様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ６ 日現在

機関番号：１５３０１
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１２０７９

研究課題名（和文）観光地の雰囲気可視化を可能とする簡易なアノテーションに基づく深層学習方式の研究

研究課題名（英文）A study of a deep learning method based on simple annotation that enables visualization of the atmosphere of tourist attractions

研究代表者
原　直（Hara, Sunao）

岡山大学・ヘルスシステム統合科学学域・助教

研究者番号：５０４０２４６７

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：地域特性のパラメータ化を進めるため，特に，ISO 12913 で標準化されているサウンドスケープの考え方を取り入れた．環境音聴取時に，ストリートビューの映像も同時に提示することで，音だけに依存しない場の印象や雰囲気にアノテーションを付与した．
そして，DNNによる音からの地域特性の分類器に関する実験を行った．入力に音源情報を併用することで，推定精度は向上する．ここで，人手による音源情報ではなく，位置情報から自動取得可能な航空写真を利用するだけでも，一定の精度向上が見られることを確認した．
さらに，コンセプトドリフトに基づく機械学習モデルの適応方式に関する研究も進めた．

研究成果の学術的意義や社会的意義
地域特性を表現するために，サウンドスケープの考え方を取り入れた．標準化された仕様を採用することは，より広範なデータ収録を容易とする．人の主観に基づくデータ収集において，基準がわかりやすいことは重要である．
DNNに基づく地域特性の推定器は，音データさえ収録できれば，人手による詳細アノテーションが十分に無くとも，自動的に取得可能な情報源から，一定の推定精度が得られる．DNNには，多くの学習データが必要である．簡易アノテーションのみで十分という事実は，低コストに大量データを集めるための重要な知見である．また，継続的なデータ収集とシステム運用に向け，コンセプトドリフトに基づく研究から多くの知見を得た．

研究成果の概要（英文）：In order to parameterize area characteristics, the concept of soundscape, which is standardized in ISO 12913, was adopted. By presenting images from Google's Street View at the same time as listening to environmental sounds, we annotated the impression and atmosphere of a place that is not dependent on sound alone.
Then, we conducted experiments on a DNN-based predictor for area characteristics based on sound. The prediction accuracy is improved by using sound source information as input. Moreover, we confirmed that the accuracy is improved by using aerial photographs, which can be automatically obtained from location information, instead of manual sound source information.
Finally, we also studied adaptive methods for machine learning models based on concept drift.

研究分野：情報学

キーワード：サウンドスケープ　地域特性　可視化　音響シーン分類

様　式　C－１９、F－１９－１、Z－１９（共通）

１．研究開始当初の背景

音環境は我々の生活に密接に関係している．実際，我々は様々な場において，様々な音を聞き，その場の状況を理解している．このような音環境理解を機械に行わせる試みは CASA（Computational Auditory Scene Analysis)と呼ばれ，盛んに研究が進められている．CASA は，スマートシティ実現のために利用されることもある．例えば，騒音公害（Noise Pollution）の可視化や，危険音検知（Abnormal Sound Detection）などの応用である．

CASA のためには，環境音データベースが必要となるが，その構築は困難である．環境音データベースには，環境音のほかに「データ中の，どの時間に，どの音が鳴っているか？」というアノテーションが必要である．しかし，このような詳細なアノテーションを付与する作業は，作業者にも専門知識が必要で，かつ，作業そのものの負荷も高い．そこで，通常より簡易な情報に限定したアノテーションから CASA が実現できれば，環境音データベースの構築コストが下がるため，　CASA を様々な場面で応用することが容易となる．

本研究では，CASA の具体的な応用として，観光地を含む周辺地域一帯の雰囲気を環境音だけで可視化するシステム(図 1)を実現する．観光客が観光地を調べる一助となるような，雰囲気の可視化を目指す．観光客の利便性向上から，観光客の増加と地域活性化につながることを期待する．

環境音とアノテーションは，クラウドソーシングによるモバイルセンシング，すなわち，クラウドモバイルセンシングの枠組みで収集する．ただし，クラウドモバイルセンシングでは，作業者の手作業を必要とするアノテーションの品質はバラツキが大きい．そこで，簡易なアノテーションとして「どの音」だけを付与させ，また，付与する音の種類も少数に限定して選びやすくすることで，アノテーションの品質を担保する．

