

# 接客時間推定に向けた従業員の位置・音声データによる 発話クラスタリング

川瀬 徹也<sup>1,a)</sup> 竹原 正矩<sup>1,b)</sup> 田村 哲嗣<sup>1,c)</sup> 天目 隆平<sup>2,d)</sup> 蔵田 武志<sup>2,e)</sup> 速水 悟<sup>1,f)</sup>

概要：我々はレストランで収録された音声について、発話クラスタリングの研究を行っている。発話データには、従業員同士の会話や、顧客との会話が含まれており、対象話者をクラスタリングすることで、従業員の接客時間などの業務に関する指標を推定できることが期待される。本稿では、マイク装着者、他の従業員、顧客の3クラス発話クラスタリングを検討している。しかし、話者の不特定性や雑音の影響により分類精度が低下する事が考えられる。そこで、従業員の位置情報を音声データと統合して精度向上を試みた。さらに、提案手法をレストラン以外の場面に応用した際の汎用性について考察した。

## 1. はじめに

近年、信号処理やパターン認識の分野に関連して、サービス工学が注目を集めている。サービス工学では、従業員の作業効率化やサービスの質の改善のために、情報の可視化やデータマイニング技術を通して様々なセンサーから得られたデータが利用される。

我々はレストランやホテルなどの施設における従業員の行動の改善に取り組んでいる [1]。従業員の行動を推定するためには、音声データが欠かせない。なぜなら、ある従業員に着目した際、その従業員が他の従業員もしくは顧客のどちらと話しているかを分類することができれば、接客しているか業務会話をしているかという、その従業員の状況を推定することができるからである。さらには、この状況推定結果から得られる接客時間も、サービスの質を決める重要な要素である。

しかし、実環境で収録された音声内の発話をクラスタリングする場合、2つの問題が生じる。1つ目は雑音の多様性である。本稿で取り上げているレストランでは、環境雑音、食器音やマイク接触音などの雑音が頻繁に発生する。これは、クラスタリングや音声認識等の精度を低下させる

要因になる。また、発話であっても対象とする従業員と関連のない、雑音としての発話も多く重畳する。2つ目はクラスタリング対象の話者が不特定であることである。接客時間を推定するには、分析対象の従業員が話している相手が、他の従業員であるか顧客であるかをクラスタリングする必要がある。しかし、他の従業員や顧客のクラスに属する人はマイク装着者と比べて多様であるから、その明確なモデルを作るのは難しい。

そこで、他の情報を音声データと統合してクラスタリングを行う手法が考えられる。先行研究として話者識別の分野では、マイクロフォンアレイやカメラと発話者の相対位置を利用して、会議場での発話者を特定する研究がある [2]。しかしながら、これらの研究では装置や発話者が特定の位置に固定されており、レストランのような従業員が頻繁に動く環境にこの研究を適用することは難しい。

我々はこれら2つの問題を解決するため、骨伝導マイクから得られた音声データ以外に、PDR センサによって計測される従業員の位置情報を用いる。位置情報は、客室や調理場といった、その座標に対応するエリアデータに変換される。従業員の滞在エリアデータを用いることで、例えば調理場では顧客との会話よりも従業員同士の会話が発生しやすいといったように、発話者をある程度限定できる。また、位置データは音声データに見られるような、環境や作業による雑音の影響を受けないため、雑音の多い環境下でも貢献できると考えられる。我々はさらに、滞在エリアデータを利用してより正しくクラスタリングされた発話ラベルを用いて、会話の種類（従業員同士の会話か顧客との会話か）を分析する手法を提案する。この結果は、従業員の行動分析、特に接客時間の推定において有用であ

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

る．そして最後に提案手法をレストラン以外のサービス現場へ適用した場合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，音声データのみでの発話クラスタリングについて述べる．3章では，提案手法である，エリアデータを合わせて用いた発話クラスタリングについて述べる．4章では，会話の種類を分類する実験について記述する．5章では，他の施設への提案手法の適用について考察する．最後に，6章で本稿をまとめる．

