音声対話システムにおけるスケーラビリティ評価モデルの検討

荒金陽助[†] 下川清志[††] 金井敦[†]

モバイル環境において，情報検索など利用者の目的を達成するためのインタフェースとして音声対話システムが注目されている．本論文では，サーバ処理能力と完全重複ユーザ数を求めることで，サーバ認識型音声対話システムのシステムスケーラビリティを推定する手法を提案する．サーバ処理能力は t 検定によって検出される応答間隔の飽和および，一定割合の遅延時間が要求値を超えないことの双方を満たすように算出される．一方，完全重複ユーザ数は，対象とするサービスの音声対話におけるユーザの発話間隔および想定する同時接続ユーザ数より算出される．そして，サーバの応答遅延時間による補正発話間隔を算出し，サーバ処理能力とによってサービスに必要なサーバ台数であるスケーラビリティが推定される．提案モデルの適用可能性を検証するために，“レストラン検索” および “駐車場検索” をサービス例とした実験評価を行い，サービス要件に応じたシステムスケーラビリティ推定の可能性を示した．

A Study for a Scalability Evaluation Model of Spoken Dialogue System

Yosuke Aragane,[†] Kiyoshi Shimokawa[††] and Atsushi Kanai[†]

In mobile environment, spoken dialogue system is suitable for input system to solve user needs such as information retrieval services. This paper proposed a system scalability evaluation method for server recognition type of spoken dialogue system. To estimate the system scalability, we focus our attention on server processing performances and completed overlapping user number. To estimate the server processing performances, we use the t-test value of server response interval saturation and the response delay time threshold. On the other hand, the completed overlapping user number is estimated by the utterance interval of target spoken dialogue services and suppositious overlapping connection user number. The utterance interval is revised by server response delay. Using these parameters, the system scalability of necessary server number for the service is estimated. To verify the usability of proposed method, we make experimental evaluation for example services, a restaurant search service and a parking search service. As a result of experimental evaluation, we indicate the usability of our method to estimate the scalability depend on the service requirement.

1. はじめに

インタフェースとしての音声は，入出力にともなう注視や手操作を必須としないことから，複合作業下のインタフェースとしての有効性が指摘されており[1)]，カーナビゲーションシステムをはじめとしたモバイル端末に利用されている[2)~5)]．しかしながら入力インタフェースとしての音声の課題として，選択肢である認識語彙をユーザに伝えることの困難性があげられており，インタラクションによってユーザ発話の誘導を行う音声対話が注目されている[6)]．

モバイル環境において音声対話インタフェースを実現させるためのシステム構成は，最も処理の重い音声認識エンジンを実装する位置によって，大きく分けてVoiceXML に代表される端末認識型と CTI システムに代表されるサーバ認識型が存在する．サーバ認識型では，潤沢な CPU パワーやメモリ容量を利用可能であることから大語彙認識における高認識率や多くのユーザ数に対する処理が期待できるとともに，Web 上の情報[7)] など他システムとの連携が容易であるといった長所が存在する一方，通信費が必要であることや通信遅延が不可避であるといった短所が存在する．また端末認識型では，限られた CPU パワーやメモリ容量による認識語彙数の制限や，音声認識に関わる他システムとの連携が困難であるなどの短所がある一方，音声認識時の通信料が不要であることや，高速な反応が可能であるといった長所が存在する．したがって，対

† 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories, NTT Corporation
†† NTT アドバンステクノロジ株式会社
NTT Advanced Technology Corporation

象とするサービスによって適切な方式を選択する必要がある．ところで，携帯端末は組み込み機器であることが多く，その性能やリソースには一定の制限が存在するため，高度な音声対話を実現することが困難である．しかしながら，20 程度の語彙の単語認識でありながらも低廉な通信料が可能となるために端末認識型を利用するサービスに優位性が存在した[2]．ところが，第三世代無線通信網の定額化の実現により，サーバ認識型の通信料に関わる短所が消失しつつあり，大規模ユーザ数への対応が可能であることや柔軟な音声対話が可能という長所がクローズアップされてきている．

音声対話については，システム挙動によるユーザの発話回数や発話時間の検討[8] など様々な研究がなされているが，昨今は，対話システムにどのような振舞いをさせるかという中身の問題よりは，それをどう実装するのか，といった問題が議論されており[9]，たとえばシステムのモジュール化の検討[10] や，対話記述方式の検討[11),12]，システムアーキテクチャの検討[13),14)] などがなされている．

一方，大規模ユーザに対応したサーバシステムの構築を検討する場合には，複数台のサーバを用いたサーバ間負荷分散による処理能力の向上や故障回避が行われることが多い[15),16)]．このような負荷分散システムを検討するためには，(1) サービス性を考慮して負荷分散するサーバ台数などのシステム規模を推定し，(2) サービス性やシステム規模などに応じたサーバ間負荷分散手法を検討し，(3) 運用開始後に需要が増大することを視野に入れて負荷検知手法およびそれに応じたサーバ増設手法の検討を行うことが必要である．(2) の負荷分散手法については，多数のサービス要求を複数のサーバに効率良く割り振ることを目的として多くの先行研究がなされており，DNS ラウンドロビンや最小負荷サーバ選択法，コンテンツ識別アドレスによる選択法，ハッシュ&スライドによる選択法など多数の手法が提案されている[17),18)]．また，サービス開始後のサーバ増設に関する (3) についても様々な研究がなされており，ミッションクリティカルなシステムなどにおいて CPU 稼働率やネットワーク帯域に応じて短時間にサーバの増設・削減が可能な手法などが提案されている[19)~21)]．

