

### 自動翻訳電話の処理システム構成とトラヒック特性に関する検討

吉田, 裕 / HAMAGUCHI, Jun / YOSHIDA, Yutaka / 浜口, 淳

---

(出版者 / Publisher)

法政大学計算センター

(雑誌名 / Journal or Publication Title)

Bulletin of Computer Center, Hosei University / 法政大学計算センター研究報告

(巻 / Volume)

8

(開始ページ / Start Page)

89

(終了ページ / End Page)

95

(発行年 / Year)

1995-03-31

(URL)

<https://doi.org/10.15002/00024702>

# 自動翻訳電話の処理システム構成と トラヒック特性に関する検討

浜口 淳

法政大学大学院工学研究科<sup>†</sup>

吉田 裕

法政大学工学部電子情報学科<sup>†</sup>

自動翻訳装置や音声翻訳システムの要素技術に関して現在様々な研究機関で盛んに研究されている。しかしながら、それらの要素技術を統合し、自動翻訳電話システムとして、ネットワークサービスを提供する場合のシステム構成については、あまり検討されていない。ここでは、種々のシステム構成を想定し、多量の処理を行う部分の共用化によりコスト削減の可能性を示し、各構成に対するトラヒック特性を計算式及びシミュレーションによって示すとともに、音声入力あたり処理不能率および処理時間の削減が今後重要であることを示す。

## 1. まえがき

自動翻訳電話システムとは、従来の機械翻訳と異なり、話し言葉(spoken languages)を扱い、機械翻訳の前後に音声認識と音声合成等の音声言語処理を行い、電話を通じて双方向で異なる言語間の会話を可能にするシステムである<sup>(1)(2)</sup>。このシステムは、言語の障壁を克服する新技術として、将来の期待が大きく<sup>(3)</sup>、ATR自動翻訳電話研究所、ATR音声翻訳通信研究所などで、盛んに研究されている。自動翻訳電話関連の研究は、10年余りを経過し、その要素技術は着々と進歩し、93年1月の国際通信実験<sup>(4)</sup>に見られるように、実用に近づいている。しかし、要素技術を総合し、通信網内においてサービスするためのシステム構成に関しては、ほとんど検討されていない。ここでは、自動翻訳電話システムを網内配置する場合のシステム構成について検討し<sup>(5)(6)</sup>、各構成に対するトラヒック特性を計算式及びシミュレーション実験結果により示すとともに、音声入力あたりの処理不能率及び処理時間の削減が今後重要であることを示す。

## 2. ネットワークサービスの処理システム

### 2.1. システム構成

自動翻訳電話システムを通信網内に設置し、ネットワークサービスを提供する場合、ネットワークユーザが利用する電話端末に対向するSLI(Spoken Language Interfacer)、翻訳などの処理を行う

ネットワークサーバ(Network Server)(以降NSと略す)と各言語に対応した辞書機能をNSに対して提供する、ディクショナリサーバ(Dictionary Server)(以降DSと略す)とにシステムをブロック分けすることができる。ここで、SLIとNS、DSを一箇所に集中配置して管理する集中配置構成(図1. a)と、NS、DS及びSLIの機能の一部を各ユーザ端末又は各国(各言語系)毎に分散して配置する分散配置構成(図1. b)が考えられる。

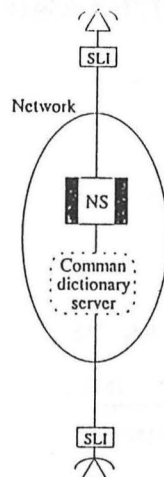


図1.a 集中配置の例

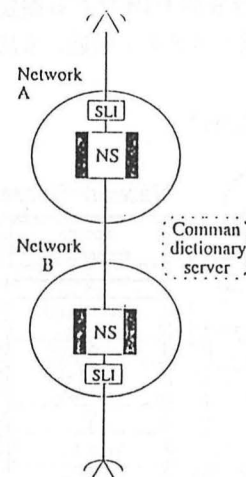


図1.b 分散配置の例

分散配置では、NS間で情報交換するか否か、分散NSにどの程度の機能を持たせるか、又はDSを共用するか否かなどで、様々な構成が考えられる(表1)。またSLIとNSが一对一のくくりつけ形式と、複数のNSを共同利用する形式が想定される。前者では1台のサーバでユーザの会話を開始から終了まで処理することから、適切な翻訳処理には

