2G2-03

方言音声空間モデルに基づいた 音声特徴量間相関を用いた方言識別法

*伊・達瓦[1,2]、井佐原　均[1]、白井　克彦[2]
[1] *National Institute of Information and Communication Technology ,Japan*
[2] *Advanced Research Institute for Science and Engineering, Waseda University, Japan*

あらまし：本研究は、国や地域によって３種類の異なる書き言葉で表記され、４つの音声によって発声される典型的な多方言言語である「モンゴル語」を研究対象とし、音声入力による方言音声－テキスト自動変換及び方言文間機械翻訳システムの実現を検討している。本論文では、あらかじめ用意された方言音声サンプル量子化距離、予備テスト音声基本周波数F_0 近似直線による推定した時間－速度パラメータ主成分分析を用いて構成した方言空間モデルと入力音声流とのユークリット距離推定により高精度方言判別手法を提案する。方言音声量子距離だけを用いた場合は方言話者判別率は83%程度だったが、候補のサンプルにおいて方言空間モデルを利用した場合は方言話者判別率は97%まで改善し、方言話者認識はほぼ14%上昇した。
キーワード：方言認識、量子化ひずみ、主成分分析、方言話者空間、モンゴル語

1．まえかき

音声波により伝達される情報には、話者が音声によって伝えようとした意図を構成する音声の音韻性のほか、話者による発声器官の違い（個人性）、発声方法の違い（方言性）がなどあり、一般に，静的なスペクトル（母音のフォルマン）パターンの違いや動的な時系列（方言による調音様式）パターンの違いとして音声波に現れる。このような話者や方言による音声の異なり、すなわち、話者による音声スペクトル特徴の変動は、音声の自動認識にとって認識率劣化の大きな要因となる[1][2]。従って、任意の話者の音声を高精度に認識するために、使用者の少量の音声サンプルを用いて認識システムをその話者に適用させる話者適用化音声認識システムに関する研究は近年盛んに行われている。

話者の音響的特徴は話者の違いにより広い範囲に変動するため、一般は把握しにくいが、まず、一定の地域方言範囲での発声特徴を推定してから（話者発話スタイル範囲推定）話者適用化を行うことはもっとも有効的手段だと思われる。

本研究は、図１に概要したように、モンゴル語諸言語音声-テキスト変換及び異なるテキスト間機械翻訳システムの実用化を目指している。本論文では、実時間入力音声に対して、話者の発声方言或いは言語の自動判別手法について検討する。

2．モンゴル語諸方言

アジアのモンゴル高原を中心に分布するモモンゴル語はモンゴル語族に属する言語の１つで、アルタイ語系言語に属する。中央アシアを中心とするモンゴル族によって話されるが、共通した標準モンゴル語(話し言葉と書き言葉)は存在しない。居住地域や国それぞれによって異なる文語や口語が使用され、特別的な人工的な翻訳手段なしに互いに言語的なコミュニケーションはほぼできない。現在でも(図１の例のように)、伝統的モンゴル文字(a)，トド文字(b)、キリル文字(c)という３種類の文字システムが、モンゴル国やロシアカルムク共和国、ブイラト共和国、中国内蒙古自治区、中国新疆自治区オイラトモンゴル地域それぞれによって使用され、国や地域によって開発された Windows 処理システムによってテキストや文書処理が行われている[3].

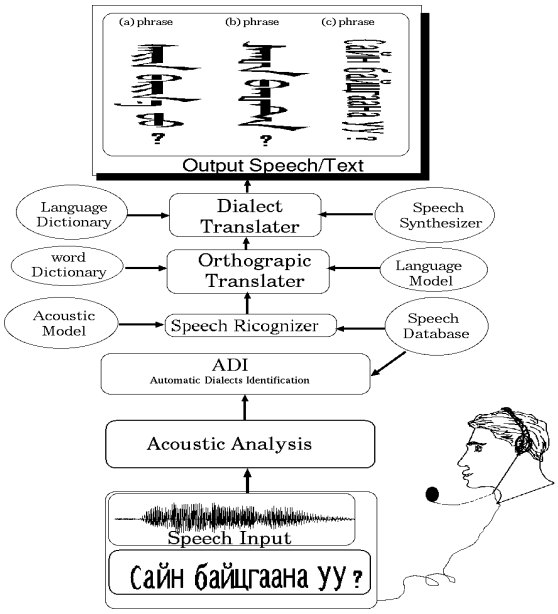


図 1. 目指すシステム概要

* 連絡先：伊・達瓦、京都府相楽郡精華町光台３－５、NICT G,idawa@nict.go.jp

2G2-03

表１：アルタイ系諸言語音響特徴相関

	Ja	Ko	Mo	Ka	Ug
Ja	1.000				
Ko	.973	1.000			
Mo	.924	.887	1.000		
Ka	.940	.865	.910	1.000	
Ug	.963	.913	.940	.980	1.000

