

科学研究費補助金研究成果報告書

平成 21 年 6 月 10 日現在

研究種目：基盤研究 (B)

研究期間：2005～2008

課題番号：17300063

研究課題名 (和文) デジタルメディアコンテンツ制作のための多様な音声の合成技術

研究課題名 (英文) Flexible Speech Synthesis for Digital Media Production

研究代表者

徳田 恵一 (TOKUDA KEIICHI)

名古屋工業大学・大学院工学研究科・教授

研究者番号：20217483

研究成果の概要：本研究は、「ト書き」付きの台本を声優の如く、自在な表現で読み上げることのできる音声合成装置の実現を目指し、種々の声質、スタイルの音声を少量の音声データから作成する“柔軟な”音声合成技術を開発した。音声のみに止まらず、歌声についても研究対象とした。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2005 年度	4,100,000	0	4,100,000
2006 年度	3,900,000	0	3,900,000
2007 年度	3,700,000	222,000	3,922,000
2008 年度	3,200,000	960,000	4,160,000
総計	14,900,000	1,182,000	16,082,000

研究分野：総合領域

科研費の分科・細目：情報学 知覚情報処理・知能ロボティクス

キーワード：音声合成，メディアコンテンツ，声質，発話様式，HMM

1. 研究開始当初の背景

映画，TV ゲーム他のエンターテインメントコンテンツの制作においては，コンピュータ

グラフィックスが広く使われ，その自動化も進んでいる．その一方で，エンターテインメントコンテンツにおいて，重要な位置を占める

登場キャラクタの音声に関しては、声優による実音声が使われているのが現状である。

音声合成器を用い、これを代用させることも考えられるが、あいにく、これまでの音声合成器は、特定話者による読みあげ調の音声を合成することはできるものの、様々な場面や状況に応じた音声を合成することはできず、このような用途には到底用いることができないものであった。

それに対し、我々はこれまでに多様な音声の合成を目指した「HMM 音声合成方式」を提案し、様々な声質、発話スタイル、感情などをもった音声を容易に合成可能であることを示した。更に、話者適応（声を真似る）、話者補間（声を混ぜる）、固有声（声をつくる）などの手法を提案し、これまでになく多様な合成音声を生成可能であることを示した。このことから、HMM 音声合成の枠組みは、本研究での目標を達成するための手法として有力な候補とすることができる。

2. 研究の目的

以上の背景より本研究では、「HMM 音声合成方式」をベースとして、「ト書き」付きの台本を声優の如く、読み上げることのできる音声合成装置の実現を目指す。このために、さまざまな類型的な感情表現や発話スタイルをモデル化可能な手法を確立する。また、感情表現や発話スタイルだけではなく、歌声、叫び声など、更に人間的で多様な音声の合成をも含めて考える。

上述の通り、現状では、映画、TV ゲームなどにおける音声メディアコンテンツの制作は、人間の実音声に頼っており、オーディション等による適切な声優の選定、音声の収録など、多大な時間と資金を要するものとなっている。また、音声収録後に、シナリオに変更があり、部分的な再収録が必要となることも多いが、当該声優が既に引退するなどの理由により、同じ話者の音声収録できないという問題も発生する。その他にも、カーナビゲーションなどに代表される ITS 関連の情報提供システム、駅構内における案内放送、電話による公共情報提供システムなどについ

ても全く同様の問題を抱えており、それらは枚挙に暇がない。

更に一般のパソコンユーザーが CG による映画作品を制作することも一般的になったが、この際にも音声を伴った作品の場合には、声優に音声収録を依頼する必要がある。このような場合にも、本音声合成装置の利用により、完全なデスクトップ上のみでの作品制作が可能となり、メディアコンテンツ制作を一般の人々にも広く開かれたものとするができる。また、電子書籍などのテキスト情報に「ト書き」を付与することにより、それらを容易に音声メディアに変換することが可能となることから、視覚障害者への福祉の意味でも大きな貢献となることが期待される。

3. 研究の方法

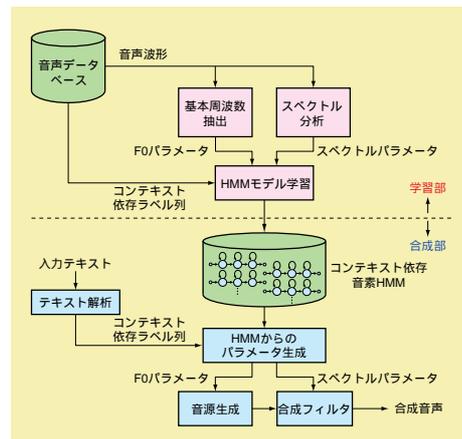


図 1 HMM 音声合成システム

既に、音素 HMM を接続し、出力確率が最大となる音響パラメータの時系列を生成して音声合成を行う HMM 音声合成方式（図 1）を確立しているが、HMM を新しい話者、新しい発話スタイル等に適応させることにより、それらの音声を合成することを進める。また、これまでに開発した音源モデル化手法を発展させ、自然音声から高精度で肉声感の高い音声の生成を可能とする手法を確立する。HMM の適応手法についても、各状態の出力分布を有効に共有し、少量の音声データでの適応学習を可能とすることを行う。さらに、適応手法によって得た HMM の変換ベクトルの内挿・外挿を行うことにより、音声

