E-021

作業推定に向けた環境雑音のクラス分類

Classification of Environmental Noises for Service Operation Estimation

宇野 太久哉† 竹原 正矩† 田村 哲嗣† 速水 悟† 蔵田 武志‡
Takuya Uno Masanori Takehara Satoshi Tamura Satoru Hayamizu Takeshi Kurata

1． まえがき

近年，サービス業における従業員の業務分析の方法として行動計測や作業推定などの客観的手法が取り入れられている[1][2]．これは，第三者による観測や経験による指導といった主観的なものと比べ，効果的な改善が期待されており，サービス工学の分野でも注目されている．

我々は，レストランの従業員の業務改善に取り組んでいる．レストランの従業員の行動データや顧客の注文情報が含まれる業務データを特徴量とした作業推定を行っている[3]. 行動データには，従業員に装着したセンサによる位置・動作データやマイクロフォンによる音声データが含まれる．

従業員の作業推定には，ビデオカメラの動画像やウェアラブルセンサによる加速度データを用いている研究が多い．しかしながら，顧客がいる場所ではプライバシーの観点からビデオカメラでの撮影が難しく，コストがかかるなどの問題点が挙げられる. 音声データは，マイクロフォンを従業員に装着することでその導入のコストを抑えることが可能であり，レストランのようなビデオカメラを用いた撮影が困難な現場でも，発話や発生する雑音から行動や状況に関する情報を得ることが可能である．作業推定に使用されている音声データの特徴量として主に発話量があり，既にその有用性が示されている[3]．しかしながら，環境雑音の情報を特徴量として使用している研究は少ない．

環境雑音を行動推定に使用する方法として，音声信号の発生頻度を利用して音源情報を推定する技術がある[4]．環境雑音の種類を分類して，その頻度情報を用いることで推定が行われている. 行動を推定するためには，複数の雑音の情報を組み合わせる必要がある．推定区間内に含まれる雑音の組み合わせと頻度からモデルを学習し，行動を推定する手法が提案されている[5]．従業員の作業推定でも，特徴的な環境雑音の組み合わせや頻度の情報によって精度の向上が期待される. しかし，従来の研究では扱う環境やマイクの位置は固定されている研究が多く，従業員に装着されたマイクロフォンのように周囲の環境が変化する状況は考慮されていない. 周囲の環境が変化する場合，行動以外が原因となって発生する雑音の種類が増加する．実環境での作業推定に活用するためには，環境による雑音と作業による雑音の違いに注目する必要がある．環境雑音の違いを考慮した手法として日常音と非日常音を別々にモデル化する手法が提案されている[6]．しかし，日常音と非日常音の分類は人手によって行われており，学習されていない雑音に対して分類する指標は定義されていない．実環境で使用するためには，音響的特徴量による指標で分類できる違いに着目する必要がある．

†岐阜大学
‡産業技術総合研究所

表 1．定常・突発雑音の例

雑音の分類	雑音の具体例
定常雑音	客室の BGM，笑い声，水流音，衣擦れの音，ふすまの開閉音
突発雑音	陶器を置く音，POS の操作音，足音，台車の音，インカム操作音

そこで本稿では，音響特徴量の突発性に注目し，従業員の音声データの非発話区間を突発・定常の 2 クラスに分類する．次に非階層クラスタリングによりさらに細かく分割する. そして，それぞれのクラスタの頻度ベクトルを作業推定の特徴量として，作業推定の精度向上を図る．

2 ． 作業推定と環境雑音

2.1 SOE の概要

SOE（Service Operation Estimation）とは，従業員の作業を推定する手法[3]であり，センサやマイクロフォンにより計測した従業員の位置・音声データや，顧客の注文・会計データのようなデータから従業員の行動を推定する手法である. SOE を現場の業務改善に活用するためには精度の向上が必要であり，推定に使用する特徴量の検討が必要である．従来，音声データの特徴量としては発話量が使用されている[3].

