社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＳＬＰ－51　（５）
2004／5／27

発話位置依存ケプストラム平均正規化による遠隔発話の音声認識

王　龍標†　北岡　教英†　中川　聖一†

† 豊橋技術科学大学 情報工学系
〒 441-8580　愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: †{xiaowang,kitaoka,nakagawa}@slp.ics.tut.ac.jp

あらまし　遠隔環境において，伝送歪みは音声認識の性能を大きく劣化させる。本稿では発話位置依存ケプストラム平均正規化（Cepstral Mean Normalization：CMN）による頑健な音声認識方法を提案する。まず，部屋をいくつかの区域に分割し，その中心位置からマイクロホンまでの伝達特性を各位置で再生されたスピーカーの音声を用いて事前に計測しておく。音声認識時には，システムはマイクロホンペア間の音声の到着時間差（Time Delay Of Arrival: TDOA）に基づいて三次元の話者位置を推定する。本システムは，四つのマイクロホンをＴ字型に配置し，相互相関法を用いてマイクロホンペアの間の到着時間差を求める。そして，推定した発話位置によって事前に計測した伝達特性を選択し，CMN によって伝送歪みを補正して遠隔発話を認識する。さらに、提案手法では，人間からの発話のケプストラム平均とスピーカーからの発話のケプストラム平均のミスマッチも補正することにより実際の発話に対応する。実験により提案手法は，遠隔環境下で音声認識システムの性能を有効に改善することを示した。
キーワード　発話位置同定，位置依存ケプストラム平均正規化，遠隔発話音声認識

Robust Distant Speech Recognition based on Position Dependent CMN

Longbiao WANG†, Norihide KITAOKA†, and Seiichi NAKAGAWA†

† Department of Information and computer Sciences, Toyohashi University of Technology
1-1, Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan
E-mail: †{xiaowang,kitaoka,nakagawa}@slp.ics.tut.ac.jp

Abstract　In a distant environment, channel distortion may severely degrade speech recognition performance. In this paper, we propose a robust speech recognition method based on position dependent Cepstral Mean Normalization (CMN). At first the system measures the transmission characteristics according to speaker positions from some grid points in the room *a priori* using a loudspeaker. In the recognition stage, the system estimates speaker position in a 3-D space based on the time delay of arrival (TDOA) between distinct microphone pairs. In our system, four microphones were arranged in T-shaped on a plane and the time delay of arrival among the microphones was estimated by using cross-correlation. And then the system selects the transmission characteristics estimated *a priori* corresponding to the estimated position and applies a channel distortion compensation method (CMN) to the speech and recognizes it. In our proposed method, we also compensate the mismatch between the cepstral means of utterances spoken by human and those emitted from the loudspeaker. Experiment results showed that the proposed method improved the performance of speech recognition system in a distant environment efficiently and it could also compensate the mismatch between voices from human and loudspeaker well.
Key words　speaker position estimation, position dependent CMN, distant speech recognition

1. はじめに

現在，音声インタフェースが急速に普及してきている。これまでは，近接マイクロホンを用いてものが主であった。しかしながら，より多くのアプリケーションにおいて自然で使い勝手のよい音声インタフェースを実現するためには，実環境下のハンズフリー音声認識[1]の技術が必要となる。しかしながら，遠隔環境下で，さまざまな距離からの音声を認識することは直接音の減衰や反射音の重畳により性能低下を招く。マイクロホンと音源の距離が1m～2mになると大きく認識精度は低下する[2]。雑音環境下ならば，S/Nが悪くなり，さらに認識は難しくなる。この最も主要な原因は実環境とトレーニング環境のミスマッチである。

