バイオメディカル・ファジィ・システム学会誌 Vol.12, No.1 , pp 19−28 (2010)

[Original article]

(2010年4月23日 Accepted)

内耳の情報処理を模倣した音声信号収集システムの性能改善

吉田 秀樹[1], 中野 正博[2], 行正 徹[3], 福地 博行[4], 進藤 覚弥[5],
有田 敏彦[1], 鞘師 守[1], 前田 康成[1], 羽山 雄偉[1], 横野 和也[1]
1）北見工業大学・情報システム工学科／地域共同研究センター知的財産本部
2）産業医科大学・産業保健学部 3）産業医科大学・医学部・心理学
4）株式会社福地工業 5）社団法人北見工業技術センター運営協会

要約： 先行報告した音声帯域(80-5,120 Hz)のデータ収集システムの性能改善を実施した。同システムは従来のPCMデータを、6帯域に分割された時間−周波数平面を構成する極大値と極小値の時系列に変換し、聴性認識の研究や応用製品の開発に資することを目的とする。帯域制限波の極値の記録から元の音響情報は再生できるので、許容誤差内で高速に極値を記録することが課題であった。解決の鍵は極値を推定することに見出され、標本化周波数を従来値の44,100 Hzから22,050 Hzに下げたことで、より少ない入力データから情報抽出が実現し、計算時間の約67%が短縮された。さらに共有メモリの設計をブロック長16,384バイト、ブロック数2とし、3種類の汎用フォーマットを提案したことで、ファイル容量の約65%が削減され、6秒毎に出力されていたファイルが約1.5秒毎に生成できる様に改良された。
キーワード： 聴性情報処理, 極値, 最小二乗法, 音声, サンプリング

Improvement of the Extremal Sampler System Modeled on Auditory Processing in the Inner Ear

Hideki YOSHIDA[1], Masahiro NAKANO[2], Toru YUKIMASA[3],
Hiroyuki FUKUCHI[4], Akiya SHINDO[5], Toshihiko ARITA[1], Mamoru SAYASHI[1],
Yasunari MAEDA[1], Yuui HAYAMA[1] and Kazuya YOKONO[1]

1) Dept. of Computer Science & Cooperative Research Center, Kitami Institute of Technology
2) Dept. of Physics & Information Science, University of Occupational & Environmental Health
3) Department of Psychology, University of Occupational and Environmental Health
4) Fukuchi, Co. 5) Kitami Industrial Technology Center

Abstract: *Estimation of the local maximal data in the band-limited waveform by using the least squared method has been a key technique in order to reduce processing time, especially for the six-channel finite-impulse-response filter-array on our proposed real-time acquisition system. According to our previous study, more than half of minimum data in the band-pass filtered waveform can be substituted by average of the contiguous maximum data, with achieving lower than 4% of phase error. Not only approximate 67% reduction of processing time but also less than half of file size, e.g., approximate 65% saving of 1-h speech data, has been attained, releasing three types of the .ext file format for the extremal data. Further, a design of shared memory has played an important role in performance of response on a single processor system of PC-AT compatible computer, offering that length and number of blocks, and sampling frequency were 16,384-byte, 2 and 22,050-hz, respectively. We have improved the acquisition system for the speech bandwidth (80-5,120 Hz), in which conventional pulse-code-modulated data series can be simply converted to the extremal data in the tempro-spectral domain, generating files every 1.5-s, for manipulation of fine acoustic structure, synthesis, succeeding processes for recognition and/or categorization, and so forth.*
Keywords: *auditory processing, extrema, least squared method, speech, sampling.*

Hideki YOSHIDA Kouen-cho 168, Kitami, Hokkaidou, 090-8507, Japan
Phone: +81-157-24-9327, Fax: +81-157-24-9344, E-mail: hy@cs.kitami-it.ac.jp

1. はじめに

環境音（生活雑音）の福祉への利用を進める時に、昼夜問わず働き続ける人口の耳のフロントエンド部の開発が必須となる。従来は大規模なマイクロフォン・アレイを使用した高度雑音抑制システム[1]や、事例毎に最大限の計算資源を投入して達成された足音[2]、拍手[3]、物を叩く音[4]、物が割れる音[5]、それに自動車のエンジン音[6]の検出技術等を統合して、最も簡素で汎用的で経済的なモジュールにする必要がある。モジュールは屋根裏か地下に組み込まれて、独居老人宅や、幼児や傷病者の居る家庭のセキュリティーに貢献するかも知れないし、近年開発のめざましいヒューマノイドロボットに組み込まれて対話や通話記録の検索に利用されるかも知れない。共通して云えることは、目立たず、電力を消費せず、電源スイッチすらないことである。これは今日完成の域にある長時間圧縮録音技術とも楽曲ファイルの高速検索ダウンロード技術とも趣旨の異なる技術であり、音声や物音を形作っている複雑な時間形状をした複合音の中から、最小限度の情報を抽出して時系列上に並べただけの、最も汎用性の高いデータ形式として公開されるべきである。従来技術で云えば、短時間周波数解析[7]、ボコーダ[8]、蝸牛フィルター[9]が相当するが、一般に周波数解析技術は時系列データを時間-周波数平面に展開することになるので一時的に著しいデータ量の増大を招き、ボコーダでは情報を削減し過ぎたことが原因で効果音発生装置に用途が限られてしまった。蝸牛フィルターでは聴覚器官の精密な模倣が改めて複雑大規模な作業を避けて通れないことを明らかにした。この意味で情報抽出とは、元の音響エネルギーに類似した構造を書き写す処理であり、時系列波形をミクロに観察することで、聴性認識に関わる部分と、関わらぬ部分に大きく二分できる知見が要求される。加えて、音響波形の改変可能な部分は後から簡易な補間によって、元の PCM データ（従来の非圧縮ディジタル録音方式）[10]相当にいつでも復元できることが望ましい。以上から提案するデータ収集装置には、従来の PCM データおよびその圧縮形式とは異なった符号化が採用されている。

