オーディオビジュアル複合情報処理 10－6
（１９９５．９．１４）

自動通訳システムにおけるマルチモーダルインタフェースの検討

鈴木 雅実　井ノ上 直己　橋本 和夫　谷戸 文廣

KDD 研究所

〒356 埼玉県上福岡市大原 2-1-15

E-mail: {msuzuki, inoue, kh, yato}@kddnews.nes.lab.kdd.co.jp

本稿では，筆者らが試作したホテル予約を対象領域とする日韓自動通訳システムのユーザインタフェースの改良事例を通して，自動通訳システムにおけるマルチモーダルインタフェースについて検討した。具体的には，(1) 音声認識結果の確認方法，(2) マルチモーダル入力の可能性，(3) タスク依存性の利用に関して考察を行なった。

A DISCUSSION ON MULTIMODAL INTERFACE FOR AUTOMATIC INTERPRETATION SYSTEMS

Masami SUZUKI　Naomi INOUE　Kazuo HASHIMOTO　Fumihiro YATO

KDD Research and Development Laboratories

2-1-15, Ohara, Kamifukuoka-shi, Saitama 356 JAPAN

E-mail: {msuzuki, inoue, kh, yato}@kddnews.nes.lab.kdd.co.jp

In this article we discussed the issue of multimodal interface for automatic interpretation (speech translation) systems, based on our experience on the improvement of user interface for our Japanese-Korean automatic interpretation system, considering the task domain: "Hotel Reservation". The topical items are (1)Confirmation of recognition result, (2)Possibility of multimodal input and (3)Utilization of task-dependency.

1 はじめに

これまでに発表されている自動通訳(音声翻訳)システムの開発事例では，ユーザインタフェースに関する言及がほとんど無いか，有ったとしてもその点に関する検討内容の報告は極めて少ない。学会誌等における解説等でも，重要性の指摘がなされてはいるが具体的な議論は同様に少ない状況である([1] [2] [3] など)。

筆者らは韓国の研究機関との共同で行なった日韓自動通訳システムの試作過程でユーザインタフェースの重要性を強く認識し，この点を考慮したシステムを作成して，韓国との共同実験およびシステムの評価実験を行なった [5]。本稿では，この試作経験を通じて得られた知見に基づいて，自動通訳システムにおけるマルチモーダルインタフェースについて考察する。

2 日韓自動通訳システムの環境条件

序章で触れたように，筆者らのシステムは，韓国の研究機関と共同して約1年半の間に双方でそれぞれ「日→韓」(日本側) および「韓→日」(韓国側) 方向の音声翻訳処理を行なうシステムを作成し，相互に接続することを前提に研究開発を開始したものである。このようなシステム間接続を行なうに当たって，各々のシステムの基本的な性能に関する目標および，システム間の接続方法についての検討を経て試作を行なった [4]。

この結果，日本(KDD)側ではハードウェア資源として汎用のワークステーション(Sparc-20)1台と音声認識の前処理(音響分析)用の4個のDSPを用いて，リアルタイムの2倍以内で音声翻訳処理を行なうシステムを実現した。処理精度についても，発話状況監視を導入することにより，上以5位内で平均90%以上の文認識率を達成したほか，柔軟な処理単位を用いた頑健な翻訳処理により，コミュニケーション上は十分な翻訳結果が得られるようになった。

しかし，このように性能面で所期の目標を満足することに成功したが，その段階ではユーザインタフェースついては明確な仕様がなく，次のような基本条件に従って実装した暫定的なインタフェースを用いていた。

- マイクからの音声入力
- 音声認識結果は複数候補をモニター画面に表示
- 正解認識候補をマウスにより選択
- 認識結果の選択と同時に翻訳処理を起動
- 認識結果と翻訳結果のテキスト情報を相手側に送信し，これに発話権の交替情報を付加する

従って，システムのユーザインタフェースとしては，音声入力の受理から音声翻訳処理結果と発話権に関する情報の送出までを円滑に実行できるような，モニター画面表示とマウス操作を伴った形態となり，必然的に一種のマルチモーダルインタフェースを指向することとなった。

