ARX音声分析合成法を用いた合成音声の声質制御*

○粕谷英樹，森大毅（宇都宮大），木戸博（東北工大）

1　はじめに

利用目的、利用環境、利用方法，利用者の要求に応じて柔軟に対応できる音声合成器が望まれている[1]。対応策の一つに，“ 声質制御の可能な音声合成器 ” がある。声質は韻質以外の音声の聴覚的特質であり，特定個人内の豊かな表現力を支える特徴であるとともに，個人性に寄与する特徴でもある。両者はまた密接に関係している。例えば，落胆した発話（個人内表現）では多くの場合気息性(breathy)発声になるが[2]，それはまた，女声（個人性）を特徴づける声質でもある[3]。

声質の記述・表現については，音声生成過程や音声学の伝統的な枠組みを考慮しながら行う方法[4]や日常生活で使われる語をよりどころにして，設定する方法がある[5]が，一般利用者への音声合成装置の普及を考えると，後者の方が適切であろう。

本報告では，主として個人性の付与を目的にした声質の表現語[5]，コーパスベースの ARX 音声合成システム[6]，表現語に基づいて行う合成音声の声質制御法[7]などについて述べる。

2　ARX 音声合成における声質制御

2.1 ARX 音声分析合成

音声生成のソース・フィルタ理論[8]を近似的な実現モデルである，ARX 音声生成モデル (Autoregressive with exogenous input speech production model) を Fig. 1 に示す。図で，$U(z)$は有声音源（放射特性を含んだ微分声門音源波）であり，数理モデルとしては Rosenberg-Klatt モデル[3]を用いる。$B(z)/A(z)$は声道フィルタ，$E(z)$は白色雑音，$S(z)$は音声信号である。無声音の生成では図中の X はゼロであり，AR (autoregressive)モデルになる。

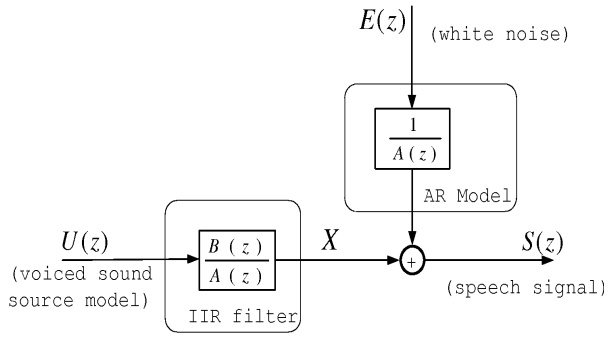

Fig. 1　ARX speech production model.

ARX モデルの合成パラメータは，以下のように音源パラメータと声道パラメータに分けられる。

音源パラメータ：基本周期（周波数），有声音源振幅，声門開放率，音源スペクトル傾斜調整パラメータ，非周期境界周波数（調波成分が優勢な低周波数帯域と不規則雑音成分が優勢な高周波数帯域を分ける境界周波数），雑音源振幅。

声道パラメータ：フォルマント，アンチフォルマント。

一定の録音条件を満たす音声データが与えられると，これらのパラメータを自動的に抽出するアルゴリズムも開発されている[9,10]。

2.2 コーパスベース ARX 合成システム

音声コーパスの ARX 分析パラメータの選択・接続と韻律（基本周波数，音声セグメントの継続時間）制御モデルに基づいた音声合成システムについては，既に報告した[6]。ARX 音声合成システムは，原理的にはフォルマント合成システム[3]であるが，他の類似のシステムとの大きな違いは，われわれの合成システムのフォルマント合成回路が母音・子音とも共通にカスケードになっているという点である。

2.3 声質制御

声質の制御機能をもつコーパスベースの ARX 音声合成形 TTS システムを Fig.2 に示す。この構成では，コーパスの話者（ここでは女性）の ARX パラメータとその統計モデルに基づく ARX 音声合成パラメータをいったん生成した後に，パラメータ変換を行って声質制御を行うという，合成パラメータ生成と声質制御とが分離したシステムになっている。しかし，将来研

* Voice quality control of synthetic speech based on the ARX speech analysis/synthesis method, by KASUYA, Hideki, MORI, Hiroki (Utsunomiya University), and KIDO, Hiroshi (Tohoku Institute of Technology).

