ヒューマンインタフェース 46－1
（1992. 1. 14）

# アクティブインタフェースの研究

岡田 孝文　山本 吉伸　安西 祐一郎
慶応義塾大学　計算機科学専攻

「パーソナルロボット」とは、近い将来オフィスや家庭などでパーソナルコンピュータのように利用されるロボットのことである。パーソナルロボットの設計では特に、エンドユーザが利用しやすいインタフェースを構築することが重要である。

本稿では、パーソナルロボットが活動する環境やパーソナルロボットの特徴を考慮し、そのインタフェースの設計について検討する。そして、それらを踏まえ「アクティブインタフェース」の概念を提案する。また、アクティブインタフェースの応用例として音声対話システム *SONIC* の実装について報告し、*SONIC* の試用を通して、提案した概念の有効性について検討する。

# A Study of Active Interface

Takanori OKADA, Yoshinobu YAMAMOTO, Yuichiro ANZAI
Department of Computer Science, Keio University

We believe that in the near future *personal robots* will be used in our office and home like personal computers in current society. One of the most important factors to realize such personal robots is to build a user friendly interface for end users.

In this paper, we discuss a user interface design for personal robots considering the environment where personal robots will be used and also discuss required features for personal robots. Then we propose a new user interface concept called *Active Interface* and introduce *SONIC* a dialogue system for human-robot interaction based on Active Interface concept. Finally we discuss the efficiency of Active Interface concept through the experimental use of *SONIC*.

# 1 はじめに

「パーソナルロボット」とは、近い将来オフィスや家庭などでパーソナルコンピュータのように利用されるロボットのことである。当研究室ではこのパーソナルロボットを中心に、人間、ロボット、計算機が混在する環境で、従来の「電子世界」から我々の活動する「物理世界」までを取り扱う研究を進めている [1]。既に開発されたパーソナルロボットのプロトタイプ Einstein I [2] では、オフィス内で書類運搬などのタスクを実行することが可能である。

ユーザがパーソナルロボットにアクセスする方法としては、次の二つが考えられる。

- ユーザが計算機 (ホスト) を介して間接的にロボットにアクセスする
- ユーザが直接ロボットにアクセスする

前者は、計算機ネットワークや無線通信を利用して遠隔地のロボットにタスクを依頼する場合や、定型業務などをバッチ的に処理させる場合、また複数のユーザで協調的に作業を行なう場合などに有効な方法であり、実際にこのような用途のためのシステムが実装されている [3]。この時ユーザは、マウス、ビットマップディスプレイ、ウィンドウシステムなど、従来の計算機のインタフェース技術 [4] を利用してロボットにアクセスすることができる。

これに対して何らかの理由で、計算機およびそのネットワークを利用できない場合や、近隣にいるロボットに対してインタラクティブに処理を依頼したい場合には、人間とロボットが直接インタラクションを行なう後者の方法が有効になってくる。

そこで本稿では、人間とロボットが直接インタラクションを行なう状況で、エンドユーザでも容易にパーソナルロボットを利用できるインタフェースの設計について検討する。

次章ではパーソナルロボットを取り巻く環境や、その特徴を考慮し、インタフェースの設計方針を述べる。次にそれらを踏まえて「アクティブインタフェース」の概念を提案する。最後に、この概念に基づいてロボット上に実装した音声対話システム *SONIC* について報告し、現在までに得られた実験データから、提案した概念の有効性について検討する。

# 2 インタフェースの設計方針

我々は、パーソナルロボットのプロトタイプを実際に利用した経験から、以下の項目を重視してインタフェースの設計を行なった。

なお本稿では「環境」という言葉を、対象となるユーザやその挙動までも含めた意味で使用する。「外界」という言葉には、ユーザに関する情報を含めずに使用する。

## 1. 環境の変化に対応する

パーソナルロボットは計算機が対象としていた「電子世界」だけでなく、我々が活動する「物理世界」までも対象にしなければならない。一般に、物理世界は動的に変化する。廊下を行き交う人々、室内の物の配置、路面状況、温度、騒音レベル、またユーザの状態も、位置、向き、発声音量など、さまざまな要素が変化する。

パーソナルロボットのインタフェースを設計する際には、ユーザを含めたこれら環境の変化に対応することを考慮に入れるべきである。具体的にはセンサ情報となって現れる環境の変化を、インタフェースに反映させる機構が必要である。環境の変化にユーザが対応するのではなく、インタフェースが対応することでユーザの負担を軽減することが望ましい。