<sup>†</sup>〒184 東京都小金井市梶野町3-7-2

欠かせない会話の履歴などの情報を管理しやすい。これに対して後者では、会話の履歴を管理する昨日を別途設けなければならない。また、分散配置では、サーバ同士が通信を行い、会話の履歴等を管理する必要があり、より複雑化すると思われる。

表1. 分散配置構成の例

| 形態                 | 各SLIにNS(DS)を<br>設置<br>NS間通信あり | 各言語にNS(DS)を<br>設置<br>NS間通信あり | 各SLIにNS(DS)を<br>設置<br>NS間通信なし | 各言語にNS(DS)を<br>設置<br>NS間通信なし |
|--------------------|-------------------------------|------------------------------|-------------------------------|------------------------------|
| 通信中の番号             | 中間言語                          | 中間言語                         | 音声、テキスト                       | 音声、テキスト                      |
| 会話の履歴              | 容易                            | 容易                           | 難解                            | 難解                           |
| エラー処理              | 複雑                            | 複雑                           | 容易                            | 容易                           |
| 逐次認識               | 可能                            | 不可                           | 可能                            | 不可                           |
| DSの供用              | 可能                            | 一部可能                         | 不可                            | 一部可能                         |
| NS内辞書の<br>必要サイズ    | 小                             | 中                            | 大                             | やや大                          |
| 想定コスト              | 小                             | 中                            | 大                             | やや大                          |
| 想定応答速度             | 低速                            | やや低速                         | 高速                            | やや高速                         |
| 想定最大負荷<br>(除:翻訳時間) | 辞書検索時間<br>エラー処理               | NS待ち時間<br>辞書検索時間<br>エラー処理    |                               | NS待ち時間                       |

## 2.2. SLI及びNSの処理

SLI及びNSの機能は音声入力、音声認識、機械翻訳（語翻訳及び言語生成）、音声合成などからなる（図2）。ただし音声合成をNSで行い、SLIに合成音を出力する機能のみを持たせた場合（※実験システム：後述）も想定する。

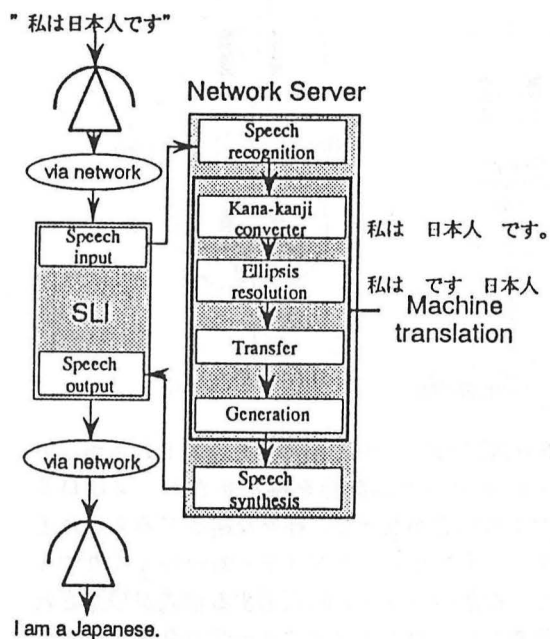


図2. 自動翻訳電話の処理内容

## 2.3. 会話のパターン

自動翻訳電話システム経由の会話パターンは、一方の話者の発声をSLIが聴き取り（SLIに音声データを蓄積）、ついでNSが音声認識、言語解析、機械翻訳等の処理を行なう。その後、NSがSLIが音声合成（speech synthesis）を行い、SLIは合成を相手話者に向けて出力する。従って、思考から相手話者が聴き取り終わるまでの1つの会話片（interaction fragment）の時間長は、対面のあるいは電話による直接の会話パターンと比べ、NS、DSによる処理と音声合成、合成音の出力に要する時間の和だけ増加する（図3）。

