

モバイル端末における複数センサ連携による生活行動内容推定技術の提案

Action reasoning based on several sensor logs of mobile terminal

楠本 昌弘^{*1} 三浦 浩一^{*2} 松田 憲幸^{*2} 瀧 寛和^{*2} 安部 憲広^{*3} 堀 聡^{*4}
 Masahiro Kusumoto Hirokazu Miura Noriyuki Matsuda Hirokazu Taki Norihiro Abe Satoshi Hori

^{*1}和歌山大学大学院 ^{*2}和歌山大学 ^{*3}九州工業大学 ^{*4}ものづくり大学

^{*1}Graduate school, Wakayama University

^{*2}Wakayama University

^{*3}Kyushu Institute of Technology

^{*4}Institute of Technologists

To assist a user adequately, a helper needs to know details of the user's situation. Recent downsizing of mobile terminals enhances the possibility of that assistance. This paper describes action reasoning based on non-voice sound logs of mobile terminal. We developed a prototype system to estimate a time which a user's activity is changed from real non-voice sound. We confirmed the effectiveness of our proposed method by an experiment.

1. はじめに

ここ数年で我々は高度に情報化された携帯電話や携帯情報端末を持つ機会が増え、生活に浸透して来ている。将来は、携帯電話と携帯情報端末の融合、パーソナルコンピュータの小型化、ウェアラブル化などが進むと言われている。

現在、日本国民の10人に7人は携帯電話を持ち、生活になくなくてはならない存在となった。携帯電話を販売する各メーカーにおいてもさらなるユーザーの確保のため、様々なサービスや機能を付加し続けている。

特に近年においては、携帯電話をユーザーが常に持ち歩いているという状況に着目し、ユーザーにとって便利で有用なサービス、つまり、ユーザーの状況や行動に応じたサービスを、いかに素早く、そして正確に提供できるかと言う顧客満足度に注目したサービスが展開されて始めている。

ユーザーが求めるサービスを即座に実現するには、ユーザーの行動の様子を把握する技術が不可欠である。

行動に即したサービスが出来れば、ユーザーの満足度も上がることになる。

このようなサービスを実現するためには、デバイスに様々なセンサ(位置情報センサ、音響センサ、圧力センサ、高度センサ、振動センサ、等)を取り付け、そこから得られる情報をシステムが取得し、連携させて活用する必要がある。

しかし、パーソナルコンピュータなどと違い、携帯性を重視した端末機器である携帯情報端末や携帯電話などでは処理能力に限界がある。高性能で大がかりなシステムを使えば、詳細で高度な情報が得られ、それに合わせた支援をユーザーに返す事が出来るだろう。しかし、高度な認識システムに置いては、「マイクロホンアレー」のような大がかりなセンサを必要としたり、高性能なデスクトップコンピュータを持ち歩かなくてはならない、など、必ずしもユーザーと常にあるわけではなく、非現実的である。

その為、新たに、小型軽量の端末機器で「ユーザーの生活行動の推定を行うための技術」が必要とされている。

本研究では「ユーザーの生活行動の推定技術」を実現するための基礎技術として「生活音からユーザーの行動や状況の変化を抽出する」システムの開発を行った。

本システムを実現するためには、生活音の特徴を示す属性、

周波数特性を見出し、行動の変化を判定するための基準を設ける必要がある。

つまり、ユーザーの行動の変化に伴う空間の変化を認識する事であり、その為には、様々な行動パターンの生活音を分析し、それらの周波数特性を調べる事によって周波数成分の特性を理解し、選択し、判定基準を判断しなければならない。

本研究は他の「音」に関する研究が、発話や音声の認識であるのに対し、「非音声音」と呼ばれる、「音声でない音」に注目した研究である。

非音声音分野では、「単発衝撃音」と呼ばれる「その発生した音は一体何の音であるか?」という研究が代表的であるが、本研究は、この材質や原因の認識を行うものに対して、生活音の流れからユーザーの行動、活動内容を推定する点で大きく違っている[yamada01]。

その他に、行動推定システムとしては、動画像から人間の行動を推定し文章化したり、サッカー選手の動きからその行動や戦術を推定する研究などがあるが、行動推定の精度が低かったり、高性能なコンピュータを必要とするなどの違いがある[kojima98]。

