解 説

モバイル端末のためのユーザインタラクション技術（後編）
－入力対話技術－

旭　敏之
仙田 修司
磯谷 亮輔
日本電気（株）共通基盤ソフトウェア研究所

ユーザインタフェース（UI）の中心的パラダイムは，30 年前に提唱されたグラフィカル・ユーザインタフェース（GUI）のそれがほぼそのまま踏襲されてきたが，近年，ユビキタス・コンピューティングの浸透とともにコンピュータの利用形態が多様化し，新たなアプローチが求められている．その中で，フルサイズ・キーボードやマウスを持たず，表示画面の小さいモバイル端末におけるユーザインタラクション技術の研究開発が喫緊の課題となっている．

前（6 月）号では表示対話系の技術として，マルチスケール表示や Web コンテンツの変換表示技術について解説した．後編にあたる本稿では，入力対話系の技術として特に付属カメラを利用したテキスト入力技術や，モバイル向け音声入力技術について詳しく説明する．また最後に，今後の技術展望について触れる．

前編でも触れたように，入力操作におけるモバイル端末の特徴は，フルサイズ・キーボードやマウスなど PC 標準の入力デバイスに比べて操作上の制約が大きい点であり，これらを克服して使いやすく快適な利用環境を提供することは，ユーザインタラクション技術の大きな課題である．もちろんモバイル端末は制約が強いだけではなく，たとえば GPS などの位置情報入力と連動したサービスや，電子チケット／マネーなど近距離無線通信による周辺端末との連携など，モバイル端末ならではの環境と連動した新機能／サービスが実現されてきている．また，ゲーム端末や一部の携帯電話では，端末の動きを検出してジェスチャ入力を可能にする UI も実用化されている．

ただし，これらを入力技術として見た場合には，限定されたサービスやアプリケーションにおいて（位置や端末連携のための情報など）特定の情報を入力するために用いられるものであり，モバイル端末が備える多様な機能を利活用するためには，より汎用的な入力対話技術が必要である．

これに対して，メニュー操作などの座標データ入力に関しては，タッチパネル，ジョグシャトル，ポインティング・スティック，タッチホイールなど，多くのポインティングデバイスが製品化されてきた．ここでは，さらに複雑な操作を必要とするテキスト入力に着目し，実用化レベルにあるユーザインタラクション技術について，その課題と解決策について解説する．

入力対話技術：モバイル向けテキスト入力技術

携帯電話や PDA（Personal Digital Assistant）などのモバイル端末では，テキスト入力にさまざまな工夫が凝らされてきた．多機能な PDA では，GUI を実現するためにタッチパネルを備えるものが多く，初期の製品ではこれを利用したソフトウェア・キーボードと，手書き文字認識によるテキスト入力が主流であった．画面にキーボードを表示してペンでタッチするソフトウェア・キーボ

図 -1　予測入力の画面例

ードは，各キーが小さく押しにくかったり，素早い入力が難しかったりする問題があった．また，手書き文字入力はリソースの制約により十分な認識性能を得ることが難しく，「機械が認識してくれる文字をユーザが学習する」必要まであった．

最近の傾向としては，コンパクトなフルキーボードを搭載する方向か，テンキーベースで入力方式を工夫するという，2 つの方向が主流となっている．前者は比較的大型な PDA もしくは PDA 機能を持つスマートフォンに多く，キートップを小さくして他キーとの干渉を避けたり，不要なキーボードを一時的に隠すなど，デバイスそのものに工夫が凝らされている．後者では，数少ないキーでも素早く正確に入力できるように，かな漢字変換やキー操作／文字コード変換のアルゴリズムが開発されてきた．以下ではその代表的な例として，予測入力方式と “T9” 方式について紹介する．

