

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25年 5月 24日現在

機関番号：13501

研究種目：若手研究（B）

研究期間：2010～2012

課題番号：22700212

研究課題名（和文）音素識別システムの要素開発による大脳聴覚野理解の精緻化

研究課題名（英文）Elemental study of vowel discernment system based on the primary auditory cortex

研究代表者

鈴木 裕 (SUZUKI YUTAKA)

山梨大学・総合分析実験センター・助教

研究者番号：40516928

研究成果の概要（和文）：

大脳第一次聴覚野の神経生理実験を基にした母音識別システムを研究開発した。具体的には、振幅変調音及び周波数変調音を刺激としたときの覚醒ネコ大脳第一次聴覚野 sustain cell の反応を基に音パルス列変法を開発し、自己組織化マップにより母音を分類させた。さらに、ヒトの聴感に合わせた波形処理である ACE-L 法 (Amplitude-bandwidth Compression / Expansion method designed with Loudness) を開発研究した。

研究成果の概要（英文）：

The vowel discernment system based on the result of the neurophysiology about primary auditory cortex was developed. The sound to pulse train signal converting processing was developed based on sustain cells response. The vowel pulse train signals were input to self-organizing maps and separated. Furthermore, the ACE-L (Amplitude-bandwidth Compression / Expansion method designed with Loudness) method was proposed for the purpose of explaining the hearing of sensorineural hearing losses.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,400,000	420,000	1,820,000
2011年度	900,000	270,000	1,170,000
2012年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,100,000	930,000	4,030,000

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：母音，大脳第一次聴覚野，持続反応細胞

1. 研究開始当初の背景

感性を科学的な視点から捉えて解明していくために、高次認知機能に対して低次となる知覚に関して理解を深めていく必要がある。本研究では、まさに感性情報処理であるヒトによる音素識別を当面の対象として、聴覚感性情報処理を深く検討することとした。

研究協力者である佐藤悠ら（山梨大学医学

部生理学第2教室・教授）は、覚醒ネコの脳第一次聴覚野における単一神経細胞活動を記録し、スパイク発火頻度と刺激音の音響特徴との関連を抽出し、大脳第一次聴覚野神経細胞における音素・音響情報処理過程を次々と解明してきた。また、母音を刺激としたヒト心理実験において、母音弁別には母音の第一ホルマントと第二ホルマントの周波

数が重要であると共に、それらホルマントが含む周辺周波数成分が重要であることを発見してきた。

一方、研究代表者はこれまでに、熟練者が音響信号を知覚、認知する機能をモデルとした音響信号処理システムを構築することで感性情報処理に関する研究を行ってきた。具体的には、内耳基底膜での周波数分析特性、感じる音の大きさに基づくラウドネス特性、人工的ニューラルネットワークを組み合わせたオリジナルの情報処理による聴診装置を開発し、熟練者と同様に音響信号の差異が微少な感性的対象の判断が可能となることを確かめていた。また、補聴器・拡声装置の開発、微小音がヒトの感性評価に与える影響を調べる研究を行っており、語音弁別能検査や、聴き取り音声品質・音色感性評価を行っていた。

これらの背景・研究成果に基づき、本研究では大脳第一次聴覚野の反応機序をモデルとした母音識別システムへと発展させることを目指すこととした。さらに、母音を刺激としたヒト心理実験結果を信号処理によって説明することを目指すこととした。

2. 研究の目的

本研究は、聴覚感性に関わる脳機能の解明を医工融合研究として目指すものである。聴覚に関する感性情報工学に従事してきた研究代表者が、神経生理学者の協力により聴覚野機能に関する知識・実験結果を参考に聴覚システムとして構築し、聴覚感性に関わる聴覚野機能理解の精緻化を行うこととした。音素の識別を対象とし、システムとして実用レベルの成果を得ることを目標とした。

3. 研究の方法

以下の(1)心理・生理学的実験を礎として(2)(3)の研究を進めることとした。また、(2)(3)による考察が(1)へとフィードバックされ、感性情報科学を発展させることを計画した。

