LG-004

狭帯域–広帯域予測モデルに基づく帯域スケーラブル LSP 量子化
Scalable LSP quantization based on a narrowband–wideband prediction model

江原 宏幸 † 押切 正浩 † 吉田 幸司 †
Hiroyuki Ehara Masahiro Oshikiri Koji Yoshida

1. まえがき

スケーラブル音声符号化は、コアとなる基本レイヤと少なくとも 1 つ以上の拡張レイヤとを用いて階層的に音声信号を符号化する方式である。階層的に符号化を行なうため、全レイヤの符号化情報がなくても音声信号を復号することができる。このようなスケーラビリティの特徴は、回線品質の異なるネットワークや再生品質の異なる端末間で多地点通信を行なったり、優先制御と組み合わせて VoIP(voice over IP) におけるパケット消失対策技術として用いたりするのに適しており、これまでにも色々な方式が検討されている [1–4]。

スケーラブル符号化には、符号化する信号の帯域は変わらずにビットレートがスケーラブルになっているもの [1,2] と符号化する信号の帯域幅がスケーラブルになっているもの [2–4] があるが、ブロードバンドネットワークの普及により、音声信号の広帯域化を実現する後者のタイプやステレオ化も考慮したスケーラブル符号化が次世代向けの音声符号化技術として注目されている [5]。本稿では、帯域スケーラブル音声符号化向けの帯域スケーラブル LSP 量子化の検討結果について報告する。

2. 帯域スケーラブル LSP 量子化

帯域スケーラブルな LSP 量子化方式としては、文献 [2,6] のものが挙げられる。これらの方法では、狭帯域 LSP に予測係数を乗じたものを広帯域 LSP の量子化に用いる。文献 [2] では狭帯域 LSP10 次を 0.5 倍したものを広帯域 LSP の低次 10 次の予測値として用いる。文献 [6] はそれを一般化し、フレーム間予測（MA 予測）とフレーム内予測（狭帯域 LSP から広帯域 LSP の低次を予測）に用いる係数の同時最適化を行なうことで性能改善を図っている。文献 [6] では、広帯域 LSP の量子化値を (1) 式で表している。

$$\hat{f}_n(i) = \sum_{p=0}^{P} \alpha_p(i)\hat{l}_{n-p}(i) + \beta(i)\hat{g}_n(i), \quad 1 \le i \le M \qquad (1)$$

ここで、$\hat{f}_n(i)$ は第 n フレームにおける量子化された広帯域 LSP（i 次）、P は MA 予測次数、$\alpha_p(i)$ は i 次 LSP パラメータに対する MA 予測係数（p 次）、$\hat{l}_{n-p}(i)$ は第 $n-p$ フレームにおける LSP コードベクトル（i 次）、$\beta(i)$ は狭帯域 LSP（i 次）から広帯域 LSP（i 次）を予測する係数、$\hat{g}_n(i)$ は第 n フレームにおいて量子化された狭帯域 LSP（i 次）、M は広帯域 LSP 分析次数である。なお、狭帯域 LSP の分析次数は普通広帯域 LSP の分析次数より小さいので、広帯域 LSP の分析次数に変換したものを $\hat{g}_n(i)$ として用いる。

†松下電器産業（株）次世代モバイル開発センター

$$D_n = \sum_{i=1}^{M} w_n(i)(f_n(i) - \hat{f}_n(i))^2 \qquad (2)$$

LSP の量子化は、ターゲットとなる広帯域 LSP と量子化 LSP との重みづき二乗誤差 D_n（n はフレーム番号）が最小となるように行なわれる（(2) 式）。

予測係数セット $[\alpha_p(i),\beta(i)]$ と LSP コードベクトルの符号帳の学習は、それぞれ (3) 式、(4) 式を用いて行なう [6]。なお、(4) 式中、$c_j(i)$ はインデックス番号 j のコードベクトル、I_n は第 n フレームにおいて（符号化により）選ばれたコードベクトルのインデックス番号である。

$$\begin{bmatrix} \alpha_0(i) \\ \vdots \\ \alpha_P(i) \\ \beta(i) \end{bmatrix} = \boldsymbol{\Phi}^{-1} \begin{bmatrix} \sum_n w_n(i)\hat{l}_n(i)f_n(i) \\ \vdots \\ \sum_n w_n(i)\hat{l}_{n-P}(i)f_n(i) \\ \sum_n w_n(i)\hat{g}_n(i)f_n(i) \end{bmatrix} \qquad (3)$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \sum_n w_n(i)\hat{l}_n(i)\hat{l}_n(i) & \cdots & \sum_n w_n(i)\hat{l}_n(i)\hat{g}_n(i) \\ \vdots & \ddots & \vdots \\ \sum_n w_n(i)\hat{l}_{n-P}(i)\hat{l}_n(i) & \cdots & \sum_n w_n(i)\hat{l}_{n-P}(i)\hat{g}_n(i) \\ \sum_n w_n(i)\hat{g}_n(i)\hat{l}_n(i) & \cdots & \sum_n w_n(i)\hat{g}_n(i)\hat{g}_n(i) \end{bmatrix}$$