本研究の最大の課題は，このような簡易なアノテーションに基づく CASA の実現である．

２．研究の目的

本研究では，以下の実現を目的として研究を開始した．(1) 簡易なアノテーションと環境音を収集するクラウドモバイルセンシングを実現し，(2) 簡易なアノテーションを教師とする深層学習モデルにより，地域の雰囲気を可視化した環境音地図を構築し，(3) 環境音地図の具体的な応用として，観光地の雰囲気の可視化システムを実現し，その有効性を実験的に評価する．

３．研究の方法

(1) 実験に用いるデータの収録は，Android 端末で動作するアプリケーションで行われている．音を保存する際にはその場の音を主観的に評価するために騒音度と混雑度の 5 段階評価を与える事ができる．さらに，その場の状況や聞こえてくる音を記録するために，任意入力のテキストエリアと 12 種類のあらかじめ設定された音の種類が用意されている．アプリケーション内の送信ボタンが押されると，直近 10 秒または 15 秒の音声波形を WAV ファイルとして作成すると同時に，主観評価の選択項目とテキスト入力の内容を表す文字列に時間をつけてログファイルに記録する．

環境音は，量子化ビット数 16 bit，サンプリングレート 32,000 Hz で収録が行われる．1 秒毎に得られる環境音サンプルに対して A 特性に基づくサウンドレベル LAeq を計算する．また，8 帯域のオクターブバンドフィルタ出力 BP（中心周波数 62.5 Hz，125 Hz，…，8,000 Hz）も同時に計算している．主観的な騒音度合いの選択肢は，L_1:とても静か，L_2:比較的静か，L_3:やや騒がしい，L_4:かなり騒々しい，L_5:とても騒々しい，の 5 段階である．

環境音は，岡山県岡山市と同県倉敷市で収録されたデータを利用した．岡山駅前の賑わった環境だけではなく，商店街や住宅街などの比較的静かな地域で収録している点が特徴である．倉敷市では，静謐な観光地である倉敷美観地区において，収録している．特に，収録日の一部においては，同地区内にある倉敷阿智神社を中心とした秋季例大祭が催されており，多くの観光客による賑わいの様子が収録されている．

本研究において主として実験に用いたデータは，合計で 11,993 個ある．主観的な騒音度合いごとの個数としては，L_1 から L_5 の 5 段階について，それぞれ，1,269 個，4,440 個，4,635 個，1,356 個，293 個である．

(2) 主観評価値の推定には，LAeq に加えて，収録の観点別に分けた三種類の特徴量セットを組み合わせて用いる．1 つ目は，収録場所の特性を表現する特徴量 LS（location specific）である．LS には，場所・時間帯・イベントの有無が含まれている．LS は収録者による操作を必要とせず，GPS などのデータから自動的に収集可能な特徴量セットである．2 つ目は，収録者の特性を表現する特徴量 PS（participant specific）である．PS には，収録者 ID・移動収録の有無が含まれている．PS は，収録者 ID に関して収録の開始時にログインなどの特定の操作により，収録者を特定する必要があるため，LS に比べて収集コストがやや高い．3 つ目は，音源の特徴量 SS（sound specific）である．SS は，11 種類の音源ラベル情報である．SS は，10 秒の音の収録を

行なうたびにアプリケーション上で環境音の種類を選択する必要があるため，本章で用いる特徴量セットの中で最も収集コストが高い．なお，本章では 12 種類の環境音の種類のうち，極端にデータ数の少なかった「踏切」を除く 11 種類を音源ラベル情報としている．これらの特徴量を DNN の入力層に与える際には，サウンドレベルは平均 0 分散 1 となるように正規化し，その他の特徴量については 1-of-k 表現を用いている．

DNN は，各層 50 ノード，5 層からなる中間層で構成される．活性化関数は ReLU 関数である．推定結果と真値の平均二乗誤差を損失関数とし，ミニバッチに基づく Momentum SGD による学習（学習率 0.07，momentum 0.09）を行う．出力層は，線形層からの連続値の出力を四捨五入により，5 段階としている．

(3) サウンドスケープに基づく地域の雰囲気可視化を進めるため，収録済みの音声データに対するアノテーション作業をおこなう．アノテーションの基準として，Axelsson らの研究に基づいた SSQP（Swedish Soundscape-Quality Protocol）を採用している．8 種類の印象評価語を用いた音環境に対する印象の評価を提案しており，ISO12913-2 でも SSQP に基づいた印象の評価方法を例として定めている．本研究では，永幡らによる先行研究を参考にしながら，8 種類の評価語に日本語の単語を 2 つずつ割り振り，評価することとしている．

