

高信頼性組み込みソフトウェア構築技術

北陸先端科学技術大学院大学 情報科学研究科 片山卓也

研究分担者 早稲田大学理工学部 コンピュータネットワーク工学科 中島達夫

研究分担者 京都大学 湯淺太一

研究分担者 国立情報学研究所 中島震

協力企業 NEC, NEC エレクトロニクス, 松下電器産業, Nokia, オムロン, オムロンソフトウェア

1. プロジェクトの背景

組み込みソフトウェアは、家電製品、自動車、携帯端末、制御機器などのあらゆる工業製品の心臓部に組み込まれ我々の生活を支えており、その機能や品質はこれら製品や機器の価値を決める最も重要な要素である。特に、ソフトウェアの不具合による経済損失などが報告される現在、高信頼性は組み込みソフトウェアに対する最も重要かつ緊急な要求である。

従来、組み込みソフトウェアはそのサイズが余り大きくなかったこと、また、比較的単純な機能の実現を行えば良かった事などもあって、その開発には最新のソフトウェアテクノロジーが用いられてこなかった。しかしながら、現在では、高度なユーザインタフェースや通信機能など製品に要求される機能が高度化すると同時に、利用可能な CPU やメモリなどのハードウェア資源に対する制約が緩和したことなどによって、組み込まれるソフトウェアが大規模化・複雑化し、これまでのソフトウェア開発方法論が十分に機能しなくなりつつある。最新のソフトウェア開発技術を組み込みソフトウェアの開発に投入し、組み込みソフトウェア開発を前進させることが強く求められている。

本プロジェクトでは、このような観点から高信頼組み込みソフトウェア構築の問題を、(1) 構築環境、(2) 実行環境、(3) 実行基盤の3つの視点から総合的に解決することを目的とし、企業との密接な連携のもと、プロジェクト終了時に産業界への技術移転を目標に、以下の課題に関する研究開発を行っている。

- (1) 組み込み用オブジェクト指向分析設計技術
(北陸先端科学技術大学院大学)
- (2) 組み込みシステム向け基盤ソフトウェア
(早稲田大学)
- (3) 組み込み用実時間 Java 技術
(京都大学)

2. 組み込み用オブジェクト指向分析設計技術

ソフトウェア開発上の問題の多くが分析・設計などの上流工程で作こまれ、それが全体の生産性や品質に支配的な影響を及ぼしていることはよく知られている。(1) では、組み込みソフトウェアの特性を適切に扱うことが可能で、上流工程段階から正しさを確認・検証しながら組み込みソフトウェアを構築できる方法論とそれを支える環境の実現に関する研究開発を行って

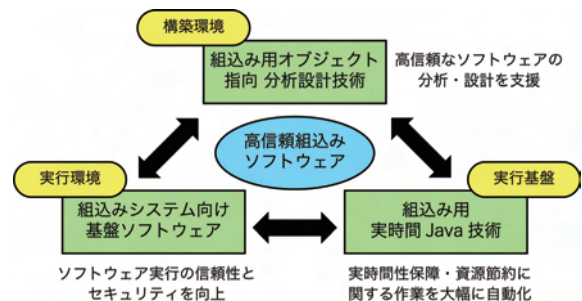
る。具体的には、UML による記述、モデル検査や定理証明技術による検証、製品系列やアスペクト概念による設計などを可能にする環境の実現を行っている。

3. 組み込みシステム向け基盤ソフトウェア

携帯電話やデジタルテレビ、あるいはこれから登場する先進情報アプライアンスにより様々な新しいサービスが提供されることが予想される。これらの次世代の情報端末や情報アプライアンスには、高度な通信や協調動作、セキュリティ、センシングなど多くの機能が要求されるが、(2) では、これらを実現する高信頼コンポーネントやミドルウェア、それらを動作させるオペレーティングシステムなどの基盤ソフトウェアの研究開発を行っている。

4. 組み込み用実時間 Java 技術

組み込みソフトウェアでは、一定の限られた時間内にレスポンスを返すこと、また、限られたメモリで動作することなど、動作時間や利用可能資源などに厳しい制約がついているのが普通である。従来はプログラマが細心の注意を払って、職人芸的技術によってこの問題を解決しており、開発コストおよび品質の両面で大きな問題となっていた。(3) では、これらの作業を最大限に自動化し、信頼性の高い組み込みソフトウェアの生産性を高めるための研究開発を行っている。



組込み用オブジェクト指向設計技術

北陸先端科学技術大学院大学 情報科学研究科 岸知二, 青木利晃, 片山卓也

国立情報学研究所 中島震

協力企業 NEC, NEC エレクトロニクス

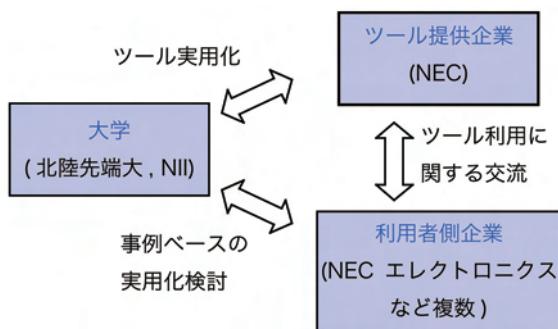
<http://kt-www.jaist.ac.jp/project/esociety>

1. プロジェクトの概要

本プロジェクトでは、組込みソフトウェアの信頼性の向上に関し、分析・設計といった開発の上流工程に焦点をあてて取り組んでいる。ソフトウェアの問題の多くが上流工程において作りこまれることはよく知られており、本プロジェクトでは高信頼性ソフトウェア開発のための分析・設計の手法や環境を研究している。

研究にあたっては、ソフトウェア開発に関する最新の工学的、科学的成果を、実際の組込みソフトウェア開発に適用することを試みている。たとえば、主としてワークステーションやパソコン上のソフトウェア開発を対象として研究や適用がなされているオブジェクト指向開発やアスペクト指向開発、プロダクトライン開発といった工学的成果を、組込みソフトウェアの開発に応用する。また従来宇宙・航空・軍事といった分野でのみ利用されてきた、高度だが高コストで利用の難しかった形式的手法やシミュレーションなどの科学的成果を、家電製品のような民需分野での組込みソフトウェア開発に適用可能にしようとする研究を進めている。

こうしたねらいを達成するためには、実際の組込みソフトウェア開発上の問題や要素技術の特性を把握し、問題解決に必要な技術が実際のソフトウェア開発に適用可能となるように、適用手法や環境を整備しなければならない。そのため、事例研究に基づく実問題の検討、それを踏まえた手法やツールの開発、中期的に必要な基本的な技術の開発という観点から、図に示す体制で研究を進めている。



2. 本年度の成果

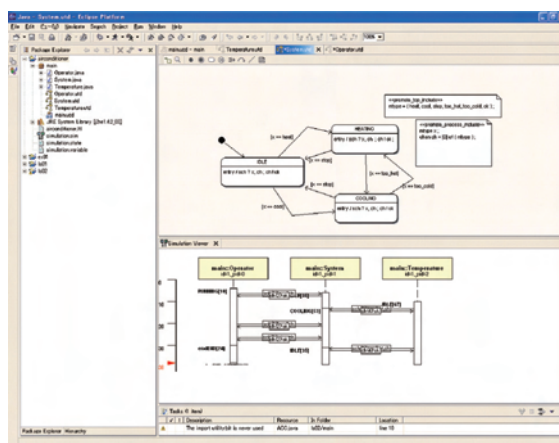
2.1. 事例研究

協力企業からの協力で、以下の事例研究を行った。

- ・テストシナリオに基づく検証: リアクティブシステムを対象としたふるまいの検証
- ・プロダクトライン開発への応用: 上記検証モデルの再利用と製品系列開発の中での体系化の検討
- ・リアルタイム OS のメカニズムの検証: スケジューリングやキューなどの動作を踏まえたシステムの検証

2.2. ツール開発

昨年度のプロトタイプ開発とその評価に基づき、実事例への適用に必要な機能を追加し、評価版として開発を行った。例えば検証結果の反例を解析するための機能を持った実行シーケンス表示機能、UML モデル上での性質を検証するために記法を拡張した LTL 定義機能、特定の状況における検証を手軽に行うための開始状態指定機能などを実装し、評価を進めた。



2.3. 各種基本技術の検討

設計検証に関わる、以下のような様々な基本的技術の研究開発を行った。

- ・分析・設計モデルの検証手法の検討: 定理証明の活用やその再利用方法、モデル検査手法によるセンサーの扱い方の検討
- ・組込みソフトウェアの設計検証の検討: 状態モデルの標準的な動作規則の検証上の問題点の検討と、新たな手法の提案
- ・並行性解析: オブジェクト指向のモデルを、応答時間などの解析が行いやすい処理中心のモデルへ変換する手法の検討

組込みシステム向け基盤ソフトウェア

早稲田大学理工学部 コンピュータネットワーク工学科 中島達夫

協力企業 松下電器産業, Nokia

<http://www.dcl.info.waseda.ac.jp/>

1. 次世代組込み環境と信頼性

組込みシステムもユビキタス時代の到来により日用品にコンピュータを埋め込み情報を扱う様々な情報アプライアンスが出現するものと思われる。情報アプライアンスは単体で機能するのではなく、組み合わせることにより様々な新しいサービスを作り出すことを可能とする。しかし、情報アプライアンスのソフトウェアは巨大なものとなり、ソフトウェアのバグによりシステム全体の信頼性を保証することが困難となる。また、情報アプライアンスはインターネットに接続されるため、様々なセキュリティに対する脅威からシステムを守る必要がある。組込みシステムは従来のエンタープライズシステムと異なり、リソースの制約を考慮したり、コンピュータの専門家以外の人でも用意に扱える必要があるため、従来の手法と異なる新しい技術の開発が必要となる。