しかしながら，サービス開始時のイニシャルコスト算出に必要となる (1) については，“新規サイト開設時にどの程度の負荷が発生するかは予測困難である” といわれており[19]，Web サーバなどの負荷分散評価におけるシミュレーション検討はなされているが[22),23)]，サーバ型音声対話システムにおける検討はほとんどなされていない．菊池らは処理時間などのシステム処理性能に着目した研究を行っているが[24]，その目的は効率的な対話戦略の選択であり，サービスの要求を満たすために音声対話サーバが何台必要であるかなどのシステムスケーラビリティについては言及されていない．また田熊らによる並列処理型計算機を用いて効率的なシステム構成の検討が行われているが[25]，サービス全体としてのスケーラビリティについては議論されていない．現在のサービスの現場においては，音声対話サーバの負荷が重く複雑なこともあり，メモリ消費量や CPU 稼働率などのシステム稼働率が 80%を超えた場合に増設を行う，といった対症療法が主流である．

そこで本論文では，カーナビゲーション装置などのモバイル端末を対象とした，数百レベルの大規模ユーザへのサーバ認識型音声対話サービスを実現するシステムに対する，サービス導入時のスケーラビリティ評価モデルを提案することを目的とする．以下，2 章では本論文で議論する音声対話およびそのシステム，スケーラビリティの定義などについて説明し，3 章でサーバ認識型音声対話システムスケーラビリティ評価モデルを提案する．4 章では，特にカーナビゲーション装置でのサービスとして採用例の多い[3),4),26)]，レストラン検索および駐車場検索を用いて提案モデルの実現性検証を行い，5 章で本論文をまとめる．

2. 音声対話システムのスケーラビリティ

本章では，本論文で議論する音声対話の対話フローおよび音声対話システムについて説明するとともに，議論対象とするスケーラビリティを定義し，スケーラビリティに影響を与える因子を示す．

2.1 音声対話フロー

音声対話はあるタスクを実現するために行われる．タスクとはレストラン検索や時刻表検索など，ユーザの求める情報を，ユーザの示す前提条件をもとに検索する形をとることが多い．音声対話は，1 対 1 の対話とグループでの対話，および一方が他方に仕事を依頼する司令型対話と対話を継続することそのものが目的となる相互交流型の対話に分類できるが[9]，本論文では，モバイル環境を対象とすることから，ユーザがシステムに対して情報検索などを行う “1 対 1 の対話” かつ “指令型対話” を対象として議論する．この場合，システムは情報を検索するのに十分な条件をユーザ発話より取得する必要があり，このために音声対話によって必要な情報をユーザから引き出していく．条件項目を並べ，ユーザに入力してもらう形態が最も単純であるが，最近の研究では “ユーザにとって自明の内容は

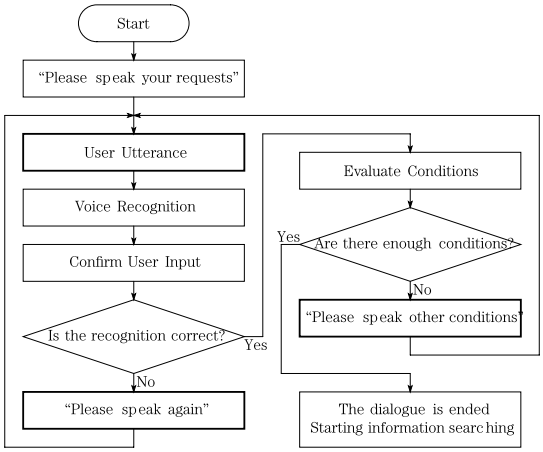

図 1 検索型音声認識の典型的な対話フロー
Fig. 1 Typical dialogue flow for information retrieval type of spoken dialogue system.

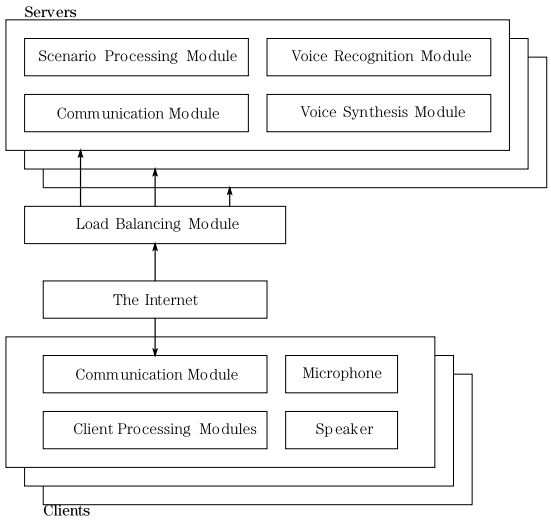

図 2 対象とするシステムの構成
Fig. 2 An architecture of target systems.

可能な限り尋ねないようにする” などのインテリジェントなシステムも提案されている．また，音声対話のフローを司るシナリオ構築を容易にするための研究もなされている[12]．

ここで検索型音声対話の典型的な対話フローを**図 1** に示す．システムはユーザ発話の内容に従ってレスポンス（多くは質問と確認）の内容を変化させ，少ないやりとりで条件を聞き出そうとする．100%の音声認識率を確保することが困難なため，ユーザ発話に対する認識結果については，逐次ユーザに確認することで認識結果を確定するフローとなっている．