(4) 10 秒の環境音から得られた各種特徴量から，サウンドスケープとしての印象情報を推定する．航空画像は，音の収録位置を元に Bing Maps から取得したレベル 20 の航空写真を用いる．印象推定をするため，いくつかの DNN モデルを試作し，その推定精度を検証する．研究終了時点においては，図 1 に示す構成の DNN モデルを基本として，そのハイパーパラメータを変えながら利用することとしている．

音響特徴量(ES)は，サウンドレベルとオクターブバンドパスフィルタ出力を利用する．このとき，1 秒ごとの値に加えて，統計量として，平均値，10%-tile，50%-tile，90%-tile も計算することで，126 次元の特徴量とする．画像特徴量(AP)は，224×224 pixel の航空画像から，ResNet-50 を用いて，2048 次元の画像特徴量とする．音源特徴量(SS)は，アノテーションで付与された 7 次元の値である．これらの特徴量を組みあわせて，DNN による推定を行う．

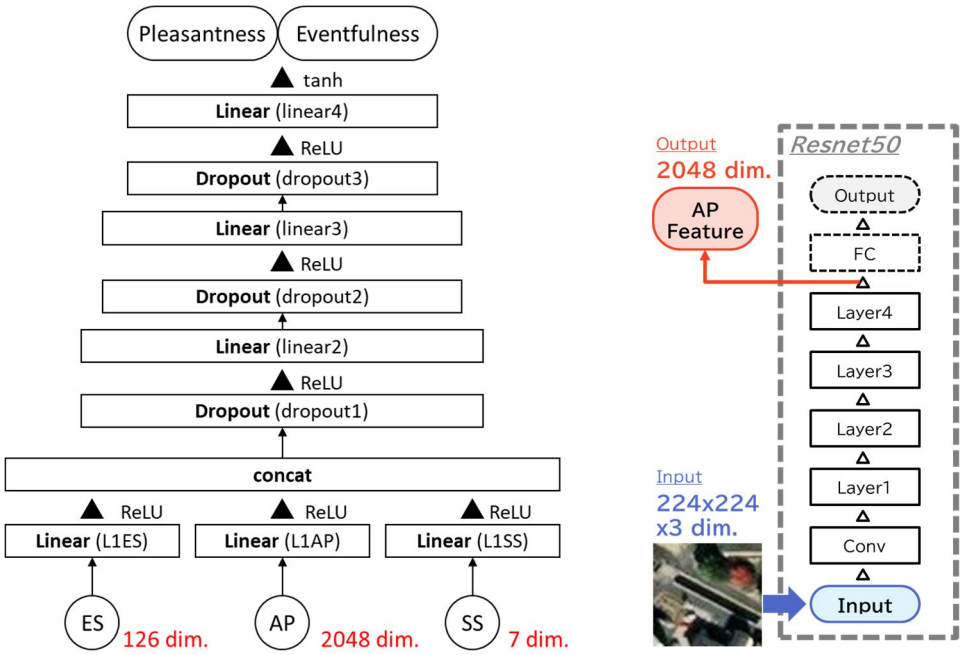

図 1　　ネットワーク構成．全体構成(左)と ResNet を用いた AP 特徴量の抽出部(右)

(5) 環境音のシーン分類タスクにおけるコンセプトドリフト条件の影響と対処に関する研究を行う．環境音のシーン分類は，音の雰囲気を表現するための一つとして有用であると考えられる．また，コンセプトドリフト条件を考慮した分類実験を進めることで，DNN のような機械学習に基づく推定システムの実用化に向けた重要な知見が得られる．

本研究では，コンセプトドリフトの検出と処置のための方式として CMGMM&KD3 algorithm を提案する．KD3 (Kernel density drift detector)は，カーネル密度推定に基づくアルゴリズムであり，コンセプトドリフトの検出に利用する．CMGMM (Combine-merge Gaussian mixture model) は，混合ガウス分布(GMM) による識別方式を基盤として，学習済み GMM モデルに新しいデータを追加する際のモデル更新（結合/新設）戦略を制御するアルゴリズムである．CMGMM アルゴリズムの概略を図 2 に示す．

