

モバイル立体音響通信に向けた信号処理技術 —音響コーデックにおけるローカル立体音響レンダリング方法—

菊入 圭 仲 信彦 大矢 智之

株式会社 NTT ドコモ マルチメディア研究所
〒239-8536 神奈川県横須賀市光の丘 3-5 NTT DoCoMo R&D センタ
E-mail: {kikuri, naka, ohya}@mml.yrp.nttdocomo.co.jp

あらまし これからのモバイル通信は、通信速度の飛躍的な向上と、それによる従来にないリッチなコンテンツを利用できることが期待されている。本稿では、次世代の通信形態のひとつとして、モバイル立体音響通信について述べる。まずモバイル立体音響通信の主要なアプリケーションを紹介し、そのシステム構成を述べる。これらシステム構成の中で、最もモバイル通信に適していると思われるサーバ・レンダリング型システムの課題は、伝送レートが増大、再生される立体音場のユーザの動きへの追従が挙げられる。これらの課題を解決するため、パラメトリックステレオ符号化を用いたサーバ・レンダリング型システムを考え、クライアント側のデコーダにおけるローカル立体音響レンダリング方法を提案する。計算機シミュレーションにより、本提案方法で修正されたバイノーラル信号は、線形補間した HRTF を畳み込んで生成したバイノーラル信号とほぼ同程度の音像定位性能であることを確認した。

キーワード モバイル立体音響通信、ローカル立体音響レンダリング、パラメトリックステレオ符号化、ステレオパラメータ

Signal Processing Technologies for Mobile Spatial Audio Communications —Local Spatial Audio Rendering Method for Audio Codecs—

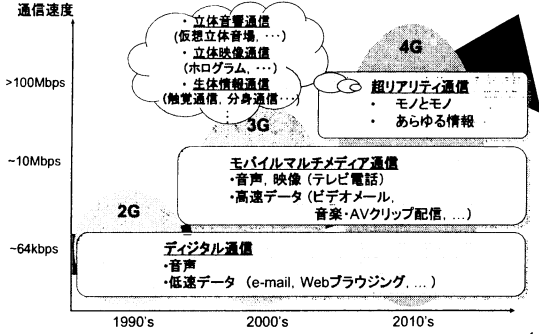
Kei KIKUIRI Nobuhiko NAKA and Tomoyuki OHYA

Multimedia Labs., NTT DoCoMo, Inc.
NTT DoCoMo R&D Center, 3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8526 Japan
E-mail: {kikuri, naka, ohya}@mml.yrp.nttdocomo.co.jp

Abstract This paper describes applications and audio signal processing technologies for mobile spatial audio communications, which will be one of the next generation mobile communications. The authors propose a local spatial audio rendering method for parametric stereo decoders. The proposed method enables server-side-rendering systems to follow user's head moving faster. From computer simulation results, a binaural signal modified by the proposed method localizes a sound image comparable to a binaural signal convolved with an interpolated HRTF.

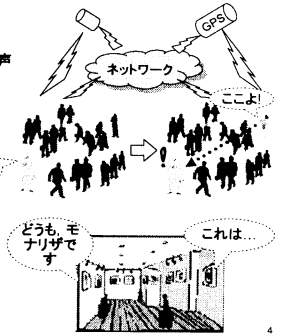
Keyword Mobile Spatial Audio Communications, Local Spatial Audio Rendering, Parametric Stereo Coding, Stereo Parameter

モバイル通信の発展

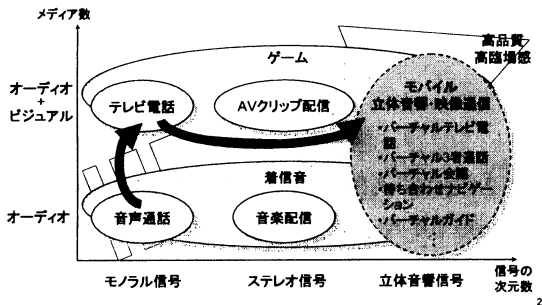


モバイル立体音響通信のアプリケーション例

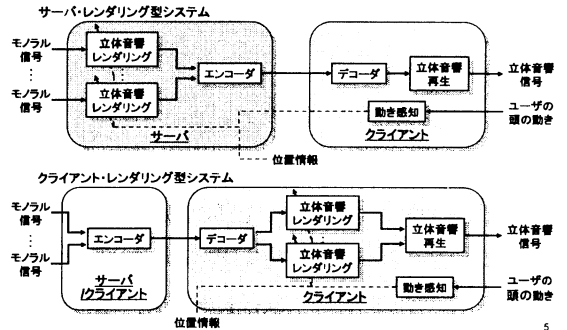
- ・ 待ち合わせナビゲーション
 - 待ち合わせ相手のいる方向から音声がかんこえてくる
 - 待ち合わせ相手の姿が見えなくてもどこにいるかわかる
- ・ バーチャルガイド
 - 博物館, 美術館でのガイド



