Short Report

異常検出システムのための音信号自動分割アルゴリズムに関する検討

安倍幸治[1]，大槻和果葉[2]，高根昭一[1]，西口正之[1]，渡邉貫治[1]

[1] 秋田県立大学システム科学技術学部情報工学科

[2] 秋田県立大学大学院システム科学技術研究科電子情報システム学専攻

近年の高齢化社会を背景として，独居老人の増加が社会問題となっている．独居する方が高齢の場合，転倒や発作など突発的なトラブルが発生した場合に，身体的な問題から，重篤な事態になる可能性が高い．そのため，異常事態が発生したことを機械的に検出し，通報などに繋げる見守りシステムの必要性が高まっている．我々は比較的プライバシーの侵害の程度が低いことが利点であることを踏まえ，音信号に基づく異常検出システムの構築を目指している．本稿では，音響特徴量からベイズ情報量規準を求め，それに基づいて，時間的な連続信号である音を発生したイベントごとに時間的に区分するシステムの構築を行った．システムによる自動分割が，人が知覚することのできるイベント分割と同等程度の精度を有していることを確かめるために，評価実験を実施した結果，人が判断することのできるイベント分割点を高い確率で検出していることが示された．

キーワード：自動分割，ベイズ情報量規準，異常検出，見守りシステム

高齢化率の上昇により，独居する高齢者，病院や介護施設を利用する高齢者が著しく増加している．このような背景から，日常生活における緊急時の対応を可能とする「見守りシステム」の需要が高まっている．見守りシステムとは，遠方にいながら高齢の親等の状況を監視し，喫緊の対応を必要とする問題が生じた場合に，それを検知，及び周知するシステムである．見守りシステムに関する研究やシステム事例はこれまでにも多く行われているが，映像情報を利用した見守りシステムがほとんどである（小林と後藤，2016；斎藤と西山，2016；関と堀，2002）．見守りシステムはなんらかのセンサを用いて監視対象者をモニタリングすることで実現されることから，映像情報以外を利用することも可能であるが，音情報を利用した事例は極めて少ないのが現状である．しかし，音情報をモニタリングのメディアとして考えると，家具などの物体による監視の死角がないことや，映像と異なり監視対象者の行為を全て映し出すことがないという点で監視対象者のプライバシー面での負担が小さいなどの利点がある．このような特徴から見守りシステムのトリガとして，音情報を利用し，その検出結果に応じて，映像情報を呼び出すなど，プライバシーに配慮した新たな見守りシステムの実現も期待できる．

音情報をモニタリングするためには，対象となる空間にマイクロホンを設置することが求められる．マイクロホンによって取得される音信号は時間的な連続信号であるため，その中からリアルタイムに異常事態を検出するためには，適切な長さで時間信号を区分し，その特徴を分析することが必要となる．ここでいうところの適切さは，その区間に含まれる音響的特徴を正確に取り出すことのできる程度の長さを持ち，事態が手遅れにならない程度のリアルタイム性を保ち，その区間に含まれる音響的イベントが出来得るなら一つに絞られていることと考える．もちろん，例えば転倒音と破壊音のようなものが同

責任著者連絡先：安倍幸治　〒015-0055 由利本荘市土谷字海老ノ口 84-4　公立大学法人秋田県立大学システム科学技術学部情報工学科．E-mail: koji@akita-pu.ac.jp

時に発生し，それの情報が時間的に重畳することも考えられるが，このような問題には区分後に分離処理等の対処を行うことを想定し，今回は考慮しないこととした．

連続した音信号を区分する手法としては，事前に学習データを準備し，学習データを手掛かりに分割する方法がある（河原，須見，緒方，後藤，2011）．しかし，この手法に基づいて分割処理をするには，分割の対象とする生活音全ての学習データを事前に準備する必要があり，現実的ではなく，システムの汎用性も低下する．この問題に対処するためには，事前に準備しなければならない学習データを必要とせず，入力である音信号の統計的な手掛かりに基づいて音を区分するが考えられる．その代表的な手法としてベイズ情報量規準に基づく方法がある．本報告では，ベイズ情報量規準を用いた音響イベントの自動分割に関する検討を行った．