## 2. ロボット固有の能力を活かす

パーソナルロボットには、物を動かす能力、移動能力など、物理世界に働きかける能力がある。ユーザと直接インタラクションを行なう際には、この能力を活かしてユーザに働きかけることが有効になると考えられる。

例えば、移動能力のあるロボットであれば、自らの位置を変更し、情報を獲得しやすい場所まで移動して、ユーザからの入力を得ることが可能になる。また、アームを搭載したロボットであれば、ユーザに物理的接触で働きかけ、注意を引き付けてから、何か重要な情報を伝えるようなインタフェースが考えられる。

このように物理世界に働きかける能力を用いて、より有利な状況でユーザから入力を得る／ユーザに出力を伝える考え方は、アクティブセンシングの概念 [5] をパーソナルロボットのインタフェースに拡張したものである。

### 3. ユーザが意識すべき入力を少なくする

パーソナルロボットが物理世界で活動するためには、数々のセンサ (特に外界センサ) が必要である。しかし、ロボットに搭載されたこれらのセンサを、エンドユーザに一つ一つ説明し理解させるのは困難な事である。ユーザに対してセンサをどのように見せるかは、パーソナルロボットのインタフェースを設計する上で重要な問題である。

我々は、可能な限りユーザにはセンサを見せない方針を取ることにした。「見せない」とは、外装のデザインによって物理的に隠すことではなく、ユーザに意識させないという意味である。

具体的には、システムに対してユーザが意識的に／明示的に入力すべきものを極力少なくし、その代わりにユーザが意識しない入力を取ってくるような能動的な機構を設ける。

これによって、システムの説明 (いわゆるユーザへの教育) を簡単なものにし、ユーザがインタラクティブに処理を進める場合でも、すべての指示を一つ一つ明示的に入力するというような負担を軽減できるのではないかと考えられる。

上記の項目を踏まえ、我々はパーソナルロボットのインタフェース設計のために、アクティブインタフェースの概念を提案する。

# 3 アクティブインタフェース

## 3.1 定義

ユーザからの明示的 (explicit な) 入力を待つだけでなく、ユーザが意識しない (implicit な) 入力や、外界の情報を取り入れ、これらの情報をもとに自ら行動を起こし、より有利な状況でユーザに接するインタフェース

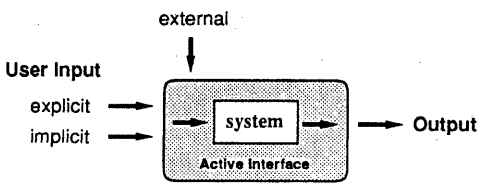


図 1: アクティブインタフェースの概念図

## 3.2 入出力

- explicit な入力:
  (キーボード入力、マウス入力、音声入力など)
- implicit な入力:
  (表情 [6]、体温、視線、音量、抑揚など)
- 外界の情報:
  (気温、騒音、湿度、現在位置など)
- 出力:
  (画面表示、発話、アクションなど)

explicit な入力は、**強い入力**とも呼ぶことができ、従来ユーザがキーボード、マウス、マイクロホンなどの入力デバイスから明示的に与えていたものがこれに相当する。

implicit な入力は、二種類の入力に大別することができる。一つは**弱い入力**と呼ぶもので、それ自身が独立して存在できるもの、つまり表情や体温、視線などがこれに相当する。

もう一つは**隠れた入力**と呼ぶもので、それ自身が独立して存在することが不可能なものである。これは前述の explicit な入力 (強い入力) に付随するものであり、explicit な入力があってはじめて存在することができる。この入力の例としては、音声入力に付随する音量や抑揚などが挙げられる。

出力に関しては、画面表示やスピーカからの音、発話などの他に、ロボットの場合には、アームや車輪の駆動なども出力に含まれる。

上記の分類は、インタフェースの設計段階における、設計者から見た解釈である。インタフェースの設計段階で、どのような入力をユーザに意識させ、またどのような入力をユーザに意識させないかを明確にするため、この分類は必要である。

## 3.3 能動性の意味

アクティブインタフェースの主張する能動性とは、ユーザに入力を促すような能動性ではない。このような能動性はユーザにとって差し出がましいものとなりがちで、快適なインタフェースとは言えないであろう。我々は以下に挙げる三点をアクティブインタフェースが主張する能動性とし、これによって最終的にユーザの負担を軽減することを目指した。

