

論 文

二つの指向性マイクロホンを用いた目的音検出に関する検討

永田 仁史^{†a)} 藤岡 豊太[†] 安倍 正人[†]

Target Signal Detection System Using Two Directional Microphones

Yoshifumi NAGATA^{†a)}, Toyota FUJIOKA[†], and Masato ABE[†]

あらまし 本論文では、2個の指向性マイクロホンを用いた目的音検出方法を提案する。本検出処理は、角度をずらして向き合うように配置した2個の指向性マイクロホンを用い、各マイクロホン出力のDFTスペクトル間の位相差とパワー比から検出パラメータを計算する。位相差としては、2チャネル間のクロススペクトルの位相の絶対値の平均を用い、パワー比としては、各チャネルのDFTスペクトルにおいて周波数成分ごとに2ch間のパワーを比較し、大きい方の成分を累積したパワーと小さい方の成分を累積したパワーの比を用いる。このパワー比はマイクロホンの指向性と各成分ごとのパワー比により強調されるため、二つのマイクロホンに同振幅で同時に到達する信号であるか否かを従来法より高精度に判別可能である。車において収集した3種類の雑音に対し、従来の2ch音声検出法で提案されているパラメータのほか、DFTに基づいた種々の検出パラメータについて音声検出精度を評価した結果、上記のパワー比と位相差の線形和による検出パラメータが総合的に最も高い検出精度を与える、提案する目的音検出方法が有効であることを確認できた。

キーワード 音声検出、2ch、突発雑音、パワー比、位相差、指向性マイクロホン

1. まえがき

音声検出処理は、音声認識システムや音声符号化処理の前処理として使われ、本体システムの性能を大きく左右するため、検出性能の向上に関して従来から多くの研究がなされている。目的音声対雑音比(S/N)が比較的良好な場合には、パワーを中心として零クロス、特定帯域のパワーなどを組み合せる方法[12]、ケプストラム[11]や線形予測残差[3], [6]を検出パラメータとして使う方法、更に、複数のパラメータを組み合せて使う方法[1]などが提案されている。また、安全上の理由から特に音声入力が期待されるような環境、例えば、走行中の車内や作業場などは室内に比べて S/N が低い場合が多く、また、突発的な雑音や目的外の音声が到来することがあるため、安定した音声検出のためにヘッドセット形などの近接マイクロホンが用いられている。

一方、雑音抑圧のため、複数のマイクロホンを用いたマイクロホンアレー処理が種々提案されており、音声

認識処理と組み合せた報告も多くなされている[18]～[20]。コストと可搬性を考慮すると、なるべく少数のマイクロホンによる処理が望ましいが、少数マイクロホンで高い性能を達成するのに適した適応ビームフォーマ処理[8], [10], [14]やブラインド信号分離[9], [21]～[23]においては、突発音のように雑音の継続長が短い場合に関する検討はほとんどなく、このような環境において、マイクロホンアレーによる雑音除去後の音声に従来の音声区間検出を適用すれば十分であるとは言えない。

これに対し、2個のマイクロホン出力に基づき、雑音抑圧処理を通さずに直接音声区間を検出する方法が提案されており、1ch出力に基づく従来方法に比べ、低 S/N でも高い検出精度を維持できることが報告されている[15]。この方法では、検出パラメータとして2ch間のコヒーレンスと時間差を用いており、 S/N が-数dB程度であっても良好な音声検出結果が得られるとしている。しかしながら、突発雑音のような非定常性の強い雑音に対する評価はなされておらず、先のような環境における実用性には問題がある。

ところで、筆者らは、先に、2ch入力の適応ビームフォーマ処理において空間的エリアシングを抑えるため、指向性マイクロホンの角度を互いにずらして配置

[†] 岩手大学工学部、盛岡市

Department of Computer and Information Science, Iwate University, Morioka-shi, 020-8551 Japan

a) E-mail: nagata@cis.iwate-u.ac.jp

する方法を用いた[16]。この配置においては、到來雑音に位相差だけでなく振幅差が生じるため、目的音検出処理にこれを利用することができる。同様なパラメータを利用した雑音抑圧処理としては、2チャネル間の位相差と振幅差を利用した音源分離処理[17]、あるいは2点間の相関を利用したスペクトルサブトラクション(SS)[7]が提案されているが、音声検出性能に関する検討はなされていない。

そこで、本論文では、2ch信号に基づいた検出パラメータについて更に詳しく検討するとともに、時間差と振幅差を組み合せた検出パラメータを提案し、窓を開けた走行中の車内や工事現場などの雑音環境を対象に検出性能評価を行う。

以下、本論文では、まず、検討に用いる検出パラメータについて述べ、続いて、雑音と単語音声を対象にパラメータの出現頻度に関して検討する。次に2ch信号に基づく目的音検出の性能評価実験を行う。

2. 二つの指向性マイクロホン出力に基づく目的音検出のパラメータ

本論文で用いたマイクロホン配置を図1に示す。図1において、目的音は正面(0°)付近から到來するものとしている。この配置の場合、目的音は同位相、同振幅で出力される一方、正面方向以外から到來する雑音に関しては位相差と振幅差が生ずるため、両者を考慮した検出パラメータが有効であると考えられる。

2ch信号に基づく音声検出法として従来提案されている方法[15]においては、2ch間のコヒーレンスと時間差が同時に所定の条件を満たした場合に目的音が到來したとみなしており、時間差は、コヒーレンスに基づいて重み付けしたクロススペクトルを逆フーリエ変換して得られる generalized cross correlation function[2]から求めている。この処理の場合、スピーカからの放射音など、方向性のある雑音の存在する環境では、コヒーレンスは目的音と同様に高い値となるので、目的音の検出精度は時間差の検出精度に依存することとなる。

これに対し、上述のマイクロホン配置ならば、2ch間の位相差に加えてパワー比も利用できるため、時間差のみによる方法よりも検出精度の向上が期待できる。パワー比に関しては、単純なチャネルごとの信号パワーの比のほか、周波数ごとにチャネル間でパワーを比較し、大きい方の成分同士の累積パワーと小さい方の成分同士の累積パワーの比も検討することとし

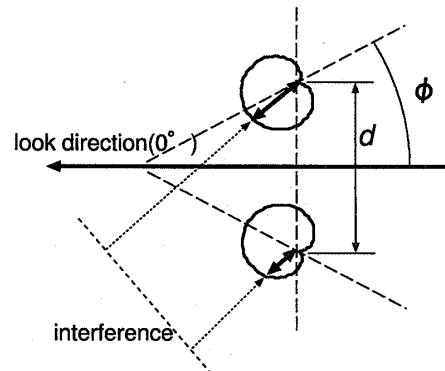


図1 マイクロホン配置
Fig. 1 Microphone arrangement.