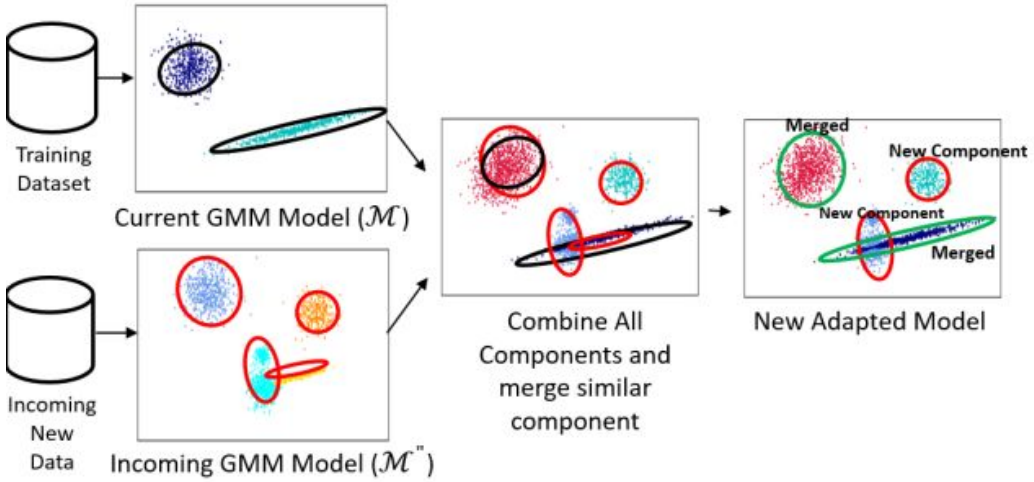

図 2　　CMGMM アルゴリズムの概略

４．研究成果

(1) 客観的な騒音マップと主観的な騒音マップ

収録されたサウンドレベル (Sound level) と主観的な騒音度 (SLL) を地図上に可視化した騒音マップを図 3 の(a)と(b)に示す．なお，本報告では約 30 メートル四方の区画ごとに，値の最頻値を可視化することで騒音マップを作成している．色が赤に近いほどサウンドレベルが大きい，あるいは，主観的な騒音度が高い区画である．

図 3(b)は，収録者が選択した騒音度を可視化したマップであり，本報告の目標とする人間の感覚を考慮した騒音マップである．図 3(a)と図 3(b)を比較すると，図の左端や右端において大きく結果が異なる地域が存在する．このことから，等価騒音レベルのみを単純に可視化しただけでは，人間の感覚を考慮した騒音マップにはならないと考えられる．

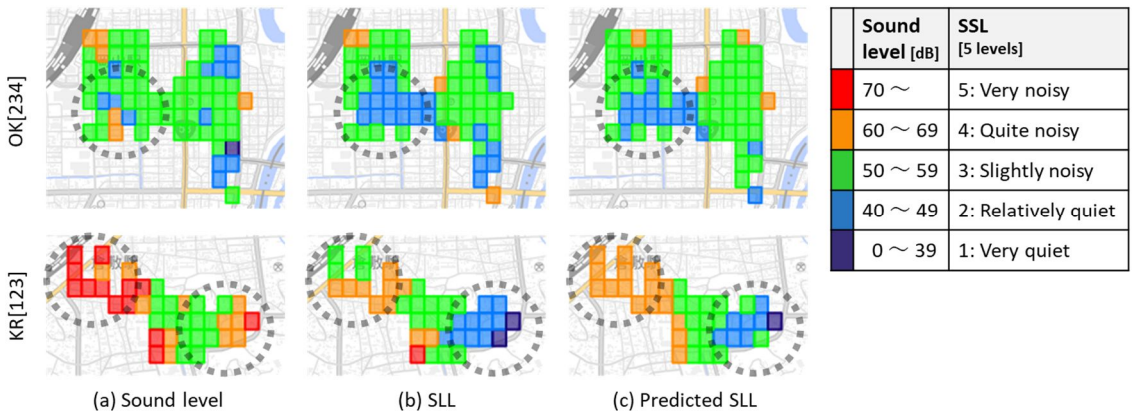

図 3　　騒音マップ

(2) 主観的な騒音度合いの推定

推定結果は F-measure で評価する．10 分割交差検証を行っており，分割後の主観的な騒音度合いの分布が元の分布とほぼ等しくなるように分割する．

特徴量の組み合わせによる推定結果の F-measure の平均値を図 4 に示す．L_2 を除けば，推定に用いる特徴量の種類を増加させるほど，推定精度が良くなる．ただし，LS の効果は限定的である．一方，L_1, L_4, L_5 における PS 追加の精度向上から，個人による感覚の違いが示唆される．また，L_1 における SS 追加の精度向上から，静かな環境における音源ラベル情報の重要性が示唆される．

