

音の時間情報変動による聴覚情報の変化 およびその知覚に関する研究

課題番号：13610079

平成 13－15 年度科学研究費補助金（基盤研究(C)(2)）

研究成果報告書

平成 16 年 5 月

研究代表者 赤木 正人
北陸先端科学技術大学院大学情報科学研究科

目次

1. はしがき		
1.1 研究組織	2
1.2 研究経費	2
1.3 研究発表	2
2. 概要	4
2.1 研究の位置付け	5
2.1.1 研究の背景	5
2.1.2 研究の目的	5
2.2 成果の概要	6
2.2.1 歌声に含まれる時間情報変動に起因する声質の変化	6
2.2.2 両耳間時間差による音源方向知覚	8
3. 研究成果（収録論文目次）	10

時間変動知覚

[1] Akagi, M. (2002). "Perception of fundamental frequency fluctuation," HEA-02-003-IP, Forum Acousticum Sevilla 2002 (Invited).	11
--	-------	----

歌声

[2] Saitou, T., Unoki, M., and Akagi, M. (2002). "Extraction of F0 dynamic characteristics and development of F0 control model in singing voice," Proc. ICAD2002, Kyoto.	17
[3] Saitou, T., Unoki, M., and Akagi, M. (2004). "Development of the F0 control method for singing-voices synthesis," Proc. SP2004, Nara, 491-494.	21
[4] Saitou, T., Unoki, M., and Akagi, M. (2004). "Control methods of acoustic parameters for singing-voice synthesis," Proc. ICA2004, 501-504.	25
[5] 齋藤、鵜木、赤木(2001). "歌声における F0 動的変動成分の抽出と F0 制御モデル"、音響学会聴覚研究会資料、H-2001-92	29
[6] 赤木、清水(2003). "STRAIGHT を用いた話声からの歌声合成"、電子情報通信学会技術報告、SP2003-37.	37
[7] 齋藤、鵜木、赤木(2003). "歌声の F0 制御モデルにおけるパラメータ決定に関する考察"、音響学会聴覚研究会資料、H-2003-111		

	43
[8] 辻, 赤木(2004). “歌声らしさの要因とそれに関連する音響特徴量の検討”, 音響学会聴覚研究会資料、H-2004-8.	49
方向知覚		
[9] Ito, K. and Akagi, M. (2003). “Study on improving regularity of neural phase locking in single neuron of AVCN via computational model,” Proc. ISH2003, 77-83.	55
[10] 伊藤, 赤木(2002). ” 神経細胞における時間情報伝達の仕組みに関する一考察 : SBC の位相同期特性と entrainment 特性について”, 音響学会聴覚研究会資料、H-2002-02.	62
[11] 伊藤, 赤木(2004). “頭内定位実験による聴覚系の位相情報伝達に関する検討”, 音響学会聴覚研究会資料、H-2004-1.	70

1. はしがき

1.1 研究組織

研究代表者：赤木正人

(北陸先端科学技術大学院大学情報科学研究科教授)

1.2 研究経費

平成 13 年度	2,100 千円
平成 14 年度	700 千円
平成 15 年度	500 千円
計	3,300 千円

1.3 研究発表

1.3.1 学会誌等

方向知覚

[1] Ito, K. and Akagi, M. (2004). "Study on improving regularity of neural phase locking in single neuron of AVCN via computational model," in Auditory Signal Processing: physiology, psychoacoustics, and models, Eds. Pressnitzer, D., de Cheveigné, A., and McAdams, S., Springer Verlag, NY, 2004 (in printing).

1.3.2 招待講演

時間変動知覚

[2] Akagi, M. (2002). "Perception of fundamental frequency fluctuation," HEA-02-003-IP, Forum Acousticum Sevilla 2002 (Invited).

1.3.3 国際会議

歌声

[3] Unoki, M., Saitou, T., and Akagi, M. (2002). "Effect of F0 fluctuations and development of F0 control model in singing voice perception," NATO Advanced Study Institute 2002 Dynamics of Speech Production and Perception.

