

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月29日現在

機関番号：15501

研究種目：挑戦的萌芽研究

研究期間：2009～2011

課題番号：21650039

研究課題名（和文） ヒト蝸牛内聴覚系の精密分子モデルとその音声認識他への応用

研究課題名（英文） Precise Molecular Model of Human Cochlea System and Its Application to Speech Recognition

研究代表者

内野 英治 (UCHINO EIJI)

山口大学・大学院理工学研究科・教授

研究者番号：30168710

研究成果の概要（和文）：

本研究の目的は、聴覚末梢系の数理モデルを作成し、それを利用して高性能な音声処理を実現することである。本研究で構築した聴覚末梢系モデルの様々な機能を検証するために、モデルから生成されたパルス時系列パターンの解析を行った。特に、ヒトの聴覚情報処理機構が有する周波数選択性、2音抑制、マスキングなどの特性について再現できることを確認し、構築したモデルの妥当性を明らかにした。次に、聴覚末梢系モデルから生成されたパルス時系列パターンを用いて雑音環境下での話者識別を行った。本提案手法と従来の周波数分析との比較検討を行い、提案したモデルの有効性を確認した。

研究成果の概要（英文）：

The purpose of this study is to achieve a high performance speech signal processing based on a mathematical model of the human peripheral auditory system. In order to verify the validity of the model, the multi-dimensional pulse signals generated by the model are analyzed. As a result, it is confirmed that the peripheral auditory model has the characteristics of the human auditory sense such as frequency selectivity, two-tone suppression, masking, and so on. Furthermore, the model is applied to speaker identification in noisy environment. Experimental results show that the proposed method based on the peripheral auditory model has a better performance than the conventional methods in case where the noise level is high.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,100,000	0	1,100,000
2010年度	1,000,000	0	1,000,000
2011年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,000,000	270,000	3,270,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：情報センシング、聴覚システム

1. 研究開始当初の背景

近年の音声認識の技術は、ヒトの聴覚機能とは異なり、隠れマルコフモデルのような強力な統計モデルを用いて、あらかじめ多数の音声データを学習し、認識するものがほとんどである。このような方法は、コンピュータの性能が近年著しく向上した結果、実用可能となったものであり、カーナビゲーションや携帯電話などの音声入力などに大いに役立っている。しかし、それは、静かな屋内で、話者が一人という理想的な環境下での話である。

統計モデルに基づく音声認識システムには、騒音下における発話や複数話者の発話に対応できないなど、ヒトの聴覚系のような柔軟性はないといえる。したがって、従来の音声認識システムでは、よりヒトに近いインターフェースとしてさらに進化するためには限界がある。

近年、ヒトの聴覚系の生物学的機構の解明が進んでいる。例えば、外耳、中耳の伝達特性を考慮したスペクトル情報である **Excitation pattern** の導出や、蝸牛内における基底膜のフィルタ特性をフィルタバンクで近似する方法や、有毛細胞における神経伝達物質の入出力の定式化、聴神経におけるパルス発生の機構の研究、パルスを入出力信号として処理を行うパルスニューロンの提案など、様々なものがある。

現在は、これらの工学的応用が待たれている。本研究では、従来の統計モデルに基づくアプローチとは異なり、ヒトの聴覚系を模擬した新しい音声処理システムの可能性について検討する。

2. 研究の目的

本研究の最終的な目的は、ヒトの聴覚系を模擬した数理分子モデルを構築し、さらにその解析を行うことにより、普段ヒトが何気なく行っている音声認識や話者識別、音声分離などが可能なシステムを構築することである。

ヒトの聴覚系は、非常に高機能な情報処理を行っている。ヒトは、パーティー会場のような多くの人がいる部屋で数メートルの距離を隔てていたとしても会話ができる。また、ヒトは、異なる距離や環境、さらには電話などの媒体を介したとしても、特に苦勞することなくコミュニケーション可能である。