2. 2004 年度プロジェクト概要

早稲田大学のグループでは、組込みシステムの信頼性を向上する 3 つのオペレーティングシステム開発の取り組みに関して取り組んでいる。ここでは、それぞれの OS における 2004 年度の成果概要を紹介する。

2.1. 組込み Linux における QOS サポート

現在の Linux は、悪意のあるプログラムをダウンロードして実行すると、そのプログラムがシステムリソースを占有することにより、システム全体が正しく動かなくなってしまうことがありえる。また、マルチメディアアプリケーション等のリアルタイムアプリケーションが CPU を占有すると非リアルタイムプロセスとして実行される GUI などの処理が動かなくなってしまう可能性がある。我々のプロジェクトにおいて開発中の Linux は、Linux に CPU アカウンティング機構と階層型スケジューラを追加することによりリソース占有を防ぐことを可能とする。

2.2. μ カーネルに基づくオペレーティングシステム

組込みシステムでは、外界からのイベントにタイムリに反応するため、一定時間内に応答を返すことが重要なアプリケーションが多数ある。組込み Linux はそれらの要求に対応するため、カーネルを横取り可能にするなどの改良がおこなわれてきた。しかし、 μ 秒オーダの高速な応答性を達成することは難しい。そのため、Linux on ITRON のようなリアルタイムオペ

レーティングシステム (RTOS) 上で Linux を動かすハイブリッドアーキテクチャが開発されてきた。しかし、従来のハイブリッドアーキテクチャは、RTOS や RTOS 上のアプリケーションが Linux カーネルのカーネルスペース内で動作するため、RTOS 上のアプリケーションのバグによりシステム全体がクラッシュしてしまう可能性がある。コード量の増大によりソフトウェアのバグも増加するため、バグが存在してもシステムが動作し続けるような仕組みをオペレーティングシステムが提供することは、ますます複雑化する組込みシステムを開発する上で非常に重要なことである。

2.3. 次世代携帯端末とミドルウェア

センサーは外界の情報を取得するアプリケーションを構築するために必要不可欠なデバイスである。センサーを利用することにより、ユーザが誰でどこにいるか、何をしようとしているかなどの実世界情報を取得し、アプリケーションの動作をユーザの振る舞いに適応することを可能とする。ユビキタスコンピューティング環境の発展に伴い、組込みオペレーティングシステムにおいてどのようにセンサーを利用するかを明らかにすることは非常に重要になってきている。

我々は、ノキアジャパンと協力して様々なセンサーを搭載した携帯端末を作成した。本システム上では Linux が動作し、センサーをアクセスするデバイスドライバを介してアプリケーションは様々なセンサー情報を取得することが可能である。センサーデータをアプリケーションから利用可能とするためのミドルウェアを利用することによりコンテキストウェアなアプリケーションを容易に構築することが可能となる。

組込みシステム用実時間 Java 技術

京都大学大学院情報学研究科 湯浅太一

協力企業 オムロン、オムロンソフトウェア

<http://www.yuasa.kuis.kyoto-u.ac.jp/>

1. 目的と目標

本プロジェクトは、オブジェクト指向言語 Java によって記述される組込み実時間アプリケーションの開発を効率化するための諸技術を開発するものであり、次の三つのサブテーマから成る。

- ・実行基盤の開発（平成 15 年度～ 16 年度）
- ・実証実験（平成 17 年度）
- ・要求仕様検証技術との統合（平成 18 年度～ 19 年度）

1.1. 実行基盤の開発

実時間組込みソフト固有の開発コストを軽減するための、自動化機能を備えた Java 処理系実装方式を開発し、試験的実装を行う。これによって、従来はプログラマの職人芸に頼っていた手作業の開発から脱却し、開発コストと開発期間の大幅な削減を目指す。

1.2. 実証実験

上記実行基盤を使用することによって、開発効率がどの程度向上するかを、実用レベルの組込みシステムの開発を対象として実証実験を行う。この実証実験の結果に基づいて、「真の実時間組込み」Java 処理系を構築する。

1.3. 要求仕様検証技術との統合

他のプロジェクトで開発される要求仕様検証のための基本的技術を、具体的に上記自動化機能を備えた処理系に適用し、要求仕様検証技術を確認する。これによって、仕様記述および検証労力の大幅な削減を目指すとともに、検証技術の実用化、安心して使える組込みソフトの実現を目指す。

2. 実行基盤の重要課題

Java による組込み実時間アプリケーション開発において特に重要な研究課題である「ごみ集め」による停止時間とメモリ制約の 2 点について、研究・開発を行った。

2.1. ごみ集めによる停止時間

Java はオブジェクト指向言語であり、アプリケーションの実行中に、データを動的に生成する。実行が進むと、メモリ領域が満杯となるために、不要なデータを回収して再利用する「ごみ集め」の処理が必要となる。通常の Java 処理系の場合、アプリケーションを一時的に中断し、ごみ集め処理を行う。し

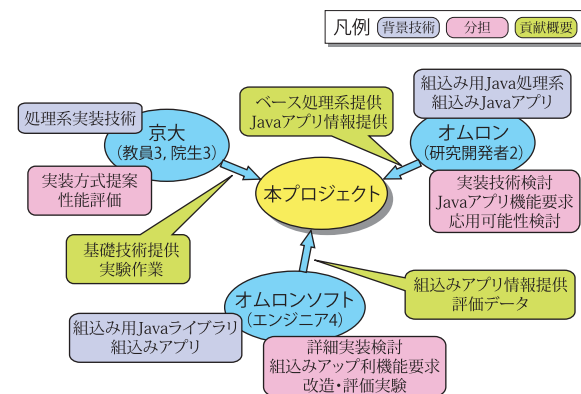
かしこの方法では、中断時間の予測が難しくアプリケーションの実時間性を保証できない。また、メモリ管理の方法によっては、データの生成時にも実時間性が問題となることがある。この問題に対応するために、拡張した「実時間対応」機能を備えた処理系を利用することがある。たとえば、ごみ集めの対象とならないメモリ領域を利用して一括廃棄するといった機能である。しかし、このようなプログラミングは、基本的に手作業で行われ、バグの原因となる可能性が高いうえに、作業工数が増加するという問題があった。これらを解決するために、本プロジェクトでは、「実時間対応」機能を使わないでも実時間性が保証できる実装技術を開発した。

2.2. メモリ制約

組込みシステムの搭載メモリ量は増加する傾向にあるが、通常のコンピュータのような仮想記憶は期待できない。限られたメモリ容量内で効率よくアプリケーションを実行する必要がある。メモリを有効利用するために、コード（クラス情報やメソッド本体など）やデータ表現を縮小する努力は必要であるが、これらの方法によるメモリ使用量の改善には限界がある。本プロジェクトでは、メモリを上手に「やりくり」することによって、与えられたメモリを有効利用する技術を開発した。従来は、この点でもプログラマの勘と試行錯誤に頼っていたのが実情であるが、アプリケーションに応じたメモリのやりくりを自動化することによって、職人芸に頼らない開発を可能にし、開発コスト削減と安定した組込みソフトの提供が期待できる。

3. 実施体制

本プロジェクトは、オムロン（株）およびオムロンソフトウェア（株）と共同で実施している。その概要を次の図に示す。



次世代高性能コンピュータシステム上の 高信頼ソフトウェアシステムの開発支援技術

東京大学 大学院情報理工学専攻 コンピュータ科学専攻 石川 裕

<http://www.il.is.s.u-tokyo.ac.jp/esociety/>

1. 次世代高性能コンピュータシステム

複数のプロセッサやI/O 機器が動的に接続 / 拡張可能なハードウェアを有する次世代高性能アーキテクチャが出現した。今後、これら機器は 10Gbps の性能を有する高速 LAN で接続されるようになる (図 1)。このようなアーキテクチャは従来のサーバコンピュータにとって変わることが可能となる。しかし、当該アーキテクチャ上で高信頼性を実現するためのシステムソフトウェアが未整備であり、現状、当該アーキテクチャの利用形態は限られたものとなっている。次世代高性能コンピュータシステム上で高信頼性を実現するためには、プロセッサ間通信機構、オペレーティングシステム、運用ツール等、多岐に渡りシステムソフトウェアを開発しなければならない。このようなソフトウェア開発のための支援技術および基盤ソフトウェアモジュールが求められている。

本研究では、次世代高性能コンピュータシステム上で信頼性を実現するシステムソフトウェアを高信頼かつ効率的に開発することを可能とするため、計算機ハードウェアやネットワークの障害等を模擬するためのソフトウェア環境を構築する。保守・運用ツールなどの基盤ソフトウェア開発に必要な基本通信機構として、保守・監視系のための通信機構、OS 核間で使用する高速通信機構の開発を行う。これにより、従来、高信頼ソフトウェアの動作検証に有していた時間を 1/10 以下に短縮することを可能とする。

2. 管理保守ソフトウェア & 故障模擬ソフトウェアシステム

近年の PC サーバでは、IPMI と呼ばれる規格に基づいたハードウェア障害をモニタリングするプロセッサ (BMC) がボード上に搭載されている。IPMI 機能を利用することにより、ハードウェア障害状況に応じた耐故障機能を組み込んだ高信頼シ