波動を表す最も簡単な基底は複素正弦波であるが[7]、元の波形へ復元するには、位相、振幅、周波数を違えた正弦波を無数に重ね合わせないと、再生した音が不自然なものとなる。先行研究の中で、帯域制限波の形状が正弦波補間に適していることを報告した[11, 12]。例えば通過帯域を１オクターブにとると、フィルタリングされた波形は瞬間周波数と振幅が同時に変化する複雑なうなり様の形状を示す。うなり様波形の極大値と極小値の時刻と振幅のみを記録に残し、隣り合う極値間を正弦波状に補間した波形が、元のフィルタリング波形の音響的な近似になっていることを、主観評価[11]と客観評価[12]の双方から報告した。ヒトの可聴域は 20-20,000 Hz であるので、対数軸上で等分割した 10 チャンネル（10 種類のフィルタリング波形）で網羅することができる。本研究で採用した基底は半周期長の正弦波をつなぎ合わせた複合基底ではあるが、チャンネルを重ね合わせる毎に再生音の厚みが増していき、必要なチャンネル数を最大 10 個と明確に定めていることに利点がある。残る問題は、マルチタスク・オペレーティング・システム(OS)下で動作する汎用電算機を使用して、極値情報の収集に要する時間を可能な限り抑制し、他のアプリケーション・プロセスのために利用できる時間を確保することである。計算時間に関しては極値探索に要する時間よりも、フィルタリング処理とファイルへの書き込み処理（I/O 時間）に費やす時間が支配的であった。先行研究で到達できた数値には余裕が無く、音声帯域（80-5,120 Hz、チャネル３から８まで）のリアルタイムな極値収集処理で、約６秒間に１回のファイル生成とブロック当たり約 0.01 秒の残り時間を実現した[13]。極値探索処理では、標本化周波数を上げないと、より高い周波数信号（例えばチャネル 10、10,240-20,480 Hz 帯域）での正確な極値が取得できないことが指摘される。これは同時に極値に該当しない不要なデータが増大し、フィルタリング処理に著しい時間を取られる原因となっている。これを解決するには極値を探索するよりも、極値のありそうな時刻を最小二乗法で推定する方が有効であることが示された[14]。本研究では、極値サンプラーシステムのひな形[13]に、極値を推定することによる改善策を施して性能向上を試みた。

2. システム構成

聴覚を始め高い時間分解能を要求される感覚情報処理は、「生産者-消費者問題」[15]に相当する。２個の独立なプロセスに同一のメモリを共有させて永久動作にする必要が生じる。本例での生産者プロセスは OS の一部（デバイスドライバ）であり、PCM 方式で録音しながら規則的に割り込みをかけて共有メモリへデー

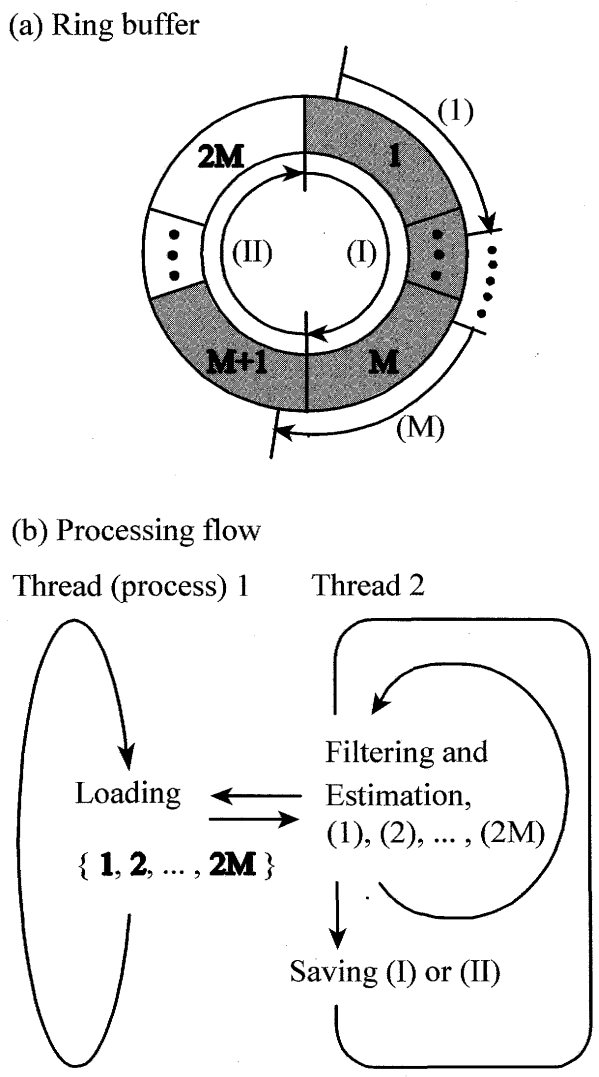


図１　共有メモリの構成と使用手順

タを転送し続ける。消費者プロセスはユーザであり、PCMデータから情報抽出して、より高次の情報処理に当るプロセスに中継するか、必要があればファイルに保存する。両プロセスが共有メモリの使用状況を監視して自身を停止できる設計にすると、あるタイミングを契機に生産者と消費者双方が実行停止に陥る場合があるので[15]、本例の共有メモリは図１aに示す様な環状の構成（リングバッファ）として使用し、生産者には共有メモリに一定間隔でデータを転送できる権限が与えられている。この場合、時間調整の責任を負うのは消費者の方で、一定時間内に情報抽出を完了した後は、生産者から次のデータが転送されるまでを待機としなければならない。厳格な時間的制約が課されているので、子プロセスとして新たに消費者プロセスを生成するのではなく、プロセス（文脈）の切り替えにかかるオーバヘッド時間がなるべく少なくて済むスレッド（軽量プロセス）技術を導入した[16]。スレッド間では大域変数、ヒープ領域それにファイル識別子が共有されるので、高速なプロセス間通信を実現しながら疑似並列動作を実現できる。リングバッファの容量(バッファ長）は２のべき乗に設定することで、アドレスの計算を簡略化できる。即ち共有メモリの末尾と先頭の境界を越える時には、アドレス更新時にキャリー(桁上がり）が発生するので、下の桁だけで共有メモリ中の全アドレスが参照できる。リングバッファを等分割する２M個の区画、一区画当たりの容量Ｌ[byte]、および生産者プロセスで録音されるPCMデータの標本化周波数Ｆがシステムの応答性能に関わるパラメータである。ブロック長Ｌが小さかったり、Ｆが大きかったりすると頻繁に割り込みがかかる様になり、正常動作している様に見えても予期せぬシステムダウンを引き起こすことがある。システムダウンに至らずとも消費者プロセスでの処理が追い付かず、生産者プロセスによるデータの上書き（オーバーランエラー）も起こり得る。ブロック数ＭはシステムのボトルネックとなるI/Oアクセス（ファイル保存）の間隔を引き延ばすための値である。一般にハードディスクへの書き込みには時間を要し、しかも不定であるので、頻繁にI/Oアクセスをすればオーバーランを招く要因となる。