表２：モンゴル語諸言語音響特徴相関

	M	I	O	K
M	1.000			
I	.987	1.000		
O	.967	.985	1.000	
K	.884	.924	.935	1.000

モンゴル語諸方言音韻の静的なスペクトルパターンの違い、音声―言語的な差違および音声認識における問題点に関して、我々は先行の研究[4][5][6]において検討し報告した。本論文では、多変量統計的な分析により方言間の相関を分析し、実時間音声入力によるモンゴル語方言話者識別システムの実用化について検討した。

2.1 言語音声相関分析

モンゴル語諸言語間の音声的な差違を説明するため、まず、モンゴル語に一番近いといわれる[7]幾つかのアルタイ系言語である Japanese(Ja), Korean(Ko), Mongolian(Mo), Kasak(Ka)及び Uigur(Ug) 諸言語音声特徴相関を調べてみた。表１はその結果を示すもので、図２は例えばモンゴル語を基準点にした場合、モンゴル語と他言語との相関を見やすく描いたものである。

同様な実験をモンゴル語諸言 Mongolian(M), Inner Mongolian,China (I), Oirat Mongolian, China(O)及び Kalmykia, Russia(K)において行うと表２及び図３が得られる。すなわち、図３より、モンゴル語諸言語間音響的な差違が相当あることがわかる。

2.2 方言音声発話速度に関する分析

音声入力によって発話話者地域方言を推定するには、上述した方言音声サンプルと入力音声特徴量との相関距離だけを用いた場合は高精度な識別率が得られない。同じ話者であっても発話速度や発話気分などの発話スタイ

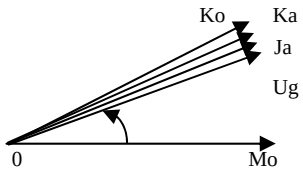


図２.モンゴル語とアルタイ系諸言語相関距離

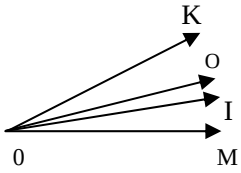


図３.モンゴル語諸言語間音響的相関距離

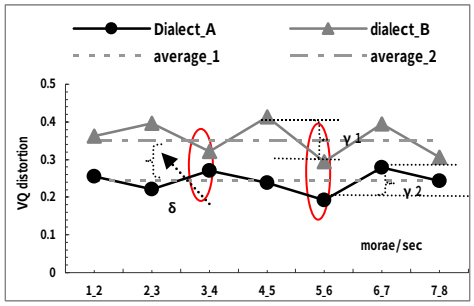


図４発話速度異なる時話者サンプルとの距離

ルが変わると図４の丸線で例に示したように、入力とサンプルとの距離が大きく変動する。普通は、話し言葉音声の発話速度は朗読音声と比較して変動幅が大きく、その影響で認識誤りが生じやすいという調査結果がある[8]。そのため、方言音声の発話速度分布の傾向を調査することは、方言認識性能の向上に有効であると考える。モンゴル語諸方言音声発話速度の違いを、各方言ごとに男女性６名それぞれが同内容の５０ 文を発声した場合の音素ラベルが付与された音声データを用いて調べた。表３はテスト音声方言ごとに含まる音素数である。

