

TR-IT-0338

正弦波モデルに立脚した物理的特徴量と音色
情報あるいは感情との関係について

**On the relation between physical
features based on a sinusoidal model
and timbral information or emotion**

小坂 直敏 & ニックキャンベル

February 14th 2000

ABSTRACT

現在の音声合成システム CHATR に、声質制御あるいは感情を付与することを目標とし、その初期検討として、声質の分析を行う。この研究の分析ツールとして、音/音声分析および合成システム「おっきんしゃい」を用いる。この研究にシステムを応用するための問題点とその解決の方針について述べた。また、現在の進捗状況について報告する。

This research's final goal is to implement a voice quality control and emotinal speech synthesis function into a speech synthesis system "CHATR." The first step is an analysis of voice quality. As an analysis tool sound/speech analysis and synthesis system named "Otkinshi" is used. This article describes the problems lying when applied to this research and its solution. Moreover, The present research situation is also reported.

©ATR Interpreting Telecommunications
Research Laboratoriess.

©ATR 音声翻訳通信研究所

1 はじめに

CHATR による音声合成方式は、音声翻訳通信の大変重要な役割を担っている。その膨大かつ豊富なデータベースにより合成される音声の品質は向上している。しかし、CHATR の基本方式の中に信号処理方式を入れず、データベースの中からラベルに最も合った音素データを選択しこれらを接続する、という基本的な枠組みのみでは、柔軟な音声合成が行えない欠点がある。すなわち、

1. 合成のために採用する音声データベースに依存するため、そのデータベース全体の持つ音声の特徴がそのまま反映される。
2. 素片の最適な選択のために、音素の音韻環境のみならず、前後のピッチができるだけ合う必要がある。しかし、ピッチまで含めると実際には、あらゆるピッチのデータベース化は不可能である。
3. さらに声質要因もある。一般に同一話者の同一データベースの中でも声質は異なるものである。したがって、現行の CHATR では合成音の前後で声質が異なることも予想される。

1. は、具体的には沈んだトーンの朗読音声データベースを採用すると、いかなるテキストの合成をさせようと、このデータベース全体に特徴的な情報はそのまま踏襲される。2. では、音韻環境が適当であっても接続部でのピッチが合致していないために音質が劣化することがある。以上の欠点を補うために、以下の検討が必要である。

1. 接続後の合成品質を向上するための音データベースからの合成音を加工するための信号処理方式。
2. 音声データベースに横たわる感情の特徴抽出
3. 音声の個人性の特徴抽出
4. 声質の特徴抽出

1. では、音声データベースで音韻環境、ピッチ、声質、感情など多くの要因のすべてについて、望むパラメータの値を持つようにするのは、その組み合わせの数の多さから不可能である。しかし、合成音声の品質を向上するためには、何らかの信号処理を施すことが必要である。

2.-4. の感情、個人性、声質はそれぞれ合成に重要な要因であり、また相互に関連している要因である。

感情は怒り、悲哀、喜び、などあるが、抽象的であり、広範である。たとえば喜びの表現には数多くの言い回しが考えられる。また、感情は音声データベース全体に関わる要因である。

個人性は、ある発声者がその発声者であるがために生ずる印象である。これも必ずしも物理的、知覚的なものに対応せず、ある発声者の音声である、という事実からそのもとでのすべての音声がある特定の個人の声とラベル付けされる。

これに対し、声質は、ハスキー、バジー、などの音声の物理的特徴と近いものと、沈んだ、すっとんきょうな声、黄色い声、押し殺した声、朗々とした声など感情に近いものもある。

これまでの韻律研究では、ピッチ、ラウドネスといった一次元尺度を用いて検討をすることが多かったが、ここでは主にスペクトル情報に基づいて声質の特徴を表現することを試みる。

2 研究の概要

以下では、音声の声質の分析、あるいは声質を制御した音声合成を目的とした研究の概要を述べる。

2.1 音声の声質の特徴抽出

声質の違いにより、ピッチのみならず、スペクトル特徴がどう変わるかを分析する。分析にあたり、音を正弦波モデル表現により調べていく。この分析方法は、音楽分析などでは一般的であるが、信号を正弦波の和として分解する。このため、調波構造とその他という分類で見ることができる。また、それら正弦波の位相情報も含めて扱うため、従来の多くの音声分析に見られるようなパワースペクトルのみに依存する分析とは異なり、きめ細かい分析が可能となる。これまでに、正弦波モデルを用いた音声の分析として、[13], [14] などがある。これらの検討の延長上で調波、あるいは残差のふるまいなどを中心に見ていく。