さらに，対象領域等から来る前提条件を説明する。本システムで自動通訳の対象としている「ホテル予約」のタスクは，部屋の予約を希望する利用者(Client)とホテル予約係(Hotel)との間で交わされる，一種の協調的な目的指向の対話で構成される。この仮定の下に，利用者側話者がホテル側の質問に適宜答える，ホテル主導型の対話を処理の対象として考えることができる。

また，音声入力は連続的な文発声を基本とするが，現段階ではシステムの基本性能に影響を与えると考えられる冗長語や言い淀み等は扱わないこととした。「ホテル予約」の対話では，宿泊日程や部屋タイプの希望，人数・名前等の確認が行なわれる。語彙は両発話者分を合わせて約1,000語の規模である。また，ホテル側話者はタスクの内容に習熟しているのに対して，利用者側話者はそうではないものと仮定した。

以上のような環境条件を考慮した，日韓自動通訳システム構成の概略を次の図1に示す。

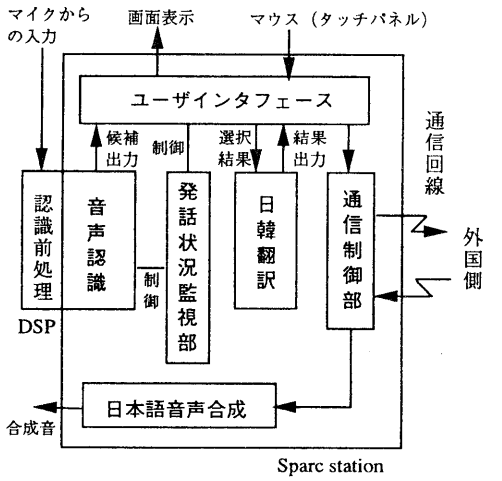


図 1: システム構成図

注) 韓国側も通信回線を挟んで対称的なシステム構成である。また上記の構成図では省略したが，韓国との共同実験の際には別途ISDN回線を用いたTV会議システムを用いて，対話相手の顔画像(システム画面の一部分に表示，図2を参照)と入力音声を相互に送受信した。

3 日韓自動通訳システムにおけるユーザインタフェースの改良

3.1 双方向の通訳対話の進行

前章で述べた日韓自動通訳システムの環境条件に基づいて，最初に実現したユーザインタフェースを通した場合の，システムの動作とユーザの操作例を順を追って示す。なお，下線部は後に改良対象となった部分である。また「選択」・「指示」操作はすべてマウスクリックにより行なう。「表示」・「出力」はシステムによりなされる。

(0) 初期状態:

(1) 発話者ロールの切替: 「切替」ボタン
利用者／ホテル予約係のいずれかを選択
(以下，利用者側を選択したと仮定)

(2) 外国側へアクセス:
「電話」ボタンのクリック

(3) 外国側の処理結果受信:
韓日翻訳結果表示部に先方の音声認識結果(ハングル)および翻訳結果表示

(4) 受信翻訳結果の音声合成
日本語合成音を出力

(5) 音声入力
「入力」ボタンのクリック後に発声

(6) 音声認識処理
発話検出後に開始，終話検出後に終了
音声認識候補(最大5位まで)の表示

(7) 正解候補の選択
表示された候補のうち一つをクリック

(8) 翻訳処理

(9) 認識・翻訳結果の送出
翻訳処理の終了と同時に相手側に自動送出

(A) 続けて発声の場合，(5) に戻る

(B) 相手側にターン(発話権の委譲)情報を送出する場合
「ターン」ボタンをクリック
(3) に戻る

このような自動通訳対話を実際に試行したところ，幾つかの問題点が明らかとなったため，次節で述べるようにそれらの改善を行なって，より使い易いユーザインタフェースの実現を図った。改良後のシステム表示画面を図2に示す。