2.2 システム構成

本論文では，**図 2** に示すような構成を持つシステムを議論の対象として考える．図 2 では音声データの伝達方法を，一般的な CTI（Computer Telephony Integration）で主流となっている音声回線ではなく，データ回線（インターネット経由）としている．これは，(1) CTI においても VoIP を利用した効率的なシステム構築が始まっていること[27]，(2) カーナビゲーションシステムなどのモバイル環境においては音声だけではなく，検索結果などのデジタルデータをあわせて扱うマルチモーダルインタフェースが重要であることによる[28]，電話回線とデータ回線の 2 回線を必須とする不効率なシステム構成を避けるためである．以下では図 2 に示すシステム動作の流れと構成要素を説明する．

ユーザの発話は通信によって負荷分散装置に伝達され，負荷分散装置は複数のサーバから適切なものを選びユーザの処理を割り振る．ユーザの音声が割り当てられたサーバで音声認識された時点でサービスが特定され，音声対話が開始される．サーバでは，サービスの音声対話シナリオに従って，ユーザの意図（たとえば検索したいレストランの種別や場所）を取得するために，ユーザの発話を引き出す質問を音声合成を用いてユーザに通知する．サービスが完了するために十分な情報が入力された段階で音声対話を終了する．

次に，システムの各構成要素について説明する．

Scenario Processing Module サービスのシナリオに従って，音声認識結果に対する出力（音声合成など）を決定し，音声合成処理などを利用してユーザに通知する．

Voice Recognition / Voice Synthesis Module 端末から送信されるユーザ発話を認識するとともに，Scenario Processing Module からの指示に従ってユーザに通知する音声の合成処理を行う．

Load Balancing Module 端末から送信されるユーザの要求を複数のサーバに割り振ることで，負荷分散を実現する．

Communication Module 端末側とサーバ側に設置され，主に音声（ユーザ発話および合成音声）のやりとりを行う．本論文では，無線通信メディアとしては，移動環境下での一定帯域確保という観点から第三世代携帯電話網を利用することとする．

Client Processing Module ユーザに対する入出力インタフェースを，マイクとスピーカを利用して制御する．

Microphone / Speaker ユーザに対するインタフェースとなる．

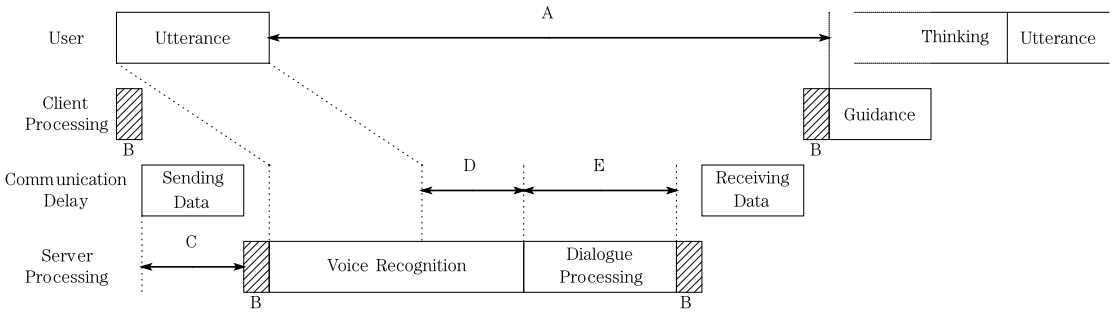

図 3 入力応答の流れ
Fig. 3 A sequence from user utterance to system response.

2.3 音声対話システムにおけるスケーラビリティ

本論文で議論するスケーラビリティとは，**“ある音声対話サービスに関して，x 人のユーザのアクセスを処理可能なサーバ台数 $= y$ 台を求めること”** と定義する．音声対話では，ユーザの意図を取得するために “1 回以上のユーザの音声入力とそれに対するシステム応答” を繰り返す．本論文では，音声対話が開始されてからシステムがユーザの意図を取得して音声対話が終了するまでをタスク，1 回の “ユーザの音声入力とそれに対するシステム応答” を入力応答と呼ぶこととする．

音声対話システムは電話の呼処理[29] などと比較して，図 1 に示すように，音声認識・合成や音声対話処理など 1 タスクの処理に必要となるシステム負荷が比較的大きいことに特徴がある．音声対話サービスを提供する際には需要に応じたシステムを構築することが必要であるが，これらの対話処理中の負荷も考慮したシステムスケーラビリティ評価が重要である．

ここで，本論文が対象とする音声対話システムの入力応答における時間に沿った処理の流れを**図 3** に示す．

ユーザ発話の開始と同時にその音声は逐次エンコードされ，パケットとして連続してサーバに送信される．サーバでは，受信したパケットから逐次音声認識処理を開始し，その音声認識結果と対話シナリオに従って，シナリオ遷移や音声合成などの対話処理を行う．合成された応答音声は端末に送信され，端末で再生されることでユーザに通知される．ユーザは，サーバからの応答音声に応じて次の発話内容を思考し，次の入力応答に遷移する．なお，ユーザ発話に対する音声認識結果の確認に関わるフローについても，「○×でよろしいですか？」というシステム応答であるガイダンスに対してユーザが思考し，「はい」または「いいえ」という発話を行う入力応答となる．したがって，このような音声認識結果の確認に関わるフローについても，図 3 に示す入力応答によってモデル化可能である．

2.4 スケーラビリティ決定要因

図 3 において，スケーラビリティ算出に影響を与えるサーバ処理が行われる部分は，音声のエンコード・デコードおよび音声認識処理，対話処理である．したがって，あるユーザの音声対話タスクを処理するすべての期間においてサーバ処理が必要ではなく，2 つの入力応答間の思考期間や送受信の通信遅延期間においてはそのユーザに関わる処理負荷を考慮する必要はない．

