

音声情報と位置情報を用いた従業員の接客作業分析とその活用

竹原 正矩^{1,a)} 野尻 弘也^{1,b)} 田村 哲嗣^{1,c)} 速水 悟^{1,d)} 蔵田 武志^{2,e)}

概要：サービス業における業務改善や従業員教育を支援するため、行動計測とデータの分析技術の向上が必要とされている。本稿では、レストランにおける従業員の接客作業に着目する。接客作業には発話の機会が多いため、音声情報を分析することで接客の状況や作業に関する指標が得られることが期待される。また、注文伺い、配膳といった接客作業を検出することで、顧客の注文後の待ち時間を推定することができる。そこで、我々は音声情報と位置情報を組み合わせ、従業員の接客作業の検出を行う枠組みを提案し、顧客の待ち時間の推定への活用を検討する。そして、接客作業の検出や待ち時間の推定が、他の店舗やサービスでどのように展開できるかを考察する。

1. はじめに

レストランや介護施設などのサービス業では従業員の生産性向上やサービスの質の向上のため、効率的な業務改善や人材教育が望まれている。しかし、人手による観測や経験に基づく指導は、コストの割に十分な改善効果が得られないことが問題であった。近年、効率的な業務改善や人材教育を支援するため、信号処理やパターン認識の技術をサービス分野で活用する研究が増えている [1]。

顧客とのコミュニケーションの多い接客を中心とするサービス業では、従業員の音声データが重要な情報を持っている。とりわけ、従業員の接客作業は多くの発話を伴うため、発話の頻度や内容、話者などの情報によって、作業の種類を判別することが可能と考えられる。また、どの接客作業がいつ行われたかが分かることで、従業員の接客時間や接客による負荷、店内の状況、サービスの質などを把握することができる。

我々は、レストランの業務改善の一つの指標である配膳待ち時間の推定を目標としている。配膳待ち時間とは、顧客が料理を注文してから配膳されるまでの待ち時間を指す。配膳待ち時間は、短ければ短いほどサービスの質とし

ては良くなり、顧客満足度の向上につながる。また、顧客の滞在時間が短くなり、顧客の回転率も上がる。さらに、料理が配膳されるまでの時間を顧客が知ることにより、時間の使い方を工夫することができ、待ち時間によるストレスの軽減につながる。配膳待ち時間は、客数や注文数、注文された料理などのデータからある程度予測可能と考えられる。多くのレストランは POS (Point-of-Sales) 端末によってこれらの注文情報を管理している。しかし、その時間帯に働いている従業員数や忙しさ、従業員のスキルなどによって配膳待ち時間は変化する。また、実際に業務を振り返る時には、より正確な配膳待ち時間が求められる。

そこで、本稿では、配膳待ち時間を従業員の作業内容から推定する枠組みを提案する。そして、音声情報とその枠組みの中でどのように利用されるか、その有効性を示す。作業内容の推定は従来から行われているが、本稿では推定精度向上のため前処理として接客作業判別を行う。接客作業判別には、従業員の音声情報と位置情報を用い、それらを平滑化して統合する手法を提案する。続いて、目標である配膳待ち時間を推定するために必要な、従業員の注文作業と配膳作業の推定を検討する。これらの実験に加え、配膳待ち時間推定に至った行動計測や作業推定などの技術を他店舗やレストラン以外の現場へ展開する際の汎用性について考察する。

本稿の構成を以下に示す。2章では、本稿の関連研究を示す。3章では、接客作業判別の手法と実験について述べる。4章では、配膳待ち時間のための従業員の接客作業の推定について検討する。5章では、これらの手法の他の店舗・業種への応用について考察する。最後に、6章で本稿

¹ 岐阜大学
Yanagido, Gifu, Gifu 501-1193, Japan
² 産業技術総合研究所
Umezono, Tsukuba, Ibaraki, 305-8568 Japan
a) takehara@asr.info.gifu-u.ac.jp
b) nojiri@asr.info.gifu-u.ac.jp
c) tamura@info.gifu-u.ac.jp
d) hayamizu@gifu-u.ac.jp
e) kurata@ieee.org

