論　文

2チャネルMUSIC法における複数音源方向の逐次的推定

永田　仁史[†a)]　岩崎　聡[††]　針山　孝彦[†††]　堀口　弘子[†††]
藤岡　豊太[†]　安倍　正人[†]

Step by Step DOA Estimation of Multiple Sound Sources Based on MUSIC with Two-Channel Input

Yoshifumi NAGATA[†a)], Satoshi IWASAKI[††], Takahiko HARIYAMA[†††], Hiroko HORIGUCHI[†††], Toyota FUJIOKA[†], and Masato ABE[†]

あらまし　MUSICに基づく音源方向推定において，頭部伝達関数の影響を受けた2chの受音信号から音声の到来方向を逐次的に推定する方法を提案する．この方法は，音声信号成分の時間–周波数軸上のスパース性を利用し，既に方向推定した音源に起因する周波数成分の重みを低下させることによって次の音源の方向推定精度を高める．筆者らは，同じ考え方を既提案の重み付きウィーナー利得 (WWG) に基づいた方向推定 [1] において導入しているが，本論文では，よく使われる高分解能法であるMUSIC法にもこの考え方を適用できるように，空間スペクトル上の最大ピークを構成する周波数成分の振幅に基づいて逐次的な処理を行う．更に，逐次処理から得られる音源方向の候補について，各候補に属する成分パワーの和を計算して音源数を推定する．提案法の性能評価のため，頭部伝達関数を用いて両耳受聴音を模擬し，様々な音源方向からの到来音を想定して音源方向の検出精度を求める計算機シミュレーションを行った．性能評価の結果，音源が3個で各音源の信号対雑音比 (S/N) が10 dBのとき，コヒーレンスに基づく成分選択を用いた通常のMUSIC法の検出率が約3%であるのに対し，提案法は音源数を既知とした場合が83%，音源数推定も同時に行った場合が78%となり，通常のMUSIC法の性能を大幅に向上できることが確かめられた．

キーワード　音源方向推定，音源数推定，2チャネル，MUSIC，頭部伝達関数，

1. まえがき

音声の到来方向推定は，音声対話システムやロボットなどにおいて，人と機械とのやり取りを円滑にする上で重要な技術である．方向推定によく用いられるMUSIC [2] 等の手法では，入力チャネル数は一般に多い程有利であるといえ，試作ロボットでは10 ch以上のマイクロホンを搭載している例 [3] もある．しかし，筆者らは，ほとんどの動物が二つの耳からの信号を用いて生命を維持していることから，2 ch処理には多くの可能性があるものと考え，2 ch入力を用いた方位–仰角の二次元の方向推定について研究している．

入力信号を2 chに限定した場合，方向推定によく用いられるのは一般化相互相関関数 (GCC) [4] である．GCCは，チャネル間の到来時間差検出に有用であるが，受音部の指向性などの振幅情報を有効に利用していないため，同じ時間差となる複数の到来方向が存在する二次元方向推定にはあまり適していないといえる．これに対し，MUSICや最小分散法 (Minimum Variance Method (MV)) などの高分解能アレー処理 [5], [6] は，振幅情報を含んだ一般のステアリングベクトルを用いれば受音部の指向性を利用できる．更に，到来音が音声であれば，信号成分の時間周波数軸上のスパース性により，入力チャネル数以上の到来音方向推定が可能となる利点がある．しかし，チャネル数が2と少ない場合，通常の処理法では性能に限界が生じる．そのため，性能向上法として，コヒーレンス検出に基づいた周波数成分の選択 [7], [8] や，音源1個の場合が対象で

† 岩手大学工学部情報システム工学科，盛岡市
Department of Computer and Information Science, Iwate University, 4–3–1 Ueda, Morioka-shi, 020–8551 Japan
†† 浜松赤十字病院，浜松市
Department of Otolaryngology, Hamamatsu Red Cross Hospital, Hamamatsu-shi, 434–8533 Japan
††† 浜松医科大学，浜松市
Department of Biology, Hamamatsu University School of Medicine, Hamamatsu-shi, 431–3192 Japan
a) E-mail: nagata@cis.iwate-u.ac.jp

はあるが到来音の調波構造を利用した仮想多 ch 化 [9] などが提案されている．

一方，筆者らは，これまで二つの指向性マイクロホンを回転対称に配置して複数音源の二次元方向推定を高精度で行える重み付きウィーナー利得 (WWG) [10] に基づく方法 [1], [11] を提案してきた．WWG は，相互相関ベースの方法であるが，チャネル間の差信号に基づいた白色化と目的音成分の強調効果により，同じ条件であれば GCC や通常の MUSIC，MV よりも方向検出性能が高いことを確認している [1], [11].

ところで，方向推定処理をロボット等に搭載する場合，マイクロホン近傍の筐体の反射や回折の影響を考慮し，各方向からマイクロホンまでの伝達関数を測定して方向推定の際のステアリングベクトルとして使うことがある [8], [12]．しかし，2 ch の方向推定において，このような伝達系を仮定した場合，例えば，ヒトの頭部伝達関数 (HRTF) の影響を受けた 2 ch の信号に対し，同じ頭部伝達関数をステアリングベクトルとして用いて処理した場合，伝達系が遅延だけで表せるような自由音場の場合と比べて性能が大幅に低下することがある [1]．これは，頭部伝達関数の矢状面上の方向において，伝達関数の差が振幅のみであまり大きくないことと，方向によって複雑に変化することが主な原因である．この結果，周波数によっては離れた方向に対応した伝達関数が近い値となることがあり，単一周波数の空間スペクトルにおいては擬似ピークが発生する場合がある．そこで，WWG に基づく方向推定において，逐次的な音源減衰 (Incremental Source Attenuation (ISA)) と呼ぶ処理 (WWG-ISA) の導入を提案し，これに対処した [1].