一般に，1 台のサーバが w 人の音声対話処理が可能なとき，x 人の音声対話処理を行うためには x/w 台のサーバが必要であると考えられる．しかしながら上述の理由により，1 台のサーバが同時処理可能な音声認識処理または対話処理の重複個数と，x 人が音声対話サービスにアクセスする際に音声認識処理または対話処理が完全に重なるユーザ数とを分けて検討する必要がある．

本論文では，前者をサーバ処理能力，後者を完全重複ユーザ数と呼ぶ．

3. スケーラビリティ評価モデル

本章では，サーバ処理能力および完全重複ユーザ数を推定し，それらとサービス用件から定義される許容値との関係からシステムスケーラビリティを評価するモデルを提案する．

3.1 サーバ処理能力推定手法

サーバ処理能力とは，完全に同時に処理可能な音声認識処理または対話処理である．しかしながら，ここでの “処理可能” であることの意味を定義する必要がある．システムの処理能力を表す最も重要な指標は，スループット（システムが単位時間あたりに処理することのできる処理要求件数）とターンアラウンドタイム（処理を要求してから処理結果が出てくるまでの時間）であるといわれている[30]．なお，図 2 に示すようなサーバ間負荷分散を行う場合には，負荷分散手法の

違いや，ユーザ発話の到着間隔などに応じて負荷分散処理自体がオーバヘッドとなり，スループットなどのシステムパフォーマンスに影響を与えるといわれている[17),31)]．しかしながら，負荷分散処理のオーバヘッドについては負荷分散手法の選択に影響を受けるため，簡単のため本論文では負荷分散処理のオーバヘッドを考慮しないこととする．

まずスループットについて考える．サーバでは有限の CPU 処理時間を各ユーザの処理に割り振ることで音声対話処理を実行する．サーバの処理能力に余裕がある状態では，到着するユーザ発話に対して即座に処理を開始できるため，到着するユーザ発話の頻度（同時処理ユーザ数）と応答処理数は比例関係にある．しかし，ユーザ発話の到着頻度がサーバの処理能力を超えた状態では，到着したユーザ発話は待ち行列に保持され，サーバが処理可能となった段階で逐次処理に移されることとなる．この段階では，到着するユーザ発話の頻度（同時処理ユーザ数）と関係なく，応答処理数はサーバの処理能力に応じて一定値ないしは減少すると考えられる．すなわち，N_C 台のクライアントからサーバに対して，思考や送受信の期間を除外した間断のない発話を繰り返し送信した際の，サーバの応答間隔 $I_R(N_C)$ と $I_R(N_C+1)$ との t 検定を行い，t の実現値 $t(N_C)$ を式 (1) によって算出する．そして，式 (2) に示される，有意水準 α%での有意な差が観測されなくなった最小のクライアント数 N_C が飽和クライアント数であり，それが 1 台のサーバで処理可能な完全に重複する入力応答数（N_S）であるといえる．なお，n_{N_C} はクライアント数 N_C における応答間隔計測データ数を示し，$s^2_{N_C}$ はクライアント数 N_C における応答間隔計測データの分散を示す．また，$t(N_C)$ は $I_R(N_C)$ と $I_R(N_C+1)$ との t 検定の結果として算出される t の実現値を示す．

$$J(N_C) = \sqrt{n_{N_C} s^2_{N_C} + n_{N_C+1} s^2_{N_C+1}}$$

$$K(N_C) = \frac{\overline{I_R(N_C)} - \overline{I_R(N_C+1)}}{J(N_C)}$$

$$L(N_C) = n_{N_C} n_{N_C+1} (n_{N_C} + n_{N_C+1} - 2)$$

$$t(N_C) = K(N_C) \sqrt{\frac{L(N_C)}{n_{N_C} + n_{N_C+1}}} \quad (1)$$

$$t(N_S) < t_\alpha \quad (2)$$

次にターンアラウンドタイムについて考える．音声対話はテキストベースのインタフェースと比較してインタラクション性が非常に高いため，ユーザの発話に対するシステム応答の遅延がサービスの快適性，ひいては有効性に大きな影響を与えることとなる．また，システムの状態をユーザに通知するという観点からも処理時間は重要なパラメータであるといわれている[24),32)]．ユーザが認識可能な応答遅延は，図 3 に示す発話が終了してからシステム応答が開始されるまでの時間 **A** である．許容できる応答遅延時間（T_M）はサービスによって異なるが，図 3 中の **B**〜**E** から構成される．

B：音声コーデック処理時間 図 3 で網掛けをした計 4 カ所に存在する．コーデックの種類や端末の処理能力に依存するが，一般的に微少な時間である．

C：送受信時間 無線通信とインターネットを経由して音声データを電送する際の遅延時間である．一般的に，無線通信区間の遅延と比較してインターネットのバックボーンにおける遅延は微少な時間となる．

D：音声認識遅延時間 音声認識処理におけるリアルタイム処理からの遅れ時間である．同時処理数に比例して増加する．

E：対話処理時間 音声認識結果に基づいて，次にユーザに通知するガイダンスを決定し，そのガイダンスを音声合成する時間である．

ユーザが認識可能な応答遅延についても，応答間隔と同様にサービスログにより計測可能である．ただし，応答遅延の分散を考慮して，N_C 台のクライアントからの発話要求に対して，遅延時間の平均値（$\overline{T(N_C)}$）ではなく，上側 r%の計測値が満足する遅延時間（$T_r(N_C)$）が許容遅延時間（T_M）を超えない N_D をターンアラウンドタイムの観点からのサーバの処理能力であると定義する．なお r については文献値を参考として，本論文では $r = 95\%$とする[33),34)]．

$$\overline{T(N_D)} < T_r(N_D) < T_M \quad (3)$$

N_D は，式 (2) の N_S をサービス性の観点から拘束する値であり，N_D と N_S のうち小さい値がサーバ処理能力を決定する同時処理数となる．

3.2 完全重複ユーザ数推定手法

本節では，同時にタスクを実行しているユーザ数と，ユーザの思考や送受信時間を除いたサーバ処理が完全に重複する処理数との関係について議論する．

ユーザがある音声対話サービスを開始してから完了するまでのタスク時間（T_T）中に n 回の入力応答を繰り返したとすれば，発話間隔（I_U）は $I_U = T_T/n$ と表すことができる．本論文では，ユーザ数が数百のサービスを対象として議論するため，サーバに到着するユーザの音声対話要求の間隔はある程度平均化されると考えられる．そこで簡単のため，完全重複ユーザ