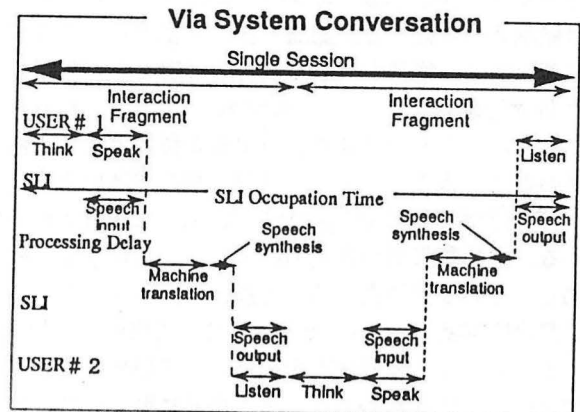
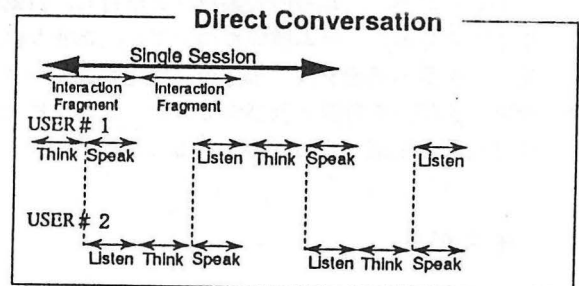


図3. 会話モデル

入力内容が複雑あるいは不適切な場合、NSが結果を出力できないことがあり、処理不能（interrupt）となる。また、音声入力が高すぎる場合も処理中断する。これを入力タイムアウトと呼ぶ。処理不能又は入力タイムアウトの場合には、システムメッセージを話者へ出力し再度入力させ、会話を継続する再送方式（resume scheme）を想定する。通信開始から終了までの処理フローを図4に示す。

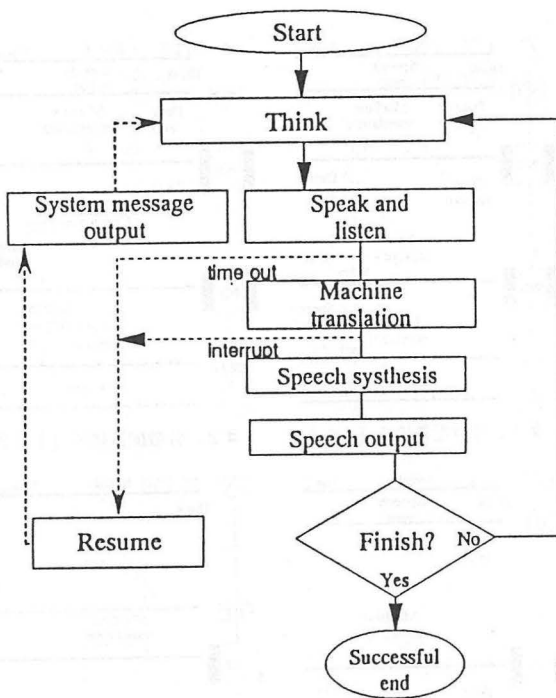


図4. 処理フロー

### 3. システム構成のモデルと平均系内時間

#### 3.1. システムパラメータの仮定

集中配置と分散配置とでは、システム構成、サービス時間が大きく異なると思われる。これらをモデル化し、上述の会話パターンを用い、平均通信時間を算出するために各種のパラメータを設定する。パラメータは、実験システム（※SL-TRANS: ATR 自動翻訳電話研究所における国際実験システム）の値を参考にして、以下のように仮定する（表2）。

