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

2005－ＳＬＰ－59（30）
2005／12／22

量子化LSPパラメータを用いた雑音下音声認識のAURORA-2Jによる検討

森田 義則† 亀田 康介‡ 船田 哲男‡ 野村 英之‡

†石川高専 〒929-0392 石川県河北郡津幡町北中条タ1
‡金沢大学自然科学研究科 〒920-1192 石川県金沢市角間町

E-mail: †morita@ishikawa-nct.ac.jp, ‡{kameda@oak.ec.,funada@, nomu@}t.kanazawa-u.ac.jp

あらまし 分散型音声認識において，音響分析にはメルケプストラムからなる特徴量を用いることが勧告されているが，音声符号化の分野ではLSPが特徴量として広く用いられている．したがってLSPを用いた場合の音声認識の性能を調べておくことは必要である．Aurora-2JによりMFCCとLSPを比較した結果，マルチコンディションで認識率80.0%が得られ,相対比で-42.2%の結果となった.また,12ビット量子化において70.5%の認識率が得られた.

キーワード LSP，AURORA-2J，ベクトル量子化

A study on noisy-speech recognition performance using quantized LSP parameters based on AURORA-2J database

Yoshinori MORITA† Kousuke KAMEDA‡ Tetsuo FUNADA‡ and Hideyuki NOMURA‡

†Ishikawa N.C.T. ta-1 Kitachujo, ,Tsubata-machi, Kahoku-gun, Ishikawa, 929-0392 Japan
‡Faculty of Engineering, Kanazawa University Kakuma-machi, Kanazawa-shi, Ishikawa, 920-1192 Japan

E-mail: †morita@ishikawa-nct.ac.jp, ‡{kameda@oak.ec.,funada@, nomu@}t.kanazawa-u.ac.jp

Abstract In distributed speech recognition (DSR), it is recommended that the MFCC (Mel Frequency Cepstrum Coefficient) is used for speech analysis. However, LSP is widely used for quantity of characteristic in speech encoding. Therefore, it is necessary to examine performance of speech recognition when LSP is used. As a result of having compared LSP with MFCC by Aurora-2J database, a recognition rate of 80.0% are provided with multicondition, relatively becomes -42.2%. And a recognition rate of 70.5% is provided in 12 bits quantization.

Keyword LSP, AURORA2J, Vector Quantization

1. はじめに

携帯電話の急速な普及に伴い，端末側では音響分析のみを行い，サーバ側で認識処理を行うという分散型音声認識(DSR; Distributed Speech Recognition)が提案されている[1]．音響分析では，メルケプストラムからなる特徴量を用いることが勧告されているが，音声符号化の分野ではLSP(Line Spectrum Pair)が特徴量として多く用いられている．

LSPには，補間特性が優れているという特徴があり，符号化（量子化）された特徴量に対しても，量子化されていない特徴量を用いた場合と同等の認識性能が期待できる[2]．

LSPパラメータを特徴量として用いた場合の認識性能をAURORA-2J[3]で評価した．NNVQ法[4]を用い，8ビット量子化雑音抑圧LSPパラメータを用いた場合の認識結果は十分な値が得られなかったので，コードブックを用い，ビット数を変化させベクトル量子化されたLSPパラメータを用いた場合の認識性能の検討を行った．

表 1 分析条件

サンプリング周波	8kHz
分析次数	8次
フレーム長	30ms
フレーム周期	30ms

2. 実験方法の概要

符号化ビット数としては8,10,12ビットを仮定し，比較に用いた符号化法はコードブックを用いる方法とニューラルネットワーク(NN)を用いる方法である．

コードブックの作成は学習用のAURORA-2Jの雑音Exhibitionに対応するクリーンな音声を用いた．分析条件を表1に示す．得られたLSPよりコホーネン自己

組織化特徴マップによりコードブックを作成した．

認識のための AURORA-2J の音声からの LSP パラメータの抽出の分析条件は表 1において，フレーム周期を 10ms とする．

HMM の学習は，コードブックまたは NN による符号化後の LSP パラメータより行った．学習およびテストに用いた LSP パラメータは，各 LSF の値から，その次数の平均値を引いた 8 次元の値を用いた．