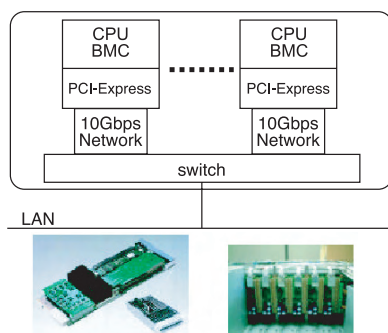


図 1 次世代高性能コンピュータシステム

テムソフトウェアが実現可能となる。このような高信頼システムソフトウェアが設計通りに実現されているかをテストするために、ハードウェアの故障模擬ソフトウェアシステムを開発している。IPMI 規格に基づくハードウェア障害情報を集約するソフトウェアモジュールを開発し、高信頼システムを支援する保守監視系ツール Tenjin を開発している (図 2)。本ツール開発において、ハードウェア障害を検知した時の対処を記述できる新しいスクリプト言語を設計開発している。

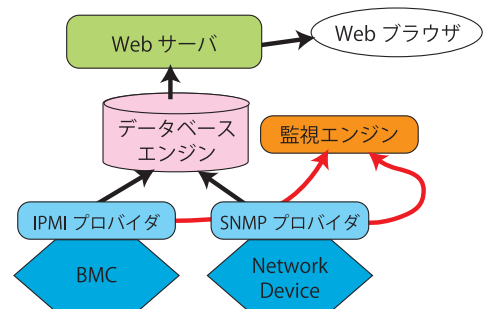


図 2 Tenjin ソフトウェアアーキテクチャ

3. 高速通信機構 & 故障模擬・開発支援システム

10Gbps 級 LAN では、1.5Kbyte 長パケット (Ethernet のパケット長は 1.5Kbyte) が 1.2 マイクロ秒で到着することになる。ネットワークプロトコル処理を従来のようにホストプロセッサで処理すると、その処理のオーバーヘッドが高くなり、物理ネットワーク性能を生かすことができなくなる。このため、ネットワークカード側でプロトコル処理を行なうとともに、リモートメモリ間通信機能の開発を行っている。

通信機器の故障に対応可能な高信頼ネットワークプロトコル処理ソフトウェアを実現するためには、通信機器の故障によるパケット喪失、CRC エラーなどを模擬し、通信ソフトウェアが頑強に、かつ性能を極端に劣化することなく故障に対処することを確認する必要がある。このような開発支援システムは、ソフトウェアだけでは実現できないため、パケット喪失、CRC エラーなどを生成可能とするハードウェア (図 3) を開発すると共に、そのハードウェアを使用した支援ツールを開発している。



図 3 故障模擬装置付 10Gbps ネットワークカード

プログラム自動解析に基づく高信頼ソフトウェアシステムの構築技術 —高信頼プログラミング言語と高信頼プログラム開発環境—

北陸先端科学技術大学院大学 情報科学研究科 大堀 淳
研究分担者 東京工業大学 渡部卓雄
協力企業 沖電気工業, 算譜工房
<http://www.jaist.ac.jp/~ohori/e-society/>

1. 高信頼言語技術の必要性

今実現しつつある高度情報化社会は、社会の制御機構の中核が、多様で膨大なソフトウェア群によって担われる社会である。この事実は、種々の社会システムの安全性や信頼性が、その社会基盤を実現するソフトウェア群の正しさと堅牢性に直接依存することを意味する。このような社会が、従来通りの信頼性と安全性を確保しながら発展していくためには、社会基盤としての高信頼ソフトウェアシステムを効率よく構築する技術を確立する必要がある。そのために、特に以下の2点の実現が緊急に求められている。

- ・プログラムの誤りをコンパイル段階で検出し高機能かつ高信頼プログラムの効率的生産を可能にする技術、
- ・CやJAVAなどの既存言語で書かれたものを含むプログラム部品を組み合わせてシステムを構築する際の不整合の自動検出やインターフェイスの自動生成を行う技術。

これら機能の実現には、従来のソフトウェア開発技術を越えた新たな技術革新を必要とする。その最も有望な基盤が、型理論に基づくプログラムの自動解析技術である。

2. 本課題の具体的な目標と実施体制

本プロジェクトの目的は、型理論とそれに基づくプログラムの静的解析技術を基礎とし、上記2点の課題を実現する上で重要な次世代高信頼プログラミング言語及び次世代高信頼言語を既存の言語とともに使用し高信頼ソフトウェアを効率よく開発するためのプログラム開発環境を構築することである。具体的には、以下の2点を目標とする。

- (1) 次世代高信頼言語 SML#. SML#(仮称)は、静的型システムを持つ高信頼言語である Standard ML を、我々の基礎研究の成果を取り入れて拡張した次世代言語であり、本プロジェクトで言語仕様、コンパイラ、実行時処理系を開発する。
- (2) 高信頼プログラム開発環境. SML# および Java を含む多言語環境での高信頼プログラム開発環境を支援する以下のツール部を開発する。
 - A) SML# のプログラミングを支援する種々のプログラミング自動化ツール
 - I) Java のプログラム部品を組み合わせて使用する際の不整合を自動検出するツール
 - ウ) 以上のツールを取り入れた多言語高信頼プログラム開発環境これら先端ソフトウェアを、システム開発現場にて活用で

きる実用性あるソフトウェア基盤として開発するためには、ソフトウェア開発を行っている企業との緊密な連携が不可欠である。本プロジェクトでは、種々のツールの有用性の現場に即した評価や分析およびプログラム開発を、沖電気工業(株)および算譜工房(有)と共同で行う体制を取る。またプログラム開発環境に関する基本方式に関する研究は東京工業大学の渡部卓雄研究室の協力を得て実施する。

3. 高信頼言語および高信頼プログラム開発環境

SML# は、ML 言語の標準仕様である StandardML を包摂し、最先端の研究成果を採り入れ拡張された次世代 ML 言語であり、現在の ML 系言語の弱点を補いその可用性を大幅に向上させる多相型レコード演算や、ランク1多層性、相互運用機能などの新しい機能を実現する。これらはいずれも、我々独自の基礎研究に基づく最先端のものであり、その理論的な有効性は確認され、認められている。SML# はこれらの機能を初めて装備したプログラミング言語である。

平成16年度は、平成15年度のコンパイラの前半部分(フロントエンド)の開発とバイトコード処理系の開発に引き続き、コンパイラの後半部分(バックエンド)の開発を行い、モジュールシステムを除く SML# 言語のバイトコード処理系のプロトタイプを完成させた。

4. 高信頼プログラム開発環境

SML# が如何にエラーの少ないプログラムの記述を可能にしようとも、既存のライブラリや既存のシステムとの連携における高信頼化技術を確立しなければ、システム全体の高信頼化は達成できない。高信頼プログラム開発環境は、SML# を含む多言語環境下でソフトウェアを開発する上で有用なツール群から構成されるプログラム開発環境である。

本年度は、平成15年度の大規模 web システムの開発事例の分析の結果明らかとなった信頼性低下の要因をターゲットとし、以下の手順で、Java で書かれたプログラム部品の整合性自動検証ツールのプロトタイプの開発を行った。

1. 昨年度に明らかとなった信頼性低下要因であるプログラムの依存性の型理論的分析
 2. プログラム部品間の依存性の整合性検証アルゴリズムの構築
 3. 検証アルゴリズム用いた Struts アプリケーションの不整合自動解析ツールのプロトタイプの開発
- 本プロトタイプは予期した不整合を検出することが確認された。

安全なシステム記述言語および高信頼 OS

東京大学大学院情報理工学系研究科コンピュータ科学専攻 米澤 明憲

協力企業 日立製作所

<http://www.yl.is.s.u-tokyo.ac.jp/e-society/>

1. 高信頼基盤ソフトウェアの必要性

高度情報化に伴い、コンピュータウイルスや不正アクセス、情報漏洩などの問題が、社会の安全性を脅かす深刻な事態となってきた。今後、さらなる情報化は不可避であり、コンピュータを利用した社会基盤の安全性を保証することが早急に必要である。しかし、すでに我々の周りには数多くの応用ソフトウェアが稼動しており、その多くに安全性の問題があると考えられている。それら個々の問題を解決していくことは対症療法にすぎない。問題の根本は、これらの応用ソフトウェアが、安全性の問題が社会的に顕在化する以前に開発された基盤ソフトウェア（言語、OS）を利用している事にある。

本プロジェクトでは、型理論をはじめとするプログラムの静的解析技術によって、既存の基盤ソフトウェアの信頼性を強化することを通して、それらを使用する応用ソフトウェア全体の安全性を向上させることを目標とする。

2. プロジェクトの概要

本プロジェクトでは、今後の実施期間を含めて次の三つの方向性から上記の目標に取り組んでいる。

・高安全な C 言語、および C++ 言語コンパイラの開発

現在、システム記述に最も広く使われている C および C++ 言語での、メモリの不完全な管理を原因とするコンピュータウイルス感染の危険性のないコードを生成する、高メモリ安全な C コンパイラを開発する。また、メモリ安全性に加えて、機密漏洩防止機構を備え、かつ危険な攻撃にさらされても安全に実行を継続できる対攻撃耐性を持ったコード生成コンパイラ VITC の設計にも取り組む。

・OS 用型付きアセンブリ言語の設計・実装

従来高級言語に使われていた型理論を適用した型付きアセンブリ言語をさらに拡張し、OS カーネルのスレッド切替機構、割込処理機構、デバイスドライバ等のアセンブリ言語で書かれた構成要素を再記述することで、OS の安全性を大幅に向上させる。また、この手法は、通常の OS だけにとどまらず様々なハードウェアにおいて、それに密着したソフトウェアの安全性保証に適用することもできる。