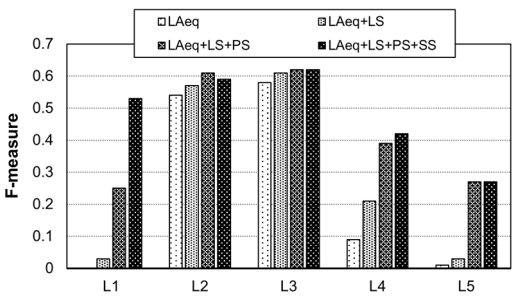

図 4　　主観的な騒音度合い推定の平均

また，推定された騒音度に基づく可視化結果は，前掲の図 3(c)に示している．従来法として示した図 3(a)と比較しても，所望の図 3(b)に近い傾向を示していることがわかる．

(3) サウンドスケープの考え方を導入したアノテーション作業

サウンドスケープに基づく地域の雰囲気可視化を進めるため，収録済みの音声データに対するアノテーション作業を行った．①音源情報は，交通関係の音，その他技術由来の音，人の声を含む音，その他人間由来の音，生物が発する音，その他自然界の音，収録雑音，の 7 種について，7 段階で評価した．②印象情報は，Pleasant, Eventful, Calm, Vibrant, Annoying, Uneventful, Chaotic, Monotonous の 8 種類について 7 段階で評価した．③本研究では利用していないが，SSQP に記載のある周囲の音環境の良し悪しや，周囲の音環境の適切さ，も評価されている．

最終的にアノテーションを付与したデータ数は，3146 個である．アノテーション作業は，のべ 6 名で行われている．アノテーション対象となった環境音は 1,560 個あり，各環境音に対して 2 名以上によるラベル付与がなされるように意図して，アノテーション作業がなされた．

(4) サウンドスケープ印象の推定

サウンドスケープとしての印象情報を，DNN により推定を行う．DNN のハイパーパラメータは Optuna ツールキットを利用して，最適なハイパーパラメータを探索している．推定する印象情報は，ISO12913-3 で定義されている式)に基づいて -1～+1 に変換された，Pleasantness と Eventfulness の 2 種の値を推定することとした．

推定結果を図 5 に示す．評価指標は，R^2 決定係数である．Pleasantness と Eventfulness のいずれも，推定結果はおおむね同じ傾向を示している．ES 単独に比べて，ES&AP&SS, ES&SS, ES&AP のいずれも性能は良い．このことから，SS や AP をサウンドスケープの印象推定に用いる有用性が示唆される．

続いて AP の有用性について考える．ES&AP&SS と ES&SS の差に比べれば，ES&AP と ES の差は大きい．このことから，AP は推定に有効ではあるものの，SS が利用できる場面での有用性は下がることが示唆される．しかし，SS には人手によるアノテーションが必要である．AP は，SS とは異なり自動的に取得できる特徴量である．SS が利用できない場面でも，AP が利用できることを考慮すれば，十分に利用価値のある特徴量であるといえる．

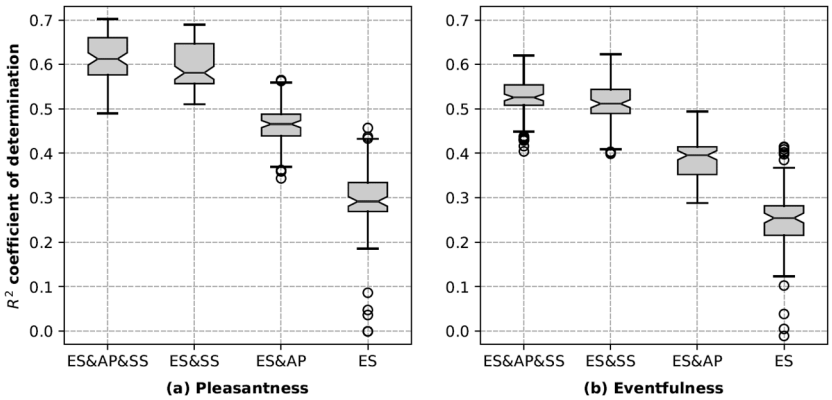

図 5　　サウンドスケープ印象の推定結果

(5) コンセプトドリフトの検出と処置に関する実験結果

音響シーンの分類タスクにおけるコンセプトドリフトの検出と処置に関する実験結果を表 1 に示す．コンセプトドリフトが発生しうるシミュレーションシナリオを 3 パターン用意して実験を行った結果，提案方式である CMGMM & KD3 algorithm は，T2 と T3 という 2 つのシナリオで高い性能を示した．残りの 1 つである T1 シナリオにおいても，従来法と同程度の精度を示していることから，提案方式の有効性が示された．