数のモデル化にあたり，発話間隔は上側 95%の計測値ではなく平均値を用いるものとする．したがって，数人の被験者について平均を求めた発話間隔（$\overline{I_U}$）により，x 人のユーザがタスクを実行している場合に，すべての入力応答の発話間隔（I_{All}）は式 (4) で表される．

$$I_{All} = \frac{\overline{I_U}}{x} = \frac{\overline{T_T}}{\overline{n} \cdot x} \tag{4}$$

3.3 スケーラビリティ評価モデル

本節では，前節までの議論を受けてスケーラビリティを推定するモデルを考察する．ユーザがある発話を開始してから，次の発話を開始するまでの時間であるユーザの発話間隔（I_U）は，ユーザの観点から以下の 4 つに分類される．すなわち，ユーザが認識語彙を発声する “1) ユーザ発話時間”，音声認識処理時間や合成処理時間，通信遅延時間などが含まれる “2) システム応答待機時間（図 3 の **A**）”，システム応答がユーザに対して再生される “3) システム応答時間”，次の発話内容をユーザが思考する “4) ユーザ思考時間” である．これらの構成要素は時相的に重複しないとは限らず，2) システム応答待機時間や 3) システム応答時間において 4) ユーザが思考することも可能であり，構成要素の時相的重複が存在する．しかしながら，音声認識失敗時のシステム応答や検索完了時のシステム応答に対しては，ユーザはシステム応答を待って思考することが要求される．このような場合，3) システム応答時間と 4) ユーザ思考時間の重複は存在するが，2) システム応答待機時間と 4) ユーザ思考時間の重複は困難であると考えられる．そこで，本論文では簡単のため，ユーザは必ずシステム応答を待って思考を開始するとし，1) ユーザ発話時間，3) システム応答時間および 4) ユーザ思考時間は 2) システム応答待機時間に影響を受けないものとして考える．このような条件下においては，式 (2)，式 (3) に示されるサーバ処理能力限界である N_S 台，N_D 台のうち小さい方（拘束条件の厳しい方）を N_A 台とすれば，この N_A 台のクライアントからの要求を処理する状況では，式 (3) に示される遅延が発生する．したがって，式 (4) に示される各ユーザの発話間隔は，各入力応答に生じる遅延時間分増加することとなる．式 (4) に示される発話間隔は x 人のユーザに対する演算であるので，式 (3) に示される平均遅延時間（$\overline{T(N_A)}$）を考慮すると，補正発話間隔（I'_{All}）は式 (5) で表される．

$$I'_{All} = \frac{\overline{I_U} + \overline{T(N_A)}}{x} \tag{5}$$

ここで，N_S 台のクライアントからの要求を処理する際には，サーバの応答間隔は式 (2) に示す $I_R(N_S)$ となる．したがって，I'_{All} の間隔での要求に応答するために必要なサーバ台数 y は式 (6) によって推定されることになる．

$$y = \frac{I_R(N_S)}{I'_{All}} = \frac{x \cdot I_R(N_S)}{\overline{I_U} + \overline{T(N_A)}} \tag{6}$$

4. 計測による検証

本章では，特にカーナビゲーション装置でのサービスとして採用例の多い，レストラン検索および駐車場検索をサービス例として，提案モデルによる音声対話システムのスケーラビリティ推定の実現性検証を試みる．

4.1 計 測 条 件

以下に計測条件として，対象としたサービス内容およびサーバ処理能力計測と完全重複ユーザ数計測の条件と計測方法について示す．

4.1.1 対象とする音声対話サービス

対象サービスとする “レストラン検索” および “駐車場検索” においては，検索の基点となる場所の指定と検索条件の指定とともに，それら入力に対する確認（はい or いいえ）を行うものとする．場所の指定は “現在地” または “目的地” とし，検索条件については，“和食” や “イタリア料理” または “（基点から）500 m 以内” や “（1 時間あたり）400 円以内” といった条件を 1 つまたは 2 つ指定するものとする．また，プリフィックスやサフィックスについてもある程度対応することで，柔軟性のある音声対話を想定する．したがって，たとえば「えーと，ラーメンかイタリア料理が食べたい」や「800 m 以内で 300 円以下で探して」といった発話を受理可能とする．さらに，多くのナビゲーションシステムにおける音声入力と同様に，ユーザ発話に対する音声認識結果の確認も音声による「はい」/「いいえ」の発話によって入力を確定するものとする．

4.1.2 サーバの処理能力測定条件

音声認識・合成およびシナリオ処理を行う 1 台のサーバに対して，1〜8 台のクライアントから間断なく音声対話処理要求を行う．ここで，間断なくとは，あるクライアントは音声認識要求に対する応答（合成された音声）を受け取った次の瞬間に次の音声認識要求を送信する，という状況とする．本計測実験におけるシステムログから，サーバが受理したユーザ発話の頻度および入力応答遅延時間を取得する．ただし，無線伝送遅延については変動要素が大きいため，無線伝送遅延のみ別に計測するものとし，サーバの処理能力計測実験においては有線 LAN 接続（100Base/T）と

