社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
NLC2011-52,SP2011-97(2011-12)

# パラ言語情報を表現可能な対話音声合成のための重回帰 HSMM の検討

永田　智洋† 　森　　大毅† 　能勢　　隆††

† 宇都宮大学大学院工学研究科　〒 321–8585 栃木県宇都宮市陽東 7–1–2
†† 東京工業大学大学院総合理工学研究科　〒 226–8502 神奈川県横浜市緑区長津田町 4259–G2–4
E-mail: †{ken1,hiroki}@speech-lab.org, ††taksahi.nose@ip.titech.ac.jp

**あらまし**　本稿では，隠れセミマルコフモデル (HSMM) に基づく音声合成方式に重回帰モデルを組み込んだ重回帰 HSMM を用いて，対話音声に見られる多様なパラ言語情報を制御可能な音声合成を目指す．本研究では，パラ言語情報を少数の次元から構成される空間上の座標として表現し，この空間を構成する次元を重回帰モデルの説明変数として用いる．次元には感情状態を表す一般的な指標とされている「快-不快」,「覚醒-睡眠」の 2 つの次元を用いる．モデルの学習時には各発話に対し次元毎に主観的に評価された評価値を用いて学習し，合成時には任意の評価値を与えて任意の感情状態の音声を合成する．合成された音声の音響的特徴量から，2 つの次元が合成音声に与える影響について検討する．また，合成された音声に対して 3 つの主観評価実験を行った．まず，自然性評価を行い，合成された音声の自然性について示した．次に，再現性評価を行い，付与した感情状態の再現性について示した．最後に，感情状態の表出について評価を行い，意図した感情状態が伝達されていることを示した．
**キーワード**　HMM 音声合成, 重回帰 HSMM, 対話音声合成, パラ言語情報

# An MRHSMM-based conversational speech synthesis with controllability of paralinguistic information

Tomohiro NAGATA†, Hiroki MORI†, and Takashi NOSE††

† Faculty of Engineering, Graduate School of Engineering, Utsunomiya University　Youtou 7–1–2, Utsunomiya-shi, Tochigi, 321–8585 Japan
†† Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology　4259–G2–4, Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa, 226–8502, Japan
E-mail: †{ken1,hiroki}@speech-lab.org, ††taksahi.nose@ip.titech.ac.jp

**Abstract**　In this paper, we aim at the realization of the speech synthesis that can control paralinguistic information using multiple regression HSMM, incorporated a multiple regression model in hidden semi-Markov model(HSMM)-based speech synthesis scheme. In this study, the paralinguistic information is expressed as a coordinate on space comprised of a small number dimension and the dimensions are used as an explanation variable of the multiple regression model. Two dimensions that considered to be a general index to express emotional state for “PLEASANTNESS” and “AROUSAL” are used. When learning model, evaluated values are used subjectively for each dimenstions. And when synthesize speech, we synthesize any speech that reflected emotion by giving arbitrary values. We examine the influence that two dimensions give synthesized speech with acoustic features of synthesized speech. Additionally, we have three subjective experiments for synthesized speech. First, the result of a naturally test show that synthesized speech are natural. Next, the result of a reproducibility test show that reproducibility of given emotion. Finally, the result of a emotional expression test show that synthesized speech transmit an aimed emotion.
**Key words**　HMM-based speech synthesis, multiple regression HSMM, conversational speech synthesis, paralinguistic information

## 1. はじめに

音声合成技術の発達により，現在では明瞭性および自然性の高い音声を合成することが可能となっている．しかし，現在実用化されている音声合成システムの多くはニュースキャスターやナレータのような読み上げ調の音声しか合成できないものがほとんどである．コミュニケーション支援インタフェースやヒューマンインタフェースでは，読み上げ調の音声よりも感情状態などのパラ言語情報や話者の個人性などの非言語情報を反映することができる音声合成，すなわち対話音声合成が必要である．

そこで，本研究では感情状態などのパラ言語情報を制御可能な対話音声合成の実現を目指す．感情状態は，従来「喜び」や「悲しみ」といった範疇的な項目で表現されることが一般的であるが，日常的な対話に含まれる感情を範疇的な項目で表現することは困難である．そのため，次元説に基づく方法によって表現されたパラ言語情報を用いる音声合成の研究 [1] [2] が行われている．これらの研究では 6 つの抽象次元を用いてパラ言語情報を表現しているが，パラ言語情報と音声の音響パラメータの関係性について明示的な関係は仮定していなかった．パラ言語情報と音声の音響パラメータの関係性について明示的な関係を仮定した音声合成の研究には文献 [3] [4] がある．これらの研究では，発話スタイルで表されるパラ言語情報と音声の音響パラメータとの間に線形の関係があると仮定し，重回帰モデル [5] を用いた音声合成によって合成音声の発話スタイルの制御を行っている．そこで，本研究では重回帰モデルと，前述した次元説に基づく方法によって記述されたパラ言語情報を用いることで，合成音声のパラ言語情報を制御することを試みる．これによって，学習データに含まれていないパラ言語情報を含む音声を合成する場合においても，合成音声の音響パラメータを補外することができると期待される．また，これまでのコンテキストクラスタリングによるパラ言語情報の制御とは異なり，自由で連続的なパラ言語情報の制御を行うことができる．