た。このパワー比は、周波数ごとのパワー差を埋もれさせずに利用できるため、単純なチャネルごとの信号パワーの比よりも雑音に対して敏感に反応するものと期待できる。

また、目的音が両マイクロホンに同時に到達すると仮定しているため、コヒーレンスに関してはコヒーレンスを求める際の時間方向の平均化に加えて周波数方向の平均化が可能であり、このパラメータも検討に加えた。

以上を考慮し、本論文ではFFT処理に基づいて得られるパラメータに関し、以下にあげる値を検討することとした。ただし下の式において、 n をFFT分析のフレーム番号、 $X_k(n)$ 、 $Y_k(n)$ を各々 n フレーム目のチャネル1、チャネル2のFFTスペクトルの k 番目の周波数成分とする。

(a) 加算パワー (P_{sum}) :

$$P_{sum}(n) = 10 \log \left(\sum_k |X_k(n) + Y_k(n)|^2 \right) \quad (1)$$

(b) SS処理後のパワー (P_{ss}) : 加算スペクトル ($X_k + Y_k$) をSS処理[4]した後の全パワー

(c) コヒーレンス (C_{oh}) : コヒーレンス関数

$$\rho_k^2(n) = \frac{|Wxy_k(n)|^2}{Wxx_k(n)Wyy_k(n)} \quad (2)$$

の平均値。

$$C_{oh}(n) = \frac{1}{M} \sum_k \rho_k(n) \quad (3)$$

論文／二つの指向性マイクロホンを用いた目的音検出に関する検討

ここで、

$$Wxy_k(n) = \overline{X_k^*(n)Y_k(n)},$$

$$Wxx_k(n) = \overline{X_k^*(n)X_k(n)},$$

$$Wyy_k(n) = \overline{Y_k^*(n)Y_k(n)}$$

である。 (— は時間平均)

(d) 時間差 (T_{df})： generalized cross correlation function

$$R_{xy}(\tau, n) = \text{FFT}^{-1}[\psi_k Wxy_k(n)] \quad (4)$$

の最大ピーク位置の絶対値。ここで $\psi_k(n)$ は、

$$\psi_k(n) = \frac{\rho_k^2(n)}{|Wxy_k(n)|(1 - \rho_k^2(n))} \quad (5)$$

である。

(e) コヒーレンス (C_{oh}) と時間差 (T_{df}) をしきい値処理した後の論理積 (C_{oh*tdf})

(f) 位相差 (P_h)： チャネル間クロススペクトルの位相の絶対値平均

$$P_h(n) = \frac{1}{M} \sum_k |\arg(X_k(n)Y_k^*(n))| \quad (6)$$

(g) パワー比 1 (A_{mp1})： チャネル 1 のパワーとチャネル 2 のパワーの比

$$A_{mp1}(n) = \frac{\max(\sum_k |X_k(n)|^2, \sum_k |Y_k(n)|^2)}{\min(\sum_k |X_k(n)|^2, \sum_k |Y_k(n)|^2)} \quad (7)$$

(h) パワー比 2 (A_{mp2})： 各周波数成分のパワーをチャネル間で比較し、大きい方の成分の累積と小さい方の成分の累積の比

$$A_{mp2}(n) = \frac{\sum_k \max(|X_k(n)|^2, |Y_k(n)|^2)}{\sum_k \min(|X_k(n)|^2, |Y_k(n)|^2)} \quad (8)$$

(i) 位相差 (P_h) とパワー比 (A_{mp2}) の線形結合 (P_{h+a})：(f) の位相差 (P_h) と (h) のパワー比 (A_{mp2}) の α 倍の値との和

$$P_{h+a}(n) = P_h(n) + \alpha(A_{mp2}(n) - 1) \quad (9)$$

(j) 全帯域にわたる相関係数 (C_{or1})：

$$C_{or1}(n) = \frac{\sum_k X_k(n)Y_k^*(n)}{\sum_k |X_k(n)|^2 \sum_k |Y_k(n)|^2} \quad (10)$$

(k) パワー比 A_{mp2} の重み付き相関係数 (C_{or2})：

$$C_{or2}(n) = C_{or1} * A_{mp2} \quad (11)$$

上の各式において、* は複素共役、 M は用いた周波数成分の数である。なお、以降の分析においては、FFT の点数 $N = 256$ 、用いた周波数成分 k の範囲は、 $4 \leq k \leq 127$ 、 $M = 124$ 、分析の時間窓は 256 点のハニング窓、フレーム周期は 128 点 (11.6 ms)、サンプリングは 11025 Hz とした。

また、処理の際、上記 (f) から (i) までの各パラメータは 9 次のメジアンフィルタにより平滑化し、(c), (j), (k) の各パラメータは、前後 5 フレーム計 11 フレームの時間平均値とし、(d) の時間差は、ラグ ±10 点までの範囲における最大ピーク位置の絶対値を 3 次のメジアンフィルタにより平滑化して用いた。また、(i) のパラメータにおける結合係数 α は予備実験により 0.6 とした。