2.2 環境雑音

本研究では，環境雑音をマイクロフォンの周囲で発生する雑音と定義する．環境雑音はマイクロフォンを装着した従業員の周囲の環境に依存するため，位置や作業毎に発生しやすい環境雑音が異なる. 従って，環境雑音の種類や発生頻度を特徴量とすることで作業推定の精度向上が期待される．

環境雑音には突発性が低い定常的な雑音（以下定常雑音）と突発性の高い雑音（以下突発雑音）が存在する．レストラン内で発生する定常雑音と突発雑音の例を表 1 に示す． 突発雑音は発生から減衰までが短い．その割合は環境雑音全体の 1 割未満である．しかし，その多くはマイクロフォン装着者の作業によって発生する雑音であり，SOE の特徴量として有用であると考えられる．そこで本研究では環境雑音を突発雑音と定常雑音に分類する．

3. 提案特徴量

3.1 特徴量の概要

環境雑音を行動推定に使用する方法として，発生頻度を利用して音源情報を推定する技術があり[4]，複数の環境雑音の組み合わせと頻度の情報から料理や掃除などのリビング周辺の生活行動を推定している[5]．従業員の作業推定においても作業ごとに発生しやすい環境雑音やその頻度が異なるためこの手法を応用できると考えられる. しかし，マイクが移動する場合は，雑音の種類が増え推定に有用な環境

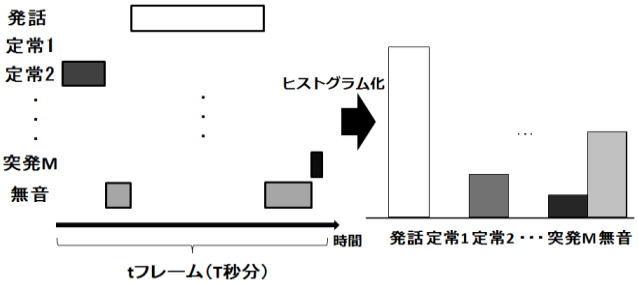

図 1. 提案特徴量の概要

雑音を分類することが困難になると考えられる．特に，割合が少ない突発雑音をクラスタリングにおいて別に考えることにする．このため，本研究では音声データを発話 1 クラス，定常雑音 N クラス，突発雑音 M クラス，無音 1 クラスの計 N+M+2 クラスに分類し，推定区間 T 秒内の頻度を求め，N+M+2 次元の特徴量とする．提案特徴量の概要を図 1 に示す．

3.2 全体の流れ

本稿で提案する音声データを用いた SOE 特徴量算出を目的とした音声データ分類の流れを図 2 に示す．まず，音声データをフレーム化し，発話区間に相当するフレームはすべて発話クラスとする．非発話区間に対して有音・無音を判別し，無音と判別されたフレームは無音クラスとする．続いて，残りの有音フレームを音響特徴量の尖度によって定常・突発の 2 種類に分類する．さらに非階層クラスタリングを行うことによって定常雑音クラスは N 個，突発雑音クラスは M 個のクラスタに分割する．以上の流れで得られた発話クラス，定常雑音 N クラス，突発雑音 M クラス，無音クラスを SOE の特徴量に使用する．

3.3 発話・無音区間の除去

音声データ中の発話区間を人手でラベル付けし，非発話区間を抽出する．そして，非発話区間において，フレーム単位で式(1)に示す対数パワーを求め，ある閾値を超えるか否かで，雑音クラスか無音クラスかを判定する．雑音クラスは 3.4 節でより詳細にクラスタリングする．なお，$s_i(l)$はiフレーム目の音声信号のl番目の値，Lはフレームのポイント数である．

$$POW_i = 10 \cdot log_{10}\left(\frac{1}{L}\sum_{l=0}^{L-1} s_i^2(l)\right) \quad (1)$$

閾値として，発話区間の 9 割が有音となる値を求め使用した．

3.4 定常・突発の分離

雑音と判定されたフレームの音声信号sに対し，式(2)に示す 4 次キュムラントを尖度の指標として分離する．

$$Cum_4^s = \frac{E[s^4(l)]}{E[s^2(l)]^2} - 3 \quad (2)$$

突発雑音は発生が瞬間的であり，定常雑音と比較してその割合は少なく，雑音全体の 1 割未満であった．2.2 節で述べたように突発雑音は作業推定に有用であると考えられるため，尖度の上位 1 割を突発雑音，残り 9 割を定常雑音とした．