## 2. 音声データのみを用いた発話クラスタリング

### 2.1 概要

話者識別の先行研究の多くは，GMM (Gaussian Mixture Model) や SVM (Support Vector Machine) を識別器として採用している [3] [4] [5]．特徴量としては，音声信号の対数パワーと MFCC (Mel-Frequency Cepstrum Coefficients) を用いているものが多い．

発話クラスタリングは，音声信号中の発話をいくつかのクラスに分類することを意味する．発話クラスタリングの処理の流れは話者識別と同じである．本稿でも，特徴量として音声信号の対数パワーと MFCC，分類器として SVM を用いる．

### 2.2 特定話者とその他の分離

本研究の最終目標は，ある会話が従業員同士の会話か顧客との会話を分類し，さらに接客時間を推定することである．これらを推定するには，音声データ内の発話を以下の3クラスに分類する必要がある．

- マイク装着者 (Own)  
その音声データの収録に使われたマイクを付けていた従業員
- 他の従業員 (Oth)  
上記の従業員と同じ時間帯に働いていた従業員
- 顧客 (Cus)  
レストランに食事に来た顧客

3クラス分類の前に予備実験として，マイク装着者とその他の2クラス分類を行う．マイク装着者は，マイクのすぐ近くで発話する分パワーが安定し，かつ特定話者であるため，分類は容易であると想定される．この2クラス分類の結果は，マイク装着者が一定の時間内でどの程度発話を行ったかを推定するのに役立つ [1]．

2クラス分類の流れを図1に示す．我々が用いた特徴量である MFCC や音声パワーはフレーム単位で抽出され，また分類器である SVM もフレーム単位で動作する．そのため発話単位の分類は，発話内のフレームの分類結果の多数決で行う．実験条件は表1に示した通りであり，また分類器である SVM には SVM-Light [6] を使用した．分析対象データはマイク装着話者が2名，それぞれ1時間の発話

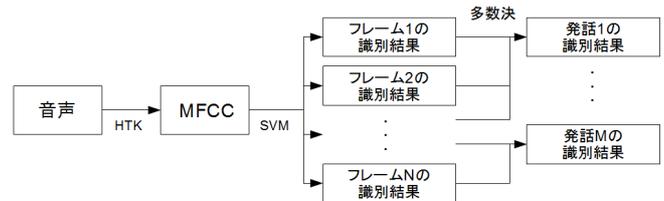


図1 2クラス分類の流れ

表1 2クラス分類の実験条件

訓練データ	positive	マイク装着者 3000 フレーム
	negative	他の従業員・顧客，各 1500 フレーム
テストデータ	positive	マイク装着者の発話
	negative	他の従業員・顧客の発話
フレームサイズ	25msec	
フレームシフト	10msec	
特徴量	39次元	
	MFCC(12), Δ MFCC(12), ΔΔ MFCC(12) Power(1), Δ Power(1), ΔΔ Power(1)	

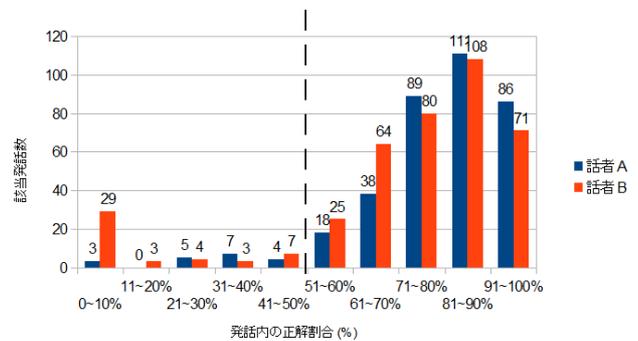


図2 2クラス分類での話者正解率ヒストグラム

である．マイク装着話者について話者 A は 361 発話，話者 B は 394 発話を含んでいる．発話の正解ラベルは人手で付けられたもので，また実験はオープン条件で行った．音声データの詳細は，文献 [7] を参考にして頂きたい．