以上に述べた 2ch の検出パラメータに関し、次章で述べる実環境雑音を対象に検討を行う。

3. 評価データ

目的音検出性能評価のため、2 チャネルで実環境雑音を収録した。雑音は、車上において収集し、

(N1) 車を停止して収集した工事現場の雑音（窓閉）
(N2) 走行雑音（窓閉）

(N3) 走行中でラジオからの音声の混じった雑音（窓閉）
の 3 種類とした。

マイクロホンは単一指向性（オーディオテクニカ AT9820X）であり、図 2 に示すように、車の助手席側の頭部正面付近に設置した。この配置において、マイクロホン間距離は $d = 12$ cm、傾きの角度は $\phi = 45^\circ$ とした。車はボンネット形のワゴン車であり、車の窓は (N1), (N2) の収集時には開放し、(N3) の収集時には閉鎖した。また、(N3) のラジオ音声は車内前後左右の四つのスピーカから出力した。

収集した雑音データのうち、(N2), (N3) は、100 Hz 以下に車の走行雑音特有のパワーの大きな低域成分を含み、この成分により全体の S/N が音声帯域の S/N より大幅に低い値となる。この場合、検出性能の目安として S/N の値の意味がなくなってしまうこと、また、このような低域は音声認識の前処理にとっては重要性が低いことにより、ここでは三つの雑音ともに 120 Hz 以下をカットした。

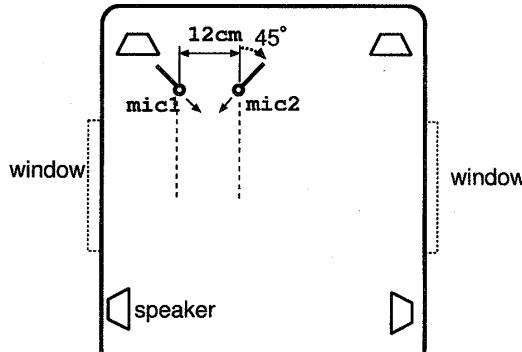


図 2 車内のマイクロホン位置
Fig. 2 Microphone arrangement in the car.

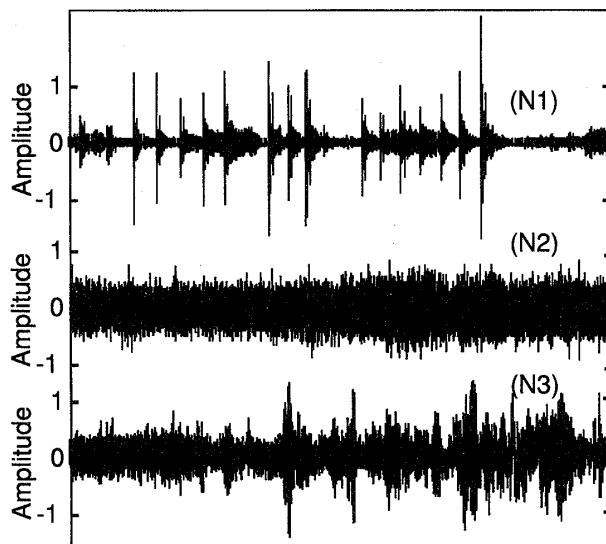


図 3 雑音波形
Fig. 3 Noise signal.

目的音声は、モノラル録音のクリア音声 492 単語 20 人分であり、1 単語の平均発声時間は 0.80 秒、発声間隔の平均は 1.38 秒である。2ch の評価データは、計算機上で両チャネルの雑音に同一の目的音声信号を重畠させることにより作成した。音声データは目視により始終端位置がラベル付けしてあり、雑音重畠の際の S/N は、目視ラベルによる音声区間内の平均パワーと雑音の全データの平均パワーとから求め、-9 dB から 20 dB まで変化させた。

図 3、図 4 に三つの雑音の波形の一部と各々の低域減衰後の平均スペクトルを示す。(N1) はハンマの打撃音などの大振幅波形を含む信号であり、(N2) は全帯域にわたる定常的な雑音、(N3) は音声帯域の非定常な雑音である。

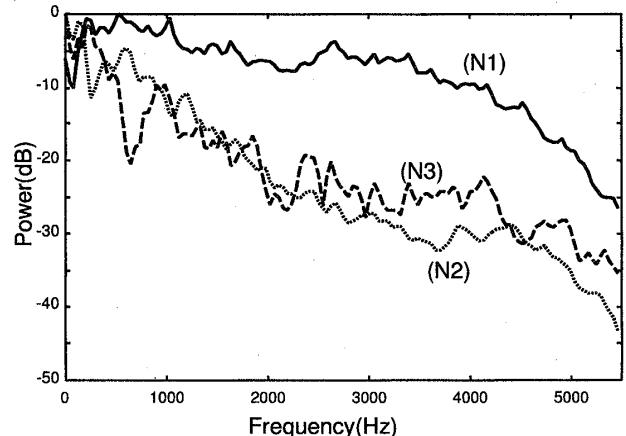


図 4 雜音のスペクトル
Fig. 4 Power spectra of noises.

4. 検出パラメータの出現頻度

4.1 出現頻度

雑音と目的音に対する検出パラメータの振舞いの違いを見るため、目視による音声区間位置に基づき、音声区間における出現頻度と雑音区間における出現頻度を各々求めた。

図 5 に雑音 (N1) の重畠音声から計算した各パラメータの相対頻度分布を示す。各図において、太線が雑音の頻度分布であり、ほかは、目的音声の S/N が -9 dB, 0 dB, 10 dB のときの音声区間の頻度分布である。

パワー (P_{sum}) の頻度分布においては、-9 dB の場合は雑音と音声の分布の重なりが大きく、目的音検出が困難であると思われるが、SS 处理したパラメータ (P_{ss}) では分布の重なりが小さくなり、検出しやすくなると予想できる。

また、コヒーレンス (C_{oh}) の分布も、-9 dB の場合分布の重なりが大きく、雑音と音声区間との区別がつかない。これは雑音中に突発的大振幅部分が多く含まれ、コヒーレンスを求めるための時間平均処理によってその近傍のフレームでコヒーレンスが高くなるためである。その他のパラメータについては、雑音と音声区間との間の出現頻度の分布形状は大きく異なつており、 P_{sum} , C_{oh} よりは検出精度が高くなるものと予想できる。