3.2 ユーザインタフェースの改良内容

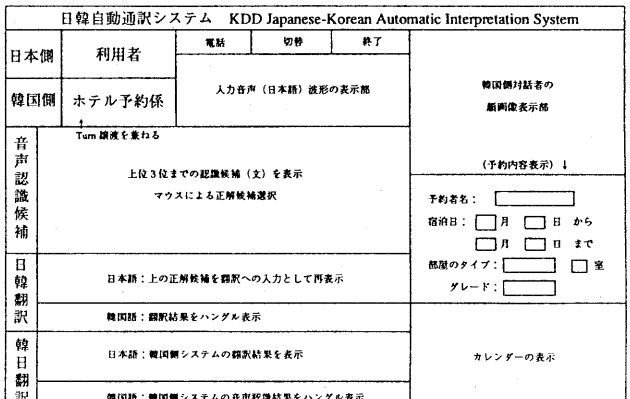
3.2.1 発話開始のタイミング

本システムでは当初，発話の開始時にマウスクリックによる指示が与えられてから発声を受理していたが，このような操作を伴なうことはユーザに負担を与えると考えられる。一方，常時システムを音声入力可能な状態にしておくことも，処理の負荷や雑音による誤動作の可能性が高くなることが予想される。そこで，日韓双方の対話の進行(すなわち発話権の移譲により交互に発声する対話環境)を考慮して，相手側からのターン信号到着後の適切なタイミングで音声入力可能表示を行なうことにより，操作無しで発声を開始できるように変更を加えた。ただし，最初の発声の認識が失敗した場合や，認識結果選択後に続けて次の発話を行なう場合は，入力可能状態に移行させるための指示(日本側のロール表示部分のクリック)を行なう必要がある。

3.2.2 発話権の交替

前記の通り，発話権の交替情報を相互にやりとりする方式を用いているが，その時点で発話権があるサイトは発話終了後に相手側の発話を促すターン(発話権交替)情報を送出する必要がある。一つの方法として考えられるのは，発話者が発声終了後にターンの送出を行なうことである。本システムでも当初はこの方法を採用していたが，発話者がターン送出を忘れたり操作自体が煩雑さを増すことになる。そこで，認識候補選択後3秒以内(可変)に続いて発声開始の指示が無い場合，自動的にターンを送出するように変更した。なお，対話の流れから相手側に意図的に発話権を譲渡する場合は，韓国側のロール表示部分をクリックする。

図2. システム表示画面



3.2.3 音声認識結果の確認

音声対話システムの現状の技術レベルとして，各発話の音声認識候補のうち第1位のものが常に正解として得られないことから，ある程度以上のタスク規模では発声内容の確認は避けられないと思われる。そこで，複数候補を尤度順に表示して発話者による確認・選択を行なうためのインタフェースを用意した。ただし，システム性能の向上 (上位3位以内に90%以上の正解率) を得たことと複数候補から一つを選択するユーザの操作上の負担を考慮して，当初5位まで表示していた候補数を3位までに削減した。さらに，計算機に不慣れな利用者を想定して，光学式のタッチパネルを実装し，マウスクリックと同様な指示選択操作を可能とした。

3.2.4 タスク依存型の対話内容表示

前述したホテル予約における発話者 (利用者／ホテル予約係) による発話内容の偏りを考慮して，対話開始時点でロールを選択する機能を設け，これと連動して発話状況監視による対話音声認識が実行される枠組を採用している。当初は，選択された認識候補 (日本語文) を対話履歴としてモニタスクリーン上に記録表示していたが，必ずしも見易いものではなかったため，予約内容のキーワードのみ (日付・部屋タイプ／グレード・名前) を予め用意した表示欄に，決定されたものから表示するインタフェースに改変した。また，日付と連動するカレンダーをこの表示欄の下に配置した。このことにより，発話者にとって部屋の予約という目標に照らして，どのような対話の状況にあるのかがより明示されるようになった。