をまとめる。

2. 関連研究

2.1 従業員のデータ計測

我々は日本食レストランの従業員を対象として、PDR (Pedestrian Dead Reckoning) 技術を用いたセンサによる従業員の位置・動作の計測と、マイクロフォンを用いた従業員の発話音声の収録を行った [2]。本稿では、前者を従業員の位置データとし、後者を音声データとする。また、これらをまとめて、従業員の行動データと呼ぶ。音声データの収集には、骨伝導マイクロフォンを用いた [3]。これは、音声データに顧客の会話や周囲の物音などの雑音が大きく重畳することを防ぎ、マイクを装着した従業員の音声をきれいに録音するためである。また、顧客の会話が入らないようにすることは、現場で計測する際に問題となる顧客のプライバシーの保護にもつながる。行動データの他に、従業員が注文時に入力する POS 端末から、顧客の客数や注文履歴などを記録した会計データを収集した。データは総じて、約 3 ヶ月、約 3000 時間分収録された。また 16 名の従業員にご協力いただいた。

2.2 音声データの分析

すでに我々は音声データに対していくつかの分析処理を行っている。文献 [4] では、VAD (Voice Activity Detection) 技術を用いて音声データ中の従業員 (マイク装着者) の発話区間を推定した。発話区間は、音声認識の前処理としてだけでなく、単位時間あたりの発話区間の割合 (以下、発話率) を求めることに使われる。発話率は、従業員の作業内容や、従業員のスキルによって違いが見られる。その応用として、従業員の作業推定の特徴量として用いられた [4][5]。VAD の特徴量には、主に低周波成分のパワーや調波構造に基づく特徴量 [6] を統合したものをを用いている。骨伝導マイクロフォンによる音声は、一般のマイクロフォン (空気伝導) と比較して発話時に低周波成分に強いパワーを観測するため、クリーン環境では低周波成分のパワーを用いることで容易に VAD ができる。しかし、実環境では骨伝導マイクロフォンであっても多くの雑音が重畳する。そのため、雑音による誤検出を抑えるために、発話によく見られる調波構造を特徴量として用いることで VAD の精度を改善している。

この他、話者識別を応用した、話者クラスタリングも行われている [7]。これは、音声データ中の発話がマイクを装着した従業員か、他の従業員か、顧客かを分類する手法である。マイクを装着した従業員がどのカテゴリ (従業員または顧客) の人と会話しているのかを知ることで、その従業員が接客をしているのか業務会話をしているのかを判断することができ、作業推定や接客量の推定につながる。文献 [7] では MFCC と SVM をベースとして音声データから

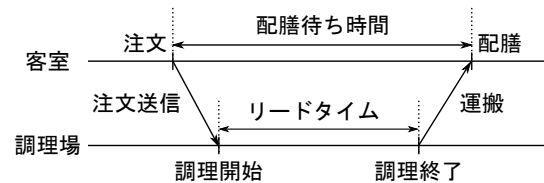


図 1 配膳待ち時間と従業員の作業の関係。

話者を分類しているが、他の従業員と顧客の 2 クラスについては不特定話者であるため、マイクを装着した従業員の位置データを新たな特徴量として加えたり、会話単位で分類したりするなどの工夫をして、分類精度の向上を図っている。

レストランの従業員が接客の際に使うキーワード (「いらっしゃいませ」「失礼します」など) の種類は比較的限られている。また、キーワードから従業員の作業や状況を把握することも難しくはない。我々は、従業員の発話から特定のキーワードの認識も検討しており、比較的クリーンな環境では認識も実現できている。しかし、従業員の発話は全体的に話速が速く状況に応じて異なるキーワード、フレーズを使うため、雑音環境下においては正しく認識できる発話が少ない。また、認識結果をどのように活用するのか、役割や店舗によって語彙が異なる場合どう対応するのか、といった課題がある。