2. 生活音からの特徴点の抽出

「生活音からユーザーの行動や状況の変化を抽出する」為には、ユーザーを取り巻く「音」に何が含まれていて、どのような特徴を持っているかを調べなくてはならない。

「生活音」とは、ユーザーが生活をしている際に発生する周囲の環境の音で、ドアの開閉音や車のエンジン音、部屋の中の音など、「非音声音」と呼ばれている音を中心とある。その中でもユーザーの普段の生活中に発生する音を「生活音」と考える。

また、ユーザーの行動や状況が変化した時点「特徴点」と呼ぶ。

ここでは、特徴点を見つけ出す為には、その音成分を分析することによって、音のどの部分に注目すべきかを調べてみた。

その為には、典型的な行動のサンプル十種類の音を用意し、そのデータを比較することによって、類似点や相違点、変化のない部分などから判断し、特徴点を抽出するために注目すべき成分を見つけ出した。

以下に、四種類のサンプルの周波数成分を掲げる。

注:横軸は周波数、縦軸はスペクトル

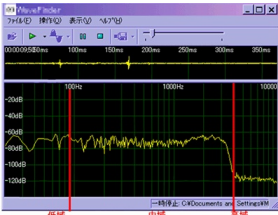


図 2.1 屋外にいる時

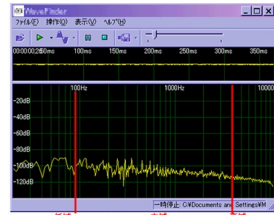


図 2.2 狭い屋内にいる時



図 2.3 『あ』の音声

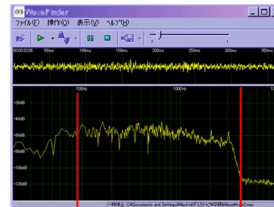


図 2.4 電車に乗っている時

まず、図 2.1 から図 2.4 を見て、大きく三つに分けた。

- 低域 : 0 ~ 90Hz
- 中域 : 90 ~ 4,000Hz
- 高域 : 4,000Hz 以上

全ての図で共通する事だが、4,000hz を越える高周波域では大きな変化が見られない。音の大きさによってはスペクトルに変化が見られるが、これでは特徴点抽出としては不十分である。

100Hz 付近から 4,000Hz では大きな変化が見られるが、この周波数帯ではベルの音や風の音、また人間の音声の中心周波数帯などが含まれている為である。特に、人間の音声は 50Hz ~ 7,000Hz まで出るとされているが、実際には 90Hz ~ 3,000Hz くらいまでが中心と考えられる。この事は、図 2.3 に顕著に現れる。

この周波数帯を省くことによって、音声を特徴点として抽出する可能性は最小限に押さえることが出来る。

80Hz 以下の周波数帯については音の強弱によるスペクトルの変化は見られるものの、あまり差がない事がわかった。

以上の考察をもとに、この 80Hz 以下の低周波帯が、録音者の行動をもっとも反映していることが分かった。

目的とする「生活音から特徴点を抽出する」システムを実現するには、以下の三つの要素が必要となる。

- ・ 取り込んだ音を数値データに変換
- ・ 数値データをフーリエ解析
- ・ フーリエ変換したデータから特徴点の抽出

フーリエ解析を行うことによって、その音が「どのような周波数成分から成り立っているのか」「その相対的な強さはいくらなのか」「各成分の位相はどの程度ずれているのか」を調べることが出来る。[kawakami97][kawakami99]

一般的に、音声認識や非音声認識では、与えられた音波形をフーリエ解析し、そのスペクトルをパターン処理させて認識している。この方法は、もっとも確実な手法であるが、高い処理能力を有する手法でもある。