この実験の評価は，フレームや発話がどの程度正しく分類されたかによって行う．フレーム単位での分類では，話者 A で 79.2%，話者 B で 71.7%の正解率が得られた．また，発話単位での分類では式 (1) で示される話者正解率を用いる．

$$\text{話者正解率} = \frac{\text{正解の話者と判定されたフレーム数}}{\text{発話中の総フレーム数}} \quad (1)$$

図2に，発話単位の話者正解率のヒストグラムを示す．発話単位の話者正解率が50%を超える発話は全体の約90%にあたり，発話単位で多数決をとることによって，約90%の発話は正しい話者として分類できることが分かる．この結果は，雑音が多い実環境であっても，マイク装着者のように話者が特定のかつその音声パワーが安定していればその話者の発話量推定が可能であることを示す．

表 2 3 クラス分類の実験結果

正解\推定	マイク装着者	他の従業員	顧客	合計発話数 (A, B)
マイク装着者	92.4%	6.0%	1.6%	(361, 337)
他の従業員	12.3%	60.4%	27.4%	(32, 25)
顧客	15.7%	51.0%	33.3%	(137, 168)

## 2.3 3 クラス分類

### 2.3.1 手法概要

本節ではマイク装着者・他の従業員・顧客の3クラス分類の実験を行う。この実験では、音声データのみで上記3クラスをどこまで分類できるかを検証する。

3クラス分類の実験を行うにあたり、2通りの手法が考えられる。1つはマルチクラスSVMを音声データに適用する手法で、もう1つは先ほどの実験でマイク装着者ではないと判定された発話を、更に他の従業員と顧客の2クラスに分類する手法である。この2つの手法を比較すると、「3クラスでの分類がうまくいけば骨伝導マイク以外で録音した音声にも用いることが出来、クラス数が増えた場合にも応用出来る」という利点が前者の手法には存在する。そのため、本稿では前者の手法を採用する。

### 2.3.2 実験条件

SVMの訓練データは3クラスとも3000フレーム(合計9000フレーム)、テストデータは各クラスの発話である。分類器にはSVM-Multiclassを使用する。他の条件は、2クラス分類の実験と同様である。

### 2.3.3 実験結果

3クラス分類の実験結果と各クラスの合計発話数を表2に示す。表中の行は正解ラベル、列は実験結果の分類ラベルに対応する。各数値は行の正解ラベルがついた発話群を列の分類ラベルに分類した割合を表し、2つの実験データの平均値である。例えば、マイク装着者の発話を誤って他の従業員と判定した割合は平均で6.0%である。表2から、マイク装着者の発話は2クラス分類の時と同様、綺麗に取り出せていることが分かる。一方、他の従業員のクラスと顧客のクラスの間での誤分類が多い。これは、マイク装着者が特定話者であり、かつその音声パワーが安定的であるのとは逆に、他の従業員や顧客は不特定話者であり、かつマイクからの距離が様々であるため音声パワーが不安定であるためだと考えられる。

この問題を解決する方法として、顧客よりは特定のである従業員全員分の個別音声モデルを作成し、より多くのクラスでクラスタリングを行う方法が考えられる。しかしこの手法を利用するには、どの従業員がどの発話を行ったかを人手でラベル付けする必要がある、とても手間がかかる。またこの手法では、接客時間の推定に必要な、顧客の発話についての精度向上は見込めない。そこで我々は、従業員の位置情報から算出したエリアデータを音声データと合わせて利用する手法を提案する。