- (1) 一会話片の終了後、次の会話片も続く確率を、互いに独立で同一の  $\beta = 0.9$  とする。
- (2) 各思考時間、各発声時間、各合成音送信時間は独立な指数分布で同一の平均値を仮定し、これを単位時間に取る。
- (3) 1回あたりのNS処理時間は、一定値  $T$  とする。標準を5とし、将来の処理速度向上を考慮して、短い場合も想定する。
- (4) NS間で情報交換する場合の処理時間の延長されるが、これを  $D_1$  とする。
- (5) DSからの情報が必要な場合の延長時間を一定の  $D_2$  とする。

- (6) SLIとNS間および、SLIとユーザ端末間が長距離の場合の伝送遅延時間を一定の  $D_T$  とする。
- (7) 各音声入力のリタイムアウト（上限）を5とする。タイムアウトになる標準の確率は、0.0067となる。
- (8) 音声入力に対する、NSの処理不能率（interrupt rate）は互いに独立で一定の確率  $\rho$ （標準0.1）とする。
- (9) 処理不能の場合のシステムメッセージ送出時間を一定値0.5とする。

表2. 数値解析に用いた記号

| 意味                | 記号             | 標準値    | 分布      |
|-------------------|----------------|--------|---------|
| 思考・発声・SLIの音声出力時間  | $Th, Sp, Slsp$ | 1.0    | 指数分布    |
| 音声入力のリタイムアウト      | $To$           | 5.0    | 一定      |
| 音声入力のリタイムアウト率     | $\tau$         | 0.0067 | 一定      |
| 会話の継続率            | $\beta$        | 0.9    | 一定      |
| 会話片の繰り返し回数        | $Nc$           | 10     | ベルヌーイ分布 |
| NSの処理時間           | $T$            | 5      | 一定      |
| 音声入力あたりのNSの処理不能率  | $\rho$         | 0.1    | 一定      |
| システムメッセージ送出時間     | $Dm$           | 0.5    | 一定      |
| 長距離の場合の伝送遅延時間     | $D_T$          | 0.1    | 一定      |
| NS ↔ NS間通信による延長時間 | $D_1$          | $0.5T$ | $T$ に依存 |
| NS ↔ DS間通信による延長時間 | $D_2$          | $0.5T$ | $T$ に依存 |

#### 3.2. システム構成と平均系内時間

分散配置と集中配置の想定されるシステム構成（図5）を示し、平均系内時間を算出する。

A：分散配置のNSに小辞書、SLIが端末側にある構成（#1）。タイムアウトの場合の延長時間

$$t_{TO} = 6 + D_m$$

非タイムアウト及び処理不能の場合の延長時間

$$t_{NTO} = \frac{1-6\tau}{1-\tau} + T + 2D_T + D_m + 1$$

非タイムアウト及び処理可能の場合の時間

$$t_S = \frac{1-6\tau}{1-\tau} + 3D_T + T + 2$$

B：分散配置のNSに小辞書、SLIが端末側にあり、NS間で情報交換を行う場合の構成（#2）。

$$t_{TO} = 6 + D_m$$

$$t_{NTO} = \frac{1-6\tau}{1-\tau} + T + 2D_T + D_1 + D_m + 1$$

$$t_S = \frac{1-6\tau}{1-\tau} + 3D_T + T + D_1 + 2$$

C : 集中配置のNSに小辞書, SLIが端末側にある構成 (#3) .

$$t_{TO} = 5 + D_m + 1 = 6 + D_m$$

$$t_{NTO} = \frac{1-6\tau}{1-\tau} + T + 2D_T + D_m + 1$$

$$t_S = \frac{1-6\tau}{1-\tau} + 2D_T + T + 2$$

D : 集中配置のNSに小辞書があり, SLIがNS側にある構成 (#4) .

$$t_{TO} = 5 + D_m + 1 = 6 + D_m$$

$$t_{NTO} = \frac{1-6\tau}{1-\tau} + T + D_m + 1$$

$$t_S = \frac{1-6\tau}{1-\tau} + T + 2$$

E : 上述A, B, C, Dの条件に加えDSの辞書を利用する構成は処理時間がだけ間延びする.