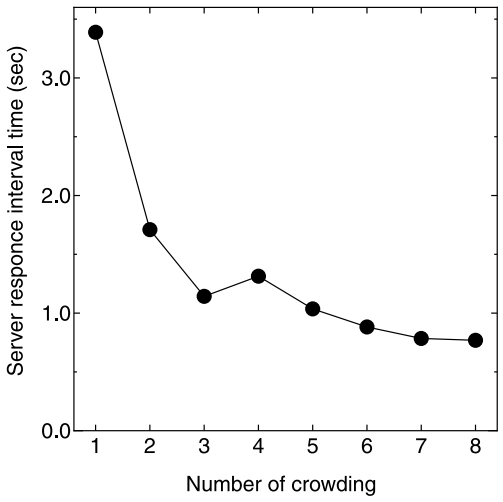

図 4 サーバ応答間隔の推移
Fig. 4 Server response interval time.

した．なお，本計測実験では Pentium 4 3.06 GHz の CPU を搭載し，1.0 GB のメモリを積んだ PC をサーバとして使用した．

4.1.3 完全重複ユーザ数測定条件

対象サービスとする “レストラン検索” および “駐車場検索” に対して，処理遅延が最小となる 1 クライアント接続条件で被験者が音声対話を行った際の発話間隔を計測する．被験者はシステム操作に習熟したユーザと簡単なインストラクションを受けただけの初心者ユーザそれぞれ 10 人の合わせて計 20 人とした．

4.2 計測結果

以下にサーバ処理能力および完全重複ユーザ数計測実験の結果を示し，提案モデルを適用したスケーラビリティ評価を行う．

4.2.1 サーバ処理能力測定結果

サーバが受理したユーザ発話間隔，すなわちサーバの応答間隔の測定結果を**図 4** に示す．

横軸に同時処理を行うクライアント数，縦軸にサーバ応答間隔の平均を秒で示している．サーバ応答間隔はクライアント数の増加とともに減少していき，クライアント数 7 のあたりで飽和していると読み取れる．このときのクライアント数増加にともなう t 検定の有意確率を**表 1** に示す．

有意確率はクライアント数に比例して増加していくが，クライアント数 7 と 8 の比較において初めて 5%を超えており，この 2 つのクライアント数における応答間隔に有意な差が存在せず，応答間隔が飽和していると考えられる．したがって，1 台のサーバが同時に処理可能なクライアント数（N_S）は 7 であり，そのときのサーバ応答間隔（$I_R(7)$）は図 4 から 770 msec であるといえる．

表 1 サーバ応答間隔の t 検定有意確率
Table 1 The t-test value of server response interval time.

client numbers of comparison	significance probability
1 and 2	approximately 0
2 and 3	2.492×10^{-104}
3 and 4	2.531×10^{-12}
4 and 5	5.904×10^{-34}
5 and 6	3.219×10^{-9}
6 and 7	5.648×10^{-5}
7 and 8	2.493×10^{-1}

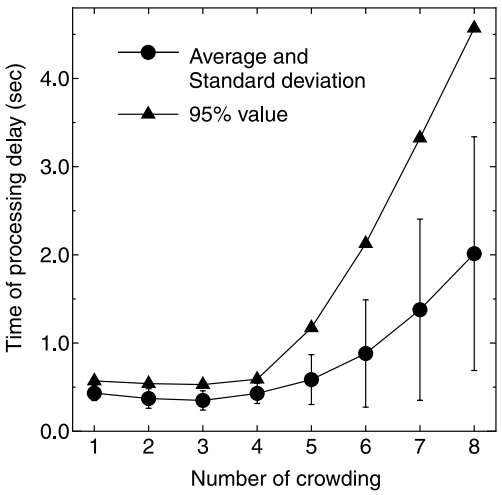

図 5 サーバ応答遅延の平均と 95%値の推移
Fig. 5 The average and 95% value of server response delay time.

一方，サーバ応答遅延の測定結果を**図 5** に示す．

図 5 では，横軸に同時処理を行うクライアント数，縦軸に遅延時間の平均値と標準偏差および 95%のデータを満たす遅延時間（$T_{0.95}(N_D)$）を示す．4 クライアント程度までの遅延時間は横ばいであるが，その後急激に立ち上がりサーバ処理能力の限界に近づいていることが示されていると考えられる．また，無線区間とインターネット区間を含めた伝送遅延を計測したところ，平均 380 msec，95%値で 450 msec であった．したがって，図 5 に示す遅延時間に往復の通信遅延としての 900 msec を加えた値が，図 3 に示すユーザが認識可能な遅延時間 **A** となる．許容可能な応答遅延時間（T_M）を約 4 秒と設定した場合，サービス性の観点から決定されるサーバが同時処理可能なクライアント数は 7 クライアントであるといえる．

表 2 被験者評価による発話間隔と発話時間割合
Table 2 The utterance interval and the utterance occupied ratio.