## 2. 合成音声のパラ言語情報制御

本節では，重回帰 HSMM を用いた音声合成によってパラ言語情報を制御する方法について説明する．宇都宮大学では自然で表情豊かな対話音声に含まれるパラ言語情報に関する研究への利用に適した音声対話データベースとして，宇都宮大学パラ言語情報研究向け音声対話データベース (以下，UUDB) [8] を開発し，広く公開している．UUDB では収録されている全ての発話に対して，音声から知覚される話者の感情状態を記述したパラ言語情報ラベルが付与されている．付与されているパラ言語情報ラベルは感情状態を表す「快-不快」,「覚醒-睡眠」に加えて話者間の対人関係を表す「支配-服従」,「信頼-不信」，そして話者自身の態度を表す「関心-無関心」,「肯定的-否定的」という 6 つの次元を用い，各次元について 7 段階の評価を施すことによって記述されている．

重回帰 HSMM は HSMM における出力確率分布および状態継続長分布の平均ベクトルが重回帰モデルによって表現されると仮定するものである [4]．文献 [4] では，HSMM の状態 $i$ における出力確率分布 $\boldsymbol{b}_i(\boldsymbol{o})$ と継続長分布 $p_i(d)$ に単一のガウス分布を仮定したとき，それぞれの分布の平均ベクトル $\boldsymbol{\mu}_i$，$m_i$ を次のように仮定する．

$$\boldsymbol{\mu}_i = \boldsymbol{H}_{bi}\boldsymbol{\xi}_b \tag{1}$$

$$m_i = \boldsymbol{H}_{pi}\boldsymbol{\xi}_p \tag{2}$$

ここで，$\boldsymbol{H}_{bi}$，$\boldsymbol{H}_{pi}$ はそれぞれ出力確率分布の平均ベクトル，状態継続長分布の平均ベクトルに対する回帰行列である．$\boldsymbol{\xi}_b$，$\boldsymbol{\xi}_p$ はそれぞれ出力確率分布の平均ベクトル，状態継続長分布に対する回帰行列の説明変数からなる制御ベクトルであり，説明変数の次元を $L$ としたとき，$(1+L)$ 次元のベクトルとなる．したがって，出力確率分布の平均ベクトルの次元を $M$ としたとき，$\boldsymbol{H}_{bi}$ および $\boldsymbol{H}_{pi}$ はそれぞれ $M\times(L+1)$ 次元，$1\times(L+1)$ 次元の回帰行列となる．また，それぞれの回帰行列の再推定式は以下の式で与えられる．

$$\bar{\boldsymbol{H}}_{b_i} = \left(\sum_{n=1}^{K}\sum_{t=1}^{T^{(n)}}\sum_{d=1}^{t}\gamma_t^d(i)\left[\sum_{\tau=t-d-+1}^{t}\boldsymbol{o}_\tau^{(n)}\right]\boldsymbol{\xi}_b^{(n)\top}\right)\cdot \left(\sum_{n=1}^{K}\sum_{t=1}^{T^{(n)}}\sum_{d=1}^{t}\gamma_t^d(i)\cdot d\cdot\boldsymbol{\xi}_b^{(n)}\boldsymbol{\xi}_b^{(n)\top}\right)^{-1} \tag{3}$$

$$\bar{\boldsymbol{H}}_{p_i} = \left(\sum_{n=1}^{K}\sum_{t=1}^{T^{(n)}}\sum_{d=1}^{t}\gamma_t^d(i)\cdot d\cdot\boldsymbol{\xi}_p^{(n)\top}\right)\cdot \left(\sum_{n=1}^{K}\sum_{t=1}^{T^{(n)}}\sum_{d=1}^{t}\gamma_t^d(i)\boldsymbol{\xi}_p^{(n)}\boldsymbol{\xi}_p^{(n)\top}\right)^{-1} \tag{4}$$