POBox ™（Predictive Operation Based On eXample）に代表される予測入力方式は，各社の携帯電話に搭載されてきたこともあり，最も普及している手法である．一般に，予測には 2 種類のフェーズがあり，1 つは “ お ” の入力で “ お願いします ” を提示するといった先頭一致に基づく単語補完機能，もう 1 つは “ よろしく ” の後に “ お願いします ” を提示するといった繋がり予測機能である．前者の例を**図 -1** に示す．ユーザが予測結果の選択操作をしない限りは通常の入力を妨げないよう設計されているため，さまざまな文字入力手段との組合せが可能である．また，補完もしくは繋がり予測する単語は自動化された機械学習で獲得することができるので，あらかじめ多人数のデータから学習しておくシステム辞書と，ユーザの入力結果をリアルタイムに学習するユーザ辞書の両方を利用できる．予測入力の効果は，特に 1 文字の入力に時間がかかる状況で顕著である．その反面，予測に頼りすぎると文章がパターン化してしまうことが欠点に挙げられている．

T9 は 1 文字を 1 キーで入力する点を特徴とする．複数の文字を 1 つのキーに割り当てているために入力に曖昧性が残るが，これを単語候補選択と組み合わせることによって解消する．たとえば，あ行を 1，は行を 6，や行を 8 のキーにそれぞれ割り当てた場合，「1681」を入力すると，それに対応する「おはよう」「いひょう」などが候補に出てくるので，ユーザはその中から入力したいものを選ぶ．文字列を入力し終えてから候補の選択を行うという一連の流れは，かな漢字変換における同音異義語の選択と同様で抵抗が少なく，効率的な入力のために学習による候補順の並び替えが重要となる点も同様である．従来手法に近い感覚で効率的な入力ができる点は大きなメリットである一方，候補に出てこない文字列の入力やかな漢字変換との組合せで特殊な操作になってしまう点が新規ユーザの敷居を高くしている．

以上，予測入力方式と “T9” 方式について述べたが，特にこの 2 つは携帯電話のテキスト入力方式として広く普及している．このほかにもさまざまな方式やバリエーションが提案されており [1)]，タッチパネルやテンキーを利用したテキスト入力は，すでにある程度成熟した技術と見なすことができる．

その一方，モバイル端末ではマイクやカメラ，GPS といった各種センサが装備されているが，これらをデータ／ユーザ操作の入力装置として利活用することで，ユーザの利便性を高めたり，新たなサービス／機能を提供できる可能性がある．以下では，その具体例としてカメラとマイク（音声入力）を用いたテキスト入力技術について詳しく解説する．

携帯カメラを利用したテキスト入力技術

最近の携帯電話や PDA ではカメラを備えるモデルが主流であり，これをキーやペン以外の新しい入力デバイスとして利用する技術が開発されている．ここでは，携帯カメラによるテキスト入力技術として，バーコード入力と文字認識（OCR）を取り上げる．

“QR コード ” に代表される 2 次元バーコードは，数字／英字／日本語などのテキスト文字列を機械可読な図形に変換したもので，読み取り位置の基準となるマーカや誤り訂正符号を付与することにより，読み取り精度が高くなっている．特に近年では多くのカメラ付き携帯電話が標準機能としてバーコード読み取りを備えることから，アクセスしてほしい URL をバーコードで印刷したり Web ページに画像として表示したりすることが多くなった．また，任意のテキストや電話帳に登録するための情報などもコード化することができるため，情報提供の際の有用性が高い．サービス提供者による積極的な情報提供手段として，今後もより一層の普及が見込まれる．

バーコードだけでなく，広く実世界にあるテキスト情報を入力するために開発されたのが，同じ携帯搭載カメラを使ってテキスト認識を行う OCR 機能である．当初，携帯に搭載されたカメラは画素数など性能的な制約が強

図 -2　携帯カメラによるテキスト入力

かったため，読み取れるテキストやその印刷媒体が限られていた．しかし最近では，高画素かつ接写可能なカメラが搭載されるようになったため，英数字だけでなく漢字まで読めるようになり，名刺に書かれた難しい名前，雑誌に書かれた URL，ポスターに書かれたメールアドレスなど，さまざまなテキスト列を手軽に入力できるようになった [2]．**図 -2** にカメラ付き携帯電話によるテキスト入力の様子を示す．

