前確率である。HMM 状態毎に、モデル θ は soft-max の活性化関数 p_θ を含む。

$$p_\theta(j|\boldsymbol{x}_t) = \frac{\exp a(j|\boldsymbol{x}_t)}{\sum_{j'} \exp a(j'|\boldsymbol{x}_t)} \tag{8}$$

a は出力層の活性であり、MMI 基準に基づき、識別的に学習される。MMI の目的関数は式 (6) の λ を θ で置き換えたものとなる。活性 a の更新式は、目的関数をそれで微分して得られる。

$$\begin{aligned}\frac{\partial \mathcal{F}^{\mathrm{bMMI}}}{\partial a(j)} &= \sum_{j'} \frac{\partial \mathcal{F}^{\mathrm{bMMI}}}{\partial \log p_\theta\left(\boldsymbol{x}_t | j'\right)} \frac{\partial \log p_\theta\left(\boldsymbol{x}_t | j'\right)}{\partial a(j)} \\ &= \kappa(\gamma_{j,t}^{num} - \gamma_{j,t}^{den}) = \kappa \Delta_{j,t}\end{aligned} \tag{9}$$

提案法においては、式 (9) のパラメータを、以下のように修正すればよい。($\gamma_{jm,t}^{num}$ は不変)

$$\begin{aligned}\Delta'_{j,t} &= (1+\alpha)\left(\gamma_{j,t}^{num} - \gamma_{j,t}^{den\prime}\right) \\ \gamma_{j,t}^{den\prime} &= \frac{\gamma_{j,t}^{den} + \alpha \gamma_{j,t}^{1}}{1+\alpha}\end{aligned} \tag{10}$$

[1]同じ音素・単語系列を実現する HMM の状態系列は複数あるので、式 (6) の第 2 項はこれらの系列に関する和になり、増幅係数 b_1 が最適化に関係する。

[2]GMM への応用は文献 [2] を参照されたい。

Algorithm 1 Construct complementary system model for DNN

Require: Initial model θ, base system models θ_q, numerator (ω_r aligned) lattice $\mathcal{A}$, and denominator lattice $\mathcal{L}$ of Eq. (2) or (5)
  for $i = 1$ to i_{eb} **do**
    Rescore $\mathcal{A}$ and $\mathcal{L}$ with θ
    $\gamma_{j,t}^{num}$ and $\gamma_{j,t}^{den}$ ⇐posteriors are gathered on $\mathcal{A}$ and $\mathcal{L}$, respectively
    $\gamma_{j,t} \Leftarrow -\gamma_{j,t}^{den} + (1+\alpha)\gamma_{j,t}^{num}$
    for $q = 1$ to Q **do**
      Rescore $\mathcal{L}$ with θ_q
      $\mathcal{L}_1$ ⇐best path of $\mathcal{L}$
      Rescore $\mathcal{L}_1$ with θ
      $\gamma_{j,t}^{1}$ ⇐ posteriors are gathered on $\mathcal{L}_1$
      $\gamma_{j,t} \Leftarrow -\frac{\alpha}{Q}\gamma_{j,t}^{1} + \gamma_{j,t}$
    end for
    $\gamma_{j,t}^{num}, \gamma_{j,t}^{den}$ ⇐positive and negative parts of $\gamma_{j,t}$
    $\theta \Leftarrow$ Update a by EBW or GD (Eq. (9))
  end for
Ensure: Complementary system model ($\theta_c \leftarrow \theta$)

すべての DNN のパラメータの勾配は、式 (9) より導かれ、後ろ向き伝搬により求められる。Algorithm 1 に、DNN の補助システムを構築する手順を示す。

4 識別的特徴量変換

音響モデルに加え、識別的基準に基づく特徴量変換が提案されている [3]。この方法では、高次元 (L 次元) の非線形な 特徴量を低次元 (K 次元) の特徴量に写像する行列 $\boldsymbol{M}$ を推定する。

$$\boldsymbol{y}_t = \boldsymbol{x}_t + \boldsymbol{M}\boldsymbol{h}_t \tag{11}$$