携帯電話におけるAVアプリケーション

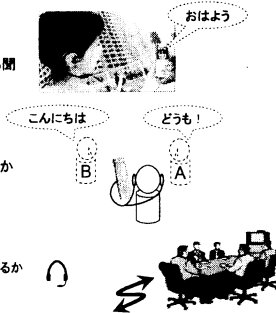


モバイル立体音響通信システム

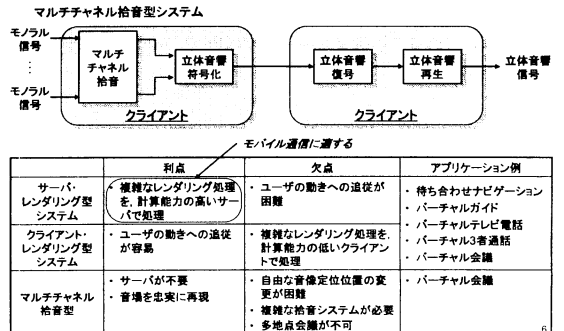


モバイル立体音響通信のアプリケーション例

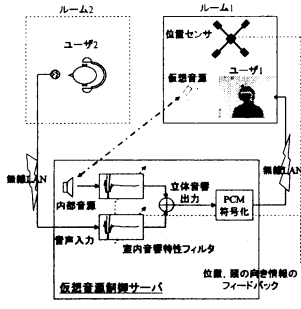
- ・ バーチャルテレビ電話
 - 通話相手の音声がかんこえてくる
 - 自然で聞き取りやすい
- ・ バーチャル3者通話
 - それぞれの話者の声異なる方向から聞こえてくる
 - 話者を認識しやすい
- ・ バーチャル会議
 - 遠隔地からの参加者も同じ会議室にいるかのように参加者の声がかんこえてくる
 - モバイルユーザーにも臨場感



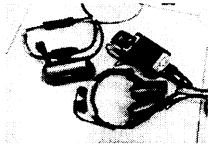
モバイル立体音響通信システム



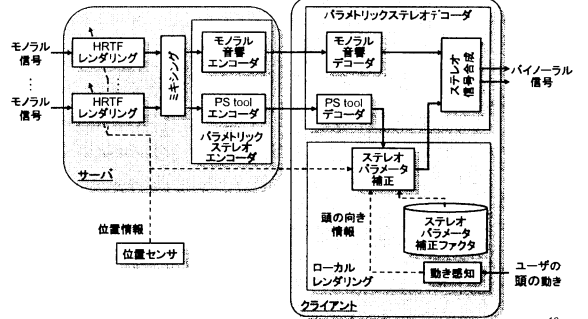
サーバ・レンダリング型システムのプロトタイプ



- ・ 仮想立体音響空間の再現
 - 受聴者の動きを、実時間処理によりフィードバック
 - デモルーム1内に仮想音源が定位
- ・ ユーザ間の位置関係を仮想的に再現することによる立体音響通信
 - ルーム2のユーザ2からの音声、実際の位置関係通りに、壁の向こうへ定位



サーバ・レンダリング型システムの提案構成



サーバ・レンダリング型システムの課題

① 立体音響信号の伝送による伝送レートの増大

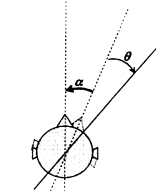
立体音響信号は、高サンプリングレートの2チャンネル以上の信号

↓
低ビットレートの符号化が必要

パラメトリックステレオ符号化

- ・ ステレオ信号を、モノラル信号とステレオパラメータに分けて低ビットレートで符号化
- ・ 復号されたモノラル信号とステレオパラメータより、ステレオ信号を合成
- ・ ステレオパラメータ ⇔ 音像定位の手がかりと対応
 - IID (Inter-channel Intensity Difference): チャンネル間の強度差
 - IPD (Inter-channel Phase Difference): チャンネル間の位相差
 - OPD (Overall Phase Difference): モノラル信号との位相差
 - ICC (Inter-Channel Coherence): チャンネル間の相関度
- ・ 2003年にMPEGにてパラメトリックステレオツールとして標準化 [4]
- ・ 同様の手法でマルチチャンネル信号の符号化も可能 [2][3][5]