4.2 脱落/誤警告の頻度

しきい値を設定して目的音を検出する際は、目的音区間を雑音と判断する脱落と、雑音区間を目的音であ

論文／二つの指向性マイクロホンを用いた目的音検出に関する検討

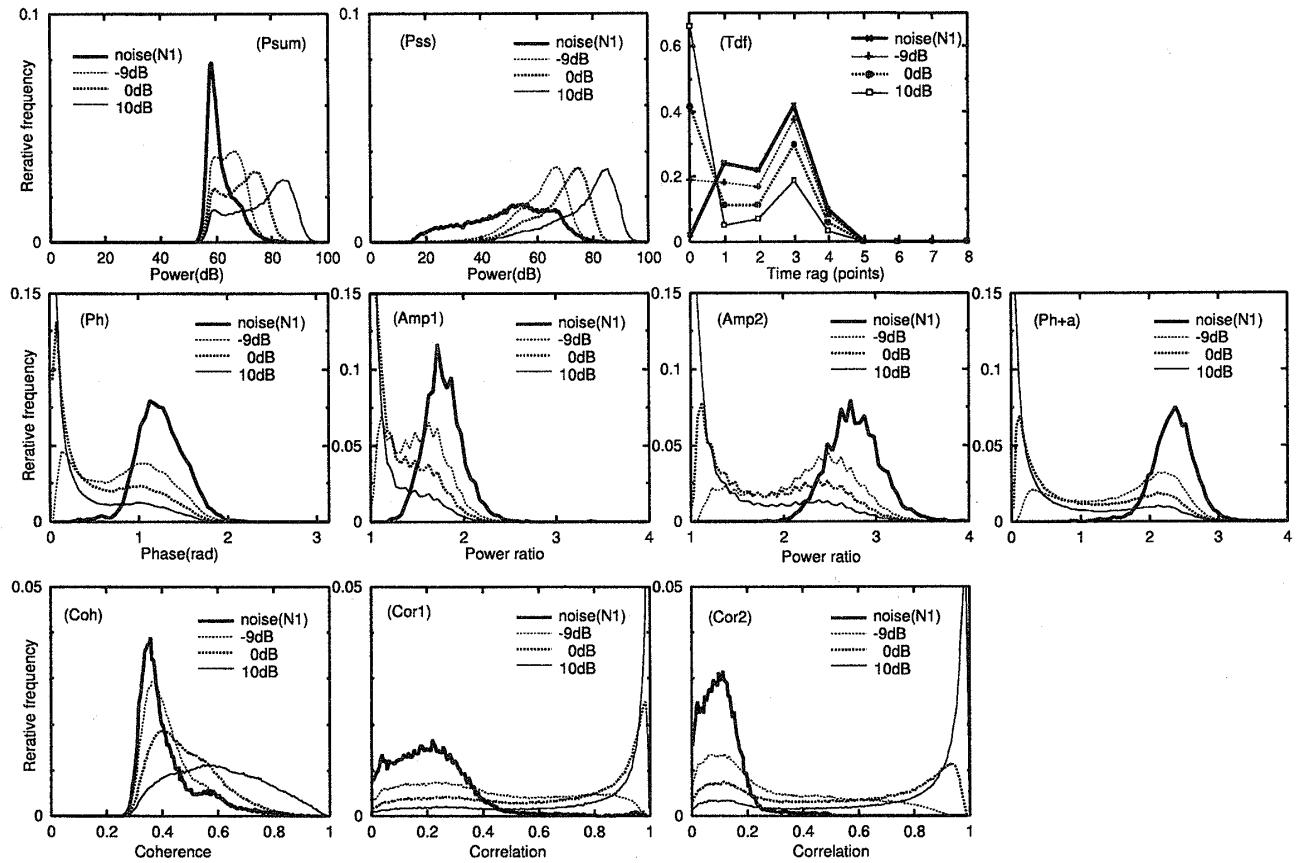


図 5 検出パラメータの出現頻度

Fig. 5 Relative frequency distribution of the endpoints detection parameter.

ると判断する誤警告の頻度の小さい方が検出パラメータとして優れていると言える。相対頻度分布において、脱落に相当する部分と誤警告に相当する部分の例を図 6 に示す。図 6 において、縦線の部分の面積が誤警告の累積相対頻度、横線の部分の面積が脱落の累積相対頻度である。なお、以降ではわかりやすくするために、脱落または誤警告の累積相対頻度を単に脱落の割合、または誤警告の割合、と呼ぶことにする。

この二つの部分の面積が等しくなるように境界を設定したときの脱落または誤警告の割合を図 7 に示す。図 7 において、(N1), (N2), (N3) は雑音の種類であり、各々の雑音について (A) はパラメータ P_{sum} , P_{ss} , T_{df} , C_{oh} , C_{or1} , C_{or2} , P_{h+a} に関する結果、(B) はパラメータ P_h , A_{mp1} , A_{mp2} , P_{h+a} に関する結果である。

パワー比に関しては、図 7 (N1~N3) の (B) からわかるように、単純なチャネル間パワー比 A_{mp1} に比べ、周波数成分間の比較を行ったパワー比 A_{mp2} の方が雑音 (N1), (N2) において常に 0.03 程度脱落 (=

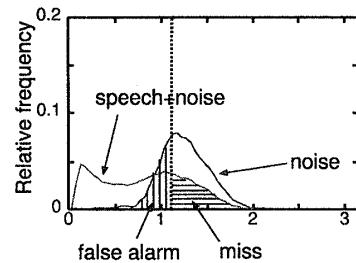


図 6 脱落割合と誤警告割合

Fig. 6 Miss rate and false alarm rate.