WWG-ISA においては，まず，WWG の空間スペクトルにおける最大ピークを第 1 の音源とみなし，次に，このピークに寄与する周波数成分を減衰させるような重み関数を求め，各周波数成分に乗じて再度空間スペクトルを計算し，得られたスペクトルの最大ピークを第 2 の音源とする．更に，第 1 と第 2 の音源ピークに寄与する成分を両方減衰させる重み関数を求め，これを用いて得られた空間スペクトルの最大ピークを第 3 の音源とする．4 個以上の音源に関しても同様に処理する．以上のように，この方法は，不要な周波数成分の影響を低下させるように周波数成分に重みをかけることから，音声強調処理において目的音源に起因する成分を選択するバイナリーマスク処理 [13], [14] に近い．音声のスパース性を利用した成分選択を用いる方法としては，先に挙げたコヒーレンス検出を使う方法のほか，クラスタリングによる周波数成分の分類に基づく方法 [8], [15] も提案されている．

一方，MUSIC や最小分散法においても，WWG-ISA と同様な考え方で逐次的に成分を選択，あるいは重み付けすることが可能であると考えられる．そこで，本論文では，特によく使われる MUSIC に関して同様な考えに基づいた複数音源方向の逐次推定法を検討する．

ところで，音源方向推定において音源数は重要な情報である．音源数は，音以外の情報から得られる場合や既知の場合もあり得ることなどから，通常は方向推定と音源数推定は分けて扱うべきであると考えられる．しかしながら，本提案の方向推定法は，音源数推定との並行処理が可能であるため，音源数推定についても検討することとした．

従来の音源数推定法としては，多チャネルの場合は，空間相関行列の固有値分布に基づいた方法 [16], [17] があり，2 ch の場合には，Mohan らの報告 [7] に示されているような，空間スペクトル上のピークを数える簡易な方法がある．しかし，頭部伝達関数の影響を受けた 2 ch 信号からの二次元推定では，一次元の場合のように際立った音源ピークが音源数個分現れるような理想的な状況とはならないため，ピーク数による推定は困難である．また，周波数ごとのピーク位置に対してクラスタリングを適用する方法も考えられるが，先に述べたように，頭部伝達関数の場合，単一の周波数の空間スペクトルには大きな擬似ピークを生じることが多いため，このような手法の適用は容易ではない．そこで，本方法においては，提案する逐次処理から得られる音源方向の候補に関し，入力信号の各周波数成分がどの候補に起因するものであるかを決定し，この結果得られる音源ごとのパワーに基づいて音源数の推定を行うこととした．

評価実験においては，従来法との比較に先立ち，理想的なバイナリーマスク処理に相当する方法も比較し，理想的に成分を選択できたとした場合に提案法がどの程度の性能を達成しているものであるかも検討した．以下，本論文では，**2.** で提案する逐次推定法について述べた後，**3.** において性能評価の方法と条件，**4.** において性能評価の計算機シミュレーション結果を述べる．**5.** は結論である．

2. MUSIC による 2 ch 信号からの音源数と音源方向の推定

2.1 受音信号モデル

方向 d_s からの到来音が伝達系の影響を受けて二つのマイクロホンに到達するものとする．伝達関数は，事前の測定等により既知であるとする．k を離散フーリエ変換 (DFT) の周波数成分の番号，$d_s = (\theta_s, \varphi_s)$ を方位 θ_s，仰角 φ_s によって表される音源方向，d_s に対応する 1 ch 目と 2 ch 目の伝達関数を $H_{x,k}(d_s)$，$H_{y,k}(d_s)$ とおくと，マイクロホン信号の DFT はこれらを用いて，

$$\begin{aligned} X_{n,k} &= V_{n,k}H_{x,k}(d_s) + N_{x,n,k}, \\ Y_{n,k} &= V_{n,k}H_{y,k}(d_s) + N_{y,n,k}, \end{aligned} \tag{1}$$

と表すことができる．ここで，n はフレーム番号，$X_{n,k}$ と $Y_{n,k}$ は，各々1 ch 目と 2 ch 目のマイクロホン信号の DFT，$V_{n,k}$ は音源信号の DFT，$N_{x,n,k}$ と $N_{y,n,k}$ は，各々1 ch 目と 2 ch 目のマイクロホン信号中の雑音成分の DFT である．

2.2 コヒーレンス検出を用いた MUSIC

ここで，上の 2 ch 信号から k 番目の周波数の MUSIC スペクトルを方向 $d = (\theta, \varphi)$ の関数として

$$P_{n,k}(d) = \frac{|a_k(d)|^2}{|a_k(d)^{\mathrm{H}} u_{n,k}|^2}, \tag{2}$$

$$a_k(d) = \{H_{x,k}(d), H_{y,k}(d)\}^{\mathrm{T}} \tag{3}$$

によって計算する [2]．上式の $u_{n,k}$ は，入力信号の相関行列

$$R_{n,k} = \begin{bmatrix} \overline{X_{n,k}X_{n,k}^*} & \overline{X_{n,k}Y_{n,k}^*} \\ \overline{Y_{n,k}X_{n,k}^*} & \overline{Y_{n,k}Y_{n,k}^*} \end{bmatrix} \tag{4}$$

の固有値展開における小さい方の固有値に対応した固有ベクトルである．ここで，$\overline{\ }$ は時間平均を表す．

マイクロホンが 2 個の場合，一つの周波数において MUSIC で推定できる音源数は一つであるが，音源信号がスパース性を有する場合は，複数の周波数成分にわたって累積した空間スペクトルから複数の音源方向が推定可能である．その際，複数の音源の成分が重畳した周波数成分，すなわち，スパース性から外れた成分は，正確な推定に寄与しない可能性が高いため，固有値の広がりを用いたコヒーレンス検出 [7], [18], [19] によって 1 個の音源による寄与が支配的であるような周波数成分を選択する．コヒーレンス検出を用いた時間・周波数累積 MUSIC スペクトルを次式で計算する．

$$S_{\mathrm{MUSIC}}(d) = \sum_n \sum_k c_{n,k} \ \log[P_{n,k}(d)], \tag{5}$$

$$c_{n,k} = \begin{cases} 1 & \text{if } \frac{e_{L,n,k}}{e_{S,n,k}} > c_{th} \\ 0 & \text{else} \end{cases} \tag{6}$$

ここで，$e_{L,n,k}$ と $e_{S,n,k}$ は，各々，$R_{n,k}$ の大きい方と小さい方の固有値，c_{th} は，固有値の比に関するしきい値である．

まえがきで述べたように，頭部伝達関数の場合，単一周波数では離れた方向の伝達関数が近い値をもつことが原因となって式 (2) の空間スペクトル上に大きな擬似ピークが出現することがある．擬似ピークの出現過程は音源ピークと同じであるため，周波数ごとに擬似ピークかどうかを判定するのは困難であり，複数の周波数成分にわたる処理が必要となる．しかし，MUSIC のピークは，零に近い値で除算するときの発散により生じるため，そのまま累積した場合は平均化による擬似ピークの低減効果が低いことから，上式のように，対数化後に累積することとした．音源ピークは複数の周波数で同じ方向に現れるため，上式によって，ほぼ音源に起因するピークのみを得ることができる．

