G-017

# 音声通信用スケーラブルステレオ音声符号化方法の検討

## A Study of Scalable Stereo Speech Coding for Speech Communications

後藤 道代[1]　吉田 幸司[1]　Teo Chun Woei[2]　Neo Sua Hong[2]
Michiyo Goto　Koji Yoshida　Teo Chun Woei　Neo Sua Hong

1 松下電器産業株式会社　次世代モバイル開発センター
Next-Generation Mobile Communications Development Center, Matsushita Electric Industrial Co., Ltd.
2 AV Systems Team, Panasonic Singapore Laboratories Pte Ltd

## 1．はじめに

筆者らは，低遅延，低ビットレートかつ高音質を目標とした音声通信用のステレオ音声符号化方法について検討を行っている．前々回の発表では，CELP(Code Excited Linear Prediction coding, コード励振線形予測符号化)の音源パラメータを左右チャネルで共有し，声道パラメータに関してはチャネル毎に有する符号化モデルとそのシミュレーション結果について報告した[1].

また，前回の発表では，L チャネル信号に対する R チャネル信号の信号チャネル間情報（遅延時間差および振幅比）を抽出し，より符号化精度を高める手法について，その内容と検討結果を報告[2]すると共に，スケーラブルステレオ音声符号化のチャネル間予測に関する予備検討を報告した[3].

今回，モノラル信号と L チャネル信号から信号チャネル間情報を抽出し，その情報を用いてモノラル信号の駆動音源信号から L チャネル信号の駆動音源信号を予測することにより，より符号化精度を高め，同時にモノラル信号－ステレオ信号間のスケーラビリティも実現する符号化モデルを検討したので報告する.

## 2．アルゴリズム

### 2-1. アルゴリズム概要

今回のスケーラブル構成のステレオ音声符号化のモデルを，前回と同様，CELP 方式ベースの 3GPP 規格の音声符号化方式である AMR-WB (Adaptive Multi-Rate Wideband) の 23.85 kbit/s モード [4]をベースとして検討した．ステレオ拡張レイヤにおける片側チャネル（L チャネル）信号の符号化において，モノラル信号と L チャネル信号から抽出した信号チャネル間情報を用いてモノラル信号の駆動音源信号から L チャネル信号の駆動音源信号を予測する

と共に，付加的に CELP の音源探索により音源の符号化を行うモデルを想定した．図 1 に，今回検討した符号化のブロック図を示す．本モデルは，L チャネル信号および R チャネル信号を入力信号とし，モノラルコアレイヤとしてモノラル信号の符号化，ステレオ拡張レイヤとして L チャネル信号の符号化を行うように構成される.

### 2-2. モノラル信号の符号化

入力された左右チャネル信号をダウンミクスすることにより，モノラル信号を生成し，AMR-WB の 23.85 kbit/s モードを用いて符号化する．出力される主なパラメータは，CELP の駆動音源信号を生成するための ACB(Adaptive Codebook)インデックス，FCB(Fixed Codebook)インデックス，ACB ゲインおよび FCB ゲインから成る音源パラメータならびに LPC パラメータ等である．符号化により得られるモノラル信号の駆動音源信号は，L チャネル信号の符号化で用いる.

### 2-3. L チャネル信号の符号化

L チャネル信号の符号化は，まず，モノラル信号と L チャネル信号とから得られたチャネル間の空間情報を用いて，モノラル信号の駆動音源信号から L チャネル信号の駆動音源信号を予測する．次に，上述の予測駆動音源信号に対して，L チャネル信号の LPC 合成信号と入力 L チャネル信号との聴覚重み付き誤差信号を最小化させるように，付加的な駆動音源符号化を行うような構成とする.

チャネル間空間情報を構成する遅延時間差および振幅比の算出方法について以下に記述する.

(1) 遅延時間差

モノラル信号と L チャネル信号の遅延時間差を次式の値を最大にする $m=m_{max}$ で定義する.

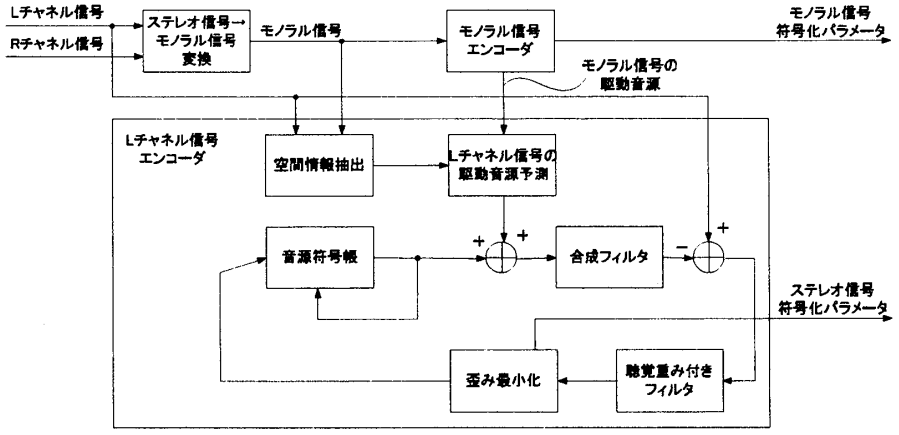


図 1　ステレオ音声符号化のブロック図

$$\phi(m)=\sum_{n=0}^{FL-1} x_{mono}(n)\bullet x_{L-ch}(n-m)$$

ここで，$x_{mono}(n)$ および $x_{L-ch}(n)$ は，モノラル信号およびＬチャネル信号，$FL$ はフレーム長である．今回，$FL$ = 320 (標本化周波数: 16 kHz)である．