誤警告) の割合が小さいことから、 A_{mp2} の方が検出性能が高くなるものと期待できる。

また、(N1), (N2) に関しては、コヒーレンス (C_{oh}) と時間差 (T_{df}) の脱落割合が他のパラメータに比べてかなり高いことがわかる。更に、この図から、どの雑音環境においても、パワー比と位相差を線形結合したパラメータが最も脱落の割合 (= 誤警告の割合) が低いことがわかる。

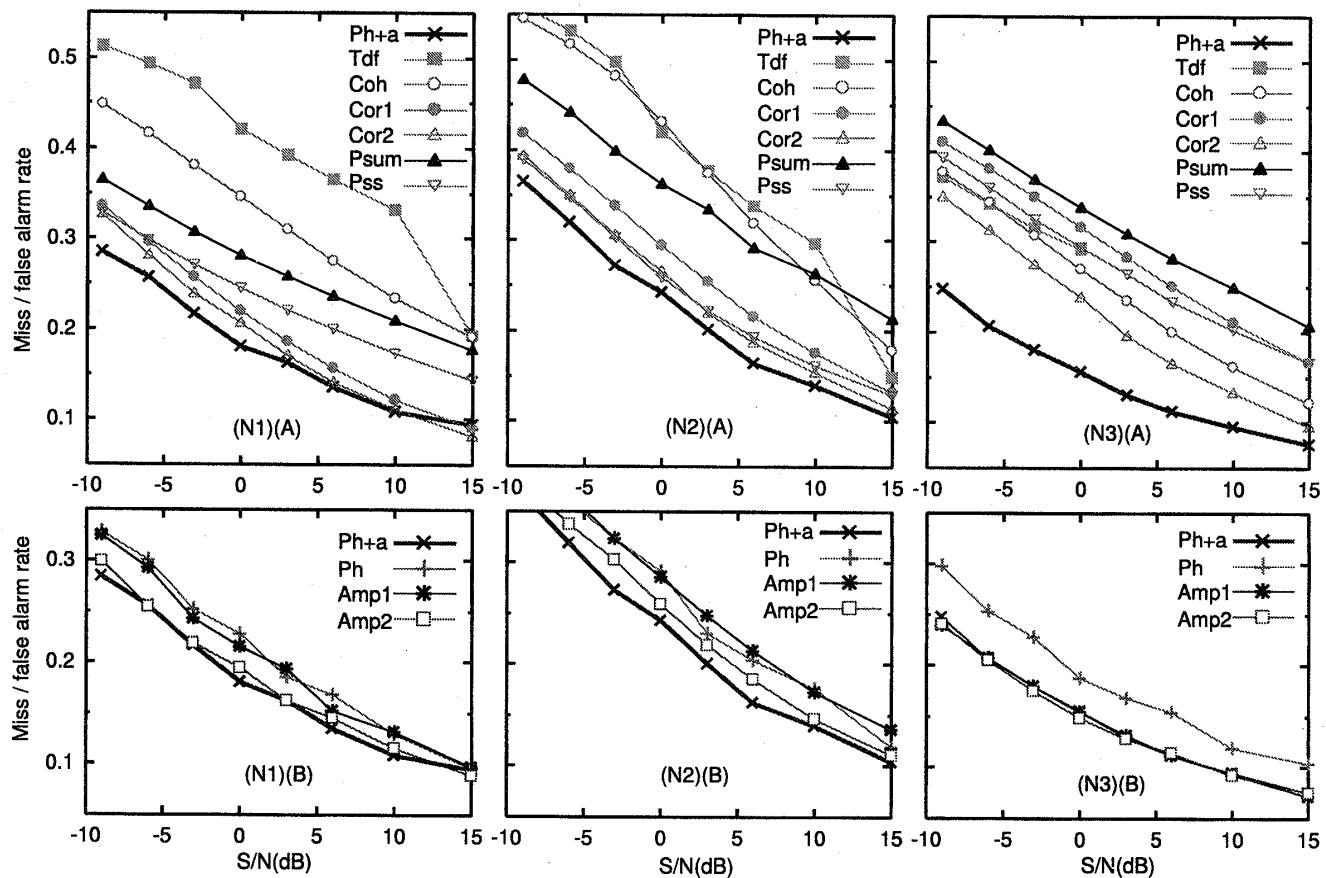


図 7 脱落割合と誤警告割合が等しくなるように設定したときの累積頻度
Fig. 7 Accumulated frequency distribution of miss/false-alarm.

5. 音声検出評価実験

5.1 音声検出のアルゴリズムと検出精度の評価方法

前章で脱落と誤警告が同程度になるようにしたときのパラメータ値の前後に検出しきい値を設定し、音声検出の評価実験を行った。音声検出のアルゴリズムは種々提案されているが、ここでは広く用いられているエネルギーのパルスの検出に基づく複数段階しきい値と時間長の制約を用いた方法[5]をパワー以外のパラメータにも当てはめて用いた。

ただし、種々のパラメータに対して複数段階のしきい値を最適に設定するのは困難であり、パラメータの検出性能ではなくしきい値設定によって検出性能が偏るのを防ぐため、ここではしきい値設定を簡単化した。すなわち、加算パワーとそのSS処理後のパワーに関しては、始終端決定に2段階のしきい値 th_1 , th_2 を適用し、エネルギーのパルスであることを確定させるの

に必要なエネルギーパルス中の最大値として一つのしきい値 $smax$ を用い、単語音声であると確定するのに必要な区間中の最大値として一つのしきい値 $tmax$ を用い、計四つのしきい値を用いた。これらは背景レベルに対する相対値で与え、その初期値を $th_{1,0} = 10\text{ dB}$, $th_{2,0} = 2\text{ dB}$, $smax_{,0} = 15\text{ dB}$, $tmax_{,0} = 25\text{ dB}$ とし、検出実験においてしきい値を変化させる際は、 $tmax$ を $th_{2,0} \leq tmax \leq tmax_{,0}$ の間で変化させた。このとき、 $th_{1,0} \leq tmax \leq smax_{,0}$ となる場合は、 $smax = tmax$ とし、 $th_{2,0} \leq tmax \leq th_{1,0}$ となる場合は $smax = th_1 = tmax$ とした。他のパラメータに関しては、同じアルゴリズムで始終端決定のしきい値1段階のみとした。