		レストラン検索	駐車場検索
タスク時間（sec）	T_T	99.6	99.9
発話回数（回）	n	6.43	6.41
発話間隔（sec）	I_U	15.48	15.58
発話時間割合（%）		8.23	11.0

4.2.2 完全重複ユーザ数測定結果とスケーラビリティ評価

被験者評価による発話間隔および発話時間割合の計測結果を**表 2** に示す．レストラン検索と駐車場検索の利用頻度がほぼ同程度であるとして，表 2 における 2 つのサービス例の発話間隔の平均である 15.53 秒および式 (4) より，x 人のユーザがタスクを実行している際の発話間隔は $15.53/x$ 秒となる．前節に示した $T_M = 4$ 秒の場合には，平均遅延時間は図 5 の 7 クライアントの平均値 1.38 秒と通信遅延の平均 760 msec（380 msec × 2）の和である 2.14 秒となる．そこで式 (5) に示される補正発話間隔（I'_{All}）は，式 (7) で表される．

$$I'_{All} = \frac{15.53 + 2.14}{x} = \frac{17.67}{x} \tag{7}$$

前節で計測結果から導いたように，同時処理可能な飽和クライアント数（$N_S = 7$）におけるサーバ応答間隔（$I_R(7)$）は 770 msec であるため，式 (6) を用いて $T_M = 4$ 秒とした場合の x 人のユーザからの入力応答を処理可能なサーバ台数（y）は式 (8) に示す値となる．

$$y = \frac{I_R(N_S)}{I'_{All}} = 4.36x \times 10^{-2} \tag{8}$$

したがって，$x = 100$ の場合は $y = 4.36$ であるので 5 台のサーバが必要となり，$x = 500$ の場合は $y = 21.79$ であるので 22 台のサーバが必要になると算出される．

5. ま と め

本論文では，モバイル環境における大規模ユーザ数を扱う音声対話サービスを対象として，システムスケーラビリティを推定するモデルを提案した．提案モデルは，スループットおよびターンアラウンドタイムから推定されるサーバ処理能力と，サービスにおけるユーザ発話間隔から推定される完全重複ユーザ数に着目することでスケーラビリティの評価を行う．サーバ処理能力は t 検定によって検出されるターンアラウンドタイムの飽和および，一定割合の遅延時間が要求値を超えないことの双方を満たすように算出される．一方，完全重複ユーザ数は，対象とするサービスの音声対話におけるユーザの発話間隔および想定する同時接続ユーザ数より算出される．そして，サーバの応答遅延時間による補正発話間隔を算出し，サーバ処理能力とによってサービスに必要なサーバ台数が算出される．最後に “レストラン検索” と “駐車場検索” というサービス例について提案モデルの適用可能性検証実験を行い，サービス要件に応じたシステムスケーラビリティの推定可能性を示した．

今後の課題としては，各種の負荷分散手法を適用した複数サーバによる実環境での負荷分散オーバヘッドを考慮した提案モデルの評価や，他のサービス例によるユーザ発話間隔の調査を行うことで，提案モデルのカバー範囲を明確化するとともに，より実サービスに近いフィールドにおける検証を行う必要がある．また，完全重複ユーザ数が数十のサービスにおける発話間隔や遅延時間のモデル化，各種のサーバ間負荷分散手法に対する適性などについても検討を深める必要がある．

参 考 文 献

1) 田中修一，中里　収，帆足啓一郎，白井克彦：複合作業下における音声インタフェースの有効性，人工知能学会言語・音声理解と対話処理研究会資料，Vol.14, pp.1–8 (1996).
2) internavi Premium Club. http://premium-club.jp/
3) G-BOOK. http://g-book.com/
4) カーウイングス. http://www.nissan-carwings.com/
5) 岩崎知弘，難波利行，石川　泰：カーナビゲーション用音声インタフェース技術，自動車技術，Vol.57, No.2, pp.65–70 (2003).
6) 速水　悟，菅村　昇：音声対話システムの研究と実用化の動向，日本音響学会誌，Vol.50, No.7, pp.574–580 (1994).
7) 神作洋一，藤江真也，宮永悠平，小林哲則：www を知識源とする音声対話システム，人工知能学会全国大会論文集，Vol.15, No.2, 3B3-02, pp.1–2 (2001).
8) 中川聖一，山本誠治：音声対話システムの構成法とユーザ発話の関係，電子情報通信学会論文誌，Vol.J79-D-II, No.12, pp.2139–2145 (1996).
9) 小林哲則：音声対話研究の現状と動向，人工知能学会誌，Vol.17, No.3, pp.266–270 (2002).
10) Polifroni, J., Seneff, S.: Galaxy-II as an Architecture for Spoken Dialogue Evaluation, *International Conference on Language Resources and Evaluation* (*LREC*) (2000).
11) 荒木雅弘：音声対話システムと VoiceXML，人