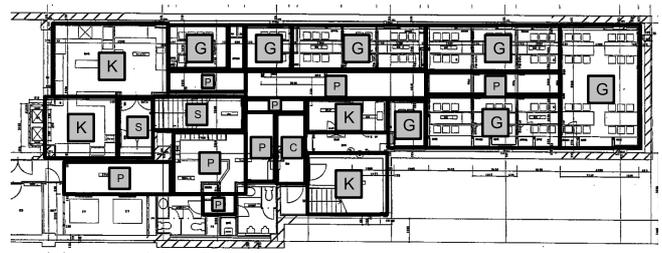


図 3 レストランの見取り図

## 3. エリアデータを用いたバイモーダル発話クラスタリング

### 3.1 エリアデータ概要

我々は音声データの他に従業員の位置情報を集めている[8]。レストランにおいて、誰と会話しているか(対象話者)は場所によってある程度限定される。例えば、顧客は通路や客席にはいるが、調理場にいることはまず無い。反対に、調理を専門とする従業員は調理場にいることが多く、通路や客席にいることは少ない。つまり、マイク装着者がどこで話しているかによって、対象話者の候補を絞り込むことができる。そのため、マイク装着者の位置情報を発話クラスタリングに用いることで、誤分類を減らすことが期待される。

本研究では、従業員の位置情報としてエリアデータを用いる。エリアデータとは、PDR センサから得られた従業員の3次元位置座標を客室や調理場といったエリアに対応させて作られたデータである。我々がデータを収集したレストランの見取り図の一部を図3に示す。このレストランのエリアを大分すると、客室(G)・調理場(K)・通路(P)・レジ(C)・そして階段(S)の5種類となる。本稿では、調理場を従業員同士の会話が起こる可能性の高いエリアとして、また、客室とレジを顧客との会話が起こる可能性の高いエリアとして捉え、発話クラスタリングの精度向上を試みた。

### 3.2 バイモーダル発話クラスタリング

本節では、提案手法の1つであるバイモーダル発話クラスタリングについて説明する。提案手法の処理は、図4に示された決定木と以下の式に基づいて行われる。

$$D_1 = \lambda_s R_s(oth) + \lambda_a R_a(kit) \quad (2)$$

$$D_2 = \lambda_s R_s(cus) + \lambda_a R_a(gr, ca) \quad (3)$$

$$D_3 = R_s(oth) - R_s(own) \quad (4)$$

$$D_4 = R_s(cus) - R_s(own) \quad (5)$$

$$D_5 = R_s(oth) - threshold \quad (6)$$

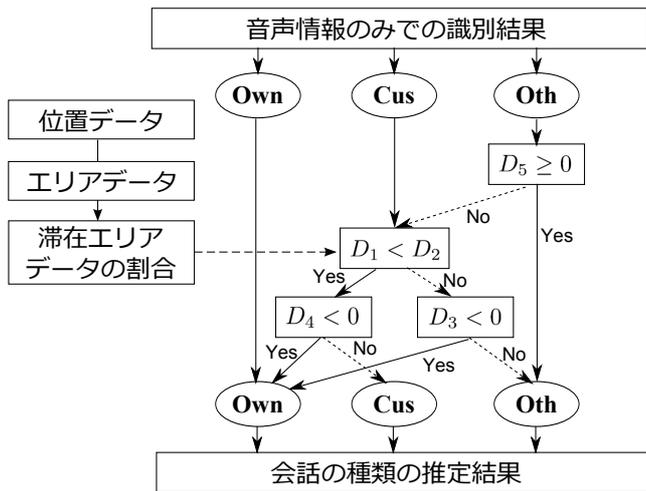


図 4 エリアデータを統合した発話クラスタリングの流れ

式 (2)(3)(4)(5)(6) において、各変数は以下の内容を表す。

- $\lambda_s, \lambda_a$  : 音声データとエリアデータの重み
- $R_s(x)$  : ある発話において、音声データのみでの 3 クラス分類でクラス  $x$  だと判定されたフレームの割合
- $R_a(kit)$  : ある発話中に、マイク装着者が調理場に行った時間の割合
- $R_a(gr, ca)$  : ある発話中に、マイク装着者が客室もしくはレジにいた時間の割合