時間長の制約に関しては、目的音声とみなすための最小継続長を40ms(3フレーム)、単語音声とみなすのに必要な全体の最小継続長を200ms(18フレーム)、音声の終端確定に必要な無音区間長を320ms(29フレーム)とした。したがって、パワー以外の検出パラ

論文／二つの指向性マイクロホンを用いた目的音検出に関する検討

メータの場合、200 ms 以上しきい値を上回るフレームが継続した場合に、しきい値を上回った部分が音声区間として検出されることになる。ただしこの区間に320 ms 以内の無音区間を含んでよいものとする。

また、次式により、検出パラメータごとに背景レベルの値をフレームごとに更新し、背景レベルに対する相対値を乗じた値を検出しきい値とした。

$$\left\{ \begin{array}{l} u_i(n+1) = u_i(n)(1 - \beta) + D_i(n)\beta \\ \quad (|D_i(n) - u_i(n)| < u_i(n)\gamma \text{ のとき}) \\ u_i(n+1) = u_i(n) \quad (\text{その他}) \end{array} \right. \quad (12)$$

ここで、 i は検出パラメータの種類を表す番号、 n はフレーム番号、 u_i は背景レベル、 D_i は観測された検出パラメータの値である。上式における β 、 γ の値は、予備実験により、 $\beta = 0.1$ 、 $\gamma = 0.2$ とした。

検出精度に関しては以下の量により評価した。

- 1) 脱落率 目視音声区間のうち、検出された音声区間に含まれなかった部分の総時間と目視音声区間の長さの総和（総発声時間）の比
- 2) 誤警告率 検出された区間のうち、目視音声区間に含まれない部分の総時間と目視音声区間外の区間長の総和（雑音総区間長）の比
- 3) 検出率 検出された区間のうち、目視音声区間と重複部分をもつ区間の数。ただし、一つの正解区間にに対して複数の検出区間が重複した場合は、目視始終端位置と検出された始終端位置と誤差の和が最も小さいもの一つとする。
- 4) 始終端誤差 3 に該当する区間の始終端位置の目視始終端位置との差の絶対値の平均

上の評価量は、検出しきい値によって変化するため、パラメータ間で比較するには比較の際の条件を一定にする必要がある。検出しきい値を次第に緩くしていくと、脱落率は減少して誤警告は増加するため、ここでは、両者の値が最も近づくときの検出率と始終端誤差及びそのときの脱落率（= 誤警告率）によって比較することにした。

5.2 音声検出実験結果

しきい値を変化させたときの脱落率と誤警告率、検出率の変化の例を図 8 に示す。しきい値の値は、各パラメータの背景レベルに対する相対値で示した。図 8 は、雑音が (N1)、 S/N が -6 dB のときの平均位相

差 (P_h) と加算パワー (P_{sum}) による検出結果である。先に述べたように、しきい値の変化によって脱落率と誤警告率が逆の傾きで変化し、交差することがわかる。

図 9 は、各 S/N においてしきい値を変化させ、脱落率と誤警告率が最も近づいたときの検出率と始終端誤差及び脱落/誤警告率を示したものである。ただし、パワー比 1 (A_{mp1}) と相関係数 1 (C_{or1}) に関する結果は、各々パワー比 2 (A_{mp2})、相関係数 2 (C_{or2}) に関する結果と似た傾向であり、これらよりも若干低い値となったものの、著しい差はなかったため、グラフの見やすさを考慮して表示を省略した。また、時間差とコヒーレンスの論理積のパラメータ (C_{oh*tdf}) の場合は、時間差のしきい値を 2 点で固定し、コヒーレンスのしきい値を変化させて評価した結果を示した。すなわち、相関関数から得られたタイムラグの点数の絶対値が 2 以下でかつコヒーレンスがしきい値より高い値となるフレームを目的音の到来しているフレームとした。時間差のしきい値は予備実験で最もよい結果を与えたものとした。

図 9において、(N1) ~ (N3) は雑音 (N1) ~ (N3) に対応し、各雑音に関する結果において (A) は検出率、(B) は始終端位置誤差の平均値である。検出率に関しては、平均位相差 (P_h)、パワー比 (A_{mp2})、位相差とパワー比の加重和 (P_{h+a}) の各パラメータがどの雑音環境に対しても高精度であるのに対し、パワー

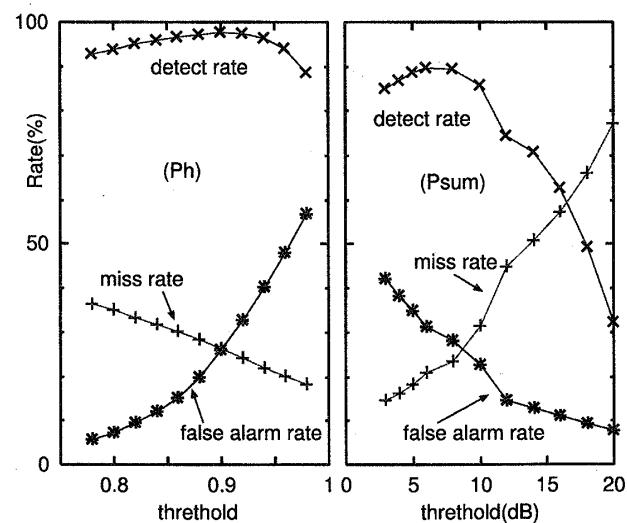


図 8 検出しきい値に対する音声検出性能の変化（雑音：N1）

Fig. 8 Endpoints detection accuracy vs. threshold.