クライアント側処理の提案方法: ステレオパラメータの補正によるバイノーラル信号の修正



頭の向きが α 変化する
ことにより、到来角度が θ
から $\theta + \alpha$ に変化する

ステレオパラメータの補正

$$\hat{x}(\theta + \alpha) = x(\theta) + d_x(\theta, \theta + \alpha).$$

$\hat{x}(\theta + \alpha)$: 補正後のステレオパラメータ
 $x(\theta)$: 補正前のステレオパラメータ
 $d_x(\theta, \theta + \alpha)$: パラメータ補正ファクタ
 $x = \{iid, ipd, opd, icc\}$

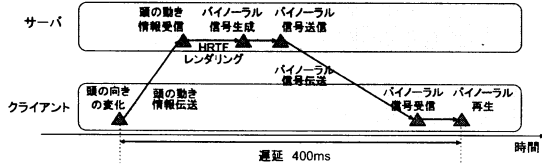
パラメータ補正ファクタの事前空算

$$d_x(\theta, \theta + \alpha) = -d_x(\theta + \alpha, \theta) = \frac{1}{N} \sum_{n=0}^{N-1} x(\theta + \alpha, n) - x(\theta, n).$$

n : フレームインデックス, N : 学習フレーム数

サーバ・レンダリング型システムの課題

② ユーザの動きに対する音像定位位置の追従

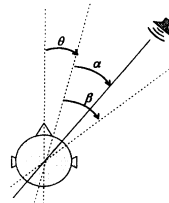


サーバが位置情報を受信してから、ユーザ側で再生されるまでに、ユーザの頭の向きが変化するため、仮想音源からユーザに対する信号到来角度が変化する

↓
頭の向きの変化に追従したバイノーラル信号の修正が必要

↓
クライアント側でのローカル立体音響レンダリング処理

ベンチマーク: HRTFの線形補間によるバイノーラル信号の修正



HRTFの線形補間

$$\hat{h}(\theta + \alpha) = h(\theta) + \frac{\alpha}{\beta} \delta_h(\theta, \theta + \beta).$$

$$\delta_h(\theta, \theta + \beta) = h(\theta + \beta) - h(\theta).$$

$\hat{h}(\theta + \alpha)$: 補間により求められた到来角 $\theta + \alpha$ の HRTF
 $h(\theta)$: 到来角 θ の HRTF
 $h(\theta + \beta)$: 到来角 $\theta + \beta$ の HRTF

計算機シミュレーション条件

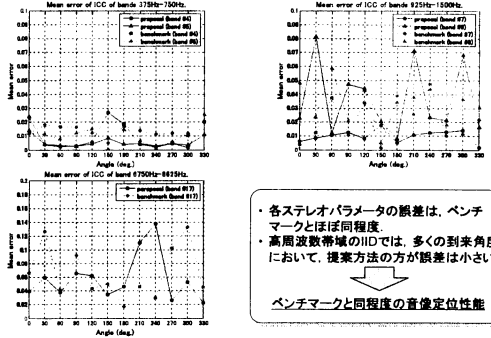
- パラメトリックステレオコーデック
 - モノラル音響符号化: MPEG-4 HE-AAC [6], PS符号化: MPEG-4 PS tool [4]
 - サンプリングレート: 48 kHz
 - ビットレート: 24 kbps (ステレオパラメータは非量子化)
- HRTF
 - CIPIC HRTF database [7]
- 提案方法
 - HRTF水平面到来角度 θ : 0, 30, 60, ..., 330 deg. (12方向)
 - 補正ファクタ
 - 修正角度 α : +30 deg.
 - 学習サンプル: 白色雑音
- ベンチマーク
 - HRTF水平面到来角度 θ : 15, 45, 75, ..., 345 deg. (12方向)
 - 修正角度 α : +15 deg.
- 客観評価方法
 - 到来角度 $\theta + \alpha$ deg. のHRTFを畳み込んだバイノーラル信号のステレオパラメータと、提案方法およびベンチマークにより、到来角度を $\theta + \alpha$ deg. に修正されたバイノーラル信号のステレオパラメータの平均誤差値を比較
 - 評価サンプル: ポップミュージック

