

博士論文

隠れマルコフモデルを利用した  
音声コーデックの性能向上に関する研究

米崎 正

2000年3月24日

奈良先端科学技術大学院大学  
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
博士(工学)授与の要件として提出した博士論文である。

米崎 正

審査委員： 鹿野 清宏 教授  
山本 平一 教授  
関 浩之 教授  
中村 哲 助教授

# 隠れマルコフモデルを利用した 音声コーデックの性能向上に関する研究\*

米崎 正

## 内容梗概

本論文では音声コーデックの性能向上を図るために隠れマルコフモデル (HMM) を適用する手法を提案し、その性能を評価する。

音声コーデックに対しては高音声品質を実現する低ビットレート化、及び、通信路誤りに対する耐性の向上が望まれている。これらは情報源符号化法、及び、通信路符号化法として独立して検討されてきた。つまり、情報源符号化は情報源を、通信路符号化は通信路をモデル化することで個別に性能向上を図り、両者を組み合わせることで高性能の音声コーデックを実現する。一方、近年、情報源モデルを利用した通信路誤り耐性の向上も検討されており有効性が確認されている。このことから情報源である音声信号のモデル化、及び、その利用法を検討することは音声コーデックの性能向上を図る上で重要な課題であるといえる。隠れマルコフモデル (HMM) は、音声認識の分野で広く用いられており、音声信号の優れた情報源モデルである。本論文では、この HMM を音声コーデックに適用することで音声コーデックの性能向上を図る。

HMM を利用した音声コーデックはこれまでも提案されてきたものの、商用として耐え得る音声品質を実現することはできなかった。そこで、高音声品質を実現する既存の音声コーデックに対して HMM を適用することで、その性能を向上させることを検討する。そのためには、HMM の利用方法、及び、音声コーデックで分析されるパラメータの効率良いモデル化に関する検討が必要である。本論文は前者を目的とし、音声コーデックに対する HMM の適用方法の提案と性能評価を行なう。性能評価は音声のスペクトル包絡パラメータをベクトル量子化する系を作成し、これに提案法を適用することで行なった。

本論文における提案法の概要は次の通りである。HMM は情報源モデルとして、あるパラメータ列が出力される確率を推定する。この出力確率は情報源の状態遷移に対応して逐

---

\*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DT9861027, 2000年3月24日.

次変化していく．そこで，まず，出力確率に対応して符号化系を適応的に構築する手法を提案する．これは，出力される確率の高いものに短い符号長を割り当てることで平均的な符号長を短くし，低ビットレート化を図るものである．しかしながら，本手法は‘適応的に可変長符号系を構築する’点で通信路での誤りの影響を受けやすい手法であるといえる．この点に関して，低ビットレートを維持しつつ誤り耐性を向上させる手法についても検討する．次に，出力確率を符号列の確からしさと考えることで通信路誤りを訂正する手法を検討する．これは，受信した符号列の制約の下，送信された符号を最尤推定するものである．以上の手法について計算機シミュレーションによって，その有効性を確認する．

## キーワード

音声符号化, 隠れマルコフモデル, 情報圧縮, 誤り訂正

# A study on adaptation of hidden Markov models to improve the performance of a speech coding system\*

Tadashi Yonezaki

## Abstract

This thesis focuses on adaptations of hidden Markov models (HMMs) for improving the performance of speech coding systems.

Low bit rate and channel robustness are required for speech coding system. These have been investigated individually as a source coding system and a channel coding system. They have been improved by modeling the encoded object. Joint source and channel coding, which augments channel robustness based on a source model, was proposed recently. Therefore, modeling of speech signal and utilization of it are important issue for an efficient speech coding system. HMMs have been widely used in application of speech recognition systems as an efficient source model of speech signal. Adaptations of HMMs to a speech coder could realize an efficient coding system.

Several speech coding algorithms, which utilize HMMs, have been proposed. However, none of them can acquire toll quality decoded speech signal. HMMs could be adapted to not a speech-coding algorithm itself but a conventional speech coding system that achieve toll quality for improving the performance. Two steps of inquisitions are required for this purpose; an adaptation algorithm of HMMs and an efficient modeling of code sequences which are derived from the coding system. This thesis focuses on the former subject, and the adaptation algorithms of HMMs are proposed and assessed. The assessments are carried on adapting the proposed algorithms to a simple coding system that quantizes spectral envelope parameter of speech signal.

The proposed algorithms are briefly described following. HMMs estimate the output probability of a symbol sequence as a source model. The probability is dynamically changed

---

\*Doctor's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT9861027, March 24, 2000.

corresponding to the status of the model. First, an adaptive coding system based on the output probability is proposed in this thesis. The shorter lengths of codes are assigned to the symbols that are likely to occur so that the average code length would be shorter. This approach realizes a low bit rate speech coder but is easy to suffer from channel errors because of its ‘adaptively constructed variable length coding system’ property. The approach that augments channel robustness, while it remains low bit rate, is also investigated. Second, error correction approach based on the output probability of a symbol sequence is presented. This approach estimates the transferred symbol in a sense of maximum likelihood approach under constraint of the received symbol sequence. The efficiencies of these approaches are evaluated by computer simulation.

**Keywords:**

speech coding, HMM, data compression, error correction

# 謝辞

本論文は、奈良先端科学技術大学院大学教授 鹿野 清宏 博士の御指導のもとに、筆者が行なった研究成果をまとめたものです。本研究は鹿野教授に頂いた御厚情なくしてはありえず、ここに深甚なる謝意を表します。同大学助教授 中村 哲 博士におかれましては、その博識は申し上げるまでもなく研究者としての姿勢に常日頃より感銘を受けること多く、更に、賜りました御助言、御鞭撻に筆者はここに至るまでの道を見い出して参りました。ここに深謝致します。

奈良先端科学技術大学院大学教授 山本 平一 博士、並びに、同大学教授 関 浩之 博士には、本論文をまとめるにあたり御教示、御助言を賜りました。ここに謝意を表します。

京都大学助教授 伊勢 史郎 博士には筆者のデジタル信号処理に関する知識に新たな視点を与えて頂き、また、奈良先端科学技術大学院大学助手 陸 金林 博士には本研究に関し貴重な御意見を頂きました。深く感謝申し上げます。山田 武志 博士(現、筑波大学助手)、並びに、滝口 哲也 博士(現、日本IBM 株式会社勤務)を初め奈良先端科学技術大学院大学情報科学研究科音情報処理学講座鹿野研究室の皆様には、本研究に対する御意見を頂いたのみならず、筆者の研究生活を様々な側面から支えて頂きました。厚く御礼申し上げます。

奈良先端科学技術大学院大学において修学する機会を与えて下さいました松下通信工業株式会社技術本部テレコム研究所所長 本間 光一 博士には深く感謝申し上げます。同研究所移動通信開発グループ主担当 滝 秀士 氏、並びに、移動通信方式開発グループマルチメディア方式チームリーダー 八木 敏男 氏には筆者の修学に纏わる多難を排し、研究生生活の道筋を整えて頂きました。両氏の御厚情には感謝の念の表しようもございません。また、プロジェクトリーダー 吉田 幸司 氏をはじめ、同チームの皆様の日頃より賜わる御協力の数々に筆者は励まされ本論文をまとめることができました。感謝の念を表します。今日に至るまで多大なる御支援を賜りました同研究所の皆様には感謝申し上げます。

筆者を温かく見守り、いかなる時も労を惜しまず協力し、励まして下さった両親、姉には大変感謝しております。最後になりましたが、本研究を妻の協力なくして続けることはできず、本論文を妻の協力なくしてまとめることはできませんでした。筆者は妻と共に本

論文を著したものの念を拭い去ることができない旨をここに記します。



# 目次

謝辞	v
第1章 序論	3
第2章 音声コーデックにおける情報源モデルの活用	9
2.1. 序言	9
2.2. 物理モデルと統計モデル	9
2.3. 量子化に対する統計モデルの適用	12
2.3.1 マルコフモデルに基づく量子化	13
2.3.2 隠れマルコフモデルに基づく量子化	15
2.4. 符号化に対する統計モデルの適用	17
2.4.1 エントロピー符号化	17
2.5. 復号化に対する統計モデルの適用	19
2.5.1 マルコフモデルを利用した復号化	19
2.5.2 隠れマルコフモデルを利用した復号化	19
2.6. 音声コーデックにおける隠れマルコフモデルの位置付け	21
2.7. 結言	23
第3章 隠れマルコフモデル	25
3.1. 序言	25
3.2. エルゴード的隠れマルコフモデル	25
3.3. フォワード-バックワードアルゴリズム	27
3.4. 結言	32
第4章 隠れマルコフモデルを用いた情報量削減	33
4.1. 序言	33
4.2. エントロピー符号化との組合せによる情報量削減	33

4.2.1	状態分布の更新	34
4.2.2	符号空間の確率分布の推定	34
4.2.3	エントロピー符号化/復号化	35
4.3.	性能評価	36
4.3.1	シミュレーション条件	36
4.3.2	HMM 状態数と圧縮効率の関係	37
4.3.3	伝送レートと歪みの関係	38
4.3.4	誤り耐性	40
4.4.	結言	40
<b>第 5 章</b>	<b>情報量削減における誤り耐性の改善</b>	<b>43</b>
5.1.	序言	43
5.2.	分布の拘束による誤り耐性の改善	43
5.2.1	状態分布の収束性	43
5.2.2	状態分布の制限による誤り耐性の向上	44
5.2.3	符号語空間上の確率分布の制限による誤り耐性の向上	44
5.3.	性能評価	48
5.3.1	状態分布量子化法	48
5.3.2	符号系量子化法	49
5.3.3	状態分布量子化法と符号系量子化法の比較	50
5.4.	結言	51
<b>第 6 章</b>	<b>隠れマルコフモデルを用いた誤り訂正</b>	<b>53</b>
6.1.	序言	53
6.2.	最尤推定との組合せによる誤り訂正	53
6.2.1	符号列モデルに基づく確率	54
6.2.2	通信路モデルに基づく確率	55
6.2.3	最尤推定を利用した確率の最大化	55
6.3.	性能評価	56
6.3.1	シミュレーション条件	56
6.3.2	符号列/通信路モデルに対する重み付け	57
6.3.3	HMM 状態数と誤り訂正能力の関係	58
6.3.4	誤り訂正能力の評価	59

6.3.5	誤り検出との組合せ . . . . .	60
6.3.6	処理量 . . . . .	62
6.4.	結言 . . . . .	62
<b>第7章</b>	<b>結論</b>	<b>65</b>
	参考文献	67



# 目 次

1.1	音声コーデックの伝送レートと音声品質の関係 . . . . .	4
1.2	通信路誤り率と音声品質の関係 . . . . .	5
2.1	音声コーデック処理の流れ . . . . .	9
2.2	音声生成モデル . . . . .	10
2.3	符号励起線形予測 (CELP) ボコーダ . . . . .	11
2.4	有限状態ベクトル量子化ブロック図 . . . . .	13
2.5	Phonetic VOCODER ブロック図 . . . . .	16
2.6	ハフマン符号系の構築例 . . . . .	18
2.7	ターボ復号化ブロック図 . . . . .	20
2.8	符号化率 1/2 の畳み込み符号化器の例 . . . . .	20
2.9	図 2.8 の符号化器に対応するトレリスダイアグラムの例 . . . . .	20
2.10	Gracia-Frias らが用いた情報源モデル . . . . .	21
2.11	情報源を考慮した図 2.8 の符号化器に対応するトレリスダイアグラムの例 . . . . .	21
3.1	5 状態 left-to-right 型 HMM の例 . . . . .	26
3.2	トレリスダイアグラムと出力パラメータ系列 . . . . .	26
3.3	複数のモデルによるトレリスダイアグラムと出力パラメータ系列 . . . . .	27
3.4	枝分かれを抑制したトレリスダイアグラムと出力パラメータ系列 . . . . .	27
3.5	8 状態エルゴード的 HMM . . . . .	28
3.6	エルゴード的 HMM のトレリスダイアグラム . . . . .	28
3.7	フォワード-バックワードパス確率分布の遷移例 : (a) 4 状態エルゴード的 HMM のトレリスダイアグラム , ⊙ 初期状態 , ● 最終状態 , (b) 出力符号列 , (c) 確率分布の遷移 . . . . .	31
4.1	HMM を利用した情報量削減アルゴリズムのブロック図 . . . . .	34
4.2	状態分布の遷移 . . . . .	35

4.3	符号空間の確率分布の遷移 . . . . .	36
4.4	8bit 符号帳に対して情報量削減手法を適用したときの HMM 状態数と平均符号長の関係 . . . . .	38
4.5	8 状態 HMM を用いて適応的にエントロピー符号系を構築したときの歪みと伝送レートの関係 . . . . .	39
4.6	8bit 符号帳に対して情報量削減手法を適用したときの HMM の状態数と誤り耐性の関係 . . . . .	40
5.1	1bit 誤り発生時の符号化-復号化に対する状態分布間距離の収束性 . . . . .	44
5.2	状態分布量子化法のブロック図 . . . . .	45
5.3	符号系符号帳の作成手順 . . . . .	46
5.4	符号系量子化法のブロック図 . . . . .	47
5.5	状態分布量子化法を用いたときの 1bit 誤りに対する状態分布間距離の収束性	48
5.6	状態分布量子化符号帳サイズと平均符号長に基づく情報量削減効果の関係 .	48
5.7	符号系量子化法を用いたときの 1bit 誤りに対するケプストラム距離の収束性	49
5.8	8 状態 HMM を用いた場合の 1 ビット誤りに対する状態分布の収束性の比較	50
5.9	8 状態 HMM を用いた場合の 1bit 誤りに対するケプストラム距離の収束性の比較 . . . . .	51
6.1	最尤推定に基づく誤り訂正法ブロック図 . . . . .	54
6.2	2 元対象通信路 . . . . .	55
6.3	8 状態 HMM を用いて誤り訂正したときの符号列モデルに対する重み付け ( $\gamma$ ) とケプストラム距離の関係 . . . . .	58
6.4	2 ~ 64 状態 HMM を用いて誤り訂正したときのケプストラム距離 (BER=3%)	58
6.5	2, 4, 8, 16 状態 HMM を用いて誤り訂正したときの通信路上のビット誤り率に対するケプストラム距離 . . . . .	59
6.6	2, 4, 8, 16 状態 HMM を用いて誤り訂正したときのアルゴリズム遅延とケプストラム距離の関係 (BER=3%) . . . . .	60
6.7	予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときの通信路上のビット誤り率に対するケプストラム距離 . . . . .	61
6.8	予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときのアルゴリズム遅延とケプストラム距離の関係 (BER=3%) . . . . .	62

# 表 目 次

2.1	音声コーデック系へのマルコフモデルの適用 . . . . .	12
2.2	音声コーデック系への隠れマルコフモデルの適用 . . . . .	12
4.1	情報量削減手法検討用パラメータ抽出諸元 . . . . .	37
6.1	誤り耐性向上における評価パラメータ抽出諸元 . . . . .	57
6.2	2, 4, 8, 16 状態 HMM を用いて誤り訂正した時のケプストラム距離の分布 (BER=3%) . . . . .	60
6.3	予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂 正したときのケプストラム距離の分布 (BER=3%) . . . . .	61





# 第1章 序論

90年代に入り情報通信はアナログ通信方式からデジタル通信方式へ大きく転換した。特にDSP(Digital Signal Processor)の発展によって、デジタル信号処理に基づく高度な信号処理の実現が容易になり、通信路の利用効率は大きく向上した。

一方、通信路利用効率の向上は1通信当たりの通信単価を低減させ、携帯電話に代表されるパーソナルモバイルコミュニケーションメディアが爆発的に普及するきっかけとなった。1999年現在、PHS(Personal Handy-phone System)と携帯電話を合わせたモバイルコミュニケーションメディアは日本国内で累計5000万台以上普及し、その普及率は40%を越えている。更に、モバイルコミュニケーションを取り巻く環境は大きく変化しており、従来の携帯電話のように音声情報のみを伝えるものではなく、動画情報の伝送や情報端末機としてデータ通信への要求も大きくなりつつある。

このように、デジタル化により通信路容量の拡大が図られているものの、通信する情報量は増大する一方であり、なお、ひとつの信号の伝送に要する情報量の削減が求められている。本論文で取り扱う音声信号に関しては、高音声品質を維持した低ビットレート化、通信路での誤りの影響を受けない高い誤り耐性を実現するコーデック方式の実現が望まれている。

これらの要望を満たすため様々な音声コーデック方式が標準化されてきた[1][2][3]。ここで、これまでに標準化されてきた音声コーデック方式についてまとめる。現在、モバイルコミュニケーションメディアが主に通信の対象としている音声信号は、有線アナログ電話のサービス開始時に定められた、通過周波数帯域が3.4kHz未満に制限され標本化周波数8kHzでサンプリングされた狭帯域音声信号である。ITU-T(International Telecommunication Union Telecommunication sector, 旧 CCITT)では有線アナログ電話と同等の音声品質を実現するコーデック方式として1972年に64kbpsの非線形PCM(Pulse Code Modulation)を標準化方式として定めた。以降、1984年にPHSで用いられている32kbps(ADPCM:Adaptive Differential PCM)音声コーデック方式、1992年に16kbps(LD-CELP:Low Delay Code Excited Linear Prediction)音声コーデック方式、1995年にインターネット電話でも用いられている8kbps(CS-ACELP:Conjugate Structure Algebraic CELP)音声コーデック方式とい

うように低ビットレート化が図られてきており、現在、4kbps 音声コーデック方式の標準化作業が進められている。これらは、元来、通信路での誤りが少ない有線系での使用を目的として標準化された音声コーデック方式であり、携帯電話等無線系での使用には通信路誤りへの対策が必要となる。

無線系での使用を目的とした音声コーデックは、日本では PDC(Personal Digital Cellular) 方式として 11.2kbps と 5.6kbps の音声コーデック方式が、欧州では GSM(Groupe Special Mobile) 規格として 22.8kbps と 11.4kbps の音声コーデック方式が、北アメリカでは 13kbps 音声コーデック方式が 1980 年代後半から相次いで標準化されている。これらの音声コーデックは通信路で生じる誤りを考慮して、ある程度の誤りが生じても復号音声の品質に影響を及ぼさない方式として標準化されている。

図 1.1 に、これまでに標準化された音声コーデックの伝送レートと音声品質の関係を通信路での誤りが少ないものとして示す。図は、左上に近いほど少ない情報量で高い復号音声

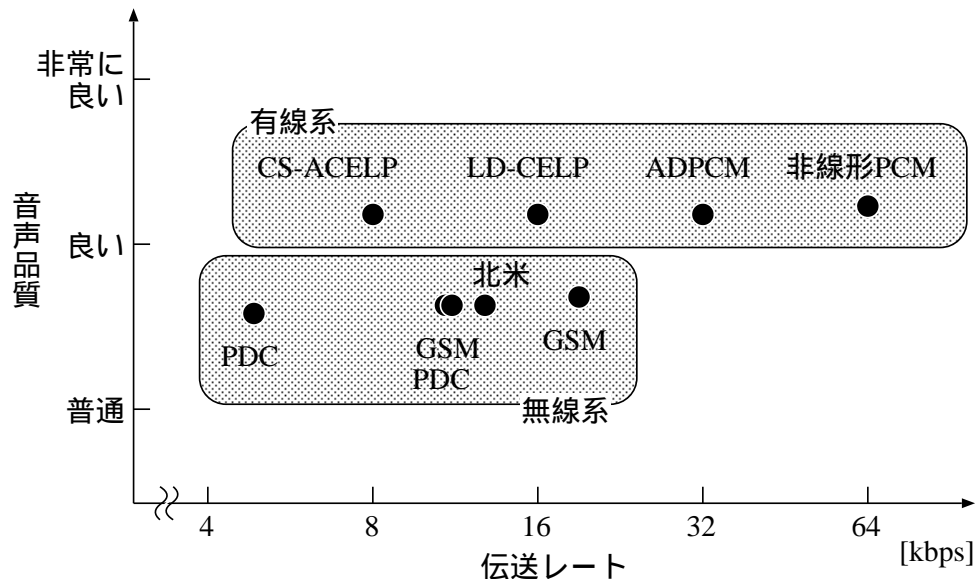


図 1.1 音声コーデックの伝送レートと音声品質の関係

品質を実現する音声コーデックであることを示している。また、図 1.2 に ITU-T8kbps 音声コーデック方式と PDC5.6kbps 音声コーデック方式について通信路誤り率と音声品質の関係を示す。図から ITU-T8kbps 音声コーデック方式は通信路での誤りが多くなると急激に音声品質が劣化することがわかる。本図においては、PDC5.6kbps 音声コーデック方式のように音声品質曲線が平坦であること、つまり通信路誤り率に関わらず誤りなし時の音声品質を維持することが望ましい。