携帯カメラによるテキスト読み取り技術は，従来の光学スキャナによる文書読み取り技術から派生したものであるが，スキャナとカメラ，バッチ処理と実時間処理，ユーザの利用環境の違いなどがあり，従来技術とは異なる特徴を持つ．具体的には，カメラでテキスト画像の入力を行う場合，環境光の影響，ピントずれ，手ぶれ，斜めからの撮影といったスキャナの場合には生じない問題に対処する必要があり，また，携帯端末の限られたリソース（ROM・RAM 容量や CPU 処理能力）においても，ユーザを待たすことのない応答を返す必要がある．

以下，携帯カメラによるテキスト読み取りの処理手順例を筆者らが開発に携わったアクセスリーダー™を例に解説する．

（STEP1）テキスト画像の撮影

従来のスキャナによる文書認識では，基本的に用紙全体を取り込んで全体（もしくはマウスで領域指定した範囲）を認識していた．それに対して携帯カメラの場合は，カメラを動かして画面を調整することで，認識させたいテキスト列を直感的に指定することができ，また処理範囲を限定できることから処理速度の向上にも繋がる．カメラによる撮影では接写することで認識に必要な解像度を確保するが，長いテキスト列は 1 画面内に入りきらない場合があるので，分割して撮影されたテキスト列の読み取り結果を繋げる機能が必要となる．アクセスリーダー™では，分割撮影時のテキスト列の重なりを自動的に探して認識結果を結合するため，連続して撮影するだけで長いテキスト列の入力ができる．

（STEP2）画像補正と 2 値化

取り込まれたテキスト画像には，ノイズを除去するための平滑化，文字と背景を黒と白で表現する 2 値化，テキスト列の傾きの補正といった処理が施される．カメラ画像に特有の処理として，手や端末などの影の影響の除去があるが，文字と背景のコントラストがはっきりしている場合はほぼ問題なく除去できるので，ユーザによる調整は必要はない．しかし，光沢のある紙面からの光の反射の影響は除去しきれない場合があるので，撮影する角度を変えるか光を遮るなど，撮影時の工夫が必要となる．

（STEP3）文字切り出しとテキスト認識

文字切り出しとテキスト認識に関しては OCR の技術がほぼそのまま利用できる．ただし，紙面の歪みやセンサ・レンズの精度の問題などから，カメラで撮影したテキストは低品質であることが多いため，そのような状況に対応した手法を適用する必要がある．たとえば，文字認識辞書に傾いた画像を登録する，かすれや潰れに強いテキスト認識手法を用いる，などである．

一般に携帯カメラによるテキスト入力では，読み取れる書体と読み取れない書体をユーザが意識することが必要になる．通常，携帯端末では大きな辞書を搭載できず，文字認識辞書に登録されていない書体は読み取り精度が極端に低く（あるいは読み取れなく）なるからである．

（STEP4）読み取り対象に応じたテキスト処理

カメラで撮影するテキスト情報の属性（“URL” や “ 電話番号 ” など）が分かれば，それに応じた処理が可能となる．たとえば URL であれば，“http://” で始まっている，“.jp” や “.com” で終わることが多い，などの性質を利用して認識結果の誤りを自動的に修正できる．また，読み取り対象が日本語の場合には，日本語らしい文字の繋がりを確率で表現した n-gram 辞書などによっても修正できる．

（STEP5）誤認識の修正

正解率 100％の認識技術は存在しないので，認識技術をアプリケーションとして提供する際には認識誤りをどのように修正するかが重要なポイントとなる．通常，OCR などで認識誤りの修正によく用いられるのは，誤り部分をユーザがキーで入力し直す，誤った認識候補から正解を選ぶ，といった手法である．