先の 2 つの実験からマイク装着者の発話は高い精度で分類できることが分かっているため、音声データのみを用いた 3 クラス分類でマイク装着者だと判定された発話は、そのまま結果を採用する。他の従業員だと判定された発話は、 $D_5$  の閾値を基準に訂正すべきかどうか決定する。客室・レジの発話でも他の従業員と会話している場合があるが、図 4 中の  $D_1$  と  $D_2$  の比較において、客室・レジの発話は顧客との会話の重みを高くしているため、他の従業員との会話を誤訂正してしまうケースがある。そこで、 $D_5$  の閾値判定によって、他の従業員の発話である可能性が高い発話はエリアデータによる訂正を行わないようにした。

音声データとエリアデータの重み付き線形結合で表される  $D_1$  と  $D_2$  の比較では、他の従業員あるいは顧客と判定された発話に対して、エリアデータを加えて再判定を行う。マイク装着者が他の従業員と話していると判定された場合は、 $D_3$  の符号により、その発話がマイク装着者のものであるか他の従業員のものであるかを判定する。マイク装着者が顧客と話していると判定された場合は、 $D_4$  の符号により、その発話がマイク装着者のものであるか顧客のものであるかを判定する。

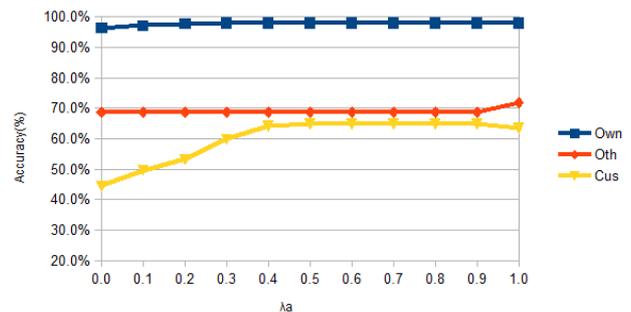
### 3.3 エリアデータを用いた 3 クラス分類実験

#### 3.3.1 実験条件

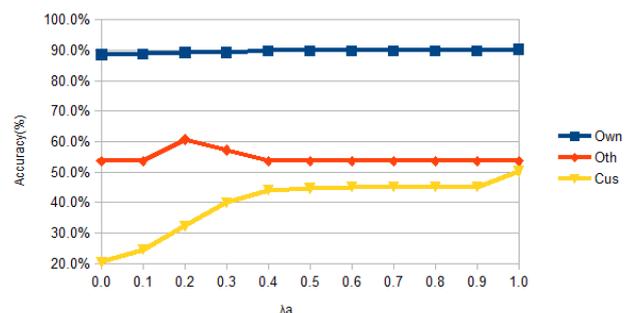
提案手法の有効性を検証するため、3 クラス分類の実験を行った。表 3 に実験条件を示す。

表 3 バイモーダル発話クラスタリングの実験条件

音声データ	訓練データ	各クラス 3000 フレーム
	テストデータ	各クラスの発話
	フレームサイズ	25msec
	フレームシフト	10msec
	特徴量	39 次元 MFCC(12), $\Delta$ MFCC(12), $\Delta\Delta$ MFCC(12) Power(1), $\Delta$ Power(1), $\Delta\Delta$ Power(1)
エリアデータ	データ間隔	1sec
	エリア種類	客室, レジ, 調理場, その他



(a) 話者 A



(b) 話者 B

図 5 バイモーダル発話クラスタリングの正解率推移

分析する音声データは先の 2 つの実験と同様である。他の従業員と判定された発話の誤訂正を防ぐ式 (6) の閾値は、経験的に 0.75 とした。音声データの重み  $\lambda_s$  とエリアデータの重み  $\lambda_a$  は、式 (7) を満たすように 0.1 刻みで変化させる。

$$\lambda_s + \lambda_a = 1 \quad (7)$$

#### 3.3.2 実験結果

エリアデータの重み  $\lambda_a$  の変化に応じた、各クラスの発話正解率の変化を図 5 に示す。図 5 より、エリアデータの重みを大きくするほど、顧客の発話クラスタリング精度が向上していることが確認できる。顧客の発話クラスタリング精度は、エリアデータを全く用いない場合と比較して、最大で話者 A の音声データで 20.4%、話者 B の音声データで 29.7% 向上した。