ここで, 各構成に共通の平均サービス時間を導出する. 1 会話片あたりの処理時間は音声入力時間や内容によって, 以下の3パターンをとる. 各パターンとその確率を示す.

1 : タイムアウト  $\tau$

2 : 非タイムアウト及び処理不能  $(1-\tau)\rho$

3 : 非タイムアウト及び処理可能  $(1-\tau)(1-\rho)$

音声入力失敗時の延長時間は,

$$t_{US} = \frac{\tau t_{TO} + \rho(1-\tau)t_{NTO}}{1 - (1-\tau)(1-\rho)},$$

音声入力失敗が回連続して起こる確率は,

$$R_n = \{1 - (1-\tau)(1-\rho)\}^n \quad n \geq 1,$$

音声入力失敗の平均繰り返し回数は,

$$\begin{aligned} \bar{R} &= \sum_{n=1}^{\infty} n R_n = \sum_{n=1}^{\infty} n \{1 - (1-\tau)(1-\rho)\}^n \\ &= \frac{1 - (1-\tau)(1-\rho)}{\{(1-\tau)(1-\rho)\}^2}, \end{aligned}$$

となるので, 1つの会話片の音声入力成功するまでの平均時間は, 次式となる

$$t_R = \{1 - (1-\tau)(1-\rho)\} \bar{R} t_{US} + t_S$$

従ってサーバ待ち時間を除く, 平均系内時間は,

$$\begin{aligned} D_R &= N_C t_R = \frac{\{1 - (1-\tau)(1-\rho)\} \bar{R} t_{US} + t_S}{1-\beta} \\ &= \frac{\{1 - (1-\tau)(1-\rho)\} \{\tau t_{TO} + \rho(1-\tau)t_{NTO}\} + t_S}{(1-\beta)\{(1-\tau)(1-\rho)\}^2} \end{aligned}$$

となる.

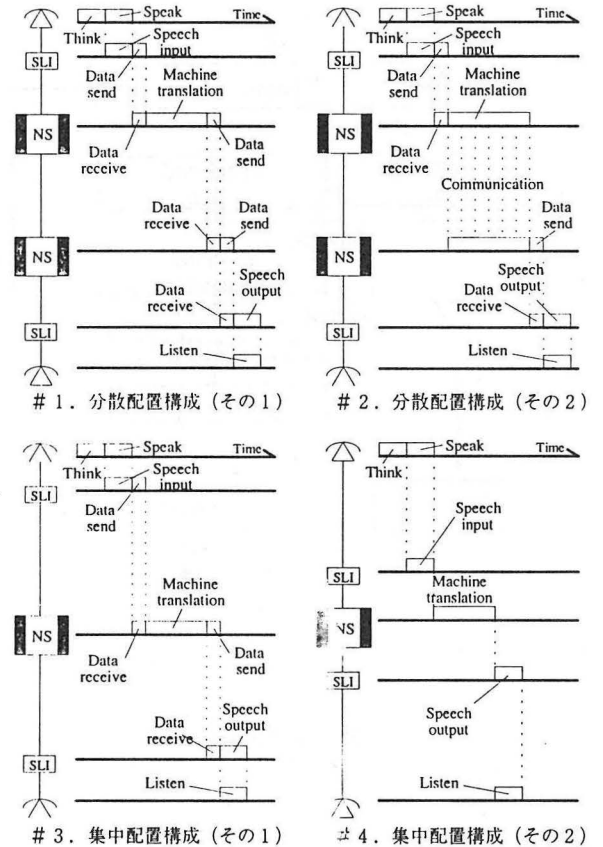


図5. 分散配置と集中配置のシステム構成

#### 4. 集中配置の場合のシステム内設備数

ここではSLI台数をN, NS台数をMとし (SLIとNSが一对一のくくりつけ形式の場合,  $N=M$ となる), 集中配置でNSを共同利用する場合について, 一定のNに対してMの大きさがどの程度になるかを検討する. そのためにシステムに対する負荷が最大になるように系内に常にN呼が存在するように設定しシミュレーション解析を行う.