このミスマッチを減少するために，入力音声の特徴量を補正するのは重要な方法である。特に，伝達特性のミスマッチを補正する簡易で効果的な解決策として，CMN（Cepstral Mean Normalization）がよく使用されている。CMNは簡単に実現できるために，現在多くのシステムに使われている。しかし，一般的なCMNでは，認識する音声の入力が終了した後に補正パラメータを推定するため，実時間処理向きではない[3]。時間同期型の認識アルゴリズムは本来入力と並行して認識処理が進められるが，一般的なCMNを併用すると並行に処理が行えなくなる。この問題に対し，リアルタイムに認識を行うために，過去に認識した音声から補正パラメータを事前に算出しておき，音声を補正する方法が提案されている[4]。この方法はreal-time CMNと呼ばれている。しかしながら，遠隔環境で異なる発話位置からの伝達特性はそれぞれ異なるため，このreal-time CMNでは話者移動によって発生する伝達特性の急速な変化に追従できず，遠隔ハンズフリー音声認識におけるミスマッチの補正には不十分である。

本稿では，話者位置によって事前推定されたパラメータを用いたreal-time CMN（位置依存CMNと呼ぶ）による頑健な遠隔音声認識手法を提案する。部屋をいくつかの区域に分割しておき、事前に区域の中心位置からマイクロホンまでの伝達特性（位置依存の補正パラメータ）を計測しておく。システムは四つのマイクロホンを使い，三次元の話者位置を推定する。本システムは，四つのマイクロホンをT字型に配置し，相互相関[5]を用いて求めたマイクロホンペアの間の到着時間差によって音源位置を推定する。そして，推定した発話位置によって事前に計測した補正パラメータを選択し，CMNによって音声を補正して遠隔発話を認識する。しかし，本手法では，事前に各区域における発話からケプストラム平均を推定する必要があり，音声収録のコストは非常に大きい。そこで，我々はスピーカーから再生した発話によって補正パラメータを推定した。しかしながら，スピーカーからの再生音の伝達特性と人間が直接発声した音の伝達特性は異なるため，スピーカーからの音声で推定した補正パラメータは人間による実際の発話に対して正しく補正することができない。そこで，人間とスピーカーの近接発話によって推定した補正パラメータを用い，人間による発話のケプストラム平均とスピーカーからの発話のケプストラム平均のミスマッチも補正する2段階補正を行った。

2. 発話位置同定

マイクロホンペアの間の到着時間差によって話者位置を推定することができる[6], [7]。本稿では，到着時間差は相互相関に基づいて求める。一般的方法では，音波を平面波と仮定して音波到来方向を求めているため，特に近接音場において誤差を生じる。そこで，本稿では，話者位置を次のように推定する。

N個のマイクロホン $M_i(i = 1, \cdots, N)$ を座標 (x_i, y_i, z_i) に置き，音源は座標 (x_s, y_s, z_s) とする。このとき，音源と i 番目のマイクロホンの距離は

$$D_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2} \tag{1}$$

となる。音源と，マイクロホン i およびマイクロホン j との距離の差 d_{ij} は次のようになる。

$$d_{ij} = D_i - D_j = c\hat{\tau}_{ij}, \tag{2}$$

ただし，c は音速，$\hat{\tau}_{ij}$ は到着時間差である。到着時間差は相互相関 $R_{x_ix_j}(\tau_{ij})$ から推定する。

$$R_{x_ix_j}(\tau_{ij}) = \frac{1}{T - \tau_{ij}} \sum_{t=\tau_{ij}}^{T} x_i(t)x_j(t - \tau_{ij}) \tag{3}$$

ただし，x_i と x_j はマイクロホンペアによって得る音声信号である。

そして，到着時間差 $\hat{\tau}_{ij}$ は式（4）になる。

$$\hat{\tau}_{ij} = arg\max_{\tau_{ij}} R_{x_ix_j}(\tau_{ij}) \tag{4}$$

三次元話者位置を推定するためには，四つ以上のマイクロホンが必要である。本稿では，四つのマイクロホンを図1のように平面上に配置する。三つのマイクロホンペア（$M1$, $M2$), ($M1$, $M3$), ($M1$, $M4$) を使えば，話者位置が推定できる。一番目のマイクロホン $M1$ は座標系の原点に置き，他の三つのマイクロホンは図1のように同じ平面にT字型に配置する。$M2$, $M3$, $M4$ と $M1$ との距離は同じである。提案した方法で，式(1)，(2) による三つの連立方程式は

$$\sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} - \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2 + (z_2 - z_s)^2} = c\hat{\tau}_{12} \tag{5}$$

$$\sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} - \sqrt{(x_3 - x_s)^2 + (y_3 - y_s)^2 + (z_3 - z_s)^2} = c\hat{\tau}_{13} \tag{6}$$

$$\sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2 + (z_1 - z_s)^2} - \sqrt{(x_4 - x_s)^2 + (y_4 - y_s)^2 + (z_4 - z_s)^2} = c\hat{\tau}_{14} \tag{7}$$