ベイズ情報量規準による分割手法

ベイズ情報量規準

提案する手法に用いるベイズ情報量規準について説明する．ベイズ情報量規準（Bayesian Information Criterion: BIC）とは，統計学における情報量規準の 1 つである．（Schwarz，1978）本提案手法では音信号を分割するための区間選択の評価規準として用いる．ある一定の長さを持つ音信号が何らかの統計モデルに従うと考える．これを一つのモデルと考えた場合，そのモデルのBIC値は以下の式により算出される．なお，L 尤度，d を特徴量の次元数，λ を分割重み，N をサンプル数とする．

$$\mathrm{BIC} = \log L - \frac{d}{2}\lambda \log N$$

上式における尤度 L は，ある区間の音信号に基づくデータセットの統計モデルへの適合性の高さを表す．本研究では，同一音響イベントから発生する音信号の音響特徴量群はある一定の値を有すると考えられることから，多変量正規分布に従うと仮定する．なお，音響特徴量は多次元とする．データセット $x_1, \cdots, x_N$ が平均 μ，分散 Σ の多変量正規分布に従う尤度は以下のように表される．

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d * |\Sigma|}} exp\left\{-\frac{1}{2}(x_i - \mu)\Sigma^{-1}(x_i - \mu)^T\right\}$$

仮定した統計モデルに対してデータセットの持つ分布の適合性が高ければ，BIC 値は大きくなる．

BIC に基づく音信号分割アルゴリズム

BIC に基づいて音信号の分割を行うために，はじめに分析対象である音信号を，音響特徴量を取り出すのに十分なサンプル長（N_{Sub}）のサブフレームに分割する．BIC 値を計算するために N 個のサブフレームから成る基本フレームを生成する．この際，分割の時間解像度影響を及ぼすパラメータとして，サブフレームのシフト長（N_{shift}）も決定することが必要となる．この基本フレームを 1 つのモデルとして考えた場合と，2 つのモデルが結合したものとして考えた場合に算出されるそれぞれの BIC 値を比較し，音信号を分割すべきかを判断する．

図 1 に BIC 値の比較の概念図を示す．基本フレームを 1 つのモデルと考えたものをモデル M_0，N 個のサブフレームからなる基本フレームを前半 j 個，後半 $N-j$ 個の 2 つに分け，それぞれ異なるモデルと考えた場合に，前半をモデル M_1，後半をモデル M_2 とする．M_0 から算出される BIC 値を $\mathrm{BIC}(M_0)$，M_1 及び M_2 から算出される BIC 値を，それぞれ $\mathrm{BIC}(M_1)$，$\mathrm{BIC}(M_2)$ とする．基本フレームが，1 つのモデル M_0 から構成されていると考えた場合と，M_1 と M_2 からなる 2 つのモデルから構成されていると考えた場合に，どちらの尤度が高くなるかを判断するために，下の式により，BIC の差である ΔBIC を求める．

$$\begin{aligned}\Delta\mathrm{BIC} &= \{\mathrm{BIC}(M_1) + \mathrm{BIC}(M_2)\} - \mathrm{BIC}(\mathrm{M_0}) \\ &= \log\frac{L_1 L_2}{L_0} - \frac{d}{2}\lambda \log\frac{N_1 N_2}{N_0}\end{aligned}$$

この ΔBIC は尤度の差であるため，ΔBIC が正の値を持てば，音信号を二つのモデルで表す方が，尤度が大きくなる．すなわち，その分割点で音響イベントが変化しており，音信号を分割した方が妥当であるということを意味すると考えられる．