(2) 振幅比

モノラル信号とＬチャネル信号の振幅比 $C$ を次式で定義する．

$$C=\sqrt{\frac{\sum_{n=0}^{FL-1} x_{L-ch}(n-m_{max})^2}{\sum_{n=0}^{FL-1} x_{mono}(n)^2}}$$

上記で得られた各々の値に対してフレーム間に渡る平滑化を行った後に量子化を行い，モノラルの駆動音源信号からのＬチャネル駆動音源信号の予測に用いる．今回，遅延時間差および振幅比の量子化ビット数は各々6bit/フレームとした．

(3) 駆動音源符号化

Ｌチャネル信号用に付加する駆動音源信号は，ACB インデックス，FCB インデックス，ACB ゲインおよび FCB ゲインから成る音源パラメータから構成される．今回，音源パラメータは AMR-WB の 23.85kbit/s モードのものを用いた．モノラル信号からのチャネル間予測駆動音源信号を含めて合成されたＬチャネル信号と入力Ｌチャネル信号との聴覚重み付き誤差信号が最小となる音源パラメータを決定する．

### 2-4. Ｒチャネル信号の符号化

Ｒチャネル信号単独の符号化は行わずに，復号信号を次式によって求める．

$$\hat{x}_{R-ch}(n)=2\bullet\hat{x}_{mono}(n)-\hat{x}_{L-ch}(n)$$

$$n=0,\ldots,FL-1$$

ここで，$\hat{x}_{mono}(n)$，$\hat{x}_{L-ch}(n)$ および $\hat{x}_{R-ch}(n)$ は，モノラル信号，Ｌチャネル信号およびＲチャネル信号の復号信号，$FL$ はフレーム長である．

## 3．評価

今回の符号化モデルの評価を空間情報の抽出結果および合成音声の品質を用いて行った．

### 3-1. 空間情報

空間情報の抽出結果を評価するための音源は，３人の話者が３方向（正面，左および右）から発声した音声を収音して編集した．

(1) 遅延時間差

遅延時間差の抽出結果の一例を図２に示す．評価音源のＲチャネル信号は，Ｌチャネル信号を時間方向に 16 サンプル（1ms）シフトさせて生成した．横軸は時間（フレーム），縦軸は遅延サンプル数を示す．

(2) 振幅比

振幅比の抽出結果の一例を図３に示す．評価音源のＲチャネル信号は，Ｌチャネル信号の振幅を３話者の各音声区間に対して各々１，0.5，２倍して生成した．横軸は時間（フレーム），縦軸は振幅比を示す．

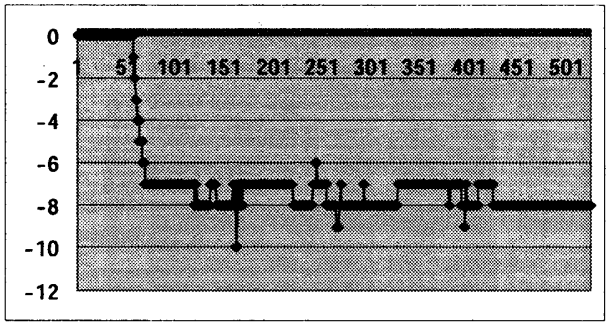


図２　Ｌチャネル信号波形に対する遅延時間差

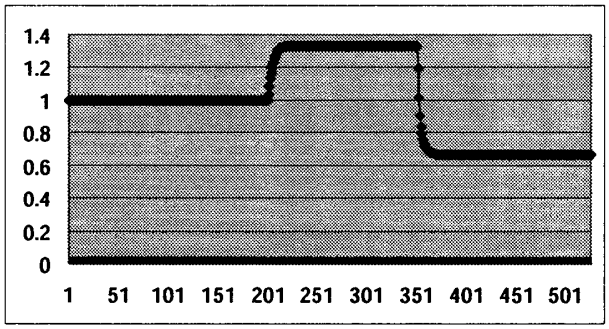


図３　Ｌチャネル信号波形に対する振幅比

### 3-2. 合成音声

シミュレーションを行って音質の確認を行った．比較対象は，(a)ステレオ原音声，(b)チャネル毎に AMR-WB 符号化したステレオ音声，とした．評価音源は，３人の話者が３方向（正面，左および右）から 発声した音声を収音して使用した．

筆者らが評価を行った結果，今回の符号化モデルによる合成音は，了解度およびステレオ感はステレオ原音声および，チャネル毎に AMR-WB 符号化したステレオ音声に比べて遜色なかった．ただし，一部音質劣化する箇所があり，その要因としては，空間情報抽出およびそのチャネル間音源予測への適用に関して改良の余地があることなどが考えられる．

## 4．まとめ

モノラル信号とＬチャネル信号から抽出したチャネル間情報を基に予測した駆動音源信号を用いてＬチャネル信号の音源符号化を行う構成のスケーラブルステレオ音声符号化方法を検討した．空間情報抽出およびそのチャネル間音源予測への適用に関して改良の余地があることや，より高能率（低ビットレート）の音源パラメータで符号化した場合の本手法の有効性の検証等が今後の課題である．

## [参考文献]

[1]後藤他，2004 信学会基礎・境界　ソサイエティ大会 A-6-6

[2]後藤他，2005 信学会総合大会　D-14-2

[3]吉田他，2005 信学会総合大会　D-14-1

[4]3GPP TS 26.190: “AMR Wideband speech codec; Transcoding functions” (2001-12)