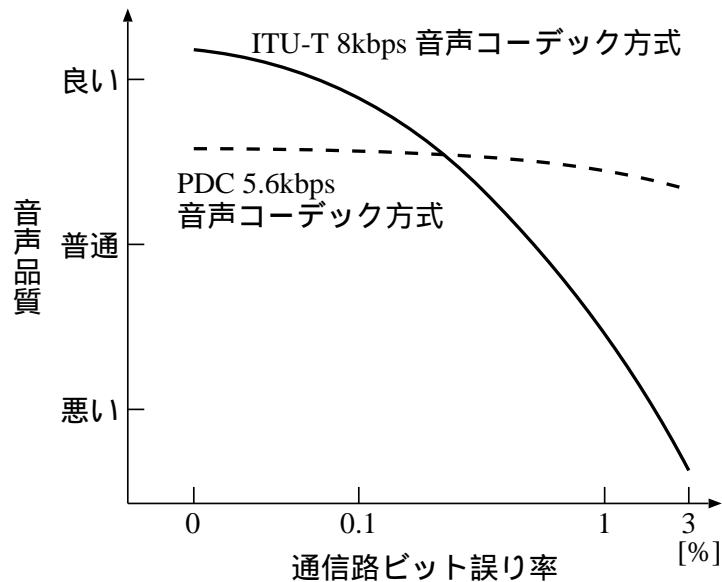


図 1.2 通信路誤り率と音声品質の関係

通信路での誤りに対して強い音声コーデック方式は，音声信号の情報を圧縮する情報源符号化部と通信路誤りに対して強い符号化をする通信路符号化部に分割し，それぞれを効率良く設計することで実現している．情報源符号化部では入力信号に含まれる冗長性，つまり信号間の相関を除去することで情報を圧縮する．信号間に相関があるということは，ある信号系列に継続する信号を予測できるということであり，情報を伝送しなくても予測によって復号できることを意味している．そこで，予測できる部分を削除することで無駄を省くことができ，情報を圧縮することができる．正確に予測するためには信号系列の精密なモデル化 (情報源モデル) が必要であり，音声コーデックでは情報源モデルとして物理的なモデルと統計的なモデルを用いて情報圧縮を図っている．一方，伝送する情報の冗長度が減ることにより伝送される情報が，全て，欠くことのできない情報となるため通信路での誤りが復号信号に大きく影響するようになる．そこで通信路符号化部では冗長性を付加することで誤り耐性を向上させる．付加する冗長性は代数演算に基づくものが多いが，通信方式に依存して通信路符号の遷移を拘束する [4] ことで冗長性を付加しても誤り耐性を向上させることができる．つまり，既知の冗長性を付加することで，冗長な部分を利用して通信路での誤りを検出することができ，誤り耐性を向上させることができる．このことは情報源符号化後に残留する冗長性についてもあてはまる．

以上のことから，情報源モデルを高精度化することで低ビットレート化を図ることができ，また，モデルによって冗長度を定量的に推定することができれば誤り耐性の向上に利

用することができると考えられる．先にも挙げたように情報源モデルには物理モデルと統計モデルがあるが，これまで標準化されてきた方式では統計モデルが十分に利用されているとはいえない．

隠れマルコフモデル (HMM) は音声認識処理において目覚しい成果を挙げている音声信号の統計モデルである．HMM を用いることで，ある信号列を情報源モデルが出力する確率を推定することができる．つまり上記の要件を満たす情報源モデルであり，本モデルを適用することで音声コーデックの高能率化を図ることができると考えられる．これまでに音声コーデックに対して HMM を適用した例は 2 章で説明する ‘認識による合成’ に基づく低ビットレート音声コーデックのみであり，商用として十分な復号音声品質 (toll quality) を実現するには至っていない．そこで本論文では従来の音声コーデック処理により得られる符号列をモデル化することで音声コーデックの高能率化を図ることを検討する．このような構成とすることで，既存の音声コーデックに対して音声品質を維持したまま低ビットレート化すること，あるいは，新たな情報を付加することなく通信路誤りに対する耐性を向上させることができる．低ビットレート化は図 1.1 において各音声コーデックの音声品質を示す黒点を左側にシフトすることに対応し，誤り耐性の向上は図 1.2 において大きな通信路誤り率における音声品質の劣化を少なくし音声品質曲線を平坦にすることに対応する．特に，情報を付加せずに性能向上を図る後者は通信路での符号系列を変更する必要がないため，復号化側のみに適用することとなる．実際に運用されているシステムでは復号化側の規格を厳密に定めることはなく，製品化において性能改善の余地が残されているため，この技術は非常に有用である．

音声コーデックへ HMM を適用するためには

- HMM を適用することで音声コーデックの性能を向上させるアルゴリズム
- 音声コーデック処理で得られる符号列の効果的なモデル化

について検討した後，実システムに対する実装時の課題を克服していく必要がある．第 1 の項目は音声コーデックに対して HMM を適用する上での基礎検討であり，第 2 の項目は実用化に際し HMM を適用する音声コーデック方式に依存する検討である．本論文は第 1 の項目に関するものであり，音声コーデックで利用されるパラメータの量子化系を仮定し，音声コーデックの高能率化を図る HMM の適用手法について提案する．

上記目的に対して本論文は以下の構成をとる．まず，2 章において音声コーデックにおける情報源モデルの利用について解説する．情報源モデルは物理モデルと統計モデルに分類することができる．本章では，動的な統計モデルであるマルコフモデルと HMM を利用することで高能率化を図った既存の手法を中心に解説し，その問題点を述べることで本論

文の位置付けを明確にする．3章ではHMMについて説明する．4章では情報源モデルを用いた低ビットレート化手法としてHMMを情報量削減に適用する手法を提案する．本手法はHMMを利用して推定されるパラメータの出力確率に基づいて最適な符号化系を逐次構築するものである．この章で提案法のアルゴリズムを説明し性能を評価する．5章では4章で提案した情報量削減手法の通信路誤り耐性について論ずる．情報量削減において冗長性を除去することで通信路誤りの影響を受け易い系となる．この章では提案した情報量削減法の構成で特に誤りの影響を受け易いブロックを誤りに対して頑健にする構成を検討する．6章では情報源モデルを用いた誤り耐性の向上手法としてHMMを誤り訂正に適用する手法を提案する．復号化側のHMMによって推定されるパラメータの出力確率は受信したパラメータの確からしさと考えることができる．そこで，提案法は，この確からしさを用いて通信路での誤りを訂正するものである．この章で本方式のアルゴリズムを説明し性能を評価する．最後に7章で本論文をまとめ，今後の課題を記述する．



## 第2章 音声コーデックにおける情報源モデルの活用

### 2.1. 序言

音声コーデックはモデル化を通して復号音声の品質を維持しつつ低ビットレート化や通信路誤りに対する耐性の強化を図る．モデルには信号の生成過程を模する物理モデルと生成された信号を確率過程として統計的に模する統計モデルがあり，それぞれ音声コーデックの高性能化に利用されている．

本章は音声コーデックで利用されている情報源モデルとして物理モデルと統計モデルを概説したのち，これまでに検討されてきた音声コーデックへの適用例を処理ブロック毎に分類して説明する．最後に従来法における問題点を述べ，本論文における情報源モデルの適用方針を説明する．

### 2.2. 物理モデルと統計モデル

音声コーデックにおける処理の流れを図 2.1 に示す．送信側では情報源から発声された

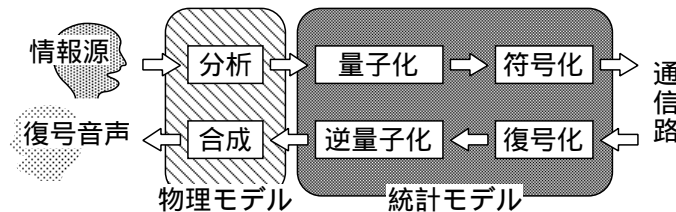


図 2.1 音声コーデック処理の流れ

音声信号を分析，量子化，符号化し通信路符号列として通信路へ送出する．受信側は受信した通信路符号列を復号化，逆量子化，合成することで復号音声を出力する．

以下に各ブロックの動作を簡単に説明する．分析ブロックでは情報源から発声された音声入力信号系列を物理モデルに基づいてパラメータへ変換する．量子化ブロックは抽出さ

れたパラメータを定められた符号語へ量子化(離散化)する。量子化にはパラメータと符号語を一対一対応させるスカラー量子化と複数のパラメータをまとめて一つの符号語へ対応させるベクトル量子化がある。符号化ブロックでは符号語に対して通信路符号を割り当てる。一般に符号語には符号語を蓄積している符号帳によるインデックス(以下、符号語インデックス)が付与されており、この符号語インデックスを通信路符号として用いることが多い。そのため量子化と符号化は同意語として用いられることが多いが、本論文では両者を区別して用いる。復号化、逆量子化ブロックでは、それぞれ、符号化、量子化ブロックの逆変換を行なう。復号信号は合成ブロックで復号パラメータを用いて物理モデルを駆動することで得る。以上のブロックは全て各ブロックに対する入力信号系列をモデル化することで効率良い系を構成している。例えば、音声コーデック処理における分析ブロックは情報源をモデル化した物理モデルに基づいて入力音声信号系列を分析することでモデルパラメータを抽出する。送受信側、双方に同じ物理モデルを有することで抽出したパラメータのみを伝送すれば入力音声信号系列を復元できることから、情報圧縮が可能となる。

音声コーデックで一般的に用いられる物理モデルは図 2.2 に示す音声生成過程をモデル化したものである。このモデルにおいて音声信号の長期的な相関(周期成分)は声帯を模し

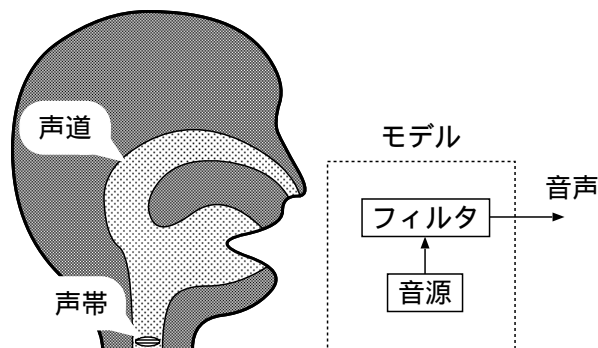


図 2.2 音声生成モデル

た音源に起因し、短期的な相関(ホルマント成分)は声道を模したフィルタによって生じる。移動体通信のように実時間通信が要求される音声コーデック処理では入力音声信号を一定の短時間単位でフレーム分割しフレーム毎にモデルパラメータを抽出する。図 2.3 に図 2.2 に示す物理モデルに基づく音声コーデック系として、1985年に Schröder らが提案し、ビットレートが中程度の音声コーデックで幅広く用いられている CELP(Code-Excited Linear Prediction) ボコーダ [5] のブロック図を示す。

CELP は長期的な相関を有する音源をモデル化した長期予測フィルタと声道フィルタをモデル化した短期予測フィルタに対し駆動信号を入力して音声を合成するボコーダである。



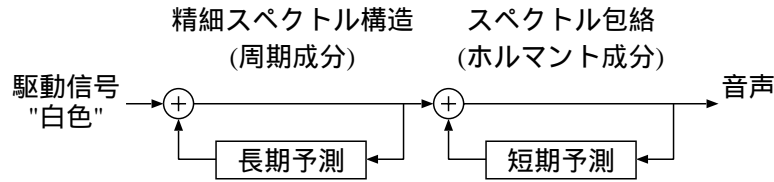


図 2.3 符号励起線形予測 (CELP) ボコーダ

駆動信号は信号間に相関がなく周波数特性が平坦な白色雑音系列を蓄積した符号帳から入力音声信号に最も近いものが選択され伝送される．長期予測フィルタではフレーム単位で再帰的に更新される符号帳を用いることで漸次変化していく音源信号のフレーム間の相関を利用した量子化系を構築する．一方，短期予測フィルタは入力音声信号の近接相関を全極型フィルタの係数として線形予測分析によって抽出することでフレーム内の相関を一組の線形予測係数で表す．抽出されたフィルタ係数はフレーム毎に量子化され伝送するのが一般的であるが，これらの係数にはフレームに跨った相関があり，更に情報圧縮することができる．特に，音声コーデック系の低ビットレート化に伴い，全体の情報量に対して大きな割合を占める線形予測係数に対してフレーム間の相関を利用した量子化方式が求められるようになってきた．大室らが提案した移動平均型フィルタを用いた量子化 [6] はこれを実現する一手法で，フレームを軸として並べたパラメータ列に対するフィルタ処理によりフレーム間の相関を削減する手法である．この手法は大きな情報量削減効果があり ITU-T による 8kbps 音声コーデック方式 [7] にも適用されている．一方，これと同様の効果は統計モデルを利用することでも実現することができる．

次に音声コーデックにおける統計モデルの利用について概観する．音声コーデックでは量子化ブロックと逆量子化ブロックを統計的に設計する．つまり，学習用データを分析し抽出したパラメータを用いて量子化系を設計する．量子化系はパラメータの時変化を考慮せずに全パラメータをまとめて学習するものが一般的で，これは静的な統計量に基づく量子化系であるといえる．しかしながらフレーム間の相関を考慮することができる動的な統計量に基づく量子化系の方が情報量削減効果があることは明らかである．音声信号処理で用いる動的な統計モデルとして代表的なものにマルコフモデルと隠れマルコフモデルがある．そこで，これら動的モデルに基づく音声コーデックについて既存の手法を検討する．

1章で述べたように，モデルに基づく音声コーデック系を構築することで情報量の削減や誤り耐性の向上を図ることができる．統計モデルは図 2.1 に示したように量子化と符号化，復号化，逆量子化の各ブロックに適用することができ，適用するブロックと適用方法によって情報量削減や誤り耐性の向上を目的とした系を構築することができる．表 2.1 と

表 2.2 にマルコフモデルと HMM を各ブロックに適用したのものとしてこれまでに提案されている手法をまとめる．なお，逆量子化は量子化の逆変換であり，ほぼ同じ構成となるため，ここでは量子化として一つにまとめる．

表 2.1 音声コーデック系へのマルコフモデルの適用

ブロック	情報量削減	誤り耐性向上
量子化	有限状態ベクトル量子化 トレリス量子化	トレリス量子化
符号化	適応エントロピー符号化	—
復号化	—	誤り隠蔽法 最尤推定法

表 2.2 音声コーデック系への隠れマルコフモデルの適用

ブロック	情報量削減	誤り耐性向上
量子化	Phonetic VOCODER	—
復号化	—	再帰的最尤推定法

以下の節で本表に示した適用例についてブロック毎に詳説する．

### 2.3. 量子化に対する統計モデルの適用

量子化ブロックでは分析ブロックで抽出されたパラメータを量子化し，量子化値に対応する符号語インデックスに変換する．一般に，量子化は系を構築する際に使用する学習データの静的な統計量に基づいて設計するが，ここでは情報源の状態変化を適応的に推定することができる動的な統計モデルに基づく量子化法について説明する．

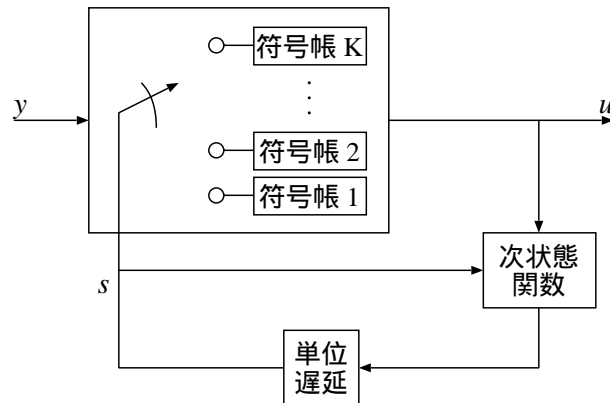
マルコフモデルに基づく量子化の代表的なものに有限状態ベクトル量子化 (以下, FSVQ) , 及び, FSVQ の状態更新にシフトレジスタを導入したトレリス量子化がある．一方, 隠れマルコフモデルに基づく量子化は Phonetic VOCODER に代表される ‘認識による合成’ 系が主である．両者の大きな違いは, マルコフモデルに基づく量子化は局所的なモデルの状態に対応して量子化系を適応させるのに対し, 隠れマルコフモデルに基づく量子化は入力信号を最もよく表すモデルを選択, 伝送し逆量子化では伝送されたモデルを用いて信号を生成する点である．

以下にこれらの量子化系について説明する．

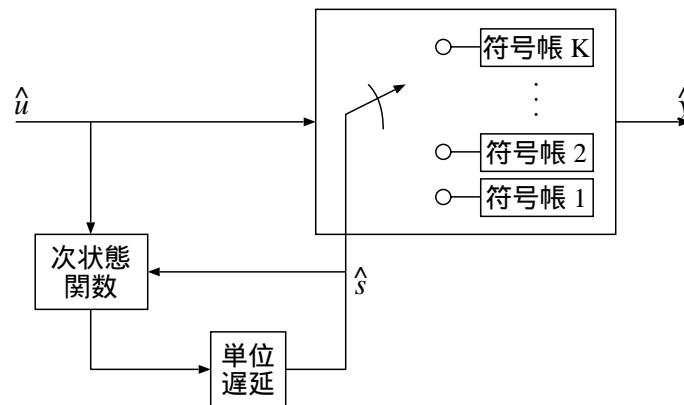
### 2.3.1 マルコフモデルに基づく量子化

過去の符号語インデックスに対応して量子化系を更新する再帰型ベクトル量子化 [8] は、入力されるパラメータ列に相関がある場合、定められた伝送ビットレートで量子化歪みを低減する上で非常に有効な手法である。入力パラメータ列をマルコフモデルでモデル化する FSVQ はこの再帰型ベクトル量子化の状態数を有限個に限定したものである。

FSVQ のブロック図を図 2.4 に示す。図のように FSVQ は量子化した符号語インデックス



(a) 量子化処理



(b) 逆量子化処理

図 2.4 有限状態ベクトル量子化ブロック図

に従って次に用いる符号帳を切替えることで動的に量子化系を更新する。ここで FSVQ を定式化する。状態数が  $S$  の状態空間  $S$  及び符号語インデックス空間  $\mathcal{N}$  を考えると、FSVQ は初期状態  $s(0)$  と全ての  $s \in S$  ならびに  $u \in \mathcal{N}$  に関して定義された入力パラメータ  $y$  に対

する符号化写像関数  $\phi(y, s)$  , 復号化写像関数  $\psi(u, s)$  そして次状態関数  $f(u, s)$  によって規定することができる .

符号化写像関数は入力パラメータ  $y(t)$  を次式によって符号語インデックス  $u(t)$  へ変換する .

$$u(t) = \phi(y(t), s(t-1)) \quad (2.1)$$

$$s(t) = f(u(t), s(t-1)) \quad (2.2)$$

復号化写像関数は受信した符号語インデックス  $\hat{u}(t)$  を次式で復号パラメータ  $\hat{y}(t)$  へ変換する .

$$\hat{y}(t) = \psi(\hat{u}(t), \hat{s}(t-1)) \quad (2.3)$$

$$\hat{s}(t) = f(\hat{u}(t), \hat{s}(t-1)) \quad (2.4)$$

式(2.2) , (2.4) より , 量子化と逆量子化の初期状態が同じで送受信通信路符号が  $u(t) = \hat{u}(t)$  であれば , 状態に関する情報を伝送せずに量子化と逆量子化の状態を同期して更新することができることがわかる . しかしながら , 通信路で誤りがあった場合  $u(t) \neq \hat{u}(t)$  となり , 誤り以後の量子化と逆量子化の状態に不整合が生じる . この状態間の不整合が再び同期することは補償されていない . このことから , FSVQ は通信路誤りに弱い量子化法であるといわれている .

FSVQ では写像関数  $\phi, \psi$  が状態  $s$  に依存しており , 状態に応じて量子化に用いる符号帳が切り替わる . 状態  $s$  で使用する符号帳を状態符号帳  $\mathcal{C}_s$  と呼び ,

$$\mathcal{C}_s = \{\psi(u, s); \text{すべての } u\} \quad (2.5)$$

で表すことができる . 上位符号帳 (super codebook)  $\mathcal{C}$  はすべての状態符号帳の和集合 , つまり状態符号帳に含まれるすべての符号語の集まりとして定義される .

$$\mathcal{C} = \bigcup_{s \in \mathcal{S}} \mathcal{C}_s = \{\psi(u, s); u \in \mathcal{N}, s \in \mathcal{S}\} \quad (2.6)$$

以上は状態符号帳が状態毎に定義されている状態ラベル型 FSVQ に基づく説明である . 他の型の FSVQ として遷移ラベル型 FSVQ があり , これは状態間の遷移に対して符号帳が定義されている . 状態ラベル型と遷移ラベル型は有限状態オートマトンの特別な型として , それぞれ , Mealy 型と Moore 型オートマトンと関連づけられている [8][9] .