2.3 MUSIC における提案逐次法

次に，提案法の処理について述べる．まず，式 (5) において，最大ピークの方向

$$d_1 = \arg[\max_d S_{\mathrm{MUSIC}}(d)]. \tag{7}$$

を 1 番目の音源方向であるとみなす．これが可能であるためには，複数の音源が存在するときに最大ピーク方向が音源方向の一つと一致していることが必要であるが，MUSIC の場合，高い頻度で音源方向の一つに対応していることが実験的に確かめてある [11]．なお，最小分散法はこの段階の性能が低いため，ここでは逐次法の適用対象とはしなかった．

ここで，方向 d_1 におけるピークを減衰させるように，d_1 における MUSIC スペクトルの周波数成分 $P_{n,k}(d_1)$ から各周波数の重みを計算し，次式の MUSIC スペクトルを新たに求める．

$$S_{\text{MUSIC-ISA}}(d, d_1) = \sum_n \sum_k \frac{c_{n,k}}{P_{n,k}^{\alpha}(d_1)} \log[P_{n,k}(d)] \tag{8}$$

上式に導入した α は重み関数の強度を制御する定数であり，実験的に定めることとする．得られたスペクトルにおける最大ピークの方向を

$$d_2 = \arg[\max_d S_{\text{MUSIC-ISA}}(d, d_1)]. \quad (9)$$

とし，これを 2 番目の音源方向の候補とする．これ以降の音源候補に対しても，同様に処理する．m 番目 $(m \geq 2)$ の候補音源方向を求める一般式は次のようになる．

$$d_m = \arg[\max_d \mathrm{S}_{\text{MUSIC-ISA}}(d, d_1, ..., d_{m-1})], \quad (10)$$

$$S_{\text{MUSIC-ISA}}\ (d, d_1, ..., d_{m-1}) = \sum_n \sum_k \frac{c_{n,k}}{\prod_{i=1}^{m-1} P_{n,k}^{\alpha}(d_i)} \log[P_{n,k}(d)]. \quad (11)$$

この処理では，α の値が小さい場合は必ずしも反復ごとに異なった方向が得られるとは限らず，同一かあるいは近い方向が複数回現れることがある．このときは，既に得られた音源方向とのなす角度が一定値 ϕ_{th} 以下である候補はその音源と同一の音源として扱うこととする．

上述の処理の各反復においては，以下に述べる音源数推定処理を行う．まず，ある回数の反復処理において，異なった M 個の音源方向の候補 d_i $(i = 1, 2, ..., M)$ が求まっているものとし，候補 d_i における空間スペクトルの値 $P_{n,k}(d_i)$ $(i = 1, 2, ..., M)$ の中で最大値を与える方向がその周波数成分の属する音源の方向であるとみなす．すなわち，

$$I_{n,k} = \arg[\max_i P_{n,k}(d_i)] \quad (12)$$

が k 番目の成分の属する音源番号である．これを用いて i 番目の音源に属する成分について時間周波数累積パワーを求め，正規化すると，

$$O_i = \frac{\sum_n \sum_k (|X_{n,k}|^2 + |Y_{n,k}|^2) g_{n,k} \cdot c_{n,k}}{\sum_n \sum_k (|X_{n,k}|^2 + |Y_{n,k}|^2) c_{n,k}} \quad (13)$$

$$g_{n,k} = \begin{cases} 1 & \text{if } I_{n,k} = i \\ 0 & \text{else} \end{cases} \quad (14)$$

となる．上式の相対音源パワーは，反復回数が大きいときに現れた音源ほど小さくなる傾向があり，新しく求まった候補の音源パワーがあらかじめ決めた相対パワーのしきい値 p_{stop} 以下となったときに反復を終了するようにする．以降，音源数も並行に推定する上述の方法を MUSIC-ISA-NOS と記すこととする．次節では音源数を既知とする場合についても評価するが，この場合は，音源個数分の異なった方向が現れた時点で反復を終了するものとし，この方法を MUSIC-ISA と記すこととする．

3. 性能評価

3.1 比較に用いた方法

提案法の評価のため，コヒーレンス検出を用いた MUSIC の通常の方法と重み付きウィーナー利得に基づいた逐次推定法を対象として性能比較する．ところで，提案法は，音源方向が 1 個推定されるごとに推定に使う周波数成分を絞っていく処理であることから，用いる成分が音源ごとに既知であるとして方向推定する場合がこの方法における理想的な場合に相当するといえる．そこで，周波数成分ごとの各音源に関する正しい信号対雑音比 (S/N) に基づいて成分選択する推定法を理想バイナリーマスク処理と呼ぶこととし，これについても検討を行った．

3.1.1 重み付きウィーナー利得に基づく逐次推定

比較に用いる WWG-ISA においては，音源数は既知とし，m 番目の音源方向は次式によって推定する [1].

$$d_m = \arg[\max_d \mathrm{S}_{\text{WWG-ISA}}(d, d_1, ..., d_{m-1})]. \quad (15)$$

$$S_{\text{WWG-ISA}}(d, d_1, ..., d_{m-1}) = \frac{\sum_n \sum_k \mathrm{Re}[G_{xy,n,k}(d)] \Psi_{n,k}(d) \Phi_{n,k}(d) \prod_{i=1}^{m-1} \Upsilon_{n,k}(d_i)}{\sum_n \sum_k G_{zz,n,k}(d) \Psi_{n,k}(d) \prod_{i=1}^{m-1} \Upsilon_{n,k}(d_i)}. \quad (16)$$