以下では図１ｂに示した生産者と消費者プロセスの流れ図と併せて図１ａを説明する。生産者プロセスはスレッド１で実現されており、図１ａに１から**2M**まで番号を付けた小区画にDMAC(Direct Memory Access Controller)を介して永久にPCMデータを転送し続ける。消費者プロセスはスレッド２で実現されており、(1)から(2M)の区画について FIR(Finite Impulse Response)フィルター処理と極値の推定処理を実現する。FIR を始め、ディジタルフィルターでは一定長の窓を設け、その中にあるPCMデータに対して積和演算を実施する。窓長はタップ数と名を変えて呼ばれ、フィルターの遮断特性に影響する。図１ａ中での**１**から**2M**までの区画と、(1)から(2M)までの区画はタップ数の半値（窓長の２分の１）だけずれた構成となっている。即ち、区画１の先頭のデータからフィルタリング処理を開始しても、フィルターをかけたデータが計算できるのはタップ数の半値だけ遅れた時刻からであり、区画**M**全てのデータについてフィルタリング処理を終えるには、区画 **M+1** のデータをタップ数の半値だけ使用しなければならないことを意味している。さらにスレッド２は、必要に応じて区画１から**M**までの領域(I)で推定さ

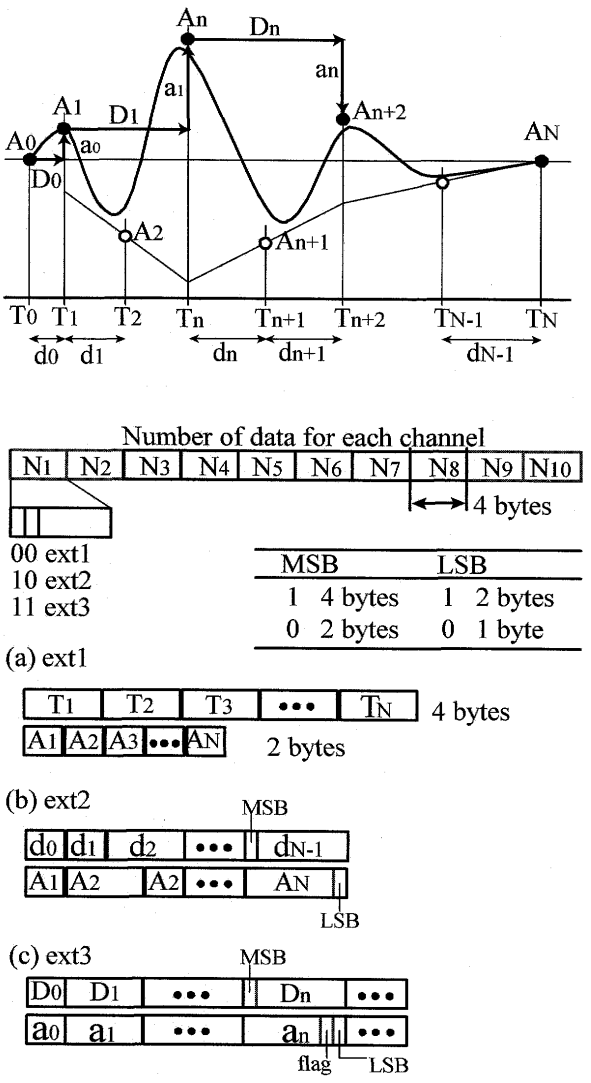


図２　極値ファイル.ext の構成

れた極値データを後述のファイル形式で保存する。以上の動作は循環して実施されるので、スレッド２が(M+1)から(2M)までの区画を処理し終われば、区画 **M+1** から **2M** までの領域(II)について極値データが保存できる。

3.　保存ファイル形式

固定長の ext1 と可変長の ext2、ext3 の併せて３種類のファイル形式を提案する。ヘッダ部は、図２中段に示す様に、各チャンネルにある極値の数をそれぞれ４バイト長で表す。本研究ではチャネル３から８までしか使用しないので、常に N_1=N_2=N_9=N_{10}=0 となる。ただしチャンネル１（N_1部）の先頭２ビットにはファイル形式のコード（00–ext1, 10–ext2, 11–ext3）を上書きするので、ファイル拡張子はファイル形式に依らず共通とし.ext を使用する。ext1 は図２上段に示す波形の極値の出現時刻 Tn を４バイト長で書き並べた後、振幅 An を２バイト長で書き並べたものである。以上の配置を、低位のチャンネル３から高位のチャンネル８に向けて逐次書き並べる。ext2 は隣り合う極値間隔 dn を書き並べた後、振幅 An を書き並べる。ext3 は極値と極値の間の時間 Dn を書き並べた後、極値から極値までの振幅変化 a_n を書き並べる。ただし Dn は極大値から極大値までの時間間隔であることもあるし、隣り合う極値間隔のこともある。前者の時は図２上段に白丸で示す極小値相当のデータは記録されず flag に１を立て、後者の時は flag を０にして a_n 部の下位から２ビット目に上書きする。dn 部と Dn 部は可変長であり、MSB（最上位ビット）が１の時４バイト長、０の時２バイト長とし、データ毎に小さな容量で保存する。時間部は正値をとるので、MSB は符号と別の意味を持たせても問題にならない。ext2 の An 部と、ext3 の a_n 部も可変長であり、LSB（最下位ビット）が１の時２バイト長、０の時１バイト長とし、データ毎に小さな容量で保存する。振幅部の下位２ビットは転用したので、振幅データは４刻み（４の倍数）でしか記録できなくなる。分解能は落ちるが、合成音の知覚の上では問題にならない[11]。