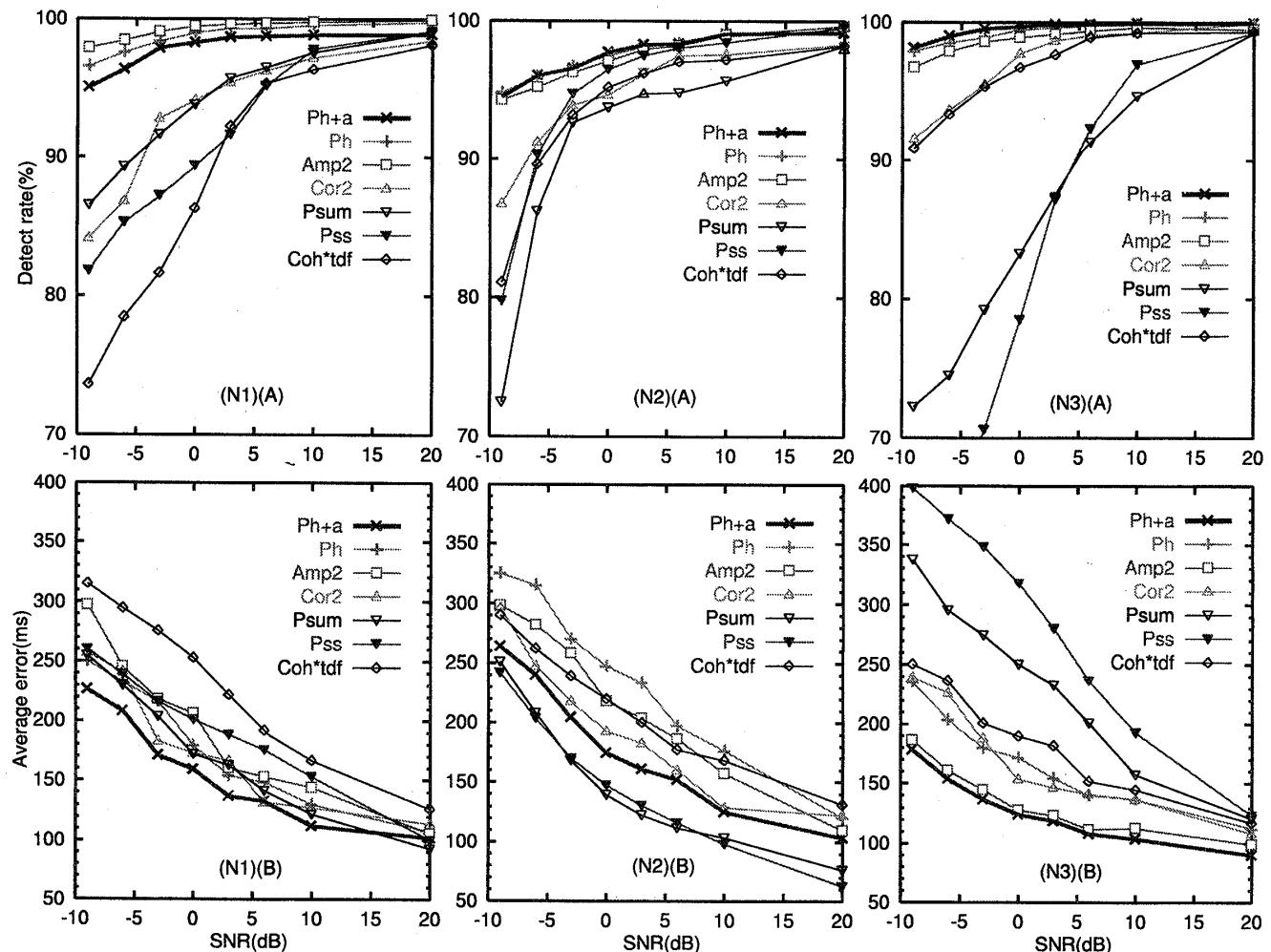


図9 音声検出結果
Fig. 9 Results of word boundary detection.

に基づくパラメータ (P_{sum}), (P_{ss}) は低 S/N 時の性能低下が著しい。また、時間差とコヒーレンスの論理積のパラメータ (C_{oh*tdf}) は、雑音 (N1), (N2)において低 S/N 時の性能低下が著しい。

始終端誤差に関しては、加重和パラメータ (P_{h+a}) が (N1), (N3) の雑音環境において最も低い値となっている。雑音 (N2) の場合、パワーに関するパラメータの始終端誤差は低いが検出率は低い。

また、雑音間で比較すると、全体的に走行雑音重畠データ (N2) に対する検出性能が低い。これは、本方法が雑音の方向性によって生ずる時間差と振幅差を利用しており、走行雑音 (N2) の場合、他の雑音より方向性が弱いためと思われる。しかしながら、(N2)の場合でも、同じ 2ch 法の従来のパラメータである (C_{oh*tdf}) より、提案したパラメータの方が良好な結果であり、特に、 S/N が 0 dB 以下において性能の差

が大きい。

以上をまとめると、いずれの雑音に対しても、提案する位相差とパワー比の加重和のパラメータが安定して高い検出精度と低い始終端誤差を示しており、例えば、 S/N が -9 dB の場合、雑音 (N1) に関しては、検出率 97.5%，始終端誤差約 230 ms、車の走行雑音 (N2) では検出率 95% 検出誤差 260 ms、車走行時かつラジオ音声がある場合 (N3) では検出率 99% 検出誤差 180 ms となった。

6. むすび

2 個の指向性マイクロホンの出力に基づいた新しい目的音声検出パラメータを提案した、車上で収集した 3 種類の評価データによる実験の結果、 S/N が -9 dB と苛酷な環境であっても、位相差とパワー比を組み合せた検出パラメータが、工事現場雑音の場合検

論文／二つの指向性マイクロホンを用いた目的音検出に関する検討

出率 97.5% 検出誤差 230 ms, 車の走行雑音の場合検出率 95% 検出誤差 260 ms, 車走行時かつラジオ音声がある場合検出率 99% 検出誤差 180 ms とどの雑音に対しても安定した高い精度を与え、従来提案されているコヒーレンスと時間差を用いたパラメータよりも提案する音声検出方法の方が高精度で目的音を検出できることを確認できた。