$$c_j(i) = \frac{\sum_{n:I_n=j} w_n(i)\alpha_0(i)\left(f_n(i) - \sum_{p=1}^{P} \alpha_p(i)\hat{l}_{n-p}(i) - \beta(i)\hat{g}_n(i)\right)}{\sum_{n:I_n=j} w_n(i)\alpha_0^2(i)} \qquad (4)$$

本検討では、上記従来の帯域スケーラブル LSP 量子化方式において、$\hat{g}_n(i)$ から $f_n(i)$ をなるべく高い精度で予測することにより量子化性能を改善することを目的とする。$P = 0$ とし、広帯域 LSP の量子化において MA 予測は用いない構成で検討を行った。

3. 提案法

3.1 提案するスケーラブル量子化法

前章のように、従来技術では狭帯域 LSP に係数を乗じて広帯域 LSP の量子化に用いている。本稿では、この係数（(1) 式の $\beta(i)$）を狭帯域–広帯域予測係数とみなし、その予測精度を上げることを考える。従来技術では、$\beta(i)$ は (3) 式を用いて設計された固定値であるが、フレーム毎に可変とした方が予測精度を改善できると考えられる。従来技術でも、係数セットを複数用意して切り替える構成が可能であるが、係数セットの符号帳にビットを配分する必要があり、LSP コードベクトルと係数セットへのビット配分のトレードオフになる。そこで本稿では、予測係数情報へのビット配分を必要最小限とし、毎フレームで適応的に狭帯域–広帯域予測係数を決定できる方法を提案する。具体的には直前のフレームにおける量子化

広帯域 LSP と量子化狭帯域 LSP との比を現フレームにおける狭帯域–広帯域予測係数の要素として用いる。式で表すと (5) 式のようになり、(1) 式の $\beta(i)$ を定数成分 $\beta'(i)$ と変動成分 $\frac{\hat{f}_{n-1}(i)}{\hat{g}_{n-1}(i)}$ の積として表す（ただし前述のように本検討では $P=0$ としている）。学習は文献 [6] と同様の方法で行なった。即ち、(3) 式、(4) 式における $\beta(i)$ を $\beta'(i)$、$\hat{g}_n(i)$ を $\left(\frac{\hat{f}_{n-1}(i)}{\hat{g}_{n-1}(i)}\right)\hat{g}_n(i)$ として学習を行なう。予測係数の初期値には、$\alpha_0(i)=\beta'(i)=1.0$ を用いた。LSP の予測残差（狭帯域 LSP から広帯域 LSP を予測した際の予測残差）ベクトルの初期符号帳は、予測係数を前記初期値にして LBG アルゴリズムを用いて作成した。

$$\hat{f}_n(i)=\sum_{p=0}^{P}\alpha_p(i)\hat{l}_{n-p}(i)+\beta'(i)\left(\frac{\hat{f}_{n-1}(i)}{\hat{g}_{n-1}(i)}\right)\hat{g}_n(i),\qquad 1\le i\le M \tag{5}$$

3.2 狭帯域 LSP 量子化器

狭帯域 LSP 量子化器は、25 ビット/フレーム（1 フレーム＝ 10ms）の MA 予測型の 2 段分割ベクトル量子化器で、MA 予測係数に 1 ビット、初段 LSP 符号帳（12 次）に 8 ビット、2 段目の 1–4 次、5-8 次、9-12 次にそれぞれ 6 ビット、5 ビット、5 ビット、を割り当てたものを用いた。後述する拡張レイヤの量子化器学習用の 7kHz 帯域音声データをダウンサンプルしたデータ（秒）に対する客観性能は平均スペクトル歪（SD）が 1.06dB であった。

3.3 拡張レイヤ LSP 量子化器

拡張レイヤ LSP 量子化器は、2 種類（バックワードモード切替型とフォワードモード切替型）を比較した。どちらの量子化器も 3.1 節で説明した予測を用いる 21 ビット/フレームの分割ベクトル量子化器であり、予測残差ベクトル（18 次）を 6 次元ずつ 3 分割でベクトル量子化する。また、どちらも (5) 式による予測を用いるモードと、(1) 式の従来法を用いるモードの 2 モード切替型である。モード切替方法が 2 つの量子化器で異なる。バックワードモード切替型では、量子化狭帯域 LSP が定常か非定常かによりモード切替を行なう。定常部でのみ (5) 式に基づく予測を用いる。モード判定は量子化狭帯域 LSP パラメータを用いて行なわれるので、モード情報の伝送は不要である。一方、フォワードモード切替型では、(5) 式による提案法のモードと (1) 式による従来法のモードの双方で量子化処理が行なわれ、歪が小さくなるモードの量子化結果が量子化器の出力となる。即ちモード情報に 1 ビット配分する。LSP コードベクトル符号帳のビット配分は、低次側から順に、バックワードモード切替型は 3，9，9 ビット、フォワードモード切替型は 3，9，8 ビット（モード情報に 1 ビット）である。