[4] Saitou, T., Unoki, M., and Akagi, M. (2002). "Extraction of F0 dynamic characteristics and development of F0 control model in singing voice," Proc. ICAD2002, Kyoto.

[5] Saitou, T., Unoki, M., and Akagi, M. (2004). "Development of the F0 control method for singing-voices synthesis," Proc. SP2004, Nara, 491-494.

[6] Saitou, T., Unoki, M., and Akagi, M. (2004). "Control methods of acoustic parameters for singing-voice synthesis," Proc. ICA2004, 501-504.

方向知覚

[7] Ito, K. and Akagi, M. (2003). "Study on improving regularity of neural phase locking in single neuron of AVCN via computational model," Proc. ISH2003, 77-83.

1.3.4 口頭発表（国内研究会）

歌声

- [8] 齋藤, 鶴木, 赤木(2001). ” 歌声における F0 動的変動成分の抽出と F0 制御モデル”, 音響学会聴覚研究会資料, H-2001-92
- [9] 赤木, 清水(2003). “STRAIGHT を用いた話声からの歌声合成”, 電子情報通信学会技術報告, SP2003-37.
- [10] 齋藤, 鶴木, 赤木(2003). “歌声の F0 制御モデルにおけるパラメータ決定に関する考察”, 音響学会聴覚研究会資料, H-2003-111
- [11] 辻, 赤木(2004). “歌声らしさの要因とそれに関連する音響特徴量の検討”, 音響学会聴覚研究会資料, H-2004-8.

方向知覚

- [12] 伊藤, 赤木(2002). ” 神経細胞における時間情報伝達の仕組みに関する一考察 : SBC の位相同期特性と entrainment 特性について”, 音響学会聴覚研究会資料, H-2002-02.
- [13] 伊藤, 赤木(2004). “頭内定位実験による聴覚系の位相情報伝達に関する検討”, 音響学会聴覚研究会資料, H-2004-1.

1.3.5 口頭発表（国内大会）

歌声

- [14] 齋藤, 鶴木, 赤木(2002). ” 歌声における F0 動的変動成分の抽出と F0 制御モデル”, 平成 14 年春季音響学会講演論文, 1-9-19.
- [15] 齋藤, 赤木(2003). “歌声の F0 制御モデルにおける制御パラメータの改良”, 平成 15 年秋季音響学会講演論文, 2-Q-21.
- [16] 齋藤, 鶴木, 赤木(2004). “歌声の F0 制御モデルにおけるパラメータ値の変動が歌声の自然性に与える影響の調査”, 平成 16 年春季音響学会講演論文, 2-7-3.
- [17] 辻, 赤木(2004). “歌声らしさに影響を及ぼす音響的特徴の分析”, 平成 16 年春季音響学会講演論文, 2-9-7.

方向知覚

- [18] 伊藤, 赤木(2001). ” 神経細胞への多重入力の時間的分布と位相同期特性の尖鋭化との関係について”, 平成 13 年秋季音響学会講演論文, 1-3-14.
- [19] 伊藤, 赤木(2002). ” entrainment 特性に基づく SBC 細胞の位相情報伝達について”, 平成 14 年春季音響学会講演論文, 2-9-11.
- [20] 伊藤, 石崎, 赤木(2002). ” 頭内定位における位相多義性の影響と位相情報伝達との関係”, 平成 14 年秋季音響学会講演論文, 1-2-4.
- [21] 伊藤, 赤木(2003). “頭内定位実験による聴覚系の位相情報伝達に関する検討”, 平成 15 年秋季音響学会講演論文, 3-7-6.

2. 概要

2.1 研究の位置付け

2.1.1 研究の背景

聴覚は時間的な情報を取り扱う、とよく言われる。

一般に、聴覚情報は1次音源の位置で音源操作の時間順序として与えられ、”事象”という形をとる。このため、聴覚にとって時間情報は重要な役割を演じており、また、聴覚自身も時間情報の変化に対して非常に鋭敏な反応を示す。たとえば、ジッタなどの時間方向への揺れが含まれる信号が聴覚に与えられた場合、聴覚はその揺れを音質の変化として知覚することが知られている。また、ジッタが大きくなれば音質のみではなく音そのものの揺れの感覚が出てくるとも事実である。ところが、時間方向への揺れが含まれる信号について、その知覚特性とか聴覚機構についての定量的な知見は未だにわずかである。時間情報の変化に対する知覚、またこれに関する聴覚機構についての研究の必要性は、上に示した聴覚情報の性質上、従来から認識されていた。しかし、これらの研究が、今までのような定常音を扱ってきた枠組みで研究できる範囲から大きく離れたものとなっていたため、今まで多くは行なわれてこなかったのである。

