

バイノーラルマイクを用いた音源方向推定による危険予測

Dangerous situation anticipation by estimate sound source with a binaural microphone

持田悠介

Mochida Yusuke

法政大学情報科学部デジタルメディア学科

E-mail: yusuke.mochida.wz@cis.hosei.ac.jp

Abstract

Recent augmented reality that uses sound information doesn't make use of sound information for knowing ambient environment. But if we know ambient environment by sound information and presentation it visually, people who can not use sound information make use of it for prediction of dangerous. In this paper, we estimate sound source direction with a binaural microphone based on human's sound image localization method. The result of experiment that used model which we made it with Head Related Transfer Function(HRTF), We made ITD models by HRTF and did sound source direction estimate experiment that used the ITD models. As a result, sound source direction estimation results appeared position of $\pm 0 \sim 30$ degrees from the right direction.

1 まえがき

現在の拡張現実には画像処理や GPS の座標を用いて情報を付与するものがある。音声を用いたものでは骨伝導スピーカーや超指向性スピーカーによる特定の場所での情報提示や音声によるナビゲーションシステムなど、音声情報を出力の方法として用いるものばかり [8] [10] [11] で、幅広い周囲環境を知るために音声情報を用いる研究はあまり行われていない。しかし、音声情報を利用し周囲環境を知ることが出来れば、得られた情報を視覚的に表現する等の方法で提示することで聴覚障害者への危険予測や、視覚から得られる情報が減る暗闇の中での危険予測等へ役立てられる。

従来の音源方向、音波到来方向推定の研究ではマイクロホンアレイを用いた方法 [12] や複数の指向性マイクロホンを用いた手法 [7] があり、3つ以上のマイクロホンを用いた手法が多く見られる。しかし、人間は2つの耳から受け取る音声情報で音源定位を行っている。そのことから、両耳から入力される 2ch の音声情報を用いれば音源方向が推定出来る可能性がある。

本研究では人間の音源定位の方法を参考に音源方向の推定を行うために、バイノーラルマイクを用い録音した音声を使用し音源方向の推定を行う。

2 バイノーラルマイクを用いた音源定位法

本研究では音源方向推定を行うために人間の音源定位法を参考にする。人間の音源定位法を参考に音源方向を推定するには、人間が聞いている音声と同等の情報を持った音声を得る必要がある。そこで、本研究では音声の収録にバイノーラルマイクを使用する。バイノーラルマイクとは外耳道に装着するマイクであり、人間が聞く音声をそのまま記録する事が出来る。そのため、立体音響の分野で使用されるマイクである。立体音響では音声を定位方向や距離感等で 3 次元的に感じる事が出

来る。そのことからバイノーラルマイクを使用する事で、人間の音源定位方法を参考に音源方向を推定するのに十分な情報を持った音声を得る事が出来ると考えられる。バイノーラルマイクで録音した音声から音源方向推定を行うために、人間の音源定位に關係する要素について知る必要がある。

2.1 人間の音源定位法

人間が水平面内にある音源の方向を知覚するには、両耳信号の差が用いられる。両耳信号の差は両耳間伝達関数と呼ばれ、両耳信号をフーリエ変換しその信号同士の割算で得られる。しかし、人間の聴覚システムは音源定位を両耳間の複雑な信号差で詳細に評価するのではなく、簡単に認識できる特性から必要な情報を引き出して行っていると考えられる [1]。それら音源定位に使用される情報として、両耳間時間差 (Interaural Time Difference : ITD)、両耳間レベル差 (Interaural Level Difference : ILD) がある。その他に音の空間情報を再現する手法として頭部音響伝達関数 (Head Related Transfer Function:HRTF) を測定し、任意の音声信号に畳み込む方法がある。HRTF には両耳間時間差や両耳間レベル差の他にも音源から両耳までの周辺環境や人体による回折・反射等の伝達系の周波数特性が含まれているため、HRTF を用いることで音の空間情報を高精度に再現できる。しかし、頭部音響伝達関数は耳介や頭部の形状等によって個人差が出てしまうため受聴者自身が測定した HRTF を使用することが望ましい。

2.1.1 両耳間時間差 : (Interaural Time Difference : ITD)