4.　極値の推定法と補間法

帯域制限波から正確な極値情報を取得するには、原波形をなるべく高い標本化周波数で記録した方が良い。しかし同時に、極値ではない夥しい不要データまでもフィルタリング処理しなければならなくなるので、リアルタイム処理の足枷となる。先行研究[14]では極値を計測するのではなく、最小二乗法を使用して極値の推定が可能なことを報告した。本研究でも計算時間削減のために最小二乗法を適用した。図３a に示す様に、標本化周波数に対して充分に周波数の低い信号波形の場合には、推定するまでもなく、複数在る頂点の標本値の中から中央の標本値を極値に選べば良い。図３b に示す様な高周波信号の場合には、頂点近傍に白丸で示した３箇所の標本値を放物線で近似して、矢印で示した放物線の頂点座標を極値とすれば良い。頂点近傍の標本値を多数採っても推定値に大きな改善は観察されなかったので[14]、最小二乗法は計算時間を抑制するために３箇所のデータのみを適用した。ファイル形

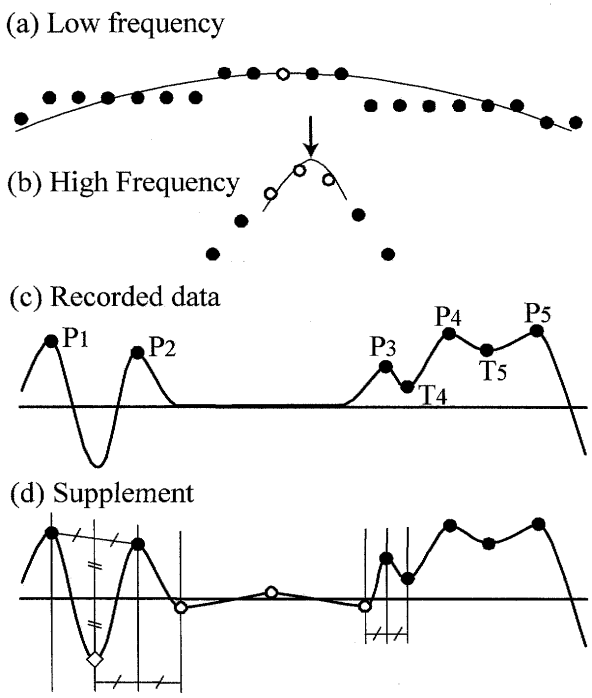


図３　極値推定と例外処理　(a)低周波信号成分（黒丸）では白丸が計測された極値データ (b) 高周波信号成分では３個の白丸を放物線で近似して矢印で示す極値を推定　(c)記録した極値（黒丸）P:peak, T:trough　(d) 極小値（菱形）は隣接する極大値を推定した後に平均して代用、白丸は低振幅になる様に補う

式ext3を使用して、図３cに黒丸で示す７個の極値データを記録に残した場合を例に挙げて再構成法を説明する。先ずP1とP2の様に極大値を連続して記録した時には、P1からP2への振幅変化を表すデータ（図２cのa_n部）のflagビットに１が立ててあるので、白抜きの菱形で示した極小値データを補う必要がある。極小値データは、極大値P1（T_1, A_1）とP2（T_2, A_2）の平均値（$(T_1+T_2)/2$, $(A_1+A_2)/2$）を求めた後に時間軸対称な点（$(T_1+T_2)/2$, $-(A_1+A_2)/2$）として算出する。次にP2からP3までは時間が開き過ぎている例外波形であるので、平均値を極小値に採用しない方が良い。この場合flagビットは０のままであるので平均値は求めず、以下の通り白丸で示した３点で補う。P2から１個前の極小値までの時間と同じ時間だけP2から進めた時刻に、なるべく０に近い負値（例えば－１）をとり、P3から１個後の極小値までの時間と同じ時間だけP3から前に戻った時刻に、同様になるべく０に近い負値をとる。残る１個は両者の中間点になるべく０に近い正値（例えば１）をとる。最後に計測された極小値T4とT5は負値をとっていない例外波形であるので、P3とP4およびP4とP5の組み合わせから平均値を算出しても誤差が大きい。極小値T4とT5はデータとしてext3ファイルに記録する以外に術はない。この場合もflagビットは０のままで平均値を求めない。尚、再生音を合成するには、こうして完成した極値データの間を正弦波状に補間する必要がある。さらにチャンネル３から８までのそれぞれの正弦波補間した波形を重ね合わせなければならない。

5．　試作機

PC/AT互換電算機（Precision 380, Dell Co., 単一プロセッサーPentium 4 (3 GHz), RAM 1 GB）に汎用アナログ／ディジタル（A/D）変換拡張ボード（LPC-320724, Interface Co., PCIバス用）を搭載させた。拡張ボードに添付のデバイスドライバ（GPG-3100）をインストールした後、２個搭載されたA/D変換器の内１個を22,050 Hz（従来値[13]の半分）で稼働させた。尚、オペレーティング・システムにはRedHat Enterprise Linux WS4 (Kernel 2.6.9-5.ELsmp) を使用し、不要なデーモンを停止して必要最小限度の構成で起動した。なるべく頻繁にファイルを生成し、かつ、システムダウンやオーバーランが生じない設計として、ブロック長 L=16,384バイト、ブロック数M＝２（２M＝４）の値を採用した。この時、１ブロックを処理するのに許される時間はL/F=16,384/22,050=0.743 秒未満となる。尚、FIRフィルターの帯域とタップ数の組み合わせは以下の通りとした。

Ch. 3, $f_3(t)$: 80-160 Hz　　tap=1,103
Ch. 4, $f_4(t)$: 160-320 Hz　　tap=553
Ch. 5, $f_5(t)$: 320-640 Hz　　tap=553
Ch. 6, $f_6(t)$: 640-1,280 Hz　　tap=553
Ch. 7, $f_7(t)$: 1,280-2,560 Hz　　tap=553
Ch. 8, $f_8(t)$: 2,560-5,120 Hz　　tap=553

予め１時間分のラジオ放送の音声を録音しておき、開発したシステムを使用してext1、ext2およびext3ファイル形式で録音し直した。１ブロックを処理するのに要した時間を計測すると共に、ファイル容量の平均値と標準偏差を算出した。ファイル容量の比較のために、大学構内の廊下の環境雑音を2009年11月17日16時30分から17時30分までの１時間、明けて18日午前４時30分から５時30分までの１時間について、