2.1.2 研究の目的

本研究の究極の目的は、波形に含まれる時間情報の変化、特にマイクロからマクロまでの揺れに対する人間の聴覚特性を調べ、聴覚系の生理学的機能との対応関係を得る、また、得られた知見をもとに人間工学的な応用をはかることにある。そこで、本研究では、これらの目的への第一歩として、時間情報の変化が人間に与える影響について、以下の項目について研究を行った。

1. 歌声に含まれる時間情報変動に起因する声質の変化
2. 両耳間時間差による音源方向知覚：ゆれを含む聴神経発火パターンからいかにして正確な時間差を検出しているか

2.2 成果の概要

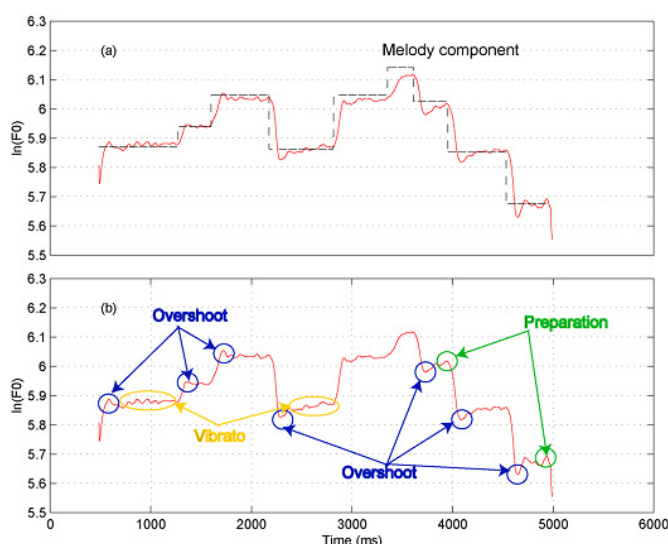
2.2.1 歌声に含まれる時間情報変動に起因する声質の変化

歌声は話声にはない特有の時間的変動を伴う音響的特徴を持つ。例えば、表現法の一つであるヴィブラートは基本周波数の変動のみならずスペクトル構造の変動をも伴う。そこで本研究では、このような時間的変動が、“歌声らしさ”という心理的特徴の重要な要因であるとの仮定を設け、歌声らしい自然な歌声を合成するために必要な時間的変動を伴う音響的特徴の特定および利用法の研究を行った。

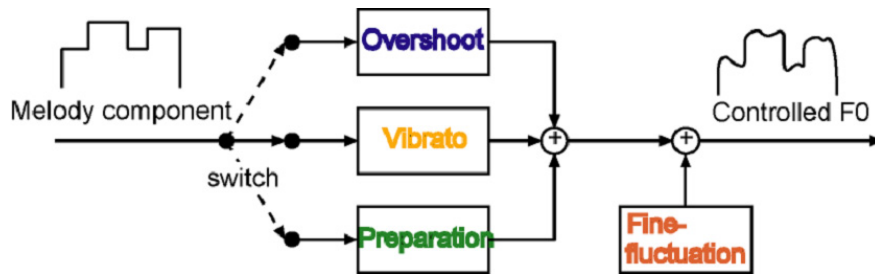
自然な歌声合成を行うには、歌声の基本周波数(F0)の特徴的な動的変動およびスペクトルを制御できるモデルが必要である。本研究では、(1) F0 コントロールモデルの提案および精緻化、また、(2) 歌声中の「ゆれ」と「歌声らしさ」の印象に関連する研究を行った。