表３：音声データの音素サンプル数

Dialect	Vowel	Consonant
Mongolian	490	606
Inner	467	635
Oirat	439	650

図５に方言による母音・子音の発話速度分布の中位値を、図６には各方言音声の発話速度分布を考査した結果を示した。

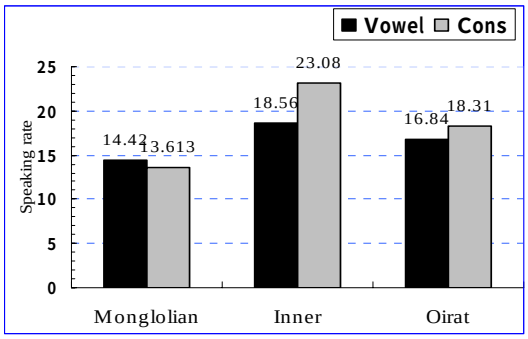


図５：方言発話速度の中位値

図６：方言音声発話速度分布（母音）

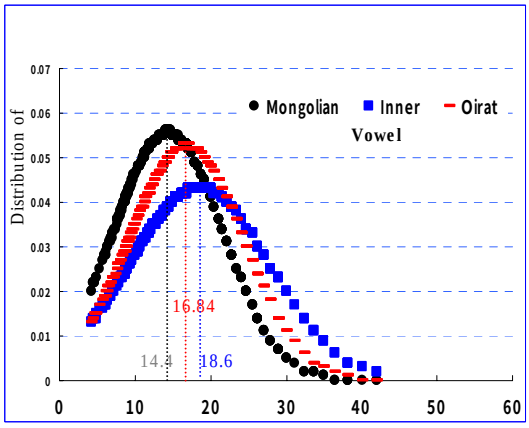
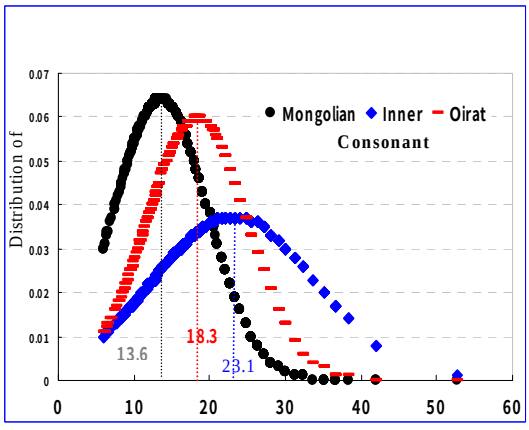


図６：方言音声発話速度分布（子音）



上表 3 及び図 5,6 より例えば、モンゴル国地域発声(Mongolian)に対して、他の各方言音声の発話速度が大きくなっていることが分かる。

3.発声音声による方言自動判別手法

モンゴル語は地方や国によって発音上差違が大きい一方、それぞれ地方の発音は異なる文字システムによって表記されるため、音声入力方式の機械翻訳の前段階では、話している話者の声において言語や方言の自動判別が必須となる。一方、実時間入力音声流に対して話者の発話速度は直接把握するのは難しいということから、本研究では、まず、方言複数話者音声サンプル量子化コードブックベクトル C_i^L と入力音声特徴量 x_j との最小距離を式(1)を用いて計算して、入力音声流に最も近い二つの方言モデルが候補対象として選択する。次は、方言音声サンプルデータ主成分分析に基いて作成された候補方言空間モデルと入力音声流基本周波数 f_0 を用いて抽出されたパラメータベクトル間との最小ユークリット距離選択によって最終的な方言が決定される方針が採用する。

$$D = \frac{1}{M}\sum_{j=1}^{M}\min_{1\le i\le V}\left[d\left(C_i^L, x_j\right)\right] \quad (1)$$

3.1 方言空間

方言空間モデル概要は図 7 に示す。この場合、方言の特徴ベクトル空間は 2 次元で、主成分分析により方言空間は 1 次元の直線として得られている。方言話者一人はこの直線の回りの部分空間上の一点として存在し、方言空間上に射影した点となり、方言全話者は直線上に分布すると考える。入力音声特徴ベクトル x_j とある方言候補サンプルとの最小量子化距離 D_k 及び基本周波数パラメータ $\Delta f_i / \Delta T_i$ が方言空間中ある点 P に位置するとき、方言空間上点 P' 又は P”までのユークリット距離 $\|P - P'\|$ が最小となる方言モデルは最終判別結果として決定される。

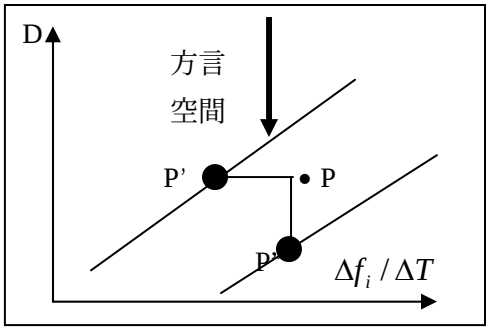


図 7.方言空間概要

方言空間主成分分析変数としては、方言複数話者テスト音声と各方言サンプルコードブックとの量子化距離 D_d、テスト音声基本周波数 F_0 近似線正負傾き及びそれぞれの変化率 $\Delta f_i / \Delta T_i$（図 8 に示す）を用いた。