1. explicit な入力を待つだけでなく、implicit な入力を取ってくる能動性
2. より有利な状況でユーザに接するために、自ら行動を起こす能動性
3. ユーザの視点から見たシステムの振舞いに関する能動性

第一の能動性は、ユーザの implicit な入力に着目し、これを積極的に獲得することである。従来の多くのインタフェースが入力を待つだけであったのに対し、この機構はユーザがすべての入力を明示的に与える必要がないため、入力に関する負担を軽減できると考えられる。また、implicit な入力や外界の情報など他の入力を参照することで、ユーザからの explicit な入力をより正確に解釈できることも期待できる。

第二の能動性は、与えられた状況に捕らわれず、implicit な入力や外界の情報をもとにロボット固有の能力を生かして自ら行動を起こし、より有利な状況でユーザに接することである。これによって、よりユーザに入力させやすく、また、システムが入力を獲得しやすくなり、出力に関しても、より確実にユーザに情報を伝達できるようになると考えられる。

第三の能動性は、ユーザに対するシステムの振舞いである。ユーザは自分で何も明示的に入力していないにも関わらず、システムは implicit な入力や外界の情報に反応し出力がある。すなわちアクティブインタフェースによって、システムに能動性があるようにユーザに対して振舞うことが可能になると考えられる。

我々はこのアクティブインタフェースの概念を、パーソナルロボット上で音声対話システムを利用する場面に応用した。次章では、音声対話システム *SONIC* について紹介する。

# 4 音声対話システムへの応用

## 4.1 ロボット上で利用する際の問題点

音声対話システムはユーザにとって親しみやすい入出力デバイスの一つである。しかし、現行の音声対話システム、特に音声認識システムはノイズに対して弱い事が指摘されている。パーソナルロボットのインタフェースとして利用することを考えた場合、システムをとりまく環境は話者も含めて動的に変化していくため、十分な認識率を得ることは難しい。

上記の問題に対する一手段として、ユーザに接話型マイクロホンやヘッドホンを利用させることが考えられる。しかし、この方法はユーザに入出力デバイスを強く意識させることになり、あまり好ましいとは言えない。

我々は単純に音声対話システムをパーソナルロボットのインタフェースに利用するだけでなく、最終的には人間同士が日常行なっているような、物理的にも意味的 [9] にも自由度の高い対話を、ユーザとパーソナルロボットとの間で実現したいと考えている。

従って、入力部には非接話型の指向性マイクロホンを、出力部には通常のスピーカを用いて、パーソナルロボット上で音声対話システムを利用する方針を取ることにした。

音声対話システムにとって問題となる環境の変化を挙げると次のようになる。

- 周囲の騒音レベルの変化
- 話者の位置変化 (方向と距離)
- 話者の発声音量の変化

これらの変化に対して音声認識システム、音声合成システムでは、それぞれ次のような問題点が予想される。

**音声認識** 認識率の低下によりユーザからの入力が得られない。

**音声合成** 騒がしい環境ではユーザに十分な情報を伝達できない。また、静かな環境では不適切な音量でユーザに不快感を与える。

近年、ノイズに強い音声認識システムが報告されている [7][8]。これらの技術に加え、ロボット固有の能力を生かしたりセンサ情報を利用することで、さらに認識率の低下を防ぐことができないであろうか。

## 4.2 アクティブインタフェースの応用

我々は、話者からの音声入力だけでなく、その音量や音源の方向、話者までの距離、周囲の騒音レ

ベルといった情報にも着目し、これらの情報をアクティブインタフェースの概念に対応させた。

そして出力としては音声認識に有利な状況を獲得するために、パーソナルロボットの移動能力を利用して話者の位置変化に対応することを考えた。また、音声合成では周囲の騒がしさ／静けさに応じて音量を変化させることを考えた。

- 入力
  - 話者からの音声入力 (explicit)
  - 話者の音量 (implicit)
  - 話者のいる方向 (implicit)
  - 話者までの距離 (implicit)
  - 周囲のノイズレベル (external)
- 出力
  - 話者への接近、追尾
  - 発話音量の変化