・プロトコル実装の形式的検証

サーバなどの基幹ソフトウェアが使用する通信プロトコルの実装が、その仕様に対して正しいかどうかを、定理証明器を用いて検証する技術を構築する。

3. 高安全 C、および C++ 言語コンパイラの開発

(1) メモリ安全 C コンパイラの開発

メモリ安全性を保障し、コンピュータウイルスによるメモリ脆弱性攻撃を検知、緊急停止するコードを生成する C 言語コンパイラの設計。前年度に引き続き、コンパイラの ANSI 規格への準拠性を高める作業を行った。

(2) メモリ安全 C++ コンパイラの開発

C 言語のオブジェクト指向拡張である C++ 言語で書かれたプログラムについても、メモリ安全性を実現するコンパイラの開発に着手、まず、C++ 言語から型情報を付加した拡張言語への変換器を日立製作所と共同開発した。

(3) 対攻撃耐性コード生成コンパイラ (VITC) の設計

上記で保証される緊急停止による安全性は、継続的にサービスを行うべきサーバなどには不十分である。そこで、脆弱性攻撃を受けても緊急停止せず安全に実行を継続できるコードを生成する耐性コードコンパイラの開発を開始、プロトタイプを実装した。またメモリ安全性にとどまらず、情報漏洩防止コードを埋め込むコンパイラのための型理論を構築した。

4. OS 用型付きアセンブリ言語の設計・実装

近年の静的プログラム解析、特に型理論の進歩により、「強く型付けされた言語」（例：Java バイトコード、OCaml、等）が広く普及している。これらの言語の長所の一つは、プログラムが実行時にエラーを生じないことを型検査によって保証できることである。

しかし、OS の作成には、未だ型理論の成果が応用されていない。そこで本研究は、OS カーネルの記述に適した強く型付けされた言語「型付きアセンブリ言語 (TAL)」を設計・実装し、OS カーネルをその TAL で記述することで、実行時にエラーを生じない安全で高信頼な OS を実現することを目指す。現在までのところ、OS カーネルの機能の内、複数スレッド切替の機能を TAL で記述することが可能なことを確認した。

5. プロトコル実装の形式的検証

サーバなどの基盤ソフトウェアの信頼性の向上には、通信に使用するプロトコルのサーバ内実装の正しさの検証が不可欠である。サーバは本質的に並列、非決定的に動作するため、従来のプログラム検証が対象としてきた逐次プログラムと比べ、システムの取り得る状態数が爆発してしまい、その検証は容易ではない。我々は、モデル記述言語として拡張された π 計算を使い、partial order reduction という探索空間削減技術を開発しこの問題を解決した。また、この技術を使用し、メール配信のプロトコルである SMTP の実装（受信部分、700 行の Java プログラム）が満たすべき種々の安全性を、定理証明器 Coq を使って形式的に検証することに成功した。

データ収集に基づくソフトウェア開発支援システム

奈良先端科学技術大学院大学 鳥居宏次

研究分担者 大阪大学 井上克郎

協力企業 日立公共システムエンジニアリング, 日立製作所, NTT ソフトウェア, SRA 先端技術研究所

<http://www.empirical.jp/>

1. はじめに

本プロジェクトでは、ソフトウェア開発の分野において、他の科学、工学分野と同様に、計測、定量化と評価、そしてフィードバックによる改善という実証的手法（エンピリカルアプローチ）の実践を目指す。以下では、本年度の主な成果の概要を述べる。

2. EPM α 版の企業プロジェクトへの適用

平成 15 年度に構築したデータ収集・分析システム EPM(Empirical Project Monitor)0.91 α 版の企業への提供を行った。EPM の導入が円滑に進むよう、EPM 研修会を開催し、EPM 概要説明、EPM デモ、環境設定、インストール実習、EPM 詳細説明、ソフトウェア工学概説、などを 3 日間で実施した。

EPM α 版の提供先は、2004 年 3 月時点で、導入準備中を含め企業 7 社、2 大学となっている。このうちの 1 社である日立公共システムエンジニアリング（日立 GP）では、2004 年 4 月から 9 月までの 6 ヶ月間、ある自社パッケージソフトウェアの開発プロジェクトにおいて EPM α 版によるデータ収集を行った。EPM の導入とデータ収集に要した工数は、EPM 研修会の受講を含めてもわずか 25 人日であり、（ソフトウェア開発の）社内標準プロセスへのデータ収集作業の組み込みもスムーズに実施できた。現在は、協調フィルタリングや潜在的意味解析法などのデータ分析法を収集データに適用し、その結果の評価や解釈を、日立 GP のプロジェクト管理者や開発者と共に実施している。

3. EPM α 版のバージョンアップ

より多くの企業ソフトウェア開発プロジェクトへの適用を実現するため、機能追加を容易にするプラグイン化、バグ収束曲線グラフの作成機能の作成、利用マニュアル等の整備などを行い、EPM0.91 α 版のバージョンアップ版である EPM0.92 α の開発を行った。あわせて、EPM のインストールキットの整備等も行った。

0.91 α 版では EPM が一体となっていたため、EPM を熟知していないと機能拡張が出来なかったが、機能追加が見込まれるアナライザ部分については、利用者が機能追加できるようプラグイン化を図り、利用者の利便性を向上させている。また、ソフトウェア信頼度成長モデル (SRGM) を用いたバグ収束曲線グラフ作成機能を作成し、ソフトウェアの信頼性並びに品質の

達成度合いや、ソフトウェア製品の出荷品質等の判断材料として利用できるようにしている。更に、利用マニュアルの整備も並行して進め、EPM0.91 α 、EPM0.92 α の利用マニュアルをプロジェクトホームページに一般公開している。

現時点で公開している EPM ファイルは次の 3 種類となる。

- ・専用サーバ (Linux) の構築を目的とした「バイナリファイル」
- ・評価のための利用を目的とした「1CD ブート Linux のイメージファイル」
- ・Windows マシン上で動作させることを目的とした「仮想化ソフト coLinux のブートイメージファイル」

4. 解析ツールの構築

ソフトウェア開発プロジェクトを潜在的意味解析法 LSA(Latent Semantic Analysis) を用いて分類するツールを開発した。LSA とは、自然言語で書かれた文書、単語の類似度を測定する方法で、ベクトル空間モデルに従った手法の一つである。ただし、ベクトル空間モデルでは検出できない間接的な関連の抽出を可能にしている。ソースコードを文書に、ソースコード上の識別子（変数名、関数名、型名）を単語に、それぞれ対応させることで、ソースコード間の類似度評価、分類（クラスタリング）が可能になる

SourceForge.net で公開されている 41 個のソフトウェアを分類するケーススタディでは、開発環境やプログラミング言語に関する深い知識なしに、複数の観点からの分類が可能であることが確認された。ケーススタディでは、ソースコードを類似度測定の対象としているが、LSA はもともと自然言語で書かれた文書を対象としたものである。要求仕様書や設計ドキュメント、プロジェクト計画書やテスト実施計画書などへの適用も考えられる。ソフトウェア産業全体、組織単位、プロジェクト単位、プロダクト単位など、多様な粒度でのベースライン開発を可能とするものである。

5. おわりに

本年度は、EPM を媒体として、ソフトウェアの生産性と信頼性の向上に向けた定量的アプローチをいくつかの企業と連携して実施することができた。また、協調フィルタリングや潜在的意味解析法といったデータ分析手法を提案することで、データ収集に基づく開発支援のより具体的なイメージを産業界に提示した。来年度は、独立行政法人情報処理推進機構ソフトウェア・エンジニアリング・センターとの連携などにより、EPM の適用先を更に拡大する予定である。

高信頼構造化文書変換技術

東京大学 情報理工学系研究科 武市正人
<http://www.psdlab.org/>

1. 高信頼構造化文書変換技術の概要

電子的な構造化文書情報の蓄積と効果的な情報利用技術は、インターネットを含む広範な情報の交換・流通にとってきわめて重要な位置を占めている。XML に代表されるこれらの技術は、発展の著しいウェブによる情報環境において既存の技術の延長線上で実務的に開発されたものであり、事実上の標準となつてはいるが、その言語的な概念が十分に整理されているわけではない。このような体系的な処理技術の欠如に起因する問題が情報交換の発展を阻害している。すなわち、既存技術の使い回しによる姑息な対処や人手による個別対応が、一般性に欠ける文書情報を蓄積する一因となっているのであり、この問題を早急に解決することが重要な課題となっている。

構造化文書はプログラミング言語のデータ構造と共通の性質を有しており、関数型言語によるアルゴリズム記述が文書処理に適している。このようなプログラムにより文書処理の信頼性を高め、情報の流通を高度化することが可能となる。本研究の中核となる Programmable Structured Document (PSD) は、プログラムの記述を含む文書を対象として、構造化文書処理を効果的に実現しようとするものである。すなわち、PSD においては構造化文書をプログラミングにおける構造化データであるとなし、プログラミング言語の理論的基盤を適用することによって、安全かつ信頼性の高い処理を実現する。また、処理を実現するプログラムは対象文書に埋め込まれており、これによって文書の高い独立性と可搬性を実現するものである。

2. PSD の基盤技術

PSD 実現のために必要となる基盤技術としては、(i) 構造化文書に必要な準構造化データの概念を型として捉える形式的枠組の定義およびデータ型に基づく効率的変換手法、(ii) PSD のための計算機構を組み込んだ構造化文書の実現手法、(iii) 関数型言語におけるデータ型の理論の発展と準構造化データに適した型理論の構築、の3つから構成される。(iii) は、さらに代数的プログラム変換(演算)の成果を構造化文書に適用し、自己参照による変換戦略を文書自体に付随させるという演算随伴機構に関する理論を含む。