2.3 配膳待ち時間推定

本研究における最終目標は、現場の従業員や経営者が行う業務改善や人材教育を支援する情報を抽出し、実際に役立てることである。前述の発話率や話者、キーワードの情報は、そのままでは現場にとって意味をなさない。それらの情報が、従業員の業務や状況の何を表すかを意味付けすることで、初めて現場に役立てることができる。我々は、現場にとって有用性のある情報として、配膳待ち時間を提案した [8]。配膳待ち時間とは、顧客が料理を注文してから、料理が配膳されるまでの時間を指す (図 1)。配膳待ち時間は短いほど顧客の満足度を高める。また、同じ料理でも配膳待ち時間にばらつきがあるなどの傾向が見えると、その原因を分析することでサービスの改善につながる可能性もある。

配膳待ち時間の類似の指標として、リードタイム [9] がある。これは、レストランにおいては料理の調理時間を指す。配膳待ち時間は、調理だけでなく注文や運搬作業などを含めるため、リードタイムよりも長い。調理場の分析であればリードタイムで十分であるが、我々は、接客従業員の行動も含めた分析を目的としており、配膳待ち時間も同様に重要な指標と位置づけている。本論では接客作業判別の後、配膳待ち時間の推定を検討する。

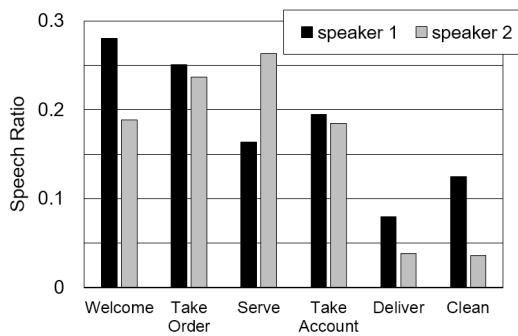


図 2 従業員の作業別の発話率 (Speech Ratio)。

3. 接客作業判別

3.1 概要

接客作業判別 (Waiting Operation Detection ; WOD) とは、従業員の行動データからその従業員が接客作業をしているかどうかを判定する手法である。従業員がいつ、どれくらいの長さ、あるいは頻度で接客作業をしているかを知ることによって、従業員の作業負荷を推定したり、作業推定へ応用したりすることが可能である。

従業員の作業推定は文献 [4][5] でも検討されているが、これらは従業員が作業をしている区間を手で抽出したものに適用している。実際には、従業員の連続した行動データに対して、(何らかの) 作業をしている区間を自動で抽出できることが望ましい。また、それが接客かどうか分かることで、推定する作業内容の絞り込みができるため、作業推定の精度も向上すると考えられる。

従業員の接客作業には 2 つの特徴がある。まず、接客中に発話することである。例えば、注文伺いという接客作業では「ご注文はお決まりでしょうか。」といった発話をする。しかし、運搬や片付けなどの非接客作業は基本的に発話を必要としない。図 2 に、2 人の従業員の 6 種類の作業における発話率の違いを示す。縦軸は発話率で、その作業中に発話した割合を示す。左側 4 つの作業 (案内、注文伺い、料理の配膳、会計) は接客作業である。右側 2 つの作業 (料理の運搬、片付け・セッティング) は非接客作業である。各従業員とも、接客作業は非接客作業に比べて発話率が高い。ただし、接客作業でなくても、従業員同士の連携のために発話したり、突発的に接客作業を行う場合もあるため、全く発話しないというわけではない。2 つ目の特徴は、客室に滞在するという点である。挨拶・案内という作業を除けば、接客は客室で行われる。また、同じ客室内に留まる場合が多く他の客室や通路を行き来するといったことはない。

3.2 処理の流れ

接客作業判別の流れを図 3 に示す。前述の接客作業の特

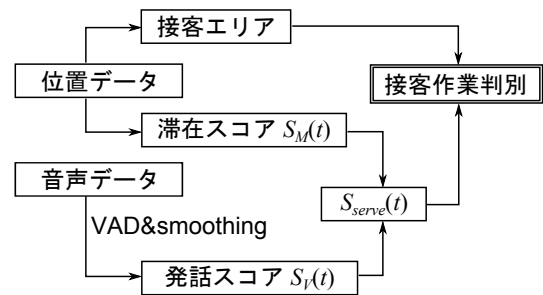


図 3 接客作業判別の流れ

徴を元に、位置データと音声データから計 3 種類の特徴量を用いて接客作業の推定を行う。まず、音声データは VAD によって発話区間を検出し、単位時間 (フレーム t) ごとの発話率を求める。そして、発話率を前後数フレームで平滑化し、全フレームで正規化したものを発話スコア $S_V(t)$ とする。接客作業中、常に発話しているわけではないため、発話していない区間に対応するために平滑化を行う。