アクセスリーダー™では，認識誤りの周辺を再度撮影し直すだけで自動的に誤りを修正する手法が搭載された．この手法と STEP1 で述べた分割撮影による自動連結と

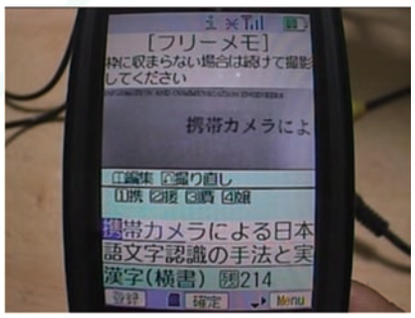


図 -3　文字撮影／入力時の自動連結と自動修正

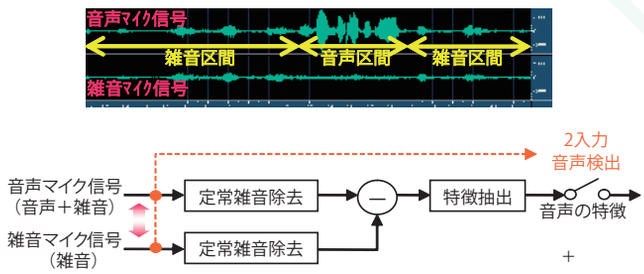


図 -4　2 マイク耐雑音音声認識

は同じ原理で実現されているため，ユーザはカメラでテキストを撮影するだけで，初回の認識，自動連結，誤認識の修正の 3 種類の処理を行うことができる（**図 -3**）．

以上，カメラを利用したテキスト入力技術について述べた．これらの技術の課題は，まず入力上のさまざまな制約を無くしていくことである．バーコード入力の場合，現状では 1 つのバーコードを適当な大きさで撮影する必要があるが，本来は，適当に撮影しただけで，画面内の複数のバーコードを読み取るようにするべきだろう．また，OCR 機能についても，カメラで捉えた任意のテキスト列を簡単に取り込めるようになることが望まれる．現在，OCR 機能は名刺などに印刷された限られた情報の入力が，応用として想定されている．看板や冊子などに書かれた情報を手軽に入力できるようになれば，単に入力の手間が省けるだけでなく，たとえば劇場ポスターから読み取った情報がその場でスケジューラに反映されるなど，実世界のテキストと連携したさまざまなサービスが生まれる可能性がある．また，OCR 機能の特長の 1 つとして，「ユーザがその文字（の読みなど）を知らなくても，撮影するだけで入力できる」ことが挙げられる．これと翻訳や辞典機能などが組み合わされば，世界のどの国でも周囲の文字が理解できるといった，新しい価値を生み出すことができる．このように，単に「使いやすくする」だけでなく，「人の能力を拡張する」ことも UI 技術の重要な課題である．

入力対話技術：音声入力技術

携帯電話は，もともとコミュニケーションのための端末であり，音声を利用したテキスト入力技術への期待は大きい．現在すでに多くの携帯電話に音声ダイヤル機能が搭載されているが，自由文の入力など，より高度な機能を持つ音声認識技術の開発が進められている．ここでは，モバイル端末で特に課題となる耐雑音と省リソースでの高機能音声認識を中心に，技術動向を解説する．

■耐雑音音声認識

モバイル端末は屋外での使用が多く，雑音下での認識精度劣化を抑えることが課題となる．基本的な方式は，音声が入る前の信号区間などを利用して雑音の周波数スペクトルを推定し，それを利用して雑音混じりの音声信号から雑音を除去してから認識を行うというもので，SS（Spectrum Subtraction）法やウィナーフィルタが代表的である．近年では，音声のモデルを利用して雑音除去の精度を高める方法などが盛んに研究されている．また，マイクロホンアレーを用いた雑音除去も多く研究されており，話者方向からの音声のみを拾うようにビームを形成する方法のほか，ICA（独立成分分析）によるブラインド音源分離技術を用いる方法等も研究されている．耐雑音音声認識においては，雑音の種類や S/N の変動に対する頑健性，非定常雑音，目的音声以外の人の声，複数方向から到来する雑音などが課題となっている．

ここでは，例として業務用端末向けの 2 マイク耐雑音音声認識技術について紹介する．本技術は，生産現場やせり市場などの高騒音下においてハンズフリーでの端末へのデータ入力を可能とするものである．ヘッドセットの口元のマイクとは別に雑音のみを拾うマイクを装着し，2 つのマイクからの信号を用いて音声区間検出と雑音除去を行う（**図 -4**）．2 つの信号のパワー比を用いて，信号中で音声が含まれる区間を検出する．検出された音声区間について，各マイクの信号に対して SS 法により定常雑音を除去した後に，2 つの信号間でのスペクトル減算を行い，非定常雑音を除去して認識を行う．このような処理により，地下鉄騒音なみの雑音下での音声認識が可能となっている．