本提案の音声検出法は、音源の方向性を利用していいるため、例えば、マイクロホン間隔と波長との兼ね合いで空間的なエリアシングが生ずる場合や、雑音の反射波と直接波との干渉により、2 ch 間の位相差が目的音に関する位相差と一致してしまう場合などに検出精度が劣化する可能性がある。しかしながら、いずれも特定の周波数において生ずるものであり、広帯域信号の場合それらの影響は軽減されるものと考えられる。更に、雑音と目的音の位相差が一致した場合でも、本方法ではパワー比が異なることによって判別が可能となる場合があるため、以上のような劣化原因の影響は緩和される可能性がある。

なお、本検出実験においては、検出しきい値を脱落率と誤警告率を等しくするような値に設定して評価したが、実際の使用時は、環境に合わせた検出しきい値の設定が重要である。今後、このしきい値設定の問題、及び複数の検出パラメータを統合した検出誤差の改善方法について検討するとともに、音声認識による評価を行う予定である。

文 献

- [1] B.S. Atal and L.R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-24, no.3, pp.201–212, 1976.
- [2] 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, 1976.
- [3] L.R. Rabiner and M.R. Samber, "Application of an LPC distance measure to the voiced-unvoiced-silence detection problem," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-25, no.4, pp.338–343, 1977.
- [4] S.F. Boll "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-27, no.2, pp.113–120, 1979.
- [5] L.F. Lamel, L.R. Rabiner, A.E. Rosenberg, and J.G. Wilpon, "An improved endpoint detector for isolated word recognition," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-29, no.4, pp.777–785, Aug. 1981.
- [6] C. Tsao and R.M. Gray, "An endpoint detector for LPC speech using residual error ahead for vector quantization applications," *Proc. of ICASSP'84*, pp.18B.7.1–18B.7.4, San Diago, CA, USA, 1984.
- [7] 金田 豊, 東山三樹夫, "2 点受音に基づいた雑音抑圧処理," *信学論 (A)*, vol.J67-A, no.4, pp.391–398, April 1984.
- [8] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust., Speech & Signal Process.*, vol.ASSP-34, no.6, pp.1391–1400, Dec. 1986.
- [9] C. Jittun and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol.24, no.1, pp.1–10, July 1991.
- [10] J.E. Greenberg and P.M. Zureck, "Evaluation of an adaptive beamforming method for hearing aids," *J.A.S.A.*, vol.91, no.3, pp.1662–1676, March 1992.
- [11] J.A. Haigh and J.S. Mason, "A voice activity detector based on cepstral analysis," *Proc. of EUROSPEECH'93*, vol.3, pp.1103–1106, Berlin, Germany, Sept. 1993.
- [12] J.-C. Junqua, B. Mak, and B. Reaves, "A robust algorithm for word boundary detection in the presence of noise," *IEEE Trans. Speech & Audio Process.*, vol.2, no.3, pp.406–412, July 1994.
- [13] 宝珠山治, 杉山昭彦, "ブロッキング行列にリーク適応フィルタを用いたロバスト一般化サイドロープキャンセラ," *信学論 (A)*, vol.J79-A, no.9, pp.1516–1524, Sept. 1996.
- [14] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech & Audio Process.*, vol.5, no.5, pp.425–437, Sept. 1997.
- [15] Agaiby and T.J. Moir, "Knowing the wheat from the weeds in noisy speech," *Proc. of EUROSPEECH'97*, vol.3, pp.1119–1122, Rhodos, Greece, Sept. 1997.
- [16] 永田仁史, 安倍正人, "話者追尾 2 チャネルマイクロホンアレーに関する検討," *信学論 (A)*, vol.J82-A, no.6, pp.860–866, June 1999.
- [17] 青木真理子, 岡本 学, 青木茂明, 松井弘行, "チャネル間情報を用いた 2 音源分離手法の検討-分離性能の方向依存性-", *音響学会春季講演論文集*, pp.535–536, March 1999.
- [18] 金田 豊, "騒音下音声認識のためのマイクロホンアレー技術," *日本音響学会誌*, vol.53, no.11, pp.872–876, Nov. 1997.
- [19] 浅野 太, 速水 悟, 松井俊浩, "話者方向同定と雑音抑圧による音声認識性能の改善," *日本音響学会誌*, vol.53, no.11, pp.889–894, 1997.
- [20] 山田武志, 中村 哲, 鹿野清宏, "マイクロホンアレーによる 3 次元ビタビ探索に基づく移動話者の音声認識," *信学技報*, SP97-22, 1997.
- [21] 武田一哉, 板倉文忠, "音源分離による音声認識性能の改

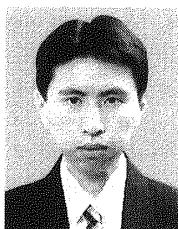
- 善,”日本音響学会誌, vol.53, no.11, pp.883–888, Nov. 1997.
- [22] 小倉久直, 中迫 昇, 濑尾訓生, “情報理論と回転変換に基づくブラインド信号分離の一提案,”信学技報, EA98-79, 1998.
- [23] 木村 隆, 佐々木秀昭, 大城美来, 尾知 博, “RLS 法を用いた 2 次統計量に基づくブラインドシステム同定,”信学技報, DSP98-113, 1998.

(平成 12 年 3 月 17 日受付, 6 月 23 日再受付)



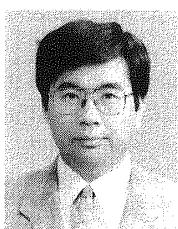
永田 仁史 (正員)

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。音声認識、デジタル音響信号処理、並列計算機用コンパイラの研究に従事。日本音響学会、情報処理学会各会員。



藤岡 豊太 (学生員)

平 4 秋田大・鶴山・電気工学卒。平 6 同大大学院修士課程了。平 9 東北大学大学院博士後期課程了。現在岩手大・工・情報工学科助手。並列計算機、情報圧縮ハードウェアに関する研究に従事。



安倍 正人 (正員)

昭 56 東北大大学院電気及び通信工学専攻博士課程了。工博。昭 58 東北大大学情報処理教育センター助手。平 1 東北大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。デジタル信号処理、並列計算機アーキテクチャ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、日本機械学会各会員。