続いて、位置データはその時間の滞在スコア $S_M(t)$ と滞在エリアを求める。滞在スコアは、従業員の位置データの変化量の分散値を逆数にしたものである。滞在スコアが高いほど、従業員がその場に留まっている可能性が高い。接客作業は基本的にその場に滞在して行うことが多い。しかし、接客作業であっても客室の中央を基準に大きく移動する場合があるため、変化量ではなく分散値を滞在スコアとして使用する。そして、発話スコアと同様に正規化する。

発話スコア $S_V(t)$ と滞在スコア $S_M(t)$ は、式 (1) のように重み付けして統合される。 $S_{serve}(t)$ はフレーム t の接客作業スコアである。重みは式 (2) のように調節する。

$$S_{serve}(t) = \lambda_V S_V(t) + \lambda_M S_M(t) \quad (1)$$

$$\lambda_V + \lambda_M = 1 \quad (2)$$

接客作業スコアをしきい値処理することで、従業員が接客しているかどうかを判定する。ただし、厨房やパントリーなどで発生する、作業しながら従業員同士で会話するパターンも、接客作業と判定されてしまうため、後処理として従業員の滞在エリアが客室かその付近でない場合は、判定結果から除外することで対処する。

3.3 実験

接客作業判別の実験を行った。表 1 に実験条件を示す。従業員 3 名、計 17 時間分の行動データに対して判別した。1 フレームの長さは 5 秒である。発話スコアと滞在スコアの正規化の段階でそれぞれ 0.5 以上が接客作業らしくなるように調節しているため、接客作業スコアのしきい値を 0.5 とした。まず、フレーム単位での判別を行った。音声データの平滑化の効果を確認するため、平滑化なし、前後 1 フレームでの平滑化、同 2 フレーム、3 フレームの 4 通りで

表 1 実験条件

#subjects	3 employees
data length	17h 2m 15s
#frames	12267
#WO	288 times
#WO frames	1545
threshold	0.5

実験を行った。続いて、接客作業単位で判別できているかを確認するため、1つの接客区間が80%以上検出できている場合は正解として判別した。また、評価尺度には、正解率 (Precision) と再現率 (Recall), F 値 (F-measure) を用いた。正解率と再現率は式 (3)(4) の通りである。また、F 値は、正解率と再現率の調和平均である。

$$\text{正解率} = \frac{\text{正しく接客と判別されたフレーム数}}{\text{接客と判別されたフレーム数}} \quad (3)$$

$$\text{再現率} = \frac{\text{正しく接客と判別されたフレーム数}}{\text{接客と判別すべきフレーム数}} \quad (4)$$

3.4 結果

フレーム単位での接客作業判別の結果を図4に示す。発話率の平滑化によって、判別精度が16ポイント改善していることが分かる。また、前後2フレーム(計25秒)で平滑化した場合、発話率の平滑化の効果が最も表れている。さらに、滞在スコア(位置データ)のみ、発話スコア(音声データ)のみの結果よりも、それぞれを統合することによって性能がわずかながら改善しているが、全体的に発話スコアの方が滞在スコアよりも接客作業の判別に影響を与えていると言える。しかし、精度は決して高くなく、作業推定や接客時間の分析のためには改善が必要である。

続いて、接客単位の判別結果を図5に示す。平滑化の条件は前後2フレームである。フレーム単位で判別した場合と比較して22ポイント改善している。しかし、再現率と比較して正解率が低いことから、接客作業を誤検出してしまふことが多いことが分かる。誤検出の原因として、客室であっても従業員と会話することや、位置データが実際とは違う場所であったことが挙げられる。対策の一つとして、文献[7]のような話者クラスタリングの結果を新たな特徴量として用いることで、そのようなケースでも対応できると考えられる。