一般に、我々の日常会話の場面では、目的とする音声だけが存在することはまれであり、様々な雑音と混在した状態で会話が行われる。しかし、ヒトは、そうした状況下においても必要な情報のみに注目して会話をしている。

現在広く用いられている、音声情報処理システムは、非常に高機能ではあるが、環境に対する柔軟性を持つヒトの聴覚系には、はるかに及ばない部分も多い。人間の聴覚における信号処理過程を模擬したモデルを取り入れることで、環境の変化に柔軟な信号処理手法が開発できる可能性がある。

本研究では、ヒトの聴覚末梢系の信号処理に注目し、その数理モデルを構築して、幾つかの音声処理への応用を目指す。

3. 研究の方法

ヒトの聴覚系は、音響振動を神経パルスに変換する聴覚末梢系、神経パルスから音の特徴要素を抽出する聴覚中枢系、そして、その特徴要素を利用して高次の認知処理を行う高次聴覚系で構成されている。

本研究ではまず、聴覚末梢系のモデルを構築する。聴覚末梢系は外耳、中耳、内耳から構成されており、耳介と外耳で集められた音は、鼓膜を振動させ、中耳の耳小骨が鼓膜の振動を内耳の蝸牛に伝える。蝸牛では、リンパ液により基底膜が振動し、有毛細胞がその振動を電気信号に変換している。この後、電気信号が聴神経に伝えられる。

これらの機能をモデル化するため、本研究では、基底膜モデル、順応特性モデル・位相固定性モデルからなる有毛細胞モデル、及び聴神経モデルを使用する。これらのモデルを直列に接続することにより聴覚末梢系モデルを構築する。

(1) 基底膜モデル

基底膜は音響刺激に誘引されて振動を起こす。この振動は音響刺激の周波数を基底膜の位置に写像する一種の関数である。このような基底膜の周波数分析機能は、工学的にはフィルタバンクで実現することができる。本研究では、基底膜の特性に由来する聴神経特性を記述するのに適した Gammatone filter を用いる。

(2) 有毛細胞モデル

蝸牛に内在する内有毛細胞と聴神経間の音響刺激に対する神経伝達物質のやり取りに注目しモデル化したのが、Meddis inner hair cell model である。有毛細胞モデルの中でも、このモデルは非常に簡単な数式で記述され、順応反応などの聴覚系が持つ性質をよく反映している。本研究では、有毛細胞モデルとしてこのモデルを参考にする。

(3) 神経パルス発火モデル

一般に内有毛細胞と聴神経間のシナプスに神経伝達物質が放出されると、聴神経の膜電位が上昇する。この電位は時間と共に減衰するが、電位があるレベルを超えると聴神経は神経パルスを発生する。しかし、聴神経は一度発火すると一定期間発火できない不応期(約 1ms)を持っている。

これは、時間と共に変化する関数を用いて膜電位の時間変化を表現し、さらに、膜電位があるしきい値以上になった場合で、かつ、不応期でない場合にパルスを発生させることで簡単に実現できる。このパルス発生モデル化において、しきい値の設定は非常に重要な要素であり、これにより発生するパルスの特性が大きく左右される。

本研究においては、膜電位の極大値の時間平均と聴神経の特性を用いた以下の3つのステップにより、このしきい値を channel (有毛細胞) ごとに適応的・自律的に自動設定する。

① 膜電位の極大値の時間平均

はじめに、膜電位の極大値の時間平均を周波数に対応した Channel において計算する。この値を用いて、膜電位の大きい Channel でより頻繁にパルスを生成させる。

② 聴神経の側抑制

聴神経には側抑制という特性がある。側抑制は、DOG(Difference of two Gaussians)フィルタで表現することができる。これを各 Channel に作用させ、周波数分解能を向上させる。

③ 高周波域でのパルス生成の抑制

一定のしきい値の場合、低周波域に比べ高周波域で、より頻繁にパルスが発生することが本研究で確認されている。そのため、高周波域でのパルスの生成を抑制させるために Channel 番号に関する高次多項式を導入する。