以上述べたような改良の結果，対話中の大部分の発話については認識結果の確認時の選択操作のみがユーザに委ねられることとなった。

3.3 対話環境におけるシステム評価実験

以上のようなユーザインタフェースに関わる改良を行なった後に，韓国の共同研究機関との間で，国際通信回線を通じた「日→韓」および「韓→日」のシステムの相互接続実験を実施した。さらに筆者らの日韓自動通訳システムを対話環境で評価するための実験を行なった[1]。また，この実験では双方向の対話環境を提供するため，本システムに対向する韓国側システムの代わりに「韓日自動通訳疑似システム」を使用した。実験方法・実験結果は表1，表2の通りである [5]。

[1]この実験自体はユーザインタフェースを直接評価することを意図したものではない。

表1. システム評価実験方法

実験環境:	実験室内 (ほぼ定常的なノイズ環境)
被験者:	20歳台〜40歳台の男女各8名
事前説明:	(1) 別室で日韓自動通訳実験の模様を収録したビデオによる説明 (5分) (2) 実験室内で，本システムと疑似システムを用いた対話のデモと操作説明 (5分)
実験条件:	次の3条件を本システムで順次試行 (対話相手側は疑似システムで対向)
(1C)	・準備したシナリオを見ながらの読み上げ発声 (利用者側)
(2A)	・被験者自身が記入した予約内容のメモを見ながらの対話発声
(2B)	・前記とは異なる内容の予約メモを渡された被験者による再試行
付随条件:	認識候補選択の手段として半数の被験者にマウスを，残りの被験者にはタッチパネルを初めに用いて試行

表2. 対話環境でのシステム評価実験結果 (平均)

成功率の尺度	1C	2A	2B
1回の発声での成功率	95.1%	94.2%	96.5%
1発話当りの発声回数	1.05	1.13	1.07
正解時の1位認識率	88.7%	86.3%	87.2%
1ターン当りの所要時間	6.3sec.	6.3sec.	6.0sec.

注)1ターン当りの所要時間は，発声可能表示から認識候補選択までを指す (この場合1回の発声で成功した時の平均).

なお，被験者に対して実験終了後アンケート調査を行なったので，その結果を記す (括弧内は回答人数)。

1. システムの処理速度についての印象
   速い (1)，普通 (11)，やや遅い (4)，遅い (0)
2. マウスとタッチパネルではどちらが快適か?
   マウス (6)，タッチパネル (8)，その他 (2)
3. この種のシステムに必要な改善要素
   (意見の集約結果)
   間投詞や明瞭でない表現の受容等の柔軟性 (10)
   語彙の大きさ等の処理範囲の拡大 (5)
   高速性 (5)，　機能的な手軽さ (4)

実験結果を見る限り，各被験者とも，日韓自動通訳システムに対しては初めての試行であったにもかかわらず，相当に高いレベルの成功率を記録した。事前説明が十分になされていたこと，ホテル側との交渉的な要素は含まない比較的容易なタスクであったことなどが理由として考えられる。ユーザインタフェース改良の寄与の度合については数値的なデータを示すことはできないが，ユーザの操作ミスがほとんど無く，極端な所要時間の遅れ等も見られなかったことから，一定の効果があったものと思われる。以下に，この実験を通じて観察された事象の中から，次章で述べるマルチモーダルインタフェースの検討に関連する項目を数点指摘する。

(1) 今回設定したホテル予約のタスクの場合，利用者側の会話1発話の所要時間は，1回の発声で成功の場合約6秒間であり，その内訳は概ね
- 平均発声時間として2秒
- 認識処理の遅れ時間が2秒
- 発声者の反応時間が2秒 (発話可能表示から発声までと，認識結果表示から候補選択まで)

であった。また，わずかではあるが，反応時間に被験者の試行回数に従った学習効果が現れているようである。

(2) 反応時間には個人差が見られたほか，日付のように視覚的にも似通った候補が複数表示される場合に，選択に要する時間が長くなる傾向が見られた。さらに，2Aおよび2Bの実験条件では，発声内容と一致しないが意味的にほぼ等価な候補が表示された場合の選択時間も同様に長めとなった。