本稿で提案する音信号分割アルゴリズムでは，分

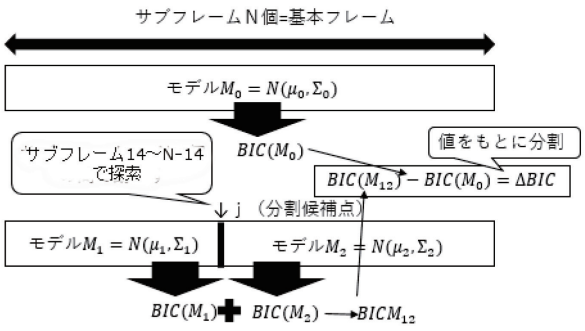

図 1 BIC に基づく音信号分割の概要

割候補点 j を設定された範囲である 14 から *N*-14 の範囲で変更し，求めた ΔBICが最大かつその値が正となる位置で信号を分割することとした．14 という数値は，分析に用いた音響特徴量ベクトルの要素が 13 であり，適切な分散値を算出するために，サンプルが要素数+1 必要であることから決定した．扱う音響特徴量が増減した場合には，この値は修正されなければならない．なお，分割重み λ の値により，ΔBIC にバイアスがかかるため，増減の傾向は変わらないが，λ の値により， ΔBIC が正となるかどうかは影響を受ける．そのため，分割重み λ は別途適切に設定する必要がある．今回は，ベイズ情報量規準に関する先行研究を参考にΔBIC = α として，各分割候補点に対する λ を算出し，その平均値を分析区間における分割重みλ とすることとした（河原ら，2011）．これにより分析区間によるΔBIC の変動を打ち消し，平均値をおおよそ一定に保持することが可能となる．完全な自動分割を実現するためには，設置した環境に最適な α の値を自動的に決定することが必要となるが，今回は，手動で適宜設定することとする．

上記の手順により，分割候補点が見つかった場合には，分割候補点の次のサブフレームが探索の始点となるように，基本フレームを更新する．また，ΔBIC の値が負となり，分割候補点が見つからなかった場合は，探索の始点は変更せず，基本フレームを倍の長さに拡張し，再度探索を実行することとする．

音響特徴量

音信号の分割は環境で発生する音響イベントに依存した音信号の変化を手掛かりに行われる．そのため，異常を検出する上で，適切な音響特徴量を設定することが必要である．見守りシステムにおいて検出が求められる異常として，転倒音や破壊音が挙げられる．このような事態においては，唐突に音の強さが大きく変動すると考えられ，音の強さを手掛かりに音響イベントの境界を検出できると考えられる．そこで，一つ目の音響特徴量を二乗平均平方根（以下，RMS（Root Mean Square））とする．また，緊急事態では，悲鳴のような日常生活ではあまり発生しない音声が収音されることを想定し，二つ目の音響特徴量を，音声の手掛かりとして良く用いられるメル周波数ケプストラム係数（以下，MFCC（Mel Frequency Cepstral Coefficient））とする．MFCC は音声の特徴を表現するのによく用いられ，主に音声認識や音声合成技術などの分野で用いられている．音声認識や音声合成技術などの分野では特徴を表現する代表例として MFCC の他にケプストラムがある．ケプストラムは周波数軸上，一様なスケールでスペクトルをサンプリングし，フーリエ変換することによって得られる．一方，MFCC は一様な周波数スケールではなく，人の聴覚特性に合わせて低周波領域のスペクトルを細かくサンプリングして得られるケプストラムである．音声信号の基本周波数は男性平均が 125 Hz，女性が 250Hz であるので，比較的低周波帯域にエネルギーが集中する信号である．この特徴から，音声認識では一般的に低次 12 次元のケプストラムを用いる．本手法でも低次 12 次元のケプストラムを用いることとする．

音信号の分割妥当性の検証

本研究で開発を目指している見守りシステムでは，日常生活における異常を検出することを目的とする．そのため，その前段として設置される音信号自動分割システムは，日常，人が聴いて異なる音響イベントであると判断することができる区間は最低限検出し，分割することが望まれる．本稿で提案する音信号自動分割アルゴリズムがこの目標を達成できているかを評価するためには，音信号を人間が聞いての手動で分割した結果が基準として必要となる．そこ