両耳間時間差とは、両耳に入力される信号の時間差のことである。

自由音場において測定される両耳間の入力信号の位相係数は非常に複雑な様相を示すが、耳間定位の実験では簡略化した形が使われる。そのためそのような実験結果を自然な空間音響に適用するには注意を要する。

両耳間時間差のもっとも単純な形は、両耳の入力信号の時間波形は同一で時間的にずれているのみの場合である。そのような信号を作成するには信号を遅延装置に通す事で図 1 のようなゆがみの無い時間遅れのついた 2 個の信号を作り出す方法が良く行われている。両方の信号の遅れが同じであれば、音源は正中面内に生じる。一方の信号に遅れが生じると音源は先に信号が到達した耳の方向へ移動する。

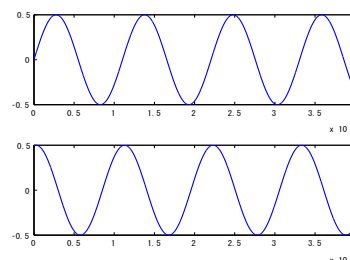


図 1 簡略化された両耳間時間差の例

ITD の代表的なモデルに c を音速、頭の半径を a として

$$ITD = \frac{a(\sin \theta + \theta)}{c} \quad (1)$$

と言うものがあり、このモデルを用いて 0~90 度の間の ITD を計算すると図 2 のようになる。

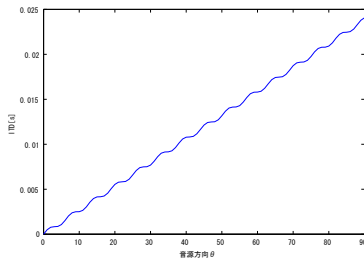


図 2 式 (1) のモデルを使用した 0 度 ~ 90 度の ITD と水平角の関係

たとえば $a = 0.09, c = 340, \theta = \frac{\pi}{2}$ とすると $ITD = 0.68ms$ となる。

2.1.2 両耳間レベル差: (Interaural Level Difference: ILD)

音源から発せられた音声に関して、左右の耳が受ける音波のレベル差を両耳間レベル差 (Interaural Level Difference: ILD) という。音源が片側からもう一方に移動するまで知覚する方向は両耳間レベル差にほぼ線形的に依存している、ILD はどんな周波数でも 10~15dB になれば、音源が片方の耳に完全に移動するとされている [1]。

ITD を使用した音源方向の推定は両耳間パワースペクトル比から得られた伝達関数を用いる。HRTF によって各方向における周波数特性は変化するため、それをモデルとして音源方向を推定出来ると考えられる。

2.1.3 頭部音響伝達関数: (Head Related Transfer Function: HRTF)

ITD, ILD だけでは前後・上下の区別ができないという問題を解決する手段に頭部音響伝達関数 (Head Related Transfer Function: HRTF) がある。HRTF とは耳介や頭部、肩までを含めた体の各部によって生じる音声の回折・反射等の音声の変化を伝達関数としてあらわしたものである。HRTF は線対象な音源に対しても異なり、音源の特性がわかれば前後・上下の区別が可能である。HRTF には ITD, ILD も組み込まれているためそれらを独立に扱う必要はない。

本研究では身体各部による HRTF の個人差に対応するために、被験者の HRTF を測定して使用した。

3 HRTF の測定およびモデル作成

3.1 HRTF の測定方法

音源方向推定に使用する ITD のモデルを作成するために HRTF の測定を行う。本研究では HRTF の個人特性に対応するために実際に HRTF の測定を行った。測定は壁に吸音材を使用し反響等の影響を抑えた部屋で行った。図 3 のように被験者にバイノーラルマイクを装着した状態で椅子に座らせ、被験者の頭部に高さを合わせスピーカーから TSP 信号を出力し測定した。スピーカーを被験者から距離 2m の位置に設置し、水平面において正面を 0 度として 10 度間隔で 36 方向の測定を行った。TSP 信号の同期加算は 8 回行い、SN 比を高めた後逆 TSP 信号を畳み込みインパルス応答を求めた。

3.2 HRTF を用いた ITD モデル作成

本研究では HRTF を用いて ITD から水平角を求める計算式を作成し、それをもとに音源方向の推定を行う。

3.2.1 ITD を用いた水平角の計算

ITD を使用した水平面上の角度の推定には式 (1) の計算式を用いることが出来る。しかしながら、両耳に届く音声情報は直接耳に入る音声だけではない。身体各部によって生じる音声