となる。三つのマイクロホンペアは対称であるから，三つの連立方程式は次のように直接解ける。

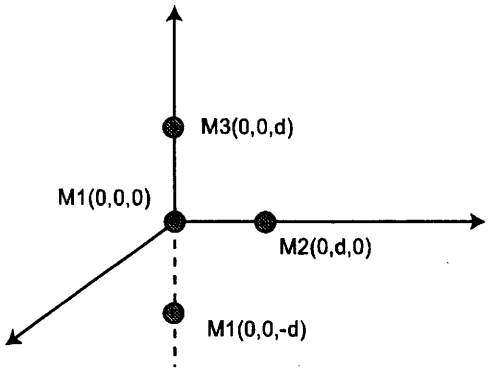

図 1　話者位置を推定するためのマイクロホン配置

$$x = \sqrt{y^2 + a_1 z^2 + b_1 z + c_1} \tag{8}$$

$$y = \frac{-b_2 + \sqrt{{b_2}^2 - 4a_2c_2}}{2a_2} \tag{9}$$

$$z = \frac{-b_3 + \sqrt{{b_3}^2 - 4a_3c_3}}{2a_3} \tag{10}$$

ただし，a_1，b_1，c_1，a_2，b_2，c_2，a_3，b_3，c_3 はマイクロホンペアの距離 d と到着時間差によって決まる値である。

本手法は計算量が少なく，到着時間差の推定が正しい場合推定される音源位置の誤差もなく，マイクロホンからの距離が近い場合にも適用できる。

3. 位置依存ケプストラム平均正規化

3.1 一般的な CMN と real-time CMN

実環境で音声認識を用いる際に考慮すべき雑音として，加算性雑音のほかに乗算性雑音がある。これは音声の伝達特性への影響のことであり，マイクロホンなどの受音系の違いや話者とマイクロホンの距離の違いなどの要因によって生じる。音声認識において学習環境とテスト環境における各ケプストラム係数の平均の差を補正するケプストラム平均正規化（Cepstral Mean Normalization：CMN）[3] は簡易で効果的な伝達特性正規化手法の一つである。この方法は伝達チャンネルと録音装置に起因する歪みを補正することができる。波形領域における乗算性雑音はケプストラム領域において，加算性雑音と考えられる。従って，認識する音声のケプストラム C_t は，乗算性雑音（伝達特性もしくはチャンネル歪み）はケプストラム領域においてテスト環境とトレーニング環境との伝達特性の差を表すパラメータ ΔC を用いて

$$\bar{C}_t = C_t - \Delta C, \qquad (t = 0, ..., T) \tag{11}$$

と補正される。

一般的な CMN において、補正パラメータ ΔC は

$$\Delta C = \bar{C}_t - \bar{C}_{train} \tag{12}$$

として求められる。ここで，$\bar{C}_t$ と $\bar{C}_{train}$ はそれぞれ認識する音声のケプストラムの平均とトレーニング音声のケプストラムの平均である。従って，従来のケプストラム平均正規化手法では，認識する音声の入力が終了し，補正パラメータを算出してからしか認識が開始できない。時間同期型の認識アルゴリズムは入力と並行して認識処理が進められるが，一般的な CMN を併用すると並行に処理が行えなくなり，実時間処理ができない。一般的なケプストラム正規化手法のもう一つの問題は，認識する発話が短い場合ケプストラム平均が正確に推定できないことが挙げられる。