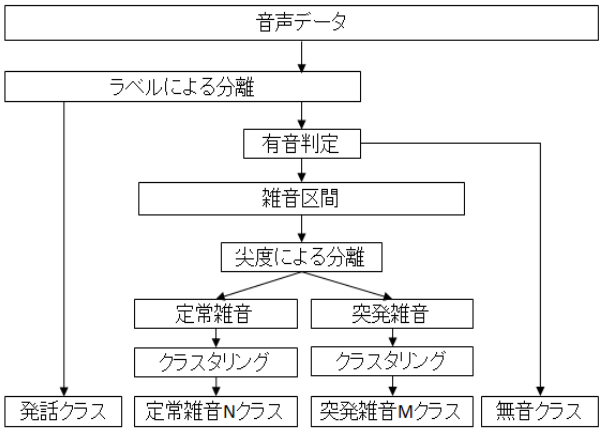

図 2. 音声データ分類の流れ

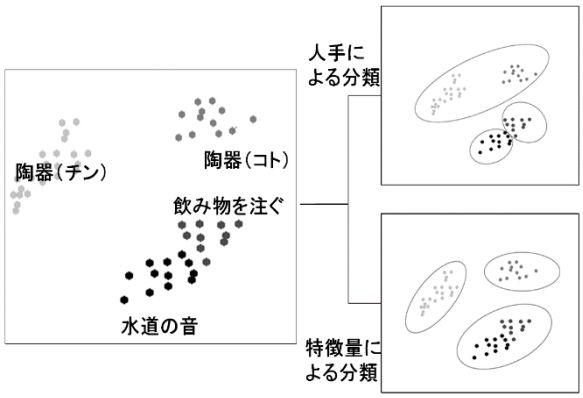

図 3. 人手，特徴量的分類の概念

3.5 雑音のクラスタリング

定常・突発に分類された雑音それぞれに対して，混合分布モデルによる非階層クラスタリングによっていくつかの特徴量的なまとまりに分割する．非階層クラスタリングによる分類は人手による分類と異なり，意味的なまとまりは存在しない．人手による分類と特徴量的な分類の概念を図 3 に示す．

図 3 において，水道の水が流れる音と飲み物を注ぐ音は特徴量的には似ているが人が捉えるまとまりとしては異なる．一方で陶器の音のように，複数の特徴を持つ雑音に関しては人が分類する場合は同一と分類するが，特徴量的な分類では別に分類される．特徴量的な分類をするメリットとして雑音のクラス数を自由に設定することが可能であり，雑音の種類が未知な場合にも対応が可能であることが挙げられる．デメリットとしては，クラスごとの雑音の意味が明確でないため，現象の解析に向かないことなどが挙げられる．

3.6 特徴量の作成

T 秒間における各セグメントに対して分類手法に基づいて分類されたクラスを割り当て，クラスごとに集計し区間のフレーム数 t で正規化することで各クラスの割合を求める．作成された N+M+2 次元の頻度ベクトルを環境雑音の特徴量とする．

表 2. 作業の種類

No.	作業内容	概要
1	挨拶・案内	客席までの案内・挨拶
2	移動・運搬	移動や料理の運搬
3	会計	レジ・客室での清算処理
4	注文伺い	注文を受け，端末操作
5	配膳	料理・ドリンクの配膳
6	片付け・セッティング	片付け，予約席の準備
7	会話	上記作業以外の会話

表 3. 実験条件

使用データ数	1136 サンプル
装着従業員数	2 名分
サンプリング周波数	16kHz
フレーム長	25msec
フレームシフト	10msec
特徴量	MFCC39 次元（12 次元＋パワー1 次元の計 13 次元及びそのΔ，ΔΔ）
1 サンプルの長さ T	5sec
1 サンプルのフレーム数 t	500 フレーム
クラスタ数 N,M	N=64,M=32

4. 評価実験

4.1 精度評価実験

提案する特徴量の有用性を示す実験として，従来の研究[3]を参考に表 2 に示す 7 種類の作業の推定を行い，推定精度を評価した．表 2 の No.は後述の Service Operation Number に対応する．

4.1.1 実験条件

実験条件を表 3 に示す．学習・評価に使用したデータは計 1136 サンプル（1 時間 34 分 40 秒分）である． closed 条件は学習と評価に同じデータを使用し，open 条件は 1 サンプルを評価データ，残りのサンプルを学習データとする Leave-one-out 法で行った．評価尺度は，作業ごとに適合率と再現率の調和平均（以下，識別精度と呼ぶ）を用いた．クラスタ数 M，N は 2 のべき乗の値で最適な精度となる値を採用した．

4.1.2 実験結果

closed 条件，open 条件での作業毎の識別精度を図 4 に，open 条件における識別結果を表 4 に示す．なお，表 4 における割合は正解作業に対する割合である．全作業の平均の識別精度は closed 条件で 81.8%，open 条件で 44.0%となった．