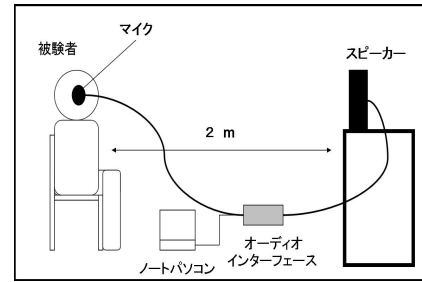


図 3 HRTF の測定系

の回折・反射などによって顔の表面を伝い到達する音声もある。これら音声の変化、伝達経路の違いによる到達時間差から ITD と水平角の関係を理論的に求めるのは難しい。

本研究では HRTF から ITD と水平角の関係を求め、その関係から近似式を計算する。得られた ITD と水平角の関係の近似式を ITD のモデルとして使用し音源方向を推定する。

まずはじめに、ITD の検出に使用する音声データを切り出す。実験的に求めた閾値を用いて、音声の発生点を定める。左右の音声データに対し図 4 のように閾値より大きなデータが入ったときに音の発生と判定し、音の発生点から 512 点 (11.6[ms]) 遡った点をデータ取得開始点とする。データ取得開始点から 1024 点 (23.2[ms]) 分、左右の音声データを抜き出す。これは非常に短い音声が入ってきたときに確実に音声のデータを取得するためである。このような音の例として手を叩く音があげられる [9]。

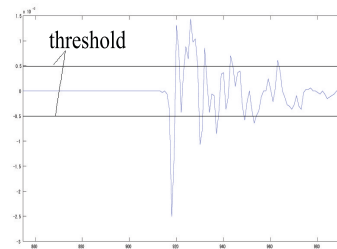


図 4 音の発生検出

ITD は両耳間の信号の位相差といえ、両耳間の信号の相互相関を計算することで求めることが出来る。

本研究では計算時間の短縮をするために、相互相関関数の計算法の一つであるクロススペクトル法を用いて相互相関関数の計算を行う。

クロススペクトルとは相互相関関数をフーリエ変換したものであり、クロススペクトル法とはクロススペクトルを計算し、逆フーリエ変換を行うことで相互相関関数を求める方法である。クロススペクトルは 1 つの信号の複素スペクトルの複素共役ともう一つの信号の複素スペクトルを掛け合わせることで求められる。 $X(f), Y(f)$ をフーリエ変換し、求められた複素スペクトル $X^*(f), Y^*(f)$ を X, Y の複素共役としたとき、クロススペクトルは次の式で求められる。

$$G_{AB}(f) = Y(f) \times X^*(f) \quad (2)$$

クロススペクトル法を用いることにより、高速フーリエ変換 FFT (Fast Fourier Transform) を用いた高速処理が可能となる。

3.2.2 ITD のモデル作成

図 5~ 図 8 に測定した HRTF のインパルス応答 (Head-Related Impulse Response: HRIR) から求められた ITD と水平角の関係を示した。ITD と水平角 α の近似式は式 (1) の ITD のモデルから、正面を 0 度とすると ITD の値が 0 度から 90 度へ向けて高くなり、90 度から 180 度へ向けて低くなる一

次式になると考えられる．近似式は線対称な音源に対しても異なるという HRTF の特徴から左右異なるものと考えられるため，左右別々に最小二乗法を用いて式 (3)～(6) のように求めた．

$$\alpha_{L-front} = \frac{ITD_{L-front}[ms] - 5.0 \times 10^{-2}}{9.1 \times 10^{-3}} \quad (3)$$

$$\alpha_{L-rear} = \frac{ITD_{L-rear}[ms] - 1.7}{-9.1 \times 10^{-3}} \quad (4)$$

$$\alpha_{R-front} = \frac{ITD_{R-front}[ms] - 9.7 \times 10^{-2}}{9.6 \times 10^{-3}} \quad (5)$$

$$\alpha_{R-rear} = \frac{ITD_{R-rear}[ms] - 1.7}{-9.1 \times 10^{-3}} \quad (6)$$