以上の基盤技術に立脚した PSD を操作するための PSD 処理システムは、文書処理ソフトウェア等による構造化文書の編集から発信にいたる処理を可能とし、構造化文書を変換するためのプログラムの自動生成を実現するものである。このような優れた特徴を持つ PSD 処理システムは、ウェブを含むオフィス環境における基本ソフトウェアとしても位置付けることができる。

3. PSD の成果

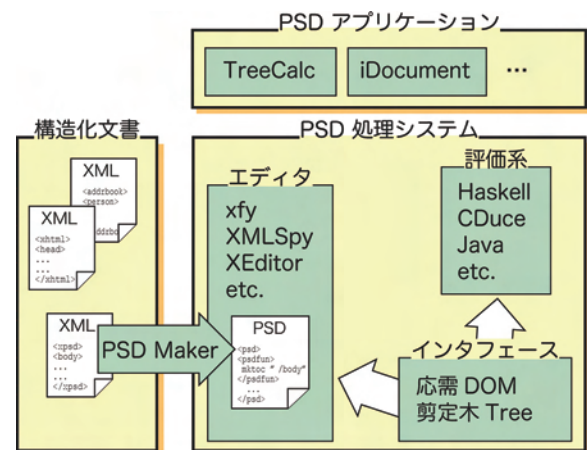
本プロジェクトでは、構造化文書に対する PSD 構造化手法と変換(演算)規則を体系化し、ソフトウェアの信頼性確保の基礎となる参照用文書(言語仕様・設計仕様)を作成する。

その成果として、一般の文書作成者が利用できる高信頼 PSD 汎用ソフトウェアパッケージ(統合的構造化文書処理システム)と PSD 構造化文書を対象とした各種アプリケーション開発用のプログラミングシステム(構造化文書変換プログラム生成システム)が挙げられる。前者は構造化文書を効率よく作成し、それらを対象としたアプリケーションを半自動的に生成することを目的としている。また、後者は主としてアプリケーションソフトウェアの開発者向けのものである。

文書に価値を付与するためには作成時から文書を構造化しておく必要があるが、統合的構造化文書処理システムでは、これを単独または従来の文書処理ソフトウェア等と連携して、付加価値の高い文書やコンテンツの作成に利用することが可能であり、情報発信のツールとしての個人、オフィスにおける利用が期待できる。特に、オフィス系ソフトウェアにおける XML 関連ソフトウェアの比率は大きくなってきており、構造化文書変換プログラム生成システムの利用は、このようなプログラムの開発の効率を高め、生産性を向上させることになる。

4. おわりに

本年度は、実用に耐えうる PSD 実行のためのインタフェースおよび PSD アプリケーションを開発し、さらに PSD の正当性や安全性を保証する理論についての研究を行った。



高信頼 WebWare 生成技術

名古屋大学 情報科学研究科 阿草 清滋

研究分担者 和歌山大学 デザイン情報学科 鯨坂 恒夫

研究分担者 愛知県立大学 情報科学部 山本 晋一郎

協力企業 富士通研究所, 富士通プライムソフトテクノロジー, 野村総合研究所

<http://www.agusa.i.is.nagoya-u.ac.jp/research/webware/index.xhtml>

1. はじめに

本研究の目標は、WebWare の信頼性と安全性を保証しつつ、エンジニアが行なう WebWare 作成作業と、デザイナーとエンジニアの協調作業の両方を支援することが可能な、統合的 WebWare 開発環境を構築することである。

今年度は、WebWare のテスト支援、解析技術、作成作業支援において以下の成果を得た。

2. WebWare のテスト支援

機能分割された Web アプリケーションの各要素の振舞いを抽象的にモデル化し、モデル検査を用いた検証を可能とする手法を提案し、手法を支援するシステムを構築した。また、JSP に対して、Web アプリケーションの仕様からテストケースを生成、テストを自動実行するツールを開発した。

モデル検査可能な開発支援システムにより、WebWare 開発において設計段階での誤りを発見可能な開発環境を実現することができた。また、テストの生成、自動実行可能なツールによって、テスト段階におけるヒューマンエラーを防ぐことによるテストの信頼性向上と自動実行によるコストの削減が実現できる。

3. WebWare の解析技術

昨年度の成果である、JavaScript 解析技術とページ遷移モデルを利用して、JavaScript に対応した Web サイト解析とページ遷移モデルに基づく Web アプリケーション解析ツールを開発した (図 1)。

JavaScript は実際の Web サイトで頻りに利用されるため、現実的な Web サイト解析には JavaScript 対応が必須である。昨年度の成果と既存ツールを融合させることで JavaScript に対応した Web サイト解析を可能とした。WebWare 解析ツールでは、ソースコードに基づき WebWare の全体構成を忠実に、

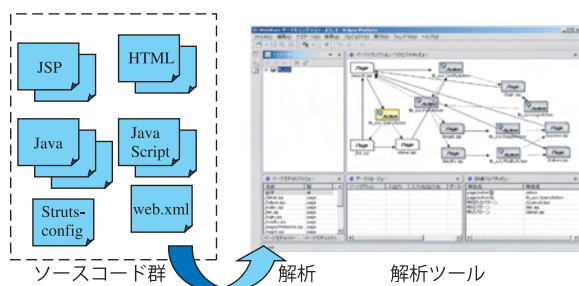


図 1 Web アプリケーション解析ツール

グラフィカルに、再現することを可能とした。また、デッドリンクやデッドページなど構成上の誤りと思われる要因の検出もツール機能として実現した。

4. エンドユーザ情報に着目した WebWare

作成支援

エンドユーザとの界面を構成する技術を対象に、高度なモジュール化による作成支援を目指して、クライアントサイドプログラムのより高度な解析、リッチクライアントを用いた WebWare のアーキテクチャの提案、表現情報主導のスタイル変換記述生成ツールの開発を行った。

クライアントサイドプログラムの高度な解析では、JavaScript の特徴の一つである、クライアントサイドでページを表示中に、プログラムの実行によって動的にページの一部を生成する機能を対象に、このようなページの解析機能を実装した。また、変数参照と制御構文への対応を行った。

リッチクライアントを用いた WebWare のアーキテクチャの提案では、リッチクライアントの実装としてよく用いられる Flash とサーバサイド技術との連携を行う標準的なフレームワークを提案した。このフレームワークは、MVC フレームワークを拡張し、View 要素をクライアントサイドに移動させ、Controller がクライアントサイドとサーバサイドに分割されたものとなっている。サーバサイドでは、Model の状態に合わせて動的に XML 文書が生成され、クライアントサイドの Controller がこれを参照する。

表現情報主導のスタイル変換記述生成ツールの開発では、スタイル変換技術に基づく WebWare フレームワークにおいて、アプリケーション実装の中心となるスタイル変換記述を効率よく行う手法を確立するために、スタイル変換記述の自動生成ツールを試作した。このツールは、変換対象の XML 文書とスキーマ記述を与えることで、エンドユーザ向けの表示文書を作成するためのスタイル変換記述を自動生成する。

5. おわりに

今年度は、(1) 設計、テスト段階における WebWare 開発の信頼性向上、(2) WebWare 解析技術を応用した CASE ツールの開発、(3) エンドユーザ情報に着目した WebWare 作成支援について成果を得た。

インターネット上の知識集約を可能にするプラットフォーム構築技術

早稲田大学 理工学部 村岡洋一

<http://www.yama.info.waseda.ac.jp/~yamana/e-society/>

1. 目的

インターネット上の WWW サーバから発信される情報量は膨大であり、2005 年 3 月時点で、テキストデータだけでも 119 億ページと推測される。現在の全世界の人口が約 65 億人であることから、一人当たり換算すると平均 1.8 ページのデータを発信していることになる。このような膨大な Web 上には、人間が一生かかっても学ぶことのできない情報、知識、ノウハウが凝縮されていると言っても過言ではない。

これまで、これらの情報の検索には、いわゆる検索エンジンが用いられてきたが、検索の結果得られる情報は必ずしも最新の情報ではない。さらにユーザは、特定のキーワードに対応した検索結果のみを利用するだけである。このように、検索エンジンは、Web の限られた利用法を提供しているに過ぎない。

これに対して本プロジェクトでは、全世界の Web ページを網羅的に、かつ、最新の情報を収集・解析することにより、従来実現不可能とされていた新しいビジネス—各種のトレンド解析、ある商品や企業の評判解析、誹謗中傷監視など—を創造することを目指している。例えば、現在、テレビの視聴率はサンプリングにより算出を行っているが、テレビで放映された番組に関する情報がどの程度 Web 上で話題になっているかを視聴率の代わりに使用することもできるし、最近問題となっている振り込め詐欺に関する事例を Web 上から自動的に抽出し、これらの詐欺に遭わないように広く周知することも可能となる。

2. 研究課題

本プロジェクトでは、商用・研究用を通じて世界最大となる 119 億の Web ページを対象に、(1) 平均して 1 ヶ月以内の新しいデータに更新することを可能とする WWW クローラーを開発すると共に、(2) 利用者の検索目的に応じて必要となる情報を抽出する知識フィルタリング技術の開発を目指している。

2 年目である 2004 年度は、国内 5 箇所にて WWW クローラーを設置し、50 億 URL の Web ページ収集を行った（研究用途として世界最大）。また、10 億規模でのデータ解析を開始した。2005 年度末までには、119 億の Web ページの収集を完了させると共に、50 億 URL 規模での解析（2004 年度までに収集した Web データを対象）を実現させる予定である。