3.2 方言判別アプローチ

①前処理

1)方言それぞれにおいて、複数話者発声データを用いて、量子化を行い、方言サンプルモデルコードブック C_i^d を作成する。

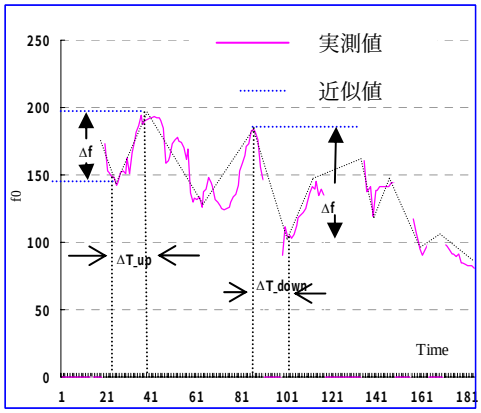


図 8.基本周波数パラメータ抽出

2)方言ごと複数テスト話者音声を用いて音声特徴量を作成し、式(1)によってサンプルコードブックと最小量子化距離を算出する。
3)テスト音声ごとに、基本周波数 F_0 近似線正負傾き及びそれぞれの変化率 $\pm\Delta f_i / \Delta T_i$ を算出し、パラメータ全てにおいて平均値を引いてデータ正規化を行う。
4)以上の各パラメータを用いて方言音声主成分分析を行い、各方言モデル空間を作成する。
②方言判別
1)被判別話者入力音声に対して、式(1)を用いて各方言サンプルコードボックと入力音声特徴量との最小量子化距離を計算され、距離が最小となる隣接二つのサンプルが候補者として選択される。
2)入力音声流に対して基本周波数パラメータを抽出する。
3)入力音声量子化距離 D_k 及び基本周波数パラメータ $\pm\Delta f_i / \Delta T_i$ ベクトルを $\mathbf{X}$、方言サンプルデータパラメータベクトルの平均ベクトルを $\bar{X}$、主成分を表現する個有ベクトルを並べて方言空間への変換ベクトル S を、

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \vdots & \vdots & & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mk} \end{bmatrix}$$

とすると、$\mathbf{X}$ から方言空間への距離 $D_{pp'}$ は

$$D_{pp'} = \left\| (I - S^T S)(X - \bar{X}) \right\| \quad (2)$$

で与えられる。ここで、k はテスト音声数、m は主成分数である。また、I は $m \times k$ の単位行列を表す。
③方言判別実験
1)実験用データとしては、本研究作成のモンゴル語多言語音声データベース MLDSD[9]中 ML_175(175 個短文)を用いた。各方言 10 名話者発声した 50 短文を方言音声サンプルデータとして利用し、方言ごとにコードブック及び主成分パラメータ D_i, F_0 を作成した。
2)各方言において、ほかの 5 名話者に、上述発声短文と同様及び異なる 50 文それぞれを発声させ、テストデータとした。
3)方言判別実験は方言音声量子化歪みのみにより実験と提案した方言空間モデルによる実験両方で close, open それぞれの方式で行った。デモによる実験結果を表 4 にまとめた。今回の実験では、k=4,m=2 とした。

表 4 方言判別	Close(%)	Open(%)
量子化おんみ	91.3	81.7
提案法	93.2	97.3

4．結び

方言音声空間モデルに基づいて話者発話声の発話スタイル、すなわち、話者口語方言を自動判別手法を提案した。方言判別はモンゴル語 4 つの方言において検討した。方言音声量子化歪み尺度のみを用いた場合と比べ、音声主成分分析を用いた方言空間モデルを利用した方法では、実時間音声入力方言判別率ははるかに上昇することを確認した。

文献

[1]C.-H Lee,. *al*, “A study on speaker adaptation of the parameters of continuous density hidden Markov models”, IEEE Trans. Acoust. Speech Signal Processing., Vol. ASSP-39, No.4, pp806-814 April 1991.
[2]山本 一公、中川 聖一、など、“発話スタイルの違いが音声認識及ぼす影響についての検討”、信学技報、SP99-31,1999.
[3]“中国少数民族多文種情報処理研究会論文集”、中国内蒙古大学、2004.8.
[4]Idomuso Dawa, Katsuhiko Shirai , *al*,. “Acoustic Feature analyses of Mongolian dialects by computer”, ACTA ACUSTIC Vo1.24, No.1 Jan.1999, pp94-97, China.
[5] Idomuso Dawa, Shigeki Okawa, Katsuhiko Shirai , “Analyzing and Classifying Mongolian Mejor Dialects by Acoustic and Prosodic Features”, Minority Languages of China, Vol. No1. 2001.1,pp26-32.
[6]I.Dawa, Shigeki Okawa, Katsuhiko Shirai, “Inquiry into a Common Acoustic Model to Realize Mongolian Dialectal Speech”, Journal of the central University for Nationalities, Vol.28,No.4 2001,pp114-121.
[7] T.Ehara, ”Mongolian to Japanese machine translation using Chasen “present at the China Japan Natural Language Processing Promotion Conference 2004, Nov., 2004.
[8]古井 貞熙など、“話し言葉認識における決定木を用いた誤り要因の分析”、日本音響学会研究発表会講演論文集、1-1-9(2001-10)
[9] I. Dawa, S. Okawa, K. Shirai “Design of Mongolian speech database considering dialectal characteristics”, Acoust.Soc.Jpn. (E) 20, 3 (1999).