■省リソースでの高機能音声認識

モバイル端末では CPU 速度やメモリ容量などリソース面の制約が強い．電話帳に登録した名前の呼び出しなど，数十～数百程度の単語の認識であれば問題ないが，自由文での情報検索やメール等のテキスト入力を行うには，数千から数万語の大語彙の連続音声認識が必要となり，認識処理に多くのリソースを要する．

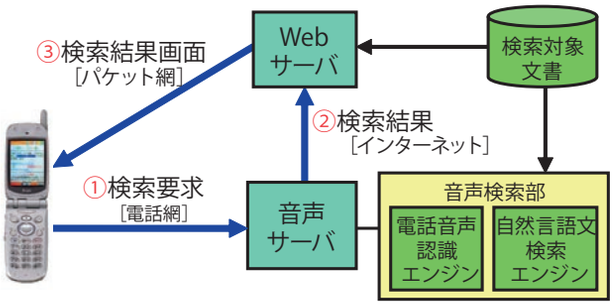


図 -5　音声／ Web 連動型検索システム

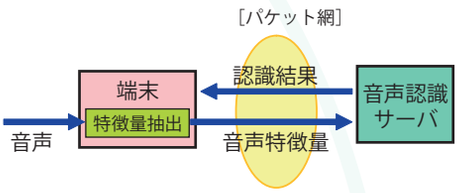


図 -6　分散型音声認識

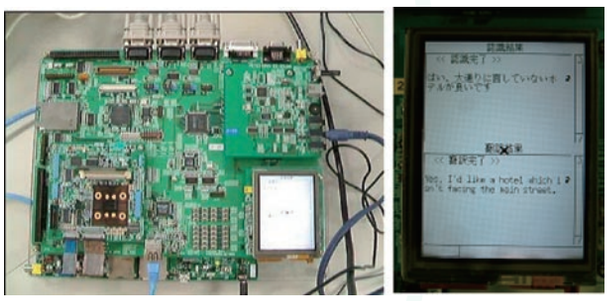


図 -7　携帯電話用プロセッサで動作する自動通訳システム

この問題に対する 1 つの解決策は，端末とサーバで処理を分散し，音声認識処理をサーバで行う方法である．特に，サーバやインターネット上の情報を音声で検索するような用途では，このような枠組みが有効である．**図 -5** に筆者らの開発した音声／ Web 連動型検索システムの構成を示す．携帯電話の電話網を用いて音声サーバに音声で検索要求を送信し，サーバで認識・検索を行い，結果の画面をパケット網経由で端末に送信する．端末には，Web 閲覧機能を有する一般の携帯電話を用いることができる．

また，端末で音声認識処理の一部を行い，残りをサーバで行う分散型音声認識（DSR：Distributed Speech Recognition）と呼ばれる方式もある．たとえば，特徴抽出処理までを端末で行い，抽出した特徴量を圧縮してパケット網でサーバに送り，残りの認識処理をサーバで行う（**図 -6**）．au から提供されている“声 de 入力™”は，本方式を用いたサービスの例である [3]．操作性や通信コストの面で優れるが，特徴量抽出などの機能を搭載した専用の端末が必要となる．

一方で，モバイル端末単体で動作するコンパクトな大語彙連続音声認識の研究開発も進められている．筆者らは，MDL（Minimum Description Length：記述長最小）基準に基づく音響モデルサイズ削減などにより，音声認識の各データやモジュールのコンパクト化，処理の高速化を図ることで，PDA 上で 5 万語の大語彙連続音声認識を実現し，これと翻訳技術を組み合わせることで旅行会話文の自動通訳を行うシステムを開発した [4]．本技術は，通訳機能を持つ携帯端末“VoToL™”として製品化されている．さらに，音声認識処理を並列化することにより，携帯電話向けの省電力マルチコアプロセッサでのリアルタイム動作が実現されている [5]（**図 -7**）．並列化の効果を最大化するため，音声認識処理の中で負荷の大きい照合処理に先読み処理を導入して処理を分割することで，負荷の均等化を図っている．このような技術により，近い将来携帯電話単体での自由文音声入力が実用化されると期待される．