文献 [4] では，チャンネル歪みが急激に変化することのないものであるとの考えに基づいて，次のように改良を行った。この方法では，過去に認識したいくつかの発話から補正パラメータを事前に算出しておく。新しい補正パラメータは

$$\Delta C = \bar{C}_{environment} - \bar{C}_{train} \tag{13}$$

になる。ここで，$\bar{C}_{enviornment}$ と $\bar{C}_{train}$ はそれぞれ特定な環境下で録音された発話のケプストラム平均とトレーニングデータのケプストラム平均である。本稿ではこの方法を従来の CMN と区別し，real-time CMN と呼ぶ。Real-time CMN により、入力音声終了を待たずに認識処理が開始できる。さらに，real-time CMN の補正パラメータは十分な発話のケプストラム係数から推定されるため，一般的な CMN より伝達特性歪みをより正確に補正することができる。

3.2 話者位置同定の real-time CMN への導入

CMN，特に real-time CMN は音声認識システムの性能の改善に有効である。しかしながら，遠隔環境下では，話者とマイクロホンの距離やそれによる部屋の反響などのために，異なる発話位置からの伝達特性は非常に違ってくる。従って，話者位置が変化する場合，伝達特性歪みが急激に変化するために，real-time CMN では十分に対応できない。そこで本稿では，話者位置同定を real-time CMN と併用する位置依存ケプストラム平均正規化 (Position Dependent CMN：PDCMN) を提案する。位置依存ケプストラム平均正規化の補正パラメータは

$$\Delta C = \bar{C}_{position} - \bar{C}_{train} \tag{14}$$

となる。ここで，$\bar{C}_{position}$ と $\bar{C}_{train}$ はそれぞれある特定の位置で発声された発話のケプストラム平均とトレーニングデータのケプストラム平均である。本稿では，部屋は図 2 のように 12 の区域に分割し，各区域の中央から発声された音声を用いてケプストラム平均 $\bar{C}_{position}$ を計測する。

3.3 位置依存 CMN の問題点と解決法

位置依存ケプストラム平均正規化では，各区域の伝達特性を事前に計測しておく必要がある。しかし，各エリアにおいて十分な人数による十分な量の発話を収録するのは非現実的である。したがって，本稿の実験では，各区域において近接マイクロホンで収録した音声をスピーカーから出力し，それを収録して用いた。しかしながら，これらのケプストラム平均はスピーカーから出力した発話によって推定されるため，実際の人間が直接発声した音声とは異なってしまい，このケプストラム平均を用いて人間が発声した実際の発話を正確に補正することができない。そこで本稿では，スピーカーと人間の音声の間のミスマッチを補正するによってこの問題を解決する。この方法の補

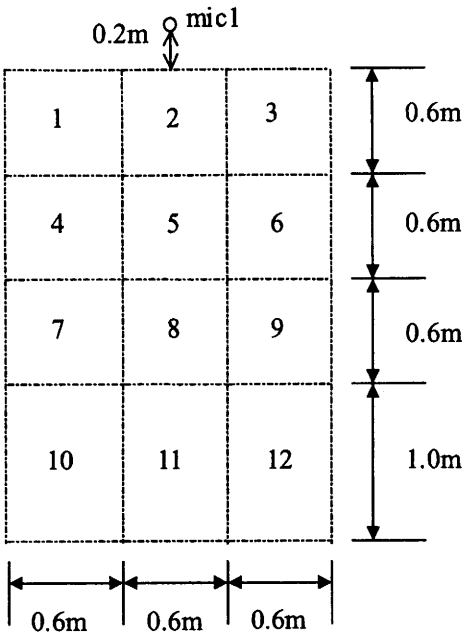

図 2 部屋の構造図（部屋の大きさ：横＝ 3m 縦＝ 3.45m 高さ＝ 2.6m）

正パラメータ ΔC は

$$\Delta C = \{\bar{C}_{position} - \bar{C}_{train}\} - \{\bar{C}_{loudspeaker} - \bar{C}_{human}\}, \quad (15)$$

となる。$\bar{C}_{human}$ と $\bar{C}_{loudspeaker}$ はそれぞれ近接マイクロホンで発声した人間の発話のケプストラム平均とマイクロホンに近接したスピーカーから再生した発話のケプストラム平均である。すなわち，第 2 項はスピーカーの補正項である。この方法では，人間は近接マイクロホンを用いた音声収録のみを行い，位置依存ケプストラム平均を推定するための発話よりかなり少ない発話を用いて $\bar{C}_{human}$ を推定しておけばよい。近接発話は現場での収録も必要なく，収録自体も容易である。