この①~③を音声フレームごとに行い、しきい値を適応的に得る。ただし、このしきい値の求め方では無音区間において不都合が起きる。無音区間においても、しきい値を同様に求めると、本来無音であるはずのフレームにおいて発火する確率が高くなってしまふ。そこで、現在のフレームが無音区間であると判断した場合は、自動的にしきい値を高く設定することとする。

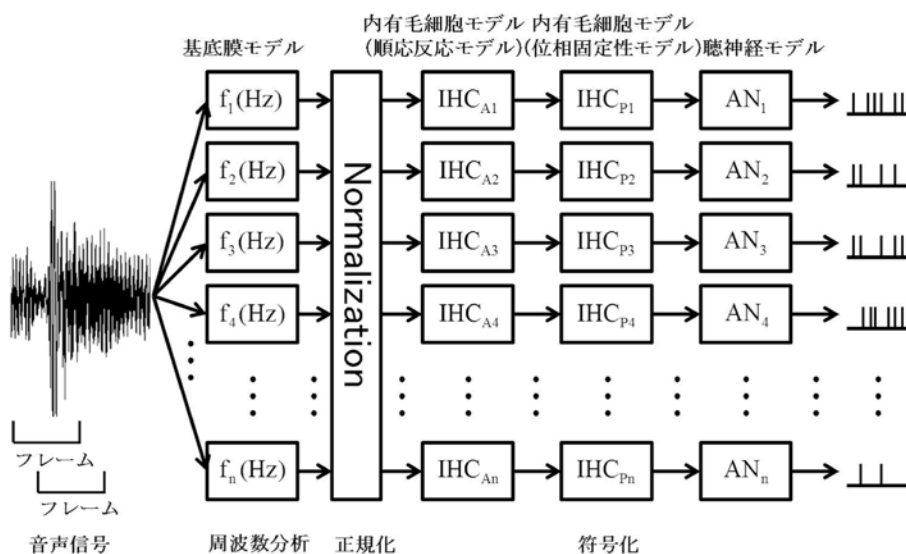


図 1. 聴覚末梢系モデルの概略図

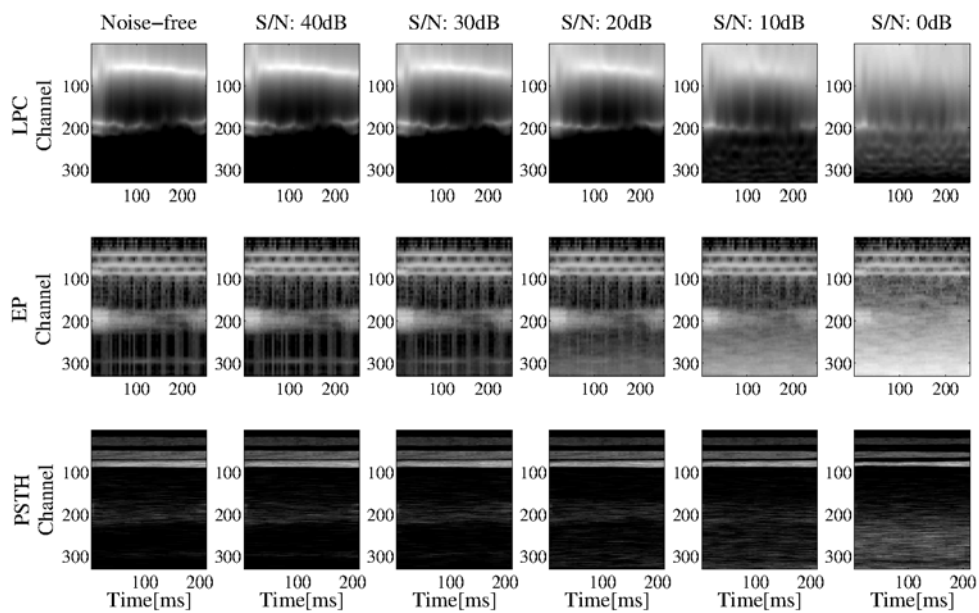


図 2. 雑音下における日本語の母音“e”の各手法による出力パターン(ホワイトノイズ)