$\boldsymbol{h}_t$ は非線形特徴量、$\boldsymbol{y}_t$ は変換された特徴量である。行列 $\boldsymbol{M}$ は、$K \times L$ 次元であり、MMI 基準により学習される。この方法は特徴量空間 MMI(f-MMI) あるいはその拡張のブーステッド f-MMI(f-bMMI) と呼ばれる。式 (11) の $\boldsymbol{y}$ を式 (5) の $\boldsymbol{x}$ に代入する ($\boldsymbol{X}$ を $\boldsymbol{Y}$ で置き換える) ことで、f-bMMI の目的関数 $\mathcal{F}^{\text{f-MMI}}$ が得られる。($\boldsymbol{Y}$ は特徴量ベクトル $\{\boldsymbol{y}_t | t = 1, \cdots, T\}$。) 目的関数を $\boldsymbol{M}$ で微分して、$\boldsymbol{M}$ を最適化する。N 個のガウス分布より、非線形特徴量 $\boldsymbol{h}_t = [\boldsymbol{h}_{t,1}; ...; \boldsymbol{h}_{t,N}]$ は、

$$\boldsymbol{h}_{t,n} = p_{t,n}\left[\frac{x_{t,1} - \mu_{n,1}}{\sigma_{n,1}}, \cdots, \frac{x_{t,K} - \mu_{n,K}}{\sigma_{n,K}}, \beta\right]^\top \tag{12}$$

のように計算される。$\mu_{n,k}$ と $\sigma_{n,k}$ は、k 番目の次元の n 番目のガウス分布の平均と標準偏差である。β は

Algorithm 2 Construct complementary system model for f-MMI

Require: Acoustic model λ, initial matrix $\boldsymbol{M}$, base system matrix $\boldsymbol{M}_q$, numerator (ω_r aligned) lattice $\mathcal{A}$, and denominator lattice $\mathcal{L}$

for $i = 1$ to i_{eb} **do**

Rescore $\mathcal{A}$ and $\mathcal{L}$ with λ using $\boldsymbol{y}_t$ $(= \boldsymbol{x}_t + \boldsymbol{M}\boldsymbol{h}_t)$

$\gamma^{num}_{jm,t}$ and $\gamma^{den}_{jm,t}$ $\Leftarrow$posteriors of $\mathcal{A}$ and $\mathcal{L}$, respectively

$\gamma_{jm,t} \Leftarrow -\gamma^{den}_{jm,t} + (1+\alpha)\gamma^{num}_{jm,t}$

for $q = 1$ to Q **do**

Rescore $\mathcal{L}$ with λ using $\boldsymbol{y}_t$ $(= \boldsymbol{x}_t + \boldsymbol{M}_q\boldsymbol{h}_t)$

$\mathcal{L}_1 \Leftarrow$best path of $\mathcal{L}$

Rescore $\mathcal{L}_1$ with λ

$\gamma^{1}_{jm,t} \Leftarrow$ posterior of $\mathcal{L}_1$

$\gamma_{jm,t} \Leftarrow -\frac{\alpha}{Q}\gamma^{1}_{jm,t} + \gamma_{jm,t}$

end for

$\gamma^{num}_{jm,t}, \gamma^{den}_{jm,t}$ $\Leftarrow$positive and negative parts of $\gamma_{jm,t}$

$\boldsymbol{M} \Leftarrow$ Update elements in $\boldsymbol{M}$ by calculating the indirect differential

end for

Ensure: Complementary system matrix ($\boldsymbol{M}_c \leftarrow \boldsymbol{M}$)

スケーリング係数である。$p_{t,n}$ はフレームごとに計算されるガウス分布の事後確率で、上位 N_1 個の事後確率のみを用いる。この仮定により、$\boldsymbol{h}_t$ がスパースになり、計算量を削減することができる。

補助システムの目的関数は、式 (1) より導出される。その際、φ を $\boldsymbol{M}_c$ で、$\mathcal{F}$ を $\mathcal{F}^{\text{f-MMI}}$ で置き換える。

$$\mathcal{F}^{\text{c}}_{\boldsymbol{M}_c}(\omega_r, \omega_1) = \mathcal{F}^{\text{f-bMMI}}_{\boldsymbol{M}_c}(\omega_r) + \alpha \ln \frac{\sum_{s_r \in \mathcal{S}_{\omega_r}} p_{\boldsymbol{M}_c}(s_r, \boldsymbol{Y})^{\kappa} p_L(\omega_r)}{\sum_{s_1 \in \mathcal{S}_{\omega_1}} p_{\boldsymbol{M}_c}(s_1, \boldsymbol{Y})^{\kappa} p_L(\omega_1) e^{-b_1 A(s_1, s_r)}} \tag{13}$$