4. 実　　験

4.1 話者位置推定

2 節で提案した話者位置同定実験を行った(注1)。図 1 におけるマイクロホンペア間の距離 d は 20 センチとした。4 マイクロホンで受音した音声は 48KHz で同期収録し，位置を求めた。音声は 0.5 秒程度の文発声である。本実験では，音源位置は (0, 120, -50), (60, 103.9, -50), (84.9, 84. 9, -50), (103.9, 60, -50), (120, 0, -50)（単位は cm）の 4 箇所とし，これらを本手法で推定した結果は (11.9, 150.4, -60.5), (83.2, 124.7, -60.5), (102.9, 109.0, -60.5), (128.7, 76.9, -60.5), (149.1, 15.7, -60.5) となった。この結果は 20cm 程度の誤差があり，あまり正確とは言えない。これは，相互相関に基づいた到着時間差推定の不正確さに起因している。しかし、位置依存ケプストラム平均正規化手法では、話者がどの区域（60cm × 60cm）で発声したのかを推定できればよく，この話者位置推定精度でも十分に有効である。ただし，現在は到着時間差をサンプリング点数単位でしか求めていないために推定が粗くなっているが，より詳細な単位で求めることも可能であり，また到着時間差の計算には CSP 法 [8], [9] などより高精度な方法を導入できるため，さらなる高精度化が可能である。

(注1)：発話位置同定は三次元の位置同定に対応しているが，本稿の位置依存 CMN の実験は高さが固定した平面上での位置に関してのみ行った。

4.2 発話位置依存 CMN の実験条件

実験は 3.45m × 3m × 2.6m の部屋で行った。部屋は図 2 のように 12（3 × 4）個の矩形の区域に分割した。後ろに示すように各区域の中心位置からマイクロホンまでの伝達特性（位置依存の補正パラメータ）をスピーカーから再生した音声のケプストラム平均を求めることにより事前に計測しておいた。提案手法を小語彙（100 単語）孤立単語認識により評価した。本実験では，発話位置に対応する区域は正確に推定できると仮定した。テストデータは東北大・松下単語データベースの 5 話者の発話を用いた。各話者は 200 孤立単語を発声した。各話者の発話に対して，最初の 100 単語はテストデータとして使用し，残りの 100 単語は式 (14) におけるケプストラム平均 $\bar{C}_{position}$ を推定するために使用した。すべての発話は各区域の中心位置においてスピーカーから再生することにより各位置において発声した発話をシミュレートした。サンプリング周波数は 12kHz である。フレーム長は 21.3ms，フレーム周期は 8ms，256 点数のハミング窓を用いた。114 個の 4 状態の left-to-right 音節 HMM を音響モデルとして使用した。HMM のパラメータは，ATR 連続音声テキスト 503 文の 8 話者による読み上げ音声および，ATR 音韻バランス 216 単語の 10 話者による読み上げ音声から切り出した音節により音節毎に学習し，日本音響学会音声データベース（30 話者）および新聞読み上げコーパス JNAS（125 話者）を用いて MAP 推定による連結学習を行った。学習データのケプストラム平均は JNAS コーパスの学習データを用いて求めた。特徴量はメル周波数 LPC ケプストラム係数 10 次元とそれらの Δ，ΔΔ 係数およびパワーの Δ，ΔΔ 係数から構成される，計 32 次元である。

4.3 スピーカー再生音によるシミュレーション実験結果

遠隔環境においてスピーカーから再生された孤立単語の認識実験を行った。表 1 では，提案手法である PDCMN（Position Dependent CMN）をベースライン（CMN なし），一般的な CMN，PICMN（Position Independent CMN）と比較している。一般的な CMN では各発話から求めたケプストラム平均に基づいて補正パラメータを算出した。PICMN では，すべての 12 区域の補正パラメータを平均して算出し，位置に関わらず同一の補正パラメータを用いてケプストラム係数を補正した。なお，PDCMN と PICMN では，5 人に対してそれぞれ共通な補正値（ΔC）を用いた。