FSVQ の設計には状態符号帳  $\mathcal{C}_s$  と次状態関数  $f(u, s)$  を最適化する必要があり , 理想的な FSVQ を設計することは不可能である [8] . 特に次状態関数の設計が難しく , これに関し条件

付きヒストグラム法[10]，最近傍法 (Nearest Neighbor design)，集合分割法 (set partitioning design)，全知法 (omniscient design)[11][12]，確率的オートマトン法等，様々な手法が提案されているものの，これらの手法が最適化過程であることの保証はない。

FSVQの音声コーデックへの適用は，Haouiらが提案した 16kbps 音声コーデック [11] に遡ることができる．この中で FSVQ は PVQ (Predeictive Vector Quantizer) として提案され，従来のベクトル量子化と同等の性能を実現する場合 PVQ は処理量，メモリ量の双方において優位であることが示された．続いて Dunham らにより FSVQ は線形予測係数の量子化に適用され従来法と比較して約 15 ~ 20% 情報量を削減できることが示された [12]．更に Cheng らは 450 ~ 600bps の音声コーデックに FSVQ を適用することで客観評価値を約 30% 向上させることを示した [13]．このように，音声コーデックにおいて動的な統計モデルあるマルコフモデルに基づく量子化は効果があることが示されている．

トレリス量子化は FSVQ の状態をシフトレジスタで表したものであり，通信路誤りの影響をシフトレジスタ長に抑えることができる点で優れている．そのため，トレリス量子化は通信路での誤りを考慮した情報源符号化法であるということから情報源-通信路統合符号化法の一実現方法として知られている．また，次状態関数がシフトレジスタによる拘束から一意に定まるため，状態符号帳のみを最適化すれば良いので学習法が確立されている．1981 年に Dunham らが誤り耐性のみならず情報圧縮においてもトレリス量子化が有効であることを示して以来 [14]，多くの研究がされてきた [15][16][17][18]．音声コーデックでは波形符号化への適用 [19][20]，CELP 符号化方式の音源への適用 [21][22]，LSP の量子化への適用 [16][23] 等が検討され，現在では ITU-T による 16kbps 音声コーデック方式にも適用されている [24]．

### 2.3.2 隠れマルコフモデルに基づく量子化

1986 年に Farges らにより初めて HMM が音声コーデックへ適用された [25]．提案された手法では，従来のベクトル量子化のように符号語インデックスを復号側に伝送するのではなく，HMM の状態系列を伝送する．量子化では入力パラメータ系列を生成する確率が最大となる状態系列を求め，逆量子化では連続性を考慮してパラメータ系列を状態系列から最尤推定する．この手法を更に押し推め低ビットレート化を図ったものが Phonetic Vocoder に代表される“認識による合成”と呼ばれる手法 [26][27][28] である．

Phonetic Vocoder では，送信側は認識過程として HMM による複数の音韻モデルから入力パラメータ系列を出力する確率が最も大きいモデルを選択しモデル系列情報を伝送する．受信側では受信したモデル系列を用いてパラメータ系列を推定し，そのパラメータを用い

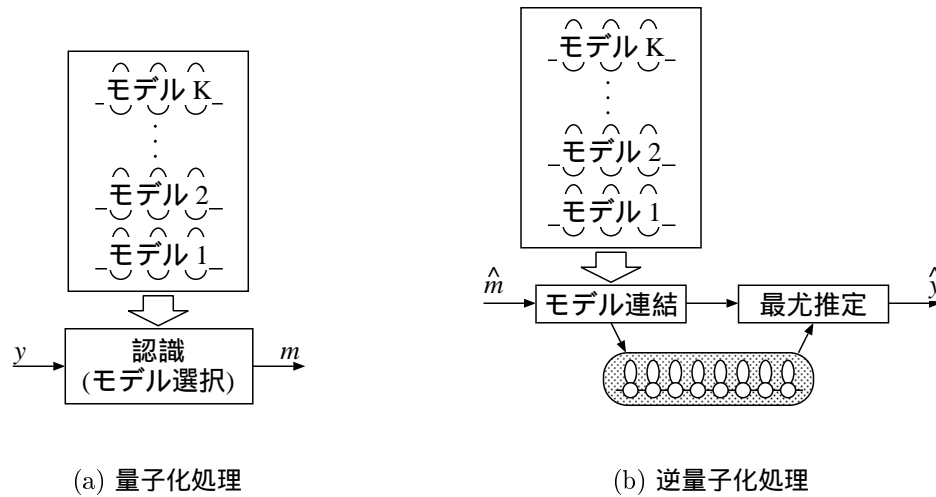


図 2.5 Phonetic VOCODER ブロック図

て音声を合成する．これにより，明瞭度を維持しつつ 100 ~ 400bps で線形予測係数を伝送することができる [26]．

送信側で行なう認識過程は従来から検討されている音声認識方法 [29][30] をそのまま適用することができる．一方，受信側で必要となるモデル系列からのパラメータ系列推定には確立された手法がなく，音韻列に対応するパラメータをあらかじめ学習しておく手法 [26]，一意に定めた状態系列から最尤推定を用いて再帰的にパラメータ系列推定する手法 [31][32][33] 等が提案されている．

これらの手法の問題点は HMM の特徴である確率的な状態遷移を，処理量上の問題から，一意に定める必要があることである．この制約により HMM が本来持っている信号源モデルとしての能力が損なわれる可能性がある．また，HMM に基づく量子化は信号源がパラメータを出力する確率を尺度として量子化するので，入力パラメータに対して歪みを最小化する従来のベクトル量子化とは異なる．これらの理由により‘認識による合成’法は極低ビットレートで明瞭度の高い音声を復号することはできるものの，入力パラメータを忠実に再現することは難しく toll quality の音声コーデック系を実現するためには多くの課題が残されている．

## 2.4. 符号化に対する統計モデルの適用

符号化ブロックでは量子化で選択された符号語インデックスに対して通信路符号を割り当てる．この通信路符号は固定長符号と可変長符号に大別できる．情報圧縮を考えた場合，音声のように無音部と有音部で情報量がダイナミックに変動する信号に対しては可変長符号の方が効率が良いことは明らかである．これに対して，固定長符号は一定の符号長で閉じた誤り検出処理が可能のため，可変長符号と比較して通信路誤りに強い符号列を構成することができる．ここでは代表的な可変長符号であるエントロピー符号化について述べる．

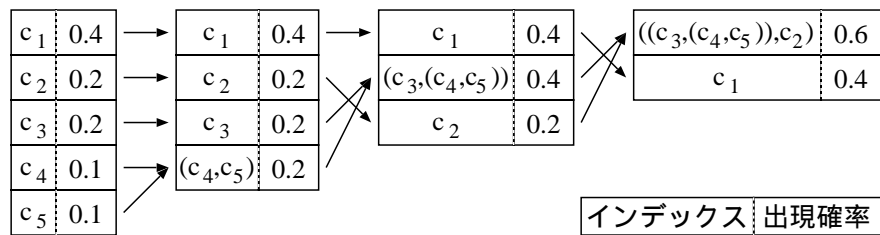
### 2.4.1 エントロピー符号化

エントロピー符号化は可変長符号を実現するもので，統計的に発生頻度の高い符号語インデックスに短い符号長，低い符号語インデックスに長い符号長を割り当てることで平均的な符号長を短くする符号化方法である．複数のパラメータを一つの符号語インデックスに対応させるベクトル量子化に対してエントロピー符号を適応する場合，符号化ブロックでの通信路符号割り当てを考慮せずに量子化用の符号帳を設計すると必ずしもシステム全体として最適な系とならない．そこで，Entropy-Constrained Vector Quantization(ECVQ)法が Chou らによって提案された [34]．これは，量子化ブロックの設計に際して符号化ブロックでエントロピー符号化することを考慮して，量子化歪みとエントロピーの和を尺度として符号帳を作成する手法であり性能の向上が確認されている．以上は静的統計量に基づく符号化系の構築に関する説明である．

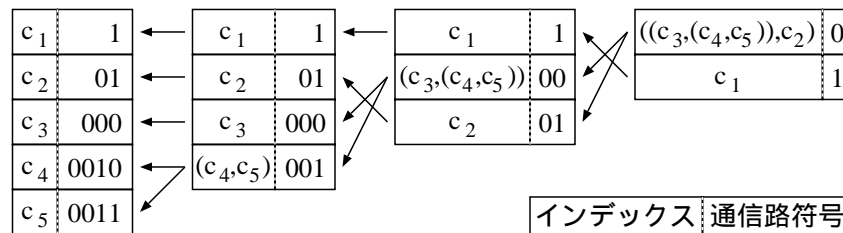
一方，動的なエントロピー符号化法も提案されている．まず，事前に学習データを用いて符号化系を構築するのではなく，符号化時に入力符号語インデックス系列に応じて符号化系を逐次更新する手法が提案されている [37][38]．これらは適応エントロピー符号化 [35][36] と呼ばれ情報源の局所的な変化に対応して符号化系を構成することができる．しかしながら，情報源モデルではなく入力符号語インデックス系列によって符号化系を更新するため必ずしも最適な符号化がされるわけではない．Chou らは，情報源モデルに基づく符号化系として入力符号語インデックス系列に対して ECVQ を導入し，入力符号語インデックス系列に関する条件付確率に基いてエントロピー符号化する Conditional Entropy Coded Conditional-ECVQ[39] を提案した．これは，FSVQ の状態符号帳をエントロピー符号化した [40] と等価であり，FSVQ とエントロピー符号化を組み合わせることによって効率良い音声情報圧縮ができることを示している．

エントロピー符号化の代表的な実現法としてハフマン符号と算術符号 [41][42] がある．こ

ここでは図 2.6 を用いて通信路符号シンボルが 2 つ  $\{0, 1\}$  の時のハフマン符号系の構築法を簡単に説明する [43] .



(a) 逐次縮退過程



(b) 逐次分割過程

図 2.6 ハフマン符号系の構築例

符号語インデックス  $\{c_1, c_2, \dots, c_5\}$  の出現確率が図 2.6 のように与えられているとする . ハフマン符号系は逐次縮退と逐次分割によって構築される . まず , 逐次縮退では出現確率が最も小さなインデックスをまとめて , 新たなインデックスを割り当てる . 図では , まず ,  $c_4$  と  $c_5$  をまとめて  $(c_4, c_5)$  というインデックスを割り当て , このインデックスの出現確率を  $c_4$  と  $c_5$  の出現確率の和とする . この処理をインデックスの数が通信路符号シンボル数となるまで繰り返す . 逐次分割では , まず , 逐次縮退で最終的に残ったインデックスに通信路符号シンボル  $u_r$  を割り当てる . 次に , 矢印に沿って , 矢印が 1 本ときは矢印の先に通信路符号  $u_r$  を複製し , 矢印が 2 本ときは矢印の先にそれぞれ  $u_r 0$  と  $u_r 1$  を書き込む . この処理を始めに与えられた符号語インデックスに対して通信路符号が割り当てられるまで繰り返す . 以上の処理によってハフマン符号系を構築する .



## 2.5. 復号化に対する統計モデルの適用

通信路符号化は送信符号列に対して既知の冗長性を付加し通信路符号とすることで通信路での誤りを検出する。この付加する冗長性は代数演算に基づいているが、更に、通信路モデルを通信路符号化に採り入れることで通信路符号間に冗長性を付加することができ誤り耐性を向上させることができる [4]。一方、情報源モデルを利用した復号化手法も検討されている。2.2 に示したように、量子化と同様用いられる情報源モデルにはマルコフモデルと隠れマルコフモデルがある。

### 2.5.1 マルコフモデルを利用した復号化

Fingscheidt らは通信路符号の復号時に軟判定復号 [44] によって得られる通信路符号の確からしさを入力として符号語インデックスのマルコフモデルを駆動し復号化することで誤りの影響を低減することができることを示した [45]。更に、動的な統計モデルである 1 次マルコフモデルは静的統計量を用いた場合より通信路誤りの影響を受けにくくする効果があることを示した [46]。これらは通信路の誤りに対して誤りを訂正するのではなく隠蔽することを目的とした復号法であり、符号語インデックスに冗長性が多く残存しているときに有効であることが報告されている [47][48]。

一方、誤り訂正を目的とした手法として、Alajaji らはマルコフモデルと通信路符号化方式を組み合わす手法を提案した [49]。この中で、米国連邦政府標準 (US Federal Standard) FS1016 規格 4.8kbps 音声符号化方式 [1] で用いられている線スペクトル対 (LSP) パラメータ量子化後の符号語インデックス列をマルコフモデルでモデル化することで、LSP パラメータに要する通信路符号の少なくとも  $\frac{1}{3}$  のビットが冗長であることを示した。そこでマルコフモデルと軟判定ビタビ復号と組み合わせる手法を提案し、最高 5 dB の符号化利得を実現することができることを報告している。

以上の報告より復号化処理にマルコフモデルを用いることで通信路誤りの復号結果への影響を低減することができることがわかる。

### 2.5.2 隠れマルコフモデルを利用した復号化

Garcia-Frias らは、HMM とターボ符号を組み合わせた誤り訂正手法を提案した [51]。ターボ符号 [50] の復号化のブロック図を図 2.7 に示す。冗長シンボル列は畳み込み符号化により付加されており、復号化には最大事後確率推定法 [52] または軟判定ビタビ復号 [53] を用

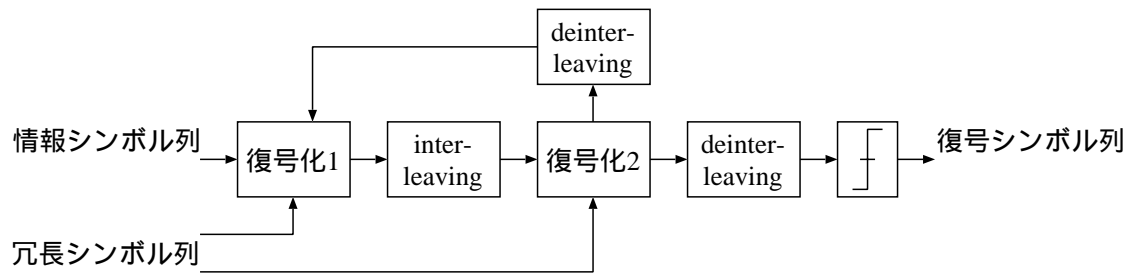


図 2.7 ターボ復号化ブロック図

いて畳み込み符号化器に対応するトレリスダイアグラムを用いて行なう．図 2.8 に畳み込み符号化器の例を示し，これに対応するトレリスダイアグラムと復号化例を図 2.9 に示す．文献 [51] では既知の HMM を情報源モデルとして用いて図 2.8 のトレリスダイアグラムを

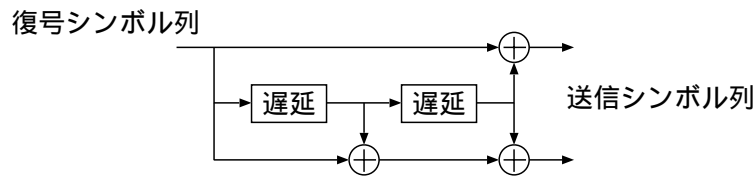


図 2.8 符号化率 1/2 の畳み込み符号化器の例

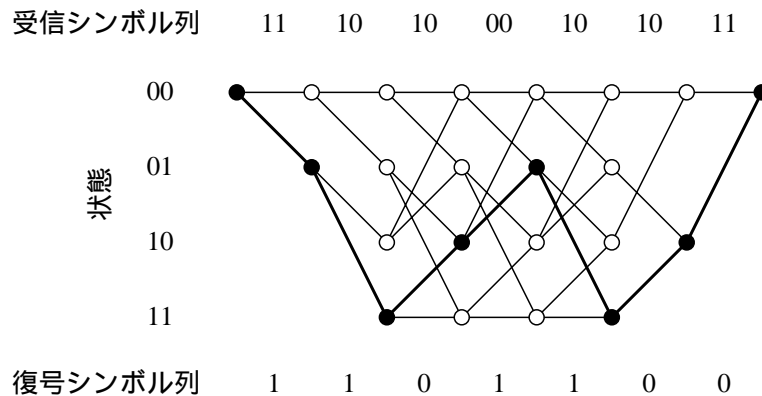


図 2.9 図 2.8 の符号化器に対応するトレリスダイアグラムの例

HMM 状態との積空間に広がるトレリスダイアグラムに拡張して最尤復号する．図 2.10 に情報源モデルの例を示し，このモデルによって図 2.9 のトレリスダイアグラムを拡張した例を 2.8 に示す．Garcia-Frias らは，この手法を既知の 2 状態 HMM が出力する 2 値符号

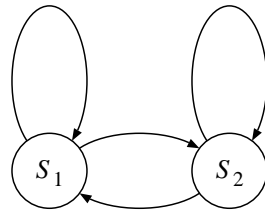


図 2.10 Gracia-Frias らが用いた情報源モデル

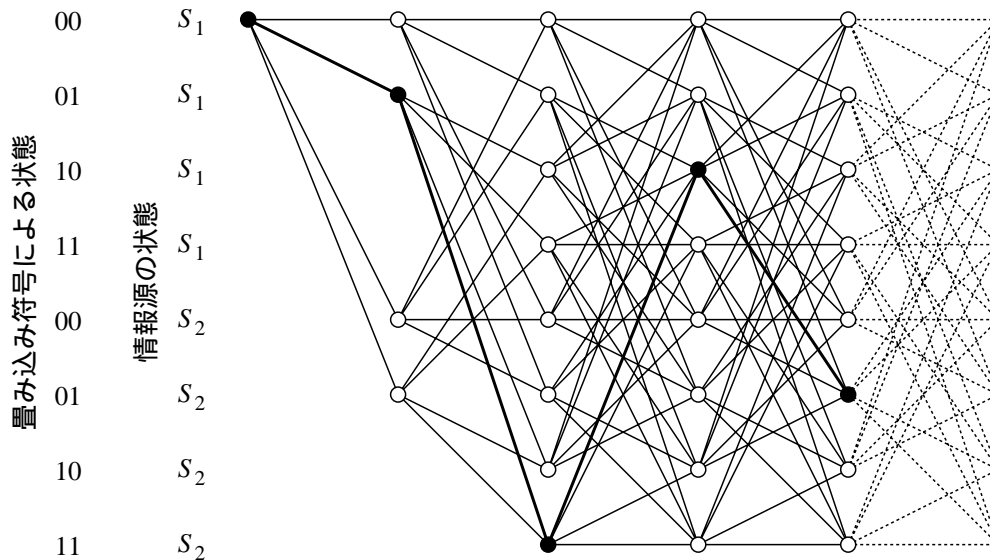


図 2.11 情報源を考慮した図 2.8 の符号化器に対応するトレリスダイアグラムの例

列に対して適用し，計算機シミュレーションにより符号化利得を向上させることができることを示した [51] .

本手法は，情報源が通信路シンボル数と同じ 2 値を出力することで容易に通信路符号化と組み合わせることを可能としている．しかしながら，音声コーデックのように情報源が複数の通信路シンボルに対応するパラメータを出力するモデルに対して適用する場合，復号化に用いるために拡張されたトレリスダイアグラムは膨大な状態数，及び，遷移を有することとなり，莫大な処理量を要することとなる．

## 2.6. 音声コーデックにおける隠れマルコフモデルの位置付け

これまで見てきたとおり，動的統計モデルを適用することで音声コーデックの効率を向上させることができる．しかしながら，その適用法にはまだまだ多くの課題が残されてい

る．ここで，2.3 節に記した‘認識による合成’に基づく音声コーデックを取り上げ，2カ国語間の同時通訳と対応させて考えることで，改めて動的モデルを適用する際の課題を考える．

2 者間での直接の会話に対して通訳を介在させる場合の問題点として通訳者に起因する遅延と誤訳がある．会話の最小単位は文であり，少なくとも一文を終えるまで他言語へ翻訳することはできず遅延が生じる．例えば，日本語では否定文と肯定文が文末で判明するため，英語のような文頭で否定文と肯定文を明確に定義する言語への変換には遅延を要する．また，文単位で正確に翻訳したとしても文脈として成り立たないこともあり，正確な翻訳には文脈を把握した翻訳が必要となるため，更に大きな遅延が生じる．つまり，遅延と誤訳にはトレードオフがあることがわかる．その他，優れた同時通訳者は事前に著作等を通じて被通訳者（話者）の思想や話し方を把握した上で通訳に望むといわれる．このことは遅延と誤訳を最小限とするためには情報源（被通訳者）モデルの高精度化が重要であることを示唆している．

次に‘認識による合成’に基づく音声コーデックについて考える．音声コーデックにおいて遅延は大きな問題である．携帯電話等に搭載される実時間処理の音声コーデックで許される遅延は 40ms 程度までである．音声認識で用いられるモデルの単位である音韻は平均 100ms[55] であり認識結果に基づくコーデック処理は実時間処理には不適切である．一方，量子化時の誤認識は選択したモデルが誤るということであり，モデルから出力されるパラメータに歪みが含まれるわけではないので復号音声品質を大きく劣化させることはないことが報告されている [26]．しかしながら，選択の誤りを復号音声品質へ影響しないようにするためには状態継続長などの付加情報が必要となることも報告されており [56]，誤認識は情報量の圧縮効率へ影響する．

以上のことから，モデル化によってコーデック系の効率を向上させるためには精度の高い情報源モデルが必要であることがわかる．マルコフモデルでは状態数が多いほどモデルの精度が高くなり，大きな性能向上が期待できる．しかしながら音声信号の場合，観測される信号列を高次マルコフモデルでモデル化することは難しく，モデルを学習するためのデータ量も大量に必要となる．一方，HMM はマルコフモデルを縮退したモデルであり以下の特徴がある．

- 小さなハードウェア規模で同等の性能を実現できる
- モデルを最適化するための学習法が確立している

つまり，マルコフモデルと比較して HMM は，モデル化の対象とする情報源からの観測信号のみが学習データとして与えられているときに，コンパクトで高精度の情報源のモデル化が可能である．そこで，本論文では HMM を音声コーデックに適用することとした．

本論文ではアルゴリズム上の遅延が小さく実時間処理で実現可能な手法による音声コーデックの効率向上を検討する．即ち，HMM を認識に用いるのではなく情報源の局所的な状態を推定するモデルとして利用し，この状態変化に対して音声コーデック系を適応させることを考える．そこで，情報量削減，及び，誤り耐性の向上に対し，それぞれ，符号化ブロック，及び，復号化ブロックへ HMM を適用することを検討した．これらは量子化ブロックと独立しており，復号音声品質を決定する分析/量子化-逆量子化/合成ブロックには既存の音声コーデックを用いることができる．

本論文では，線形予測パラメータの量子化系を仮定し，これに対して提案法を適用することで HMM を利用することによる情報量削減，及び，誤り耐性の向上の効果を確認する．

## 2.7. 結言

本章では音声コーデックへの適用を中心に，これまでに提案されてきた情報源モデルの利用方法をまとめた．情報源モデルは物理モデルと統計モデルに大別でき，音声コーデック処理における各ブロックへ適用することができる．