コーパスの得られない声質等の実現を図る。これらにより個性豊かな声質や発話スタイルでしかも感情が込められた音声を中心に制御・生成することが可能となる。

4. 研究成果

研究成果を以下に列挙する。

- (1) 音声データの自動ラベリング：読み上げ調の音声データに関しては、自動ラベリングが可能との見通しを得ていたが、更に、会話調等の多様な音声データの自動ラベリングを考慮したラベルなしデータからの音響モデルの自動学習法について提案・検討した。
- (2) 合成音声の肉声感の確保：HMM 音声合成は、音声の分析合成（いわゆるボコーダ）に基づいた手法であるため、人間の実音声に比べた場合、肉声感という面では一定の限界があった。この主たる原因は、音声生成モデルにおける励振源モデル化が単純すぎたためである。この問題を解決するため、自動学習が可能という HMM 音声合成の特徴を活かしたまま、励振源モデルを高精度化する新手法を開発した。受聴試験により、有効性を確認した。
- (3) 話者性と発話様式の分離：感情表現などを伴った音声を収録・ラベリングすることができれば、元音声と同じ感情表現を伴った音声の合成を合成可能であることがこれまでにわかっているが、特定の感情表現を安定に表出しながら発声できるのは、一部のプロフェッショナルだけである。従って、通常音声から感情音声へのモデルの変形規則を別にモデル化することが必須となる。つまり、話者の声質と発話様式のモデルを分離して学習する必要がある。このことを可能とするため、様々な要因の加算的な構造を自動抽出し、モデル化可能な新しい手法を定式化した。
- (4) 多様性の拡大：更に表現力豊かな音声の合成を目指し、様々な検討を行った。特に、歌声に注目し、ピブラート等の

歌唱表現の個人性をモデル化・再現可能なシステムを構築した。また、本方式は言語依存部分が極めて少ないことを利点を活かし、システムを英語、中国語、ポルトガル語等の多言語へ拡張した。

これらの成果のまとめとして、計算機上にバーチャルなアニメーションキャラクターを実現し、音声情報案内端末に組み込んだ。また、成果の一部をオープンソースソフトウェアとして取りまとめ、インターネット上で公開した（図 2 他）。

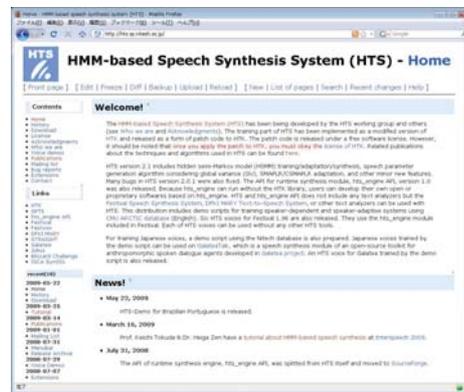


図 2 HMM 音声合成システム (HTS) 公開ウェブサイト (<http://hts.sp.nitech.ac.jp/>)

5. 主な発表論文等

〔雑誌論文〕(計 10 件)

- (1) Junichi Yamagishi, Takashi Nose, Heiga Zen, Zhen-Hua Ling, Tomoki Toda, Keiichi Tokuda, Simon King, Steve Renals, “A robust speaker-adaptive HMM-based text-to-speech synthesis,” IEEE Transactions on Audio, Speech and Language Processing (accepted). 査読有
- (2) Tomoki Toda, Alan W. Black, and Keiichi Tokuda, “Statistical mapping between articulatory movements and acoustic spectrum with a Gaussian mixture model,” Speech Communication, vol.50, no.3, pp.215-227, March 2008. 査読有