(3) 被験者へのアンケート結果を見ると，本システムの処理速度はタスクの遂行に関して遅過ぎるという印象はないものの，柔軟性の向上が要求されていることが分かる。ポインティングデバイスとしてのマウスとタッチパネルの選好については，マウス使用に慣れた被験者10名のうち各5名がそれぞれマウス／タッチパネルの方が快適としているのに対し，不慣れな被験者4名では1名以外はタッチパネルを好んでいる。また，タッチパネルの快適性は画面の大きさ等によるとの指摘も見られた。

4 自動通訳システムにおけるマルチモーダルインタフェースの検討

前章では，筆者らの日韓自動通訳システムにおいて当初用いていたユーザインタフェースの問題点を改善し，利用者の視点に立ってより使い易いインタフェースを実現した結果について説明した。本章では，前章までに述べた筆者らの経験に基づいて，今後の自動通訳システムにおけるマルチモーダルインタフェースの在り方について検討することにする。

4.1 音声認識結果の確認方法

自動通訳システムだけでなく，翻訳処理を伴わない対話型のマンマシンインタフェースにも共通する問題点であるが，ある程度以上の規模の音声対話システムでは発声内容の確認を何らかの形で行なう必要がある。第1位の認識候補のスコアが顕著に高い場合など，確認を行なわずに済むような状況も考えられるが，同一カテゴリーの認識候補 (例えば日付／部屋の種類／人名等) が多い場合の認識結果に対しては，適当な数の候補を表示して選択を促すためのインタフェースが必要である。この場合，何が認識対象となっているかを判断して，あるいはスコアの分布等を考慮して候補の数を可変にすることも考えられる。

さらに，本システムでは認識結果をそのまま文候補として表示したが，認識結果の表示形式の問題を検討する必要がある。前に述べた評価実験の際には類似の候補が複数表示された場合，自分の発声した通りの表現がどこにあるのか，あるいはそのものでなくても意味的に等価な候補があるのかを判断するのに多少の時間を要することが観察された。これは認識文法の記述にも関連する問題である。日本語の話し言葉に顕著に見られる助詞の脱落の有無に関して，両方の候補を表示してもユーザは戸惑うことが多いことから，本システムでは助詞付きの候補のみを出力するようにした。しかし，「～にして下さい」と「～をお願いします」のように，ほぼ同一の意味として用いられる別個の表現を一つにマージすることはしていない。また，このように複数の意味的に等価な候補が表示される場合，評価実験の項で述べた通り，ユーザが直ちに確認できない状況も見られる。ただし，この種の問題はユーザがシステムに慣れるに従って解消される部分もあると思われる。

4.2 マルチモーダル入力の可能性

本稿前半で述べた実験等を通じて得られた経験では，個人の名前や日付等の数表現については，音声認識技術では十分な性能を実現できない可能性がある。このため音声認識処理よりも確実で容易な入力手段があれば，それを組み合わせたマルチモーダル入力機能を備えたユーザインタフェースを導入する必要性を感じている。前項でも述べた通り，これらの入力対象は確認自体も手間がかかる要素であることから，簡易な方法による入力手段が期待される。

この場合に問題となるのは，入力対象によって最適な入力手段を提供するための方法である。すなわち，対象カテゴリー毎に，例えば日付ならばカレンダー型のタッチボードないし数字キー入力，名前ならばアルファベット／カナのタッチボードまたはキー入力を用意するといったことが考えられる。しかし，一回の発話で認識すべき対象が，基本的に単一のキーワードのみの場合は比較的容易と思われるが，複数のキーワードを含む文であった場合は入力操作の回数が増える可能性がある。さらに，名前等の固有名詞に関して言えば，初出時に音声以外のチャネルからの入力により登録された後は，音声による入力を可能とすることが望ましいと思われる。ただし，この際カナ以外の表記と発音の関係が明確でない場合も考えられるので，その対処も必要となる([6]等が参考になる)。