平均誤差値の定義

- IID: $\epsilon_{iid} = \frac{1}{N} \sum_{n=1}^{N-1} |i(d(n)) - iid(n)|$
- IPD: $\epsilon_{ipd} = \frac{1}{N} \sum_{n=1}^{N-1} |\text{ipd}(n) - \text{ipd}(n) \bmod 2\pi|$
- OPD: $\epsilon_{opd} = \frac{1}{N} \sum_{n=1}^{N-1} |\text{opd}(n) - \text{opd}(n) \bmod 2\pi|$
- ICC: $\epsilon_{icc} = \frac{1}{N} \sum_{n=1}^{N-1} |icc(n) - icc(n)|$

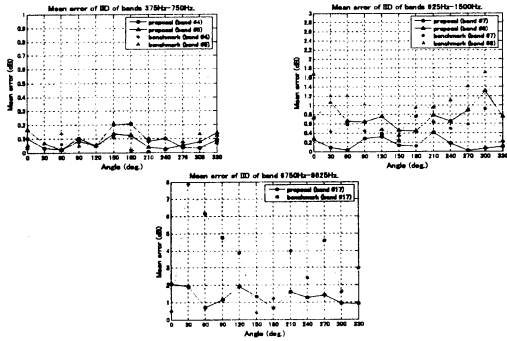
$i(n)$: フレーム n におけるHRTFを畳み込んだバイノーラル信号のステレオパラメータ
 $\text{ipd}(n)$: フレーム n における提案方法、ベンチマークにより修正されたバイノーラル信号のステレオパラメータ
 N : フレーム数

計算機シミュレーション結果: ICC



各ステレオパラメータの誤差は、ベンチマークとほぼ同程度
 高周波数帯域のIIDでは、多くの到来角度において、提案方法の方が誤差は小さい
 ↓
 ベンチマークと同程度の音像定位性能

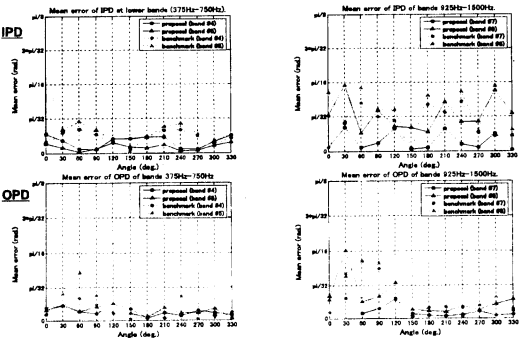
計算機シミュレーション結果: IID



まとめ

- サーバルエンダリング型システムの課題(立体音響信号の伝送による伝送レートの増大、ユーザの動きに対する音像定位位置の追従)を解決する、パラメトリックステレオ符号化を用いた、ステレオパラメータの補正によるバイノーラル信号の修正方法の提案
 - 計算機シミュレーションにより、HRTFを畳み込んで作成したバイノーラル信号のステレオパラメータ (IID, IPD, OPD, ICC) との平均誤差値を評価。
 - HRTFの線形補間によるバイノーラル信号修正と比較し、各ステレオパラメータにおいて、ほぼ同程度の誤差であることを確認
 - 音像定位性能は、HRTFの線形補間により修正されたバイノーラル信号とほぼ同程度。
- 今後の課題
 - 主観評価試験による音像定位性能の評価

計算機シミュレーション結果: IPD, OPD



参考文献

- Y. Yasuda, T. Ohya, D. McGrath, and P. Flanagan, "3-D Audio Communications for Future Mobile Networks," AES 23rd Int. Conf., Copenhagen, May 2003
- C. Faller and F. Baumgarte, "Binaural Cue Coding - Part I: Psychoacoustic Fundamentals and Design Principles," IEEE Trans. on Speech and Audio Proc., pp. 509-519, Vol. 11, No. 6, Nov., 2003.
- C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and Applications," IEEE Trans. on Speech and Audio Proc., pp. 520-531, Vol. 11, No. 6, Nov., 2003.
- ISO/IEC, "Coding of audio-visual objects - Part 3: Audio, AMENDMENT 2: Parametric Audio Extension," ISO/IEC Int. Std. 14496-3:2001/Amd 2:2004, 2004.
- J. Herre, et al., "The Reference Model Architecture for MPEG Spatial Audio Coding," AES 118th Convention, Barcelona, May 2005.
- ISO/IEC, "Coding of audio-visual objects - Part 3: Audio, AMENDMENT 1: Bandwidth Extension," ISO/IEC Int. Std. 14496-3:2001/Amd 1:2003, 2003.
- V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," Proc. IEEE WASPAA 2001, New York, 2001.
- B. C. J. ムーア, 聴覚心理学概論, 大串徹吾(監訳), 誠信書房, 東京, 1994.