ここで K は観測系列の総数，$T^{(n)}$ は n 番目の観測系列 $\boldsymbol{O}^{(n)}$ の総フレーム数，$\boldsymbol{o}_\tau$ は $\boldsymbol{O}^{(n)}$ の時刻 $\tau$ における観測ベクトル，$\gamma_t^d(i)$ は状態 $i$ で観測系列 $\boldsymbol{o}_{t-d+1}^{(n)},\ldots,\boldsymbol{o}_t^{(n)}$ を出力する確率であり，状態占有確率と呼ばれる．

本研究では重回帰モデルの説明変数に，UUDB に記述されている 6 次元のパラ言語情報を用いることで，合成音声のパラ言語情報の制御を行う．したがって，重回帰モデルの説明変数によって構成される低次元のベクトルを $\boldsymbol{v}$ で表すと，式 (1)，式 (2) は次のようになる．

$$\begin{aligned}\boldsymbol{\xi}_b = \boldsymbol{\xi}_p &= [1,\boldsymbol{v}^\top]^\top \\ &= [1, v_{\text{pleasantness}}, v_{\text{arousal}},\ldots,v_{\text{positivity}}]^\top\end{aligned} \tag{5}$$

式 (5) を用いて，出力確率分布及び状態継続長の平均ベクトルを計算し，これらを用いて HSMM 音声合成を行う．また，従来の重回帰 HSMM では回帰行列は初期化を行うために共有決定木コンテキストクラスタリング [6] を用いる．しかし，本手法では文献 [3] および文献 [4] であるところの単一スタイルによる音声合成となるため，共有決定木コンテキストクラスタリングを用いることができない．そこで，文献 [7] の初期化手法を用いた．文献 [7] の初期化手法は，EM アルゴリズムによる回帰行列の再推定において状態 s 占有確率の近似を用いて重回帰

モデルの初期化を行うものであり，単一スタイルの場合においても重回帰モデルの初期化を行うことができる．

## 3. 合成音声の作成

### 3.1 合成条件

実験には UUDB の対話セッション C002 から C007 までの話者 FTS についての発話を用いた．スペクトルパラメータはメルケプストラムとし，サンプリング周波数 16kHz の音声信号から，分析周期 5ms，分析窓長 25ms のハミング窓を用いて求めた 0 次から 24 次のメルケプストラム係数を用いた．F0 パラメータは対数基本周波数とした．したがって，特徴ベクトルはこれらのパラメータにそれぞれのデルタ，デルタデルタパラメータを加えた 78 次元とした．HSMM は 5 状態の left-to-right モデルとし，前述した話者 FTS の 550 発話を用いて学習を行った．音素単位は無音とポーズを含めた 32 種類の音素を用い，各音素の継続時間は大語彙連続音声認識エンジン Julius [9] を利用した強制アライメントを用いて求めた．学習時にはコンテキストクラスタリングにより決定木を構築し，この決定木により作成された分布に対して出力分布および状態継続長分布の再推定を行った．ここで，再推定に必要となる初期回帰行列は文献 [7] による初期化手法を用いた．制御ベクトルは UUDB に付与されている 6 次元のパラ言語情報から，感情状態を表す一般的な指標とされている「快-不快」,「覚醒-睡眠」を用いた．UUDB に付与されているパラ言語情報の各次元には相関があるため，多重共線性による影響を避けるためである．したがって，式 (5) のベクトル $\boldsymbol{v}$ は 2 次元となり，制御ベクトルは式 (6) とした．

$$\boldsymbol{\xi} = [1, v_{\mathrm{pleasantness}}, v_{\mathrm{arousal}}] \tag{6}$$

また，各次元の値にはラベラ 3 名による平均評定値を用いた．合成内容は UUDB の対話セッション C001 の話者 FTS についての 25 発話とした．合成時に付与する感情状態値は「快-不快」,「覚醒-睡眠」の各次元について，3.0 から 5.0 まで 1.0 刻みで変化させた 9 通りの値を与えた．

### 3.2 発話継続時間

各評価値を与えた場合の発話時間の平均を図 1 に示す．図より,「快-不快」の次元に与えた値が大きくなるほど発話時間は短くなり,「覚醒-睡眠」の次元に与えた値が大きくなるほど発話時間は長くなった．したがって，発話時間が最も長くなったのは $(v_{\mathrm{pleasantness}}, v_{\mathrm{arousal}}) = (3.0, 5.0)$ の場合の 0.681[s] であり，最も短くなったのは $(v_{\mathrm{pleasantness}}, v_{\mathrm{arousal}}) = (5.0, 3.0)$ の場合の 0.625[s] であった．また，発話時間の変化は,「快-不快」の次元よりも「覚醒-睡眠」の次元による影響が大きいという傾向が得られた．