一方で、他の従業員の発話クラスタリングの精度はあまり向上していない。これは、顧客との会話の多くが客室や

レジで発生しているのに対して、他の従業員との会話は様々なエリアで発生しているためだと考えられる。

会話の発生する場所が限定される場合は提案手法は有効である。一方で、限定されない場合はエリアによってどのような会話が発生するかを確率的に定義したり、他の従業員のクラスを役割ごと（例えば調理場専門、客室専門など）に細分化するなどの対策が考えられる。

## 4. 会話種類の分類

### 4.1 会話種類の定義と実験条件

3章の実験では、エリアデータを音声データと組み合わせることにより、発話クラスタリングの精度を向上させることができた。本章では、接客時間の推定に向けて、会話単位での推定を行う。会話とは、数秒以上の無音区間を境界として複数の発話をまとめたものと定義する。会話内に含まれる発話は表4に示すように4通り存在する。

従業員の行動や接客時間の推定にはどのタイプも何らかの形で正しく推定できることが重要だが、タイプ1（マイク装着者のみ）は元々のマイク装着者の精度が高く、タイプ2（3クラス混合）のケースは少ないため、ここではタイプ3とタイプ4（2クラス混合）の場合に着目する。会話の種類がタイプ3であるかタイプ4であるかは、その会話内で、他の従業員と推定された発話の合計時間が顧客のそれを上回っていればタイプ3、そうでなければタイプ4とする。

会話の境界には、一定以上の長さの無音区間を用いる。本稿では7秒・5秒・3秒の3種類を試した。実験の評価は、2章の3クラスでのクラスタリング結果と3章のクラスタリング結果をそれぞれ会話単位にまとめ、2章の結果をまとめた会話のうち誤判定された会話の合計時間を、3章の結果をまとめた会話でどれだけ訂正できたかとし、それは式(8)で表される。

$$recovery\ rate(\%) = \frac{\text{訂正できた会話の合計時間}}{\text{誤判定された会話の合計時間}} \quad (8)$$

### 4.2 実験結果

タイプ3とタイプ4を合わせて、全体で回復できた時間の割合を図6に示す。図6より、エリアデータを統合することで、音声データのみでの分類で間違っ分類された会話の約70%を訂正できたことが確認できる。また提案手法は、会話の境界とする無音区間の長さによってその訂正の性能があまり変化しなかった。この事は、提案手法が会話の境界となる無音区間の長さ、更にはそれらの境界により作られる会話の長さに依存せず効果を発揮できることを表す。このようにエリアデータを用いることで、発話クラスタリングのより正確な接客時間の推定への貢献が期待される。

さらに、3章と4章の結果を合わせて考察する。3章で

表4 会話種類の定義

	マイク装着者	他の従業員	顧客	会話内容の例
タイプ1	○			挨拶, 独り言
タイプ2	○	○	○	複数人での接客
タイプ3	○	○		業務連絡
タイプ4	○		○	注文, 会計