4.3 タスク依存性の利用

日韓自動通訳システムの事例では,「ホテル予約」のタスクに処理対象を限定したことで，タスクの特徴を利用して効果的なユーザ向けインタフェースの実現を図った。例えば，タスクに依存した予約内容の確認のためのフォームを利用することにより，ゴール達成に向けての対話の進行状況がより明確となる。

この点をさらに発展させて，現在話題となっている事象をシステム側からユーザに対して明示したり，発話に先だって，これから話題とする事象をユーザがタッチパネルによる選択等によってシステムに指示できるようにすると，音声理解率を向上させることが期待できる。音声対話においては，特に話題が転換する場合に次発話を予測することが困難になることから，このような話題の明示は有効であると考えられる。

この際，話題となる事象の表示や選択には，キーワードの文字列表現よりも視覚的なアイコンの方が一般ユーザに受け入れられ易いものと考えられる。また，話題の階層化や粒度等をどのように設計すればよいかも重要な検討項目と思われる。また，タスクに依存する制約を利用することにより，認識結果を一意に確定することも可能となる。実際，筆者らのシステムでも，ホテル予約における日付の整合性や，すでに予約内容として認識された結果を利用して音声認識候補を限定する試みを行なっている。

一方，翻訳の質的向上のためにも，現在話題となっている事象のほか，既知情報／新情報等の正確な把握が有効である場合が多い。しかし，上記のような音声認識対象の範囲指定としての話題情報の明示と比較した場合，翻訳処理に必要な(有益な)情報が発話者にとって自明でない場合が多いと思われるので，システムからユーザへの問い合わせ等のインタラクションがより複雑となることが予想される。

いずれにしても，対象領域を限定した上で自動通訳システムの利用可能性を拡大するためには，このようなタスク(領域)依存性を積極的に活かす方法を考案する必要がある。

5 まとめ

試作した日韓自動通訳システムのユーザインタフェースの改良内容とシステムの評価実験結果について報告した後，この経験に基づき自動通訳システムにおけるマルチモーダルインタフェース設計に関して検討を行なった。考察内容を整理すると，次のようになる。

(1) 他の音声対話システムと共通する問題として，音声認識結果の確認方法があり，これを効率化することによりユーザにとってより快適なシステムとなる。

(2) 当面の対象領域を限定した自動通訳システムに関して，音声以外の入力モダリティを備えたインタフェースを導入することにより，対話の進行がより円滑となることが予想される。

(3) 同様に，対象領域を限定することにより，その範囲に生起する話題をユーザがシステムに対して明示することにより，対話音声認識率および翻訳品質の向上が期待できる。

また，以上に関連して，マルチモーダルインタフェースを評価する方法も重要な研究課題と位置づけられる。

謝辞

本研究の遂行に際してご指導・ご助言を頂いたKDDの村谷常務取締役，浦野研究所長と，適切なコメントを寄せられた研究所各位に感謝いたします。また，システム作成の過程で桶谷・河井の両氏にはたいへんお世話になりました。さらに，日韓自動通訳共同実験の実施に際してご協力頂いた社内外の方々に，この場で厚くお礼申し上げます。

参考文献

[1] Woszczyana, M., et al.: "Recent Advances in JANUS: A Speech Translation System", *EUROSPEECH '93*, 1993.

[2] Morimoto, T., et al.: "ATR's Speech Translation System: ASURA", *EUROSPEECH '93*, 1993.

[3] 森元 逞: "自動翻訳電話の実現に向かって", 情報処理学会誌, Vol.35 No.1, pp.1-10, 1994.

[4] 鈴木・井ノ上・谷戸："日韓音声翻訳システムの試作", 人工知能学会, SIG-SLUD-9403-4, 1995.

[5] 鈴木・井ノ上・谷戸:"タスク環境を考慮した日韓自動通訳システムのインタフェース改良", 信学技報, NLC95-29, 1995.

[6] 住吉・相沢："英語固有名詞の片カナ変換", 情報処理学会論文誌, Vol.35 No.1, pp.35-45. 1994.