## 4.3 音声対話システム *SONIC*

*SONIC* は、センサ情報とロボットの移動能力を利用したパーソナルロボットのための音声対話システムである。システムは図2の自立移動ロボット上に実装されており、下記の各モジュールはVMEバスを介して図3に示す構成を成している。音声認識モジュール、および音声合成モジュールは、ボード型のシステムをロボット本体内に搭載することも可能であるが、現在は据え置き型のシステムをロボットの外部に設置して利用している。これらはRS-232C によってメインボードに接続されている。

システムの構成：

- 音声認識システム (VC-171)[1]
- 音声合成システム (しゃべりん坊 HG)[2]
- 騒音レベル監視モニタ
- 3 つのマイクによる音源センサ
- 8 方向赤外線熱源センサ
- 4 方向超音波距離センサ
- 左右輪 2 つのモーター

*SONIC* の主な機能は次の通りである。

[1]松下技研 (株)
[2]NTT インテリジェントテクノロジ (株)

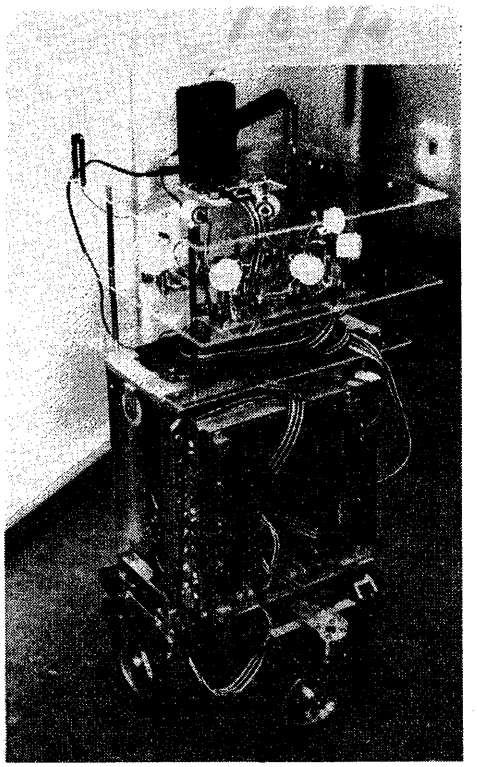

図 2: 自律移動ロボットと *SONIC* の概観

- ユーザとの対話中に周囲の騒音レベルが上昇すると、ロボットが自らアクションを起こし、十分な認識率を得られる位置まで話者に接近する。
- 話者の位置が変化した場合、各センサから得られた情報に基づき、現在の状態を対話に適した状態に近づけるように自ら行動を起こす。
- 音声合成の発声音量も、周囲の騒音レベルに応じてユーザが聞きとりやすいように変化する。

話者のいる方向を決定する機構では、音源センサの情報と熱源センサの情報を利用している。

音源センサでは一辺が30[cm] の正三角形の頂点に配置された 3 つのマイクロホンに到達する音の時間差から音源の方向を算出する。これは音源定位の中でも特に方向定位としてよく知られた手法であり、近年ではより高度な定位を実現したシステムが報告されている [10]。実際のサンプリングはロ

