HMM 音声合成を用いた表情豊かな対話音声の合成におけるコンテキストの検討*

☆佐藤航, 森大毅 (宇都宮大)

1 はじめに

人間と機械との円滑なコミュニケーションを実現する対話システムには、話者交代などの様々な要因を考慮した人間味のある音声合成システムが必要になる。特に私達人間同士の対話における発話の末尾の変動には多くの要因が関係しており、句境界ピッチ運動などの音声の意図や態度に関する韻律的特徴は、HMM 音声合成において自然な音声の合成に寄与することが示されている [1]。そして同じように、発話の話者交代もしくは継続という情報も、発話の韻律的特徴との関係が示されている [2, 3]。特に対話音声における句末音調の上昇調や上昇下降調などは、対話における話者交代の有無と非常に密接な関係にあると考えられるため、これらの情報を利用することで、HMM 音声合成での発話末尾の f_0 の変化を制御することができると考えられる。

本論文では、表情豊かな対話音声について HMM 音声合成を行った場合に、対話の句末音調や話者の交代・継続を記述したラベルをコンテキストに追加することで、合成音声の自然性がどのように変化するのかを明らかにすることを目的とする。

2 句末音調、話者交代コンテキストの追加

HMM 音声合成において、コンテキストはスペクトル、f_0、継続長といった音響的特徴を表す変動要因として音素単位にラベル付けされ、これらのコンテキストの組み合わせに対してコンテキスト依存 HMM が学習される。本論文では、文献 [1, 6, 7] を参考に、HMM 音声合成に使用するコンテキストとして、表 1 に示した 7 種類の情報を用いた。通常の読み上げ音声の合成に広く用いられているものとして、音素、モーラ、アクセント、発話長の 4 つを用いている。また、拡張コンテキストとして、パラ言語情報ラベルを付与している [5]。

本論文では、以上のコンテキストの他に、拡張コンテキストとして句末音調と話者交代に関する情報を付与した 2 つのコンテキストを追加した。句末音調コンテキストでは、ラベルとして X-JToBI[4] によって定義された tone ラベルに属する句末音調ラベルのうち、L%,L%H%,L%LH%,L%HL%の 4 つを用いた。また、話者交代のコンテキストには、各発話末の最終モーラのコンテキストラベルが対応する発話において、発話終了後に話者交代をする場合には”T” を、発話話者が同じで発話が継続する場合には”F” を付与している。これらを追加したコンテキストラベル例を図 1 に示す。

Table 1 コンテキストのカテゴリ

音素	{ 先行 当該 後続 } 音素の種類
モーラ	アクセント句内のモーラ位置
アクセント	{ 先行 当該 後続 } アクセント句の長さ、アクセント型、位置、ポーズの有無
発話長	文長
パラ言語情報	「快-不快」、「覚醒-睡眠」
句末音調	L%,H%,L%HL%,L%H%
話者交代	話者の交代・継続

本論文では、音声合成に使用する音声コーパスとして、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 (UUDB)[5] を用いた。UUDB には親しい友人同士による日常的な会話音声が収録されている。特に話者 FTS に関しては、感情がよく表出されており、表情豊かな音声が収録されている。本論文では UUDB の XML データ及び X-JToBI に基づき作成した韻律ラベルから言語・韻律情報を抽出し、それをコンテキストとして HMM 音声合成に適用している。

3 主観評価実験

3.1 実験条件

作成したコンテキストを用いて HMM 音声合成を行う。学習には UUDB の対話セッション C002 から C006 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としてサンプリング周波数 16 kHz の音声信号から、分析周期 5 ms、分析窓長 25 ms 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。f_0 パラメータは対数基本周波数とし、特徴ベクトルはこれらのパラメータにそれぞれの Δ、ΔΔ パラメータを加えた 78 次元のベクトルとした。

評価実験はヘッドフォンによる両耳聴取により行った。被験者は 9 人の男性とし、音声の研究室に所属する大学院生 7 人及び大学生 2 人の合計 9 人とした。

3.2 自然性評価実験

作成したコンテキストを適用した合成音声の自然性を評価するために、AB 法による主観評価実験を行った。合成する発話は UUDB の対話セッション C001

*Investigation of contexual factors for the synthesis of expressive dialogue speech using HMM-based speech synthesis. by SATO, Wataru, MORI, Hiroki (Utsunomiya University)