音声コーデックに対しては，復号音声の品質を維持しつつ低ビットレート化することと通信路での誤りの影響をできるだけ受けないことが望まれている．これらは動的統計モデルを用いることで効率良く実現することができる．本章では，統計モデルとしてマルコフモデルと HMM を用いることで情報量削減，あるいは通信路誤り耐性の向上を図る手法としてこれまでに提案されてきた手法について概説しその問題点を述べることで，本論文におけるモデル適用の位置付けを示した．



## 第3章 隠れマルコフモデル

### 3.1. 序言

1970年代中頃に Baker[29] や Jelinek[30][54] によって音声認識への適用が紹介された HMM は 1980 年代後半から急速に普及することとなった。更に、1990 年代に入り HMM の適用は音声認識のみならず音声合成、音声符号化等あらゆる音声信号処理へ広がり、現在では音声信号の統計モデルとして重要な役割を果たしている。

本章では HMM の音声認識への適用法を説明し、本論文で用いるエルゴード的 HMM について説明する。その後、モデルの学習、及び、音声コーデックへの適用上重要なフォワード-バックワードアルゴリズムについて説明する。

### 3.2. エルゴード的隠れマルコフモデル

HMM は信号系列を確率的に出力する情報源モデルである。音声認識では認識対象とする観測信号系列を出力する確率が最も大きくなるモデルを選択し、モデルと結びつけられたラベルを認識結果とする。本節では音声認識に対する HMM の適用について音声符号化の観点から考察する。本論文では HMM を用いて信号源の状態を逐次推定することで音声コーデックの効率の向上を図る手法を検討する。そこで、以下に、音声認識でよく用いられる left-to-right 型 HMM を用いて HMM の状態の遷移を説明した後、本論文で用いるエルゴード的 HMM について説明する。

図 3.1 に 5 状態 left-to-right 型 HMM の例を示す。図 3.1 の model 1 によりパラメータ系列  $\{p_1 p_2 \dots p_6\}$  が出力されたときの状態遷移とトレリスダイアグラムとの関係を図 3.2 に示す。HMM を利用して符復号化ブロックの状態を逐次更新する場合、各フレームにおけるパラメータの出力確率分布が必要となる。model 1 が図 3.2 において  $p_4$  を出力する確率は図の太線で示した遷移で  $p_4$  を出力する確率を足し合わせることで求めることができる。

ところで、実時間処理による音声コーデックのように連続音声入力信号を逐次処理する場合、モデルの終了状態に対応するフレームを規定することができない。従って全てのモデ

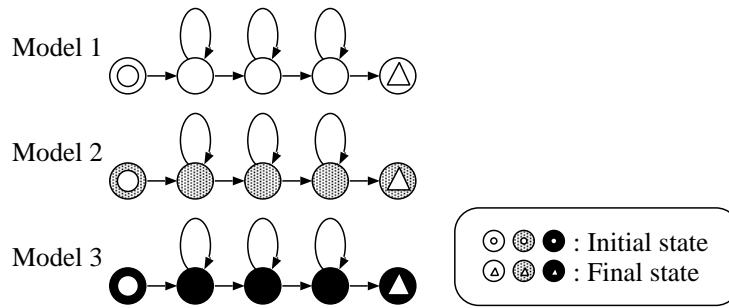


図 3.1 5 状態 left-to-right 型 HMM の例

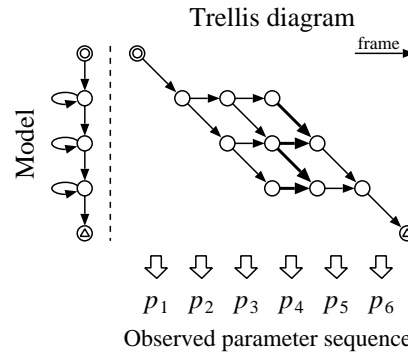


図 3.2 トレリスダイアグラムと出力パラメータ系列

ルの初期状態と終了状態を連結することで音声信号源モデルを構成する．図 3.3 に図 3.1 に示した 3 つのモデル系列を認識する場合のトレリスダイアグラムを示す．このように HMM による状態遷移は時間の経過と共に枝分かれが増大し膨大なトレリスダイアグラムとなっていくことがわかる．実際には，図 3.4 のように終了状態と初期状態を同じ状態にすることで，爆発的な枝分かれを避け，演算量を削減することができる．

以上が left-to-right 型 HMM を用いたときの議論である．次に，図 3.5 にエルゴード的 HMM を示す．エルゴード的 HMM は全ての状態間の遷移を許す最も一般的な情報源モデルである．1 つのエルゴード的 HMM で音声信号全体をモデル化し，全ての状態が初期状態，及び，終了状態を取り得るとすれば図 3.6 のトレリスダイアグラムのみを考えることで情報源の状態を推定できる．left-to-right 型 HMM は状態の遷移に指向性があるため音韻単位でのモデル化に優れているが各音韻に対してモデルが必要となり，複数のモデルで音声信号全体を表すこととなる．これに対してエルゴード的 HMM は指向性はないものの学習により状態毎に役割が分担され音声信号源の十分なモデルとなり得ることが報告されており [9]，この場合一つのモデルのみを考えればよい．そこで，本論文ではエルゴード的



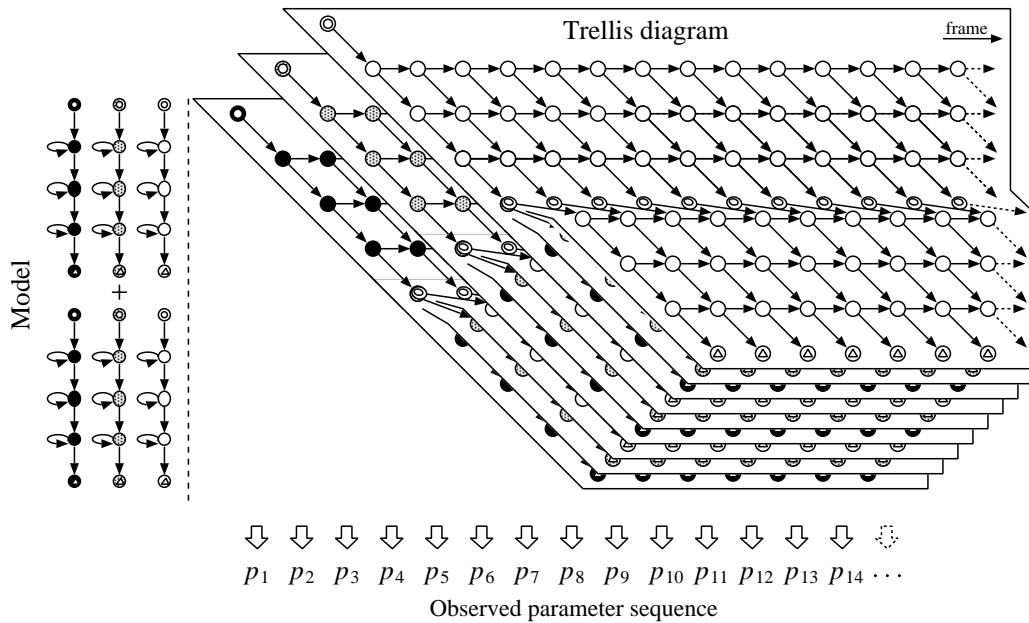


図 3.3 複数のモデルによるトレリスダイアグラムと出力パラメータ系列

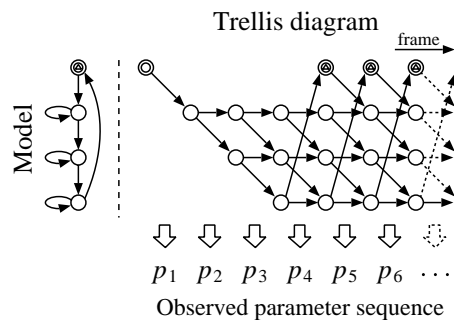


図 3.4 枝分かれを抑制したトレリスダイアグラムと出力パラメータ系列

HMM を用いることとした。

### 3.3. フォワード–バックワード アルゴリズム

本節では本論文で検討する方式で重要な部分を占め，HMM の学習に欠くことができないフォワード–バックワード アルゴリズムを説明する。

状態数  $S$  を有する HMM の状態空間  $S$  を定義し，時間と共に遷移する状態系列を  $x$ ，出力される符号語インデックス系列を  $y$  とする。

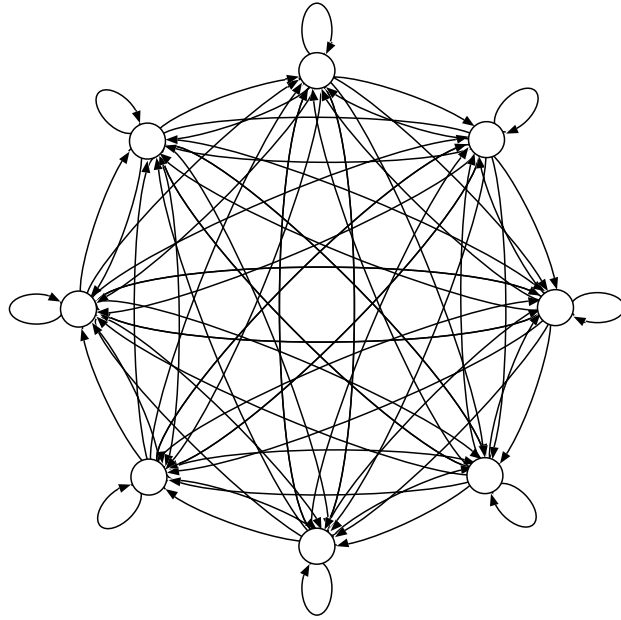


図 3.5 8 状態エルゴード的 HMM

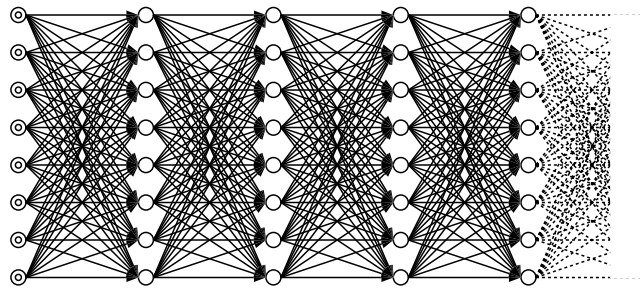


図 3.6 エルゴード的 HMM のトレリスダイアグラム

本論文ではモデルの状態が遷移すると共に信号を出力する Mealy 型 HMM[9] を用いる。HMM ではモデルの状態は式 (3.1) で示すように状態空間  $S$  に対する確率分布により表される。本論文ではこれを状態分布  $\pi$  と呼ぶこととする。

$$\pi(t) \stackrel{def}{=} \begin{bmatrix} P(x(t) = 1) \\ P(x(t) = 2) \\ \dots \\ P(x(t) = S) \end{bmatrix} \quad (3.1)$$

図 3.5 に示すエルゴード的 HMM の初期状態分布が  $\pi(1)$  のときに  $N$  個の符号語インデックス系列  $\mathbf{y} = \{y(t)|t = 1, \dots, N\}$  を HMM が出力する確率  $P(\mathbf{y})$  を考える。ただし、任

意の  $t$  に対して  $y(t) \in \mathcal{C}$  である．ここで， $\mathcal{C}$  は  $K$  個の符号語インデックスからなる符号語インデックス空間  $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$  である．また，要素が全て 1 の  $S \times 1$  行列  $U$  を定義し， $U^T \pi(1) = 1$  とする．ただし， $U^T$  は  $U$  の転置行列を表す．

HMM は状態を遷移する確率とその遷移において符号を出力する確率によって情報源をモデル化する．状態  $i$  から状態  $j$  に状態が遷移する確率を  $a_{i \rightarrow j}$ ，この遷移において符号  $k$  を出力する確率を  $b_{i \rightarrow j}(k)$  と表記することとする．これらの確率に対応して次の行列を定義する．

$$\mathbf{A} \stackrel{def}{=} \begin{bmatrix} a_{1 \rightarrow 1} & a_{1 \rightarrow 2} & \dots & a_{1 \rightarrow S} \\ a_{2 \rightarrow 1} & a_{2 \rightarrow 2} & \dots & a_{2 \rightarrow S} \\ \dots & \dots & \dots & \dots \\ a_{S \rightarrow 1} & a_{S \rightarrow 2} & \dots & a_{S \rightarrow S} \end{bmatrix} \quad (3.2)$$

$$\mathbf{B}(k) \stackrel{def}{=} \begin{bmatrix} b_{1 \rightarrow 1}(k) & b_{1 \rightarrow 2}(k) & \dots & b_{1 \rightarrow S}(k) \\ b_{2 \rightarrow 1}(k) & b_{2 \rightarrow 2}(k) & \dots & b_{2 \rightarrow S}(k) \\ \dots & \dots & \dots & \dots \\ b_{S \rightarrow 1}(k) & b_{S \rightarrow 2}(k) & \dots & b_{S \rightarrow S}(k) \end{bmatrix} \quad (3.3)$$

更に，行列要素毎の積演算を  $\otimes$  と表すこととする．

$$\mathbf{A} \otimes \mathbf{B}(k) \stackrel{def}{=} \begin{bmatrix} a_{1 \rightarrow 1} b_{1 \rightarrow 1}(k) & a_{1 \rightarrow 2} b_{1 \rightarrow 2}(k) & \dots & a_{1 \rightarrow S} b_{1 \rightarrow S}(k) \\ a_{2 \rightarrow 1} b_{2 \rightarrow 1}(k) & a_{2 \rightarrow 2} b_{2 \rightarrow 2}(k) & \dots & a_{2 \rightarrow S} b_{2 \rightarrow S}(k) \\ \dots & \dots & \dots & \dots \\ a_{S \rightarrow 1} b_{S \rightarrow 1}(k) & a_{S \rightarrow 2} b_{S \rightarrow 2}(k) & \dots & a_{S \rightarrow S} b_{S \rightarrow S}(k) \end{bmatrix} \quad (3.4)$$

また，

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B}(\mathbf{y}) &= \prod_{t=1}^N [\mathbf{A} \otimes \mathbf{B}(y(t))] \\ &= [\mathbf{A} \otimes \mathbf{B}(y(1))] [\mathbf{A} \otimes \mathbf{B}(y(2))] \cdots [\mathbf{A} \otimes \mathbf{B}(y(N))] \end{aligned} \quad (3.5)$$

と定義する．

以上の定義より，任意の  $t$  に対して符号  $y(t)$  を出力して状態  $i$  から状態  $j$  へ遷移する確率  $P_{i \rightarrow j}(y(t))$  は  $a_{i \rightarrow j} \times b_{i \rightarrow j}(y(t))$  となる．従って，状態分布  $\pi(t)$  が与えられたとき，HMM が  $y(t)$  を出力し状態  $j$  へ遷移する確率  $P(x(t+1) = j | y(t), \pi(t))$  は

$$P(x(t+1) = j | y(t), \pi(t)) = \sum_{i=1}^S P(x(t) = i) \times a_{i \rightarrow j} \times b_{i \rightarrow j}(y(t)) \quad (3.6)$$

となる．よって、 $y(t)$  を出力することで更新される状態分布  $\pi_{y(t)}(t+1)$  は

$$\pi_{y(t)}(t+1)^T = \begin{bmatrix} P(x(t+1) = 1|y(t), \pi(t)) \\ P(x(t+1) = 2|y(t), \pi(t)) \\ \dots \\ P(x(t+1) = S|y(t), \pi(t)) \end{bmatrix}^T = \pi(t)^T \{ \mathbf{A} \otimes \mathbf{B}(y(t)) \} \quad (3.7)$$

となる．HMM は信号系列  $\mathbf{y}$  に基づいて状態分布が遷移しており，HMM  $\mathcal{M}$  が  $\mathbf{y}$  を出力する確率  $P(\mathbf{y}|\mathcal{M})$  は  $y(N)$  を出力した後の状態分布が最終状態に遷移している確率の総和となる．ここでは，全ての状態が最終状態を取り得るエルゴード的 HMM を用いているため，モデルが  $\mathbf{y}$  を出力する確率は式 (3.8) となる．

$$\begin{aligned} P(\mathbf{y}|\mathcal{M}) &= \pi(N+1)^T \mathbf{U} \\ &= \pi(N)^T \{ \mathbf{A} \otimes \mathbf{B}(y(N)) \} \mathbf{U} \\ &= \pi(1)^T \{ \mathbf{A} \otimes \mathbf{B}(y) \} \mathbf{U} \end{aligned} \quad (3.8)$$

ところで，式 (3.7) は  $t$  の昇順に従う遷移に対する状態分布の定式化であった．同様の定式化は  $t$  の降順に従う遷移に関しても可能である．これをバックワードパス確率分布  $\beta(t)$  と呼び，以下のように定義する．

$$P(x(t) = i|y(t), \beta(t+1)) = \sum_{j=1}^S a_{i \rightarrow j} \times b_{i \rightarrow j}(y(t)) \times P(x(t+1) = j) \quad (3.9)$$

$$\beta_{y(t)}(t) = \begin{bmatrix} P(x(t) = 1|y(t), \beta(t+1)) \\ P(x(t) = 2|y(t), \beta(t+1)) \\ \dots \\ P(x(t) = S|y(t), \beta(t+1)) \end{bmatrix} = \{ \mathbf{A} \otimes \mathbf{B}(y(t)) \} \beta(t+1) \quad (3.10)$$

このバックワードパス確率分布  $\beta(t)$  に対応させて，先に定義した状態分布  $\pi(t)$  をフォワードパス確率分布  $\alpha(t)$  と呼ぶこととする．

図 3.7 に 4 状態のエルゴード的 HMM によるフォワードパス，バックワードパス確率分布の遷移例を示す．(a) は 4 状態 HMM のトレリスダイアグラムである． $\odot$  は初期状態， $\bullet$  は最終状態を表している．(b) は状態間の遷移により HMM が出力する符号列である．(c) はフォワードパス確率分布，及び，バックワードパス確率分布の遷移を示しており，出力される符号に対応して式 (3.7)，(3.10) より算出される．

$\beta$  の初期状態として  $\beta(N+1) = \mathbf{U}$  と定義すると式 (3.7) より明らかなように，

$$P(\mathbf{y}|\mathcal{M}) = \alpha(N+1)^T \beta(N+1) \quad (3.11)$$

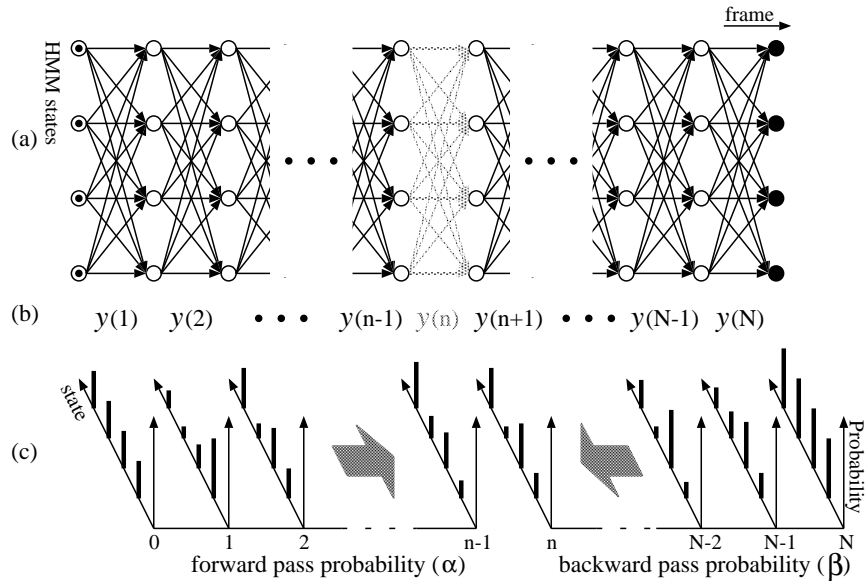


図 3.7 フォワード-バックワードパス確率分布の遷移例 : (a) 4 状態エルゴード的 HMM のトレリスダイアグラム,  $\odot$  初期状態,  $\bullet$  最終状態, (b) 出力符号列, (c) 確率分布の遷移

であり, 任意の  $t(1 \leq t \leq N + 1)$  に対して

$$P(\mathbf{y}|\mathcal{M}) = \boldsymbol{\alpha}(t)^T \boldsymbol{\beta}(t) \quad (3.12)$$

が成り立つ.

HMM の学習では,

$$\gamma(t) = \frac{\{\boldsymbol{\alpha}(t)\boldsymbol{\beta}(t+1)^T\} \otimes \mathbf{A} \otimes \mathbf{B}(y(t))}{P(\mathbf{y}|\mathcal{M})} \quad (3.13)$$

が  $y(t)$  を出力して状態間を遷移する確率行列を表しており, この値を用いて遷移確率行列  $\mathbf{A}$  と出力確率行列  $\mathbf{B}(c)$  を更新する. 更新した確率行列に基づき, 再度, 学習データ系列のフォワードパス, バックワードパス確率を算出し確率行列を更新する. この再帰処理を収束するまで繰り返すことで学習データ系列を出力する確率が高くなるように HMM を学習していく.

### 3.4. 結言

本章では, 本論文で用いるエルゴード的 HMM について説明し, HMM の学習で用いられるフォワード-バックワードアルゴリズムの定式化を行なった.