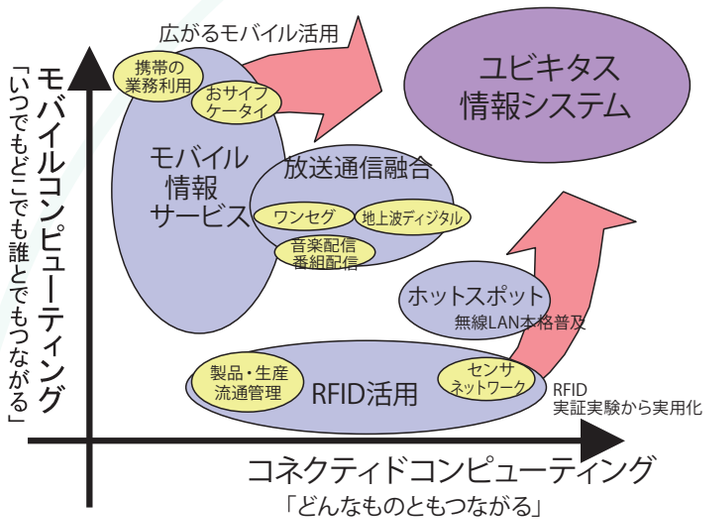
音声入力は自然で使いやすいインタフェースとしての期待は高いが，声に出すことに心理的抵抗がある場合もある．これに対して近年，小さな声やささやき声も拾うことのできる骨伝導マイクや肉伝導マイクなどの開発が進んでいる．これらは耐雑音の面でも有利であり，今後の高性能化と音声認識への活用が期待される．

本稿では説明を割愛したが，ユーザインタラクション技術としては高品質で自然な音声合成も重要である．音声合成は，一般に波形の素片や発音のための辞書情報など多くのデータを必要とするが，モバイル端末の限られたリソースでいかにこれらのデータを効率よく保持し，自然な音声合成を実現するかが課題である．

ユビキタス時代に向けたユーザインタラクション技術の展望

メディアを賑わす「ユビキタス」の掛け声とともに，携帯の普及や TV によるインターネット利用など，端末の多様化が進んだ．過去 30 年にわたって PC&GUI 環境で培われた膨大な資産（これには利用者の「GUI 対話に対する経験や慣れ」も含まれる）を，これら非 PC 端末で快適に利活用できるようにすることは，「いま，ここで」必要とされている技術である．本稿では，そのためのユーザインタラクション技術の例を，表示系と入力系に分け 2 回にわたって解説した．ただし前編冒頭でも述べたように UI 技術領域はスコープが広い．対象をモバイル端末に絞ったとしても，UI デバイス，コンテキスト利活用，サービス連携，設計・評価，UI 開発環境，人間特性の探求などなど，本稿以外のさまざまなテーマの取り組みが必須である．

図 -8 に示されるように，現状のモバイル・コンピュ



文献 6）から許諾を受けて転載

図 -8　ユビキタスシステムの現況

ーティングやコネクティッド・コンピューティングの進展は，本格的なユビキタス社会に向けての 1 ステップであると考えられる．その中で，本稿で解説したようなモバイル端末向け UI 技術は，いわば過渡的な技術と見なすことができる．今後，端末のモバイル化やウェアラブル化のさらなる進展，機器／環境埋込み型端末の普及，RFID やセンサ情報を活用した実世界融合型サービスの実用化などが予想される中，提供される多様な情報／サービスに快適かつ自然にアクセスするため，GUI に代わる新たな UI パラダイムの創出が望まれる．

筆者らは今後のユーザインタラクション技術の進化には，大別して 2 つの方向があると考えている．1 つは生活空間や周囲環境／機器に埋め込まれた CPU や端末を駆使し，そのときの状況に適したサービスや情報を利用者が享受できるようにする UI である．これまでのような「使いやすさ」や「学習容易性」や「効率」だけでなく，「さりげなさ」「自然さ」「魅力」などの属性が重要視される．また，対話技術に加えて，各種 UI デバイスやセンサ，端末間連携，実世界連携，状況推定などの技術が重要になるだろう．