4. 実験

予測係数および LSP コードベクトル符号帳の学習は、7kHz 帯域の日本語クリーン音声 53200 フレーム（532 秒、10 話者 200 短文章）を用いて行なった。学習データに対する客観性能比較結果を表 1 に示す。提案する予測モデルの適用によって、いずれのモード切替型でも SD 性能が 0.2dB 以上改善している。モード切替単独の効果も確認するため、従来法にモード切替のみを導入（適応予測（(5) 式）は行なわない）した場合についても性能比較を行なった。結果を表 2 に示す。表 2 において、mode0 は非定常モード、mode1 は定常モードである。表 2 から分かるように、モード切替単独による SD 性能の改善は 0.04～0.05dB であるのに対し、適応予測の SD 性能改善効果は 0.16～0.18dB あることが分かる。また、適応予測（(5) 式）が適用される定常部（mode1）に関して言えば、適応予測による SD 性能の改善効果は 0.20～0.22dB である。これらの結果より、提案する「狭帯域 LSP →広帯域 LSP」予測の導入効果が確認できた。

表 1: 学習データに対する SD 性能比較

方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.65	76.2%	23.7%	0.1%
提案法 B*	1.45	84.2%	15.8%	0.0%
提案法 F*	1.42	84.9%	15.0%	0.1%

* B=バックワード型、F=フォワード型

表 2: モード切替と適応予測の効果

方式	SD[dB]		
	全平均	mode0	mode1
従来法	1.65	—	—
+モード切替	1.61	1.83	1.53
+適応予測 (提案法 B*)	1.45	1.82	1.31
+モード切替	1.60	1.69	1.53
+適応予測 (提案法 F*)	1.42	1.64	1.33

* B=バックワード型、F=フォワード型

学習外データに対する客観性能についても比較した。比較に用いたデータは、クリーン音声（64 秒、8 話者 8 文章対）および背景雑音条件（カーノイズ 15dB、64 秒、8 話者 8 文章対）である。結果を表 3 に示す。結果より、学習外データに対しても従来法と比較して 0.19～0.23dB の SD 性能改善が得られていることが分かる。全般的にフォワードモード切替型の方がバックワードモード切替型よりも若干（0.04dB）性能が良い。

また、フレーム消失条件（フレーム消失率（FER）1%、2%）についても SD 性能を確認した。音声データには前述のクリーン音声を使用し、100、50 フレームに 1 回周期的にフレームを消失させた。また、拡張レイヤのみが 100、50、25 フレームに 1 回周期的に消失する場合についても確認した。結果を表 4 に示す。結果より、バックワードモード切替型は誤り率の上昇とともに急速に性能が劣化するが、フォワードモード切替型は FER2%では従来法より高い性能（平均 SD、SD が 4dB を超えるフレームの割合）を維持していることが分かる。例えば、バックワードモード切替型は FER1%で SD が 4dB を超えるフレームの割合が従来法を上回っているが、フォワードモード切替型では FER2%でも平均 SD 性能で 0.16dB の差を維持するとともに、SD が 4dB を超えるフレームの割合も従来法以下である。また、拡張レイヤのみのフ

表 3: 学習外データに対する SD 性能比較

クリーン音声				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.72	75.8%	23.9%	0.2%
提案法 B*	1.53	83.8%	16.0%	0.1%
提案法 F*	1.49	85.6%	14.3%	0.1%
背景雑音条件（カーノイズ）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.54	86.6%	13.4%	0.1%
提案法 B*	1.35	91.5%	8.5%	0.0%
提案法 F*	1.31	93.4%	6.6%	0.0%

* B=バックワード型、F=フォワード型

レーム消失とした場合、バックワードモード切替型ではモード判定誤りがなくなるため（コアレイヤも消失する場合に比べて）性能劣化が小さくなるが、依然として誤り率の上昇による性能劣化は大きく、FER4%で従来法に劣る結果となった。