### 3.3 基本周波数

各評価値を与えた場合の F0 最大値の平均を図 2 に示す．図より,「快-不快」の次元の値が高いほど合成音声の F0 の最大値は高くなる傾向が得られた．また,「覚醒-睡眠」の次元については，評価値 4.0 を与えた場合に F0 の最大値が最も小さくなり，評価値 4.0 を基準として低い評価値および高い評価値のどちらを与えた場合についても F0 の最大値は大きくなった．

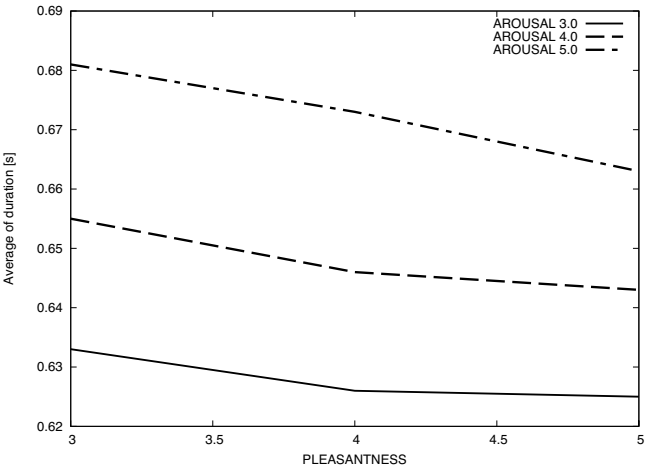


図 1 各評価値における発話時間の平均
Fig. 1 Average of the duration in each score.

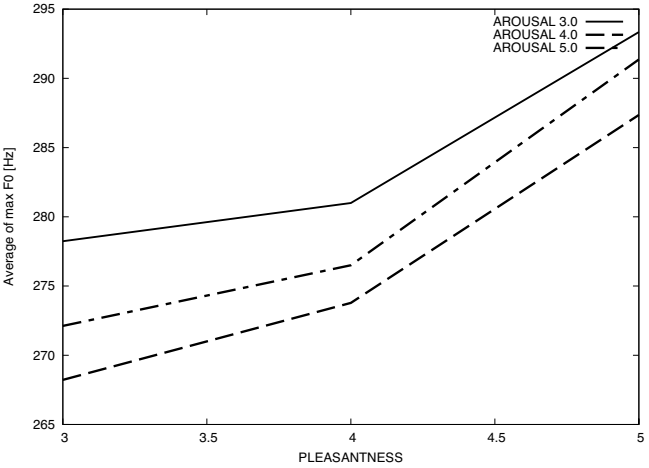


図 2 各評価値における F0 最大値の平均
Fig. 2 Average of max F0 in each score.

各評価値を与えた場合の F0 レンジの平均を図 3 に示す．F0 レンジについては,「覚醒-睡眠」の次元に与える評価値が高いほど合成音声の F0 レンジは小さくなる傾向が得られた．また,「快-不快」の次元については，評価値 4.0 を与えた場合に F0 レンジが最も小さくなり，評価値 4.0 を基準として低い評価値および高い評価値のどちらを与えた場合においても F0 レンジは大きくなる傾向となった．

また，F0 最大値については「快-不快」，F0 レンジについては「覚醒-睡眠」の次元の値の影響を大きく受ける傾向が得られた．

### 3.4 フォルマント周波数

感情状態を表す各次元の値の変化が，合成音声の聴音的特徴に与える影響を調べる目的で，合成音声のスペクトルについての分析を行った．各次元の値の変化によって，合成音声の各母音のフォルマント周波数が受ける影響について調べた.「快-不快」および「覚醒-睡眠」による，母音のフォルマント周波数の変化を図 4 に示す．また，それぞれのフォルマント周波数は，各母音の継続時間における各フレームのフォルマント周波数の平均としている．図では，与えた「快-不快」および「覚醒-睡

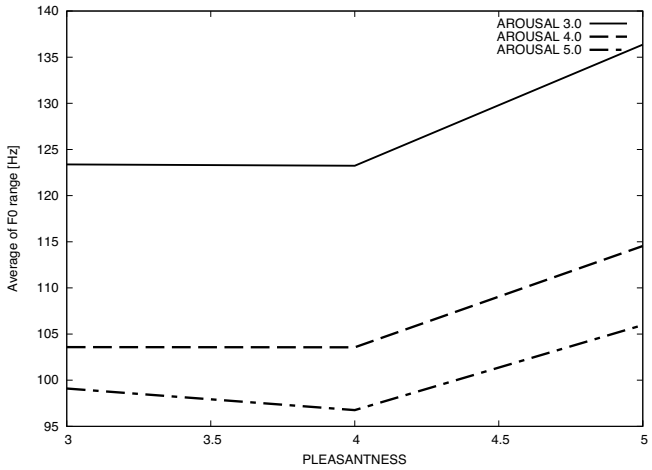


図 3　各評価値における F0 レンジの平均
Fig. 3　Average of F0 range in each score.