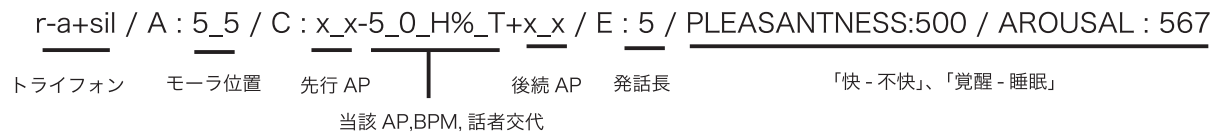

Fig. 1　新要素を追加したコンテキストラベル例

Table 2　主観評価実験の結果

	提案法	従来法
選択割合 (%)	55.2	44.8

に収録されている発話を対象とし、さらに実験に使用する発話は、合成した音声から相槌などを除いた 63 発話とした。合成音声はベースコンテキスト (音素、モーラ、アクセント、発話長) 及びパラ言語情報と、句末音調と話者交代のコンテキストを適用した音声、及び、従来のベースコンテキスト及びパラ言語情報のみによって合成された音声 (従来法) の 2 種類を用意した。

実験の結果を表 2 に示す。2 種類の音声を対比較した結果、新しいコンテキストラベルを適用した合成音声の方が、従来法の合成音声と比べて自然性が高いということが分かった。このことから、提案法による合成音声の自然性の向上を確認できた。

自然性向上の要因を確かめるために、自然性の高いと選択された音声と、提案法で付与された句末音調ラベルの関係を調査した。その結果、自然性の向上した発話の多くは、提案法において L%HL%のラベルを付与した合成音声であった。この例として、発話「でもビーではあー今日はあったかいなって言ってるから」の f_0 軌跡を図 2 に示す。図の上段は従来法における合成音声の f_0 軌跡であり、下段は提案法における合成音声の f_0 軌跡である。図より、従来法では句末音調が表現できていないのに対し、提案法では L%HL%が表現できているのが分かる。L%HL%は対話音声の特徴的な句末音調であるため、この句末音調が表現可能となったことが、自然性向上の一因になっていると考えられる。

一方、元々L%HL%の句末音調が表現されていた音声も存在し、従来法の方が音声の自然性が高いと評価されたものもあった。従来法のコンテキストで表現可能だったものについては、提案法の新しいコンテキストによって音声の自然性が損なわれてしまったことも考えられる。

4　おわりに

本論文では、句末音調や話者交代の情報を HMM 音声合成のコンテキストに追加することが、表情豊かな対話音声の合成において合成音声の自然性に与える影響を評価した。主観評価実験による自然性評価の結果、句末音調や話者交代のコンテキストを用いることで、表情豊かな対話音声の合成において、より自然性の高い音声が合成できることが分かった。

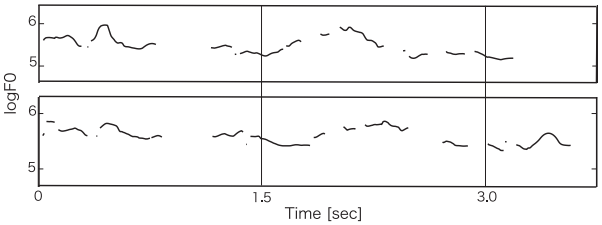

Fig. 2　「でもビーではあー今日はあったかいなって言ってるから」

今後は、今回自然性が低下してしまった音声について、コンテキストによる影響をより詳しく調査することをを考えている。

参考文献

[1] 郡山知樹 他, “HMM に基づく対話音声合成における多様な韻律生成のためのコンテクストの拡張,” 電子情報通信学会論文誌, Vol.J95-D, No.3, pp.597–607, 2012.

[2] 木村太郎 他, “F0 モデルを用いた日本語対話における韻律と話者交代の分析,” 電子情報通信学会技術研究報告, SP2007-75, pp.25–30, 2007.

[3] 千田みのり 他, “話者交代に対するプロソティ情報を利用した聞き手による予測認知の検討,” 人工知能学会研究会資料, SIG-SLUD-A803-11, pp.57–62, 2009.

[4] 前川喜久雄 他, “自発音声の韻律ラベリングスキーム,” 電子情報通信学会技術研究報告, SP2001-106, pp.25–30, 2001.

[5] H. Mori et al., “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.

[6] 吉岡元貴 他, “HMM 音声合成における韻律の変動要因の検討,” 信学技報, SP2001-80, pp.51–56, 2001.

[7] 横溝秀始 他, “HMM 音声合成における韻律コンテキストの評価,” 日本音響学会講演論文集, pp.403–404, 2010.