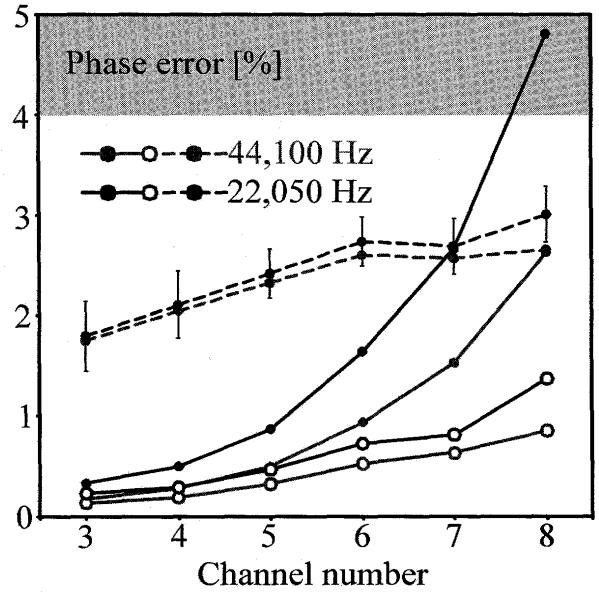


図４　精密計測された極値との位相誤差
実線に黒丸は計測した極値、実線に白丸は推定した極値、破線に黒丸は極大値を推定した後に、極小値を隣接する極大値の平均値で代用

ext1、ext2およびext3ファイル形式で同時に録音した。ファイル容量は、ext1形式で録音された音声ファイルを基準にして百分率で表した。比較のため、ext1形式で録音時には極値の推定処理を実施しなかった。またext3形式での録音時に極小値を省略できる（flagに１を立てる）条件として、計測した極小値と平均値で代用した極小値の間の位相誤差が４%以下であることにした[11]。

６．　極値推定の効果

３種類のパラメータであるブロック長Ｌ、ブロック数Ｍ、標本化周波数Ｆは、位相誤差を算出して決定された。位相誤差Ｐとは、厳密に計測された極値と推定された極値の平均的な時間差ｄを百分率で表したものであり、当該極値に隣り合う極小値から極小値（あるいは極大値から極大値）までを一周期Ｄと考えて以下の様に定義した[11]。

$$P[\%] = \Sigma(d/D)/N \times 100 \qquad (1)$$

ここでＮは平均するのに要した極値数である。図４にはラジオ放送から入手した１時間の音声を予め充分に高い標本化周波数１MHz でサンプリングして厳密な極値を計測した後に、同データを44,100 Hz（灰色の線）と22,050 Hz（黒色の線）にダウンサンプリングして取得した極値との位相誤差を示す。ダウンサンプリングしたデータから極値を計測した場合の位相誤差を実線に黒丸、極値を推定した場合を実線に白丸、極大値を推定した後に極小値を平均値で代用した場合を破線で示した。計算時間を節約するために低い標本化周波数（22,050 Hz）で録音したデータを使用すると、高い周波数帯域のチャンネル８では、位相誤差の閾値となる４%を超えることが観察された。この理由で先行研究では、標本化周波数に44,100 Hz を採用せざるを得なくなり、ブロック数も16にまで拡大することでI/Oアクセス数を減らしてシステムを安定動作させた[13]。その時の性能として約６秒に１回のファイル生成とブロック当り約0.01秒の残り時間を報告した。図４の白丸を見れば極値を計測するのではなく推定することで、どちらの標本化周波数を採用しても全チャンネルについて位相誤差の閾値４%を下回っていることがわかる。また計算時間とファイル容量の削減のために、極小値には推定した極大値の平均値を代用しても全チャンネルについて閾値４%を下回っていることが観察された。以上の結果により標本化周波数Ｆ＝22,050 Hzが採用された。ブロック長Ｌは16,384 あるいは32,768 のどちらかの値しかシステムの動作が安定しなかった。フィルタリングと極値データの推定はブロック長Ｌ毎に実施されるので、なるべく頻繁に極値データが出力できる様にＬ＝16,384 を採用した。ファイル生成間隔に直結するＭは１からでも動作はするが、余裕を持たせてＭ＝２を採用した。これはファイル生成間隔の観点だけからみればＬ＝32,768、Ｍ＝１に等しい。この時のファイル生成間隔はML/F≒1.5 秒に１回となる。

７．　ファイルフォーマット別にみる性能

図５ａにブロック当たりの処理時間を、図５ｂにブロック当たりのファイル容量比（平均と標準偏差）を示す。濃淡付けしたグラフが音声データ、白塗りのグラフが夕方の環境雑音、黒塗りのグラフが明け方の環境雑音を示している。いずれのファイル形式で録音しても処理時間には充分な（> 0.4 s）余裕が観察されており、極値の推定を実施したext2形式（中央のグラフ）ではやや推定処理のために計算時間が増大した。極小値に平均値を代用したext3形式では幾分計算時間が減少しており、推定の手間がいくらか省けたことが理由に挙げられる。ファイル容量比は音声をext1形式（左