滞在スコアの重みを発話スコアの重みよりも大きくした場合、精度が急激に落ちている。これは、その場に静止しているながら発話のない非接客作業を多く誤検出してしまったことが主な原因である。これは、発話が全くない作業は接客作業としないなど、新たな判別条件を加えることによって改善すると考えられる。

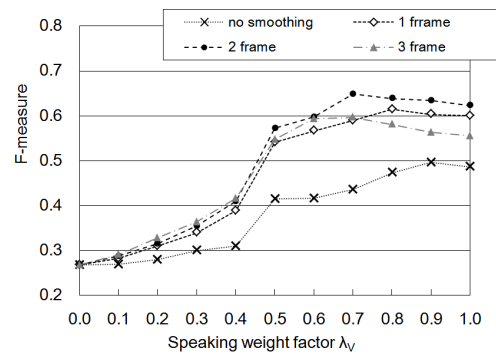


図 4 発話スコアの重みと平滑化の違いによるフレーム単位の接客作業判別結果。

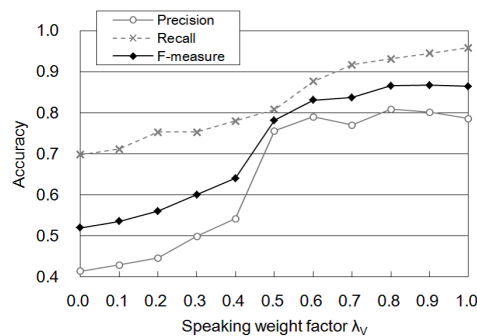


図 5 接客作業単位で見た接客作業判別の結果。

4. 配膳待ち時間推定に向けて

4.1 概要

配膳待ち時間とは、顧客が料理を注文してから配膳されるまでの時間であり、レストランにおいてサービスの質を表す指標の一つである。配膳待ち時間は短いほど顧客の満足度を高める。配膳待ち時間を推定する方法はいくつか考えられる。例えば、料理によって調理時間が決まっているため、料理名と従業員の忙しさ(客数や注文数、従業員数など)からおおよその配膳待ち時間が予測可能とみられる。しかし、従業員の忙しさの推定は容易ではない。別の方法として、従業員の作業内容から推定することが考えられる。従業員の注文作業と配膳作業がその客室で行われた時刻が分かれば、その時刻の差分を取ることによって配膳待ち時間を計算することができる。配膳待ち時間は、予測だけでなく、実際どのくらいの時間かかったか、現状を知るといふ点でも重要な指標である。その点においても、従業員の作業という視点から配膳待ち時間を推定できることが望ましい。本章では、接客作業判別の結果を元に従業員の注文作業と配膳作業の推定を検討する。

4.2 注文作業

本研究で実験したレストランの従業員は、注文作業の際に POS 端末を用いて顧客情報や注文情報を記録する。POS 端末には注文時刻が記録されており、おおよその注文

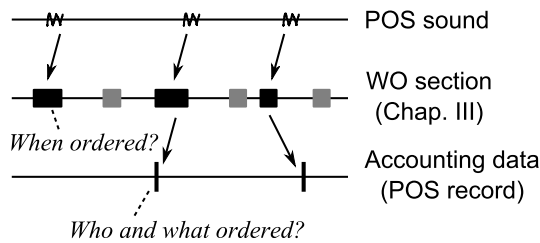


図 6 従業員の注文作業の推定の流れ。

作業の時刻と、どの客室で注文作業を行ったかが分かる。また、音声データ中には POS 端末の操作音が含まれるため、音声データから操作音を検出することにより注文を入力した正確な時刻が分かる。そこで、図 6 のように、3 章で判別した接客区間情報と、POS 端末音、注文情報を元に、従業員の注文作業を推定した。従業員は注文作業後すぐに POS 端末を操作して記録するため、操作音が鳴る直前の接客区間を注文作業とみなすことで、ほとんどの注文作業を特定できることが分かった。