(1)心理・生理学的測定による大脳第一次聴覚野機序の探求

ラウドネス感覚は音波の振幅に関係することが広く知られているものの、聴覚野における振幅情報分析仮定に関しては、ほとんど研究されていなかった。そこで協力者によりこれら振幅情報分析が調べられ、さらに、音素に反応するネコ大脳第一次聴覚野神経細胞活動計測、音声に対する心理実験についても研究が進められていた。

(2)大脳第一次聴覚野における神経機序の情報工学的モデル化

これまでに、ラウドネス関数を参考にしたラウドネス知覚モデルを構築し、シングモイド関数を用いたフィードフォワード三層型ニ

ューラルネットワークを用いた音響信号識別システムを開発してきた。本研究では更に細部の大脳第一次聴覚野における振幅情報分析過程および周波数分析過程をインプリメントする。具体的には音刺激による大脳第一次聴覚野神経細胞の発火タイミングを算定し、周波数に順応する各神経細胞間を荷重加算で結ぶ自己組織化マップニューラルネットワークモデルを構築することとした。

(3)音素識別システムの開発による大脳聴覚野理解の精緻化

心理・生理実験で確認されている各音素(母音を対象とする)に対する識別キューとシステムの反応を考察することとした。さらに、ヒトは難聴や低音量下ではしばしば異聴を引き起こすが、そのメカニズムを解明するためにヒトの聴感に合わせた波形表示信号処理法である ACE-L 法 (Amplitude-bandwidth Compression / Expansion method designed with Loudness) を開発した。

4. 研究成果

(1) 生理実験の結果

協力者らにより以下の生理実験が行われていた。覚醒状態のネコにおいて大脳第一次聴覚野から単一神経細胞活動(スパイク電位)を記録することで、大脳第一次聴覚野細胞の反応様式が調べられた。ヒトの母音やネコ鳴き声(meow)に反応するネコ大脳第一次聴覚野細胞には反応多様性があり、phasic cell は母音の開始と終了、振幅の一過性変化を分析し、sustain cell は音声のホルマント周波数分析を行っていることが明らかにされた。音の大きさの変化に対する大脳第一次聴覚野の反応様式を詳しく調べるために、ランプ音(音圧が時間とともに上昇する)とダンプ音(音圧が時間とともに下降する)刺激に対するスパイク反応様式が調べられた。その結果、刺激の持続時間が 40 ms 以上のとき sustain cell はオンセット反応がある上に、ランプ音刺激ではダンプ音刺激よりも反応の順応が遅い反応が見られた。ゆえに、sustain cell は、刺激音圧の上昇はスパイク頻度の上昇、刺激音圧の下降はスパイク頻度の下降によって符号化されることが示唆された。

音の高さに関しては、大脳第一次聴覚野では最適周波数の異なる細胞が地図状配列していることがすでに分かっていた。周波数が時間とともに上昇する周波数上昇音、下降する周波数下降音によってその反応様式がさらに詳しく調べられ、細胞の周波数応答領域を通過するとき sustain cell の持続反応が得られた。これらの生理実験結果により大脳第一次聴覚野の反応モデルを作成した。

(2) 振幅変調音に対する反応モデル

本研究では母音分析において支配的である sustain cell の音圧変化と周波数変化に関する反応様式から音声-神経パルス発生信号処理を構築した。

まず、入力音の音圧情報を取り扱うために、振幅の包絡線を得て、その包絡線振幅値の関数として刺激パルス列を決定する。パルス発生時刻 t を包絡線振幅値 V_e による線形関数として定式化した。

$$t = \frac{\theta_2 - \theta_1}{(f_2 - f_1)V_e + (\theta_2 f_2 - \theta_1 f_1)} \quad (1)$$

ここで、 θ_1 、 θ_2 はそれぞれ、音反応パルスが発生する振幅最小閾値と最大閾値を表す。また、 f_1 、 f_2 はそれぞれ、 θ_1 、 θ_2 の振幅値が入力されたときの反応パルス頻度である。最小閾値 θ_1 より小さな V_e の信号はパルスを発生させることはなく、また、 θ_2 より大きな V_e の信号が入力されたときの出力パルス周波数は f_2 とする。