以下に 2004 年度の研究課題と成果概要を示す。

2.1. 研究用途として世界最大の 50 億 URL の Web データ収集

2003 年度の 10 億 URL 収集に続き、2004 年度は収集拠点数を国内 3 箇所から国内 5 箇所（早稲田大学内 2 箇所、インター

ネットデータセンター、NTT 未来ねっと研究所、国立情報学研究所）に増やし 50 億 URL の Web ページ収集を行った。収集においては、複数の Web クローラーを連携するインテグレーションサーバを導入し分散収集を実現させた。さらに、同一ページの重複収集回避、相手 Web サーバへの負荷軽減を実現する技術を開発した。また、収集した Web ページの更新を高速に行うため、2003 年度に提案した「Web ページの新鮮度を 1 ヶ月以内に仕上げる仕組み」を用いて実験を行った。その結果、26% の Web ページを収集することにより、更新された Web ページの 75% を収集可能であることを確認した。なお、新鮮度とは、Web ページの内容が、収集済の Web ページの内容から更新された時点からの経過日数である。

2.2. 10 億のデータを対象とした Web マイニング

2003 年度に収集済みの 10 億 URL の Web データを対象に、2003 年度に提案した手法を用い、製品を対象とした競合製品の自動抽出及び評判情報の抽出を行った（図を参照）。その結果、Web データのコンテンツを分析することなく、ページ間のリンク関係のみから、これらの情報が抽出可能であることを確認した。本手法は、Web データのコンテンツを分析する必要がないことから、100 億レベルの超大規模データにもアルゴリズム的に適用可能であることを確認した。さらに、超大規模データを対象としたマイニングを高速に実現するための手法としてマイニングの並列化手法、及び、TF2P-growth と呼ぶ新しいデータマイニング手法を開発した。

3. 実施体制

本プロジェクトは、プロジェクトリーダー村岡洋一（早大・理工）、サブリーダー山名早人（早大・理工）のもと、富士通株式会社、アクセラテクノロジー株式会社、NTT 未来ねっと研究所、国立情報学研究所と共同で実施し、技術移転を進めている。



先進的なストレージ技術

東京大学 生産技術研究所 戦略情報融合国際研究センター 喜連川 優
協力企業 日立製作所, NTT プラットフォーム研究所
<http://www.tkl.iis.u-tokyo.ac.jp/project/e-society/>

1. はじめに

情報通信技術の革新が進む中で、IT システムの一層の高信頼化、高性能化、管理容易化が求められている。とりわけ9.11テロ以降、災害時にもデータを失うことなく業務を継続させることができるディザスタリカバリ (DR) 技術が注目を集めている。また、爆発的に増加し続けるデータの高速な検索と低コストでの管理を実現するためには、ストレージの高性能化・管理容易化が必須である。本研究では、ストレージをサーバ上のアプリケーションと融合させる次世代ストレージ技術 (Storage Fusion) の確立を目指す。戦略的競争力を実現すべく、上記の課題の解決に焦点を絞って研究を進めている。平成 16 年度は、安価なシステムコストで広域災害時のデータ保護を実現する高度ディザスタリカバリ機構、各種知識を利用し大幅に I/O 性能を向上させるストレージ超高速アクセス機構、ストレージと DBMS の挙動を可視化する管理コスト低減機構を開発した。

2. 高度ディザスタリカバリ機構

データ保護とオンライン性能確保を両立する DR システムを安価なシステムコストで実現することを目的に、昨年度に検討したハイブリッド転送方式を製品化すると共に、ログのみを同期リモートコピーで転送し DB を副サイトでリアルタイム適用する「ログのみ同期転送方式」を開発・評価した。DB の転送を不要とすることにより、回線コストを削減すると共に、ログ適用処理を簡易サーバで稼働させることにより、平常時の副 DB サーバの稼働を不要とし、機器・管理コストの削減も可能とした。正サイトのオンライン性能について、回線シミュレータを用いたシミュレーション環境で評価を実施した結果、100Mbps 以下の帯域幅の回線で広域災害を考慮した DR システムを実現可能なことが確かめられた (図 1)。

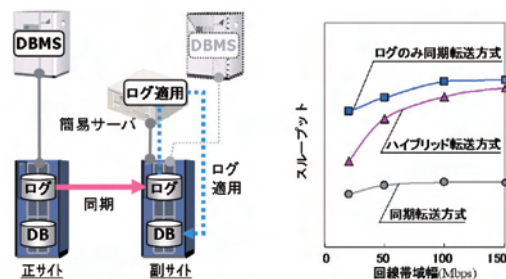


図 1 システム構成とオンライン性能シミュレーション結果 (サイト間距離：東京～大阪間)

3. ストレージ超高速アクセス機構

ストレージ超高速アクセス機構では、昨年度は実験機を用いていたのに対し本年度は商用ストレージを用いたクエリブ

ラン利用先読み技術のプロトタイプを開発した (図 2)。更に 2 次先読み方式を新たに開発し、本プロトタイプに実装した。TPC-H Q8 ベンチマークを用いて評価し、図 3 に示すように大幅に処理時間が短縮されることを確認した。



図 2 クエリプラン利用先読み技術プロトタイプの構成概略



図 3 プロトタイプの評価結果 (TPC-H Q8 Benchmark)

4. ストレージ管理コスト低減機構

これまで個別に運用管理されてきたストレージと DBMS を統合的に管理するために、ストレージと DBMS の性能情報を統合的に表示するシステムモニタリング技術を開発した。本技術により、ストレージのリソースを DBMS がどのように使用しているか、及び SQL 文が DB サーバ及びストレージのリソースをどのように使用しているかの内訳を可視化することが可能となった (図 4)。開発した技術を用いてストレージ競合による性能劣化問題解決の時間を計測、既存機能と比較して作業時間を最大 28% 削減できることを確認した。また、本技術の一部を製品に反映した。

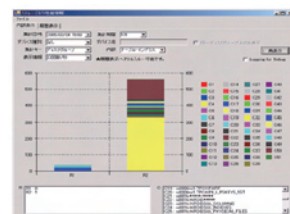


図 4 ディスクグループに対する I/O の内訳

5. まとめ

ストレージとデータベースアプリケーションを融合することにより高信頼かつ高性能なデータ格納プラットフォームを実現する技術を開発した。高度 DR 機構においては、データ保護とオンライン性能確保を両立する DR システムを安価で実現可能な「ログのみ同期転送技術」を開発した。次年度以降は、副サイトのログ適用処理を専用装置で実現可能とするための DR 方式の検討を進める。ストレージ超高速アクセス機構においては、商用ストレージを用いてクエリプラン利用先読み技術のプロトタイプを開発・評価した。来年度以降、実社会のシステムに近い負荷を用いてより詳細な評価を実施する計画である。ストレージ管理コスト低減機構においては、ストレージと DBMS の挙動を統合的に可視化するシステムモニタリング技術を開発した。今後は本技術を用いたデータ配置最適化の評価、ボルトネック検出技術の検討を進める。

先進的な Web 解析技術

東京大学 生産技術研究所 戦略情報融合国際センター 喜連川 優
 協力企業 三菱電機
<http://www.tkl.iis.u-tokyo.ac.jp/>

1. はじめに

Web 上では企業や省庁、個人による情報発信が刻々と行われており、それらはハイパーリンクの複雑な網の目を形成し、互いに強く影響を与え合っている。Web の大域的な構造を把握し、その変化を追跡することは、実社会に起こる事象の背景や予兆を探る上で極めて有効と考えられるが、現状の検索エンジンによる近視眼的な検索方法ではその目的を達することはできない。

本プロジェクトでは、日本国内 Web 情報の過去から現在に到る履歴を蓄積する Web アーカイブを構築し、リンク解析およびテキスト解析（自然言語処理）を核とした分析技術により、Web の空間構造および時系列変化の分析を可能にする新たなツールの実現を目的とする。

2. Web アーカイブ基盤の構築

Web アーカイブ基盤は、日本国内 Web のスナップショットを随時大容量ストレージに蓄積し縦横なアクセスを可能としたものであり、Web 時空間解析のプラットフォームを提供する。平成 15 年度はアプリケーションインタフェースの検討に重点を置き、基本機能を持つプロトタイプの開発を行った。

平成 16 年度は複数スナップショットの統合管理および横断的アクセスに向けた課題に取り組み、キーワードと期間を指定して複数スナップショットの検索を行える時系列全文検索エンジンのプロトタイプを開発した。また、スナップショット間やスナップショット内の重複コンテンツの検出機構を実現した。さらに、Web ページ間のリンク関係を効率的に格納するリンクデータベースにおいても複数スナップショットの横断的分析に対応し、Web ページに関するポピュラリティの推移、ページ間の関連の時系列変化などを解析する基盤を実現した。

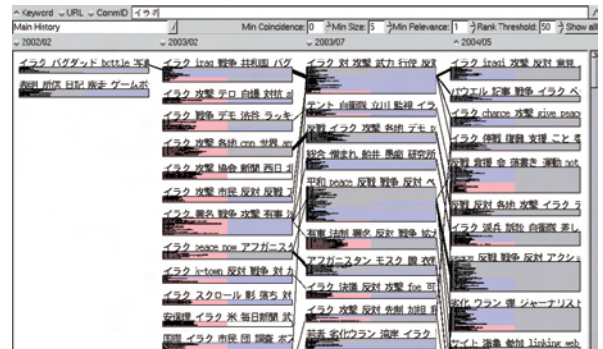
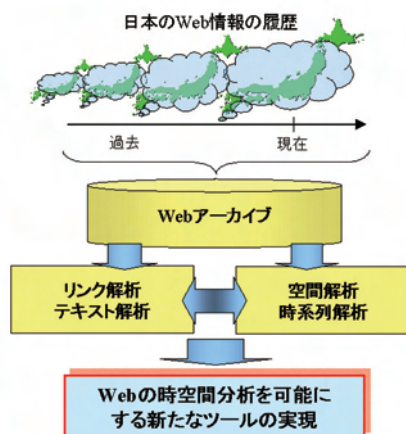