工知能学会言語・音声理解と対話処理研究会資料, Vol.34, pp.39–44 (2002).
12) 関口真理子，荒金陽助，阿部伸浩，下川清志：テレマティクス用音声対話システムのための 1 bit 状態遷移表シナリオ記述方式の提案と評価，情報処理学会論文誌次世代移動体通信システム特集号, Vol.45, No.12, pp.2720–2731 (2004).
13) 河口信夫，松原茂樹，長森 誠，稲垣康善：複数の音声対話システムの統合制御機構とその評価, 情報処理学会研究報告，2001-SLP-36, pp.63–70 (2001).
14) 大淵康成, 畑岡信夫, 赤堀一郎, 立石雅彦, Judy, S., Mitamura, T., Nyberg, E.: VoiceXML をベースにした頑強な音声対話管理アーキテクチャ, 情報処理学会研究報告，2003-SLP-46, pp.49–54 (2003).
15) 巳波弘佳, 熊谷和則, 能上慎也, 阿部威郎, 堀之内剛史：コンテンツ配信システムにおけるサーバ間の負荷分散制御法の検討，電子情報通信学会大会講演論文集，Vol.2001, B-7-217, p.350 (2001).
16) 宮田篤人, 石丸 浩, 大久保公博, 青山春巳, 村田達彦：クライアント・サーバシステムにおける危険分散・負荷分散方式，電子情報通信学会大会講演論文集，Vol.2003, B-6-112, p.112 (2003).
17) 巳波弘佳，熊谷和則，能上慎也，阿部威郎：負荷分散制御アルゴリズムの性能評価と適用領域，電子情報通信学会技術研究報告，Vol.101, No.8, NS2001 1-10, pp.15–20 (2001).
18) 佐竹伸介，稲井 寛：Web サーバシステムにおける古い負荷情報に基づく負荷分散方式, 電子情報通信学会技術研究報告，Vol.104, No.182, IN2004 36-42, pp.25–30 (2004).
19) 増田峰義, 垂井俊明, 庄内 亨, 吉村 裕, 杉江衛：Web アクセス集中対応 3 層データセンタ制御方式，情報処理学会研究報告，Vol.2002, No.60, OS-90, pp.103–110 (2002).
20) 須藤勝弘, 小倉広実, 三上秀秋, 深瀬政秋：汎用負荷分散装置を用いたメールサーバの多重化，情報処理学会研究報告，Vol.2004, No.96, DSM-35, pp.49–52 (2004).
21) 笠井秀一，毛利公一，吉沢康文：システムの負荷を考慮した負荷分散ルータの開発，プログラミング・シンポジウム報告書，Vol.42, pp.99–106 (2001).
22) 中岡範之，中川俊夫，山本 真：放送連動インターネットサービスにおける時間方向アクセス負荷分散手法の検討と評価，電子情報通信学会技術研究報告，Vol.104, No.187, CQ2004 49-59, pp.41–46 (2004).
23) 永野壮太，中里秀則：自律分散型サーバクラスタにおけるリクエスト振り分け手法，電子情報通信学会技術研究報告，Vol.104, No.354, NS2004 133-144, pp.1–4, (2004).
24) 菊池英明，小林哲則，白井克彦：システムの処理性能を考慮した対話制御方法の検討，人工知能学会研究会資料，SIG-SLUD-9804-1, pp.1–6 (1999).
25) 田熊竜太，岩野公司，古井貞熙：並列処理型計算機を用いた音声対話システムの検討，人工知能学会研究会資料，SIG-SLUD-A201-04, pp.21–26 (2002).
26) カロッツェリアカーナビゲーションシステム. http://www.pioneer.co.jp/carrozzeria/
27) 千村保文：VoIP のイントラネットへの適用 — IP を用いた企業情報通信システム，情報処理, Vol.42, No.2, pp.132–135 (2001).
28) 大杉 淳，竹下 潤，野呂影勇，高尾秀伸，境薫：車載情報端末における音声操作の最適化に関する検討，ヒューマンインタフェース学会研究報告，Vol.2, No.3, pp.17–20 (2000).
29) 山口開生，藤田史郎：電気通信工学，オーム社 (1974).
30) 平山 博：新版データ通信，オーム社 (1980).
31) 小野村哲也，稲井 寛：複数の Web サーバにおける DNS を用いた負荷分散，電子情報通信学会技術研究報告，Vol.102, No.252, IA2002 9-17, pp.21–28 (2002).
32) 伊藤克亘：音声対話システム，電子情報通信学会技術研究報告，SP92-38, pp.23–30 (1992).
33) 四宮光文，村上英世，浅谷耕一：マルチメディアネットワークサービスと品質 B-ISDN 網品質研究課題と標準化動向，電子情報通信学会論文誌, Vol.J80-B-I, No.6, pp.305–312 (1997).
34) 大原一浩，間瀬憲一，能上慎也，柄沢直之：異なるパケット遅延配分を考慮したスケジューリングアルゴリズム，電子情報通信学会技術研究報告, CS99-146, pp.37–42 (2000).

(平成 17 年 1 月 31 日受付)
(平成 17 年 7 月 4 日採録)

荒金　陽助（正会員）

平成 7 年東京工業大学工学部電気電子工学科卒業．平成 9 年同大学大学院総合理工学研究科博士前期課程修了．同年日本電信電話株式会社マルチメディアネットワーク研究所に入所．以来，ITS におけるコミュニケーションおよび情報セキュリティの研究に従事．現在，同社情報流通プラットフォーム研究所情報セキュリティプロジェクト研究主任．博士（工学）．平成 12 年，平成 16 年情報処理学会高度交通システム研究会優秀論文賞，平成 15 年マルチメディア，分散，協調とモバイル（DICOMO2003）シンポジウムベストカンパーサント賞，平成 16 年 DICOMO2004 優秀論文賞，平成 17 年 DICOMO2005 優秀プレゼンテーション賞各賞受賞．電子情報通信学会，IEEE 各会員．

下川　清志

昭和 51 年東北大学工学部電子工学科卒業．昭和 53 年同大学大学院工学研究科前期課程修了．同年日本電信電話公社（現 NTT）電気通信研究所入所．以来，衛星通信システム，アクセス系無線システム，テレマティクス分野の研究実用化，技術企画等に従事．現在，NTT アドバンステクノロジ株式会社アクセスネットワーク事業本部部長．電子情報通信学会会員．

金井　　敦（正会員）

昭和 55 年東北大学工学部通信工学科卒業．昭和 57 年同大学大学院工学研究科情報工学教室博士前期課程修了．同年日本電信電話公社入社．以来，クロスコンパイラ，ソフトウェア開発プロセス，ソフトウェア設計技法，ソフトウェア開発環境，超高速 Web 検索技術，ネットワークコミュニティ，情報セキュリティの研究開発に従事．現在，NTT 情報流通プラットフォーム研究所主席研究員．博士（情報科学）．電子情報通信学会会員．