表 1　　コンセプトドリフトの検出と処置に基づく音響シーンの分類精度

ACTIVE APPROACH											
ADAPTOR	DETECTOR	ACCURACY			F1 SCORE			EXECUTION TIME			Drift
		T1	T2	T3	T1	T2	T2	T1	T2	T3	
CMGMM*	KD3*	0.8373	**0.7962**	**0.7409**	0.8432	**0.7993**	**0.7460**	128.06	115.07	110.49	39
	ADWIN	**0.8401**	0.6415	0.6332	**0.8418**	0.6379	0.6379	83.07	84.22	85.44	23
	HDDM	0.2762	0.2627	0.2990	0.3184	0.2992	0.3406	84.81	84.53	83.11	373
IGMM	KD3*	0.8283	**0.7574**	**0.6622**	0.8173	**0.7499**	**0.6488**	120.04	128.75	120.50	35
	ADWIN	**0.8419**	0.5711	0.6057	**0.8423**	0.5722	0.6063	82.80	84.219	83.08	21
	HDDM	0.2363	0.2507	0.2032	0.2436	0.3055	0.2675	84.37	87.55	84.87	350

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 原直，阿部匡伸	４．巻 46
２．論文標題 機械学習による環境音からの主観的な騒音マップ生成	５．発行年 2022年
３．雑誌名 騒音制御	６．最初と最後の頁 619-623
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Id Ibnu Daqiqil、Abe Masanobu、Hara Sunao	４．巻 6
２．論文標題 Acoustic Scene Classifier Based on Gaussian Mixture Model in the Concept Drift Situation	５．発行年 2021年
３．雑誌名 Advances in Science, Technology and Engineering Systems Journal	６．最初と最後の頁 167～176
掲載論文のDOI（デジタルオブジェクト識別子） 10.25046/aj060519	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　4件）

１．発表者名 Yusuke Ono, Sunao Hara, and Masanobu Abe
２．発表標題 Prediction method of Soundscape Impressions using Environmental Sounds and Aerial Photographs
３．学会等名 APSIPA Annual Summit and Conference (APSIPA-ASC 2022)（国際学会）
４．発表年 2022年

１．発表者名 Ibnu Daqiqil Id, Masanobu Abe, and Sunao Hara
２．発表標題 Incremental Audio Scene Classifier Using Rehearsal-Based Strategy
３．学会等名 IEEE 11th Global Conference on Consumer Electronics (GCCE 2022)（国際学会）
４．発表年 2022年

１．発表者名
Ibnu Daqiqil Id, Masanobu Abe, and Sunao Hara

２．発表標題
Concept drift adaptation for audio scene classification using high-level features

３．学会等名
2022 IEEE International Conference on Consumer Electronics (ICCE)（国際学会）

４．発表年
2022年

１．発表者名
小野祐介，原直，阿部匡伸

２．発表標題
環境音と航空写真を用いた場所の印象を推定する方式の検討

３．学会等名
第24回 日本音響学会関西支部 若手研究者交流研究発表会

４．発表年
2021年

１．発表者名
小野祐介，原直，阿部匡伸

２．発表標題
SSQPによる場所の印象情報を環境音と航空写真から推定する方式の検討

３．学会等名
電子情報通信学会 2022年3月LOIS研究会

４．発表年
2022年

１．発表者名
Ibnu Daqiqil Id, Masanobu Abe, Sunao Hara

２．発表標題
Concept Drift Adaptation for Acoustic Scene Classifier Based on Gaussian Mixture Model

３．学会等名
The 2020 IEEE Region 10 Conference (IEEE-TENCON 2020)（国際学会）

４．発表年
2020年

１．発表者名
平田瑠，原直，阿部匡伸

２．発表標題
GPSデータのクラスタリングによる日常生活における場所の重要度の分析

３．学会等名
マルチメディア，分散，協調とモバイルシンポジウム（DICOMO2020）

４．発表年
2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

HARA, Sunao (原 直)
https://www.a.cs.okayama-u.ac.jp/~hara/index.html

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究協力者	阿部　匡伸 (ABE Masanobu)	岡山大学・学術研究院ヘルスシステム統合科学学域・教授 (15301)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関