本論文では HMM を局所的な情報源の状態推定に用いる．エルゴード的 HMM はマルコフ情報源として最も一般的なモデルであり，一つのエルゴード的 HMM によって情報源をモデル化することができるのでアルゴリズムの定式化が容易になる．

次章からエルゴード的 HMM を用いた音声コーデックの高能率化について検討する．

# 第4章 隠れマルコフモデルを用いた情報量削減

## 4.1. 序言

本章では HMM を符号化ブロックと復号化ブロックに適用することで情報量削減を図る手法を提案する。

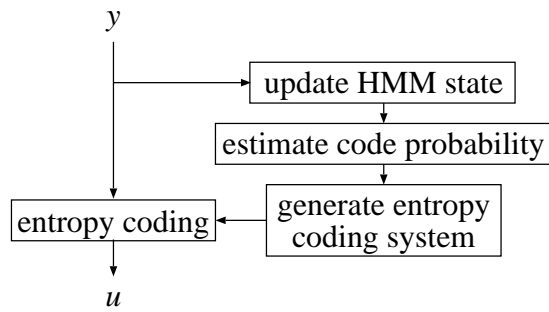
HMM で推定される局所的な状態に基づき符号化系を適応させることで情報量を削減できる。本章では符号化系としてエントロピー符号系を用いることとした。以下、4.2 に提案する情報量削減アルゴリズムを説明し、4.3 で評価する。

## 4.2. エントロピー符号化との組合せによる情報量削減

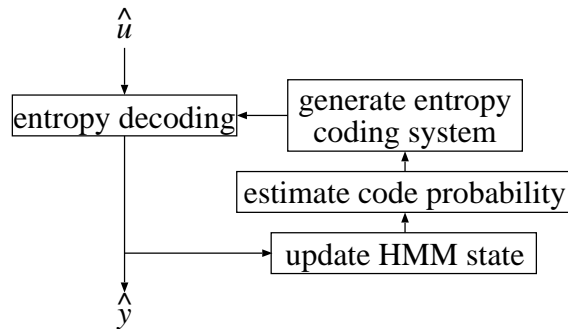
図 4.1 に HMM を利用した情報量削減のブロック図を示す。本方式は量子化ブロックから出力される符号語インデックスをエントロピー符号系で符号化/復号化することで情報量の削減を図る。エントロピー符号系は HMM によって推定される符号空間の確率分布の変化に適応して HMM の状態遷移毎に生成する。符号化、復号化の状態は量子化、及び、復号された符号語インデックスによって同期しつつ更新される。情報量削減アルゴリズムは次の 3 つのブロックからなる。

- 状態分布の更新
- 符号空間の確率分布の推定
- エントロピー符号系の生成、及び、符号化/復号化

本アルゴリズムは FSVQ と同様に再帰型符号化系であり初期状態分布  $\pi(1)$ 、次状態関数  $f(y, \pi)$ 、エントロピー符号化写像関数  $\phi_H(y, \pi)$ 、エントロピー復号化写像関数  $\psi_H(u, \pi)$  によって規定される。以下に各ブロックの動作アルゴリズムについて説明する。



(a) 符号化処理



(b) 復号化処理

図 4.1 HMM を利用した情報量削減アルゴリズムのブロック図

#### 4.2.1 状態分布の更新

状態分布の更新は式 (3.7) に基づいて行なう。つまり次状態関数  $f(y(t), \pi(t))$  は

$$f(y(t), \pi(t)) = \pi(t)^T \mathbf{A} \otimes \mathbf{B}(y(t)) \quad (4.1)$$

となる。

#### 4.2.2 符号空間の確率分布の推定

$\pi(t)$  が与えられたときに次の遷移で HMM が  $y(t)$  を出力する確率  $P(y(t)|\pi(t))$  を考える。これは、状態遷移を制約しなければ、全ての状態間の遷移で  $y(t)$  を出力する確率の総和に



より算出することができる．

$$\begin{aligned}
 P(y(t)|\boldsymbol{\pi}(t)) &= \sum_{j=1}^S \sum_{i=1}^S P(x(t) = i) \times a_{i \rightarrow j} \times b_{i \rightarrow j}(y(t)) \\
 &= \boldsymbol{\pi}_{y(t)}(t+1)^T \mathbf{U} \\
 &= \boldsymbol{\pi}(t)^T \mathbf{A} \otimes \mathbf{B}(y(t)) \mathbf{U}
 \end{aligned} \tag{4.2}$$

よって，状態分布  $\boldsymbol{\pi}(t)$  が与えられたとき，符号語空間  $\mathcal{C}$  の出力確率分布  $P(\mathcal{C}|\boldsymbol{\pi}(t))$  は

$$P(\mathcal{C}|\boldsymbol{\pi}(t)) = \begin{bmatrix} \boldsymbol{\pi}(t)^T \mathbf{A} \otimes \mathbf{B}(c_1) \mathbf{U} \\ \boldsymbol{\pi}(t)^T \mathbf{A} \otimes \mathbf{B}(c_2) \mathbf{U} \\ \dots \\ \boldsymbol{\pi}(t)^T \mathbf{A} \otimes \mathbf{B}(c_K) \mathbf{U} \end{bmatrix} \tag{4.3}$$

となる．

図 4.2 に状態分布の遷移例を図 4.3 に状態分布の遷移に伴う符号空間の確率分布の遷移例を示す．

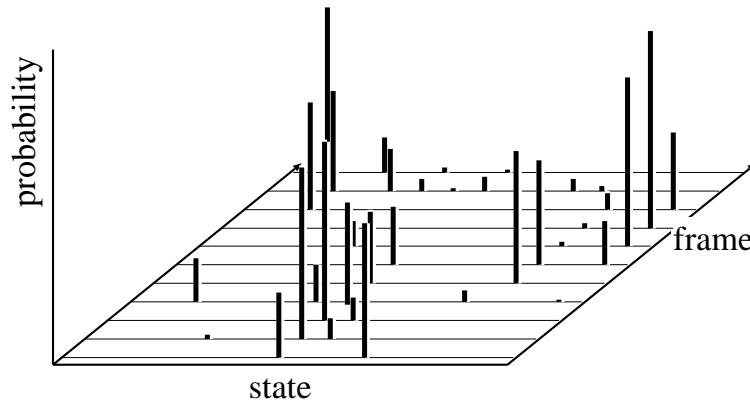


図 4.2 状態分布の遷移

### 4.2.3 エントロピー符号化/復号化

エントロピー符号系は 2.4 に記述したハフマン符号を用いる．

符号化では，まず，出力確率分布  $P(\mathcal{C}|\boldsymbol{\pi})$  に基づき符号語に対するハフマン符号系  $C_h\langle P(\mathcal{C}|\boldsymbol{\pi}) \rangle$  を生成する．生成した  $C_h\langle P(\mathcal{C}|\boldsymbol{\pi}) \rangle$  を用いて符号語インデックス  $y$  に対して通信路符号  $u$  を割り当てる．

$$u = \phi_H(y, \boldsymbol{\pi}) = C_h\langle P(\mathcal{C}|\boldsymbol{\pi}) \rangle(y) \tag{4.4}$$

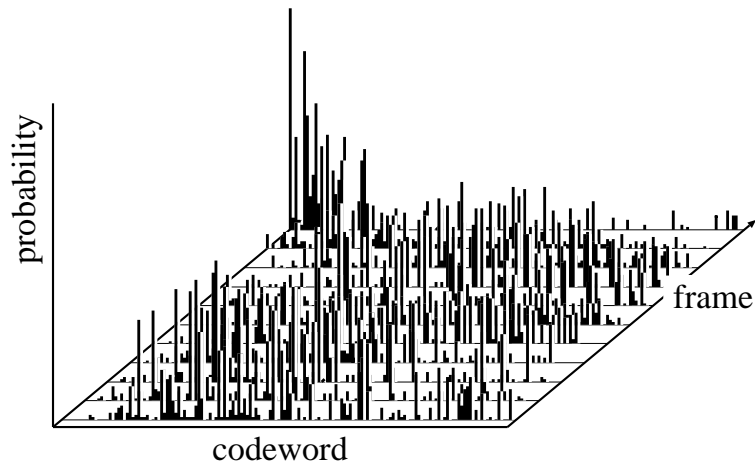


図 4.3 符号空間の確率分布の遷移

復号化も同様に，出力確率分布  $P(C|\hat{\pi})$  に基づき符号語インデックスに対するハフマン符号系  $C_h \langle P(C|\hat{\pi}) \rangle$  を生成し，受信した通信路符号  $\hat{u}$  に対応する符号語インデックス  $\hat{y}$  に変換する．

$$\hat{y} = \psi_H(\hat{u}, \hat{\pi}) = C_h \langle P(C|\hat{\pi}) \rangle^{-1}(\hat{u}) \quad (4.5)$$

ただし， $C_h \langle \cdot \rangle^{-1}$  はハフマン符号系  $C_h \langle \cdot \rangle$  による復号化処理を表す．

### 4.3. 性能評価

本節では HMM を用いた情報量削減アルゴリズムの性能についてシミュレーションした結果を示す．まず，4.3.1 節にシミュレーション条件を示した後，4.3.2 節で HMM 状態数と情報圧縮効率の関係を，4.3.3 節で伝送レートと歪みの関係をシミュレーションにより求める．

#### 4.3.1 シミュレーション条件

まず，評価に用いた符号語インデックス列について説明する．

音声コーデック系への入力音声データは音響学会音声データベースを 8kHz ヘダウンサンプリングしたものをを用いた．量子化/逆量子化ブロックで使用する符号帳と HMM を学習するためのデータとして話者 44 名 (男性 20 名，女性 24 名) 計 6639 発話を用いた．また，学習内として話者 44 名 (男性 20 名，女性 24 名) から，学習外として話者 20 名 (男性 10 名，

女性 10 名) から各 100 発話をランダムに選択し評価用データとした。これらの音声データから図 2.2 の物理モデルに基づいて表 4.1 に示した分析法によりパラメータを抽出し、量子化ブロックでベクトル量子化することで評価用の符号語インデックス列を作成した。

表 4.1 情報量削減手法検討用パラメータ抽出諸元

分析	分析法	LPC ケプストラム分析
	次数	15 次
分析窓	種類	ハミング窓
	フレーム長	20ms
	窓長	32ms

ベクトル量子化に用いる符号帳は LBG アルゴリズム (Splitting & the Generalized Lloyd アルゴリズム [8]) により学習し、HMM はフォワード-バックワードアルゴリズムに基づいて学習した [9]。

比較のために、マルコフモデルに基づく符号化系として、含まれる符号語は同じであるが生起する確率分布が異なる状態符号帳を有する FSVQ を用いて、状態毎に異なるエントロピー符号系を構成する符号化方式のシミュレーションを行なった。FSVQ は状態ラベル型 FSVQ とし、学習データから得られる符号語間の遷移ヒストグラムを用いて次状態関数を設計する条件付きヒストグラム法 [8] で学習した。状態符号帳毎のエントロピー符号系は学習データに対する状態毎の符号語生起確率を用いて作成した。従って、状態数は符号帳サイズと等しく、状態符号帳は、状態毎に符号語の生起確率が異なるのみで、全て同一の符号語で構成されている。更に、状態符号帳は HMM に基づく方式で用いる符号帳と同一のものとした。

#### 4.3.2 HMM 状態数と圧縮効率の関係

図 4.4 に 8 bit 符号帳 (符号語数 256) を使用したときの HMM の状態数と圧縮効率の関係を示す。FSVQ の状態数は符号語数に対応しているため本節の検討では 256 状態である。図中、実線 (CLOSE) は学習内話者音声データの平均符号長を、破線 (OPEN) は学習内話者音声データの平均符号長を示している。

図 4.4 より、HMM の状態数を増やすことで圧縮率は向上し、8 状態 HMM で約 1.5bits 削減できることがわかる。また、256 状態 FSVQ と同程度の圧縮効率を 2 状態 HMM で実現しており、HMM はマルコフモデルよりも少ないハードウェア規模で高精度のモデル化が

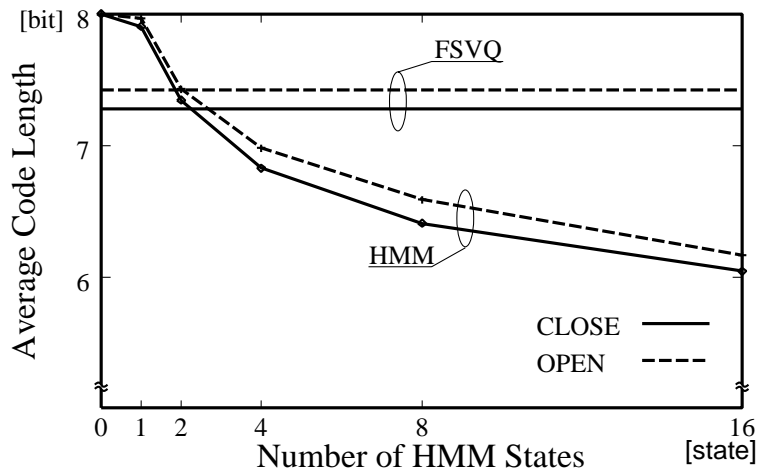


図 4.4 8bit 符号帳に対して情報量削減手法を適用したときの HMM 状態数と平均符号長の関係

可能であることを示している。

1 状態 HMM は学習データにおける符号語インデックスの出現頻度に従ってエントロピー符号化したものであり，静的統計量に基づいて最適な符号系を構築したものと等価である．従って，本章で検討している情報量削減法は，入力音声信号源の状態に追従し，動的に最適な符号化を行なっていると考えられる．更に，状態数を増やすと共に追従性も良くなっている．これらの結果は，HMM とエントロピー符号化の組合せによる情報量削減法の有効性を示している．

### 4.3.3 伝送レートと歪みの関係

HMM の状態数を 8 状態に固定した時，評価用データ学習内外 200 発話に対する平均歪みと伝送レート（符号帳サイズ）の関係を図 4.5 に示す．縦軸は送信側の分析ブロックで抽出されたパラメータと受信側の復号パラメータ間のケプストラム距離，横軸は一つの符号語インデックスを伝送するのに要するビット数を表している．ケプストラム距離  $CD$  は式 (4.6) で定義される客観的なスペクトル距離尺度である [1]．

$$CD = \frac{10}{\log 10} \sqrt{2 \sum_{i=1}^q (C_x(i) - C_y(i))^2} \quad (4.6)$$

ただし， $C_x(i)$ ， $C_y(i)$  は LPC ケプストラムの  $i$  次の係数である．また  $q = 15$  としてケプストラム距離を算出した．図 4.5 の破線 (VQ) は符号帳インデックスを通信路符号として使用

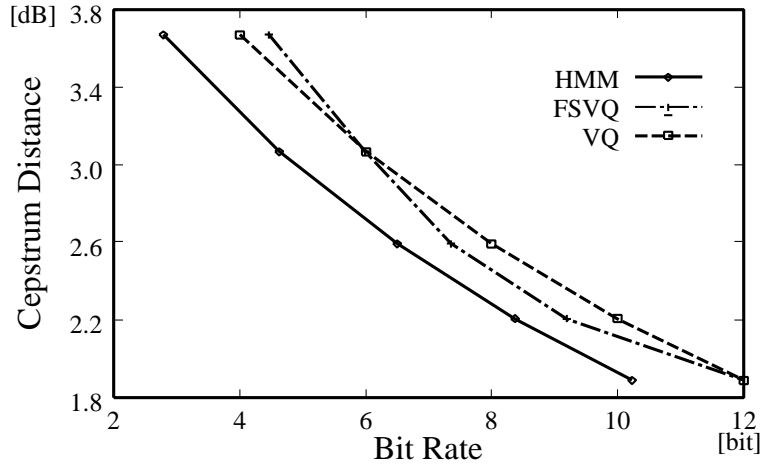


図 4.5 8 状態 HMM を用いて適応的にエントロピー符号系を構築したときの歪みと伝送レートの関係

する従来の固定長符号による性能を示している．図 4.5 より，HMM を用いて適応的に符号系を構築することで  $0.3 \sim 0.4$  [dB] の符号化利得があることが示された．また，FSVQ は状態ラベル型 FSVQ を用いており符号語の増加と共に状態数も増加するので，約 9.2bits 迄は伝送レートと共に性能が向上している．しかしながら，符号帳サイズが大きくなりすぎると FSVQ の性能は劣化することがわかる．本節で遷移ラベル型 FSVQ の設計に用いた条件付きヒストグラム法は全ての状態間の遷移頻度を用いて学習するため，状態数の増加に伴って十分に学習するために必要となるデータ数が指数的に増加する．従って，状態数が増すにつれ入力音声信号の状態によりよく追従するようになるものの，状態数を増やし過ぎるとモデルを十分に学習するためのデータが不足するた情報源モデルの性能が劣化することとなる．このことから，HMM を符号化に適用することは符号化利得だけでなく学習効率の面でもマルコフモデルと比較して優れていることがわかる．

#### 4.3.4 誤り耐性

図 4.6 に，HMM の状態数と誤り耐性の関係を示す．符号帳サイズは 8bits(符号語インデックス数 256)である．また，この結果は，評価用データ学習内 100 発話に対して，ひとつの誤り率毎に 9 種類の誤りパターンを加えた時のケプストラム距離の平均値である．縦軸の距離尺度はケプストラム距離を用い，横軸は誤り率を表している．

図 4.6 より，HMM の状態数が増すとともに誤りに対する耐性が劣化することがわかる．

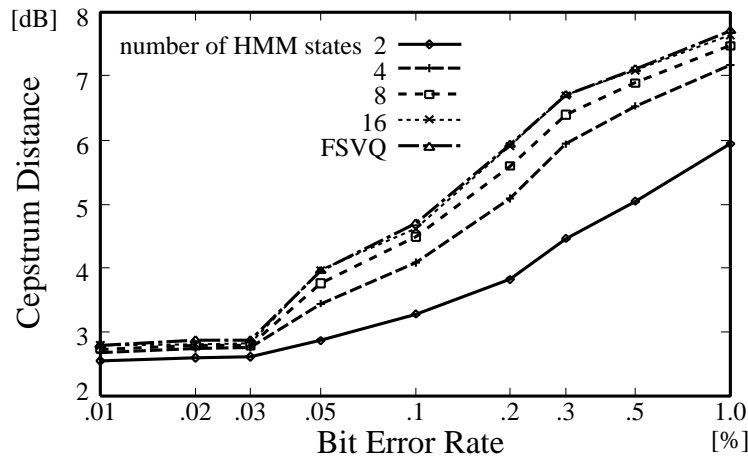


図 4.6 8bit 符号帳に対して情報量削減手法を適用したときの HMM の状態数と誤り耐性の関係

また，16 状態 HMM は状態数 256 が FSVQ と同等の誤り耐性を有している．FSVQ と同程度の圧縮効率を有する 2 状態 HMM は，FSVQ と比較して誤り耐性に優れており，誤り耐性の面でも HMM に基づく符号系はマルコフモデルと比較して優れていることがわかる．

#### 4.4. 結言

本章では，信号源を HMM でモデル化し，局所的な HMM の状態における符号語インデックスの生起確率を用いてエントロピー符号化する手法を検討した．シミュレーション結果より，HMM に基づく符号系の適応は，入力音声信号の状態を模擬した状態分布に基づき，状態分布毎に最適な符号化系を構成することができ，情報量を削減することができる点で優れた性能を有することが示された．例えば，8 状態 HMM を用いて 8bit 符号帳を符号化したとき，平均符号長を約 80%(6.5bits) に圧縮することができる．

一方，HMM の状態数を増すことによって圧縮率を向上させることができるが，誤りに対して弱くなることも示された．

次章では情報圧縮率を維持しつつ本方式の通信路誤りに対する耐性を強化する手法について検討する．

# 第5章 情報量削減における誤り耐性の改善

## 5.1. 序言

4章で検討した方式は通信路誤りの影響を受けやすい手法であった。これは状態が符号化ブロックと復号化ブロックで独立して更新される再帰型符号化法であること、適応的に構築する符号系が冗長性を可能な限り除去するエントロピー符号であることに起因していると考えられる。

本章では、4で検討した情報量削減法の通信路誤りに対する耐性の強化について論ずる。以下、5.2で誤り耐性の向上手法を提案し、5.3で評価する。

なお、本章におけるシミュレーションはHMMの状態数8、符号語数256のシステムを基本構成としている。

## 5.2. 分布の拘束による誤り耐性の改善

### 5.2.1 状態分布の収束性

4章で検討したアルゴリズムは、符号化の状態分布  $\pi(t)$  と復号化の状態分布  $\hat{\pi}(t)$  が通信路誤りによって不整合となった場合、同期が回復することを保証していない。そこで本節では通信路誤りが永久に伝搬し、系が発散あるいは振動することがないことをシミュレーションにより確かめる。

図5.1に、通信路上での1bit誤りにより不整合が生じた後、符号化と復号化の状態分布の同期が回復する様子を示す。図中、縦軸は符復号化それぞれの状態分布間の Kullback-Leibler Divergence(以下、K-L情報量)[57]である。なお、K-L情報量  $KLD$  は次式で定義され、値が小さいほど確率分布間の距離が近いことを意味している。

$$KLD = \sum_{s=1}^S P(x(t) = s) \log \frac{P(x(t) = s)}{P(\hat{x}(t) = s)} \quad (5.1)$$