図 1 にランプ音を入力したときのモデルの出力を示す。入力信号 (図 1(a)) は 400 Hz 純音のランプ音であり、図 1(b) は抽出した包絡線を示す。図 1(c) は式 (1) を用いて変換した出力パルス列である。各パラメータの条件は、 $\theta_1=0.01$ 、 $\theta_2=0.8$ とし、スパイク頻度は $f_1=8.8$ Hz、 $f_2=44$ Hz としている。これらの条件は、パルス列を図示するための条件である。ランプ音によって頻度が上昇するパルス列が作成されていることが確認できる。

(3) 周波数変調音に対する反応モデル

上記のパルス変換法は純音に限った場合の、音の大きさ変調音に対応する sustain cell の簡易反応モデルであり、周波数変調音や複合音には対応しない。そこで、フィルタバンクにより周波数帯域分割し、帯域ごとに式 (1) を適用し、出力チャンネルに対応させることとした。これは、大脳第一次聴覚野の周波数選択性細胞に対応する。

本研究では、音の高さに関する聴覚モデルとして一般的に用いられる周波数尺度である ERB で配分したフィルタバンク、及び対数配分、線形配分、冪乗配分の比較検討を行った。規格化したときの各尺度の特性を図 2 に示す。ERB は対数的な特性であり、冪乗処理 ($f^{0.6}$) は中間的な特性である。対数的な尺度ほど低周波数成分の分割は細くなる。

入力信号を 50 Hz から 4 kHz において 8 個の帯域に分割するフィルタバンクを用い、各帯域の成分の包絡線を式 (1) を適用することで 8 チャンネルのパルス列を得る。

図 3 に、周波数軸を冪乗としたフィルタバンクを用い、日本語母音 /i/ または /u/ を入力したときの出力結果を示す。各パラメータは、

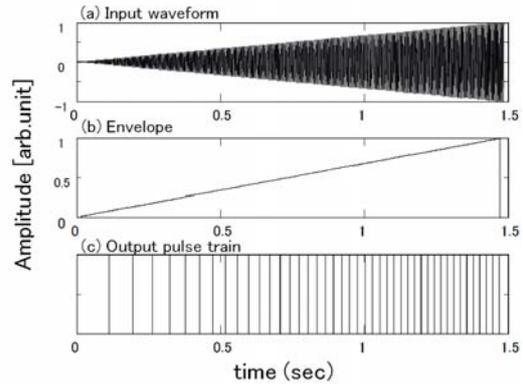


図 1. ランプ音刺激をしたときのモデル出力

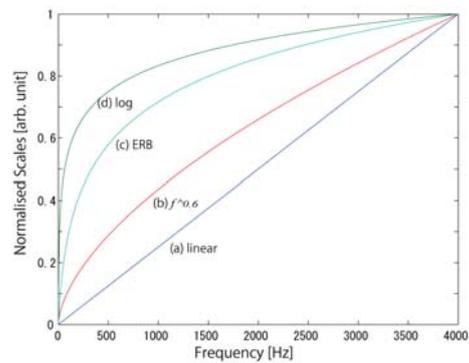


図 2. 各フィルタバンクの周波数配分

$\theta_1=0.02$ 、 $\theta_2=0.2$ とし、パルス周波数は $f_1=100$ Hz、 $f_2=400$ Hz とした。Ch. 1 は低周波数を示し、Ch. 2、Ch. 3 になるにつれ高周波帯域の出力となる。結果ではチャンネルごとに出力されるパルス数は異なり、/i/ では Ch. 3 の成分が大きく、Ch. 6、7 においてもパルスが出力されることが分かる。一方、/u/ を入力したときでは Ch. 1、Ch. 2、Ch. 3 にパルス出力が得られている。これは、入力音声の第一ホルマントおよび第二ホルマントの成分の情報が保存されていることを示している。

(4) 自己組織化マップによる母音分類

自己組織化マップとは教師なし競合近傍学習アルゴリズムを用い、入力層と出力層により構成された 2 層のニューラルネットワークであり、データマイニング手法として用いられる。高次元の入力ベクトルの集合を学習させることで、正負加算演算の結果、類似したサンプル群はマップ上の同一のノードまたは近傍に集まった出力マップが得られる。