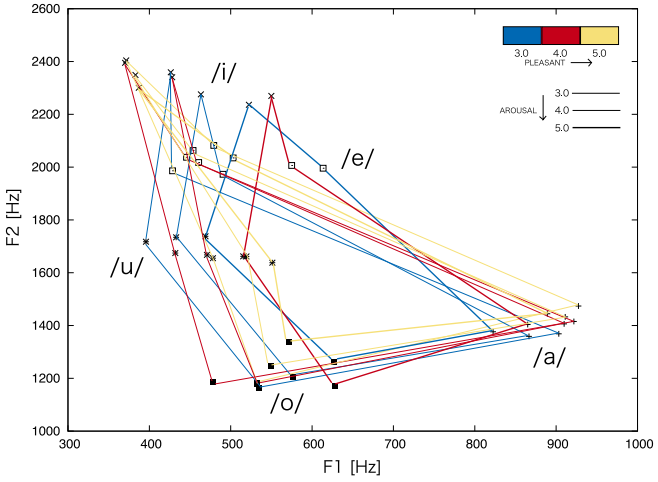


図 4　各母音のフォルマント周波数
Fig. 4　Formant frequency of each vowels.

眠」の値が等しい母音を線で結んでいる．また,「快-不快」の次元は線の色によって表現され,「覚醒-睡眠」の次元は線の太さによって表現される．

比較した母音/a/は合成音声「エーは，うんとね」の/e:wa/の/a/とした．得られた結果より,「快-不快」の次元の値が大きいほど，第 1 フォルマント周波数および第 2 フォルマント周波数が高くなるという傾向が得られた.「覚醒-睡眠」の次元については，値が大きくなるほど第 1 フォルマント周波数は低く，第 2 フォルマント周波数は高くなった．したがって，母音/a/は「快-不快」の次元の値を小さく,「覚醒-睡眠」の次元の値を大きくほど母音/o/のフォルマント周波数に近づいた．

母音/i/は合成音声「で，ディーは」を用いて比較した．母音/i/では,「快-不快」の次元の値が大きくなるほど，第 1 フォルマント周波数は低く，第 2 フォルマント周波数は高くなる傾向が見られた.「覚醒-睡眠」の次元については，値が大きくなるほど，第 1 フォルマント周波数は高く，第 2 フォルマント周波数は低くなった．したがって，母音/i/では,「快-不快」の次元の値を小さく,「覚醒-睡眠」の次元の値を大きくするほど，母音/e/のフォルマント周波数に近づいた．

母音/u/は合成音声「うん」を用いて比較した．母音/u/では,「快-不快」の次元の値が大きくなるほど，第 1 フォルマント周波数が高くなり，第 2 フォルマント周波数は低くなる.「覚醒-睡眠」の次元については，値が大きくなるほど，第 1 フォルマント周波数が高くなり，第 2 フォルマント周波数は低くなる傾向が得られた．しがたって，母音/u/では「快-不快」の次元の値を小さく,「覚醒-睡眠」の次元の値を小さくするほど，母音/i/のフォルマント周波数に近づいた．また,「快-不快」の次元の値を大きく,「覚醒-睡眠」の次元の値を大きくするほど，母音/o/のフォルマント周波数に近づいた．

母音/e/は合成音声「エーは，うんとね」の最初の/e/を用いた．母音/e/では,「快-不快」の次元の値が大きくなるほど，第 1 フォルマント周波数は低く，第 2 フォルマント周波数が高くなる傾向が得られた．また,「覚醒-睡眠」の次元については，値が大きくなるほど第 1 フォルマント周波数が高く，第 2 フォルマント周波数が低くなるという傾向が得られた．したがって，母音/e/では,「快-不快」の次元の値を大きく,「覚醒-睡眠」の次元の値を小さくするほど母音/i/に近づいた．また,「快-不快」の次元の値を小さく,「覚醒-睡眠」の次元の値を大きくするほど，母音/a/に近づく傾向が得られた．

母音/o/は合成音声「そのよこにじいちゃんがすわってて」の最初の/o/を用いた．母音/o/では,「快-不快」の次元の値が大きくなるほど，第 2 フォルマント周波数が大きくなる傾向がみられた．また,「覚醒-睡眠」の次元については，値が大きくなるほど，第 1 フォルマント周波数は高く，第 2 フォルマント周波数も高くなった．したがたって，母音/o/は「快-不快」の次元の値を大きく,「覚醒-睡眠」の次元の値を大きくするほど，母音/a/のフォルマント周波数に近づく結果となった．