4.3 配膳作業

注文作業とは対照的に、従業員が配膳作業をするときは POS 端末を使用しないことが多く、配膳時刻も記録されていない。そのため、POS 端末のデータを用いず、従業員の行動データと接客区間の情報から配膳作業を特定する必要がある。また、配膳待ち時間を求めるためには、配膳作業がどの顧客に対して行われたか、対応付けの必要がある。

図 7 に、手で音声データから配膳作業区間を抽出した際の、配膳作業と判断した理由の統計を示す。配膳作業の約半数では「こちら（料理名）です。」というように料理名を発話している。また、22%は配膳終了後に料理の説明をしている。よって、接客区間の最初と最後において発話の頻度や発話率が高ければ、配膳作業の可能性があるとと言える。また、配膳作業の際に使われるキーワードやお盆や食器をテーブルに置くときの音なども頻度として特徴量に用いることで、配膳作業を推定する手がかりとなる。

配膳作業と顧客との対応付けについては、基本的に従業員の位置データと POS 端末の注文情報にある客室のデータ（どのテーブルの注文か）を組み合わせることで実現可能である。しかし、位置データの精度が十分ではなかったため、新たな対応付けの手法が必要である。

5. 他店舗・業種への応用

本章では、作業推定や接客作業分析がどのような店舗、業種、シーンに応用できるかを考察する。

5.1 行動計測と作業推定

我々の技術は、大型レストラン、旅館、医療施設など比較的規模の大きい店舗であれば適用する効果があると考えられる。これらの施設は店舗の広さや従業員数の多さなど

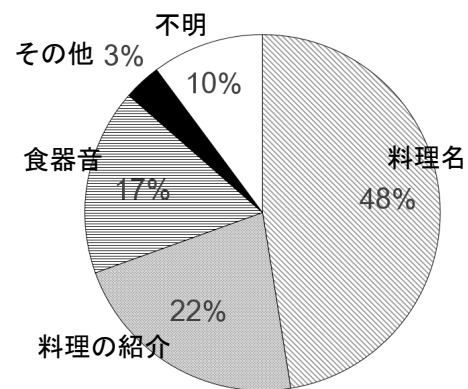


図 7 音声データから配膳作業と判断した理由の内訳。

で、従業員の仕事ぶりや問題点について十分な把握が難しい。そこに、行動計測によってデータを記録し、客観的な分析を施すことによって、業務中の問題点の発見や従業員自身の振り返りに貢献できる。一方で、コンビニなどの小規模の店舗では、ある程度作業内容が決まっていたり、作業を行う範囲が限られるため、従業員の作業を分析・評価することは人手でもそれほど難しくないと考えられる。しかし、従業員の行動が顧客の行動にどう影響を与えたかという観点で、従業員の作業と顧客の行動と比較する場合、行動計測や作業推定は同様に必要性が高いと予想される。

5.2 待ち時間推定

本稿では、顧客の注文時刻のみ会計データに存在する場合の待ち時間推定を検討したが、店舗によっては配膳時刻の情報を持っていたり、すでにおおよその待ち時間を予測して顧客に開示する店舗も見られる。もし、従業員の作業に基づいて待ち時間を推定できれば、配膳時の待ち時間にかぎらず、様々なシーンの待ち時間の推定に応用できる可能性がある。例えば、顧客が入店してから着席できるまでの待ち時間や、注文以外の要望に対する待ち時間などがある。また、時間を推定するだけでなく、従業員の忙しさや接客スキルがどう待ち時間に影響するのかを調査することで、現場の業務改善をさらに支援できることが期待される。

5.3 従業員の接客量

従業員がより多くの顧客に対して関わりを持つことで、顧客満足度や売上の上昇につながる。本稿で求めた接客区間は統計的に分析することで、顧客単位の接客・サービス量や、従業員の接客量、それに伴う負荷を推定することが可能である。例えば、高齢者介護施設の場合、入居者との会話量や声掛けの回数の分析が要望としてある [10]。これは本稿で推定した接客作業区間から分析が可能である。量販店でも、おすすめ商品の紹介やサポート対応など、従業員が積極的に接客することによって顧客の購買心理を向上させることが、利益の向上につながる。その点で、従業員の接