日本語を母語とする 20 代から 50 代の話者 20 名 (男 15 名、女 5 名) から日本語母音を録音した。発声時間はおおよそ 0.3 秒であり、1 母音あたり、おおよそ 30 回発声した。

録音した音声のおおよそ 0.54 秒に音声-パ

ルス変換法を適応した。各パラメータは、 $\theta_1=0.02$, $\theta_2=0.2$, $f_1=100$ Hz, $f_2=400$ Hz とした。フィルタ分割数は8帯域として50Hzから4kHzであり図2で示した尺度でそれぞれ等分割する。その後、帯域ごとの総パルス数を算出し、自己組織化マップへと入力して学習させた。

各尺度による処理ごとの自己組織化マップの学習結果を図4に示す。学習終了条件は学習回数500回とした。出力マップのサイズは10×10個の六角格子型マップである。色づけられた六角形は出力されたサンプルの個数を大きさで表しているが、4個以下ときは表示していない。自己組織化マップは入力ベクトルが特徴空間に写像されたものであるため、縦軸及び横軸の概念はなく、近傍の出力ノードには類似した入力ベクトルをもったデータが出力され、離れた出力ノードほど入力ベクトルの類似しないデータが出力される。

尺度を冪乗とした図4(b)の出力マップが、母音ごとの分類が最も良好であり、母音ごとに1つ、または2つのエリアを形成している。エリアを形成できないデータに関しては、今後検討が必要である。続いて良好なのは線形としたときであるが、冪乗よりも/u/がマップ上に広く分散しており、分類が難しいことが分かる。対数、ERB尺度を用いたときは、さらに母音ごとのエリアが曖昧になる。特に/i/と/u/、/e/と/o/、/a/と/o/の分類が難しくなる。ERB尺度は聴覚模擬として用いられるが、今回のフィルタ分割数と周波数帯域の条件では、冪乗の方が母音識別には適切であった。

(5) ACE-L 処理による聴感に合わせた波形表示法

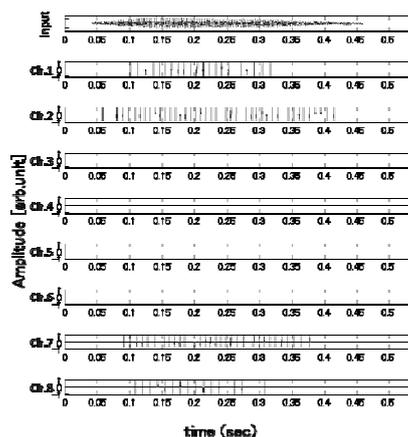
ヒトは難聴や低音量下や騒音下ではしばしば異聴を引き起こすが、そのメカニズムを解明するためにヒトの聴感に合わせた波形処理法であるACE-L法を開発した。

ヒトが感じる音の大きさ(ラウドネス) L は強度 I に対して非線形であり、純音の場合、その関係は次式で表される

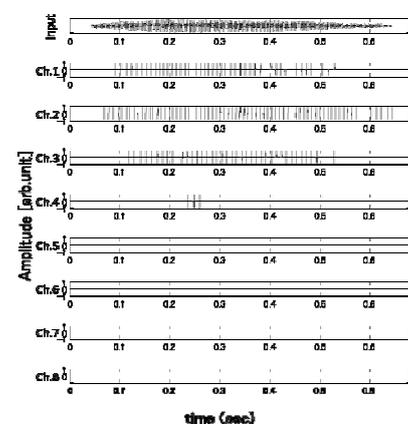
$$L = c(I^n - I_c^n) \quad \dots (2)$$

ここで、 c は線形増幅係数、 n は健聴者の感知指数、 I_c は聴覚閾値音圧である。ラウドネス L と強度 I の関係は周波数によって異なり、健聴者の場合(2)の各パラメータは周波数ごとに等ラウドネスレベル曲線ISO226(2003)に従って求まる。(2)を参考にし、聴取に合わせた波形処理を再現するために、ACE-L処理では次式を用いる。