一部の母音については，予想に反した結果が得られた．特に，母音/a/では「覚醒-睡眠」の次元の値を大きくするほど，第 1 フォルマント周波数が上昇すると予測されたが，得られた結果は対照的なものとなった．

## 4.　主観評価実験

### 4.1　実 験 条 件

後述する 3 つの主観評価実験に共通する条件を以下に示す．また，実験ごとに異なる条件は各々に示す．

モデルの学習時の条件は，3 節で行った分析時の条件と同様の条件とした．主観評価実験の被験者は 3 名の男性大学院生と 5 名の男性大学生の計 8 名であり，ヘッドホンによる両耳聴取によって評価を行った．また，被験者は 2 節で示したパラ言語情報を表す次元については十分に理解している．

### 4.2　合成音声の自然性評価

合成された音声が自然であるかの評価を行った．合成する音声の内容は UUDB の対話セッション C001 の話者 FTS についての発話内容 70 発話とした．与える「快-不快」,「覚醒-睡眠」の値は UUDB に付与されている 3 名のラベラによる平均評定値を与えた．評価は「肉声らしさ」と「対話音声らしさ」の観点を総合的に評価し,「自然：5」,「やや自然：4」,「どちらでもない：3」,「やや不自然：2」,「不自然：1」の 5 段階から選択す

表 1　全被験者の MOS と標準偏差
Table 1　MOS and standard deviation at all examinee.

| | MOS 値 | 標準偏差 |
|---|---|---|
| 全被験者 | 3.55 | 0.95 |

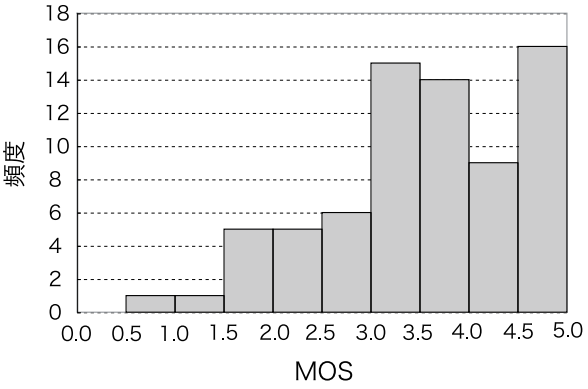


図 5　全被験者の平均のヒストグラム
Fig. 5　Histogram of the evaluation level in all subjects

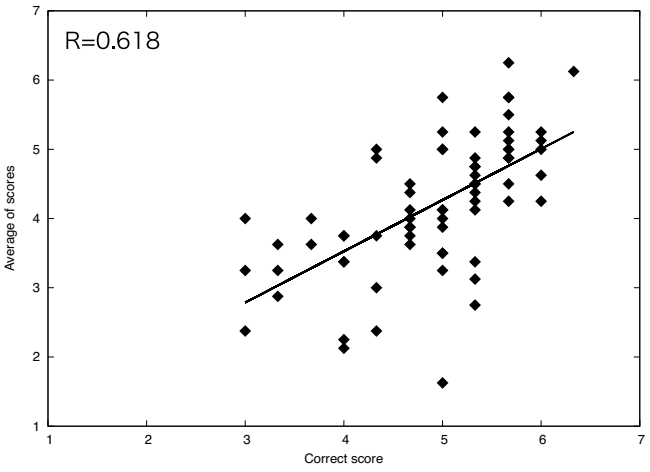


図 6 「快-不快」における正解評価値と平均評価値
Fig. 6　Correct score and average score

る形式で行った．

全被験者の平均評価値 (MOS: Mean Opinon Score) と標準偏差を表 1 に示し，図 5 に全被験者の MOS 値のヒストグラムを示す．表 1 より，全被験者の MOS 値は 3.55 であり，中央値である 3 よりも高い値を示している．また，図 5 より，MOS 値は 4.5 よりも大きく，5.0 未満の部分に最も多く分布し，全体的に見て MOS 値が 3.0 よりも大きい部分に多く分布している．これらの結果から，合成された音声は比較的自然であるといえる．

### 4.3　感情状態の再現性評価

音声合成時に与えた感情状態の再現性の評価を行った．評価に用いる合成音声は自然性評価に用いたものと同様に UUDB の対話セッション C001 の話者 FTS の発話内容 70 発話とした．与える「快-不快」,「覚醒-睡眠」の値は UUDB に付与されている 3 名のラベラによる平均評定値を与えた．評価方法は合成音声から受ける印象を,「快-不快」,「覚醒-睡眠」の項目について