もう 1 つの方向は，広大な情報／サービス空間の案内役として，仮想的なエージェントが自律的に振る舞うような形態である．キャラクタやロボットという形で UI 自身が仮想的な人格を持ち，利用者のパートナーとして適材適所な情報／サービスを提供する．このためには自然な対話のためのマルチモーダル技術や自然言語理解技術，周囲の環境や利用者の意図を理解する状況／意図推定技術などとともに，膨大な情報空間の構造化／ハンドリング技術が必須になる．

これら 2 つの対話様式を仮に対立概念として捉えると，1990 年代に起こった「知的 UI か直接操作か」論争が想起される [7]．ここ 30 年の GUI の普及を見ると，少なくともこれまでは「直接操作」が主流であったと言えるが，今後，サービスや端末や利用シーンの多様化が進めば，“ 何が本流か？ ” といった議論は意味を失っていくだろう．そこでは，利用者の置かれた環境やサービスの内容によって，適材適所な UI を提供することが最も肝要になる．今後の UI 研究者には，インタフェースの原点である人間特性に立ち返り，状況やサービスに応じてどこでどんな UI で提供すべきかといった，フレームワークの構想力が問われている．

参考文献

1) 増井俊之：携帯端末のテキスト入力手法 , ヒューマンインターフェース学会誌 , Vol.4, No.3, pp.131-144 (2002).
2) 仙田修司，西山京助，旭　敏之：携帯カメラによる日本語文字認識の手法と実現，電子情報通信学会技術報告書，パターン認識とメディア理解研究会，PRMU2004-124 (2004).
3) 加藤恒夫，河井　恒，宇都宮栄二：分散型音声認識の商用システム構築，情報処理学会研究報告 2006-SLP-63，pp.39-44 (2006).
4) Isotani, R., Yamabana, K., Ando, S., Hanazawa, K., Ishikawa, S., Emori, T., Iso, K., Hattori, H., Okumura, A. and Watanabe, T.: An Automatic Speech Translation System on PDAs for Travel Conversation, Proc. ICMI'02, pp.211-216 (2002).
5) 石川晋也，山端　潔，磯谷亮輔，奥村明俊：携帯電話用プロセッサで動作する大語彙連続音声認識の並列処理，FIT2005 (第 4 回情報科学技術フォーラム)，pp.121-122 (2005).
6) 原　良憲，山田敬嗣：情報システム基盤における Symbiotic Computing，情報処理 Vol.47, No.8, pp.844-850 (2006).
7) Maes, P., Shneiderman, B. and Miller, J.: Intelligent Software Agents vs. User-Controlled Direct Manipulation: A Debate, ACM SIGCHI97 Extended Abstracts, pp.105-106 (1997).

（平成 19 年 6 月 11 日受付）

旭　　敏之（正会員）
t-asahi@bx.jp.nec.com

1984 年大阪大学大学院基礎研究科修士課程修了．同年 NEC 入社．1997 年奈良先端科学技術大学院情報科学研究科博士後期課程修了．ユーザインタフェースの研究開発に従事．現在，NEC 共通基盤ソフトウェア研究所勤務．ヒューマンインタフェース学会会員．

仙田　修司（正会員）
s-senda@ap.jp.nec.com

1996 年京都大学大学院工学研究科情報工学専攻博士課程修了．工学博士．同年 NEC 入社．手書き／印刷文字認識，ユーザインタフェース，画像圧縮応用の研究に従事．現在，NEC 共通基盤ソフトウェア研究所勤務．電子情報通信学会会員．

磯谷　亮輔
r-isotani@bp.jp.nec.com

1985 年東大・工・計数卒業，1987 年同大学院修士課程修了．同年 NEC 入社．以来，音声認識，自動通訳システムの研究開発に従事．現在，NEC 共通基盤ソフトウェア研究所主任研究員．電子情報通信学会，日本音響学会各会員．