- (3) Tomoki Toda, Alan W. Black, and Keiichi Tokuda, "Voice conversion based on maximum likelihood estimation of speech parameter trajectory," *IEEE Transactions on Audio, Speech and Language Processing*, vol.15, no.8, pp.2222-2235, November 2007. 査読有
- (4) Rannierey Maia, Heiga Zen, Keiichi Tokuda, Tadashi Kitamura, and Fernando G. V. Resende, "An HMM-based Brazilian Portuguese speech synthesizer and its characteristics," *Journal of Communication and Information Systems*, vol.21, no.2, pp.58-71, Aug. 2006. 査読有
- (5) Heiga Zen, Keiichi Tokuda, Takashi Masuko, Takao Kobayashi, and Tadashi Kitamura, "Hidden Semi-Markov Model Based Speech Synthesis System," *IEICE Transactions on Information Systems*, vol.E90-D, no.5, pp.825-834, May 2007. 査読有
- (6) Tomoki Toda and Keiichi Tokuda, "Speech Parameter Generation Algorithm Considering Global Variance for HMM-Based Speech Synthesis," *IEICE Transactions on Information Systems*, vol.E90-D, no.5, pp.816-824, May 2007. 査読有
- (7) Heiga Zen, Tomoki Toda, Masaru Nakamura, and Keiichi Tokuda, "Details of Nitech HMM-based speech synthesis system for the Blizzard Challenge 2005," *IEICE Transactions on Information and Systems*, vol.E90-D, no.1, pp.325-333, Jan. 2007 (電子情報通信学会情報・システムソサイエティ論文賞および電気通信普及財団賞). 査読有
- (8) Heiga Zen, Keiichi Tokuda, Tadashi Kitamura, "Reformulating the HMM as a trajectory model by impos-

ing explicit relationships between static and dynamic feature vector sequences," *Computer Speech and Language*, vol.21, no.1, pp.153-173, Jan. 2007. 査読有

〔解説論文〕(計3件)

- (1) Heiga Zen, Keiichi Tokuda, Alan W. Black, "Statistical parametric speech synthesis," *Speech Communication* (accepted). 査読有
- (2) Heiga Zen, Keiichi Tokuda, "TechWare: HMM-Based Speech Synthesis Resources," *IEEE Signal Processing Magazine*, July 2009.
- (3) 徳田恵一, アランブラック "音声合成研究も協調と競争の時代に -The Blizzard Challenge-, " *日本音響学会誌*, vol.62, no.6, pp.466-472, June 2006. 査読有

〔国際会議論文〕(計43件)

- (1) Heiga Zen, Keiichiro Ooura, Takashi Nose, Junichi Yamagishi, Shinji Sako, Tomoki Toda, Takashi Masuko, Alan W. Black, Keiichi Tokuda, "Recent development of the HMM-based speech synthesis system (HTS)," 2009 APSIPA Annual Summit and Conference, Sapporo Convention Center, Sapporo, Japan, October 5 - 7, 2009 (to be published). 査読有
- (2) Kei Hashimoto, Heiga Zen, Yoshihiko Nankaku, Keiichi Tokuda, "Variational Bayesian method for HMM-based speech synthesis," 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), pp.4029-4032, Taipei, Taiwan, April 19-24, 2009. 査読有
- (3) Yi-Jian Wu, Keiichi Tokuda "Minimum generation error training by using original spectrum as reference for log spectral distortion measure,"

- 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), pp.4013-4016, Taipei, Taiwan, April 19-24, 2009. 査読有
- (4) Kaori Yutani, Yosuke Uto, Yoshihiko Nankaku, Akinobu Lee, Keiichi Tokuda, "Voice conversion based on simultaneous modeling of spectrum and F0," 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), pp.3897-3900, Taipei, Taiwan, April 19-24, 2009. 査読有
- (5) Keiichiro Oura, Yoshihiko Nankaku, Tomoki Toda, Keiichi Tokuda, Ranniery Maia, Shinsuke Sakai, Satoshi Nakamura, "Simultaneous Acoustic, Prosodic, and Phrasing Model Training for TTS Conversion Systems," International Symposium on Chinese Spoken Language Processing (ISCSLP2008), SPE1.1, pp.1-4, Kunming, China, December 16-19, 2008 (**Best Student Paper Award**). 査読有
- (6) Simon King, Keiichi Tokuda, Heiga Zen, Junichi Yamagishi, "Unsupervised Adaptation for HMM-Based Speech Synthesis," Interspeech 2008, pp.1869-1872, Brisbane, Australia, September 22-26, 2008. 査読有
- (7) Yoshihiko Nankaku, Kazuhiro Nakamura, Heiga Zen, Keiichi Tokuda, "Acoustic modeling with contextual additive structure for HMM-based speech recognition," 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp.4469-4472, Las Vegas, Nevada, U.S.A., March 30-April 4, 2008. 査読有
- (8) Tomoki Toda, Keiichi Tokuda, "Statistical approach to vocal tract transfer function estimation based on factor analyzed trajectory HMM," 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp.3925-3928, Las Vegas, Nevada, U.S.A., March 30-April 4, 2008. 査読有
- (9) Ranniery Maia, Tomoki Toda, Keiichi Tokuda, Shinsuke Sakai, Satoshi Nakamura, "ON the state definition for a trainable excitation model in HMM-based speech synthesis," 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp.3965-3968, Las Vegas, Nevada, U.S.A., March 30-April 4, 2008. 査読有
- (10) Yoshihiko Nankaku, Kenichi Nakamura, and Keiichi Tokuda, "Spectral conversion based on statistical models including time-frequency matching," Proc. of 6th ISCA Speech Synthesis Workshop, Bonn, Germany, August 22-24, 2007 (CD-ROM proceedings). 査読有
- (11) Heiga Zen, Yoshihiko Nankaku, Keiichi Tokuda, Tadashi Kitamura, "Speaker adaptation of trajectory HMMs using feature-space MLLR," Interspeech 2006 - ICLSP, pp.1141-1144, Pittsburgh, PA, Sept. 17-21, 2006. 査読有
- (12) Keiji Saino, Heiga Zen, Yoshihiko Nankaku, Akinobu Lee, Keiichi Tokuda, "HMM-based singing voice synthesis system," Interspeech 2006 - ICSLP, pp.2274-2277, Pittsburgh, PA, Sept. 17-21, 2006. 査読有
- (13) Maria Joa o Barros, Ranniery Maia, Keiichi Tokuda, Fernando Gil Resende, Diamantino Freitas, "HMM-based European Portuguese TTS system," INTERSPEECH 2005 - EU-