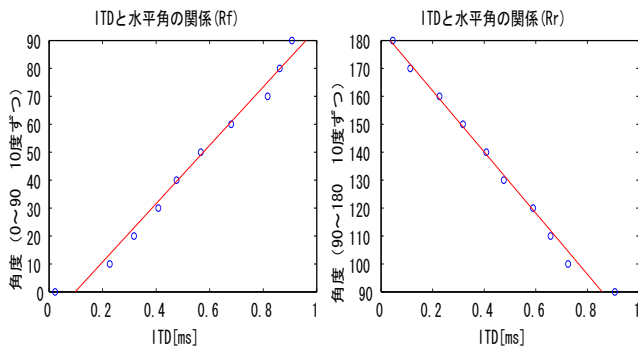


図 5 正面 0 度から右側 90 度までの水平角と ITD の関係

図 6 右側 90 度から右側 180 度までの水平角と ITD の関係

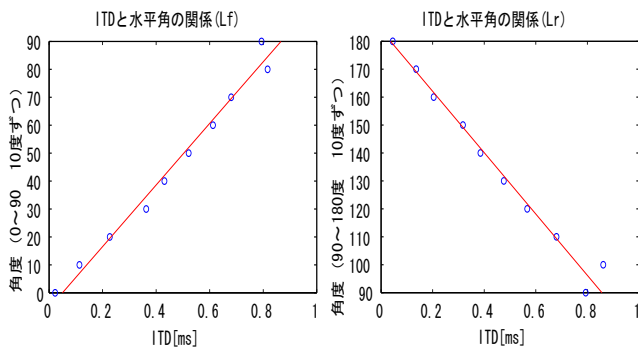


図 7 正面 0 度から左側 90 度までの水平角と ITD の関係

図 8 左側 90 度から左側 180 度までの水平角と ITD の関係

4 HRTF を用いて作成したモデルを使用した音源方向推定の評価

4.1 実験方法

室内で測定した HRTF を用いて作成したモデルと，同じ室内で録音した定位対象音 11 種類を用いた音源方向推定実験を行う．

定位対象音は 11 種類の音声 (サンプリング周波数 44.1[kHz]) で男性 2 名の母音 (/a/, /i/, /u/, /e/, /o/) と拍手を使用した．定位対象音の測定は HRTF の測定と同様，壁に吸音材を使用し反響等の影響を抑えた部屋で行った．また，室内の暗騒音は 30.6dB(A 特性)50.6dB(C 特性)であった．定位対象音は被験者の頭部に高さをあわせたスピーカーから出力され，図 3 の様に椅子に座った被験者に装着したバイノーラルマイクによって録音される．測定方向は正面を 0 度として回転する椅子に座った被験者を動かし，10 度間隔で 36 方向での録音を行った．

録音したデータに対してモデル作成時と同様に，実験的に求めた閾値を用いて音声の発生点を求める．音声の発生点から 512 点 (11.6[ms]) 遡った点をデータ取得開始点とし，取得開始点から 1024 点 (23.2[ms]) 分の両耳信号を抜き出し，ITD を求める．求めた ITD を用いて作成したモデルから音源方向の推定を行う．

同様の方法で屋外 (暗騒音 55.4dB(A 特性)68.1dB(C 特性))にて測定したデータとの比較を行い，音源方向推定のモデル作成を実環境で行う必要があるか，本研究の手法を用いて実環境での音源方向推定が行えるか検討する．

4.1.1 実験結果と考察

図 9～図 11 に録音した音声の方向と推定結果の関係を示す．図の横軸は測定した音声の音源方向を示し，縦軸は本研究の手法で推定した音源方向を示している．収録した音声の音源方向と推定した音源方向が一致した場合，グラフの左下から右上に向かう対角線上に $y=x$ が表示される．

図 9 は室内で録音した HRTF から作成したモデルを用いて室内で被験者の左側に音源がある場合の推定結果，図 10 は屋外で録音した HRTF から作成したモデルを用いて室内で被験者の左側に音源がある場合の推定結果，図 11 は屋内で録音した HRTF から作成したモデルを用いて屋外で被験者の左側に音源がある場合の推定結果である．

また表 1～表 3 には左側，右側での推定結果が実際の角度に対して ± 10 度， ± 20 度， ± 30 度の範囲でどれほど含まれているかを示した．

実験から，誤差を ± 10 度以内に収めるには背景雑音や反響が同様の環境で測定したモデルを使用する必要があると考えられる．しかし，誤差 ± 20 度からはどちらのモデルにおいても大きな差は現れていない．

