

Title	声道の共振特性を考慮した歌声合成システムの構築に関する研究
Author(s)	長田, 和也
Citation	
Issue Date	2015-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12634
Rights	
Description	Supervisor:赤木正人, 情報科学研究科, 修士

声道の共振特性を考慮した歌声合成システムの構築に関する研究

長田 和也 (1310049)

北陸先端科学技術大学院大学 情報科学研究科

2015 年 2 月 12 日

キーワード: 歌声合成, ソースフィルタモデル, 声道の共振特性, フォルマント分析, フォルマント制御.

歌声には話し声と比較して多様な発声法が存在する. 人は声の生成系を利用することによってそれらの多様な表現を行うことが出来る. しかし, その仕組みは完全には解明されていない. それらの多様な表現がどのような要因によって行われているのかを解明したいと考え, 歌声についての研究を行う. 歌声についての研究は古くから行われており, 特にヨーロッパを中心としてオペラ歌唱についての研究が盛んに行われている. Farinelli という歌声合成システムはフランス国立音響音楽研究所 (IRCAM) において作成されたものであり, 去勢された男性オペラ歌手 (カストラート) の歌声を映画「カストラート」において再現している.

歌声についての研究方法は, 大きく分けて 2 つあり, 声の生成系を MRI 等をもちいて物理的に計測し, 計測結果を元に物理モデルを制御する手法が存在するが, 物理量変化を計測するために大掛かりな装置が必要であったり, 被験者の負担も大きい. そこで, 本研究はもう一方の手法である, 収録した歌声を分析し, その分析結果をもとに音響的特徴量を制御する手法を使用する.

歌声合成システムの手法として, VOCALOID に代表される波形接続方式や隠れマルコフモデル (HMM) 合成方式といった手法が存在するが, これらの手法はより自然な歌声を合成することに重点がおかれており, 人がどの部分を制御することで歌声の表現を行っているかといった知見を得ることは出来ない. そこで, 人の声帯や声道を模擬したモデルとしてソースフィルタモデルを利用した手法がある. ソースフィルタモデルを利用した先行研究として, 歌声らしさの特徴量としてヴィブラートと呼ばれる, 基本周波数 (F_0) と振幅包絡が 4~6 Hz の変調周波数で振動する現象やプレパレーションやオーバーシュートといった歌声らしさに影響を与える F_0 の変化と歌唱フォルマントを適用することによって歌声合成を行う研究が行われている. また, ARX モデルと Liljencrants-Fant (LF) モデルを組み合わせた ARX-LF モデルを利用し, 声帯モデルである LF モデルを制御するこ

とにより Vocal fry, Modal, Falsetto といった 3 つの声区を表現することを可能としている。しかし、これらの研究では声道の共振特性についてはまだ十分に検討されていない。

そこで、本研究では、先行研究において十分に検討されていない声道の共振特性を考慮することで、より自然な歌声を合成する歌声合成システムを提案する。本研究の目的は、声道の共振特性を考慮して歌声合成を行うことにより、より自然な歌声を合成することを目的とする。また、本研究によって自然な歌声の合成が可能になるだけでなく、声道の共振特性の解明につながる可能性がある。提案する歌声合成システムでは合成器として、Klatt Formant Synthesizer を使用する。この合成器を使用する理由は各フォルマントのパラメータを独立して入力として制御が出来る合成器であり、共鳴器を独立して制御することで声道の共振特性を表現するには適していると考えたからである。また声帯音源の作成には歌声らしさに関係する音響的特徴量の分析によって構築された F0 制御モデルと声区表現を可能にした LF モデル制御を使用する。

はじめに、声道の共振特性を考慮したフォルマント制御モデルを構築するために分析対象の歌声を収録した。収録音声は声楽経験のある男性 1 名の歌声と話声を計 60 音声収録した。収録時に被験者の首に EGG を装着し、声門の最大開口部、最大閉口部を記録できるようにした。次に、収録した歌声を音高ごとに分割した 108 音声に対してフォルマント分析を行った。分析音のサンプリング周波数は 8kHz であり、分析にはハミング窓と LPC を用い LPC 次数は 10 とした。そして、F0 変化によるフォルマントの変化を見るために 10 次のメディアンフィルタを使用して 10 近傍のサンプルに対する中央値を求めた。その結果、F0 の変化によるフォルマントへの影響は第 1 フォルマント (F1)、第 2 フォルマント (F2) とともに大きく分けて 4 つの変化の傾向が見られた。1 つ目の区間は F0 が 300Hz 以下の音域で、これは modal で歌われている音域である。この区間では、F0 の変化 1Hz あたり F1 は 0.32Hz 減少しており、F2 は 0.41Hz 増加している。2 つ目の区間は F0 が 300Hz から 336.64Hz の音域である。この区間は modal から falsetto への変化が起きている音域である。F0 の変化 1Hz あたり F1 は 3.32Hz 減少しており、F2 は 4.23Hz 減少している。3 つ目の区間は F0 が 336.64Hz から 454.64Hz の音域である。この区間は falsetto のうち F0 が低い音域である。F0 の変化 1Hz あたり F1 は 0.15Hz 増加しており、F2 は 0.82Hz 減少している。そして、4 つ目の区間は F0 が 454.64Hz 以上の音域である。この区間は falsetto のうち F0 が高い音域である。F1 は F0 に合わせて増加しており、F0 との差の中央値は 15Hz である。また、F2 は第二高調波に合わせて増加しており、第二高調波との差の中央値は 35Hz である。これらの分析結果をもとに 4 つの区間に分けたフォルマント制御モデルを構築した。

そして、構築したフォルマント制御モデルの有効性を評価するために被験者 6 名に対して評価実験を行った。評価実験では先行研究を元に合成した 3 種類の歌声合成音と提案手法により合成した歌声合成音 1 種類、計 4 種類の歌声合成音を使用した。合成音のメロディには分析結果により得られた 4 つの区間の音がそれぞれ 1 つずつ、計 4 音を低い方から並べたものである。4 種類の音声のうち 2 つの音声の組み合わせ 12 種類をそれぞれ 10

回ずつランダムに呈示した．評価基準は二つの音を順に呈示し，modal から falsetto への変化が自然な歌声を自然な歌声として評価を行い，他の音声より自然であると判断された割合を比較した．その結果1名については提案手法がもっとも自然であると評価されたが，総合すると先行研究による手法より不自然であるという結果となった．そこで，先行研究による歌声合成音と比較することでその原因を検討した．その結果，提案手法による歌声合成音は先行研究による歌声合成音と比較して2つ目と3つ目の音の振幅が極端に小さく，また，3つ目の音から4つ目の音への振幅の変化が極端に大きくなっていた．そこで，提案手法による歌声合成音の振幅を評価実験において最も自然であると評価された音声の振幅包絡線に合わせて制御した音声を作成した．また，評価実験の結果を見ると modal から falsetto への変化のない歌声合成音に対して最も modal から falsetto への変化が自然であると評価する被験者が半数見られた．そこで，評価基準を modal と falsetto という主観的なものではなく，収録した歌声に近いかどうかという基準で自然さを判断するように変更した．以上のことを踏まえて，振幅制御を行った音声を加えた5種類の音声に対して評価実験を再度行った．その結果，制御を行った音声は制御前と比較して他の音より自然であるという評価が大きく向上し，以前の評価実験で最も自然であると評価された歌声合成音と同等に自然であると評価された． n