提案法は一般化された目的関数から始めて、識別的特徴量変換にも適用できる。Algorithm 2 は最急降下法を用いた補助システムのモデルの更新手順である。

5 音声認識実験

5.1 実験条件

提案法の検証のために、2 つのコーパス (第 2 回 CHiME チャレンジ (トラック 2) と日本語話し言葉コーパス (CSJ)) を用いた。前者で、提案法の音響モデル (DNN)、識別的特徴量変換への有効性を示し、後者で、提案法が異なるタスクにおいても有効に働くことを示す。前者は、残響･非定常高騒音環境における中程度語彙 (5,000 単語) のタスク (Wall Street Journal 0) である [5]。Kaldi ツールキット [6] を使った。学習セットは 83 話者の 7,138 発話、開発セット (si_dt_05) は 10 話者の 409 発話、評価セット (si_et_05) は 12 話者の 330 発話からなる。音響モデルは学習セットで学習し、音響スケール κ は開発セット (si_dt_05) で調整した。騒音は他の話者の妨害、家庭内の騒音、音楽といった非定常性のもので、SNR は $\{-6, -3, 0, 3, 6, 9\}$dB の 6 段階である。事前分布に基づくバイナリマスク [7] による騒音抑圧後のデータを使った。

音響特徴量は、0-12 次 MFCC + Δ + $\Delta\Delta$ で、これに、特徴量変換手法 (線形判別分析 (LDA)、最尤線形変換 (MLLT)) と話者適応手法 (話者適応学習 (SAT)、特徴量空間最尤線形回帰 (fMLLR)) を使った。

コンテクスト依存 HMM の状態数は 2,500 で、ガウス分布の総数は 15,000 である。DNN の学習には、Povey による Kaldi の実装を用いた。DNN は隠れ層 3 層からなり、100 万のパラメータを持つ。学習率は 0.01 から始めて、最終的に 0.001 まで低減した。識別的特徴量変換においては、400 のガウス分布を用い、オフセット特徴量にはそれぞれ 9 フレームコンテキスト拡張した 40 次元の特徴量を与えた。よって特徴量ベクトル $\boldsymbol{h}_t$ の次元は、$400 \times 40 \times 9$ となる。事後確率の上位 2 つに対応する特徴量だけを選択した。β は 5 とした。提案法のパラメータ α と b_1 はそれぞれ開発セットで調整し、0.75 と 0.3 に設定した。

CSJ は、講義形式の大語彙 (70,000 単語) 連続音声認識のタスクである。テストセット 1 は 10 人の男性話者による 10 から 15 分程度の講演である。HMM の状態数は 3,500、ガウス分布の総数は 96,000 とした。提案法のパラメータは CHiME チャレンジと同じものを用いた。複数システムの出力仮説を統合する際には、信頼度による重み付きの ROVER を用いた。

5.2 第 2 回 CHiME チャレンジ (高騒音下音声認識)

f-bMMI と DNN の開発セットにおける有効性を検証した。GMM システムの場合は、文献 [2] に示してある。Table 1 (左列) は、MFCC 特徴量に特徴量変換 (LDA+MLLT) と話者適応 (SAT+fMLLR) の変換を施した特徴量に対して、f-bMMI を行った場合の WER を示している。上段、上中段、下中段、下段はそれぞれ、従来の単一システム (S1-S4)、従来法による複数システムの ROVER(R1,R2)、提案法による補助システム (P1,P2)、提案法を含む ROVER(RP1,RP2) である。f-bMMI は通常同時に音響モデルも識別学習することが多い。この場合、補助システムを 2 つの方法により構築した。1 つ目は、f-bMMI と bMMI、双方の目的関数を修正した場合 (式 (13) かつ式 (6)) である (f-bMMI$_{\text{c}}$ + bMMI$_{\text{c}}$)。2 つ目は、 f-bMMI に対して

Table 1 Average WER[%] for isolated speech (development set (si_**dt**_05) and evaluation set (si_**et**_05)) on discriminative feature transformation. (MFCC with LDA+MLLT + SAT+fMLLR) (upper: conventional Single systems (S), upper middle: ROVER among conventional multiple systems (R), lower middle: single Proposed complimentary systems (P), and lower: ROVER including Proposed complementary system (RP))

ID	bMMI	f-bMMI	f-bMMI$_c$ + bMMI$_c$	f-bMMI$_c$ +bMMI	WER (dt)	WER (et)
S5	✓				35.86	29.46
S6		✓			**33.19**	**27.00**
R3	✓	✓			33.80	27.15
P3			✓		35.38	28.27
P4				✓	33.88	27.86
RP3		✓	✓		32.75	**26.60**
RP4		✓		✓	**32.67**	26.62

Table 2 Average WER[%] for isolated speech (si_**dt**_05 and si_**et**_05) on acoustic modeling (DNN). (MFCC with LDA+MLLT)

ID	CE	bMMI	bMMI$_c$	WER (dt)	WER (et)
S7	✓			36.59	30.84
S8		✓		**32.40**	**26.91**
P5			✓	33.09	27.97
RP5		✓	✓	**31.38**	**26.48**

Table 3 WER[%] in terms of SNR[dB] for isolated speech (si_**et**_05) on f-bMMI (S6→RP3) and DNN (S8→RP5).