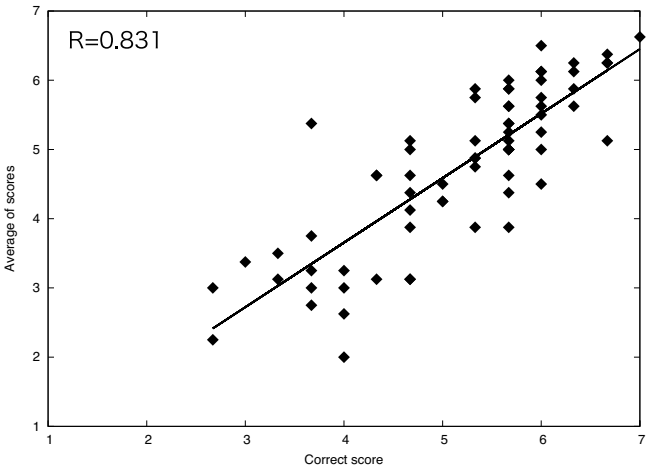


図 7 「覚醒-睡眠」における正解評価値と平均評価値
Fig. 7　Correct score and average score

7 段階で評価する形式で行った．

合成音声を作成する際に付与した平均評価値と被験者の評価値を比較するために，付与した評価値と被験者の評価値の平均値を求めた．図 6 と図 7 にそれぞれの次元の正解評価値に対する被験者の評価値の平均を示す．図の横軸は合成時に与えた正解の評価値であり，縦軸は被験者によって評価された評価値の平均である．これらの図より，どちらの次元についても被験者による評価値は与えた評価値よりも低く知覚される傾向が見られた．また，各次元の相関係数を図中に示す．この結果から,「快-不快」,「覚醒-睡眠」のどちらの次元についても高い相関係数が得られたことが確かめられた．また，これらの分布に対して，回帰直線を求め，決定係数 $R^2$ を求めた．回帰直線はそれぞれ図 6，図 7 に示す.「快-不快」の次元の分布に対する決定係数 $R^2$ は 0.38 と比較的小さい値であるが,「覚醒-睡眠」の次元の分布に対する決定係数 $R^2$ は 0.69 と比較的大きい値が得られた．

相関係数および決定係数は「快-不快」の次元のものと比較して,「覚醒-睡眠」の次元の方が高くなるという傾向が得られた．この原因としては,「覚醒-睡眠」の次元の値を変化させた場合の発話の継続時間および F0 レンジの変化の度合が「快-不快」の次元を変化させた場合の度合と比較して大きいことが考えられる．

### 4.4　感情状態の表出評価

音声合成時に与える感情状態を変化させることにより，合成音声に様々な感情状態が反映されているか評価を行った．評価に用いる合成音声は UUDB の対話セッション C001 の話者 FTS の発話内容 25 発話とした．与える「快-不快」,「覚醒-睡眠」の値は図 8 にプロットされている 7 点の組み合わせとし，各合成音声について付与した感情状態値の重複がないようランダムに入れ替えた 25 発話 7 セットの合成音声を呈示した．評価方法は合成音声から受ける印象を,「快-不快」,「覚醒-睡眠」の次元について 7 段階で評価する形式で行った．

発話毎の各被験者の評価値の平均と与えた評価値との相関係数の平均を表 2 に示す．表より,「快-不快」,「覚醒-睡眠」のど

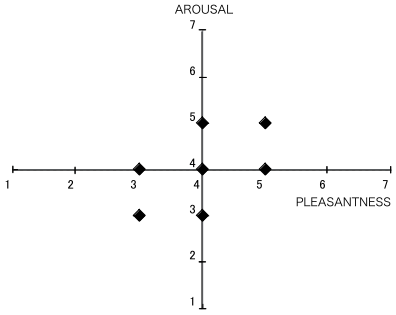


図 8　付与した感情状態値

Fig. 8　The emotional score that I give.

表 2　付与した評価値と被験者の平均評価値の相関係数

Table 2　Cofficient of correlation with average score and the score that I give.