ここで，

$$G_{xy,n,k}(d) = \overline{Y_{n,k} X_{n,k}^*} H_{y,k}^{-1}(d) [H_{x,k}^{-1}(d)]^*, \quad (17)$$

$$\Psi_{n,k}(d) = 1/G_{dd,n,k}^{\beta}(d), \quad (18)$$

$$\Phi_{n,k}(d) = \mathrm{Max}\left[1 - \frac{\gamma G_{dd,n,k}(d)}{|G_{xy,n,k}(d)|},\ 0\right], \quad (19)$$

$$\Upsilon_{n,k}(d_i) = \mathrm{Min}\left[\left(\frac{G_{dd,n,k}(d_i)}{|G_{xy,n,k}(d_i)|}\right)^{\mu},\ 1\right] \quad (20)$$

$$G_{dd,n,k}(d) = \overline{|X_{n,k} H_{x,k}^{-1}(d) - Y_{n,k} H_{y,k}^{-1}(d)|^2}, \quad (21)$$

$$G_{zz,n,k}(d) = \overline{|[X_{n,k} H_{x,k}^{-1}(d) + Y_{n,k} H_{y,k}^{-1}(d)]/2|^2}, \quad (22)$$

であり，Re[ ] は実部をとる操作，$*$ は複素共役，Max[ ] は [ ] 内の値のうち大きい方を選択する操作，Min[ ]

は [ ] 内の値のうち小さい方を選択する操作である．また，β，γ，μ は，各々，重み関数 $\Psi_{n,k}(d)$，$\Phi_{n,k}(d)$，$\Upsilon_{n,k}(d)$ の強度を調節する定数であり，実験的に定める．なお，上式に含まれる逆伝達関数 $H_{x,k}^{-1}(d)$，$H_{y,k}^{-1}(d)$ は，$H_{x,k}(d)$，$H_{y,k}(d)$ の周波数特性に谷がある場合，性能劣化の原因となる可能性があるが，評価に用いる帯域においては顕著な谷が存在しないため，特に対策は施していない．

3.1.2 理想バイナリーマスク

理想バイナリーマスク処理においては，周波数成分に占める各音源成分の S/N に基づき，音源ごとにどの周波数成分を用いるかの対応を決めて方向推定を行う．すなわち，MUSIC の場合，m 番目の音源方向は，

$$S_{\text{MUSIC-BM}_m}(d) = \sum_n \sum_k b_{m,n,k} \log[P_{n,k}(d)] \tag{23}$$

の最大ピークの方向とし，WWG の場合は，

$$S_{\text{WWG-BM}_m}(d) = \frac{\sum_n \sum_k \text{Re}[G_{xy,n,k}(d)] \Psi_{n,k}^{\beta}(d) b_{m,n,k}}{\sum_n \sum_k G_{zz,n,k}(d) \Psi_{n,k}^{\beta}(d) b_{m,n,k}} \tag{24}$$

の最大ピークの方向とする．ここで，

$$b_{m,n,k} = \begin{cases} 1 & \text{if } \frac{P_{m,n,k}}{Z_{n,k} + \sum_{i \neq m} P_{i,n,k}} > \nu_{th} \\ 0 & \text{else} \end{cases} \tag{25}$$

は，第 n フレームにおいて，m 番目の音源に関する空間スペクトル計算に k 番目の成分を使うか否かを決める理想バイナリーマスク，$P_{m,n,k}$ は m 番目の音源のパワー，$Z_{n,k}$ は背景雑音のパワー，ν_{th} は 2 値化のためのしきい値である．$P_{m,n,k}$ と $Z_{n,k}$ は，複数音源の状況と同じ音源が各々単独で存在した場合，及び，雑音のみ存在した場合をシミュレートして計算する．MUSIC の場合，コヒーレンス検出による重み $c_{n,k}$ が理想バイナリーマスク $b_{m,n,k}$ の効果を分かりにくくするため，これを $b_{m,n,k}$ で置き換えた．一方，WWG の場合は $\Phi_{n,k}(d)$ が目的音源の S/N に相当する重みであり，$\Upsilon_{n,k}(d)$ が既に推定した音源による寄与分を減衰させる重みであるため，これらをまとめて $b_{m,n,k}$ で置き換えた．以降，式 (23) による方法を MUSIC-BM，式 (24) による方法を WWG-BM，また，式 (5) による従来法を MUSIC-CD と記すこととする．

表 1　HRTF の測定方向
Table 1　Directions measured for HRTF.

elevation ($-60^\circ \leq \varphi \leq 90^\circ$)	azimuth step
$+90^\circ$	360°
$+85^\circ, +80^\circ$	30°
$+75^\circ, +70^\circ$	15°
$+65^\circ, \pm 60^\circ$	10°
$-55^\circ - +55^\circ$	5°

3.2 評価実験のセットアップ

3.2.1 頭部伝達関数

想定する伝達系として，以前に測定した 4 体の頭部伝達関数 [1] を用いた．表 1 に示すように，測定した仰角は $-60^\circ \leq \varphi \leq 90^\circ$ の範囲で 5° おきであり，方位角は $-180^\circ \leq \theta \leq 180^\circ$ の範囲で仰角に対応して刻み角が異なる．ここで，正面方向は，$\theta = 0^\circ$，$\varphi = 0^\circ$ である．性能評価シミュレーションにおける観測信号の生成の際は，実際に測定した方向の中からランダムに音源方向を選ぶようにしてあり，一方，方向推定処理の際は，測定していない方向に関しては HRTF の線形補間によって処理している．

なお，システムの使用場所によっては残響に対する性能が重要となるが，残響については本論文の範囲を超えるため，検討の対象とはしていない．以降のシミュレーションにおいては，直接音のみが HRTF による変形を受けて受音されるものと仮定している．

3.2.2 評 価 尺 度

様々な音源方向と信号内容に対する平均的な方向推定性能を求めるため，次式の音源検出率 (Source Detection Rate (SDR)) を性能評価の尺度とした．

$$SDR(M_s, e_p) = K_{\text{success}}(M_s, e_p) / K_{\text{total}}(M_s). \tag{26}$$

上式において，M_s は音源数，e_p は推定方向の許容誤差，$K_{\text{total}}(M_s)$ は方向推定の試行回数，$K_{\text{success}}(M_s, e_p)$ は，成功した試行回数である．以前の実験結果 [11] から，性能差が現れやすいと考えられる $e_p = 5^\circ$ を用い，各試行において得られた M_s 個の音源方向推定値と正しい音源方向との角度差がすべて許容誤差内にある場合に成功と判定した．音源数 M_s は 1 から 3 とし，音源方向は試行ごとに一様乱数を用いて生成した．以降のシミュレーションにおいて，SDR を求めるための試行回数は，各頭部伝達関数あたり 500 とした．頭部伝達関数は 4 セット用いたため，$K_{\text{total}}(M_s)$ $(M_s = 1, 2, 3) = 2000$ である．