4.1.3 考察

配膳（SO No.5）や片づけ・セッティング（SO No.6）の推定精度が比較的高い結果となっている．これは陶器の音や客席の BGM といった特徴的な雑音が多いことが理由として考えられる．一方で，注文伺い（SO No.4）は精度が低く，その多くが会話（SO No.7）や配膳（SO No.5）に誤推定されていた．注文伺いには特徴的な環境雑音が少なく他の SO との違いが少ないためうまく分類ができず，発話部分が会話として，その他の部分が客席に多い雑音の影響を受けて配膳として誤推定されてしまったと考えられる．誤判定の傾向として会話が含まれる作業が会話（SO No.7）として誤推定されている．これは区間における雑音の割合が少なく，発話量のみでの判断となったためであると考えられる．改善のためには発話に関する特徴量の検討が必要であり，環境雑音の情報から改善するのは困難であると考えられる．

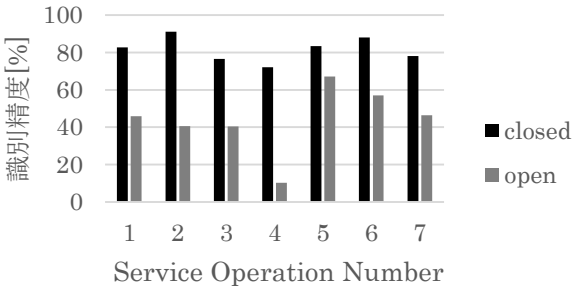

図 4. 環境雑音特徴量による識別精度

表 4. open 実験の推定結果

		推定された作業						
		1	2	3	4	5	6	7
正解の作業	1	44.4%	2.6%	7.3%	6.0%	6.6%	12.6%	20.5%
	2	6.0%	37.3%	1.5%	3.0%	14.9%	29.9%	7.5%
	3	17.9%	0.9%	35.9%	2.6%	6.0%	12.8%	23.9%
	4	11.6%	5.0%	7.4%	7.4%	38.0%	9.1%	21.5%
	5	3.9%	1.9%	2.7%	2.7%	76.0%	5.4%	7.4%
	6	2.2%	4.5%	3.1%	3.6%	17.5%	58.7%	10.3%
	7	10.1%	2.5%	7.0%	8.5%	8.5%	13.1%	50.3%

4.2 従来の特徴量との比較実験

従来の特徴量に提案特徴量を加えた場合の有用性を示す実験として，作業推定を行い，推定精度を比較した．

4.2.1 実験条件

実験条件は前実験と同様，表 2，表 3 に従う．推定に使用する特徴量として音声データの特徴量に加えて，センサデータと業務データから抽出した特徴量 18 次元を用いる．センサデータからは位置，方位，加速度をもとに，区間におけるエリア毎の滞在割合 8 次元，向いている方位の割合 5 次元，歩数 1 次元を求めた．業務データからは注文発生件数や客数などの 4 次元のデータを求めた．音声データの特徴量は，使用しない（none），従来法である発話量 1 次元（SR），提案特徴量 N+M+2 次元（NR）の 3 種類で closed 条件と Leave-one-out 法による open 実験を行い精度の比較をした．評価尺度は前実験と同様に作業毎の適合率と再現率の調和平均を用いた．

4.2.2 実験結果

closed 条件，open 条件での作業毎の識別精度を図 5，図 6 に open 条件における識別結果を表 5 に示す．closed 条件における全作業の平均の識別精度は音声データを使用しない場合が 60.0%，発話量を使用した場合が 60.3%，提案特徴量を使用した場合が 72.3%となった．open 条件では使用しない場合が 37.7%，発話量を使用した場合が 40.4%，提案特徴量を使用した場合が 51.6%となった．

