早稲田大学大学院理工学研究科

博 士 論 文 概 要

論 文 題 目

高品質音声合成のためのスペクトル包絡の推定 及び変換に関する研究

Studies on Spectral Envelope Estimation and Conversion for High Quality Speech Synthesis

申 請 者

氏 名	望月 亮
	Ryo Mochizuki

専攻・研究指導 (課程内のみ)	情報・ネットワーク専攻 知覚情報システム研究

2005 年 12 月

近年，コーパスベースの音声合成方式によって，音質の良い音声の合成が可能となった．特に大規模な音声コーパスを用い，韻律変換をまったく行わない波形接続合成方式では，読み上げ口調の音声に限れば自然発声と比較してほとんど遜色の無い合成が可能である．一方，音質の改善が進むにつれ，最近では感情や態度，話者性，発話口調を自由に制御するための技術が求められている．例えば音声合成を音声対話システムへ応用する場合，ユーザとシステムとの自然なやり取りを実現するためには単なる読み上げ口調ではなく，システムの発話意図や態度の多彩なパラ言語表現が必要不可欠である．

音声合成によって多彩な発話を実現する手段としては，(1)発話スタイルや話者ごとに音声を録音する，(2)少量データの学習によって適応する，等のアプローチが考えられる．前者は高品質を実現するという意味では有効であるが，現在の波形接続合成方式では録音やラベル情報の付加に膨大な人手の作業が発生するため，発話スタイルや話者ごとにデータベースを構築するのは現実的な方法とは言いがたい．一方，後者においては，現時点では十分な適応・変換方法が存在しないため変換処理を施すと音質劣化が目立ったり，変換自体が不十分だったりといった問題がある．しかし，この問題は今後検討が進むにつれて改善されることが期待される．

現在，高品質な合成を実現している波形接続合成方式は，合成時に元となる音声データを一切加工しない方式であり，このことが高品質を実現するカギとなっている．しかし，発話の多様化を目指すためには，少なくとも適応や変換処理が施せるレベルまで「音声信号処理」に踏み込んだ方式を採用する必要がある．PSOLA（Pitch Synchronous OverLap Add）法は波形接続合成より変換に対する自由度が高く，変換率が低い場合は高品質な韻律変換が可能であり，従来の線形予測を代表とするパラメトリックな方式よりも格段に音質が良いという長所を持つ．本研究では高品質な音声合成が期待できる PSOLA 法をベースに，音質の改善，及び多彩な発話表現の実現に必要不可欠なスペクトル包絡の抽出，補正，及び変換に関する要素技術を提案・検討する．

以下に本論文の構成を示す．

第１章では，本研究の目的と，その背景について述べる．また，関連する従来研究を紹介する．

第２章では，歪の少ないスペクトル包絡の推定を目的とし，ピッチ同期で短時間波形を抽出する方法について提案する．PSOLA 法は短時間窓を利用して基本周期の影響を含まない短時間波形を抽出し，この短時間波形を所望する基本周期で再配列することによって F0 変換を行うことができる．しかし，安定したピッチ同期分析が行えない場合，波形抽出位置がふらつき，韻律変換処理によって音質劣化を引き起こす．従来，短時間波形の抽出は基本周期の２倍の窓長を持つハニング窓で抽出するのが一般的であったが，先行研究ではどの位置を窓関数の中

心に設定するのが音質として良いのか明確な回答を持っていなかった．そこで変形自己相関によって線形予測残差波形のピーク抽出を行い，このピーク位置を短時間波形抽出の基準位置（ピッチマーク）として波形抽出する方法を提案する．また，提案方法によって決定したピッチマークを基準に，どの程度遅延した位置にスペクトル歪が最小となる波形抽出位置が存在するのか，音声信号モデルを用いて最適な波形抽出位置を実験的に探索する．ここで決定した波形抽出位置を用いて F0 変換音声を作成し，試聴評価によって提案する波形抽出位置の有効性を評価する．また，ピッチマーク決定の頑健性についても F0 変換音声の試聴実験によって評価する．

