音響情報と位置情報を用いたライフログデータのクラスタリング

山野　貴一郎[†1]　伊藤　克亘[†2]

個人の生活や体験をカメラ，マイク，GPS などのセンサを用いて収録し利用するための研究が行われている．収録された個人体験記録はライフログと呼ばれる．ライフログは常時記録をしているためデータ量が膨大かつ冗長であり，効率的に利用をするためには，データの検索や要約が必要となる．本論文ではウェアラブルなマイクによって収録された，音響ライフログのクラスタリングや提示手法について述べる．本研究では音響ライフログを場所，話者，時間で分類することを目的としている．音響ライフログが収録された場所は GPS により取得ができるが，建物内の部屋までは取得ができない．そこで，本論文では音響ライフログをセグメントに分割し，スペクトル包絡を用いて部屋のクラスタリングをする実験を行った．実験は GPS により場所を取得したと仮定した場合と，GPS を用いない場合の 2 つである．結果，本論文の実験の範囲では GPS の利用によりクラスタリングの精度が向上することがわかった．また，音響ライフログを効率的に扱うにために，GPS を用いた地図上でのブラウジングを提案した．

Clustering of Life-log Data using Audio Information and Location Information

Kiichiro Yamano[†1] and Katunobu Itou[†2]

The use of personal life experiences recorded by cameras, microphones, GPS devices, etc. is studied. A record of a person's personal life is called life-log. Since the amount of data in a life-log system is enormous and since the data also include redundant data, methods for the retrieval and summarization of the data are required for the effective use of the life-log data. In this paper, audio life-log recorded by wearable microphones is described. A purpose of this study is classifying audio life-log according to places, speakers and times. However places of audio life-log recorded are obtained by GPS devices, rooms in buildings cannot be obtained. In this paper, experimentations were carried out about audio life-log was divided into segments and clustered by spectrum envelopes according to rooms. The experimentations are two situations that location information are captured and not captured. Results of the experimentation showed location information aid room clustering by audio infomation. Audio life-log browsing on a map using GPS is also suggested.

1. はじめに

個人の生活や体験を様々なセンサ (カメラ，マイク，GPS など) を用いて記録し，利用するための研究が行われている[1)]．このような個人の記録をライフログと呼び，備忘録や自動の日記作成，商品のレコメンドシステムやマーケティングなどへの利用が期待されている．しかし，ライフログは常時記録をしているためデータ量が膨大かつ冗長であり，そのままでは利用が難しい．従って，効率的な利用のためには要約や検索の必要があり，様々な試みがなされている．

本研究ではウェアラブルなマイクで日常生活の音を常時記録した音響ライフログと GPS によって取得した位置情報について扱う．音響ライフログは様々な音響情報を含んでおり，多くの情報が得られる．例えば，音声からはその時の会話の内容や話者情報などが得られる．また，騒音や音楽からは雑踏や店にいるなど周辺の情報が得られ，キーボード打鍵音やクリック音などの PC の操作音からはユーザの行動がわかる．しかし，音響ライフログはほぼ環境音や音声などが含まれていない部分や，含まれている音が聴取しただけでは不明な部分があり，冗長な情報源である．従って，収録をしたデータをそのまま提示しただけでは所望の情報を探すのが困難である．このような問題に対し，従来研究では時系列での情報の提示が行われてきた[5)]．本研究では，音響ライフログの音声部分に着目し，場所や話者についてデータのクラスタリングを行う．さらに，従来の時系列の情報提示だけではなく，GPS 情報を併用して地図上で時系列情報をブラウジングする手法を提案する．

2. ライフログデータの処理

2.1 先行研究

ライフログを有効に扱うための処理についてこれまで多くの研究が行われている．文献[2)]では脳波，加速度，位置情報などのセンサ情報とインターネットの履歴，e-mail などから検

†1 法政大学大学院情報科学研究科
†2 法政大学情報科学部

索キーを抽出しライフログ映像の検索を行うシステムが提案されている．また, Life Pod[3] という携帯電話を利用したシステムも提案されている．Life Pod では携帯電話で取得した画像や位置情報，ユーザが入力したメモの利用をしている．