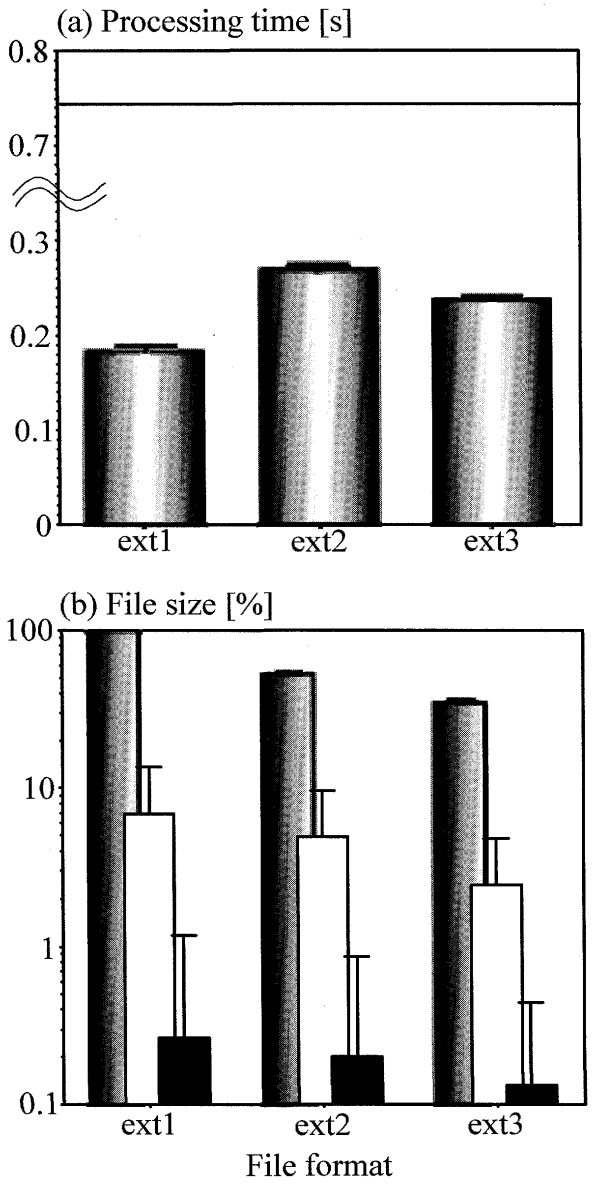


図５　計算時間とファイル容量　ext1 形式の時のみ極値を推定していない、ext3 形式では計測した極小値と代用値との位相誤差が 4%以内の時に極小値を記録から外した、濃淡付けグラフは音声データ、白色は夕方の環境雑音、黒色は明け方の環境雑音それぞれ１時間

端のグラフ）で録音した場合を 100%としているが、ext1 より ext2 が減少し、ext2 より ext3 が減少する様子が観察された。加えて、音声が最もファイル容量を増大させ、環境雑音の録音では学生や教職員の行き交う夕方で 10%未満、往来の途絶えた明け方では 1%未満の結果を得た。

8．　考察

心理学の研究報告によれば、聴覚情報を統合する上で、脳内には短気記憶回路が設けられていることが知られている[17, 18]。リングバッファと同じく非常に短い時間だけ聴覚情報を蓄え、例えば直前に受信した音響信号と、現在受信したての信号との物理的な差異を自動的に検出する上で不可欠の役割を担っている。刺激と刺激の間隔を変更しながら誘発脳波を計測した研究によれば、聴覚情報が 160 ms から 300 ms 程度維持されることが報告されている[17-19]。これは現在の音響データ収集システムをヒトの聴覚性能に近づけるための目標値であり、記録した極値データは遅くとも 300 ms 以内には高次の情報処理にあたるプロセスへ連絡する必要がある。デバイスドライバを改良しない限りＬの値は小さくできないので、例えば、L＝16,384 のブロックをF＝44,100 Hz でサンプリングしたデータで埋めておいて、フィルタリングには２個おきにダウンサンプリングしたデータを使用する手法が考えられる。さらにM＝１とすれば、ML/F≒0.372 秒毎にファイルを転送できるので分散コンピューティング方式で高次の情報処理を実現できるかも知れない。しかし簡素なシステムとするにはファイル保存を取り止めて時間を節約し、第３のスレッドを生成して極値データを第２のリングバッファで連絡する方が現実的かも知れない。

アナログ信号をディジタルに変換するには、一定間隔で離散化と標本化を実施することになる。標本化の時間幅（サンプリング周波数）によって取り扱える信号周波数の上限が決定され[20]、量子化の幅（分解能）が性能を表す１つの指標になっている。提案する極値サンプリングを実施するには、一定周期ではなく、任意の時刻に標本化を実施できる特別な装置[21]が必要となるが、これは同時に普及の上で課題を残した。そこで本研究では、従来の非圧縮ディジタル録音方式である PCM データから、チャンネル３から８まで（80-5,120 Hz）の周波数帯域について極値データを推定することにした。先行報告[13]では約６秒に１回であったファイル保存回数が、約 1.5 秒に１回にまで向上し、約 0.372 秒間の PCM データを処理するのに約 0.36 秒要していた計算時間が、同一の電算機を使用して約 0.119 秒（換算値、ext3 形式で保存時）約 67%の短縮にまで改善した。汎用の電算機には標準で PCM 音源ボードを搭載している機種が多いので、ハードウエアの制約も克服されたことになる。同システムを利用して音声や音楽、あるいは足音、機械音、ガラスの

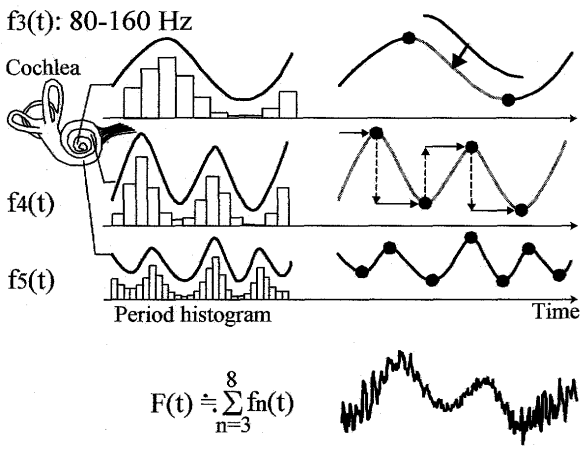


図６　極値ファイルの時間−周波数構造と再構成法

割れる音等の物音の識別と云った高次の情報処理を実施する場合には、さらに電算機単体の性能改善を待つか、あるいは計算時間の要するFIRフィルター部を暫定的に並列処理させる必要がある。共通のPCMデータを使用して、各チャンネルは独立にフィルタリング演算と極値推定処理が実施できるので、音声帯域の処理では６個並列、可聴域の処理では10個並列の外付けプロセッサー（DSP[21]やGPU[22]）の導入も容易である。