5. 考察

バックワードモード切替型の提案法（提案法 B）は、誤りに弱いことが確認されたが、前章の実験結果より、拡張レイヤのみのフレーム消失にする（提案法 B のモード誤りをなくす）ことで、性能が大きく改善（FER2%で SD0.15dB）されることが分かる。このことから、提案法 B が誤りに弱い主要因はモード誤りを発生することであると考えられる。実際、提案法 B のモード判定誤り発生数はフォワードモード切替型（提案法 F）の 6 倍近かった（FER2%の場合）。その一方、提案法 F は、エラーフリー、誤り条件、ともに提案法 B を上回る性能となった。この理由は主として以下の 2 点にあると考えられる [7]：1) モードの選択を閉ループで行なうことにより、開ループでモード判定を行なう場合に比べて 1 ビット以上の LSP コードベクトル符号帳のビット削減効果が得られる、2) 消失フレーム以外ではモード誤りがなく、量子化結果の誤りに対しても (1) 式を用いるモードが選択された時点で誤りの伝播がリセットされる。また、提案法 B がモード誤りのない状態（拡張レイヤのみのフレーム消失）でも提案法 F の性能を下回る結果になった理由としては、エラーフリーでの性能に差がある（平均 SD で 0.04dB）ことと、mode0 が選択されるフレーム数が提案法 F の方が多かった（実際に 10%程度多いことを確認した）ことが挙げられる。

6. まとめ

過去の量子化広帯域 LSP と量子化狭帯域 LSP の比を用いて現在のフレームの狭帯域–広帯域予測を行なう帯域スケーラブル LSP 量子化方法を提案し、その有効性を客観評価により確認した。フレームあたり狭帯域コア 25bit、拡張レイヤ 21bit で構成した場合、提案法による SD 性能改善効果はエラーフリーで 0.23dB、FER2% 想定で 0.16dB であった。主観的品質の確認は今後の課題である。

表 4: フレーム消失条件における SD 性能比較

フレーム消失条件（FER1%）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.73	75.2%	24.4%	0.5%
提案法 B*	1.62	79.2%	20.0%	0.9%
提案法 F*	1.53	84.1%	15.6%	0.3%
フレーム消失条件（FER2%）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.75	74.6%	24.6%	0.8%
提案法 B*	1.81	73.6%	23.1%	3.3%
提案法 F*	1.59	81.1%	18.2%	0.7%
フレーム消失条件（拡張レイヤのみ FER1%）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.73	75.2%	24.3%	0.5%
提案法 B*	1.58	80.9%	18.7%	0.5%
提案法 F*	1.53	84.2%	15.4%	0.5%
フレーム消失条件（拡張レイヤのみ FER2%）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.75	74.7%	24.5%	0.8%
提案法 B*	1.66	76.9%	22.2%	0.9%
提案法 F*	1.59	81.3%	17.9%	0.8%
フレーム消失条件（拡張レイヤのみ FER4%）				
方式	SD[dB]	～2dB	2～4dB	4dB～
従来法	1.78	73.5%	25.3%	1.2%
提案法 B*	1.82	68.6%	29.6%	1.9%
提案法 F*	1.69	76.3%	22.4%	1.4%

* B=バックワード型、F=フォワード型

謝辞：狭帯域 LSP 量子化器の開発にご協力頂いた松下電器産業（株）AV コア技術開発センターの森井利幸主席技師に感謝いたします。

参考文献

[1] R. D. D. Iacovo and D. Sereno: “Embedded CELP coding for variable bit-rate between 6.4 and 9.6 kbit/s”, Proc. IEEE ICASSP-91, pp. 681–684 (1991).

[2] T. Nomura, M. Iwadare, M. Serizawa and K. Ozawa: “A bitrate and bandwidth scalable CELP coder”, Proc. IEEE ICASSP-98, pp. 341–344 (1998).

[3] 片岡, 林：“G.729 を構成要素として用いるスケーラブル広帯域符号化”, 信学論 D-II, **J86-D-II**, 3, pp. 379–387 (2003).

[4] K. Koishida, V. Cuperman and A. Gersho: “A 16-kbit/s bandwidth scalable audio coder based on the g.729 standard”, Proc. IEEE ICASSP-2000, pp. 1149–1152 (2000).

[5] 日和崎, 森, 大室, 池戸, 徳元, 片岡：“高品質なユビキタス通信を実現するスケーラブル音声符号化技術”, NTT 技術ジャーナル, **16**, 1, pp. 10–13 (2004).

[6] K. Koishida, J. Lindén, V. Cuperman and A. Gersho: “Enhancing MPEG-4 CELP by jointly optimized inter/intra-frame LSP predictors”, Proc. IEEE Workshop on Speech Coding, pp. 90–92 (2000).

[7] T. Eriksson, J. Lindén and J. Skoglund: “Interframe LSF quantization for noisy channels”, IEEE Trans. on Speech and Audio Processing, **7**, 5, pp. 495–509 (1999).