	−6dB	−3dB	0dB	3dB	6dB	9dB	Avg.
S6	44.14	35.42	28.56	21.46	17.41	**14.98**	27.00
S8	43.86	33.36	28.13	22.01	17.75	16.36	26.91
RP3	43.21	34.24	28.25	21.58	**17.17**	15.13	26.60
RP5	**42.85**	**32.43**	**27.91**	**21.56**	17.75	16.40	**26.48**

だけ、目的関数を修正した場合 (式 (13) かつ式 (5)) である (f-bMMI$_c$ + bMMI)。bMMI と f-bMMI の組み合わせ (R3) は、f-bMMI だけの場合よりも性能が低かったが、提案法と組み合わせることにより (RP3 と RP4) 認識率が向上した。「f-bMMI$_c$ + bMMI」と「f-bMMI$_c$ + bMMI$_c$」の間には顕著な差は見られない。

Table 2 は、MFCC と PLP に LDA+MLLT の特徴量変換を施した場合 (話者適応なし) の DNN の WER である。識別学習により認識率は 4.19%向上し (S7→S8)、提案法と組み合わせでさらに性能が向上した (RP5)。評価セットの場合にも、傾向は類似している。Table 3 は、SNR 毎に WER を調査したものである。S6 と RP3 (f-bMMI の場合)、S8 と RP5 (DNN の場合) を比較すると、ほぼすべての場合で、提案法は WER を改善しており、特に SNR が低い場合に有効である (最大 1.2%)。よって、提案法はさまざまな環境において効果が安定していて頑健であり、音響モデルや識別的特徴量変換といった幅広い系列識別学習に有効であることが示された。

Table 4 WER[%] (CSJ, test set 1) on acoustic modeling (GMM). (MFCC)

ID	ML	bMMI	bMMI$_c$	WER
S1	✓			21.00
S2		✓		**18.64**
R1	✓	✓		18.69
P1			✓	18.81
RP1		✓	✓	18.52
RP2	✓	✓	✓	**18.28**

5.3 CSJ(大語彙連続音声認識)

2 つ目のコーパスである CSJ を用いて性能評価を行った。Table 4 は、GMM システムのテストセット 1 における WER である。この場合にも、従来の ROVER(R1) は単一システムの場合 (S2) よりも性能が低下しているが、提案法は 2 つあるいは 3 つのシステムを組み合わせることで、0.36%の改善が見られた。

6 まとめと今後の課題

システム統合のための一般的な識別学習の枠組みを提案し、補助システムを識別学習の枠組みに基づき構築した。実験により、高騒音下および大語彙連続音声認識タスクにおいて提案法の有効性が示された。さらに音響モデルの識別学習と識別的特徴変換いずれにおいても効果が見られた。今後の課題としては、他の識別学習の手法との組み合わせが考えられる。

参考文献

[1] J. Fiscus, "A post-processing system to yield reduced error word rates: Recognizer output voting error reduction (ROVER)," Proceedings of ASRU, pp.347–354 (1997).

[2] 太刀岡勇気, 渡部晋治, J. Le Roux, J. Hershey, "システム統合のための音響モデルの相互情報量最大化識別学習," 音講論 (春), pp.35–38 (2014).

[3] D. Povey, D. Kanevsky, B. Kingsbury, B. Ramabhadran, G. Saon, and K. Visweswariah, "Boosted MMI for model and feature-space discriminative training," Proceedings of ICASSP, pp.4057–4060 (2008).

[4] K. Veselý, A. Ghoshal, L. Burget, and D. Povey, "Sequence-discriminative training of deep neural networks," Proceedings of INTERSPEECH (2013).

[5] E. Vincent, J. Barker, S. Watanabe, J. Le Roux, F. Nesta, and M. Matassoni, "The second 'CHiME' speech separation and recognition challenge: Datasets, tasks and baselines," Proceedings of ICASSP, pp.126–130 (2013).

[6] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, M. Petr, Y. Qian, P. Schwarz, J. Silovský, G. Stemmer, and K. Veselý, "The Kaldi speech recognition toolkit," Proceedings of ASRU, pp.1–4 (2011).

[7] Y. Tachioka, S. Watanabe, J. Le Roux, and J. Hershey, "Discriminative methods for noise robust speech recognition: A CHiME challenge benchmark," Proceedings of the 2nd CHiME Workshop on Machine Listening in Multisource Environments, pp.19–24 (2013).