図 イラク戦争による Web の変化

3. Web 時空間分析

互いに関連する Web ページは直接あるいはリンク集などを介して間接的にリンクされる傾向にあり、この特徴を利用して Web 空間から稠密なリンク構造を抽出すると Web コミュニティと呼ばれる関連ページの集合が得られる。平成 15 年度には Web のスナップショットからほぼ全てのコミュニティを網羅的に抽出し、それらを互いの関連度に従って配置した Web 空間の俯瞰図（コミュニティチャート）を作成した。

本年度はチャートの時系列的な発展過程を追跡する手法を開発し、視覚的に変化を確認できる発展過程ビューアのプロトタイプを構築した（上図）。本システムは、Web におけるトピックの推移、新たな情報の発生、Web 上の社会的な現象などの調査に利用することができる。

また、利用者の分析意図に応じたコミュニティの抽出を目指して、リンク解析と自然言語解析を併用した解析方式の検討および評価実験を行い、コミュニティの抽出精度の向上を確認した。

4. まとめ

日本国内 Web 情報の履歴を蓄積した Web アーカイブ基盤上で Web 時空間分析の実現に取り組んだ。日本国内 Web 全体の空間構造を抽出し、その時系列変化を可視化することで Web 空間の発展や変遷を調べることが可能になった。また、リンク解析技術と自然言語解析技術を併用した新しい手法により、空間構造抽出の高品質化を進めた。

今後は Web アーカイブ基盤へのスナップショット追加を高頻度化すると共に、実応用分野における Web 時空間分析の有効性を明らかにすべく実証システムを構築して行く予定である。

ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術

奈良先端科学技術大学院大学 情報科学研究科 鹿野清宏

研究分担者 京都大学 河原達也, 名古屋大学 武田一哉, 和歌山大学 河原英紀, 奈良先端大 猿渡洋,
名古屋工業大学 徳田恵一, 立命館大学 西浦敬信

協力企業 松下電器, 旭化成, 日立製作所, 松下電工, ASTEM, オムロン

<http://cif.iis.u-tokyo.ac.jp/e-society/database/index.html>

1. プロジェクトの目標

携帯電話, 携帯端末 (PDA), PC の入出力, カーナビ, 家電制御, 秘書ロボットなどを, 誰でも容易に利用できることが望まれる。これらを実現するには, 人と機械との自然な対話として, 音声認識・合成技術が有望である。本格的に利用されるためには, 頑健かつ高精度の音声認識基盤ソフトウェアを開発して, かつ廉価に利用できるようにすることが不可欠である。この音声認識・合成技術の基盤ソフトウェアの普及により, 誰でもが気軽に, 情報機器の利用ができるようになり, デジタルデバイドの社会問題の軽減, IT 市場の活性化につながる。

このプロジェクトでは, 大語彙連続音声認識プログラム, 話者環境適応プログラム, ハンズフリー音声認識プログラム, ハンズフリー音声収録チップ, 大語彙連続音声認識プログラムのマイコン上への実装を行う。また, 音声合成では, 多様な音声合成プログラムを作成する。これらのプログラムは, 単に開発するだけでなく, 実環境での応用システムでの実証試験による評価も行って改善をはかる。さらに, 音声認識技術の利用法のノウハウも蓄える。

2. プロジェクトの概要

音声認識を本格的な商用化につなげるには, 頑健な音声認識システムとして, 以下の技術を研究開発することが必要である。
(i) 利用環境およびユーザに対する負担をかけない適応技術, (ii) マイクを意識しない自然なハンズフリー音声認識技術, (iii) 高精度連続音声認識プログラムおよび音声認識システム研究開発ワークベンチが必要となる。これらのソフトウェアを研究開発し, 廉価に誰もが利用できるプログラム, あるいは DSP/マイコンとして提供する。さらに, 開発したソフトウェアを用いて, (iv) 人と機械の音声対話の実証実験を行い, プログラム, システムの評価を行うとともに, 利用に関するノウハウを蓄積する。音声合成では, (v) 多様な声質を実現できることが重要であり, 高精度音声分析合成系 STRAIGHT を用いた声質変換プログラムを開発する。

本プロジェクトの研究開発の概要を図 1 に示す。さらに, 研究開発項目ごとの中間目標, 最終目標を表 1 にまとめておく。

3. H16 年度の進捗

研究項目ごとに H16 年度の進捗をまとめておく。

- ・教師なし話者適応プログラム: 実環境音声対話システムによ

るデータ収録と, 高精度子供音韻モデルの構築を行った。教師なし話者適応プログラムの高速化を達成した。つぶやき声 (NAM) の音声認識と電話での伝送の可能性を実証した。

- ・大語彙連続音声認識ソフトウェア: 音声対話システムに対応する音韻モデル, 言語モデルを構築した。音声認識エンジン Julius の機能・性能強化 (信頼度, 雑音モデル) を行った。Julius の簡易版のマイコンへの実装を行った。
- ・マイクロフォンアレイによるハンズフリー音声認識: 空間サブトラクションアレイ (SSA) の改良と評価を行った。ブラインド音源分離の研究を進め, PC 実装を行った。
- ・音声認識のフィールドテスト: インターネット楽曲検索システムによりユーザの学習効果を調べた。各種対話システムで音声データの収録を継続している。
- ・多様な音声合成プログラム: 感情付音声データベースを拡充し, これに基づき, 感情音声変換アルゴリズムの簡易版を開発した。さらに, ピッチ抽出アルゴリズムを改善して, 音声分析合成システム STRAIGHT の性能向上を達成した。

表 1 研究開発の目標の概要

研究開発項目	中間目標	最終目標
ユーザ負担の少ない話者・環境適応	教師なしオンライン話者適応	ハンズフリー環境・話者適応プログラム
大語彙連続音声認識ソフトウェア	丁寧な話し言葉認識, マイコンに実装	話し言葉認識ソフトの完成と普及
ハンズフリー音声認識	ハンズフリー音声収録プログラム	ハンズフリー音声収録 DSP
実環境の音声対話システムの構築	各種実環境音声対話システムの構築	実環境音声対話システムの運用と評価
多様な声質の音声合成ソフトウェア	多様な音声変換プログラム	多様な音声変換プログラムの普及

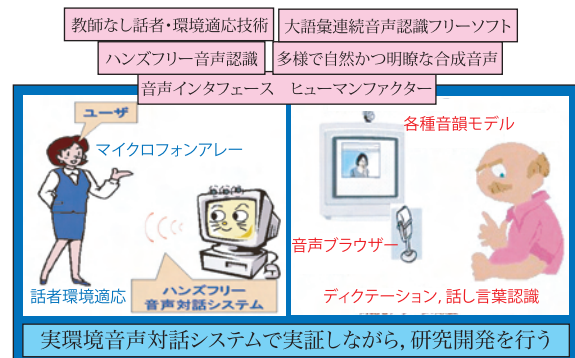


図 1 一人にやさしい自然な音声インタフェース

教師なし話者・環境適応技術

奈良先端科学技術大学院大学
情報科学研究科 鹿野清宏
<http://isw3.naist.jp/IS/Shikano-lab/home.html>

1. ねらい

人と機械との自然な対話を実現するためには、高精度な音声認識技術が必須である。とくに、利用環境とユーザへの適応技術が必要となり、かつ、ユーザに負担をかけない適応技術が望まれる。このような環境と話者への適応を実現のために、教師なし話者・環境適応技術に関して開発を行う。

2. 研究の概要

雑音環境下で、任意の1文発声に基づくユーザに負担をかけない教師なし話者適応アルゴリズムの研究開発を進める。具体的には、雑音に頑健な音声認識アルゴリズムとHMM（隠れマルコフモデル）十分統計量に基づく教師なし話者適応アルゴリズムの研究開発を行う。最終的に、オンラインで動作する話者・環境適応プログラムを完成させる。さらに、新しい静かな音声メディアとして、つぶやき声の認識（無音声認識）、電話による伝送（無音声電話）の研究を行う。

3. H16年度の進捗

平成16年度における研究進捗を以下に示す。

- (i) 音声情報案内システム「たけまるくん」(図1)により、20万発話以上の音声データを収集して書き起こしを行った。特に、従来データが少なかった子供のデータが収集できた。子供の音声データを用いて、高精度子供音韻モデルの構築と、話者適応の効果を確認した。
- (ii) より広範な話者層として、高齢者と成人音声、合計600人を用いて、話者層、性別ごとの音韻モデルを用いて話者適応アルゴリズムを改良した。
- (iii) 教師なし話者適応プログラムの高速化プログラムを作成した。数秒で動作するオンライン話者適応プログラムパッケージの作成が可能であることが確認できた。
- (iv) つぶやき声 (NAM: Non-Audible Murmur) は、話し手の近くでも聞こえない声である。このNAMを音声認識（無音声認識）したり、電話で送る（無音声電話）ことができることの可能性を実証した。



図1 音声情報案内システム「たけまるくん」

ハンズフリー音声認識

奈良先端科学技術大学院大学
情報科学研究科 猿渡洋
<http://isw3.naist.jp/IS/Shikano-lab/home.html>

1. ねらい

ユーザに負担をかけない自然な音声入力系として、ハンズフリー音声認識システムを構築する。特に、マイクロフォンアレー(図2参照)による音声収録技術に着目し、音声認識性能の向上を目指す。また、コンパクトかつ廉価なマイクロフォンアレーアルゴリズムの開発を行う。