$$w = \text{sgn}(v)k(|v|^{2n} - v_c^{2n})^{\frac{1}{2}} \quad \dots (3)$$

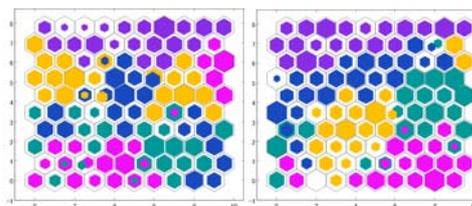


(a)母音 /i/



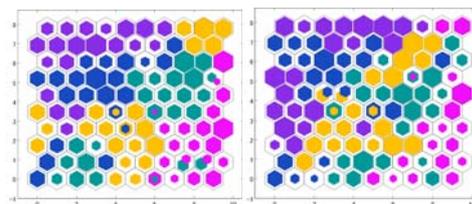
母音 /u/

図3. パルス列出力結果



(a) Linear

(b) Exponential ($f^{0.6}$)



(c) ERB

(d) Logarithm



図4 SOM の出力結果

ここで、 sgn は符号関数、 k は線形増幅係数、 v は入力音声を電圧に変換した瞬時値、 v_c は聴覚閾値、 n は処理に応じた圧縮伸張係数、 w は音の大きさ L を知覚させる瞬時値を示す。例えばラウドネス L が 1 [sone] のとき w の実効値を 1 [stev] とする。つまり、 $L=1$ [sone] の実効値を持つ正弦波の w は $\sqrt{2}$ [stev] として出力される。健聴者の場合 (3) の各パラメータは、ISO226 (2003) のパラメータを基に設定することで、健聴者の聴取の感覚に合わせた波形が示される。

外有毛細胞が損傷した感音性難聴者の聴こえ方の特徴として、小さな音は聴くことができない一方、大きな音の聴こえは健聴者に近づく (補充現象がある)。難聴の度合いが大きいほど聴取できる最小の音圧は上昇すると共に、小さな音ほど大きく非線形増幅する機能は失われる。本研究では健聴者と難聴者で 90dB SPL の音圧で同程度のラウドネスと仮定し、健聴者と難聴者のそれぞれの最小可聴値 I_{cn} 及び I_{ci} を用いて難聴者の特性を決定した。得られた特性を健聴者の圧縮伸張係数 n からの非線形性の損失 m とし、さらに聴覚閾値 I_{ci} を設定することで強度 I に対する感音性難聴者のラウドネス L を推定することとした。この場合、難聴者の ACE-L 処理は次式となる。

$$w_i = \text{sgn}(v) k' \left(|v|^{2nm} - v_{ci}^{2nm} \right)^{\frac{1}{2}} \quad \dots (4)$$

ここで、 sgn は符号関数、 k' は線形損失係数、 v は入力音声を電圧に変換した瞬時値、 v_{ci} は難聴者の聴覚閾値、 n は健聴者の感知指数、 m は健聴者から比較した難聴者の感知指数、 w_i は音の大きさ L を難聴者に知覚させる瞬時値を示す。

入力された音はバンドパスフィルタで周波数帯域ごとに分割される。分割する周波数帯域は 160~320, 320~640, 640~1280, 1280~2560, 2560~5120 [Hz] の 5 つとした。次に帯域ごとに (3) または (4) の処理を行う。ここで周波数帯域ごとに異なったパラメータを設定することができる。その後再度バンドパスフィルタを通し (3) または (4) により生じた高調波の除去を行い、最後に分割した帯域を合成し出力とする。

ACE-L 処理によって得た健聴者と感音性難聴者の聴取の波形表示法を用いて比較を行った。入力として母音を用いた ACE-L 処理波形および、その周波数スペクトルを比較した。健聴者の ACE-L 処理波形では小さな音ほど相対的に大きく増幅されていることが表現され、一方、難聴者をモデルとした ACE-L 処理波形では、小さな音は健聴者ほどは増幅されず、聴覚閾値以下の信号は表示されない結果が得られた。また、周波数スペクトルの比較では、異聴が引き起こされる母音ではホルマ

ント情報が欠如していることが示される結果が得られた。ここで示した ACE-L 処理手法は大脳聴覚野での知覚を理解するための有効な手法であると考えられる。

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)