チャンネル３から８まで（80-5,120 Hz）の周波数帯域を従来方式で録音するには定理に従って、標本化周波数が10,240 Hzあれば足りる[20]。１ワードが16ビットであるから、データ転送速度は 16×10,240＝163.84 kbpsとなる。図４では音声信号をext1形式で録音したファイル容量を基準に100%としたが、データ転送速度に換算すると708.44 kbps（Hi-Fi Audio: 16-bit×44.1-kHz＝705.6-kbps相当）と非常に大きな値となる。極値の推定方法とファイルの構成を工夫したext3形式にするとやや改善がみられ245.1 kbpsとなり、録音対象によっては夕方の環境雑音17.1 kbps、明け方の環境雑音0.72 kbpsと動的に変化して、PCM方式にはない特徴がみられる。これは情報抽出と情報圧縮の差異を端的に表しており、音声信号の圧縮に特化すれば原音8-bit×8-kHz＝64-kbps を波形符号化方式で圧縮して16-kbps程度[23]、分析合成方式で圧縮して1-kbps程度[8]に下限があることが報告されている。他方、情報抽出の場合では他のアプリケーションで利用できる最も簡明な時間−周波数構造が提供できる様に情報変換技術に重点が置かれている。図６を使って神経生理学と音響心理学を加味しながら、提案したファイル形式を考察する。入力音波の周波数の分析は、内耳の蝸牛内部に在る基底膜の局所的な振動に因ることが報告されている[24]。３回転半した蝸牛管の奥へ進む程、より低い周波数成分が電気信号に変換される仕組みは、FIR フィルターによる帯域制限波からの情報抽出に相当する。基底膜上に在り、振動−電気変換センサーである有毛細胞に接続された聴神経の発火頻度（周期ヒストグラム）の観察報告によれば、発火のピークは振動波形（帯域制限波）と位相が同期することが知られている[25]。即ち、図６に黒丸で示した帯域制限波の極大値あるいは極小値のどちらか一方の出現時刻に合わせてセンサーが応答していると云うことは、時系列データの瞬時値を意味の有るデータと無いデータに選別していることに他ならない。この解釈は神経線維内での情報伝達を全か無かのどちらかディジタルなものだとする学説に適うものである[26]。提案するファイル形式ext1では各極値の出現時刻と振幅をチャンネル毎に簡素に書き並べただけの構成であり、ext2形式では隣接する極値の時間間隔を振幅と共に可変長で記録することでファイル容量の簡易な削減を試みた。残るファイル形式ext3では、可能な範囲で極小値を記録から削除することを認めた。失われた極小値の情報は、近傍にある極大値の情報を単純平均して回復できる場合があり、その見積もりには位相情報の許容誤差４%を適用した[11]。最終的にファイル形式は図６中段にある複合基底 $f_4(t)$の表現の様に、隣り合う極値間の差分をとって時間変分（実線の矢印）と振幅変分（破線の矢印）で表すことになる。例えば振幅変分（量子化幅）を対数軸にとることは[23]、ラウドネス特性[27]の表現に適うものであり、情報圧縮法の１つとして利用されている。差分情報の利用は情報統合の上で必要不可欠であり、ヒトの聴覚情報処理を模倣する上では標本化の時間間隔についても適応化することを提案する[28]。その際、局所的な時間構造（位相情報）を保持できる目安は主観評価の結果から許容誤差4%から9%程度であり[11]、これは既存技術で云えば、高周波信号に（適応）デルタ変調（Adaptive Delta Modulation）[23]した時に発生する傾斜過負荷歪みの抑制評価にも適用できることを示唆している。極値の推定に要する大凡の計算時間を観たかったので、実験では比較のためにext1形式で保存する際に極値の推定を実施しなかった。標本化周波数22,050 Hzでの極値の計測では図５に示す通り位相誤差が充分に小さくならないので、極値は推定

してファイル保存することが望ましい。推定に最小二乗法が適用できたのは、音声波が自己回帰過程（Auto-Regression Process、全極モデル）として記述できるからある[29]。即ち、n個の計測されたデータに最小二乗法を使用して、n+1 番目のデータの値を予想することで、APC(Adaptive Predictive Coding)[30]、ADPCM(Adaptive Differential Pulse Code Modlation)[31]、あるいは LPC(Linear Predictive Coding)[32]と云った圧縮符号化が実用化された。本研究では複雑な予測器の設計を避け、簡易に放物線の形状を利用した。

図６に示す通り、音声データは時間-周波数平面上６チャンネルの極値情報に変換される。精密な蝸牛モデルに従えば、可聴域は24個の臨界帯域[33]に分けられたフィルターバンクで実現されるべきであるが、汎用電算機単体の性能ではフィルター計算のリアルタイムな実現が困難であった。各チャンネルを元の PCM データ F(t)に再構成する手順は、隣接する極値間に失われたデータを正弦波状に補間した後に fn(t)を６チャンネルについて重ね合わせることである。各チャンネルとフォルマント（声道の伝達関数の極）[34]、即ち構音器官によって共振エネルギーが集中する周波数との大凡の対応付けであるが、声帯振動数 F_0 がチャンネル３から４、第１フォルマント F_1 がチャンネル５から６、および第２フォルマント F_2 以上がチャンネル７から８と考えてよい。各チャンネルのデータを任意の時間長で平均すればスペクトルとなるので、音声／音韻認識システムのフロントエンド部としても接続が良い。

9. まとめ

チャンネル８（5,120 Hz）までの周波数成分であれば、ナイキストレートの2×5,120 Hz での標本化により保持される[20]。本研究ではその２倍の 22,050 Hz で標本化したにもかかわらず、5,120 Hz までの周波数分析しか実現できなかったのは、瞬間的な位相同期を実現したかったからである。時間にして帯域制限波の中の隣接する極値を正弦波で近似すれば、正弦波の一周期の２分の１に相当する。ナイキストレートでの標本化では、不確定性関係により、充分な時間（窓長）の信号入力がなければ周波数分解能を低下させてしまう。チャンネル９以上の信号に対してもより高い標本化周波数を採用すれば、聴神経での位相同期上限である４kHz よりも機械の性能が上回る。果たして 10,000 Hz と 10,001Hz の差異を瞬間的に言い当てる能力を望むか否かは設計者の手に委ねられる。先行研究と比較して計算時間の約67%短縮を達成できたのは、極値を計測するのではなく、最小二乗推定を実施したからである。本研究では汎用の電算機でリアルタイムな極値情報の収集ができる様にリングバッファの設計をすると共に、汎用の PCM データから、極値を使用した時間-周波数構造に変換して記述するための簡素なファイル形式を３種類提案した。約 0.743 秒単位で生成した極値データは、約 1.5 秒毎にファイル出力され、情報通信速度は録音対象により変化し、16-bit×44.1-kHz＝705.6-kbps 相当かそれ以下となる。

参考文献

[1] Kaneda, Y. and Ohga, J. (1986): "Adaptive Microphone-Array System for Noise Reduction", IEEE Trans. ASSP, Vol.34, No.6, pp. 1391-1400.

[2] Ekimov, A. and Sabatier, J. M. (2006): "Vibration and Sound Signatures of Human Footsteps in Buildings", J. Acoust. Soc. Am. Vol.120, pp. 762-768.

[3] Repp, B. H. (1987): "The Sound of Two Hands Clapping: An Exploratory Study", J. Acoust. Soc. Am., Vol.81, No.4, pp. 1100-1109.