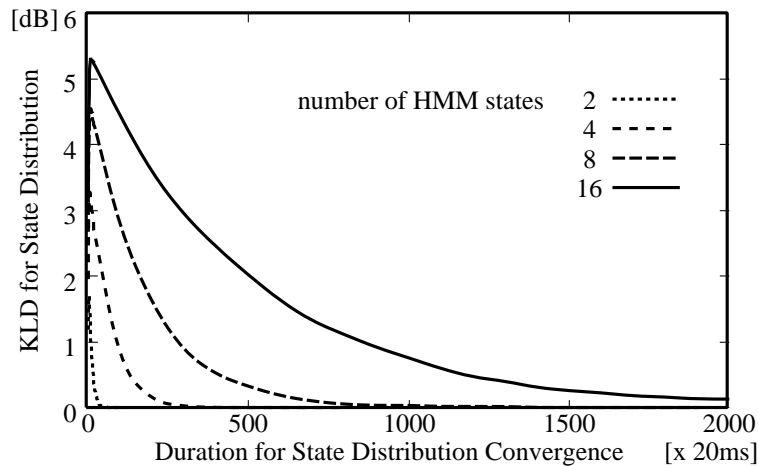


図 5.1 1bit 誤り発生時の符号化-復号化に対する状態分布間距離の収束性

図 5.1 より、通信路誤りの影響は有限時間で収束し系が発散しないことがわかる。また、状態数が少ない程、短時間で状態分布は収束している。このことは 4.3.4 で評価した HMM の状態数と誤り耐性の関係と一致する。

### 5.2.2 状態分布の制限による誤り耐性の向上

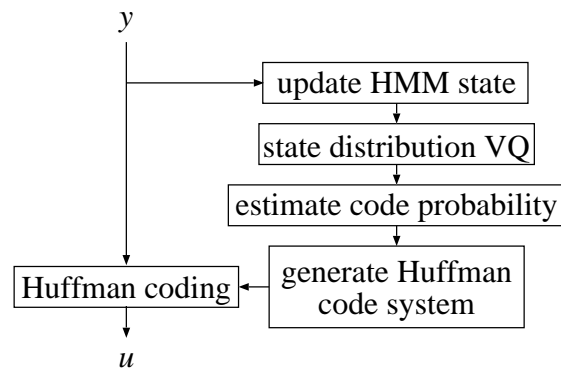
再帰的な状態更新は補助情報を付加することなく符復号化の状態を同期させることができる点で情報量削減に適している。しかし、通信路で誤りが生じたとき送信符号  $u$  と受信符号  $\hat{u}$  が異なるため、これらに基づいて更新する状態分布  $\pi$ ,  $\hat{\pi}$  にも不整合が生じる。そのため誤りが後続する復号化処理にも影響し復号結果の誤りが数フレームに渡り伝搬する。特に、HMM のように確率分布で情報源の状態を表す場合、状態分布を無限に取り得るために同期の回復が困難になる。従って、取り得る状態分布数を制限することで、誤り耐性を向上させることができると考えられる。

そこで、図 5.2 のように、状態分布をベクトル量子化し、取り得る状態分布数を制限する構成を考える。これを状態分布量子化法と呼ぶこととする。また、状態分布をベクトル量子化するための符号帳を状態分布符号帳と呼ぶ。

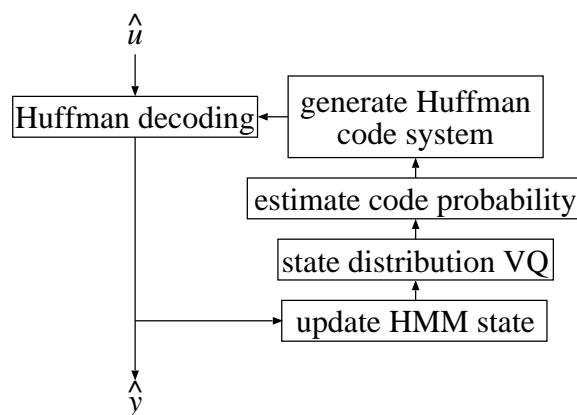
### 5.2.3 符号語空間上の確率分布の制限による誤り耐性の向上

ハフマン符号化は冗長性をほぼ完全に排除する符号系であるため、符号語空間の確率分布の微小な相違にも敏感な符号系である。つまり、少しでも異なる状態分布に対して大き





(a) 符号化处理

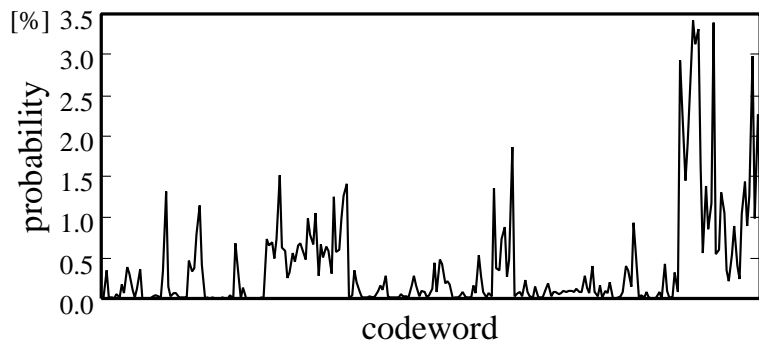


(b) 復号化处理

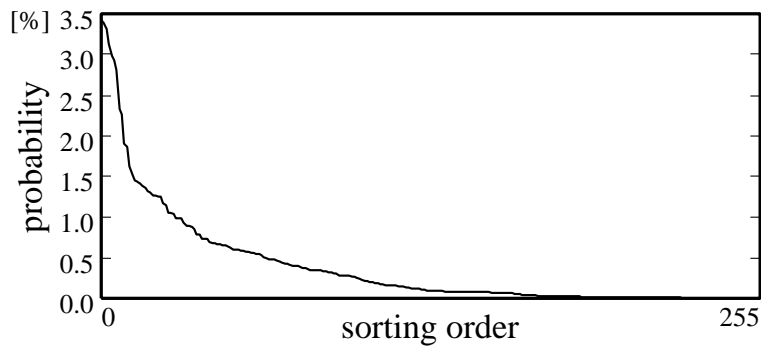
図 5.2 状態分布量子化法のブロック図

く異なる符号系が生成されることとなる。従って、通信路誤りが生じると、符復号化、双方の状態分布の同期が失われ、同期が回復するまで異なる符号系で符復号化される。そのために通信路符号化した時点で符復号化で異なる符号系を用いていることに起因する誤りが重畳し、符復号化の状態分布が収束するまでに長時間が必要となると考えられる。そこで、使用する符号系を制限し、状態分布のわずかな相違による符号系の変動を低減することで通信路誤り耐性が増すことが期待される。

上記目的を達するため、学習データで生じる符号語空間の確率分布をクラスタリングし各クラスタに対応してエントロピー符号系の符号帳(以下、符号系符号帳)を作成した。符復号化には、状態分布で決まる符号語空間の確率分布に対して平均符号長が最短となる工



(a) 符号語空間の確率分布



(b) 確率分布を確率の降順に並べ替え



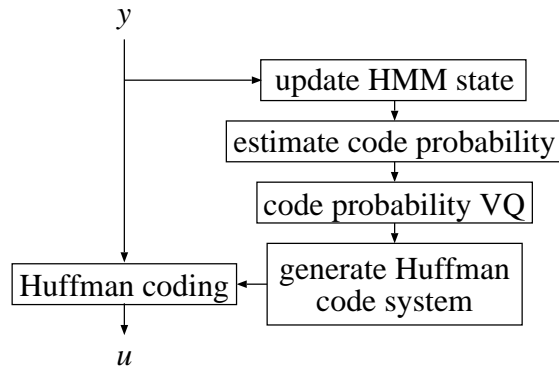
(c) 割り当てられた符号長の分布

図 5.3 符号系符号帳の作成手順

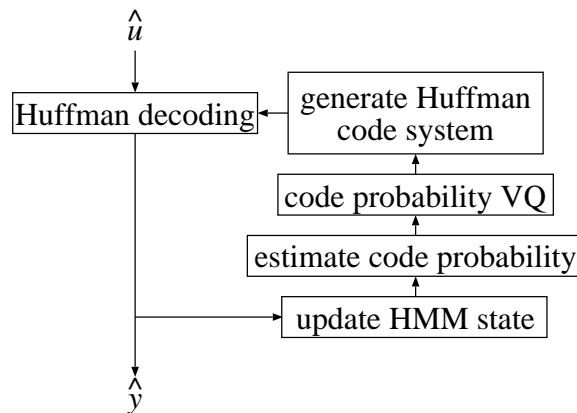
ントロピー符号系を符号系符号帳から選択し用いることとした。ここで、符号系符号帳の作成手順を図 5.3 を用いて説明する。

学習データから図 5.3 (a) に示す符号語空間の確率分布がフレーム毎に得られる．これを図 5.3 (b) のように確率の降順に並べ替え，クラスタリングする．このようにして得られた確率分布のクラス毎にエントロピー符号を求め，符号系符号帳とする．参考として，図 5.3 (b) に対するハフマン符号の符号長分布を図 5.3 (c) に示す．

作成した符号系符号帳を用いてエントロピー符号系を拘束する構成を図 5.4 に示す．これを符号系量子化法と呼ぶこととする．



(a) 符号化处理



(b) 復号化处理

図 5.4 符号系量子化法のブロック図

## 5.3. 性能評価

### 5.3.1 状態分布量子化法

本節では状態分布量子化法による通信路誤り耐性の向上についてシミュレーションにより確認する．図 5.5 に 1bit 誤りが生じたときに符復号化の状態分布が収束する様子を，図 5.6 に状態符号帳サイズと平均符号長の関係を示す．

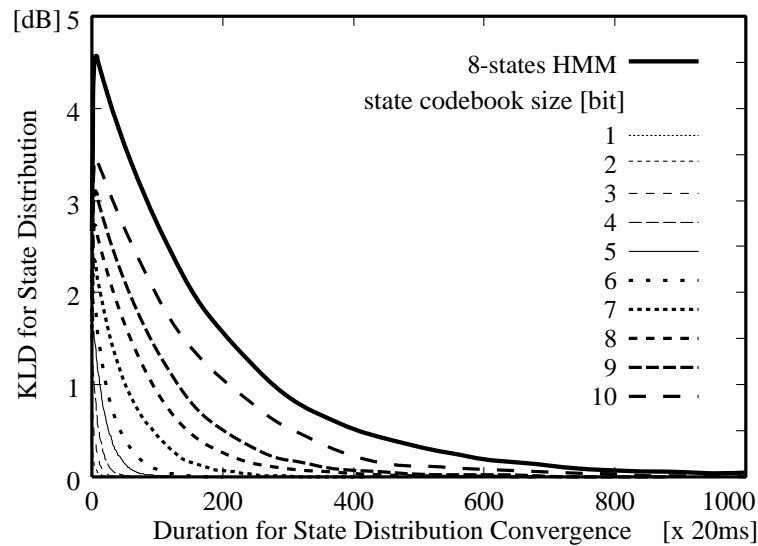


図 5.5 状態分布量子化法を用いたときの 1bit 誤りに対する状態分布間距離の収束性

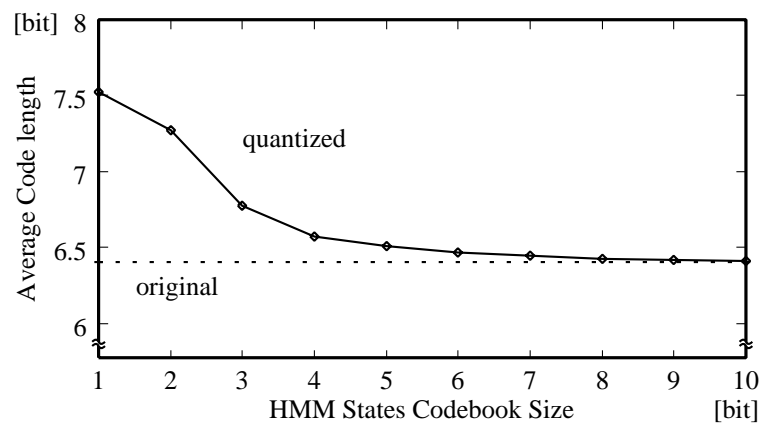


図 5.6 状態分布量子化符号帳サイズと平均符号長に基づく情報量削減効果の関係

図 5.5 より，取り得る状態分布を制限することで伝送路誤りに対する状態の収束性，つ

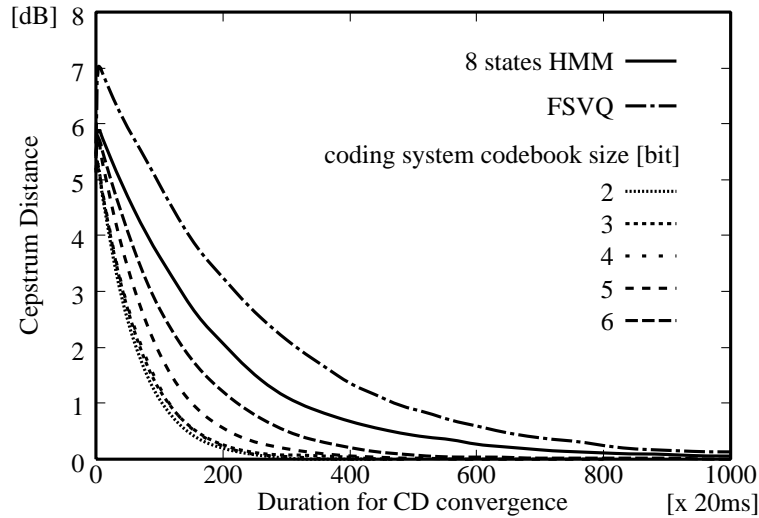


図 5.7 符号系量子化法を用いたときの 1bit 誤りに対するケプストラム距離の収束性

まり，誤り耐性が向上することが示せた．しかし，図 5.6 に示すように，状態符号帳サイズが小さくなるに従って，圧縮効率は劣化する．

図 5.1 の結果と比較すると，状態分布符号帳サイズが 128(7bits)，及び，16(4bits) の時，それぞれ，4 状態 HMM，2 状態 HMM と同程度の収束性を実現している．状態分布を量子化しない場合の平均符号長は，図 4.4 より，4 状態 HMM で約 6.9bits，2 状態 HMM で約 7.4bits である．これに対し，状態分布を量子化した場合，図 5.6 より，状態分布符号帳サイズが 16(4bits) のときの平均符号長は約 6.6bits である．従って，状態分布量子化法を用いることで，誤りに対して同じ収束性で高い圧縮効率を得ることができる．

### 5.3.2 符号系量子化法

本節では符号系量子化法による通信路誤り耐性の向上についてシミュレーションにより確認する．図 5.7 に符号系量子化法を用いたときの 1 bit 誤りに対する応答を示す．図の縦軸は量子化と逆量子化で得られるパラメータ間のケプストラム距離を表す．用いた符号系符号帳サイズは，2，3，4，5，6 bits の 5 種類である．

平均符号長は，オリジナルの 6.41bits に対して，6.44 ~ 6.41bits(2bit ~ 6bit) となり，符号系符号帳のサイズによる圧縮効率の大きな劣化は見られなかった．比較として，FSVQ の収束曲線も示す．

図 5.7 より，符号系を制限することで誤りに対する状態の収束性が向上することが示さ

れた．また，マルコフモデルに基づく FSVQ と比較して誤りの影響を受けにくいことがわかる．これは，HMM では状態への確率分布を用いているため，誤りの伝搬が分散されることに起因すると考えられる．ケプストラム距離が 1dB 未満となるまでの時間を比較すると，FSVQ の約 9.4 s に対して，8 状態 HMM は約 6.4 s，更に符号系量子化法を用いることで約 2.2 s となる．

### 5.3.3 状態分布量子化法と符号系量子化法の比較

本節では状態分布量子化法と符号系量子化法の性能を比較する．図 5.8 に両手法の状態分布の収束曲線を示す．縦軸は状態分布間の K-L 情報量である．参考として，4 状態 HMM の収束曲線も表示している．図中，括弧内に示している値は，それぞれの手法により削減できる bit 数である．

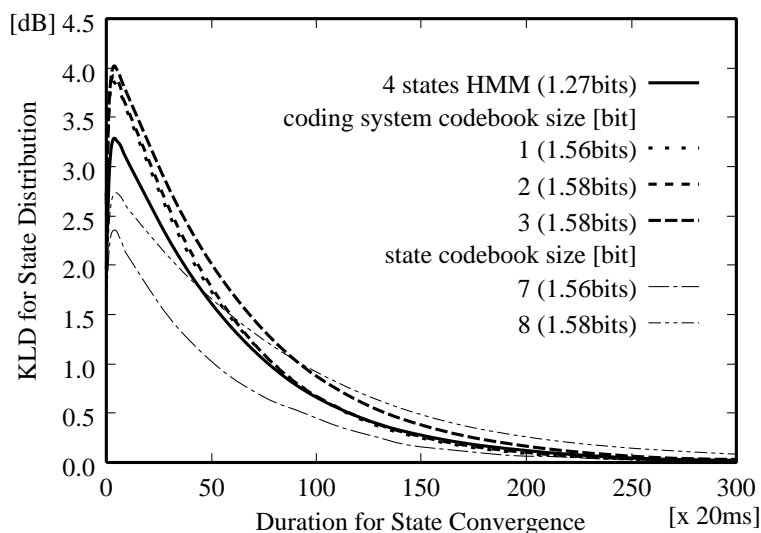


図 5.8 8 状態 HMM を用いた場合の 1 ビット誤りに対する状態分布の収束性の比較

図 5.8 より，符号系量子化法は，状態分布量子化法と比較して，誤りによる影響を大きく受けるが，その後の収束性が良いことがわかる．また，4 状態 HMM がこれらの手法と同じ収束性であることから，多くの状態数でモデル化した HMM を用いて状態分布量子化，または，符号系量子化をしたシステムの方が圧縮効率・誤り耐性の面で高性能な符号化システムが実現できるといえる．

符号系量子化法は，符号系符号帳サイズを小さくしても，ある程度以上，収束性を高めることはできない．これは，符号語の生起確率を並べ替えて量子化していることで選択し

た符号系を用いて符号を割り当てる上で自由度が大きいためである．一方，状態分布量子化法には，そのような自由度がないため，収束性を高めることができる．

次に，図 5.9 に，誤りによる歪みの収束曲線を見る．縦軸は量子化と逆量子化で得られるパラメータ間のケプストラム距離を表している．6bit 符号帳による状態分布量子化法では約 1.53bits，2bit 符号帳による符号系量子化法では約 1.56bits の削減を実現している．図

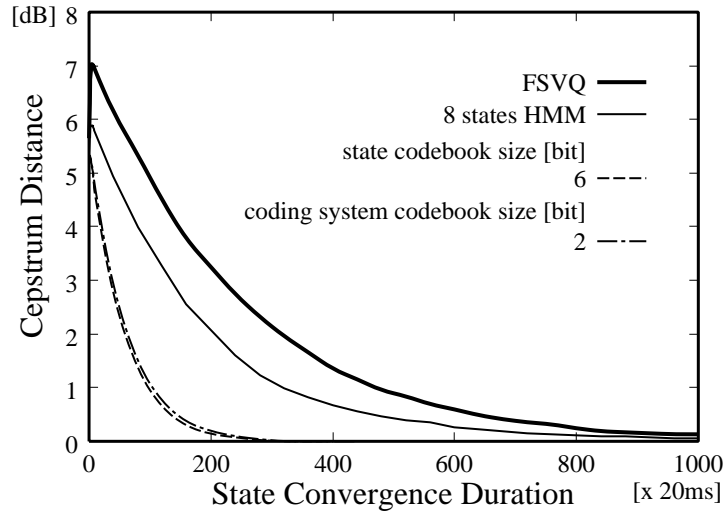


図 5.9 8 状態 HMM を用いた場合の 1bit 誤りに対するケプストラム距離の収束性の比較

5.9 では状態分布量子化法の方が符号系量子化法と比較して，多少，収束が早い．しかしながら，符号系量子化法の方が圧縮率が良いことを考慮すると両手法とも性能上大きな差はないといえる．

## 5.4. 結言

本章では，4 章で検討した情報量削減法の誤り耐性について調べ，状態分布または符号語空間の確率分布をベクトル量子化することで誤り耐性を向上させる手法について検討した．状態分布量子化法では，状態分布符号帳サイズを小さくするに従い，符復号化の状態分布の同期を早く回復することが可能であるが，圧縮効率が劣化した．符号系量子化法では，状態分布量子化法と同様，符号系符号帳サイズによって符復号化の状態分布の同期回復を早めることができたが，状態分布量子化法と比べて誤りによる影響を大きく受けることがわかった．





# 第6章 隠れマルコフモデルを用いた誤り訂正

## 6.1. 序言

本章では HMM を復号化ブロックへ適用することで通信路誤りに対する耐性を向上させることを検討する。

検討する手法は HMM による符号列モデルと通信路モデルから，受信符号列の制約の下，再帰的に最尤推定を行なうことで送信符号列を推定し通信路での誤りを訂正するものである．同様の手法として 2 値符号列を出力する HMM とターボ符号を組み合わせた復号手法が提案されている [51] が，本章では音声コーデックでより一般的な多値符号列を出力する HMM を扱う．

以下，6.2 に提案する誤り訂正アルゴリズムを説明し，6.3 で評価する．

## 6.2. 最尤推定との組合せによる誤り訂正

符号列をモデル化した HMM  $\mathcal{M}$  が与えられたとき，モデル  $\mathcal{M}$  が符号列  $\mathbf{y} = \{y(t)|t = 1, \dots, N\}$  を出力する確率  $P(\mathbf{y}|\mathcal{M})$  を式 (3.8) により算出することができる．一方，符号列  $\hat{\mathbf{u}} = \{\hat{u}(t)|t = 1, \dots, N\}$  を受信したとき，送信された符号列が  $\mathbf{y}$  である確率  $P(\mathbf{y}|\hat{\mathbf{u}})$  は通信路モデルを定義することで算出することができる．本章ではこの 2 つの確率を  $\gamma$  の重み付けでかけ合わせた式 (6.1) を符号列  $\mathbf{y}$  の尤度とし，送信符号列を最尤推定に基づいて推定する手法を検討する．

$$P(\mathbf{y}) = P(\mathbf{y}|\mathcal{M})^\gamma \cdot P(\mathbf{y}|\hat{\mathbf{u}})^{(1-\gamma)}, \quad (6.1)$$