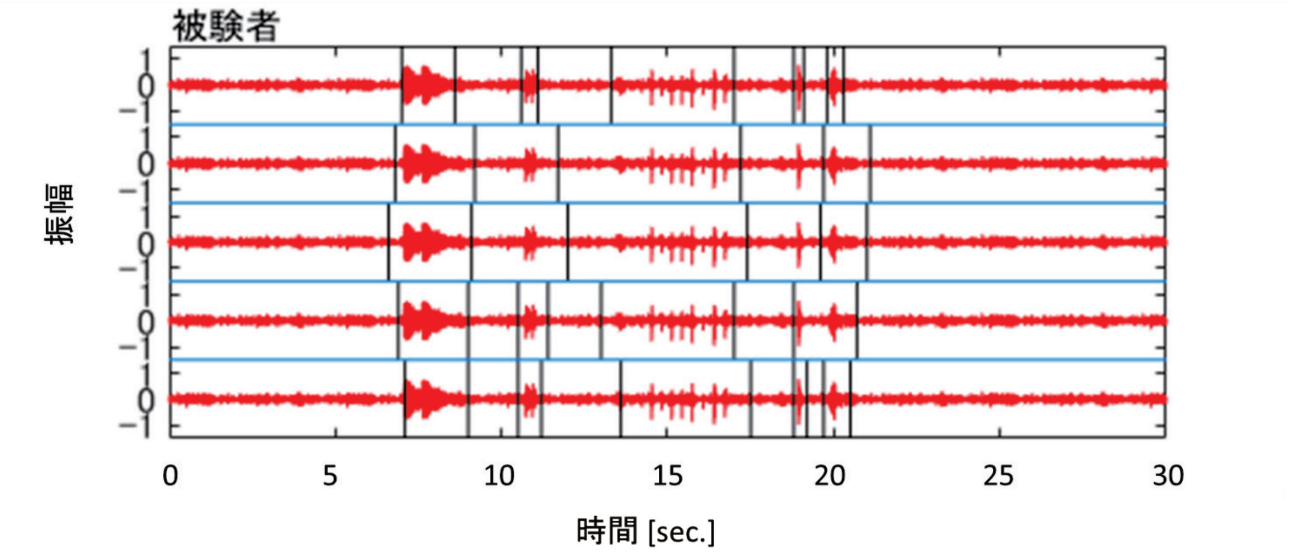

図 2　人間による音信号の分割結果

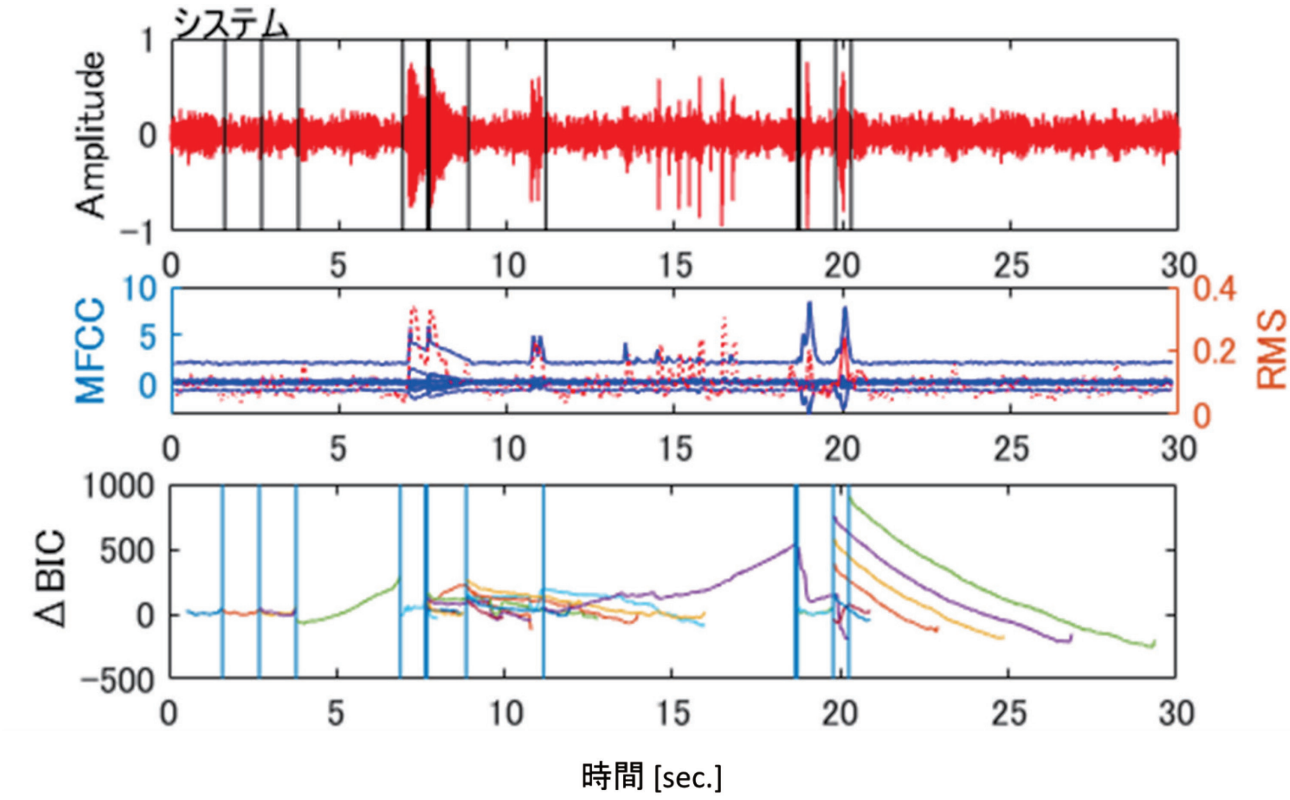

図 3　提案アルゴリズムによる音信号の分割結果

で, いくつかの音信号を用いて聴取実験を実施した. 実験手順は次の通りである. 実験に使用する音信号はパソコンのサウンドエディタで波形が見える状態にし, 初めに, 被験者にその音信号を一度通して聴かせる. 続いて, 被験者にパソコンの操作を委ね, 自由に音信号を停止, 再生, 巻き戻しが出来る状態で, 事前に渡した音信号波形を記載した回答用紙に音信号に含まれる音響イベントの内容が変わったと感じた境界に線を引くよう教示した. 被験者は, 正常な聴力を有する 20 代の女性 3 名, 男性 2 名である.

実験に用いた刺激音は 4 種類で, 継続時間はそれ

ぞれ 約 30 秒である．3 つの刺激音は，環境音のデータベースから引用した刺激である（効果音ラボ ver. 3.0）．1 つ目の音は電子レンジの動作音であり，電子レンジの扉が開く音，閉まる音，動作音が含まれる．2 つ目の音は玄関における環境音であり，足音，チャイム音，ドアのノック音，ドアの開閉音が含まれる．3 つ目の音はオフィスにおける環境音であり，話し声，電話の音が含まれる．残り 1 つの刺激音は，本学の研究室において実際に録音したものであり，足音や扉の開閉音，本が落ちる音が含まれる．

実験結果の一例を図 2 に示す．図 2 は，玄関の環境音に対する人による分割結果を示したものである．それぞれの波形は，音信号の波形を表しており，グラフ内の縦線は，被験者によって描かれた音信号の区分の境界を表している．波形の前半の大きな二つの包絡は，チャイムの音であり，その次の山はノックの音となっている．さらにその次の細かいパルス状の信号は，足音となっている．最後に見える二つのピークは，ドアの開閉音である．図 2 から分かるように前述した大まかな音響イベントの区分が，被験者によって，およそ同一の傾向で分割されていると言える．特に足音のような音信号は，パワーとしてはオンオフがはっきりしているが，人間にとっては一塊のイベントとして認識されていることが分かる．一方，同じ刺激音を提案する自動分割システムに入力した結果を図 3 に示す．