[4] Freed, D. J. (1990): "Auditory Correlates of Perceived Mallet Hardness for a Set of Recorded Percussive Sound Events", J. Acoust. Soc. Am., Vol.87, pp. 311-322.

[5] Warren, W. H. Jr. and Verbrugge, R. R. (1984): "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics", Journal of Experimental Psychology: Human Perception and Performance, Vol.10, No.5, pp. 704-712.

[6] Choe, H. C., Karlsen, R. E., Gerhert, G. R. and Meitzler, T. (1996): "Wavelet-Based Ground Vehicle Recognition Using Acoustic Signal", Wavelet Applications III, Vol.2762, pp. 434-445.

[7] Cooley, J. W. and Tukey, J. W. (1965): "An Algorithm for the Machine Calculation of Complex Fourier Series", Math. Comput. Vol.19, pp. 297-301.

[8] Dudley, H. (1939): "The Vocoder", Bell Labs Record, Vol.18, No.4, pp. 122-126.

[9] Seneff, S. (1988): "A Joint Synchrony/Mean-Rate Model of Auditory Speech Processing", Journal of Phonetics, Vol.16, pp. 55-76.

[10] Reeves, A. H. (1939): US Patent 2.272.070.

[11] 吉田秀樹、角井健二、前田康成、藤原祥隆 (2008): "極

値サンプリング技術と許容誤差-wavファイルからの情報抽出-" バイオメディカル・ファジイ・システム学会誌 Vol.10, No.2, pp. 123-131.

[12] Yoshida, H., Kakui, K., Maeda, Y. and Fujiwara, Y.: "Evaluation of the Synthesized Speech by Using Mismatch Negativity", International Journal of Biomedical Soft Computing and Human Sciences, under revision.

[13] Yoshida, H., Fukuchi, H., Kakui, K., Maeda, Y., and Fujiwara, Y.: "Design of the Extremal Sampler System by Using a Thread", International Journal of Biomedical Soft Computing and Human Sciences, under revision.

[14] Yoshida, H., Kakui, K., Maeda, Y. and Fujiwara, Y.: "Least-Squares Estimation of the Extrema in the Narrow-Band Music Data", International Journal of Biomedical Soft Computing and Human Sciences, under revision.

[15] Tanenbaum, A. S. (2007): "Modern Operating System", Prentice Hall.

[16] Andrews, G. R. (2000): "Foundations of Multithreaded, Parallel, and Distributed Programming", Addison-Wesley.

[17] Cowan, N. (1984): "On Short and Long Auditory Stores", Psychological Bulletin, Vol.96, pp. 341-370.

[18] Baddeley, A. D. (1992): "Working Memory", Science, Vol.255, pp. 556-559.

[19] Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998): "Temporal Window of Integration of Auditory Information in the Human Brain", Psychophysiology, Vol.35, pp. 615-619.

[20] Shannon, C. E. and Weaver, W. (1949): "The Mathematical Theory of Communication", University of Illinois Press, Urbana.

[21] Yoshida, H., Xie, W., Iriba, T. and Fujiwara, Y. (2005): "Design of the 10-channel Extremum Sampler System", Proceedings of SICE Annual Conference 2005, Electronic published.

[22] Moreland, K. and Angel, E. (2003): "The FFT on a GPU", Proceedings of the ACM SIGGRAPH/EURO-GRAPHICS Conference on Graphics Hardware, pp. 112-119.

[23] Jayant, N. S. (1970): "Adaptive Delta Modulation with a One-Bit Memory", Bell Syst. Tech., Vol.49, No.3, pp. 321-342.

[24] Greenwood, D. D. (1990): "A Cochlear Frequency -Position Function for Several Species- 29 Years Later", J. Acoust. Soc. Amer., Vol.87, pp. 2529-2605.

[25] Palmer, A. R. and Russel, I. J. (1986): "Phase-Locking in the Cochlear Nerve of the Guinea-Pig and It's Relation to the Receptor Potential of Inner Hair-Cells", Hearing Research, Vol.24, pp. 1-15.

[26] Adrian, E. D. (1934): "The Basis of Sensation -The Action of the Sense Organs-", London. Christophers.

[27] Fletcher, H. and Munson, W. A. (1933): "Loudness, Its Definition, Measurement and Calculation" J. Acoust. Soc. Am., Vol.5, No.2, pp. 82-108.

[28] Yoshida, H., (2004): The Japanese patent application No.2004-25437.

[29] Itakura, F. and Saito, S. (1968): "Analysis Synthesis Telephony Based on the Maximum Likelihood Method", Reports of the 6th Int. Cong. Acoust., C-5-5.

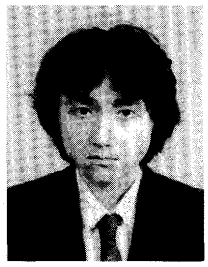
[30] Atal, B. S. and Schroeder, M. R. (1970): "Adaptive Predictive Coding of Speech Signals", Bell Syst. Tech. J., Vol.49, No.8, pp. 1973-1986.

[31] Cumminskey, P., Jayant, N. S. and Flanagan, J. L. (1973): "Adaptive Quantization in Differential PCM Coding of Speech", Bell Syst. Tech. J., Vol.52, No.7, pp. 1105-1118.

[32] Atal, B. S. and Hanauer, S. L. (1971): "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", J. Acoust. Soc. Amer., Vol.50, No.2, pp. 637-655.

[33] Fletcher, H. (1940): "Auditory Patterns", Reviews of Modern Physics, Vol.12, pp. 47-65.

[34] Chiba, T. and Kajiyama, M. (1942): "The Vowel: Its Nature and Structure", Tokyo-Kaiseikan, Tokyo.



吉田秀樹（よしだひでき）
現職　北見工業大学工学部助教授
1991年 九州大学工学部電子工学科卒
1996年 博士（工学）（九州大学）
学会活動　BMFSA 正会員
研究分野　医用生体工学

中野正博（なかのまさひろ）
現職　産業医科大学・産業保健学部
准教授・BMFSA 学会長
1971年 九州大学理学部物理学科卒
1979年 理学博士（九州大学）
2009年 医学博士（産業医科大学）
学会活動　日本物理学会他多数
研究分野　原子核理論物理学・
統計学・情報科学