F0 コントロールモデルの提案

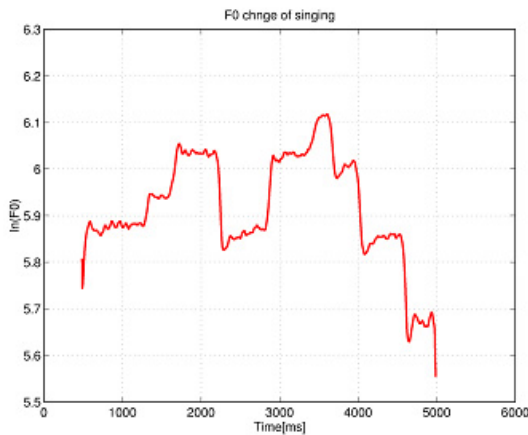
(1)については、様々なジャンルの歌声の F0 にフィッティングを行い、歌声の F0 における特徴的な動的変動の抽出を行った。そして、聴取実験によって、それら動的変動の歌声知覚に与える影響を明らかにした。また、動的変動成分を付加できる歌声の F0 制御モデルを提案した。その結果、F0 動的変動成分であるオーバーシュート・アンダーシュート、ヴィブラート、予備的変動成分が歌声知覚に大きな影響を与えていることが明らかとなった。さらに、モデルから生成された F0 を用いて、自然性の高い歌声合成を行うことができる可能性を確認した。このモデルにより生成した歌声を用いて、F0 の揺れに含まれる周波数成分とその大きさについて検知限・許容限を測定し、相互の関連および他の物理量との関係について考察した。