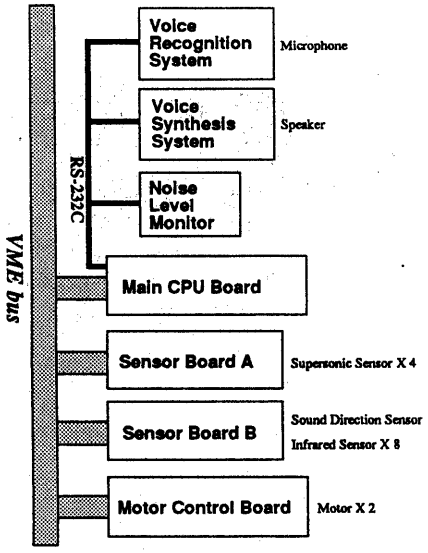


図 3: *SONIC* の内部構成

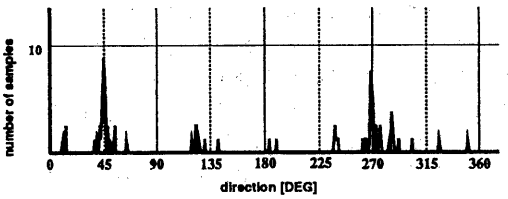


図 4: 音源センサの情報

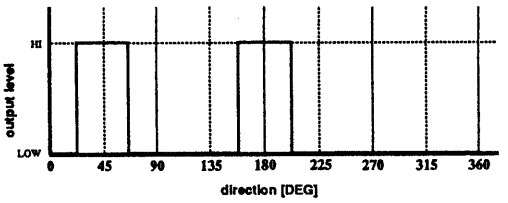


図 5: 熱源センサの情報

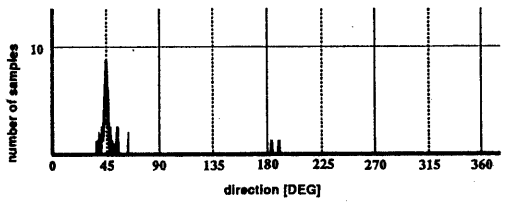


図 6: 掛け合わせた情報

ボットが現在位置に停止してから開始し、1[DEG] きざみで図4のようにデータが蓄積されていく。

熱源センサは人体が放出する赤外線に反応するもので、現在は全方位を 8 個のセンサで分担している。実際のデータは図5に示すように、1 個のセンサが担当する 45[DEG] の範囲内において、反応がある／ないの二値出力で得られる。

話者のいる方向は、上記の二種類のデータを掛け合わせることによって決定する。図6は、前述の音源センサの情報 (図 4) と熱源センサの情報 (図 5) を掛け合わせたデータである。このデータから最もサンプル数の多かった方向が話者のいる可能性が高い方向として、モーターの制御部へ伝達される。

方向を決定した後は、超音波センサからの距離情報、および音声認識システムの出力を参照することによって、対話に十分な位置まで話者に近づく。

# 5 実験による評価

*SONIC* で提供された機能の有効性を検証するため、我々はまず認識率に関する実験を行なった。この実験では周囲の騒音レベルが上昇した場合に、ロボットが自ら話者に近づくことによって、どの程度認識率の低下を抑えることができるかを測定した。

## 5.1 実験環境と方法

実験で使用した騒音のサンプルは、本研究室内で収録した実雑音で、足音、話し声、コピー機の動作音などの非定常雑音を含んだものである。このサンプルは本研究室のパーソナルロボットが実際に活動する環境の騒音そのものであり、一般のオフィスにおける騒音と特に差異はないと考えられる。

ノイズレベルは上記の収録音を再生する際の音量によって変化させ、これを JIS C-1502 の A 特性に準拠した騒音計を用いて測定、およそ 30-70[dB(A)] までの範囲を利用した。

認識に用いた単語候補数は 10 個で、この辞書に

は書類運搬のタスクで実際に利用される人名が登録されている。音声入力は、辞書中から任意の一単語を選択、発声し、これを録音したものを用いた。再生は騒音レベル30-70[dB(A)] の範囲を5[dB(A)] ごとに9段階について、それぞれ20回づつ一定の音量で繰り返し行なった。

話者への接近は50[dB(A)] から機能させ、2.0-0.5[m] までを50[cm] ごとに、実験的に4段階に変化させた。

### 5.2 結果

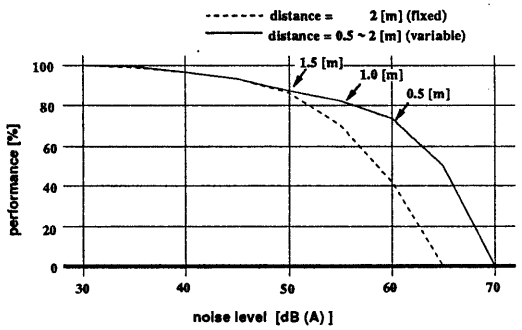


図7: 話者に接近した場合の認識率

図7において、横軸は実験環境の騒音レベル、縦軸は30[dB(A)] における音声認識システムの認識率を100[%] とした場合の相対値である。点線は話者とロボット間の距離を2[m] に固定した場合の認識率を表している (固定)。実線は接近機能を働かせた場合の認識率を表しており、注釈はその際の話者とロボットの距離を示している (接近)。

### 5.3 考察

今回の実験では、固定／接近それぞれの場合で認識率の比較を行なうため、両者の実験条件ができるだけ同じになるように配慮した。従って、音声や騒音の再生に用いたテープレコーダの音質、スピーカの向き、また入力で用いたマイクロホンの性能等の影響によって、認識率の絶対値は参考にならない。この理由から図7の縦軸は相対値による表現を用いた。