3.2.3 音声データ

評価実験に用いる音源の音声として，電総研音声データベース「ETL-WD I & II」中の 10 人の男性の単語発声音声を用いた．このデータは標本化周波数が 16 kHz であるため，測定した頭部伝達関数の標本化周波数の 48 kHz に合わせるためにアップサンプリングした．各話者の 492 単語の発声データを話者別に接続し，更に，パワーに基づいて無音区間を除去し，話者ごとの長時間音声信号を作成した．信号の時間長は 1 音源当り約 230 秒となった．また，背景雑音として用いるため，多数の計算機のある部屋で計算機の冷却ファンの雑音を頭部モデルの一つを用いてバイノーラル録音した．次項で述べるように，評価に用いる周波数帯域は 260 Hz から 4 kHz としたため，音源の S/N 計算の簡単化を考慮し，音声と雑音はいずれも，301 点の FIR フィルタにより 260 Hz から 4 kHz の帯域に制限した．

音源検出率計算においては，試行ごとに前述の 10 人分の長時間信号を順に各音源に割り当て，割り当てた長時間信号から 2 秒分の区間をランダムに切り出して音源信号とした．背景雑音についても長時間の録音データから試行ごとに 2 秒分の区間をランダムに選んだ．切り出した音源の区間信号に対し，まず，正面方向の頭部伝達関数をフィルタリングして頭部正面に到来した信号を生成し，この信号の S/N が所定の値となるように音源区間信号の振幅を調整した．この後，各音源方向の頭部伝達関数をフィルタリングして両耳の受音音声を生成し，背景雑音に重畳した．また，複数音源の場合，各音源の区間信号は等パワーとなるようにした．

3.2.4 分析条件

対象とするすべての方法において，標本化周波数は 48 kHz，FFT 点数は 2048 点，フレーム周期は 1024 点，時間窓は Hanning 窓，評価に用いる周波数帯域は 260 Hz から 4 kHz，空間スペクトルの時間平均化のための時間長は 2 秒とした．また，予備実験から，$R_{n,k}$ の成分スペクトル，及び，WWG-ISA に用いる時間平均スペクトル（式 (17)，(21)，(22)）のための時間平均フレーム数は 5 (128 ms)，WWG-ISA の定数 β，γ，μ については，各々，$\beta = 1$，$\gamma = 1$，$\mu = 1.2$ とした．また，式 (6) のコヒーレンス判定のためのしきい値 c_{th} は，後に述べる検討結果から 20 dB とした．

MUSIC-ISA に関しては，重み関数の指数 α を 0.06，求まった音源候補を同一の音源とみなすか否かを決める際の音源間角度 ϕ_{th} を 5° とした．また，反復終了のしきい値 p_{stop} は，予備実験より −16 dB としたが，p_{stop} を下回るのが非常に遅い場合もあることを考慮し，反復回数が 20 に達した場合は処理を打ち切ることとした．

表 2　音源と空間スペクトルの角度範囲
Table 2　Angular ranges for spatial spectrum calculation and source direction set.

	spectrum calculation	source direction sets
azimuth	$-180^\circ \le \theta \le 180^\circ$	$-180^\circ \le \theta \le 180^\circ$
elevation	$-60^\circ \le \varphi \le 90^\circ$	$-55^\circ \le \varphi \le 90^\circ$
	step = 5°	min. separation = 10°

音源の存在を仮定する角度範囲と空間スペクトルを計算する角度範囲は，表 2 に示すとおりとした．また，複数音源を仮定する場合は，音源間の角度が 10° 未満にならないように制限した．

4. シミュレーション結果

4.1 理想バイナリーマスク処理

最初に，理想バイナリーマスクに基づいた方向推定の結果について述べる．この処理は，2 値化のためのしきい値 ν_{th}（式 (25)）が処理に使う周波数成分を決定し，推定精度を変化させるため，ν_{th} を −50 dB から 50 dB の範囲で変えながら，式 (23) の MUSIC-BM と式 (24) の WWG-BM の音源検出率 (SDR) を求めた．この結果を図 1 に示す．図 1 は，音源が 3 個，S/N が 10 dB のときと 0 dB のときの結果であり，実線は MUSIC-BM，点線は WWG-BM の結果を示す．

図 1 を見ると，MUSIC-BM の場合，S/N が 10 dB のときは $\nu_{th} = -20$ から 30 dB の範囲で SDR が最大，かつ，一定になり，S/N が 0 dB では，最大性能を示す範囲が少し狭まり，$\nu_{th} = -20$ から 25 dB の範囲となる．この結果は，MUSIC は高 S/N の成分だけでも最大性能を得られるが，しきい値を高くし過ぎると推定に使える成分が少なくなって全く推定できない試行が増え，SDR は低下することを表している．一方，WWG-BM の場合，S/N が 10 dB のときも 0 dB のときも $\nu_{th} = -20$ dB 付近で SDR が最大となることから，S/N の非常に低い成分以外はすべて使う方がよいことが分かる．また，同じ S/N では，MUSIC-BM の方が WWG-BM よりも高い SDR を達成でき，例えば，$S/N = 10$ dB のとき，MUSIC-BM の SDR の最大値は約 93%，WWG-BM の SDR の最大値は約 82%であった．したがって，MUSIC の方が潜在的な

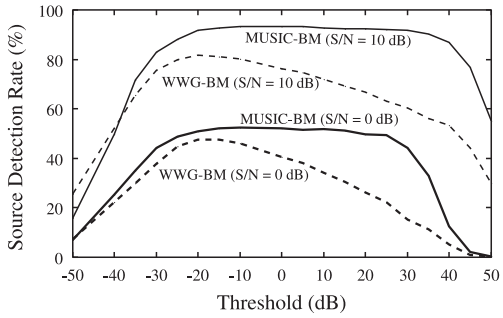


図 1　理想バイナリーマスク法による音源検出率の成分選択しきい値 ν_{th} による変化（音源 3 個のとき）

Fig. 1　Source Detection Rate (SDR) obtained by ideal binary mask methods vs. threshold ν_{th} ($M_s = 3$).