図 6.1 に本章で検討する誤り訂正法のブロック図を示す．

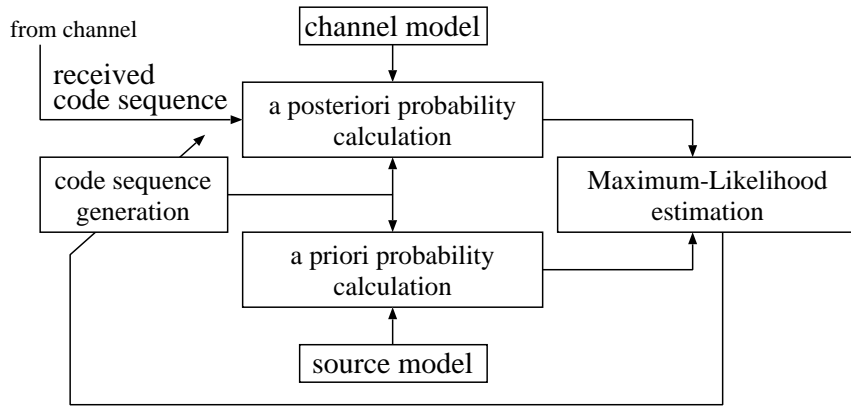


図 6.1 最尤推定に基づく誤り訂正法ブロック図

### 6.2.1 符号列モデルに基づく確率

符号列モデルとして HMM を用いることで， $P(\mathbf{y}|\mathcal{M})$  は 3.3 で定式化したフォワードパス確率分布，及び，バックワードパス確率分布を用いて効率良く算出することができる．

式 (3.12) に示したように，符号列  $\mathbf{y}$  がモデル  $\mathcal{M}$  より出力される確率  $P(\mathbf{y}|\mathcal{M})$  は任意の  $t(1 \leq t \leq N + 1)$  に対してフォワードパス確率分布  $\alpha(t)$  とバックワードパス確率分布  $\beta(t)$  を用いて次式で表わされる．

$$\begin{aligned} P(\mathbf{y}|\mathcal{M}) &= \alpha(t+1)^T \beta(t+1) \\ &= \alpha(t)^T \mathbf{A} \otimes \mathbf{B}(y(t)) \beta(t+1) \end{aligned} \quad (6.2)$$

式 (6.2) における符号列  $\mathbf{y}$  のうち符号  $y(t)$  のみを他の符号  $\tilde{y}(t)$  へ置換した符号列を  $\tilde{\mathbf{y}}$  とする．このとき，モデルが符号列  $\tilde{\mathbf{y}}$  を出力する確率  $P(\tilde{\mathbf{y}}|\mathcal{M})$  は，符号列  $\mathbf{y}$  に対して  $\{\alpha(t), \beta(t) | t = 1, \dots, N + 1\}$  が予め算出されていれば，式 (6.2) の  $\mathbf{B}(y(t))$  を  $\mathbf{B}(\tilde{y}(t))$  に置換することで容易に算出できることがわかる．

ところで， $N$  が大きくなると式 (3.7)，(3.10) の演算で桁落ちを防ぐために必要とされる精度が増大する．そこで，次式で示す正規化した  $\bar{\alpha}(t)$ ， $\bar{\beta}(t)$  を  $P(\mathbf{y}|\mathcal{M})$  の算出に用いる．

$$\begin{aligned} \bar{\alpha}(t) &= \frac{1}{S_{\alpha}^{(t)}} \alpha(t) \\ \bar{\beta}(t) &= \frac{1}{S_{\beta}^{(t)}} \beta(t) \end{aligned} \quad (6.3)$$

ただし，

$$\begin{aligned} S_{\alpha}^{(t)} &= U^T \alpha(t) \\ S_{\beta}^{(t)} &= U^T \beta(t) \end{aligned}$$

このとき式 (6.2) は次式となる .

$$\begin{aligned}
 P(\mathbf{y}|\mathcal{M}) &= \prod_{i=1}^{t+1} S_{\alpha}^{(i)} \bar{\alpha}(t+1)^T \bar{\beta}(t+1) \prod_{j=t+1}^{N+1} S_{\beta}^{(j)} \\
 &= \prod_{i=1}^t S_{\alpha}^{(i)} \prod_{j=t+1}^{N+1} S_{\beta}^{(j)} \bar{\alpha}(t)^T \mathbf{A} \otimes \mathbf{B}(y(t)) \bar{\beta}(t+1)
 \end{aligned} \tag{6.4}$$

### 6.2.2 通信路モデルに基づく確率

符号列  $\hat{u}$  を受信したときに送信符号列が  $\mathbf{y}$  である確率  $P(\mathbf{y}|\hat{u})$  は変復調方式や通信路のモデルを定義することで算出することができる . 通信路モデルとして , 図 6.2 に示す記憶のない 2 元対称通信路を仮定すると  $P(\mathbf{y}|\hat{u})$  は次式となる .

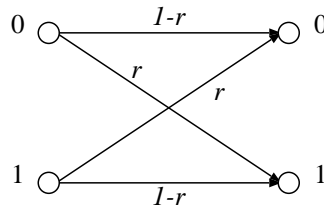


図 6.2 2 元対象通信路

$$P(\mathbf{y}|\hat{u}) = \prod_{t=1}^N P(y(t)|\hat{u}(t)) = \prod_{t=1}^N r^{D_h^{(t)}} (1-r)^{(b-D_h^{(t)})}, \tag{6.5}$$

ただし ,  $D_h^{(t)}$  は  $y(t)$  と  $\hat{u}(t)$  のハミング距離 ,  $b$  は 1 符号を表すビット数である . また ,  $r$  は通信路でビットが誤る確率 (BER:Bit Error Rate) である .

$D_h^{(t)}$  は  $t$  について独立なので ,  $P(\tilde{\mathbf{y}}|\mathcal{M})$  と同様に , 任意の  $y(t)$  の置換に対して  $P(\tilde{\mathbf{y}}|\hat{u})$  は容易に算出することができる .

### 6.2.3 最尤推定を利用した確率の最大化

式 (6.1) の尤度を最大にする  $\mathbf{y}$  を再帰的アルゴリズムにより求める . ここで ,  $\mathbf{y}^{(k)}$  を  $k$  回目の繰り返しにより推定される符号列とし , 符号列  $\mathbf{y}^{(k-1)}$  と  $\mathbf{y}^{(k)}$  はあるひとつの符号  $y^{(k)}(t)$  のみが異なることとする . 従って , 1 回の繰り返しでは元の符号列に対して 1 符号のみを置換したときに最大の尤度を与える符号列を求めることとなる . 更に , 本検討では式 (6.1)

の尤度を直接求めるのではなく， $\frac{P(\mathbf{y}^{(k)})}{P(\mathbf{y}^{(k-1)})}$  を求め， $\frac{P(\mathbf{y}^{(k)})}{P(\mathbf{y}^{(k-1)})} \leq 1$  をアルゴリズムの終了条件とした．これは  $k-1$  回目の置換により尤度が極大値を取ることを意味しており， $\mathbf{y}^{(k-1)}$  を本アルゴリズムにより誤り訂正された符号列として出力することとなる．このように，本検討では非現実的な処理量を要する全探索ではなく，再帰的アルゴリズムにもとづいて符号列を推定するため極大値の検出をアルゴリズムの終了条件とした． $\frac{P(\mathbf{y}^{(k)})}{P(\mathbf{y}^{(k-1)})}$  は式 (6.6) により任意の  $\mathbf{y}^{(k)}$  に対して算出することができる．

$$\begin{aligned} \frac{P(\mathbf{y}^{(k)})}{P(\mathbf{y}^{(k-1)})} &= \frac{P(\mathbf{y}^{(k)}|\mathcal{M})^\gamma P(\mathbf{y}^{(k)}|\hat{\mathbf{u}})^{1-\gamma}}{P(\mathbf{y}^{(k-1)}|\mathcal{M})^\gamma P(\mathbf{y}^{(k-1)}|\hat{\mathbf{u}})^{1-\gamma}} \\ &= \left( \frac{\bar{\boldsymbol{\alpha}}(t)^T \mathbf{A} \otimes \mathbf{B}(\tilde{\mathbf{y}}^{(k)}(t)) \bar{\boldsymbol{\beta}}(t+1)}{S_{\bar{\boldsymbol{\alpha}}}^{(t+1)} \bar{\boldsymbol{\alpha}}(t+1)^T \bar{\boldsymbol{\beta}}(t+1)} \right)^\gamma \cdot \left( \frac{P(\tilde{\mathbf{y}}^{(k)}(t)|u(t))}{P(\bar{\mathbf{y}}^{(k-1)}(t)|u(t))} \right)^{(1-\gamma)} \end{aligned} \quad (6.6)$$

式 (6.6) において， $\{\bar{\boldsymbol{\alpha}}^{(k)}(t)|t = 1, \dots, N\}$ ， $\{\bar{\boldsymbol{\beta}}^{(k)}(t)|t = 1, \dots, N\}$  は  $\bar{\mathbf{y}}^{(k)}$  に対するフォワードパス確率分布，及び，バックワードパス確率分布である．また， $\bar{\mathbf{y}}^{(0)} = \hat{\mathbf{u}}$  を初期条件とする．

本アルゴリズムでは  $\bar{\mathbf{y}}^{(0)}$  が最大の尤度を与えるとき通信路誤りを検出しなかったといえる．

## 6.3. 性能評価

### 6.3.1 シミュレーション条件

まず，評価に用いた符号語インデックス列について説明する．音声コーデック系の処理対象とする音声データは，4.3 と同じく，音響学会音声データベースを 8kHz ヘダウンサンプリングしたものを用いた．量子化，逆量子化ブロックで使用する符号帳や HMM を学習するためのデータとして話者 12 名 (男性 6 名，女性 6 名) 計 1200 発話，評価用として学習内 100 発話 (男性 6 名，女性 6 名)，学習外 100 発話 (男性 14 名，女性 18 名) を用いた．これらの音声データから図 2.2 の物理モデルに基づいて表 6.1 に示した分析法によりパラメータを抽出した．

抽出したパラメータは線スペクトル対パラメータ (LSP) へ変換 [58] し量子化ブロックで 8 bit 符号帳でベクトル量子化し，更に，符号化ブロックで符号語インデックスを通信路符号として割り当てることで評価用の符号列を作成した．得られた符号列に対してビット誤り率 0.1 ~ 3% の誤りをランダムに付加した符号列  $\hat{\mathbf{u}}$  を入力として誤り訂正を行なった．通信路は 2 元対称通信路を仮定し付加した誤り率  $r$  を既知とし式 (6.5) より  $P(\mathbf{y}|\hat{\mathbf{u}})$  を算出した．

表 6.1 誤り耐性向上における評価パラメータ抽出諸元

分析	分析法	線形予測分析
	次数	10 次
分析窓	種類	ハミング窓
	フレーム長	20ms
	窓長	25ms

HMM を用いた誤り訂正法について，

- (1) 1 音声データに相当する全符号の列に対して誤り訂正した場合
- (2) 符号列の一部に対して誤り訂正し，決定した符号を逐次出力した場合

の 2 つの条件の下で性能を評価した．条件 2 はリアルタイム処理を想定したものであり，決定した符号でフォワードパス確率分布の初期状態分布  $\alpha(1)$  を更新することとした．また，各符号に対して通信路誤りが予め検出されている条件下で HMM による誤り訂正の性能についても調べた．この場合，アルゴリズムの終了条件に  $k > 0$  を加え，更に，誤りが検出されていない符号は置換せず，検出されている符号に対してのみ符号を置換し誤り訂正することとした．

評価尺度として，誤りなしの符号列より得られる符号語からのケプストラム距離<sup>1</sup>を用い，平均歪みと 2 ~ 4[dB] 及び 4[dB] を越える歪みが全符号数に占める割合 (Outlier) を測定した．

### 6.3.2 符号列/通信路モデルに対する重み付け

本節では 6.3.4，6.3.5 の評価で用いる式 (6.1) の重み付け  $\gamma$  の値を決定する．図 6.3 に符号列モデルとして 8 状態 HMM を用いたときの  $\gamma$  の値と平均歪みの関係を示す．通信路モデルを 2 元対称通信路としたので  $\gamma = 0.0$  が誤り訂正を行わず受信符号を用いて復号した時の通信路誤りによる歪みを示している．また， $\gamma = 1.0$  は符号列モデルのみを用いて誤り訂正した結果であり，過度に符号列モデルに依存することで歪みが大きくなってしまふことを示している．これは誤っていない符号を間違えて‘誤り’と検出したために生じたもので，統計モデルに基づく誤り検出を行なう上で符号列を完全にモデル化できないことに起因する．図 6.3 より  $P(y|M)$  に対する重み付け  $\gamma$  として 0.5 を用いることとする．

<sup>1</sup>聴感上，1dB 以上の歪みを検知できるとされる

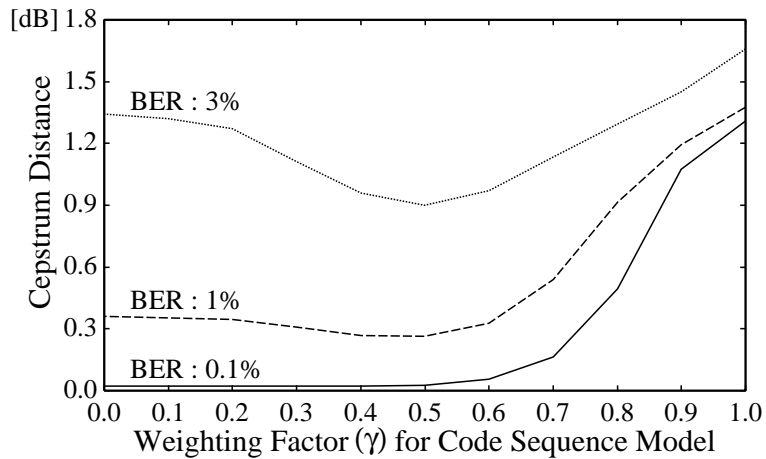


図 6.3 8 状態 HMM を用いて誤り訂正したときの符号列モデルに対する重み付け ( $\gamma$ ) とケプストラム距離の関係

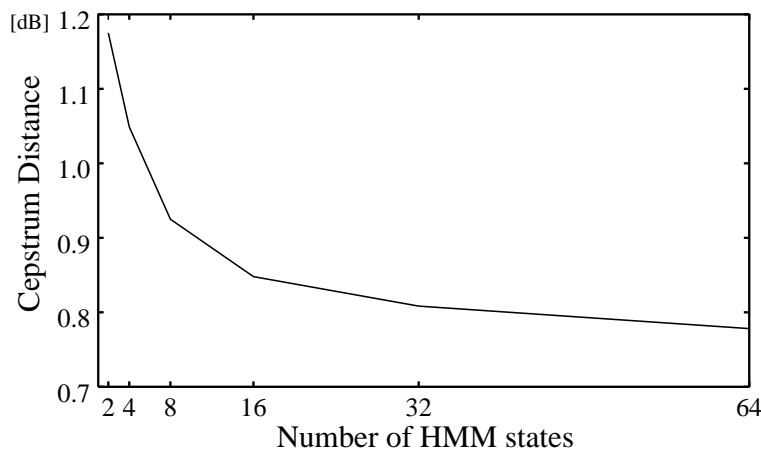


図 6.4 2 ~ 64 状態 HMM を用いて誤り訂正したときのケプストラム距離 (BER=3%)

### 6.3.3 HMM 状態数と誤り訂正能力の関係

本節では，HMM の状態数と誤り訂正能力の関係を調べる．図 6.4 に 3% 誤りに対して HMM 状態数を 2 ~ 64 と変化させたときの平均歪みを示す．図より，状態数を増加させるに従い歪みが減少し誤り訂正能力が向上していることがわかる．しかしながら，状態数を増加させるとともに，状態数に対する歪み低減効率は劣化している．

ところで，本論文ではエルゴード的 HMM を用いているため全状態間に対して遷移確率と出力確率を蓄積する必要がある．つまり，状態数の 2 乗に比例して要するメモリ量が増大する．また，尤度算出においても全状態間の遷移確率を利用して符号の出力確率を算出

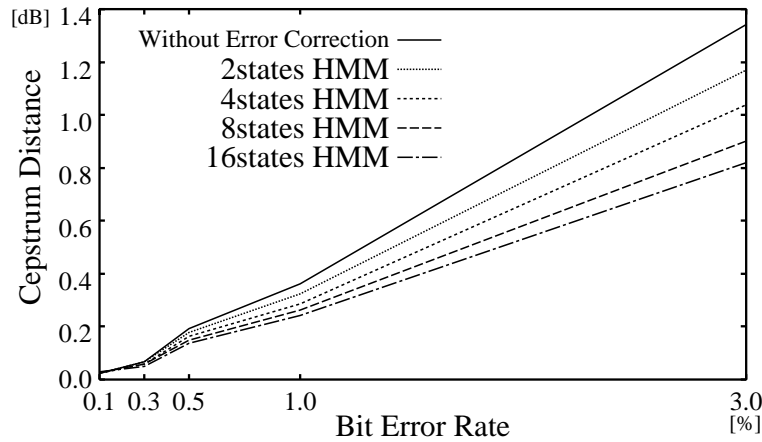


図 6.5 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときの通信路上のビット誤り率に対するケプストラム距離

する必要があるため，状態数が増加するにつれ処理量も増大する．

以上より，実現性を考慮して以下の実験では 2 ~ 16 状態 HMM を用いて誤り訂正法を評価する．

### 6.3.4 誤り訂正能力の評価

本節では，6.2 に記した誤り訂正法の訂正能力を評価する．まず，6.3.1 の条件 (1) について HMM の状態数を 2, 4, 8, 16 としたときの誤り訂正能力を図 6.5 に示す．図より，HMM の状態数を増すにしたがって誤り訂正能力が向上することがわかる．3% 誤りでは，状態数に対して，それぞれ，13%，23%，33%，39% 程度の歪み低減効果が確認された．次に，3% 誤りに対する Outlier を表 6.2 に示す．表 6.2 では，2 ~ 4[dB] の割合はほとんど変化していないが 4[dB] を越える符号の割合が大きく減っている．従って，性能が大きく劣化する符号誤りに対して HMM を利用した誤り訂正法が有効であることがわかる．

次に，6.3.1 の条件 (2) に対応して，実システムでの利用を考慮して演算量及びアルゴリズム遅延に影響するバッファ長と誤り訂正能力の関係について評価した．誤り率 3% におけるアルゴリズム遅延フレーム数と性能の関係を図 6.6 に示す．図 6.6 は音声データの全フレーム数に対する遅延フレーム数の割合を横軸としている．テスト用 200 音声の平均フレーム数は 227.38 フレーム (約 4.5sec) であった．図 6.6 より誤り訂正能力は，HMM の状態数に関係なく 2 ~ 3 フレーム (約 1%) の遅延で大きく向上した後，遅延フレーム数が増えるに従って全符号を用いた誤り訂正性能に漸近していくことがわかる．このことから，4

表 6.2 2, 4, 8, 16 状態 HMM を用いて誤り訂正した時のケプストラム距離の分布 (BER=3%)

Type	CD [dB]	Outlier [%]	
		2 - 4 [dB]	> 4 [dB]
訂正なし	1.34	5.78	15.81
2states	1.17	5.77	14.01
4states	1.04	5.73	12.49
8states	0.90	5.78	10.91
16states	0.82	5.79	9.90

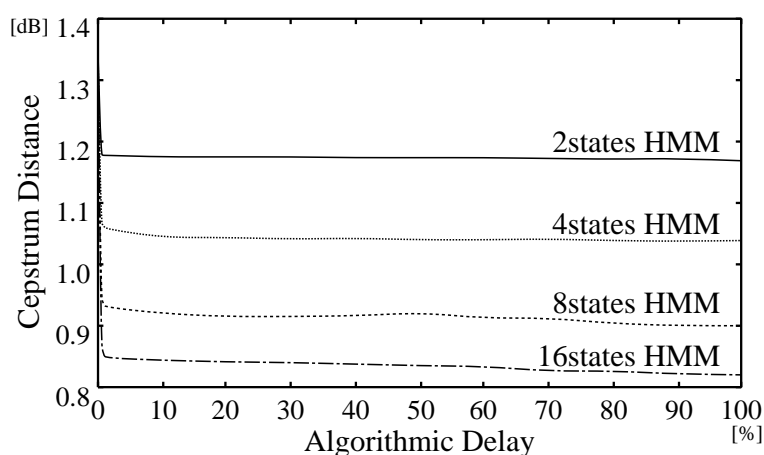


図 6.6 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときのアルゴリズム遅延とケプストラム距離の関係 (BER=3%)

フレーム (現フレーム+3 遅延フレーム) のバッファ長により十分な誤り訂正が実現できることがわかる。

### 6.3.5 誤り検出との組合せ

本節では、誤りがあらかじめ検出されているとしたときの誤り訂正能力について調べる。まず、6.3.1 の条件 (1) について実験した。図 6.7 は HMM の状態数と性能の関係を示したものである。図 6.5 と同様、HMM の状態数を増すにしたがって誤り訂正能力が向上している。3% 誤りにおける歪み低減効果は状態数に対し、それぞれ、19%、31%、42%、49% 程