文献[4] ではユーザの記憶を支援するためのシステムとして，位置情報や会話データに音声認識を行った結果を利用している．しかし，会話に対する音声認識の結果は誤りを含む可能性があるため，認識をした単語の信頼度も併せて提示することで，ユーザの想起を支援するシステムが提案されている．文献[5] では収録時のユーザの負担を最小にするため，センサは無指向性マイクと GPS のみを利用し，62 時間のデータを収録して音響情報のスペクトル情報に着目しクラスタリングを行うことで，図書館，レストラン，授業，会議などの 16 の場所や環境に関するセグメントの分類を行っている．また，セグメントの平均時間は 26 分であり，最短の場合でも 15 分程度である．そのようなセグメントをクラスタリングするために，1 フレームの長さを 1 分として特徴量の抽出を行っている．

2.2 音響ライフログと位置情報の利用

音響ライフログに含まれる音響情報には様々な利用法がある．例えば，文献[6] では音声，キーボード打鍵音，紙をめくる音の出現頻度を利用して，オフィス環境でのデスクワークとミーティングの分類をしている．文献[7] では駅ホームにおける電車の発着音，通過音を用いて，ライフログ映像のシーン分割を行った．

音響ライフログで収録された音響情報の中で，多くの応用に役立つ情報としては音声が挙げられる．音声からは会話の内容や話者情報などを含んでいる．このような音声記録は備忘録としての用途が期待できる．しかし，収録した音響ライフログデータは音声を含んでいない部分も多い．実際に収録したデータの 3 時間分を 1 分のセグメントに分割し，聴取したところ 180 セグメント中に音声を含むものは 91 セグメントであり，半分は音声が含まれていなかった．また，状況によっては数時間にわたり音声が含まれていない場合もある．本研究で収録したデータでは，自宅や研究室で 1 人で PC 作業などをしている場合にほぼ音声が含まれていなかった．また，音声が含まれる部分が抽出できても，それが長時間になると所望の部分を探すのは困難である．そこで，検索や整理をするため音声にインデクスやタグを付けるといった処理が必要である．音声に与えられるインデクスとしては，時間，話者，場所，会話の内容などが考えられる．会話の内容を音声認識によりインデクスとする手法が考えられるが，音声認識の誤りや辞書に登録されていない単語を認識できないなどの問題がある．特に未知語が会話のトピックとなる場合，有効な検索が行えなくなる．そこで本研究では時間，場所，話者情報での音声データの提示を行う．

話者情報はクラスタリングにより，同じ話者の会話データを 1 つのグループにまとめる．位置情報は GPS により取得が可能であるが，建物内どの部屋にいるかという情報は取得ができない．そこで，同じ部屋内では背景音が似ていることを利用し，クラスタリングにより詳細な場所の分類を行う．従来の研究では 1 分のセグメントに分割をして特徴量を抽出し場所のクラスタリングを行っている．本研究でも 1 分のセグメントでのクラスタリングを行う．話者や部屋の情報はクラスタリングを行った段階では，同じ話者や場所でのクラスタができているだけである．クラスタへのインデクスはユーザにより付けられることを想定している．

3. データ収録

音響ライフログは 3 種類の IC レコーダー (EDIROL R-09，R-09HR，YAMAHA POCKETRAK CX) とバイノーラルマイク (Adphox BME-200) を用いて収録した．レコーダーは日によって異なるものを使用した．異なるレコーダーを用いることで，デバイス間で録音時の音量が統一できなくなるため，処理が難しくなる場合がある．しかし，ライフログの収録期間は非常に長く，レコーダーの製品寿命よりも長くなると考えられる．従って，デバイスの違いによる影響に頑健な処理が必要である．バイノーラルマイクはイヤホン型のマイクで，本来は両耳に装着をして利用するマイクである．しかし，両耳に装着をした状態での長時間の収録はユーザの負担となるため，図 1 のように肩から提げて収録した．サンプリング周波数は 48kHz，量子化ビット数はレコーダーによって異なり，24 ビットもしくは 16 ビットで収録をした．以上のデバイス，条件で日常生活の音を収録した．主な収録音を表 1 にまとめた．

本研究では処理をした音響ライフログは地図上に提示するため，音響情報の収録と同時に位置情報の記録も行っている．位置情報は GlobalSat DG-100 を用いて 5 秒間隔で取得している．時間は GPS とレコーダーを同時に収録し始めることで，GPS に記録された時間が利用できる．