ROSPEECH, pp.2581-2584, Lisbon, Portugal, September 4-8, 2005. 査読有

〔招待講演〕(計5件)

- (1) Keiichi Tokuda, "An HMM-based approach to flexible speech synthesis," The 5th International Symposium on Chinese Spoken Language Processing (ISCSLP 2006), Kent Ridge, Singapore, December 13-16, 2006.
- (2) Keiichi Tokuda, "Hidden Markov model-based speech synthesis as a tool for constructing communicative spoken dialog systems," Proc. of The 4th Joint Meeting of The Acoustical Society of America and The Acoustical Society of Japan, Honolulu, Hawaii, 28 November-2 December 2006.
- (3) Keiichi Tokuda, "The trajectory HMM: Reformulating the HMM as a trajectory model," Trajectory Models for Speech Processing, Supported by EPSRC (The Engineering and Physical Sciences Research Council, U.K.), Edinburgh, U.K., August 31, 2005.

〔研究会発表〕(計11件)

- (1) 山田知彦, 武藤聡, 南角吉彦, 酒向慎司, 徳田恵一, "HMMに基づく歌声合成のためのピブラートモデル化," 情報処理学会研究報告「音楽情報科学(MUS)」, vol.2009-MUS-80, no.5, IPSJ-MUS09080005, May 14, 2009.

〔全国大会発表〕(計35件)

- (1) 油谷かおり, 南角吉彦, 戸田智基, 徳田恵一, "声質変換における時系列マッチングを含む統計モデルの拡張," 日本音響学会 2008 年秋季研究発表会, vol.I, 2-p-24, pp.411-412, September 2008 (日本音響学会ポスター賞).
- (2) 中村和寛, 全 炳河, 南角吉彦, 李 晃伸, 徳田恵一, "HMM 音声認識におけるコンテキストの加算的構造を考慮した音響モデリング," 日本音響学会 2007 年

春季研究発表会, vol.I, 1-P-13, pp.149-150, March 2007 (日本音響学会ポスター賞).

- (3) 才野慶二郎, 全 炳河, 南角吉彦, 李 晃伸, 徳田恵一, "因子分析に基づく固有声モデルを用いた HMM 音声合成," 日本音響学会 2007 年秋季研究発表会, vol.I, 3-4-8, pp.365-366, Sep. 2007.
- (4) 全 炳河, 南角吉彦, 徳田恵一, 北村正, "トラジェクトリ HMM の制約付き最尤線形回帰による話者適応," 日本音響学会 2005 年秋季研究発表会講演論文集, vol.I, 3-7-6, pp.113-114, September 2005 (日本音響学会粟屋潔学術奨励賞).

〔図書〕(計1件)

- (1) 徳田恵一, "4.3 HMM 音声合成における統一的な韻律の制御," 韻律と音声言語情報処理-アクセント・イントネーション・リズムの科学-, 広瀬啓吉 編著, 丸善, pp.118-127, Jan. 2006. (ISBN 978-4-621-07674-3)

6. 研究組織

(1) 研究代表者

徳田 恵一 (TOKUDA KEIICHI)
名古屋工業大学・大学院工学研究科・教授
研究者番号: 20217483

(2) 研究分担者

北村 正 (KITAMURA TADASHI)
名古屋工業大学・大学院工学研究科・教授
研究者番号: 60114865

南角 吉彦 (NANKAKU YOSHIIKO)
名古屋工業大学・大学院工学研究科・助教

研究者番号: 80397497

戸田 智基 (TODA TOMOKI)
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号: 90403328

(3) 連携研究者

ブラック アラン (BLACK ALAN)
カーネギーメロン大学・言語技術研究所・准教授