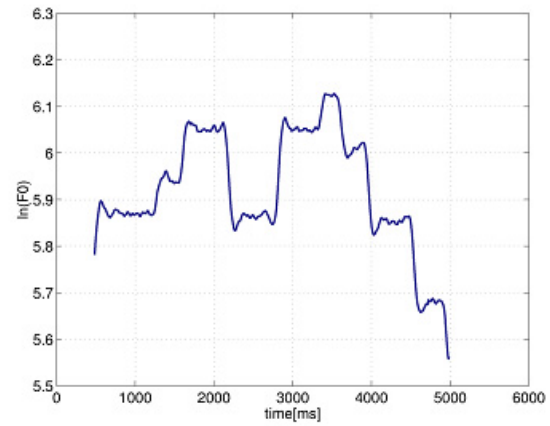
図：歌声の基本周波数抽出結果（上）、F0 動的変動成分（下）



図：F0 制御モデル



Real F0 contour

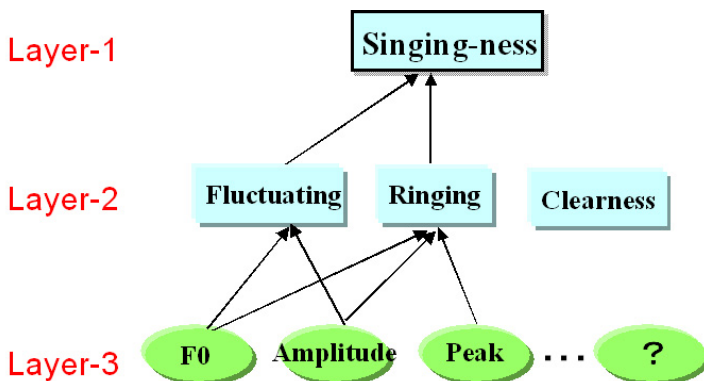


Synthesized by the model

図：F0 制御モデルの出力。(左) 歌声からの抽出 F0, (右) モデルから生成された F0

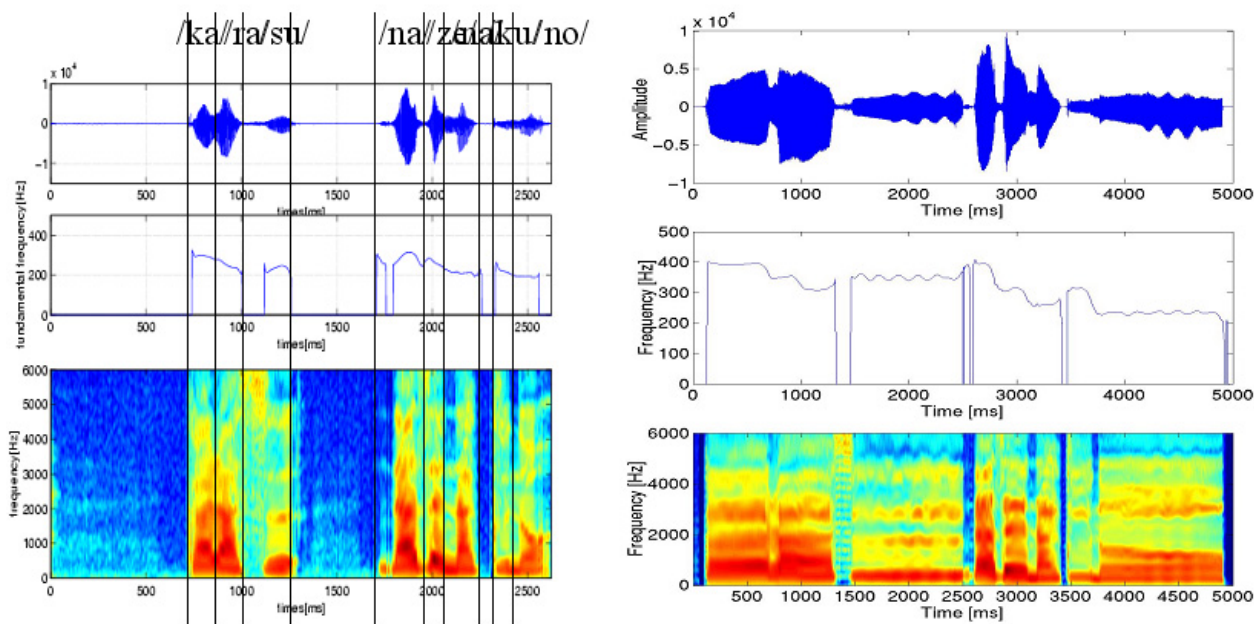
歌声らしさに関連する物理量

(2)については、歌声らしさ知覚を説明するための三層構造モデルを提案し、歌声らしさに関連する基本的心理量と基本的心理量に関連する物理量の検討を行った。この結果、ゆれは「歌声らしさ」の印象に最も関係する基本的心理量であること、ゆれの知覚に関連する物理量は F0 の変動とこれと同位相で変化するスペクトルの AM, FM 変調であることが明らかとなった。



図：三層構造モデル

これらの結果を歌声合成システムに組み込み、楽譜と歌詞の朗読音声から歌声の合成を試みた結果、自然な歌声に匹敵する合成歌声が得られた。



図：（左）歌詞朗読音声のスペクトル（童謡「七つの子」/karasunazekuno/），（右）合成された歌声スペクトル。

2.2.2 両耳間時間差による音源方向知覚：ゆれを含む聴神経発火パターンからいかにして正確な時間差を検出しているか

ヒトによる音源定位は、低周波数帯域の場合、主に左右耳の時間差（ITD）により成されており、900Hz 純音の定位能力は時間換算で10-20マイクロ秒である。一方、生体での時間情報の媒体である神経インパルスは、聴覚の知覚が達成する時間的な精密さに比べると、一桁あるいは二桁以上の時間的な揺らぎを伴っていたり、冗長であったりする信号である。聴覚系は、このような生体信号の持つ時間的な変動を如何にして克服し、どのようにロバストな時間情報処理のシステムを構築しているのだろうか。