4. 音響情報を用いた場所のクラスタリング

音響ライフログが収録された場所をクラスタリングにより分類する実験を行った．実験に使用したデータは 2 日分で約 9 時間 30 分である．このときのレコーダーは 1 日が YAMAHA POCKETRAK CX で別の 1 日が EDIROL R-09HR である．データに出現した場所は研究室，廊下，大学構内の屋外，路上，自宅，コンビニエンスストア，スーパーマーケットで

図 1　収録時のバイノーラルマイクの装着方法

表 1　主な収録場所と収録された音

場所	主な収録音
研究室	音声，PC 操作の音，紙をめくる音，ファンの騒音
教室	音声，ファンの騒音
廊下	足音，音声
大学構内 (屋外)	工事，排気ダクト，音声
自宅	TV，音楽
レンタルビデオ店	音楽，音声
ファストフード店	音声
コンビニエンスストア	音楽，音声
スーパーマーケット	音声，音楽
路上	車，音声，音楽，踏切の警告音

ある．この評価データを 1 分のセグメントに分割し，セグメントから特徴量を抽出しクラスタリングを行った．総セグメント数は 571 セグメントである．

以上のデータを用いて 2 つの実験を行った．1 つは GPS を使わないと仮定しデータに現れた全ての場所をクラスタリングする実験で，もう 1 つは GPS で大学という場所を取得できたと仮定し，大学内のデータだけをクラスタリングするものである．

4.1　特　徴　量

本論文で用いた特徴は正規化平均スペクトル包絡である．スペクトル包絡は短時間スペクトルにメル周波数軸上でフィルタバンク分析をして求めた．短時間スペクトルは 85.3ms のハニング窓を 42.7ms ずつシフトさせて切り出した波形に FFT をして求めた．メル周波数は対数軸の周波数で人間の聴覚の感覚尺度に近いとされている．フィルタバンク分析では短時間スペクトルをメル周波数軸上で 600 の幅の三角窓を 300 ずつシフトして切り出し，各帯域の和を求めた．このような処理を行うことで短時間スペクトルが 12 点に集約され，図 2 のような概形が得られる．1 分のセグメントからスペクトル包絡は複数得られるのでそれらの平均を求め，正規化を行いセグメントの特徴量とした．正規化は平均スペクトル包絡の 12 点の平均を減算することで行った．

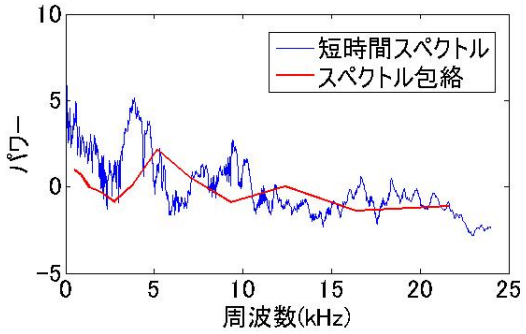

図 2　スペクトル包絡

以上のようにして得た特徴量を k-means 法によってクラスタリングした．クラスタ数は全ての評価データをクラスタリングする場合には 7，大学内のデータをクラスタリングする場合には 3 とした．

4.2　クラスタリング結果

クラスタリングの結果を表 2，3 に示す．評価は各クラスタに手動でラベル付けを行い，再現率と適合率を用いて行った．

4.3　考　　察

研究室と屋外のクラスタにおいて場所が取得できたと仮定した方が再現率と適合率が高い結果となった．屋外と研究室はスーパーのセグメントと混同されやすく，場所情報を利用することでクラスタリング精度を向上させる可能性があることがわかる．廊下の再現率は

表 2　大学内のデータのみをクラスタリングした結果．クラスタには手動で場所のラベルを付けた．各行がクラスタに含まれるセグメント数である．例えば，研究室のクラスタには研究室のセグメントが 210，廊下のセグメントが 3，屋外のセグメントが 1 含まれる．

	研究室	廊下	屋外	適合率	再現率
研究室	210	3	1	98.1%	51.5%
廊下	180	6	0	3.2%	50.0%
屋外	18	3	18	46.2%	94.7%