##### 4.1. SLIの設備台数

SLIの必要設備台数については, 構成に関らず, 呼の継続期間中回線に接続し続けなければならないことから, サービス要求呼の発生率, 呼の平均継続時間及び基準呼損率を与えることにより, アーランB式を用いて算出可能である.

##### 4.2. NSの設備数

NSの設備数についてSLIとNSが一对一のくくりつけ形式と, 通信中の呼が必要に応じて待ち行列を形成し, 複数のNSのうち, 空いたものを利用

する形式が想定される。ここでは後者について、図6に示す閉待ち行列モデルを用いてシミュレーション解析する。ここでは  $N=20$  とした。

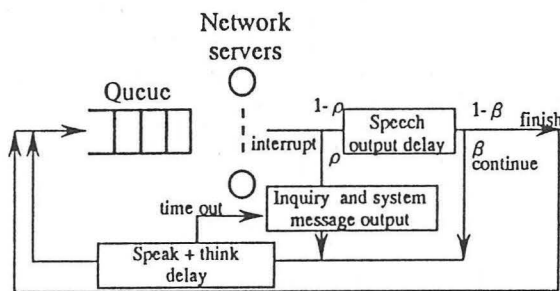


図6. 閉待ち行列モデル

#### 4.3. シミュレーション解析

図7にくくりつけ形式の処理不能率に対する平均系内時間を示し、図8に前節の仮定で処理不能率0.1の場合のNS台数に対する平均系内時間を示す。

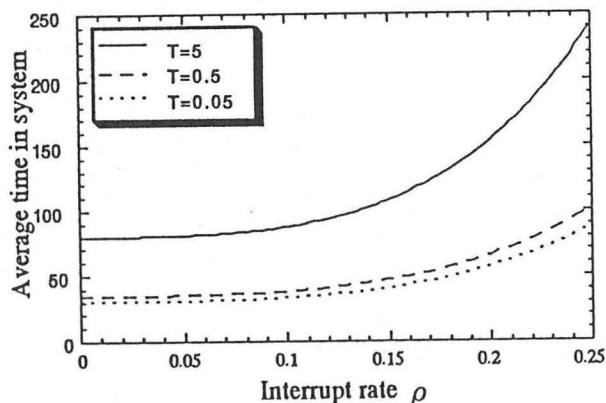


図7. 集中配置の場合の処理不能率と平均系内時間

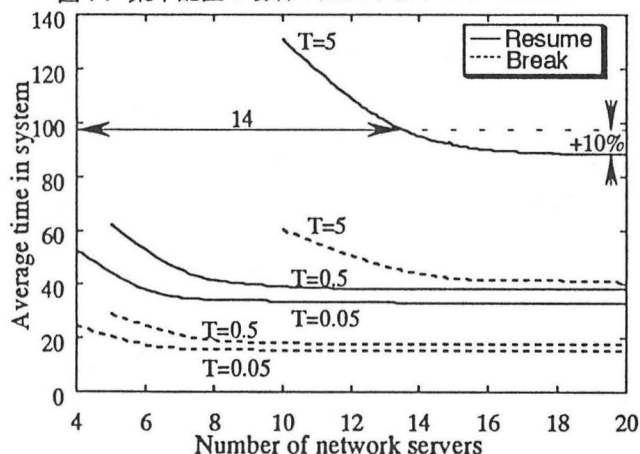


図8. 平均系内時間

これらのシミュレーション結果から以下のことが明らかとなった。

(1) 現在の研究レベルでの会話片あたり処理不能率は、0.05～0.22であり、0.1と設定した場合は、呼の半数は、処理不能に陥り、再送などのオーバー

ヘッドにより系内時間が増加していることになる。

(2) 再送する場合、システムメッセージ送信及び音声入力のくり返しに要する延長時間のため、再送しない場合の約2倍に増加する。

(3) NSの性能が向上し、処理時間が短くなると、NSの処理時間および待ち時間が減少するため平均系内時間が減少するが、 $T=0.5$ 以下に減少しても効果が少なくなる。