| | 快 - 不快 | 覚醒 - 睡眠 |
|---|---|---|
| 相関係数 | 0.692 | 0.797 |

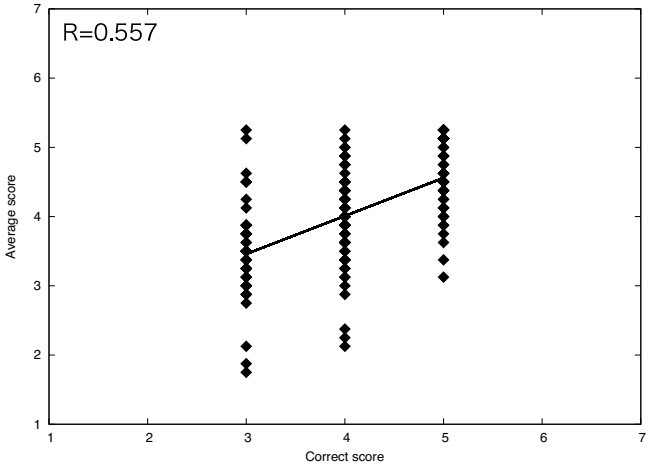


図 9 「快-不快」における正解評価値と平均評価値

Fig. 9　Correct score and average score

ちらの次元においても 0.6 以上の相関係数が得られたことから，意図した感情状態が伝達できたことが確認できた．また，各次元に付与した評価値に対して，被験者が評価した評価値の平均を図 9 および図 10 に示す．これらの相関係数においても「快-不快」の次元では 0.577,「覚醒-睡眠」の次元では 0.679 と比較的高い値を得られた．

## 5. おわりに

本研究では，重回帰 HSMM に基づく音声合成方式を用いて，対話音声合成におけるパラ言語情報を制御することを試みた．制御するパラ言語情報には,「喜び」や「悲しみ」といった範疇的に表現されるものではなく，次元説に基づく方法によって表現されたパラ言語情報を用いた．具体的には，感情状態を表す一般的な指標とされている「快-不快」および「覚醒-睡眠」の 2 つの次元を用いた．

合成された音声の各音響特徴量を分析し，パラ言語情報の制御に用いた 2 つの次元が合成音声に与える影響について検討した．得られた結果から，また，合成音声の主観評価実験を行い，合成された音声に現れた各音響特徴量の変化が，付与したパラ言語情報を再現しているか評価した．加えて，パラ言語情報の表出度合についても評価を行い，意図したパラ言語情報が伝達されているかを評価した．また，合成された音声の自然性についても評価を行い，比較的自然な音声が合成されたことを確かめた．

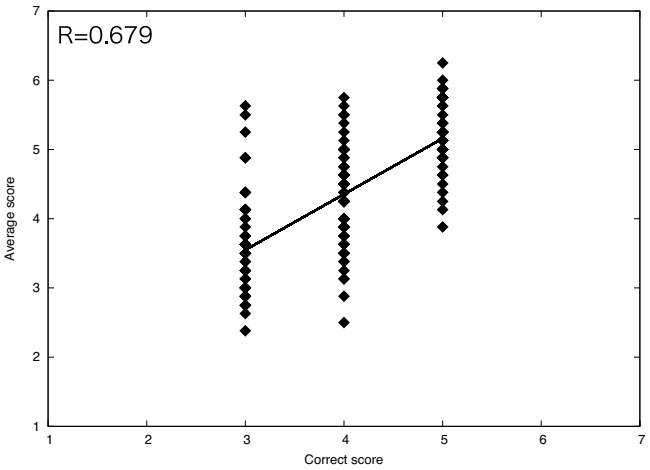


図 10 「覚醒-睡眠」における正解評価値と平均評価値

Fig. 10　Correct score and average score

今後の課題としては，学習データを確保するための話者適応などが挙げられる．

### 文　献

[1] 森大毅, “相槌音声合成におけるパラ言語情報の感情次元による制御,” 日本音響学会 2009 年秋季研究発表会講演論文集, 1–2–P, pp.253–254, 2009.

[2] 人見貴嗣, “表情豊かな対話音声の合成に関する研究,” 宇都宮大学修士論文, 2011.

[3] 宮永圭介, 益子貴史, 小林隆夫, “HMM 音声合成における多様なスタイル実現のための制御法,” 信学技報, SP2004–7, pp.35–40, 2004.

[4] 能勢隆, 山岸順一, 小林隆夫, “重回帰 HSMM を用いた合成音声のスタイル制御,” 信学技報, SP2005–160, pp.61–66, 2005.

[5] 藤永勝久, 中井満, 下平博, 嵯峨山茂樹, “連続量を変形要因とする重回帰モデルを内包する HMM,” 信学技報, SP2000–83, pp.49–54, 2000.

[6] J. Yamagichi, M. Tamura, T.Masuko, K. Tokuda, and T. Kobayashi, “A context clustering technique for average voice models,” IEICE Trans. Information and Systems, vol.E86–D, nol3, pp.534–542, March 2003.

[7] 能勢隆, 小林隆夫, “感情音声合成における主観的表出度合のモデル化と制御の検討,” 日本音響学会 2011 年秋季研究発表会講演論文集, 3–8–1, pp.329–330, 2011.

[8] H. Mori, T. Satake, M. Nakamura, H. Kasuya, “Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics”, Speech Communication Vol. 53 pp. 36–50, 2011.

[9] 李 晃伸, “大語彙連続音声認識エンジン Julius ver.4,” 情報処理学会研究報告, 2007–SLP–69–53, 2007.