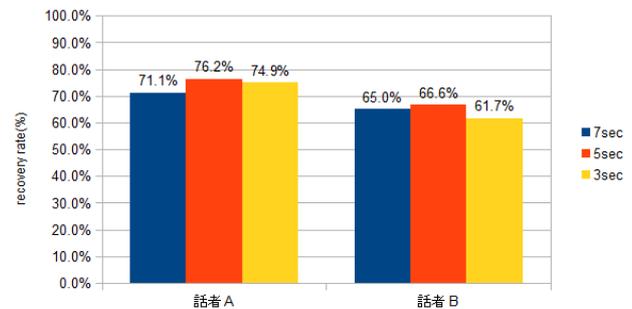


図6 エリアデータの利用による会話種類の誤分類回復割合

はバイモーダル発話クラスタリング、4章ではバイモーダル会話種類分類について述べたが、そのどちらでもエリアデータは正確なクラスタリングに貢献した。また、その有効性は3章の発話単位での分析よりも4章の会話単位での分析の方が大きかった。このように、発話という小さい単位でエリアデータを利用して、その結果を用いて会話というより大きい単位のクラスタリングを行うことは有用であった。この事実は、他のバイモーダル信号処理の分野にも共通するかもしれない。

## 5. 提案手法の汎用性考察

4章まででは、日本食レストランの業務改善のために提案手法を適用した場合に、発話者や会話種類の分類精度が向上することを確認した。本章では、この提案手法を他の施設に適用し、業務の改善に役立てることが可能であるかを考察する。施設の例としては、我々の先行研究で取り上げている介護付き老人ホームや温泉旅館の他に、量販店や病院を取り上げる。考察点は、現場の要望等に対して提案手法が適用できた場合に役立つか、データ収集や適用の際の問題点は何か、の2点である。

### 5.1 介護付き老人ホームの場合

我々の先行研究 [9] において、業務改善のために介護主任・看護師・介護ヘルパーのそれぞれについて合計移動距離やフロア移動回数などを提示したところ、以下の様な要望が得られた。

- 1週間単位・1時間単位と時間の尺度を変更した際の、従業員が各エリアに居る時間の比率を見たい。
- 従業員の音声を記録する際に、入居者との会話量や声掛けの回数等を分析して欲しい。

1つ目の要望について、直接的にはより広い時間範囲で観測された位置データを分析することで要望を叶えることができる。更に細かい、「あるエリアに滞在している時間の中で、待機している時間や介護を行っている時間の割合が見たい」といった要望が出た場合には、提案手法を適用することで単純な位置情報だけでは推定が難しい従業員の行動内容もある程度推定できると考えられる。

2つ目の要望について、入居者との会話量の推定には、会話種類の分類の提案手法をそのまま応用することができる。また声掛けの回数も、提案手法中の無音区間に相当する、会話の区切り目を表す指標を得ることができれば同時に実現可能である。

次に、データ収集に関する問題点を考える。レストランでは不特定多数の顧客に食事を提供するという営業スタイルであるのに対し、介護付き老人ホームでは施設に登録された入居者に対して介護やサービスを行う。従って、親族を含めた入居開始時の話し合いのような個人情報をやりとりする場面が生じる。このような場面のデータについては使わないようにするなど、利用者への配慮が必要である。

## 5.2 量販店の場合

量販店で利益を上げるには、価格の調整や話題商品の取り入れ等の他に、商品に関する知識を持った店員による接客が欠かせない。これは、顧客が商品の知識を得ることで購買行動に結びつく可能性が高くなるからである。

そのため、業務改善には本稿で扱った接客時間も重要な要因となり得る。提案手法の適用には、調理場を従業員事務所、客室をレジ周辺以外の販売フロア（従業員が固まりやすいためレジを除く）のように対応させればよい。さらに、本稿で扱ったレストランでは「通路では従業員同士の会話も顧客との会話も起こり得る」という問題があったが、量販店では業務連絡にインカムを中心に用いる店舗もあり、インカムの音質の特徴を特定することで、接客エリアでの従業員同士の会話を除外することができれば、従業員の滞在エリアを接客エリアと従業員同士の会話エリアにより明確に分離して利用することが可能である。

データ収集については、商品の取り置きのような場面では個人情報が入る可能性が考えられるため、その時間だけマイクをオフにするなどの配慮が必要である。

## 5.3 温泉旅館の場合

我々の先行研究 [9] で取り上げた温泉旅館では、スキル伝達のサポートやサービスのモデル化のために、客室係の計測当日の業務内容について回顧インタビューが行われた。回顧インタビューの際に、従業員の位置や姿勢の計測データから推定した一人称視点の擬似映像を提示するのだが、会話がある部分については提案手法を適用し、顧客対応しているかどうかを合わせて提示すればインタビューさ