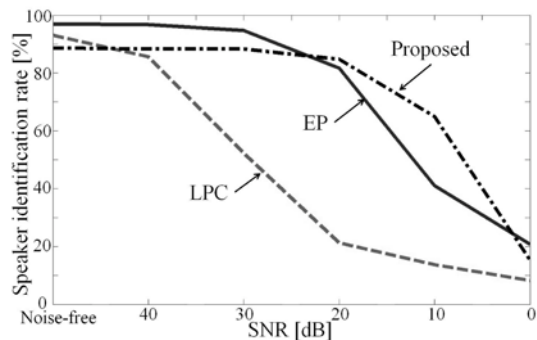


図 3. 雑音下における話者識別結果 (ホワイトノイズ)

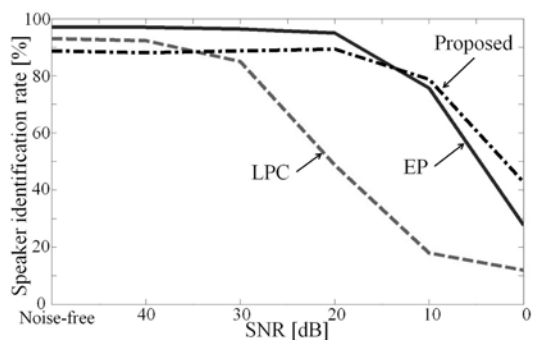


図 4. 雑音下における話者識別結果 (ピンクノイズ)

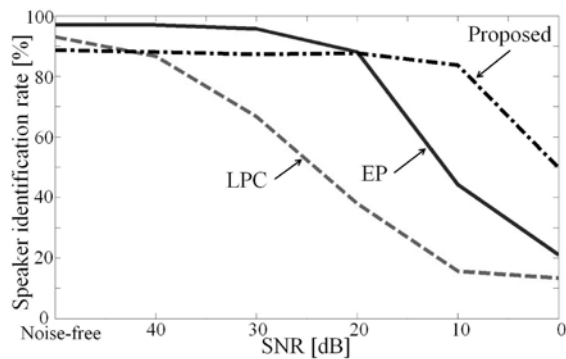


図 5. 雑音下における話者識別結果 (ブルーノイズ)

無音区間の判定は、膜電位の極大値の時間平均の最大値と最小値を用いる。その差の大きさによってフレームごとに判定を行い、その結果に基づいて、しきい値を決定する。

図1に、構築した聴覚末梢系モデルの概略図を示す。本研究で構築した聴覚末梢系モデルの様々な機能を検証するために、モデルから生成されたパルス時系列パターンの解析を行う。

特に、ヒトの聴覚情報処理機構が有するマスキング、周波数選択性、2音抑制などの特性について解析を行い、構築したモデルの妥当性を明らかにする。

さらに、聴覚末梢系モデルから生成されたパルス時系列パターンを用いて雑音環境下での話者識別を行い、その結果と従来の周波数分析との比較検討を行う。

4. 研究成果

構築した聴覚末梢系モデルでヒトの聴覚系で見られる幾つかの特性・現象を示すことができた。具体的には、得られたパルス時系列信号を解析することにより、聴神経の自発放電現象、順応現象、マスキング現象、周波数選択性、2音抑制現象などが、生理データとの比較により、本モデルで良く表現できていることを確認した。

さらに、提案したしきい値決定法を導入した聴覚末梢系モデルの有効性を確認するため、男女計12名（男性9名、女性3名）の各被験者から採取した5つの母音音声5セットを用いて雑音環境化での話者識別実験を行った。パターン識別法としては部分空間射影法を用いた。このとき用いた特徴量は、聴覚末梢系モデルから得られた PSTH (Post-Stimulus Time Histogram) である。