[雑誌論文] (計 2 件)

1. Yutaka Suzuki, Osamu Sakata, Shun-ichi Imamura, Shuichiro Endo, Akihito Mizukoshi, Asobu Hattori, Nozomu Iida and Hayato Waki, "Evaluation of ACE method based on OSL as hearing aid for sensorineural hearing losses," Journal of Signal Processing, Vol.16 No.6, pp.581-586, (2012) 査読有
2. Yutaka Suzuki, Osamu Sakata, Ling Qin and Yu Sato, "Voice-Pulse Conversion Method Based on the Response of the Primary Auditory Cortex -For Auditory Input-Type BMI by using Electric Micro-stimulation of the Primary Auditory Cortex-," Proc. of IEEE TENCON 2011, pp.23-27, DOI : 10.1109 / TENCON.2011.6129308, (2011) 査読有

[学会発表] (計 8 件)

1. 飯田 望, 鈴木 裕, 遠藤 周一郎, 今村 俊一, 水越 昭仁, 服部 遊, 阪田 治, "ACE ラウドネス処理を用いた難聴者の音声聴取評価法に関する要素研究," 電子情報通信学会 2013 年総合大会, (2013/3/19), (岐阜)
2. 飯田 望, 鈴木 裕, 遠藤 周一郎, 今村 俊一, 水越 昭仁, 服部 遊, 阪田 治, "振幅圧伸法を用いた難聴者の音声聴取評価法に関する要素開発," 日本福祉工学会 第 16 回学術講演論文集, pp.43-44, (2012/12/1), (千葉)
3. 鈴木 裕, 飯田望, 阪田 治, 脇 隼人, 今村 俊一, 遠藤 周一郎, 服部 遊, 水越昭仁, "波形処理による音声聴取の評価法に関する要素研究(I)," 平成 24 年度電気学会電子・情報・システム部門大会講演論文集, (2012/9/5), (青森)
4. Yutaka Suzuki, Nozomu Iida, Osamu Sakata, Asobu Hattori, Hayato Waki, Shun-ichi Imamura, Shuichiro Endo, Akihito Mizukoshi, "Elemental Study of the Visualization Method for Evaluating Voice-hearing Power by Using Amplitude Compression / Expansion," 34th Annual International Conference of IEEE in

- Medicine & Biology Society (EMBC2012), THD08.104, (2012/8/30), (San Diego U. S. A)
5. Yutaka Suzuki, Osamu Sakata, Shun-ichi Imamura, Shuichiro Endo, Akihito Mizukoshi, Asobu Hattori, Nozomu Iida Hayato Waki, “Elemental Study of ACE method based on OSL as hearing aid for sensorineural hearing losses,” 2012 International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’12), pp.249-253, (2012/3/3) (Hawaii, U. S. A)
 6. 鈴木 裕, 阪田 治, 今村 俊一, 遠藤 周一郎, 水越 昭仁, 服部 遊, 飯田 望, 脇 隼人, “最適語音レベルに基づく圧縮処理音の難聴者による評価 (I),” FIT2011, (2011/9/9), (北海道)
 7. 鈴木 裕, 秦 嶺, 劉 永春, 王 驚宇, 阪田 治, 佐藤 悠, “大脳聴覚野直接電流刺激による聴覚入力 BMI のための音声-パルス信号変換法に関する要素研究,” 第 50 回日本生体医工学会大会, P3-4-6, (2011/5/1), (東京)
 8. 鈴木 裕, 加藤 隆也, 阪田 治, 今村 俊一, 遠藤 周一郎, 水越 昭仁, 服部 遊, “振幅圧縮処理した音声の聞き取り検査の自動化,” 日本福祉工学会第 14 回学術講演論文集, pp.81-82, (2010/11/27), (岩手)

[その他]

ホームページ等

http://erdb.yamanashi.ac.jp/rdb/A_Displnfo.Scholar/9/BC8438194151434C.html

6. 研究組織

(1) 研究代表者

鈴木 裕 (SUZUKI YUTAKA)
山梨大学・総合分析実験センター・助教
研究者番号：40516928

(2) 研究分担者

なし

(3) 連携研究者

なし