表 3　全ての評価データをクラスタリングした結果．クラスタには手動で場所のラベルを付けた．表 2 と同様に各行がクラスタに含まれるセグメント数である．

	研究室	廊下	屋外	自宅	コンビニ	路上	スーパー	適合率	再現率
研究室	113	1	1	4	2	2	3	89.7%	27.0%
廊下	37	4	1	0	1	1	1	8.9%	33.3%
屋外	40	2	8	0	1	13	2	12.1%	42.1%
自宅	0	0	0	78	0	0	0	100%	94.0%
コンビニ	51	0	2	0	2	0	0	3.6%	33.3%
路上	67	1	0	1	0	10	0	12.7%	29.4%
スーパー	100	4	7	0	0	8	3	2.5%	33.3%

場所を大学に限定した方が高いが，適合率は場所を限定することで低くなった．これは，他のクラスタと誤りやすく，セグメントの数も多い研究室セグメントが，大学に限定することで，むしろ占める割合が大きくなったためである．また，本実験で用いたデータを収録した環境ではスーパーやコンビニ，自宅，路上は部屋の分類がないため，GPS で得た場所がそのままクラスタとなる．以上のことより位置情報を利用することで音響情報のクラスタリングの精度を補完できると考えられる．また，路上については位置は取得可能であるが，移動することが多いためクラスタリングには工夫が必要である．本論文では用いたデータでは 10 分程度の移動なので，それを 1 つのクラスタとしても問題はないが，移動が数時間になった場合のクラスタリング方法も考えなければならない．

研究室のセグメントが分散した原因としては，収録される音が状況により異なることが理由として考えられる．研究室で主に現れる音には表 1 に示したような音が挙げられ，その中で特に音声が含まれているときと含まれていない時で音響的特徴の差が大きい可能性がある．会話の場合，数分間にわたり音声が収録されている場合がある．従って，会話をしていない時のデータとはセグメントの特徴量が大きく異なる．特に収録者の音声は音量が大きく，スペクトルの形に大きく影響すると考えられる．音響ライフログの応用によっては，これを利用して会話時と会話をしていないときという状況で，クラスタを分けることが有効であるかもしれない．また，1 分毎にセグメントのクラスタが頻繁に変化することはほとんどない．従って，文献[8] のように時間的に近いセグメントを同じクラスタに分類されやすくすることで，研究室のセグメントを 1 つのクラスタにまとめることが可能かもしれない．

本論文の実験では異なるレコーダーを用いた 2 日間のデータを用いたが，正規化によりレコーダーの違いに関わらずクラスタリングができていた．これは表 2 の屋外のセグメントがほぼ 1 つのクラスタに含まれたことからもわかる．しかし，天候や部屋の空調などの条件が異なると音響的な特徴が変化する可能性もある．このような違いによる影響を検証するには，長期のデータを用いた実験が必要である．

5. データの提示手法

5.1 地図上での音声提示の問題点

ユーザのデータの提示要求としては，

1. 5 月 1 日に研究室で A さんが話していたことを聞きたい
2. 日付は定かではないが，夕方ごろに A さんと話していたことを聞きたい
3. 5 月 1 日に A さんと B さんと 3 人で話していた時の会話を聞きたい

というような様々な要求が考えられる．そのような場合に，本論文で提案をした部屋でのクラスタリングが役立つ．話者のクラスタリングについては，1 分以上のセグメントを用いると 1 つのセグメントに複数の話者の音声が入り込む場合が多い．従って，より短いセグメントでクラスタリングを行う必要がある．理想的には固定長のセグメントではなく，可変長で音声区間のみを抽出することが望ましい．このような処理によりインデクスを付けることで，上記の 1 の要求に対しては 5 月 1 日の研究室での A の発話を提示することで要求に応えられる．また，2 の要求に対しては夕方の A の発話を提示することができる．3 に対しては，追加の処理として A，B とユーザの発話が集中している時間帯を探すことで，要求に一致した音声データを提示することができる．

5.2 音響ライフログ提示の例

実際に場所，日付，時間，話者から音響ライフログを提示を行う例を図 3 に示す．図 3 のシステムには Google Maps API[⋆1] を用いている．はじめに，ユーザは GPS により取得さ

⋆1 Google Maps API http://code.google.com/intl/ja/apis/maps/

れた場所のマーカーを選ぶ．さらに部屋を選ぶと，左部分に日付別にデータがツリー構造で表示される．日付を選ぶと，その日の音響ライフログに登場した話者が表示され，話者を選択するとその話者の発話が時間別に表示される．時間をクリックするとその時の音声が再生される．

