講演番号： 183

音声入力による環境音合成

D-14

Effective Sound Synthesis from Speech Input

坂本　昌寛†　　小坂　直敏†

Masahiro SAKAMOTO†　　Naotoshi OSAKA†

† 東京電機大学大学院未来科学研究科

†Graduate School of Future Science, Graduate School of Tokyo Denki University

1. はじめに

映像作品やゲーム，Web 上のコンテンツなどのマルチメディアコンテンツにおいて，効果音は重要な表現要素である．環境音データベースから，製作者の時間長，抑揚，あるいは聞こえなどの条件をすべて満たす音を直接得ることは一般に困難である．そこで，本稿では，コンテンツ制作者の所望する効果音を得るために，その特徴を音声で模擬（声真似）し，これを入力として，所望の音色を持つ環境音を変形合成する問題を検討する．

所望の効果音をターゲットとし，効果音の音色を保ったまま，音声入力により欠落している部分を補って合成する方法は，環境音を音声に置き換えて考えると，声質変換技術である．そこで，ここでは声質変換技術を効果音へ適用することにより，データベースにない音を合成する．下図 1 に犬の鳴き声を例として，変換システムの概念図を示す．

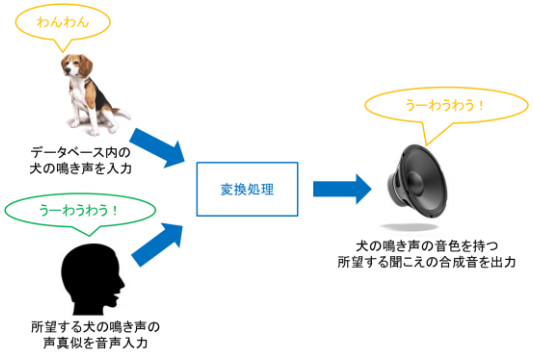

図 1　変換システムの概念

ユーザは，犬の「うーわうわうっ」と聞こえる音を所望する場合，始めに所望する犬の鳴き声を模擬した音声を入力する．次にデータベース内の音色が同じで，これに近い犬の鳴き声の波形を入力する．それぞれの波形に対して変換処理を施すことにより，ユーザが所望する犬の鳴き声の音色を持ち「うーわうわうっ」と聞こえる合成音を得られる．

2. GMM に基づく声質変換法

声質変換技術は，音声合成の主要テーマのひとつである．以下では，戸田らの GMM(Gaussian Mixture Model)に基づく声質変換手法[1]を応用し，音声を入力として環境音合成について提案する．

本稿では，ソースを音声，ターゲットを環境音とし，音声特徴量を変換する．音声と環境音からスペクトル包絡を抽出し，GMM でモデル化する．従来の GMM に基づく声質変換法では，変換関数の学習にパラレルデータを用いている．環境音の聞こえとユーザが聞こえを模擬した声真似を同文章発話とみなし，それらをパラレルデータとして学習に用いる．また，従来の声質変換法では，スペクトル包絡のみを変換の対象としていたが，本方式ではターゲットが環境音となるため，音源については正弦波モデル[2]を用いてモデル化する．

3. 変換結果

音声と環境音からスペクトル包絡を算出し，GMM でモデル化した結果を図 2,図 3 に示す．GMM のクラス数は 64 とし，正規分布のパラメータの推定には EM アルゴリズムを使用した．ここで使用した環境音は雷の音であり，話者の音声はそれを模擬した声真似音声である．

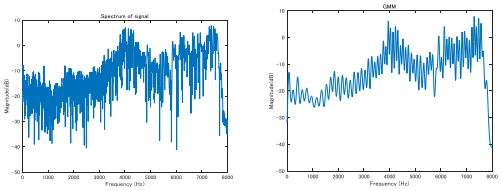

図 2　話者のスペクトル（左）と GMM（右）

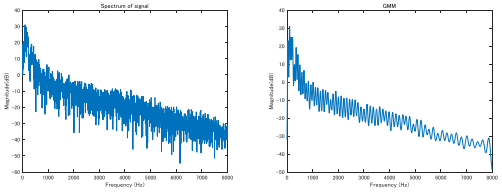

図 3　環境音(雷)のスペクトル（左）と GMM（右）

話者のスペクトル距離は 6.30[dB]，環境音のスペクトル距離は 6.60[dB]であった．

4. まとめと今後の課題

声質変換法を人物と環境音に適応し，音声入力からの環境音合成手法を提案した．声質変換法を適応させることでデータベースにない音を、合成によって得ることができた．ただし、音質及び変換精度のために提案手法の再検討が必要である．

参考文献

[1]　戸田智基ほか ”周波数軸伸縮を用いた混合正規分布モデルに基づく声質変換法” 信学誌　D-Ⅱ　No.10 pp.2181-2189 (2001-10-01) ．

[2]　R. J. McAulay and T. F. Quatieri: Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE Trans. On ASSP, vol.ASSP-34,No4,Aug.1986.