図 3 の上段のパネルは図 2 と同じ形式で分割点を図示したものである．中段のパネルは，各サブフレームで算出された音響特徴量をプロットしたものであり，下段のパネルは，その特徴量を基に算出されたΔBIC の値をプロットしたものである．図 3 の上段パネルを見て分かるように，自動分割システムでは，人間が分割した結果である図 2 と比べて，比較的多くの分割がなされていることが分かる．特に，刺激音の開始部及び終了部のチャイムの音，ドアの開閉音は，大きなパワー変動がみられる部分で二つの音響イベントとして分割される結果となった．

人間によって分割された分割位置が，自動分割システムの分割で得られているかを定量的に確認するために，各々の手法による分割位置を基準として，その前後±1 s 内にもう一方の手法による分割が存在した場合に同じ分割点が得られたとし，再現率を求めた（Muller と Guido，2017）．その結果を表 1 に示す．システムによる分割位置を基準として求めた再現率 R_m を見ると高くても 75 %程度となっている．一方，人による分割位置を基準として求めた再現率 R_h を見ると，例として示した玄関の環境音においては，80 %とやや低かったが，その他の 3 つの刺激音においてはほぼ 100 %に近い再現率が得られている．このことから，自動分割システムによる分割は，少なくとも人間の手による分割が発生するポイントを高い確率で得られていると言える．また，R_m がR_h より低いことから分割がやや過剰であるともいえる．これは，設定した分割重み λ（$\alpha = 265$ として決定）が，やや検出感度が高くなる方向に設定されていたことが原因と考えられる．しかし，異常検出システムとしては，検出漏れを可能な限り低減したいため，検出がやや過剰であることはそれほど問題ではないと考える．ただ，異常検出システムの計算負荷の面から考えると妥当な感度が自動的に設定されることが望まれる．今回のシステムによる分割結果である図 3 を見ると，背景雑音によるものであると思われる音刺激開始部分の定常状態でも分割が発生しているのが見て取れる．このような定常部では分割を行う必要性は低いと思われるため，定常状態を手掛かりに，より適切な分割重み λ の設定が可能となるのではないかと考える．