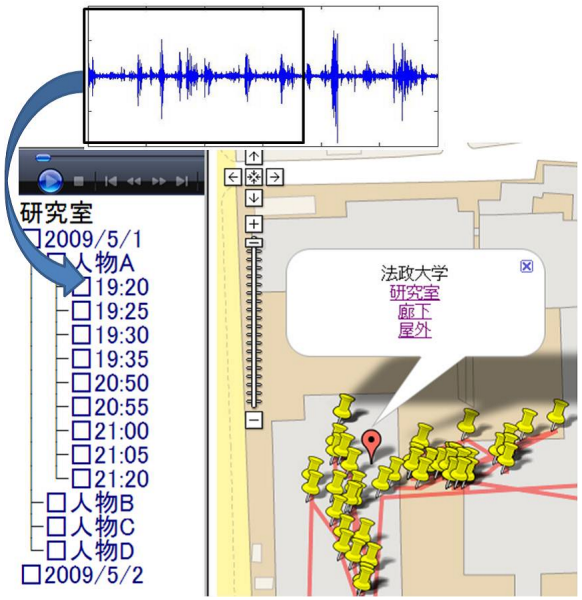

図 3　データの提示

6. あ と が き

本論文では音響ライフログの効率的な利用法として，位置情報，話者，時間による情報の提示手法を提案した．また，建物内での位置情報を得るための手法として，音響情報を用いた場所のクラスタリングを提案した．提案手法の実験として，2 日分の音響ライフログをセグメントに分割し，スペクトル包絡を用いて 3 つの場所のクラスタリングを行った．実験は GPS により場所を取得したと仮定した場合と，GPS を用いない場合の両方を行った．その結果，本論文の実験の範囲では GPS の利用によりクラスタリングの精度が向上することが確認できたが，より長期間収録を行ったデータに対する実験も必要と考えられる．

本論文では GPS から得た位置情報に誤りがないという前提で実験を行った．しかし，GPS により取得される位置情報は，数メートルから十数メートル程度の誤差を含んでいることがある．稀に数キロメートル以上の誤差を含む場合もあるが，継続的にこのような大きな誤差が出ることはないため，誤りとして処理することは可能である．場所や建物を GPS で特定する場合には，GPS の信号が途切れた地点や観測点が集中している場所が重要となる．今後はこのような情報から場所を特定する手法や実験も行わなくてはならない．また，屋外にいる場合でも位置情報が取得できなかったこともあるので，その現象の頻度も調査しなければならない．

その他の課題点として，場所や話者のクラスタリングに適切な特徴量の探索が考えられる．また，本論文では k-means 法を用いたが，実際のデータではクラスタ数が不明であるので，自動でクラスタ数の決定ができるクラスタリング手法も考えていかなければならない．

参　考　文　献

1) J Gemmell et. al., “MyLifeBits: A PERSONAL DATABASE EVERYTHING”, *COMMUNICATIONS OF THE ACM*, Vol.49, No.1, pp.88-95, Jan. 2006
2) K Aizawa, “Digitizing Personal Experiences: Capture and Retrieval of Life Log”, *Proceedings of the 11th International Multimedia Modelling Conference*, pp.10-15, Jan. 2005
3) Atsunori Minamikawa et. al, “RFID Supplement for Mobile-Based Life Log System”, *Proceedings of SAINTW’07*, pp.50-50, Jan. 2007
4) V Sunil et. al., “An Audio-Based Personal Memory Aid”, *Ubicomp 2004*, Vol.3205, pp.400-417, Oct. 2004
5) DPW Ellis et. al., “Minimal-impact audio-based personal archives”, *CARPE’04*, pp.39-47, Oct. 2004
6) 志村他, “行動状況により検索可能な体験映像提示手法の検討”, **情処論講** *68* **回**, pp.4_81-4_82, 2006
7) 山野他, “バイノーラルマイクを用いたライフログ映像のショット識別”, **第** *23* **回信号処理シンポジウム**, Nov. 2008
8) Wei-Hao Lin et. al, “Structuring Continuous Video Recordings of Everyday Life Using Time-Constrained Clustering”, *SPIE Symposium on Electronic Imaging,*, Jan. 2006