第３章では，ピッチ同期で抽出した短時間波形の低域におけるスペクトル包絡を，スペクトル傾斜と F0 変換率に応じて動的に再構築する方法を提案する．PSOLA 法によって韻律変換を行う場合，抽出した短時間波形をそのまま利用すると変換音声に著しい音質劣化が生じる場合がある．この音質劣化は原音声から抽出した短時間波形のスペクトル包絡が韻律変換後の環境に適合していないことが原因として考えられる．この原因の一つとして，PSOLA 法では元の F0 より低域において信頼できるスペクトル情報が得られないという問題が存在する．本来，周波数分析によって求められるスペクトルは，F0 の整数倍にあたる高調波のみで構成される線スペクトルとなるのが理想であるが，実際は短時間波形抽出に用いる窓関数の漏れが隣接する高調波間で重畳され，滑らかなスペクトル包絡が形成される．しかし F0 より低い帯域においては，F0 における窓関数の漏れの影響が観測されるのみで，正しいスペクトル包絡情報が観測できない．この低域スペクトルの問題により， F0 を低い方へ変換した場合に音質劣化が顕著になっているものと考えられる．そこで本研究では，F0 変換を行ってもスペクトル傾斜は保存されるという仮定に基づいて，動的に低域におけるスペクトル包絡を再構築し，音質劣化を軽減する方法を検討する．実際に提案方法によって生成した F0 変換音声の試聴評価を行い，F0 を低い方へ変換した場合に，提案方法の有効性を確認する．

第４章では，韻律特徴量を利用し，統計的な手法によってスペクトル特徴量をターゲットの環境にあったスペクトル特徴量へ変換する方法について提案する．音声合成によって多様な発声を実現するためには，音声収録時の発話から，ターゲットの発話へ変換するための適応技術が必要となる．話者の発話スタイルや話者性を決定づける要因としては，イントネーションやアクセントなど韻律的な特徴が重要であるが，それに劣らず，声質を決定するスペクトル包絡に関しても精度の良い変換が強く望まれる．この適応・変換を実現するために，今まで統計的な手法を用いた様々な方法が検討されているが，従来方法のほとんどの研究では変換元となるスペクトルとターゲットのスペクトルとの１対１の対応学習によって変換が行われていた．しかし，スペクトル変換を音声合成へ応用した場合を考

えると，変換関数の入力にはスペクトル以外にも韻律や音素系列などのコンテキスト情報を利用することが可能である．特にスペクトルは韻律特徴量との間にある程度の相関があるため，変換モデルに韻律情報を考慮することで変換精度の改善が期待できる．そこで本研究では PSOLA 法に基づくピッチ同期処理において，韻律情報を利用した統計的なスペクトル変換手法を提案し，話者変換実験によってその有効性を検証する．また，従来の変換モデルの学習に同一発話文コーパスを用いる方法が一般的であったが，非同一発話文コーパスを学習データに使って変換モデルを学習する方法についても検討する．

第 5 章では，PSOLA 法をベースにしたスペクトル包絡の推定方法，及び変換方法に関する取り組みに対して結論を述べ，今後の課題について議論する．

研 究 業 績

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
論文	○望月亮,大久保雅史,小林哲則, ``韻律情報を用いたスペクトル変換方式の検討,'' 電子情報通信学会誌,D-II, Vol.88, No.11, pp.2269-2276, Nov. 2005.
論文	○R.Mochizuki and T.Kobayashi, ``A low-band spectrum envelope reconstruction method for PSOLA-based F0 modification,'' IEICE Trans. INF.&SYST., Vol.87, D, No.10, pp.2426-2429, Oct. 2004.
論文	○R.Mochizuki and T.Kobayashi, ``A Low-band Spectrum Envelope Modeling For High Quality Pitch Modification,'' Proc. ICASSP2004, SP-P9.5 Vol.1 pp.645-648, May 2004.
論文	○望月亮,新居康彦,西村洋文,本多高, ``駆動点同期型ピッチ波形抽出法'', 音響学会誌 53 巻 10 号,pp.772-778, Oct. 1997.
講演	望月亮,小林哲則, ``GMM によるスペクトル変換モデルの非パラレルコーパスを用いた学習,''音響学会講演論文集 3-6-20, Sep. 2005.
講演	望月亮,小林哲則, ``ＰＳＯＬＡ法における音質改善のための低域スペクトル包絡の補正方法,'' 音響学会講演論文集 2-Q-4,pp.319-320, Sep. 2003.
講演	望月亮,本多高,新居康彦, ``駆動点同期型ピッチ波形抽出法の頑健性評価,''電子情報通信学会総合大会 D-141-1,pp.243, March 1997.
講演	望月亮,三浦成充,本多高,新居康彦,蓑輪利光, ``ピッチ波形抽出位置と一様ピッチ変換音声の音質との関係,'' 音響学会講演論文集 2-P-22,pp.345-346, Sep. 1995.
講演	望月亮,本多高,新居康彦,吉田博子,蓑輪利光, ``短区間変形自己相関係数を用いたピッチ波形抽出法の検討,'' 音響学会講演論文集 3-4-6,pp.285-286, March 1995.