究が進めば，図の韻律パラメータ制御部と分節特徴接続・制御部の中に組み込まれるべきものと考えている。

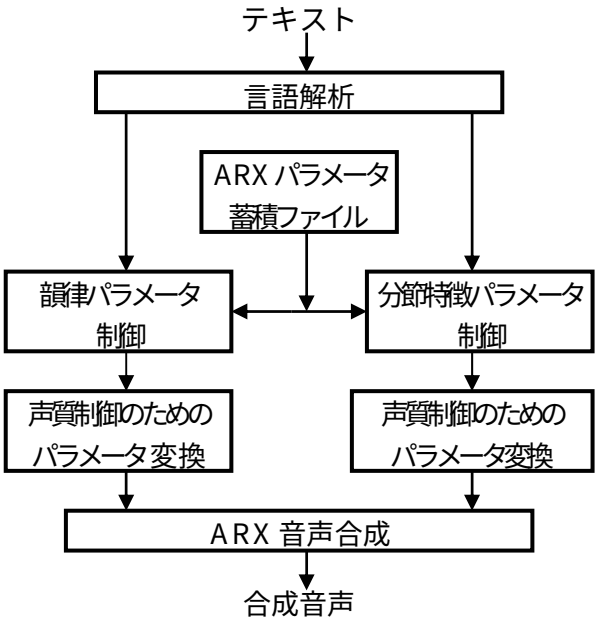

Fig.2 ARX-model based TTS system being able to control voice quality of synthetic speech

2.4 声質表現語

われわれは膨大なアンケートならびに文献資料調査に基づいて声質表現語を収集した上で，それらを音声研究者でない一般市民が自分自身の声質を評価したデータに基づいて，統計的手法を用いて，８語対と対をもたない１語の声質表現語を抽出した[5]。それらは, 低い/高い, 男性的/女性的，かすれた/澄んだ，落ち着きなる/落ち着きのない，迫力のある/弱々しい，若い/老けた, 張りのある/張りのない, 太い/細い, の８語対と鼻声（閉鼻性と開鼻性）という１語である。

2.5 声質表現語の音響関連量

音源パラメータの中で，基本周波数(F0)に関連にしたパラメータとして, F0 ベースライン周波数，F0 レンジ，音源スペクトル傾斜調整パラメータ，声道パラメータとしてフォルマント周波数，その他の韻律制御パラメータとして話速, 合計５種類の音響パラメータを取り上げた。それらを３段階に変えて，合計 243 種類の合成音声を用いて，聴覚実験を行った[7]。実験では各合成音声を先に述べた声質表現語８語対についてそれぞれ７段階で評価させた。例えば，男性的/女性的という表現語対については, 非常に男性的(1)，かなり男性的(2)，やや男性的(3)，どちらとも言えない(4)，やや女性的(5)，かなり女性的(6)，非常に女性的(7)，の７段階である。得られた評価結果を用いて，各表現語対を目的変数，5 種類の音響パラメータを説明変数として重回帰分析を行ったところ，以下のような結果が得られた。低い/高いについては，F0 レンジ，F0 ベースライン，フォルマント周波数がいずれも大きい（あるいは高い）方が“ より高い ”声質になる。男性的/女性的では，F0 レンジ，F0 ベースライン，フォルマント周波数が大きい（あるいは高い）方がより “ 女性的 ”，落ち着きのあり/なしについては, 話速, F0 レンジ，F0 ベースライン, フォルマント周波数が大きい（あるいは高い）ほど，落ち着きがなくなる，太い/細いについては，F0 レンジ，F0 ベースライン，フォルマント周波数が大きい（あるいは高い）ほど，細くなる，という比較的明確な関係が分かった。しかしその他の声質表現語は２種類の F0 パラメータと弱い関係があるものの，取り上げた音響パラメータだけでは不十分であった。

3　おわりに

統計的に抽出した声質表現語に従って，合成音声の声質を制御する ARX 音声合成法について述べた。今後はその他の ARX パラメータと声質との関係も含めて，より精密な検討を行いたい。

謝辞

関連する研究課題でこれまで協力して頂いた宇都宮大学粕谷研究室の卒業生に感謝する。本研究の一部は科研費(16300061)によった。

参考文献

[1] 粕谷，音響学会誌，48（1），46-51，1992.
[2] Kasuya, H., Maekawa, K. & Kiritani, S. , Prco. ICPhS, 2505-2512, 1999.
[3] Klatt,D.& Klatt,L., JASA,87(2),820-857,1990.
[4] Laver, J., “The phonetic description of voice quality,” Cambridge Univ. Press, 1980.
[5] 木戸,粕谷,音響学会誌, 55(6),405-411, 1999.
[6] Mori, H., *et al*., Proc. ICSLP 2002, Denver, 4: 2365- 2368, 2002.
[7] 川股，他，音講論（秋），289-290, 2004.
[8] Fant,G., “Acoustic theory of speech production,” The Hague:Mouton, 1960.
[9] 大塚，粕谷,音響学会誌,58(7), 386-397, 2002.
[10] Ohtsuka, T. & Kasuya, H., Proc. Eurospeech, Aalborg, 3, 2267-2270, 2001.