実験は当研究室内において早朝の静かな時間帯を選んで行なったが、無響音室のように厳密な環境ではないため、実験環境そのものの非定常雑音の影響が固定／接近の両者に対して全く同じ条件ではなかった。

以上の点から、数値の詳細に関して多少の誤差を含むことが考えられる。しかし実験結果のあらわす全体の傾向としては、明らかに接近機能を働かせた場合の方が、固定よりも高い認識率を得られることが分かった。

## 6 まとめ

本稿では、人間とロボットが直接インタラクションを行なう状況において、パーソナルロボットのインタフェースの設計について検討した。そしてアクティブインタフェースの概念を提案し、この概念に基づいた音声対話システム *SONIC* について報告した。

現在までの実験評価では、話者への接近機能による優位性があらわれ、提案した概念の有効性について見通しを得ることができた。

## 今後の課題

今回の実験では、*SONIC* で提供した機能のうち、接近機能に関する評価だけを行なった。今後は残りの機能に関する評価も行なっていく予定である。

なお *SONIC* では話者の位置を決定する機構に、音源センサと熱源センサの情報を用いているが、これらの情報だけではユーザを認識するのに十分とは言えない。特に複数の話者がロボットの周囲にいる場合は、対象となるユーザの方向を決めるのが困難である。今後はこれらの問題に対処するため、画像情報を取り入れたシステムの改良を考えていきたい。

また、アクティブインタフェースはユーザの意識しない入力を扱うため、時としてユーザが意図しないような結果をもたらす可能性がある。ユーザから能動的に入力を獲得する機構は、ユーザの入力負担を軽減できるという利点がある反面、ユーザを不安にさせる等の心理的問題が生じる可能性もある。今後は、これらの問題点についても十分な検討をしていきたいと考えている。

## 謝辞

*SONIC* の実装に関して協力して頂いた安西・天野研究室の皆様、また日頃から活発な討論をして下さったHUROBINT (学生によるヒューマンインタフェース研究会) の皆様、そして音声対話システムに関して御協力頂いた松下技研 (株)、NTT インテリジェントテクノロジ (株) の皆様に感謝します。

## 参考文献

[1] Y.Anzai, "Towards a New Paradigm of Human-Robot-Computer Interaction", *Proc. of IEEE Int. Workshop on Robot and Human Communication*, pp.11-17 (1992).

[2] 山崎信行, 安西祐一郎, "パーソナルロボットのためのアーキテクチャの提案", 日本機械学会 [No.920-33] ロボティクス・メカトロニクス講演会'92(川崎), 講演論文集, Vol.A, pp.51-56 (1992).

[3] Y.Nakauchi, T.Okada, N.Yamasaki and Y.Anzai, "A multi-agent interface architecture for human-robot cooperation", *Proc. of IEEE Int. Conf. on Robotics and Automation*, vol.3, pp.2786-2788 (1992).

[4] D.C.Smith, C.Irby, R.Kinball, B.Verplank, and E.Harslem, "Designing the Star User Interface", *Byte*, Vol.7, No.4, Apr. (1982).

[5] R.Bajcsy, "Active Perception", *Proc. of IEEE*, vol.76, No.8, pp.996-1005 (1988).

[6] H.Kobayashi and F.Hara, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network", *Proc. of IEEE Int. Workshop on Robot and Human Communication*, pp.381-386 (1992).

[7] 平岡省二, 二矢田勝行, 木村達也, "ワードスポッティング手法を用いた不特定話者・小数語向け音声認識装置", 信学技法, SP88-18 (1988).

[8] 竹林洋一, 金澤博史, "ワードスポッティングによる音声認識における雑音免疫学習", 信学論 (D-II) Vol.J74-D-II, No.2, pp.121-129, (1991).

[9] 佐藤倫太, 開一夫, 安西祐一郎, "ロボットとの対話: センサ情報を利用した音声対話システム Linta の設計と実装", 情報処理学会 研究グループ資料, 92-SLP-1, 電子情報通信学会 時限専門委員会資料, SPREC-92-1, 人工知能学会 研究会資料, SIG-SLUD-9202-3, 「音声言語処理と対話理解」に関する共催研究会 (諏訪),pp.19-26, (1992).

[10] J.Huang, 大西 昇, 杉江 昇, "生体に示唆を得た音源定位システム–反響のある環境での単一音源定位–", 信学論 (A) Vol.J71-A, No.10, pp.1780-1789, (1988).