2. 研究の概要

ユーザからの距離1m以下で高性能に動作するハンズフリー音声認識システムを、8チャンネル以下のマイクロフォンアレーを用いて開発する。認識性能は、1m離れた音声入力で、従来の接話マイクとほぼ同等の認識性能を目指す。さらに、マイクロフォンアレーのコストを20分の1以下にするため、ハンズフリー音声収録用DSPを開発する。

3. H16年度の進捗

平成16年度における研究進捗を以下に示す。

- (i) 従来の遅延和型・Griffith-Jim 適応型マイクロフォンアレー信号処理アルゴリズムと、昨年度に提案した空間スペクトル演算アレー SSA (Spatial Subtraction Array) の音声認識性能評価・比較を行い、SSAの優位性を様々な雑音環境において確認した。また、素子誤差補正、帯域別素子間隔選択法、既知雑音重畳音響モデルなどに関して検討を行い、性能改善を達成した。更にユーザ方位のずれに対して頑健にするため、実時間方位推定法の検討も行った。
- (ii) DSPによるSSAの実時間処理化に先立ち、通常のPCをベースにして実装検討を行った。想定される規模(4~8素子アレー)において、実時間処理が可能であることを確認できた。
- (iii) 音源間の独立性に基づいて音声を抽出するブラインド音源分離に関して検討を行い、音質劣化が少なく安定なフィルタ学習アルゴリズムの提案・PC実装を行った。



図2 直線状マイクロフォンアレー

大語彙連続音声認識プログラム

京都大学 学術情報メディアセンター 河原達也
<http://www.ar.media.kyoto-u.ac.jp/>

1. ねらい

ユーザに負担のない自然な音声対話を実現するには、音声認識システムが、できるだけ広範な話者層や言い回し、特に話し言葉に対応できる必要がある。このような大語彙連続音声認識を行うオープンソースのプログラムを開発する。また、組み込み機器にも利用できるように、マイコンへの実装も行う。

2. 研究の概要

これまでのディクテーションシステムは主に成人の読上げ音声を対象としていたが、音声対話を指向して発展させる。音韻モデルについては、多様な話者および対話調の音声に対応できるようにする。言語・発音モデルについても、話し言葉に対応できるようにする。さらに、音声認識エンジン Julius についても音声対話システム向けの機能強化を行うとともに、マイコンに移植できるような効率化を図る。

3. H16 年度の進捗

平成 16 年度における研究進捗を以下に示す。

- (i) 音声対話システムに自然に発話される音声に対応するための不特定話者音韻モデルを作成し、評価・比較を行った。
- (ii) 多様なタスク・ドメインに適用できる会話調の発音・言語モデルを構築した。
- (iii) 音声認識エンジン Julius の機能・性能強化、及び Web ブラウザ連携対応 (SALT のサポート) を行った。
- (iv) 音声認識ソフトウェアを Web ベースで容易にカスタマイズできるツールを作成した。
- (v) 音声認識プログラムの SH-4 マイコンへの実装を進め、簡易版の動作を確認した。
- (vi) ソフトウェア普及のための講習会を行った。

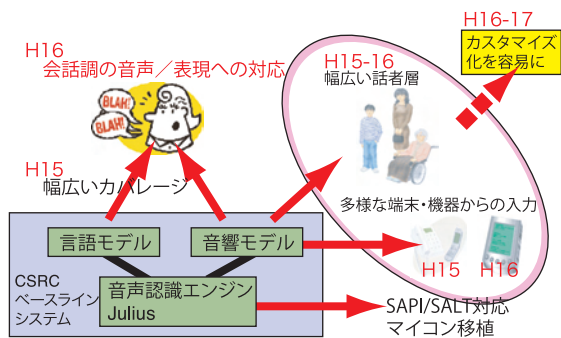


図 3 大語彙連続音声認識プログラムの開発概要

音声認識システムのフィールドテスト

名古屋大学 情報科学研究科 武田一哉
<http://www.sp.m.is.nagoya-u.ac.jp/>

1. ねらい

ユーザにとっての音声対話システムの性能は、認識性能だけでなく、様々なヒューマンファクタに支配されている。開発したプログラムやモデルを実環境下で運用することで、開発成果の検証を行うとともに、運用結果からユーザ負担の少ない音声対話システムの設計指針を得る。

2. 研究の概要

開発した音声認識プログラム、各種モデルを用いて、実環境下での音声対話システムのフィールドテストを行い、システムの使い方の教示方法や学習効果など、音声認識利用のヒューマンファクタに関するノウハウを蓄積する。さらに得られたノウハウを整理して、音声対話システム構築のためのガイドラインを作成する。

3. H16 年度の進捗

平成 16 年度における研究進捗を以下に示す。

- (i) 平成 15 年度までに行ってきたフィールドテスト (バス運行情報案、公共施設情報案内、車内情報検索、インターネット楽曲検索) に加え、新たにコンピュータ利用案内 (QA) システムのテストを開始した。アンケート調査の結果を分析し、発話可能な内容および発話のタイミングの理解が、利用感の向上において重要な要素であること等の知見を得た。
- (ii) 同一の音声対話システムを継続的に利用することで、利用性能が改善する「学習効果」について分析を行い、5 日間の使用により平均 60% の誤り改善率が得られることを性能改善の経時パターンとともに明らかにした。
- (iii) 音声対話システム構築のためのガイドライン項目を選定した。

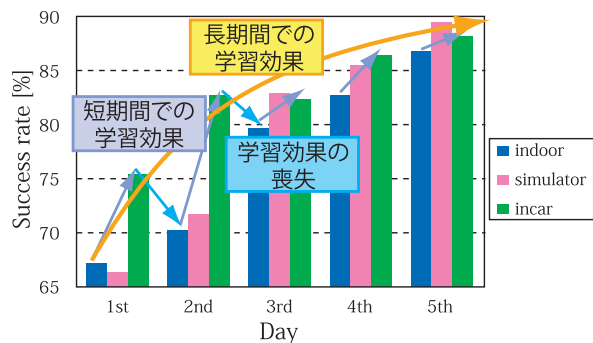


図 4 長時間利用による認識性能の改善

多様な音声合成プログラム

和歌山大学 システム工学部 河原英紀
<http://www.wakayama-u.ac.jp/~kawahara/>

1. ねらい

人間の音声には、文字で表すことのできる言語情報だけではなく、話し手の感情やその人らしさなどを反映する情報が含まれている。本プロジェクトでは、機械と人間との対話を自然なものとするために、人間のように多様な声質を有する音声を作成することの出来るプログラムを開発する。

2. 研究の概要

多様な声質の合成音声を作成するためには、人間の肉声に匹敵する品質の良い音声を作成でき、操作できる技術が必要である。そのための基盤として、本プロジェクトでは、我々の開発した高精度音声分析合成系 STRAIGHT を用いる。STRAIGHT は、人間の音声を (1) 声の高さを表す基本周波数、(2) 音韻や音色に関連するスペクトル、(3) 有声音や無声音を区別する音源情報に精密に分解し、加工・再合成を可能とする方法である。この研究手段を用いて、まず、話し手による声質の違いや話し手の感情や話し方による声質の違いを取り出すためのデータベースを整備する。これらのデータベースの解析結果を利用して、様々な声質を合成音声に付与することの出来るプログラムを開発する。開発するプログラムは、高品質で精密な声質の付与が可能ではあるが非リアルタイムのものと、品質には制限があるがリアルタイムで動作するものの二系統とする。

3. H16 年度の進捗

本年度も研究計画に従い下記の項目について研究を推進した。

- (i) 感情付与音声データベースを活用した感情制御パラメータ変換関数の近似および依存性の研究を進めた。STRAIGHT を用いたモーフィングで得られる感情制御パラメータ変換関数は、対象とする感情と音韻、文脈、話者に依存する。本年度は、母音部分での変換関数の統計処理により簡易な感情音声への変換法を開発した (図 5)。
- (ii) 声質変換パラメータの体系化においては、STRAIGHT スペクトルに含まれる音源波形依存部分と、声道形状依存部分とを明らかにした。また、声質変換パラメータの体系化として、STRAIGHT をベースとして正弦波モデルと波形ベースのモデルを統一的に表現できる AMALGAM 法を開発した。
- (iii) 隠れマルコフモデルを用いた韻律制御モデルの学習の素材として用いることのできる音声素材データベースを構築した。
- (iv) これらの当初計画の推進に加え、STRAIGHT に基づく

音声変換法を、他人には聴こえないレベルのつぶやき声 NAM(Non Audible Murmur) から通常の音声への変換に応用する技術を開発した。

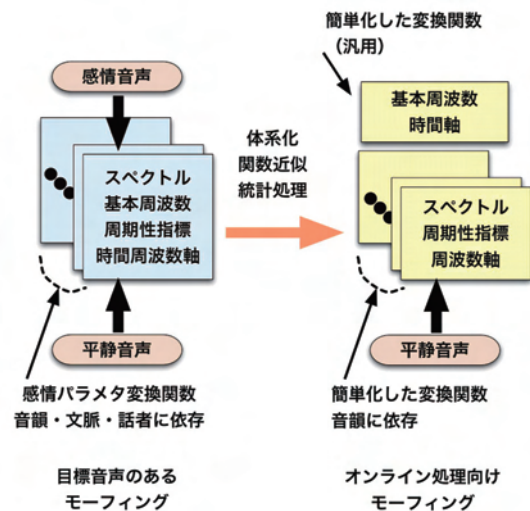


図 5 感情制御パラメータ変換関数の体系化と近似に基づく簡易な感情音声変換法