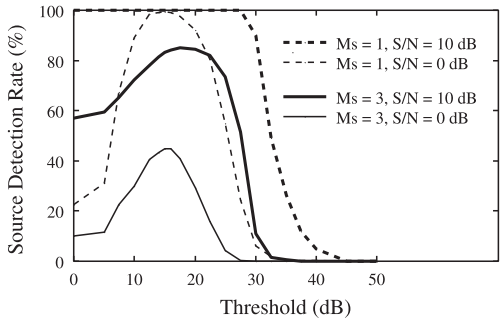


図 2　MUSIC-CD（音源 1 個の場合）と MUSIC-ISA（音源 3 個の場合）の音源検出率のコヒーレンス検出しきい値 c_{th} による変化

Fig. 2　Source Detection Rate (SDR) obtained by MUSIC-CD ($M_s = 1$) and MUSIC-ISA ($M_s = 3$) vs. coherence detection threshold c_{th}.

性能は高いといえる．

4.2　コヒーレンス検出しきい値

次に，音源数を既知として処理する MUSIC-ISA に関して，コヒーレンス検出のしきい値 c_{th} を 0 dB から 50 dB まで変化させながら同様に SDR を求めた．この結果を図 2 に示す．なお，音源数 $M_s = 1$ のときは，MUSIC-ISA と MUSIC-CD は同一である．また，しきい値 c_{th} は，小さい方の固有値に対する比であるため，$c_{th} \geq 0$ dB である．図 2 において，点線は音源数 $M_s = 1$，実線は $M_s = 3$ のときの結果である．この図から，MUSIC-ISA のコヒーレンス検出は，理想バイナリーマスクと似たような働きをし，$S/N = 0$ dB と 10 dB の場合の両方ともしきい値が 15 dB から 20 dB あたりで性能が最大になっていることが分かる．しかし，コヒーレンス検出に音源を区別する働きはないため，SDR の値は理想バイナリーマスク処理ほどは高くなく，$M_s = 3$，$S/N = 10$ dB のときの最大値は約 83%であった．

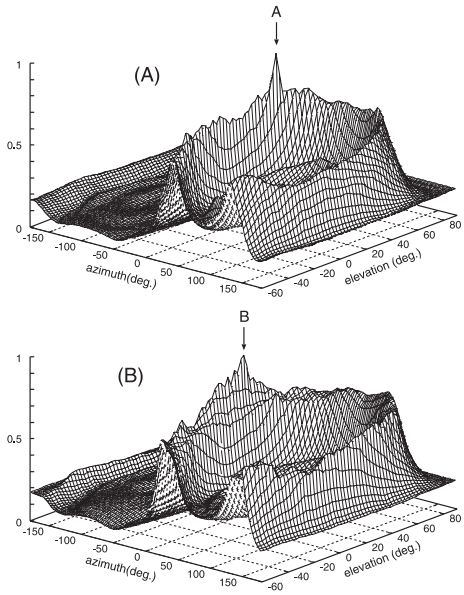


図 3　MUSIA-ISA における空間スペクトル，(A) 初期スペクトル，(B) 音源 A の成分減衰後のスペクトル（$S/N = 10$ dB，音源数 2，音源 A $(20°, 30°)$，音源 B $(10°, 10°)$）

Fig. 3　Spatial spectrum obtained from MUSIC-ISA, (A) initial spectrum, (B) source A is attenuated ($S/N = 10$ dB, source A is present at $(20°, 30°)$, source B is at $(10°, 10°)$).

4.3　MUSIC-ISA の空間スペクトル

ここでは，定量的な評価の前に，MUSIC-ISA によって得られる空間スペクトルについて検討した．2 個の音源 A，B が存在して音源 A が $(20°, 30°)$，音源 B が $(10°, 10°)$ に位置するときに得られた空間スペクトルを図 3 (A)，(B) に示す．$S/N = 10$ dB であり，(A)，(B) とも最大値で正規化してある．図 3 (A) は，式 (5) の初期スペクトルであり，図において矢印で示した位置 A に音源 A があり，この位置に最大ピークが現れている．一方，図 3 (B) は，式 (11) によって最大ピークの音源を減衰させたスペクトルであり，音源 A のピークは消え，音源 B のピークが現れている．図 3 (A) の初期スペクトルにおいては，音源 B のピークが全く現れていないことから，初期スペクトルからの複数の音源ピーク検出は困難であり，MUSIC-ISA の処理によって埋もれている音源ピークが浮かび上がる様子が見て取れる．

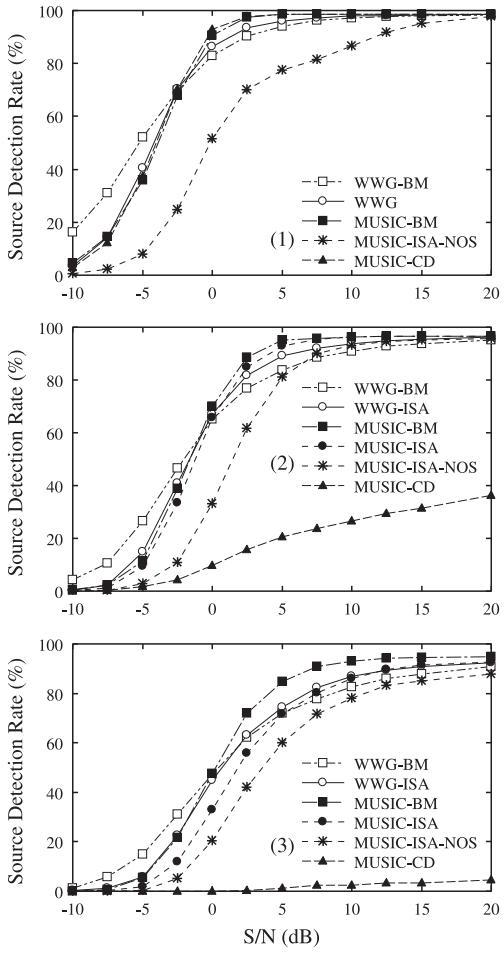


図 4　音源検出率の S/N による変化（(1) 音源 1 個，(2) 音源 2 個，(3) 音源 3 個）

Fig. 4　Source Detection Rate (SDR) vs. S/N. ((1) single source, (2) two sources, (3) three sources)

4.4　MUSIC-ISA の音源検出率

次に，提案法の性能を WWG に基づく方法，及び，従来の MUSIC 法と比較するため，S/N を変えながら SDR を求めた．得られた結果を図 4 (1)，(2)，(3) に示す．図において，(1)，(2)，(3) は，各々，音源数が 1 個，2 個，3 個のときの結果である．従来法の MUSIC-CD と理想バイナリーマスク処理の結果も同じ図に示してある．前節の検討から，MUSIC-ISA，MUSIC-ISA-NOS，MUSIC-CD におけるコヒーレンス検出のしきい値 c_{th} を 20 dB，MUSIC-BM と WWG-BM のバイナリーマスクしきい値 ν_{th} を各々，20 dB，-20 dB とし，MUSIC-ISA-NOS 以外の方法では，音源数は既知とした．