表 1 *自動分割アルゴリズム及び人による分割位置を基準とした再現率*

刺激音	R_m ［%］	R_h ［%］
電子レンジ	45.3	100
玄関	70.0	80.3
オフィス	46.7	94.4
研究室	75.9	97.6

結言

見守りシステムの前段に必要と考える学習データを必要としないベイズ情報量規準に基づく音信号自動分割システムを提案した．いくつかの音刺激サン

プルを用いて，評価実験を実施し，その分割の妥当性を検討したところ，人間の主観的により検出される音響イベントの分割位置は本システムにより検出可能であることが示された. しかし, 現在のところ，分割感度を決定する分割重み λ を事前に決定しておく必要があることが問題として残されている．背景の定常雑音部で分割の必要がないことを手掛かりとして，分割重みの適切な値の自動設定が今後の課題と考える．

2019 年 6 月 30 日受付
2019 年 7 月 9 日受理

謝辞

本研究は秋田県立大学平成 30 年度学長プロジェクト研究費「科研費チャレンジ研究」の支援を受けて行った．ここに記して謝意を表する．

文献

小林甲一，後藤健太郎（2016）.「高齢者見守りシステムの展開, 現状そして新たな取組 ―広島県福山市と福岡市を事例に―」『名古屋学院大学論文集，社会科学篇』52（4），23-38.

斎藤裕佑，西山裕之（2016）.「人物動作および顔の表出情報による見守りシステムの提案と実装―咀嚼認識を例として―」『電子情報通信学会技術報告』115（414），193-198.

関弘和，堀洋一（2002）「高齢者モニタリングのためのカメラ画像を用いた異常動作検出」『電気学会論文誌. D, 産業応用部門誌』122 (2)，182-188.

河原達也，須見康平，緒方淳，後藤真孝（2011）「音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出」『情報処理学会論文誌』52（12），3363-3373.

Andreas C. Muller，Sarah Guido（2017）『Python ではじめる機械学習 ―scikit-learn で学ぶ特徴量エンジニアリングと機械学習の基礎』オライリージャパン.

Schwarz, G. E. (1978). Estimating the dimension of a model.*Annals of Statistics*, 6, 461-464.

Investigation of an automatic segmentation algorithm of acoustic events for an anomaly detection system

Koji Abe[1], Wakaba Ohtsuki[2], Shouichi Takane[1], Masayuki Nishiguchi[1], Kanji Watanabe[1]

[1] *Department of Information and Computer Science, Faculty of Systems Science and Technology, Akita Prefectural University*

[2] *Course of Electronics and Information Systems, Graduate School of Systems Science and Technology, Akita Prefectural University*

The number of elderly people living alone is a social problem that is increasing due to an aging society. For elderly people, sudden problems encountered alone lead to serious consequences. Therefore, there is an increasing demand for monitoring systems that detect abnormal situations mechanically. We constructed an anomaly detection system using sound signals which has the advantage of a reduced violation of privacy. In this paper, we explain how we constructed a system to divide sound signals automatically using the Bayesian information criterion. The Bayesian information criterion used in this system can be obtained from calculations using multi-dimensional acoustic features in a short time frame. An evaluation experiment was carried out to confirm that the results of automatic segmentation were equivalent to manual segmentation based on human perception. As a result, we found that if the appropriate segmentation weight is set, it has the same performance as segmentation based on human perception.

Keywords: automatic segmentation of acoustic events, Bayesian information criterion, anomaly detection

Correspondence to Koji Abe, Department of Information and Computer Science, Faculty of Systems Science and Technology, Akita Prefectural University, 84-4 Ebinokuchi, Tsuchiya, Yurihonjo 015-0055, Japan. E-mail: koji@akita-pu.ac.jp