2.2 音声の声質の制御（付与、変換、加工）

音楽では、クロスシンセシス、あるいはハイブリッド合成（混成音）などと呼ばれる合成がある。これは、ひとつの音／音声を合成するのに、複数個の音／音声から別々に韻律、音韻、声質などを取り出して再合成するものである。この結果喋るクラリネット、別の人のイントネーションの付与などが可能となる。モーフィングもハイブリッド合成の一つである。このモーフィングを使うことにより、声質変換や韻律の多様な制御に応用できる。声質変換はターゲットの声質に変換するのみならず、その中間の音声に変えることができる。また、音声の接続に際しモーフィング技術を適用することでより自然な音声が可能である。また、単に自然な音声のみならず、これらの制御の幅を大きくすることにより、強調された音声、非日常的な音声（演劇の場のような）、あるいは芸術的な音声などへも応用できる。

2.3 声質のデータベース

これらの研究を進めるにあたり、音声データベースが必要であり、男声発話（声優）4 テキストでさまざまな発話のデータベース（文献 [1]）あるいは ATR データベースのうちの飯田作成の感情データベースを利用する。

3 音合成システム「おっきんしゃい」の応用

コンピュータ音楽の音合成システムはもともと音声合成とは大変関係深い。代表的なシステムには例えば [2],[3] などがある。これらは実時間音合成の枠組みである。一方実時間合成でなくとも、よりきめ細かい合成を行えるものとして、筆者はこれまで、コンピュータ音楽ための楽音合成ソフト「おっきんしゃい」“O^tkinshi”(O^to to Koega isshoni naru shisutemu) を検討してきた [7],[8],[9],[10],[11]。これら文献 [8]-[11] は、バージョン I として NeXT コンピュータの上で Objective C で書かれていた。このシステムは第 1 義的には作曲家の制作用、あるいは、マルチメディアコンテンツ制作時の音響合成を目的としている。一方、ここで対象とする音楽は非常に広く、声では歌声のほかに文章、詩歌の朗読も含まれる。

そのため、音声合成にも応用可能である。特に前章で述べた声質あるいは韻律の研究を行うためのツール、あるいはこれらの制御方式を実装するためのシステムとして位置付けることができる。類似したシステムとしては、これまで [4],[5],[6] などがある。本研究では、音声の声質の分析を行うツールとして、またその結果明らかになった事実を反映するためのシステムとして位置づける。

「おっきんしゃい」は現在新バージョンとして、PC の Windows 上で動作している。そのシステムコンセプトは引き継がれ、さらに、正弦波モデルに基づく分析機能と加工機能、および新しい GUI(Graphical User Interface) を実装している。このシステムの特徴を次節で述べる。また、詳細は付録に記す。これは文献 [12] を添付したものである。

3.1 “O^tkinshi” の特徴

システムの概要を以下に述べる。目的：音声／音を正弦波モデル表現し、この上でのエディットを行う。この機能の中の最も発展した上位機能としてモーフィングがある。モーフィングは二つの音を与えて、その間の音色を連続的に合成する技術の総称である。現在までに、信号モデルによる方法と物理モデルによる方法がある。ここでは特に信号モデルによる方法を用いていく [15],[16],[17],[18]。これらの機能は研究面で要素技術を反映したもので、他のツールと差別化できる。一方、音／音声波形加工ツールとしては、こうした本ツールの特徴となる機能のほか一般的な加工機能もあった方が使い勝手がよい。そこで、波形エディタとしての基本機能も搭載している。また、UNIX 上で動作する研究用ソフトと、一般ユーザ用の Windows 用ソフトの 2 種類ある。

- UNIX 用の特徴
 - ・ 研究用に筆者が作成したもので言語は C
 - ・ コマンドベース
 - ・ 研究用ツールのため、常に研究の最新状況を反映。
- Windows 用
 - ・ 外注による。言語は C++。
 - ・ GUI ベース
 - ・ UNIX ベースのものより機能の搭載は遅れている。

4 研究進捗状況

研究スタート時点で Windows 上でフルート音から女声へのモーフィングなど一部のモーフィング音デモができていた。これから上記研究へ発展させるために達成すべき点は以下である。