客作業の時間を把握することは重要である。仕入れなど接客以外の業務や受動的に接客するレジ業務と区別し、それぞれの時間の割合や接客作業が売上向上に結びついたかを分析することが業務改善につながると考えられる。

6. まとめ

本稿では、顧客の配膳待ち時間を推定するために、従業員の接客作業に着目した。従業員の音声データや位置データを統合することで接客らしさを数値化し、接客作業判別を行った。音声データから得た発話率は平滑化することにより、接客作業判別の精度が向上した。

接客作業判別の結果と会計データ、POS 端末音から従業員の注文作業のほとんどを推定可能であることが分かった。従業員の配膳作業については、食器音や発話の頻度、キーワードなどを分析し、総合的に判断することが必要である。また、どの顧客に対する作業かの対応付けも課題である。

接客作業は、その内容を明らかにすることによって、待ち時間などより業務改善に有用な指標を得ることが可能である。一方で、他店舗や業種へ応用する際、接客作業かどうか、あるいは接客作業の回数、長さなどの情報は汎用性が高いと考えられるため、推定精度の向上や、他データとの比較を検討していきたい。

謝辞 本研究は平成 21 年度および 22 年度経済産業省委託事業「IT とサービスの融合による新市場創出促進事業」として実施されました。計測実験にご協力頂きました産業技術総合研究所サービス工学研究センターの皆様にご感謝いたします。また、調査にご協力頂きましたがんこフードサービスの皆様にご礼申し上げます。

参考文献

- [1] M. Takehara, S. Tamura, R. Tenmoku, T. Kurata and S. Hayamizu: *The role of speech technology in service-operation estimation*, Proc. O-COCOSDA 2011, pp.116-119, 2011.
- [2] F. Tomohiro, R. Tenmoku, T. Okuma, R. Ueoka, M. Takehara and T. Kurata: *Improving service processes based on visualization of human-behavior and POS data: a Case Study in a Japanese Restaurant*, Proc. ICServ 2013, 2013.
- [3] M. Takehara, S. Tamura, S. Hayamizu, R. Tenmoku, T. Okuma, T. Fukuhara and T. Kurata: *Measurement and analysis of speech data toward improving service in restaurant*, Proc. O-COCOSDA 2013, 2013.
- [4] 竹原正矩, 加藤狩夢, 田村哲嗣, 天目隆平, 蔵田武志, 速水悟: 業務音声の発話区間検出による作業推定の改善, 電子情報通信学会論文誌 Vol.J97-D, No.10, 2014.
- [5] 宇野太久哉, 竹原正矩, 田村哲嗣, 速水悟, 蔵田武志: 作業推定に向けた環境雑音のクラス分類, 第 13 回情報科学技術フォーラム (FIT2014), 2014.
- [6] Y. Guo, Q. Qian and Y. Yan: *Robust voice activity detection based on adaptive sub-band energy sequence analysis and harmonic detection*, Proc. Inter-

- speech 2007, 2007.
- [7] T. Kawase, M. Takehara, S. Tamura, S. Hayamizu, R. Tenmoku and T. Kurata: *Improvement of utterance clustering by using employees' sound and area data*, Proc. ICASSP 2014, pp.3071-3075, 2014.
 - [8] 野尻弘也, 竹原正矩, 前山賢人, 田村哲嗣, 蔵田武志, 速水悟: 位置・発話・会計データを用いた配膳待ち時間推定の試み, 第 12 回情報科学技術フォーラム (FIT2013), pp.171-174, 2013.
 - [9] 新村猛, 竹中毅, 上岡玲子, 蔵田武志, 大浦秀一: 外食産業におけるサービス工学の導入事例, 精密工学会誌, Vol.78, No.3, pp.208-211, 2012.
 - [10] 天目隆平, 竹原正矩, 速水悟, 蔵田武志: 労働集約型サービス-従業員行動計測技術に基づく分析と可視化, HCG シンポジウム 2010 論文集, pp.443-448, 2010.