図 9，図 10 では誤差が山なりに現れている．これは図 5～図 8 から音源方向が 90 度では無くその前後で ITD の値が最大になるためであると考えられる．本研究ではモデルの作成時に 90 度で ITD の値が最大になると仮定し，最小二乗法を用い 1 次式で近似した．そのため図 9，図 10 のように誤差が現れたと考えられる．

図 11 の屋外で録音したデータを用いた実験結果では，推定結果が広範囲に広がってしまっている．これは本研究の手法では両耳に入ってきた音声の閾値を越えた時点として音声の発生点としているためと考えられる．屋外での測定では周辺の環境音やマイクにあたる風の音など，方向を推定したい音源が発する音声ではない音声が閾値を越えた点を音声の発生点としてしまう．環境音を検知し音声の発生点としてしまうと推定したい音源の方向ではない方向を推定してしまう．また，マイクにあたる風のように片側だけに音声の入力があると両耳信号で相関が正しく取れなくなってしまうため図 11 のような結果になってしまうと考えられる．そのような理由で両耳間時間差が検出できなくなる事に対しては音源分離を行うことで，目的の音声に対して ITD の検出を行うことができると考えられる．

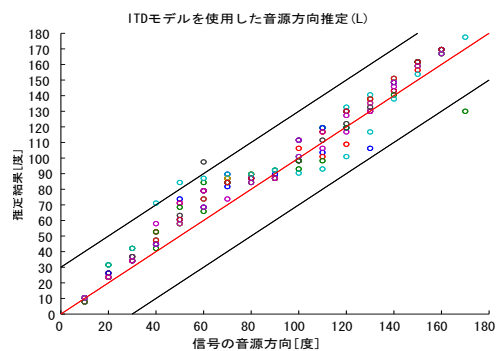


図 9 室内で測定した HRTF から作成した ITD モデルを用いた室内で録音した音声の音源方向推定実験結果 (L)

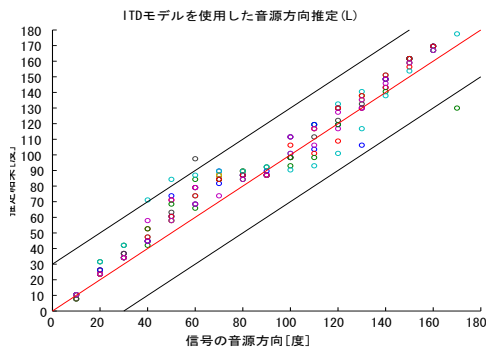


図 10 屋外で測定した HRTF から作成した ITD モデルを用いた室内で録音した音声の音源方向推定実験結果 (L)

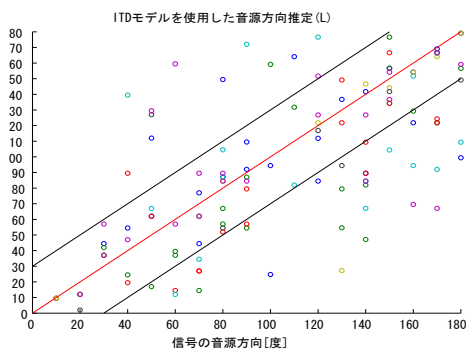


図 11 室内で測定した HRTF から作成した ITD モデルを用いた屋外で録音した音声の音源方向推定実験結果 (L)

表 1 室内で測定した HRTF から作成した ITD モデルを用いた室内で録音した音声の音源方向推定実験結果

音源方向	左側			右側		
誤差 [度]	±10	±20	±30	±10	±20	±30
%	81	95	98	71	86	93

表 2 屋外で測定した HRTF から作成した ITD モデルを用いた室内で録音した音声の音源方向推定実験結果

音源方向	左側			右側		
誤差 [度]	±10	±20	±30	±10	±20	±30
%	66	95	98	67	89	93

表 3 室内で測定した HRTF から作成した ITD モデルを用いた屋外で録音した音声の音源方向推定実験結果

音源方向	左側			右側		
誤差 [度]	±10	±20	±30	±10	±20	±30
%	25	33	40	10	22	31

5 まとめ

本研究では、測定した HRTF を用いて ITD のモデルを作成し音源方向を推定することを目指した。HRTF を測定した室内での音源方向推定実験では音源を設置した方向から誤差 ±10 度の範囲で 70% 程度、±20 度の範囲で 80% 程度、±30 度の範囲で 90% 程度の音源方向推定を正しく行うことが出来た。しかし、屋外で測定した音声の推定実験では ±30 度の範囲では 30% 程度の推定率となった。