1. 現在までの UNIX 上のソフトはシステムを作ることを意識されていない。そのため、Windows 上のソフトを書くために、合成部のエンジンは書き直している。これは C++ で書かれている。その結果、ここで作成あるいは変更されたソフトが UNIX 上で機能することができず、ソフトの作成効率が悪い。
2. 声質を分析するためのツール、尺度というものがまだない。そのため声質に影響を与えるものの定量化、尺度化などの検討が必要となる。
3. モーフィングは音楽の枠組みでやってきたため、母音と母音などの事例はあるが、文章音声と文章音声などの例ではまだ達成されていない。音声用のチューニングが必要である。
4. 主に調波を対象としてモーフィングができているが、残差も対象とし音声品質の良い方式を作る必要がある。
5. データベースについても吟味する必要がある。声優による文音声は spontaneous な音声とはいえない、しかし、多くの明確な特徴が期待できる。一方、spontaneous なデータベース、特に感情データベースは全体としては特徴が知覚できるものの、文単位で考えた場合、特に一文一文に特徴が出ているわけではないため、感情などの特徴のラベリングも重要な課題となる。

これらの課題の中で、今期は 1,2 を中心に検討を進めてきた。2. の定量化、尺度化についてはまだ達成されていないが、以下の要因を初期的に検討している。

1. 定常母音の周波数対応での雑音性
2. 母音などの調波の現れ方（強度、持続時間）
3. 調波の強度の相対関係

すなわち、声質の物理的特徴として、調波を中心にして見ていく。今後は 1. を達成し、ソフトの流通をよく研究効率をあげる。一方上記の項目の定量化、尺度化をはかって声質分析を進めていく。

謝辞 本研究を進めるにあたり研究討論、ツールの使用でお世話になりました第 2 研究室のみなさんに感謝します。

References

- [1] 小坂直敏, 「文音声の発話方法と韻律情報との関係」, 音講論 H2 年度春季 2-4-11, 平 2. 3
- [2] Eric Lindemann, Francois Dechelle, et al., "The Architecture of the IRCAM Musical Workstation", Computer Music Journal, Vol.15, No.3, pp41-pp49, Fall 1991.
- [3] Miller Puckette, "Combining Event and Signal Processing in the MAX Graphical Programming Environment", Computer Music Journal, Vol.15, No.3, pp68-pp77, Fall 1991.
- [4] 信号処理ソフトウェア "ILS", 音響学会誌 45, pp. 129-130, 1989.
- [5] 天白成一, 平原達也, "UNIX 音声研究用ツール -SpeechTools-", 音響学会講演論文集 3-P-20, 1991,10.
- [6] 河原英紀, "音声知覚過程研究支援環境のユーザインタフェース", 音響学会 聴覚研究階資料、H87-21,1987
O^tkinshi コンピュータ音楽関係
- [7] 小坂直敏, 「NeXT 上のサウンドエディタ "O^tkinshi" の開発」, 情処学会 音情研 2 年度 12 月研究会 平 2. 12
- [8] 小坂直敏, 「音素材開発システム "O^tkinshi" について」 情処学会 音情研 3 年度夏シンポジウム予稿, 平 3. 8
- [9] 小坂直敏, 「コンピュータミュージックシステム展示 "O^tkinshi"」, 情処学会 音情研, 3 年度 12 月研究会予稿 平 3. 12.
- [10] 小坂直敏, 「楽音生成・演奏システムについて」, 信学技報 EA92-31, pp. 27-32, 1992. 6.
- [11] Naotoshi Osaka, "O^tkinshi": A sound generation and performance system," Proc. of ICMC 92 pp. 406-407. San Jose, California, 1992.
- [12] 小坂直敏, 「Windows 上の音合成システム「おっきんしゃい」の検討」, 情報処理学会研究報告, 99-MUS-31, pp.19-26, 1999.8. 正弦波モデルによる分析
- [13] 小坂直敏, 「Sinusoidal model を用いた定常母音の声質分析」, 音講論 6 年度秋季 1-5-6, 平 6. 10.
- [14] Naotoshi Osaka, "An analysis of voice quality using sinusoidal model," Proc. of ICSLP 94, pp. 1647-1650, Yokohama, Japan, 1994.
- [15] Naotoshi Osaaka, "Timbre interpolation of sounds using a sinusoidal model," ICMC 95 Proceeding, Banff, Canada, Sep. 1995.

- [16] 小坂直敏, 「Sinusoidal model を用いた母音の声質補間」, 音講論 H7 年度秋季 2-1-10, 平 7. 9.
- [17] 小坂直敏, 「Sinusoidal model による音色補間」, 情処学会 音情研 13-9, 7 年度 12 月研究会予稿 pp. 45-50, 平 7. 12.
- [18] 小坂直敏, 「Sinusoidal model による音色の補間」, 信学技報 SP95-130 pp. 9-16, 平 7. 12.