学習と認識には AURORA-2J の標準スクリプトを用いており、評価カテゴリは 0（バックエンドの変更なし）である．

3. 認識結果

表 2(a)に量子化なし LSP パラメータとエネルギー項を用い，Δ と ΔΔ を加えた 27 次元の場合の結果を示す．Clean Training で 40.0%，Multicondition training で 80.0% という認識精度となり，Clean Training でベースラインの値より 11.4%，Multicondition training で 42.2% 低下した．Δ と ΔΔ を加えない場合は，それぞれ認識率 35.4%と 58.7%となった．

表 2(b)(c)(d)に量子化した場合の結果を示す．コードブックを用い，12 ビット量子化した場合は，それぞれ 26.0%と 70.5%となった．量子化なしの(a)よりかなり悪く，量子化ビット数が足りないためと考えられる．

SNR 20dB の音声で学習した 3 層のニューラルネットワークを用い，LSP パラメータを 8 ビット量子化した場合の認識結果はそれぞれ 9.1%と-0.1%となった．雑音抑圧 NN を用いた認識結果は，雑音抑圧の効果が得られず，非常に低い値となった．雑音抑圧処理を行っても認識精度の向上が得られなかった理由として，男女および多人数による特徴量の変動，またいろいろな種類の雑音による特徴量の変動にニューラルネットワークの学習がうまく適応できなかったためと思われる．

4. 結果の検討

図 1に量子化ビット数を変化させたときの認識率を示す．Δ と ΔΔ を用いると認識率の向上が得られることがわかる．12 ビット量子化で低い認識率となったのはコードブックの作成時にデータが少なかったことが原因と考えられる．

提案法の NNVQ 法を適用し評価を行ったが，低い認識率しか得られなかった．解決策としてはニューラルネットワークのサイズを大きくし，符号化ビット数を増やすことが考えられ，どの程度のビット数が必要かをさらに検討することが今後の課題である．

5. まとめ

LSP パラメータを用いたとき，マルチコンディションで認識率 80.0%が得られた．また,12 ビット量子化において 70.5%の認識率が得られた

本研究では，IPSJ SIG--SLP 雑音下音声認識評価 WG の雑音下音声認識評価環境(AURORA-2J)を利用した．

参 考 文 献

[1] ETSI standard document, Speech processing, Transmission and Quality aspect(STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithm, ETSI ES 202 050 v.1.1.3 (2003-11).

[2] A. Bernard, A. Alwan, Source and Channel Coding for Remote Speech Recognition over Error-Prone Channels, ICASSP-2001, Vol.4, pp.2613-2616.

[3] 山本一公ら，“AURORA-2J/AURORA-3J データベースとその評価ベースライン，” 情報処理学会研究報告, SLP-47-19, 2003.

[4] 森田，船田，野村，“LSP パラメータを用いた雑音下音声認識の AURORA-2J による評価，”音講論集, 1-1-18, pp.35-36, Sep. 2004.

表 2 各種方式の比較

(上から(a)量子化なし，(b)12 ビット VQ，(c)10 ビット VQ, (d)8 ビット VQ)

%Acc				
	A	B	C	Overall
Clean Training	43.44	32.98	47.34	40.04
Multicondition training	88.74	74.71	73.03	79.99
Average	66.09	53.85	60.19	60.01

%Acc				
	A	B	C	Overall
Clean Training	28.92	16.95	38.14	25.98
Multicondition training	74.36	71.84	60.05	70.49
Average	51.64	44.40	49.09	48.23

%Acc				
	A	B	C	Overall
Clean Training	26.99	17.60	36.16	25.07
Multicondition training	73.62	71.74	59.28	70.00
Average	50.30	44.67	47.72	47.53

%Acc				
	A	B	C	Overall
Clean Training	17.95	4.44	31.36	15.23
Multicondition training	68.25	65.39	49.73	63.40
Average	43.10	34.92	40.54	39.31

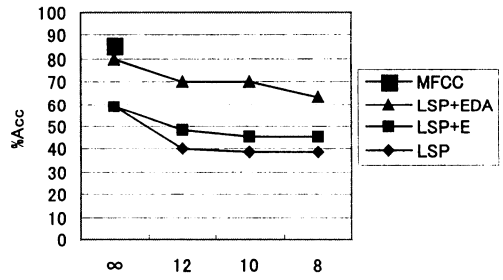

図 1 量子化ビット数による変化

(MFCC:ベースライン, LSP+EDA:27 次元 LSP+E:9 次元, LSP:8 次元ベクトル)