表 1 の最後に，各手法のベースラインに対する相対的なエラー削減率（Relative Error Reduction : RER）を示した。まず，CMN を用いない場合，認識率はマイクロホンからの距離が大きくなるにつれて認識性能が大きく劣化していることがわ

表 2　人間による発話の認識実験結果（%）

区域	話者	CMN なし	一般的な CMN	人間の発話による CMN	スピーカー音声による CMN	提案手法
5	1	90.0	77	90.0	86.0	92.0
	2	96.0	85	95.0	97.0	95.0
	3	92.0	81	95.0	96.0	95.0
	4	93.0	81	94.0	97.0	92.0
	5	83.0	71	85.0	86.0	87.0
	平均	90.8	79	91.8	92.4	92.2
9	1	88.0	64	86.0	86.0	86.0
	2	87.0	76	92.0	90.0	92.0
	3	93.0	77	94.0	93.0	93.0
	4	92.0	84	95.0	93.0	94.0
	5	76.0	66	87.0	88.0	85.0
	平均	87.2	73.4	90.8	90.0	90.0
10	1	75.0	62	82.0	74.0	78.0
	2	78.0	78	87.0	85.0	90.0
	3	85.0	68	93.0	87.0	91.0
	4	88.0	77	93.0	93.0	93.0
	5	70.0	59	78.0	76.0	80.0
	平均	79.2	68.8	86.6	83.0	86.4
全平均		85.7	73.7	89.7	88.5	89.5

表 1　*CMN* なし，一般的な *CMN*，位置独立ケプストラム平均正規化（*PICMN*），位置依存ケプストラム平均正規化（*PDCMN*）の単語認識率の比較（%）。

区域番号	CMN なし	一般的な CMN	PICMN	PDCMN
区域 1	90.4	86.4	96.0	95.4
区域 2	91.8	88.8	96.4	96.4
区域 3	86.0	84.4	95.6	96.4
区域 4	81.8	82.2	92.4	92.8
区域 5	83.0	85.2	96.0	95.8
区域 6	79.6	83.0	94.4	94.4
区域 7	80.6	81.0	94.4	95.0
区域 8	83.8	83.2	94.6	95.2
区域 9	82.6	82.2	95.0	94.8
区域 10	68.4	76.4	90.4	92.0
区域 11	68.6	75.2	88.2	88.8
区域 12	68.6	75.8	90.4	92.0
平均	80.4	82.0	93.7	94.1
RER	-	8.2	67.9	69.9

かる。部屋の反響などの影響により伝達特性が大きく変化し，認識に影響していることが確認できる。一般的な CMN を用いた場合，あまり効果が得られていない。特に，マイクロホンから遠い区域では認識率が改善されるが，近い区域では逆に劣化してしまう。本実験で用いたテストデータの発話長が短く，音韻的なバランスが発話内でとれていないため，補正値を求めるためのケプストラム平均の信頼性が低い。したがって，伝達特性の補正効果と音韻バランスの悪さが相殺される形となった。もちろん，長い発話でテストすればあまり問題とならないと考えられるが，実用を考えた場合にコマンド等短い発話が用いられることは多いと考えられ，問題になる。一方，PICMN や PDCMN では，事前に十分な量の発話から補正値を求めておくため，この問題が生じない。すべての区域においてベースラインや一般的な CMN の性能を上回った。さらに，補正値を位置依存にすることによって，位置独立な補正値を用いた場合よりも性能が向上した。その結果，CMN なしと比較して，PDCMN では相対的に 69.9%のエラー削減を達成した。