この結果からあらかじめ測定した HRTF を用いてモデルを作成し、音源方向の推定を行う事が可能であると考えられる。また、室内で測定した HRTF を用いて作成したモデルを使用した音源方向推定実験と屋外で測定した HRTF を用いて作成したモデルを使用した音源方向推定実験から、±20 度程度の誤差を許容するのであれば、実環境下で測定した HRTF を用いずとも音源方向推定のための ITD モデルを作成することが可能であることが確認された。

本研究では水平面において 10 度ずつ HRTF を測定したが適当な HRTF の補間を行うことで、従来のバイノーラルマイクを用いた音源方向推定に関する研究で行われていたように [9]、細かい間隔ですべての方向の ITD,ILD を計測するのではなく、少ない点数で HRTF を測定しモデルを作成することで少ない HRTF のデータから全方位に対するモデルを作成することが出来る。そのため、個人の体質に合わせて音源方向推定モデルの作成を行うことが容易になると考えられる。

本研究では屋外で録音した音声の音源方向推定結果において良い結果が得られなかった。今後の課題として音源分離を行う事で目的の音声を抽出し ITD を検出する事により雑音・環境音の影響を克服する必要がある。また、本研究では HRTF から ITD のモデルを作成し音源方向の推定を行ったが、同様に各方位での HRTF の違いから生まれる両耳間パワースペクトル比を用いたモデル作成とモデルの有効性の検討を行う必要がある。

参考文献

- [1] イェンス・ブラウエルト 他, “空間音響” 鹿島出版会 (1986)
- [2] 伊藤一仁・赤木正人, “音源方向定位のための聴覚モデルの検討”, 電子情報通信学会技術研究報告. SP, 音声 97(587), (1988-03-06), pp57-64
- [3] 田村 智道 他, “ステレオ信号による音像の定位方向と音の大きさの関係についての一検討”, “電子情報通信学会技術研究報告. US, 超音波”, (2004-01) pp.57-62
- [4] 別府大幸 他, “音像定位受聴試験における刺激音に関する検討”, “電子情報通信学会技術研究報告. EA, 応用音響” 100(397), pp.9-16, (2000-10)
- [5] 岩谷幸雄, “バイノーラル収録・再生における課題: ヘッドホンの自由空間等価特性の影響 (<小特集>バイノーラル技術による立体音響再生の動向)”, 日本音響学会誌 61(7), (2005), pp.386-391
- [6] 小玉 亮介 他, “HRTF の低域の周波数特性の簡略化が定位に及ぼす影響に関する基礎的検討”, 電子情報通信学会技術研究報告. EA, 応用音響 109(166), (2009), pp.55-59
- [7] 田嶋 誠 他, “複数マイクロホンを用いた簡易な呼びかけ音方向同定システム” 電子情報通信学会技術研究報告. DSP, デジタル信号処理 102(333), (2002) pp.19-24
- [8] 馬渡 隆行 他, “拡張現実感における音声注釈情報取得のための仮想マイクロフォンの提案” 電子情報通信学会総合大会講演論文集, (2004), p.346
- [9] 章 忠 他, “バイノーラルモデルを用いた音源方向定位”, 日本機械学会論文集, C 編 74(739), pp642-649
- [10] “「拡張」方法の多様化する「拡張現実」”, <http://blogs.itmedia.co.jp/akihito/2010/04/post-6e8a.html>, (参照日 2011/1/14)
- [11] “空襲の記憶 - 音の拡張現実”, <http://akihitok.typepad.jp/blog/2008/10/-3586.html>, (参照日 2011/1/14)
- [12] 王 輝 他, “音波到来方向推定におけるマイクロホン間の位相差による影響について”, “電子情報通信学会技術研究報告. EA, 応用音響”, 101(479), (2001), pp43-50
- [13] 馬渡 隆行 他, “拡張現実感における音声注釈情報取得のための仮想マイクロフォンの提案”, “電子情報通信学会総合大会講演論文集”, (2004), p346