まず，音源数 $M_s = 1$ のとき（図 4 (1)），MUSIC-CD は WWG とほぼ同等であるが，MUSIC-ISA-NOS は，$S/N < 15$ dB において S/N 低下に伴う性能低下が大きく，MUSIC-CD と比べると，$S/N = -2.5$ dB のときに約 40%の検出率低下になっていることが分かる．また，理想バイナリーマスク処理については，S/N が -2.5 dB 以下では WWG-BM が，それ以上では MUSIC-BM の方が性能が高いが，その差はわずかである．

$M_s = 2$ のとき（図 4 (2)），MUSIC-ISA は WWG-ISA とほぼ同等であるが，MUSIC-ISA-NOS に関しては，$S/N = 10$ dB 以上で MUSIC-ISA と同等の性能であり，S/N の低下に伴う MUSIC-ISA との検出率の差は，最大で $S/N = 0$ dB のときの約 30%であり，$M_s = 1$ のときよりは差が小さいことが分かる．従来法の MUSIC-CD は，大幅に性能が低下し，$S/N = 10$ dB のときの SDR は，MUSIC-ISA が 96%，MUSIC-ISA-NOS が 93%であるのに対し，MUSIC-CD は 26%となった．

$M_s = 3$ のとき（図 4 (3)）は，低 S/N において WWG-ISA の方が MUSIC-ISA よりも若干 SDR が高いが，それほど顕著な差はなかった．また，MUSIC-ISA-NOS は，全体に MUSIC-ISA よりも性能が低いが，MUSIC-ISA に対する検出率の差は，$S/N = 2.5$ dB のときに最大の 12%程度となっており，音源数が 1 個と 2 個のときほど性能低下していないことが分かる．

MUISC-ISA-NOS の性能は，反復終了のしきい値 p_{stop} の値を大きくする程 $M_s = 1$ のときの性能は高く，逆に $M_s = 3$ のときの性能は低くなる．これは，複数音声の重畳した信号に対する時間分解能を考慮して相関行列推定に使う信号長を小さくしてあることと，音源 1 個の場合，雑音の支配的な成分が多いことから，雑音の支配的な周波数成分がコヒーレンス検出において残る場合があり，反復処理によって音源ピークが減衰されると，このような成分に起因するピークが候補として出やすい状況になるためと考えられる．したがって，雑音成分除去により，音源 1 個のときの MUISC-ISA-NOS の性能を改善できると考えられる．

MUSIC-CD，MUSIC-ISA，MUSIC-ISA-NOS を比較すると，音源数増加に伴う MUSIC-CD の性能低下が著しいのに対し，MUSIC-ISA と MUSIC-ISA-NOS の性能低下は小さく，これらの方が大幅に性能が高い．

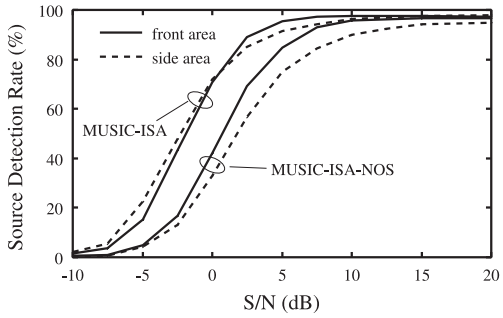


図 5　音源方向による音源検出率の違い（音源 2 個）
Fig. 5　Source Detection Rate (SDR) vs. S/N for two areas of source directions. (two sources)

例えば，音源数 3 個で $S/N = 10\,\mathrm{dB}$ のとき，MUSIC-CD は SDR が約 3%であるのに対し，MUSIC-ISA は約 83%，MUSIC-ISA-NOS は約 78%となった．

4.5　推定性能の方向依存性

前節では，HRTF の測定範囲全体を対象として音源検出率を求めたが，HRTF においては，側方よりも正面付近の方が，到来方位の違いに対するチャネル間の時間差の変化が大きいため，正面付近の方が方向推定精度が高いと思われる．そこで，本節では，正面付近と側方における推定性能の違いについて検討した．性能評価のため，正面付近については，$-25^\circ \leq \theta \leq 25^\circ$，$-25^\circ \leq \varphi \leq 25^\circ$，側方については，$65^\circ \leq \theta \leq 115^\circ$，$-25^\circ \leq \varphi \leq 25^\circ$，の各々の角度範囲について，音源が 2 個存在する場合の音源検出率を求めた．この結果を図 5 に示す．

図 5 は，前節と同様，S/N を $-10\,\mathrm{dB}$ から $20\,\mathrm{dB}$ まで変化させたときの MUSIC-ISA と MUSIC-ISA-NOS の音源検出率である．図中の実線は正面方向，点線は側方に関する結果である．正面と側方の結果を比べると，MUSIC-ISA についてはほぼ同等であり，MUSIC-ISA-NOS については正面方向の場合の方が若干性能が高かったが，正面と側方における SDR の差は，差の大きい $S/N = 5\,\mathrm{dB}$ 付近で約 9%であり，それほど顕著な差とはならなかった．

5. むすび

理想バイナリーマスク処理とコヒーレンス検出しきい値を変えた場合の MUSIC 法に関する検討結果から，高い音源検出率を得るためには，MUSIC においては 15 dB から 20 dB 以上の S/N の高い成分だけを用いる方がよく，逆に，相互相関ベースの WWG の場合には，高 S/N の成分だけでは高性能とはならず，$-20\,\mathrm{dB}$ 以上の周波数成分をすべて用いる方がよいことが分かった．次に，提案法の MUSIC-ISA と MUSIC-ISA-NOS を既提案の WWG-ISA，及び，従来法の MUSIC-CD と比較した結果，複数音源の環境において MUSIC-ISA と MUSIC-ISA-NOS は MUSIC-CD よりも大幅に高い性能となり，MUSIC-ISA と WWG-ISA はおおむね同等の性能となることが分かった．また，音源数推定を同時に行う MUSIC-ISA-NOS は，音源数を既知として処理する MUSIC-ISA と比較すると，音源が 1 個の場合は $S/N = 10\,\mathrm{dB}$ 以下における性能差が大きかったが，音源が 2 個と 3 個の場合，$S/N = 10\,\mathrm{dB}$ 以上で MUSIC-ISA と大差ない性能が得られ，いずれも，従来法の MUSIC-CD より大幅に高い性能となった．以上の結果から，本提案法の有効性が確認できた．

文　献

[1] Y. Nagata, S. Iwasaki, T. Hariyama, T. Fujioka, T. Obara, T. Wakatake, and M. Abe, “Binaural localization based on weighted Wiener gain improved by incremental source attenuation,” IEEE Trans. Audio Speech Language Process., vol.17, no.1, pp.52–65, Jan. 2009.