4.4　人間による直接発話の認識実験

人間が実際に部屋の各位置から発声した発話を認識する実験を行った。5 名の男性話者により区域 5，9，10 から発声された，200 単語（4.3 節の実験と同じ単語）を収録し，人手で切り出して，音響分析し，最初の 100 単語をテストに用いた。また，式（15）のスピーカーと人間の差異の補正値を算出するため，close-talk 条件で人間およびスピーカー再生音の収録も行った。結果を表 2 に示す。人間の発話による CMN とは，収録した 200 単語の後半 100 単語を位置依存補正値の算出に用いた場合であり，理想的な条件といえる。スピーカー音声による CMN とは，4.3 節で用いた位置依存補正値をそのまま用いた場合である。提案手法は，式（15）により位置依存の補正と，スピーカーと話者間の差異の補正を行ったものである。なお，実際の人間の音声を用いて補正を行った実験（人間の発話による CMN および提案手法）では，テストに用いる話者以外の 4 名から求めた補正値（ΔC）を用いて補正を行った。結果より，まず一般的な CMN は発声が短いために発話内容に依存した CMN となり，性能を劣化させていることがわかる。またスピーカー再生音から求めた補正値では，人間の音声による補正に若干劣るが，スピーカーと人間の補正により，その劣化を改善し，ほぼ同等の認識性能が得られることがわかる。

5. ま と め

本稿では，位置依存ケプストラム平均正規化による頑健な遠隔発話音声認識方法を提案した。提案手法は一般的なCMNや位置独立ケプストラム平均正規化により認識率が有効に改善できた。また人間による発話ケプストラム平均とスピーカーからの発話のケプストラム平均のミスマッチの補正方法も提案し，実験により，提案手法は人間による遠隔音声も効率的，効果的に補正できることを示した。

今後は，より正確な音源位置を推定するため，到着時間差の計算にCSP法（Crosspower Spectrum Phase）[8], [9]など，より高精度な方法を導入する予定である。さらに現在は到着時間差をサンプリング点数単位でしか求めていないために推定が粗くなっているが，より詳細な単位で求める予定である。また伝達特性の計測方法としてTSP（time-stretched pulse）法[10]により部屋の異なる区域からのインパルス応答を計測する方法などがある。例えば，文献[11]では，TSP法によって得られたインパルス応答を畳み込んだ音声でHMMをトレーニングすることで，クリーンな音声のHMMより良い性能を達成している。これらの手法との比較検討を行っていく。最終的に，話者位置同定を位置依存ケプストラム平均正規化と統合する予定である。

文 献

[1] B.H. Juang, F.K. Soong, "Hands-free telecommunications", HSC2001, pp. 5–10, 2001.

[2] 中村 哲 "外乱に強い音声認識を目指して", 日本音響学会誌, Vol. 57, No. 10, pp. 662-667, 2001.

[3] S. Furui, "Cepstral analysis technique for automatic speaker verification", J. Acoust. Soc. Amer., Vol. 55, pp. 1204–1312, June, 1974.

[4] N. Kitaoka, I. Akahori, S. Nakagawa, "Speech recognition under noisy environments using spectral subtraction with smoothing of time direction and real-time cepstral mean normalization", HSC2001, pp. 159–162, 2001.

[5] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans. Acoust. Speech Signal Process., ASSP24(4):320–327, Aug. 1976.

[6] Y. Huang, J. Benesty, G.W. Elko and R.M. Mersereau, "Real-Time Passive Source Localization: A practical Linear-Correction Least-Squares Approach", IEEE Trans. Acoust. Speech Signal Process., VOL. 9, pp. 943–956, Nov. 2001.

[7] 岡村 敦介, 稲刈 治英, 島田 正治, "任意配置された複数マイクによる単一音源位置推定", 電子情報通信学会論文誌（A）, Vol. J80-A, No. 11, pp. 2054-2058, Nov. 1997.

[8] M.Omologo, P. Svaizer, "Acoustic Event Location using a Crosspower Spectrum Phase based Technique", Proc. of ICASSP, 1994, Vol. 2, pp. 273–276.

[9] 西浦 敬信, 山田 武志, 中村 哲, 鹿野 清宏 "マイクロホンアレーを用いたCSP法による複数音源位置推定", 電子情報通信学会論文誌, Vol. J83-D-II, No. 8, pp. 1713-1721, 2000.

[10] Y. Suzuki, F. Asano, H.Y. Kim and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses", J. Acoust. Soc. Am., Vol. 97(2), pp.1119–1123, February 1995.

[11] D. Giuliani, M. Matsssoni, M.Omologo and P. Svaizer, "Training of HMM with filtered speech material for hands-free recognition", Proc. of ICASSP, March 1999, Vol. 1, pp. 449–452.