比較のため、従来の LPC スペクトルおよび、Excitation pattern の2つを用いた話者識別実験も合わせて行った。

無雑音の場合と、雑音を加えその大きさを変えながら実験を行った。雑音は人工的に作った雑音として、1. ホワイトノイズ、2. ピンクノイズ、3. ブルーノイズの3種類と、自然環境下で収録を行った、1. 車のエンジン音、2. 小川の水の流れの2種類の計5種類である。雑音の大きさは SNR が 40dB, 30dB, 20dB, 10dB, 0dB の5パターンである。

図2に各手法における出力パターンを示す。

図2より、提案した聴覚末梢系モデルによる出力は、他の手法に比べ、雑音の強度が上がっても本提案モデルの出力に影響が少ない、すなわち、高レベル雑音の影響を受けにくいことが分かる。

図3～図5に、ホワイトノイズ、ピンクノイズ、ブルーノイズ下における本提案モデルを用いた話者識別結果を示す。この結果より、雑音の強度が低い場合、聴覚末梢系モデルは他の方法よりも識別率が劣ってはいないが、雑音の強度が上がるにつれての話者識別結果の劣化は緩やかであり、さらに高雑音下においては本モデルの従来モデルに対する耐雑音性の優位性が確認できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計6件)

- ① T. Azetsu, M. Abuku, N. Suetake and E. Uchino:
"Speaker Identification in Noisy Environment with Use of the Precise Model of the Human Auditory System,"
Proceedings of the 2012 IAENG International Conference on Artificial Intelligence and Applications, pp.92-95, 2012-3-16 Regal Kowloon Hotel Hong Kong China
- ② 安福正啓, 畔津忠博, 末竹規哲, 内野英治:
"話者識別における耐雑音性をもつ聴覚末梢系モデル," 2011年日本知能情報フュージ学会中国四国研究会予稿集, pp.21-26, 2011-12-23 山口大学 山口市
- ③ M. Abuku, T. Azetsu, E. Uchino and N. Suetake:
"Application of Peripheral Auditory Model to Speaker Identification,"
Proceedings of the Second World Congress on Nature and Biologically Inspired Computing (NaBIC 2010), pp.673-678, Best Paper Award, 2010-12-7 Kitakyushu International Conference Center Kitakyushu Japan

- ④ 安福正啓, 畔津忠博, 内野英治, 末竹規哲:
“ヒトの聴覚特性の模擬とその実験的検証,” 第 15 回日本知能情報ファジィ学会中国・四国支部大会講演論文集, pp. 17-18, 2010-12-4 広島大学 東広島市
- ⑤ 安福正啓, 畔津忠博, 内野英治, 末竹規哲:
“ヒトの聴覚特性を模擬したモデルの話者識別への応用,” 第 19 回計測自動制御学会中国支部学術講演会論文集, pp. 168-169, 2010-11-28 島根大学 松江市
- ⑥ 安福正啓, 畔津忠博, 内野英治, 末竹規哲:
“聴覚末梢系モデルから得られる多次元パルス信号を用いた話者識別,” 第 23 回バイオメディカル・ファジィ・システム学会年次大会講演論文集, pp. 35-40, 2010-10-9 産業医科大学 北九州市

[その他]
ホームページ等
<http://www.ic.sci.yamaguchi-u.ac.jp/>

6. 研究組織

(1) 研究代表者

内野英治 (UCHINO EIJI)
山口大学・大学院理工学研究科・教授
研究者番号：30168710

(2) 研究分担者

末竹規哲 (SUETAKE NORIAKI)
山口大学・大学院理工学研究科・准教授
研究者番号：80334051

畔津忠博 (AZETSU TADAHIRO)
山口県立大学・情報化推進室・准教授
研究者番号：70285451