れる従業員がより思い出しやすくなる可能性がある。

また温泉旅館では、食事時間など接客を中心に行う場合と布団敷きなど接客がない作業があり、従業員の役割分担から接客しているであろう時間帯を決定することで、時間情報も特徴として利用することが可能である。

## 5.4 病院の場合

医師や看護師の、患者への十分な時間の発話は、患者に病気の症状を詳しく説明する、もしくは医療処置に対する患者の不安を和らげようという努力だと捉えることができる。この部分には、予備実験の2クラス分類で対応が可能である。また、医療従事者間の会話と、患者対応の会話とが分離できれば、長時間の音声ファイルからその部分を抜き出して集めて分析することで、新米医師の学習に使う指標を作ることも可能である。

我々と同じようなデータ収集を行った医療分野の研究に、E-ナイチンゲール [10] がある。このプロジェクトは、行動に基づく知識（行動経験知識）を従業員の間で共有して医療業務を改善しようという目的で行われているため、本稿の提案手法が貢献できる可能性がある。

## 6. 結論と今後の課題

本稿では、エリアデータを音声データと組み合わせることで発話や会話のクラスタリング性能を上げる、バイモーダル処理について提案した。発話単位でのクラスタリングでは、音声データのみを用いた場合と比較して、顧客のクラスタリング精度を20~30%向上させることが出来た。従業員同士の会話か顧客対応かを判別する会話種類の分類では、音声データのみで間違っただけの時間のうちの約70%を、エリアデータを合わせることで回復することが出来た。

今後の課題としては、他の指標との更なる統合や理論的な決定木の作成が挙げられる。他の指標の例として注文や会計の時刻があり、これによりエリアデータが間違っている場合も訂正できることが期待される。エリアデータと音声データを統合する決定木については、エリアデータの効果が強めに出てしまう可能性がある構造になっているため、「客室にいたら何%の確率で接客をしている」といった確率的なデータの扱いを検討する。また、4章で分類した会話種類を基本として接客時間推定を行っていく。

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へ感謝いたします。また、調査にご協力頂きましたがんこフードサービス、スーパーコート平野、西村屋招月庭に御礼申し上げます。

## 参考文献

- [1] Takehara, M. and Tamura, S. and Tenmoku, R. and Kurata, T. and Hayamizu, S.: *The role of speech technology in service-operation estimation*, Speech Database and Assessments (Oriental COCOSDA), 2011 International Conference on, pp.116-119 (2011)
- [2] Araki, S., et al. : *Low-latency meeting recognition and understanding using distant microphones*, Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on, pp.151,152 (2011)
- [3] Matsui, Tomoko, and Kunio Tanabe.: *Comparative study of speaker identification methods: dPLRM, SVM and GMM*, IEICE transactions on information and systems 2006, pp.1066-1073 (2006)
- [4] Reynolds, D.A., Rose, R.C.: *Robust text-independent speaker identification using Gaussian mixture speaker models*, Speech and Audio Processing, IEEE Transactions on , vol.3, no.1, pp.72,83 (1995)
- [5] Chen, Shi-Huang, and Yu-Ren Luo.: *Speaker verification using MFCC and support vector machine.*, Proceedings of the International MultiConference of Engineers and Computer Scientists. ol.1. (2009)
- [6] SVM-Light,  
入手先 <<http://svmlight.joachims.org/>>
- [7] 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)
- [8] 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. ICServ2013 (2013)
- [9] 天目隆平, et al. : *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, pp.443-448 (2010)
- [10] Noma, H.; Ohmura, A.; Kuwahara, N.; Kogure, K.: *Wearable sensors for auto-event-recording on medical nursing - user study of ergonomic design*, Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on , vol.1, pp.8-15 (2004)