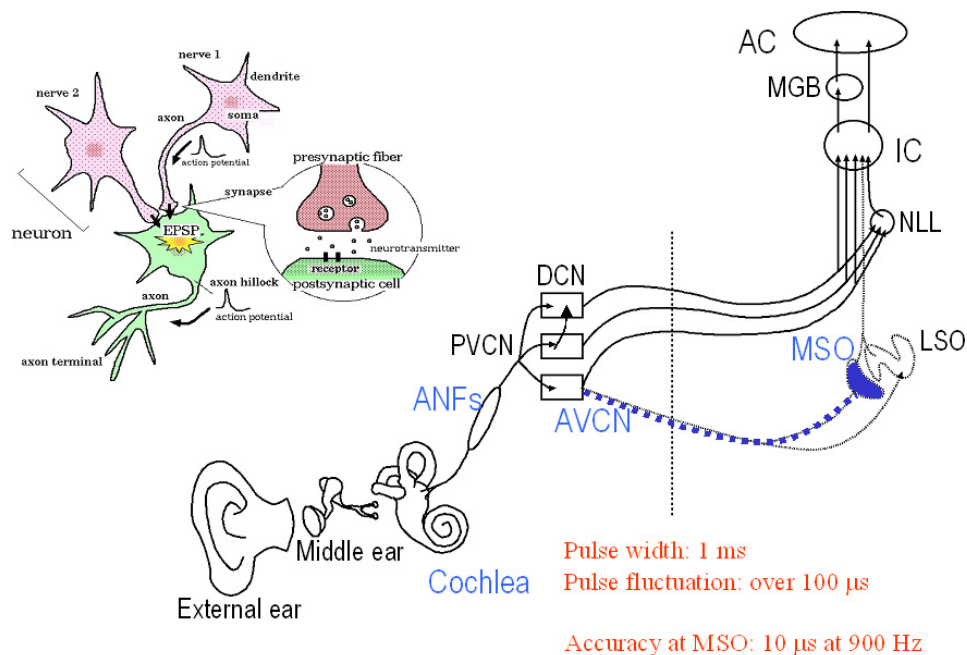
本研究では、生体信号の特性を背景に達成される脳の情報処理の戦略を理解するために、感覚器官の一つである初期聴覚系の神経細胞レベルで行われる時間情報処理のメカニズム、特に、時間的な揺らぎや冗長性を伴う神経インパルスに基づく時間情報が、音源方向知覚のために、どのように伝達され、整形され、利用され、そして喪失されるのかを、近年の生理学・解剖学的な知見に基づく聴覚モデルを用いたシミュレーションを通して、数理的に検討した。

聴覚やそれを含む脳は、膨大なネットワークからなる高度の情報システムであり、この極めて複雑なシステムにおける情報の表現や計算のアルゴリズムを解明するには、現在の生理学的あるいは心理学的実験的方法のみでは難しい。生理学的実験で判ることの多くは、個々の神経

細胞に関するもので、多くの神経細胞によって作られる複雑な神経回路の機能原理を容易に掴めない。一方、知覚現象を対象とする心理実験のみで、神経系での情報処理仕組みを説明することも難しい。要素レベルの知見を総合して一つの概念にまとめたり、既知の現象からブラックボックス的な神経系のシステムを逆推定したりするための理論的な手助けとして、計算機によるモデル研究が重要な役割を担うと考えられる。本研究では、計算機上の聴覚モデルを積極的に活用し、次の知見を得ている。

聴覚系は、神経インパルスの jitter や休止による影響を多重な入力機構によって緩和したり（整形）、逆に jitter を積極的に利用して一見ぼやけた情報群の中から本当に必要な情報を抽出する（利用）という、生体信号が元来持っている時間的なゆらぎや冗長性を前提とした巧みな時間情報処理のメカニズムを構築している。

本研究は、初期聴覚系の神経細胞レベルで行われる時間情報処理の過程を、一貫して総合的に検討した点で新規である。そして、このような聴覚系での時間情報処理のメカニズムを理解することは、脳の時間情報処理の戦略を理解することにつながり、延いては、ノイズにロバストなシステムの構築など、生命工学や情報工学での多くの応用が期待できる。



図：聴覚経路。External ear（外耳）に入った音響信号情報は、Cochlea（蝸牛）で神経パルスに変換され、左上に示したシナプス結合を介して神経パルスとして中枢へ運ばれる。時間情報を用いた方向知覚はMSOで処理されるとされており、その精度は900 Hzの音響信号で1度（時間差にして10 μs）である。ところが、神経パルスの幅は約1 ms（時間差の1000倍）、パルスの時間揺らぎは数百μsである。このような信号からいかにして10 μsの精度を得るのだろうか？

3. 研究成果