(4) NS台数MをSLI台数Nより少なくすると、一対一のくくりつけ形式より、平均系内時間はNS処理待ち時間分だけ増加する。しかし図7から見られるように、平均系内時間を1割増まで許容できるとすれば、NS台数を  $M = 0.7N$  台以下に抑えられ、コスト削減が期待できる。

#### 5. 結論

自動翻訳電話をネットワークサービスとして提供するために、網内に配置する種々のシステム構成について検討した。シミュレーションの結果から、次のことが明らかとなった。

(1) 待ち時間の増加をある程度許容できるのであれば、多量の処理を行うサーバを共用することにより、サーバの所要設備数を減らすことができる。

(2) 会話片当りの処理不能率が0.1の場合、処理不能に陥る呼の割合は約0.5とかなり大きくなり、平均系内時間が処理不能率に伴って急激に増加する。

分散配置の場合は、更にNSによる処理時間が延びる。従って、実用的に供するためには、サーバの処理時間をさらに短縮すること、処理不能率をさらに下げることが重要となる。これには、プロセッサの高速化、自由発話音声認識方式の開発、用例ベース翻訳などによる言語翻訳の高速化、などの対策が有効と思われる。

#### 参考文献

- 1) 樽松明, "自動翻訳電話のための音声処理と言語処理" 信学誌, 75, 10, pp.1050-1057 (1992-10).
- 2) 樽松明, "自動翻訳電話の基礎研究...7年間の区切りとこれからの期待" ATR ジャーナル, No.13, pp.4-5 (1993-05).
- 3) Kurematsu, A., "Future Perspective of Automatic Telephone Interpretation" IEICE Trans. Commun., E75-B (1992-01).
- 4) 谷戸文廣, 竹沢寿幸, "音声翻訳システム ASURA と自動翻訳電話国際共同実験" ATR ジャーナル,

No.13,pp.14-19(1993-05).

- 5) 浜口, 吉田, 谷戸, ” 自動翻訳電話システムの構成  
に関する検討” 信学春季全大, B-734, p.3.385(1993).
- 6) 浜口, 吉田, 谷戸, ” 自動翻訳電話の処理システム  
構成について” 信学論 (B-I), J77-B-I, 7, pp.483-  
486(1994).



## キーワード

自動翻訳電話, ネットワーク, SLI, ネットワークサーバ, 集中配置, 分散配置  
.....

## Summary

### A Study on Configurations and Traffic Characteristics of Automatic Interpreting Telephony System

Jun HAMAGUCHI\*

\* Department of Electrical Engineering, Hosei University<sup>†</sup>

Yutaka YOSHIDA\*\*

\*\* Department of Electronic Informatics, Hosei University<sup>†</sup>

Automatic interpreting telephony system is under research and its component techniques are evolving greatly, recently. Its final objective might be provision with service facility in networks. However, its system configuration as network facilities has never been investigated yet. This paper studies on both centralized and decentralized configurations of functions including spoken language interfacers, network servers, and dictionary servers. Especially, extension of call duration time due to via network facilities is taken into account and it is concluded that network servers which are required to have very large processing power should be used commonly for economy.

## Key Words

Automatic interpreting telephony system, network service, SLI, network server, centralized configuration, decentralized configuration

---

<sup>†</sup>3-7-2, Kajino-cho, Koganei-shi, Tokyo 184, Japan