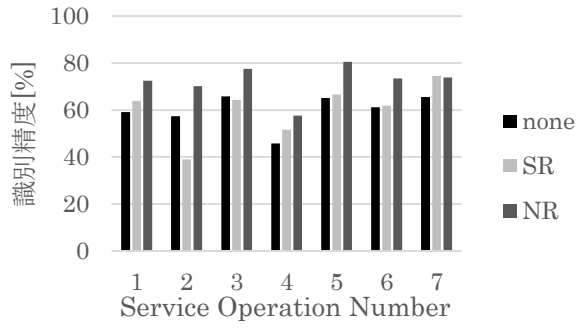

図 5. 識別精度（closed 条件）

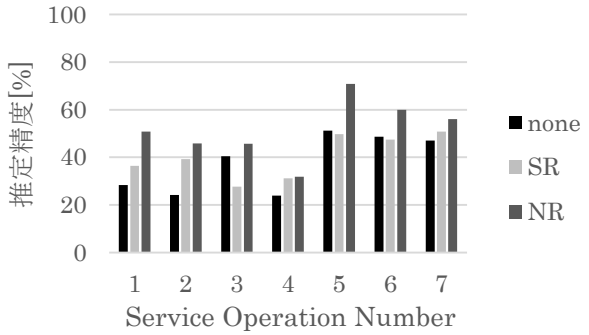

図 6. 識別精度（open 条件）

4.2.3 考察

音響特徴量として提案特徴量を用いることにより，発話量のみを使用した場合と比較して closed 条件で 12.0%の向上が得られた．open 条件でも，11.2%と同様の精度向上が見られた．発話量と比較して精度が向上した理由として，発話量は発話の有無のみの判別を行っているため，発話の内容を扱うことができず，作業による違いを十分に判別出来ないのに対して，提案特徴量は複数の雑音の組み合わせの情報があり，判別可能になったためだと考えられる．

5. まとめ

本稿では，作業推定（SOE）に使用する音声データの特徴量として環境雑音の特徴量化を検討した．突発性によって定常・突発に分類し，それぞれを複数のクラスにクラスタリングすることで雑音の種類を分類し，発生頻度を特徴量として使用した．従来の特徴量である発話量と比較して作業推定の精度を 11.2%改善した．提案特徴量は特に特徴的な雑音が発生しやすい作業の推定に向いていると考えられる．

今後の課題として，環境雑音の分類に用いる音響特徴量の検討が挙げられる．本稿では音響特徴量として MFCC を使用したが，この特徴量は一般には音声認識など発話の分析に用いられる．雑音の特徴をよく表現する特徴量を用いることで，クラスタリングした際により意味のあるクラスが生成され，作業推定の特徴量としての有用性も高まると考えられる．また，本稿では環境雑音を定常雑音と突発雑音の 2 種類に分割したが，雑音の突発性にもう少し着目していくべきである．さらに，発話クラスの扱いや発話量との組み合わせも検討していきたい．

表 5. open 実験の推定結果

		推定された作業						
		1	2	3	4	5	6	7
正解の作業	1	50.3%	4.6%	6.6%	9.3%	9.3%	7.3%	12.6%
	2	6.0%	40.3%	6.0%	4.5%	7.5%	29.9%	6.0%
	3	13.7%	1.7%	42.7%	4.3%	1.7%	18.8%	17.1%
	4	9.9%	2.5%	4.1%	25.6%	32.2%	11.6%	14.0%
	5	4.3%	1.6%	2.7%	2.7%	78.3%	4.7%	5.8%
	6	2.7%	4.5%	4.5%	3.1%	15.7%	61.9%	9.4%
	7	11.6%	1.0%	8.0%	3.5%	8.0%	10.1%	57.8%

謝辞

本稿の執筆にあたり，データ収集にご協力頂きましたがんこ銀座 4 丁目店の従業員の皆様に感謝致します．

参考文献

[1] 赤松幹之，新井民夫，内藤耕，村上輝康，吉本一穂，サービス工学―51 の技術と実践―，朝倉書店, 2012.

[2] T. Tomiyama.”A manufacturing paradigm toward the 21st century,” Computer Aided Engineering, Vol.4, pp.159-178, 1997.

[3] R. Tenmoku, R. Ueoka, K. Makita, T. Shinmura, M. Takehara, S. Tamura, S. Hayamizu, T. Kurata: ”Service-Operationestimation in a Japanese restaurant using multi-sensor and POS data,” Proc. Adbances in Production Management Syastems 2011 Conference, Parallel 3-4: 1, Stavanger, Norway, Sep. 2011.

[4] P. Smaragdis, B. Raj,“Topic Molels for Signal Processing”, IEEE International Conference on Acoustics, Speech and Signal Processing 2011

[5] 井本桂右，野口賢一，島内末廣，大室仲，羽田陽一，“音響イベントを用いた人の行動の確率的生成モデルによる行動識別手法の検討,” 日本音響学会研究発表会講演論文集（春）, pp.825-826, 2013.

[6] 小川順平，林田亘平，森勢将雅，山下洋一，“マルチステージ環境音識別に関する検討,” 日本音響学会研究発表会講演論文集（秋）, pp.101-102, 2011.