研 究 業 績

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
その他 （論文）	大久保雅史,望月亮,小林哲則, ``心的態度表現に寄与する韻律／スペクトル包絡特徴の評価,'' 電子情報通信学会誌,D-II, Vol.88, No.2, pp.441-444, Feb. 2005.
その他 （論文）	望月亮,蓑輪利光, ``平滑化特徴ベクトルを用いたアクセント句の F0 パターン選択方法,'' 電子情報通信学会誌,D-II, Vol.87, No.2, pp.475-486, Feb. 2004.
その他 （論文）	T.Minowa, R.Mochizuki, and H.Nishimura, ``Improving the naturalness of synthetic speech by utilizing the prosody of natural speech,'' Proc. ICSLP2000, Vol.1 pp.609-612, Oct. 2000.
その他 （論文）	R.Mochizuki, Y.Arai and T.Honda, ``A study on the word synthesis method by using the VCV-balanced word database,'' J. Acoust. Soc. Jpn (E)21, pp.17-24, Jan. 2000.
その他 （論文）	R.Mochizuki, Y.Arai, and T.Honda, ``A study on the natural-sounding Japanese phonetic word synthesis by using the VCV-balanced word database that consists of the words uttered forcibly in two types of pitch accent,'' Proc. ICSLP98, Vol.5 pp.2011-2014, Dec. 1998.
その他 （論文）	Y.Arai, R.Mochizuki, and T.Honda, ``A Study on Natural-sounding Japanese Phonetic Word Synthesis Based on the Pitch Waveform Concatenation,'' Proc. ICA98, Vol.1 pp.267-268, June 1998.
その他 （論文）	Y.Arai, R.Mochizuki, H.Nishimura, and T.Honda, ``An Excitation Synchronous Pitch Waveform Extraction Method and Its Application to The VCV-Concatenation Synthesis of Japanese Spoken Words,'' Proc. ICSLP96, Vol.3 pp.1437-1440, Oct. 1996.
その他 （講演）	大久保雅史,望月亮,小林哲則, ``HMM 素片選択を用いた話者変換方式の検討,'' 信学技報 SP2004-139,pp.13-18, Jan. 2005.
その他 （講演）	大久保雅史,望月亮,小林哲則, ``波形重畳型音声合成におけるＨＭＭを用いた素片選択,'' 音響学会講演論文集 3-2-14,pp.343-344, Sep. 2004.
その他 （講演）	大久保雅史,望月亮,小林哲則, ``心的態度表現における韻律的／分節的特徴の影響,'' 音響学会講演論文集 2-P-23,pp.375-376, March 2004.

研 究 業 績

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
その他 （講演）	大久保雅史,望月亮,蓑輪利光,小林哲則, ``波形重畳型音声合成における心的態度の再現性評価,'' 情報科学技術フォーラム FIT2003, Vol.2, pp.285-286, Sep. 2003.
その他 （講演）	望月亮,蓑輪利光, ``属性ベクトルを用いたF0パタン選択方法の検討,'' 音響学会講演論文集 3-10-21,pp.369-370, Sep. 2002.
その他 （講演）	蓑輪利光,望月亮, ``テキスト音声合成に対する大規模コンテキストの利用に関する一考察,'' 信学技報 SP2002-24,pp.1-6, May 2002.
その他 （講演）	蓑輪利光,望月亮, ``コーパスサイズに最適な韻律制御方法の検討,'' 音響学会講演論文集 2-10-19,pp301-302, March 2002.
その他 （講演）	望月亮,蓑輪利光, ``波形重畳型の合成方式に用いる代表ピッチ波形生成方法の検討,'' 音響学会講演論文集 2-1-4,pp.181-182, Sep. 2000.
その他 （講演）	蓑輪利光,望月亮,西村洋文,釜井孝浩, ``韻律のベクトルを利用した音声合成方式,'' 信学技報 SP2000-4,pp.25-31, May 2000.
その他 （講演）	望月亮,西村洋文,蓑輪利光,新居康彦, ``波形接続合成に用いるＶＣＶ素片データベースの構築方法,'' 信学技報 SP99-1,pp.1-8, May 1999.
その他 （講演）	望月亮,西村洋文,蓑輪利光,新居康彦, ``ターゲットピッチパターンに着目したＶＣＶ素片データベースの検討,'' 音響学会講演論文集 2-3-3,pp.231-232, March 1999.
その他 （講演）	望月亮,西村洋文,蓑輪利光,釜井孝浩, ``韻律ベクトルを用いた高音質規則合成方式,'' 音響学会講演論文集 1-3-22,pp.227-228, Sep.-Oct. 1999.
	その他（講演） 12件（1995-1998）