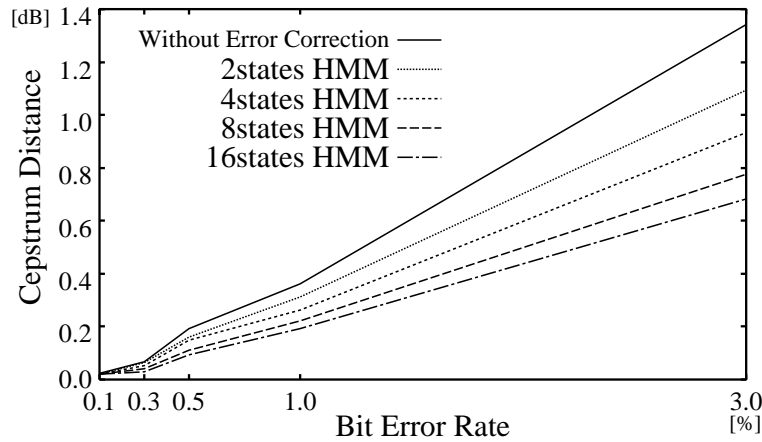


図 6.7 予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときの通信路上のビット誤り率に対するケプストラム距離

度あった。このことから、誤りが確実に検出されたことで 6.3.4 と比較して約 10% 性能が向上していることがわかる。表 6.3 に 3% 誤り時の Outlier を示す。表 6.3 より、6.3.4 と同

表 6.3 予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときのケプストラム距離の分布 (BER=3%)

Type	CD [dB]	Outlier [%]	
		2 - 4 [dB]	> 4 [dB]
訂正なし	1.34	5.78	15.81
2states	1.09	5.77	13.14
4states	0.93	5.69	11.09
8states	0.78	5.69	9.21
16states	0.68	5.57	8.03

様に大きく性能が劣化する符号誤りに対して HMM を用いた誤り訂正法が有効であることがわかる。また、2 ~ 4[dB] Outlier も減少していることから、6.3.4 で訂正することができなかった劣化の少ない符号誤りも訂正していることがわかる。

次に、6.3.1 の条件 (2) に対応して図 6.8 に誤り率 3% での遅延量と性能の関係を示す。6.3.4 の結果を踏まえ、3 フレーム遅延までを評価した。本実験では予め誤りが検出されているため、16 状態の場合、遅延なしでも全符号を用いて訂正した場合に対して 70% 程度の性能を実現しており、また 1 フレーム遅延では 97% もの性能を実現している。このことが

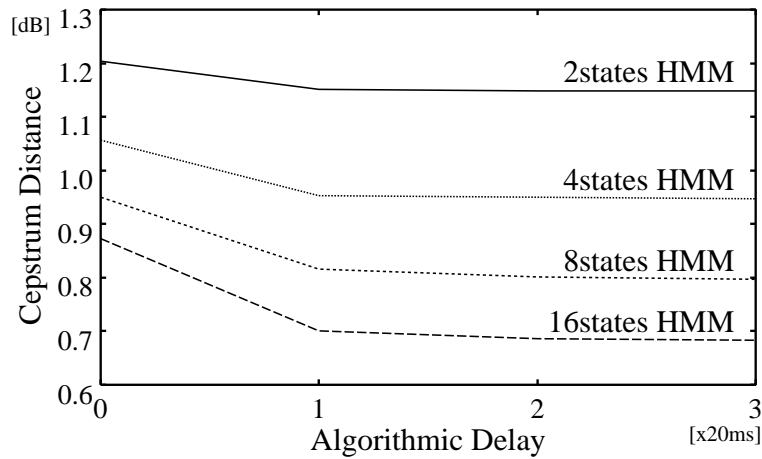


図 6.8 予め誤りが検出されている符号列を 2, 4, 8, 16 状態 HMM を用いて誤り訂正したときのアルゴリズム遅延とケプストラム距離の関係 (BER=3%)

ら，誤り訂正のみを行なう場合バッファ長は 2 フレーム (現フレーム+1 遅延フレーム) あれば十分であるといえる．

### 6.3.6 処理量

本節では 1 フレーム遅延で 8bit 符号を訂正することを仮定して，HMM を用いた誤り訂正法の演算量を見積もった．以下に 1 符号置換の処理量を示す．対象としている符号数は 2 つ (現フレーム+1 遅延フレーム) なので，複数の符号の誤り訂正では，この値を 2 倍した処理量で十分に収束すると考えられる．

まず，16 状態 HMM を用いて全てのビット誤りの組み合わせに対して最尤推定した場合，1 フレームあたり約 457k ステップが必要であった．つまり，分析周期 20ms では約 23MOPS となる．この値は 1bit 誤りのみを対象に訂正を行なうことで約 5MOPS となる．次に，8 状態 HMM を用いた場合，1 フレームあたり約 181k ステップ (約 9MOPS) が必要であるが，1bit 誤りのみを対象とすると約 1.7MOPS で実現できることがわかった．

## 6.4. 結言

本章では HMM を利用して誤り訂正を行なう手法を検討した．本アルゴリズムは符号列モデルと通信路モデルから算出される生起確率を最大化する符号列を送信された符号列として推定するものである．

LSP パラメータを符号化した通信路符号列に対して評価した結果, 3% ランダム誤りに対して, 16 状態 HMM を用いることで約 39% ケプストラム距離を低減できることを示した. 特に, 劣化の大きな誤りに対して本手法が有効であることを確認した. 更に, あらかじめ誤りが検出されている場合, 約 49% のケプストラム距離低減効果があることを示した.



## 第7章 結論

本論文では HMM を音声コーデックへ適用することで情報量削減，あるいは，通信路誤りに対する耐性の強化を図る手法を提案し，その効果を評価した．

情報量の削減については，4章において HMM で推定される符号語空間の生起確率分布に基づいてエントロピー符号系を動的に生成することで情報圧縮が可能であることを示した．本手法は冗長性を大幅に削除するため誤りの影響を受け易くなる．そこで，情報圧縮性能をできるだけ維持したまま通信路誤りにも強くする手法について5章で検討した．

通信路誤り耐性の向上については，6章において HMM から推定される符号語空間の生起確率分布を利用して復号したパラメータ列の正しさの尺度として誤り訂正する手法を検討した．受信した通信路符号による拘束の下，本手法を適用することで復号パラメータの歪みが低減されることを示した．

これらの手法は，情報源符号化，及び，通信路符号化アルゴリズムと独立しており，既存の音声コーデック方式と通信路符号化方式間のインターフェースとして利用することで音声コーデックの効率を向上させることができる．本論文では基本的な音声コーデックの量子化系に対して提案法を適用することで，その効果を確かめた．提案法を実システムへ適用するにあたり次の2点に関して効率の良いモデルを検討する必要がある．

- モデルの精度
- ハードウェア規模

音声コーデックは方式によって量子化ブロックが異なる．つまり出力される符号語列の特性が異なるため，構造が同じモデルを使用しても，モデルの精度は音声コーデックによって異なる．モデルの精度は提案法の性能に影響するため，高精度のモデル化は欠かすことができない．従って，本論文で提案した手法を適用するにあたって，対象とする音声コーデックが出力する符号語列の高い精度でのモデル化を検討する必要がある．本論文ではモデル化の対象として線形予測パラメータのみを用いたが，複数のパラメータを出力するモデルを考えることで精度が向上することが考えられ，そのような観点からもモデルの精度を向上させる検討をする必要がある．

一方、モデルの精度向上に伴い必要となるメモリー量、処理量等のハードウェア規模が増大することが考えられる。音声コーデックが搭載されることが多いモバイルコミュニケーションメディアでは小型、かつ、実時間処理が要求されるためハードウェア規模は大きな問題となる。そこで、精度を劣化させることのない簡素なモデル構造についても検討する必要がある。

これらの検討、及び、最適化を通して提案法を実システムへ適用することが可能となる。ハードウェア規模の制約を考慮しない場合、HMMを利用することで音声コーデックは更に高能率化を図ることができる。情報源符号化に対しては、現在は音声コーデックの設計時に構築している量子化系を情報源モデルの状態に応じて逐次設計することで情報量圧縮効率を向上させることができる。この手法は提案法における符号系の逐次更新と異なり、量子化系を最適化するため固定長符号系で実現できる点で優れる。通信路符号化に対しては、[51]のように通信路復号化処理において情報源モデルから推定される尤度を考慮することで誤り訂正能力を向上させることができる。これは情報源-通信路統合符号化のひとつの実現法であり、従来検討されてきた通信路符号化の制約を情報源符号化へ適用する手法と対をなすものである。この手法は提案法と同じく復号化側のみの変更で性能向上を図ることができる。

以上のことから、今後、DSP等の処理能力の向上に伴いHMMのような動的情報源モデルを利用した情報源符号化法、及び、通信路符号化法の研究は更に発展するものと期待される。

## 参考文献

- [1] 守谷 健弘. 音声符号化. 電子情報通信学会, 1998.
- [2] 小澤 一範. 移動体通信における音声符号化の現状と今後の課題. 音講論集, 271–272, September 1995.
- [3] R. V. Cox. Speech coding standards. In W.B. Kleijn and K. K. Paliwal, editors, *Speech Coding And Synthesis*, chapter 2, 49–78. Elsevier, 1995.
- [4] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon. *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, 1991.
- [5] M. R. Schröder and B. S. Atal. Code-excited linear prediction (CELP): High-quality speech at very low bit rates. *Proc. ICASSP*, 937–940, 1985.
- [6] 大室 伸, 守谷 健弘, 間野 一則, 三樹 聡. 移動平均型フレーム間予測を用いる LSP パラメータのベクトル量子化. 信学論, J77-A(3):303–313, March 1994.
- [7] ITU-T. *ITU-T Recommendation G.729 – Coding of speech at 8kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, March 1996.
- [8] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
- [9] J. R. Deller, Jr., J. G. Proakis, and J. H.L. Hansen. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, 1993.
- [10] J. Foster, R. M. Gray, and M. O. Dunham. Finite-state vector quantization for waveform coding. *IEEE Trans. on Inform. Theory*, IT-31:349–359, May 1985.
- [11] A. Haoui and D. G. Messerschmitt. Predictive vector quantization. *Proc. ICASSP*, 1:10.10.1–10.10.4, 1984.

- [12] M. O. Dunham and R. M. Gray. An algorithm for the design of labeled-transition finite-state vector quantizers. *IEEE Trans. on Commun.*, COM-33(1):83–89, January 1985.
- [13] Y.-M. Cheng and D. O’Shaughnessy. On 450-600 b/s natural sounding speech coding. *IEEE Trans. on Speech and Audio Process.*, 1(2):207–219, April 1993.
- [14] J. G. Dunham and R. M. Gray. Joint source and noisy channel trellis encoding. *IEEE Trans. on Inform. Theory*, IT-27(4):516–519, Jul. 1981.
- [15] E. Ayanoglu and R. M. Gray. The design of joint source and channel trellis waveform coders. *IEEE Trans. on Inform. Theory*, 33(6):855–865, November 1987.
- [16] D. Rowe and P. Secker. A robust 2400 bit/s MBE-LPC speech coder incorporating joint source and channel coding. *Proc. ICASSP*, II141–II144, 1992.
- [17] C. R. Nassar and M. R. Soleymani. Codebook design for trellis quantization using simulated annealing. *IEEE Trans. on Speech and Audio Process.*, 1(4):400–404, October 1993.
- [18] J. Hagenauer. Source-controlled channel decoding. *IEEE Trans. on Commun.*, 43(9):2449–2457, September 1995.
- [19] M. W. Marcellin, T. R. Fischer, and J. D. Gibson. Predictive trellis coded quantization of speech. *IEEE Trans. on Acoust. Speech, and Signal Process.*, 38(1):46–55, January 1990.
- [20] P. Sriram and M. W. Marcellin. Performance of adaptive prediction algorithms for trellis coded quantization of speech. *IEEE Trans. on Commun.*, 42(2/3/4):1512–1517, February/March/April 1994.
- [21] S. Kang and T. R. Fischer. Trellis excitation speech coding at low bit rates. *IEEE Trans. on Commun.*, 42(2/3/4):1902–1910, February/March/April 1994.
- [22] A. Popescu, N. Moreau, and C. Lamblin. CELP coding using trellis-coded vector quantization of the excitation. *IEEE Trans. on Speech and Audio Process.*, 3(6):464–472, November 1995.



- [23] K. T. Malone and T. R. Fischer. Trellis-searched adaptive predictive coding of speech. *IEEE Trans. on Speech and Audio Process.*, 1(2):196–206, April 1993.
- [24] ITU-T. *ITU-T Recommendation G.728 Annex I – Frame or packet loss concealment for the LD-CELP decoder*, May 1999.
- [25] E. P. Farges and M. A. Clements. Hidden Markov models applied to very low bit rate speech coding. *Proc. ICASSP*, 433–436, 1986.
- [26] J. Picone and G. R. Doddington. A phonetic vocoder. *Proc. ICASSP*, 580–583, 1989.
- [27] K. Yi. A vocoder based on speech recognition and synthesis. *GLOBECOM*, 3:1605–1609, 1995.
- [28] 広井 順, 徳田 恵一, 益子 貴史, 小林 隆夫, 北村 正. HMM に基づいた極低ビットレート音声符号化. 音講論集, 351–352, Autumn 1997.
- [29] J. K. Baker. The DRAGON system – an overview. *IEEE Trans. on Acoust. Speech, and Signal Process.*, 23(1):24–29, February 1975.
- [30] F. Jelinek, L. R. Bahl, and R. L. Mercer. Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Trans. on Inform. Theory*, 21(3):250–256, May 1975.
- [31] K. Tokuda, T. Kobayashi, and S. Imai. Speech parameter generation from HMM using dynamic features. *Proc. ICASSP*, 660–663, 1995.
- [32] 徳田 恵一, 益子 貴史, 小林隆夫, 今井聖. HMM からの音声パラメータ生成アルゴリズム  $\Delta$ . Technical Report SP95-122, 信学技報, January 1996.
- [33] 立和 航, 古井 貞熙. HMM による規則音声合成の検討. 音講論集, 239–240, 1999.
- [34] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Trans. on Acoust. Speech, and Signal Process.*, 37(1):31–42, January 1989.
- [35] L. D. Davisson. Universal noiseless coding. *IEEE Trans. on Inform. Theory*, IT-19(6):783–795, Nov. 1973.

- [36] L. D. Davisson, R. J. McEliece, M. B. Pursley, and M. S. Wallace. Efficient universal noiseless source codes. *IEEE Trans. on Inform. Theory*, IT-27(3):269–279, May 1981.
- [37] T. M. Cover. Enumerative source coding. *IEEE Trans. on Inform. Theory*, IT-17(1):73–77, Jan. 1973.
- [38] R. G. Gallager. Variations on a theme by Huffman. *IEEE Trans. on Inform. Theory*, IT-24(6):668–674, Nov. 1978.
- [39] P. A. Chou and T. Lookabough. Conditional entropy-constrained vector quantization of linear predictive coefficients. *Proc. ICASSP*, 197–200, 1990.
- [40] T. Kim. New finite state vector quantizers for images. *Proc. ICASSP*, 1180–1183, 1988.
- [41] C. B. Jones. An efficient coding system for long source sequences. *IEEE Trans. on Inform. Theory*, IT-27(3):280–291, May 1981.
- [42] J. Rissanen and G. G. Langdon, Jr. Universal modeling and coding. *IEEE Trans. on Inform. Theory*, IT-27(1):12–23, January 1981.
- [43] 嵩 忠雄. 情報と符号の理論入門. 情報工学入門選書 6. 昭晃堂, 1989.
- [44] J. Hagenauer, E. Offer, and L. Papke. Iterative decoding of binary block and convolutional codes. *IEEE Trans. on Inform. Theory*, 42(2):429–445, March 1996.
- [45] T. Fingscheidt and O. Scheufen. Robust GSM speech decoding using the channel decoder's soft output. *EUROSPEECH*, 1315–1318, 1997.
- [46] T. Fingscheidt and P. Vary. Robust speech decoding : A universal approach to bit error concealment. *Proc. ICASSP*, 1667–1670, 1997.
- [47] K. Sayood and J. C. Borkenhagen. Use of residual redundancy in the design of joint source / channel coders. *IEEE Trans. on Commun.*, 39(6):838–846, June 1991.
- [48] T. Fingscheidt, P. Vary, and J. A. Andonegui. Robust speech decoding : Can error concealment be better than error correction ? *Proc. ICASSP*, 1:373–376, 1998.

- [49] F. I. Alajaji, N. C. Phamdo, and T. E. Fuja. Channel codes that exploit the residual redundancy in CELP-encoded speech. *IEEE Trans. on Speech and Audio Process.*, 4(5):325–336, September 1996.
- [50] C. Berrow and A. Glavieux. Near optimum error correcting coding and decoding : Turbo-codes. *IEEE Trans. on Commun.*, 44(10):1261–1271, October 1996.
- [51] J. Garcia-Frias and J. D. Villasenor. Combining hidden Markov source models and parallel concatenated codes. *IEEE Commun. Lett.*, 1(4):111–113, July 1997.
- [52] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. on Inform. Theory*, 20:284–287, March 1974.
- [53] J. Hagenauer and P. Hoeher. Viterbi algorithm with soft-decision outputs and its applications. *Proc. of GLOBECOM*, 1680–1686, 1989.
- [54] F. Jelinek. Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 64(4):532–556, April 1976.
- [55] L. Rabiner and B.-H. Juang. 音声認識の基礎 (上). In *Modern Methods of Speech Processing*. NTT アドバンステクノロジー株式会社, 1995.
- [56] 広井 順, 徳田 恵一, 益子 貴史, 小林 隆夫, 北村 正. HMM に基づいた極低ビットレート音声符号化. 信学論, J82-D-II(11):1857–1864, November 1999.
- [57] 坂本 慶行, 石黒 真木夫, 北川 源四郎. 情報量統計学. 情報科学講座 A.5.4. 共立出版, 1983.
- [58] 板倉 文忠, 嵯峨山 茂樹. 線スペクトル周波数をパラメータとした音声合成法とその LSI 化. 日経エレクトロニクス, 128–157, 2nd February 1981.



# 对外発表

## 論文リスト

- [1] 米崎 正, 鹿野 清宏, “隠れマルコフモデルとエントロピー符号化を用いたベクトル量子化,” 信学論, Vol.J79-D-II, No.12, 2199-2206, December, 1996
- [2] 米崎 正, 吉田 幸司, 八木 敏男, 鹿野 清宏, “隠れマルコフモデルを用いた最尤推定に基づく誤り訂正,” 信学論, Vol.J83-A, No.1, 1-8, January, 2000

## 国際会議リスト

- [1] T. Yonezaki, K. Shikano, “Entropy coded vector quantization with hidden Markov models,” Proc. ICSLP, 310-313, October, 1996
- [2] T. Yonezaki, K. Yoshida, T. Yagi, “An error correction approach based on the MAP algorithm combined with hidden Markov models,” Proc. ICASSP, I-33-36, May, 1998

## その他の発表リスト

- [1] 米崎 正, 鹿野 清宏, “HMM 確率モデルとエントロピー符号を用いた誤り耐性のあるベクトル量子化,” 信学技報, SP95-131, 17-24, 1995
- [2] 米崎 正, 吉田 幸司, 八木 敏男, “HMM を用いた最尤推定に基づく誤り訂正,” 信学技報, SP97-43, 45-51, 1997
- [3] 米崎 正, 吉田 幸司, 八木 敏男, 鹿野 清宏, “MA 予測型量子化におけるフレーム損失パラメータの HMM に基づく補間,” 信学技報, SP98-92, 41-48, 1998

- [4] 米崎 正, 鹿野 清宏, “時間周波数マスクングを利用したスペクトル包絡のベクトル量子化,” 音講論集, 3-1-8, 283-284, September, 1995
- [5] 米崎 正, 鹿野 清宏, “HMM 確率モデルに基づくエントロピー符号を用いたベクトル量子化,” 音講論集, 1-4-17, 251-252, March, 1996
- [6] 米崎 正, 吉田 幸司, 八木 敏男, “MBE 音声符号化に対する HMM とハフマン符号を用いたベクトル量子化法,” 音講論集, 3-4-5, 269-270, September, 1996
- [7] 米崎 正, 吉田 幸司, 八木 敏男, “離散 HMM により符号帳を選択するベクトル量子化手法,” 音講論集, 3-7-16, 289-290, March, 1997
- [8] 米崎 正, 吉田 幸司, 八木 敏男, “HMM を用いた誤り訂正法の検討,” 音講論集, 3-2-11, 287-288, September, 1997
- [9] 米崎 正, 吉田 幸司, 八木 敏男, “RS 符号と HMM を組み合わせた誤り訂正法の検討,” 音講論集, 2-7-10, 253-254, March, 1998
- [10] 米崎 正, 吉田 幸司, 八木 敏男, 鹿野 清宏, “HMM に基づく損失したゲインパラメータの補間に関する検討,” 音講論集, 3-2-14, 275-276, September, 1998
- [11] 米崎 正, 吉田 幸司, 八木 敏男, 鹿野 清宏, “STRAIGHT スペクトルパラメータの 2 次元 DCT に基づく量子化に関する検討,” 音講論集, 1-3-6, 191-192, September, 1999
- [12] 吉岡 大祐, 陸 金林, 中村 哲, 鹿野 清宏, 米崎 正, “高調波成分を利用した STRAIGHT の極低ビットレート音声符号化,” 音講論集, 1-3-7, 193-194, September, 1999