[2] R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propag., vol.AP-34, no.3, pp.276–280, March 1986.

[3] 石井カルロス寿憲，シャトッ・オリビエ，石黒　浩，萩田紀博，“実環境における music 法を用いた 3 次元音源定位の評価，” 人工知能学会研究会資料，pp.27–32, SIG-Challenge-A802-5, Nov. 2008.

[4] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-24, no.4, pp.320–327, Aug. 1976.

[5] S.U. Pillai, ed., Array signal processing, Springer-Verlag, 1989.

[6] 菊間信良（編），アレーアンテナによる適応信号処理，科学技術出版，1998.

[7] S. Mohan, M.L. Kramer, B.C. Wheeler, and D.L. Jones, “Localization of nonstationary sources using a coherence test,” Proc. IEEE Workshop on Statistical Signal Processing, pp.470–473, 2003.

[8] S. Mohan, M.E. Lockwood, M.L. Kramer, and D.L. Jones, “Localization of multiple acoustic sources with small arrays using a coherence test,” J. Acoust. Soc. Am., vol.123, no.4, pp.2136–2147, April 2008.

[9] 谷川真一，浜田　望，“2 チャネルマイクロホンアレーの仮想多チャネル化による音声の到来方向推定法，” 信学論（A），vol.J85-A, no.2, pp.153–161, Feb. 2002.

[10] Y. Nagata, T. Fujioka, and M. Abe, “Speech enhancement based on auto gain control,” IEEE Trans. Audio

Speech Language Process., vol.14, no.1, pp.177–190, Jan. 2006.
[11] Y. Nagata, F. Fujioka, and M. Abe, “Two-dimensional doa estimation of sound sources based on weighted Wiener gain exploiting two directional microphones,” IEEE Trans. Audio Speech Language Process., vol.15, no.2, pp.416–429, Feb. 2007.
[12] 浅野　太，“ロボットにおける音源位置推定，” 音響誌，vol.63, no.1, pp.41–46, Jan. 2007.
[13] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, “Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones,” Acoust. Sci. and Tech., vol.22, no.2, pp.149–158, 2001.
[14] N. Roman, “Binaural segregation in multisource reverberant environments,” J. Acoust. Soc. Am., vol.120, no.6, pp.4040–4051, Dec. 2006.
[15] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Doa estimation for multiple sparse sources with normalized observation vector clustering,” Proc. ICASSP2006, pp.V-33–V-36, 2006.
[16] 山本　潔，浅野　太，山田剛志，北脇信彦，“音源分離における svm を用いた音源数推定について，” 信学技報，EA2002-4, April 2002.
[17] A. Quinlan, “Automatic determination of the number of targets present when using the time reversal operator,” J. Acoust. Soc. Am., vol.119, no.4, pp.2220–2225, April 2006.
[18] D. Wang and G.J. Brown, ed., Computational audio scene analysis, John Wiley & Sons, 2006.
[19] C. Faller and J. Merimaa, “Sound localization in complex listening situations,” J. Acoust. Soc. Am., vol.116, no.5, pp.3075–3089, Nov. 2004.

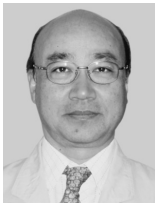
（平成 21 年 2 月 5 日受付，5 月 27 日再受付）

永田　仁史　（正員）

昭 59 東北大・工・電子卒．平 2 同大大学院情報工学専攻博士課程了．工博．同年（株）東芝入社，研究開発センター勤務．平 6 同社関西研究所，平 9 岩手大学工学部講師，平 13 岩手大学工学部准教授．音声認識，ディジタル音響信号処理の研究に従事．日本音響学会，情報処理学会各会員．

岩崎　　聡

1980 三重大・医入学．1986 浜松医科大学耳鼻咽喉科学入局，1998 米国ハウス耳科学研究所留学，2000 浜松医科大学講師．2007 愛知医科大学教授．2008 浜松赤十字病院部長，聖隷クリストファー大学客員教授．

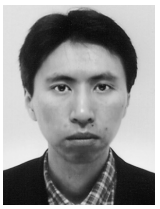


針山　孝彦

昭 58 東北大学大学院医学研究科博士課程退学．理博．昭 58 東北大学応用情報学研究センター・改組により平 5 同大・大学院情報科学研究科・助手．平 13 浜松医科大学医学部・助教授．平 16 同大医学部・教授．動物生理学，光生物学，神経行動情報学の研究に従事．日本動物学会，日本比較生理生化学会，日本応用動物昆虫学会，日本進化学会各会員．

堀口　弘子

平 8 静岡大・理・生物卒．平 10 同大大学院生物地球環境科学専攻博士前期課程了．同年浜松医科大学生物学教室教務補佐員，平 12 浜松医科大学生物学教室教務員．無脊椎動物の感覚生理学の研究に従事．日本動物学会，日本進化学会各会員．



藤岡　豊太　（正員）

平 4 秋田大・鉱山・電気工学卒．平 6 同大大学院修士課程了．平 9 東北大学大学院電気・通信工学専攻博士後期課程了．平 9 岩手大学工学部助手．ディジタル音響信号処理，能動騒音制御に関する研究に従事．情報処理学会会員．

安倍　正人　（正員）

昭 56 東北大学大学院電気及び通信工学専攻博士課程了．工博．昭 58 東北大学情報処理教育センター助手．平元東北大学大型計算機センター助教授．平 8 岩手大学工学部情報工学科教授．ディジタル信号処理の音響，振動への応用に関する研究に従事．IEEE，ACM，米国音響学会，日本音響学会，日本騒音制御学会，日本機械学会，情報処理学会各会員．