軽量のデバイスによる処理は困難であり、また全ての周波数成分に行動の特徴が現れる訳ではないので、ここでは別の手法を考えた。

図 3.1 にシステム構成図を示す。

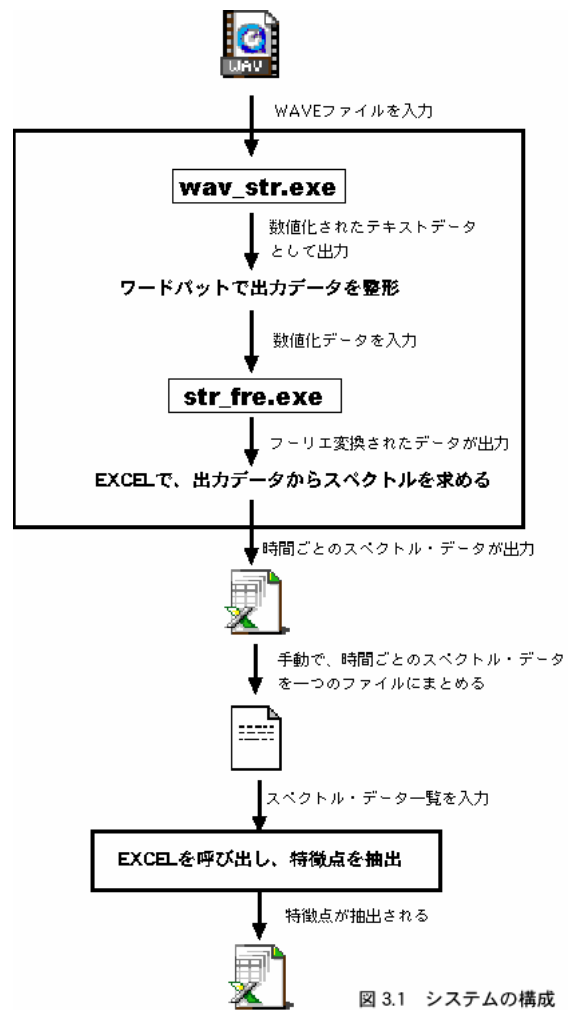


図 3.1 システムの構成

尚、取り込む WAVE データは、データ容量を軽くするために、モノラル・データを使用している。ステレオデータの場合、L、R の二つのチャンネルデータを有するために、モノラルの二倍のデータ容量になる為である。

サンプリング周波数は、これはレコーダの性能上 22,050Hz となった。当然、44,050Hz にすれば音の質は良くなるが、その分データ容量も増える。また、初期実験から 22,050Hz でも十分音の特徴が抽出可能である事を確認した。

ここで数値化する生活音のデータは「1 秒毎、100 ミリ秒間」の音データである。

100 ミリ秒は非常に短い時間ではあるが、サンプリング周波数が 22,050hz の場合、サンプル数は 2205 個となり十分な数値であると考えられる。

このデータ取得方式では「1 秒 ~ 1.1 秒」「2 秒 ~ 2.1 秒」「3 秒 ~ 3.1 秒」...での音の変化は認識できるが、「3.2 秒」や「4.5 秒」に発生した音の変化は認識できない。

ここでは、そのように一秒に満たない音の変化はないものとしてデータを取得しているが、当然この取得する音の時間を「200 ミリ秒」「300 ミリ秒」にすれば特徴点抽出の精度が向上する。

最も理想的な音の取り方は「一秒毎、一秒間」であるが、それでは現在のシステムの 10 倍以上の負担がかかるため、ここではシステムの処理を軽くするために、計算式を出来る限り少なくしている。

また、取得する音の時間の間隔は、一秒毎が最も良いと考えられる。仮に二秒毎にすると、それだけの長さに見合った時間の長さでデータを取得しなければならず、また、間隔が広がれば広がるほど特徴点の抽出精度は落ちる。逆に、短くすると取得するデータの時間の長さを短く出来るが、全体的な計算量は増え、非常に効率が悪くなる。

よって、取得する時間の間隔は「一秒間」とする。ただし、取得する時間の長さに関しては実際のデバイスなどの処理能力を考慮する必要がある。

次に、数値化されたデータからフーリエ解析し、得られたスペクトルを分析する。

フーリエ解析により、サンプリング周波数 22,050Hz でサンプリングされた 2206 データ (約 100 ミリ秒) から直流成分を含めて基本周波数から 1,103 倍の高調波の周波数をもつ sin と cos の波形の振幅が求められる。

ここで、フーリエ解析されたデータの基本周波数は、以下の式によって求められる。

$$\text{「サンプル周波数} \div \text{データのサンプル数」}$$

以上までの事をふまえて、実際に結果を用いて特徴点の抽出を行う。

生活音は行動の変化に伴ってスペクトル値が変化するので、その変化の値を求める。

具体的には、80Hz までのスペクトルについて比較を行うのだが、ここでは、0 ~ 80Hz までのグラフの面積を求め、その値を「その時点の生活音のデータ値」として比較する方法を用いた。

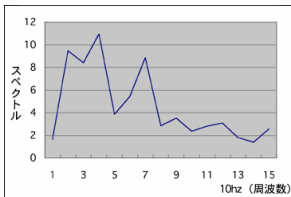


図 3.2 狭い屋内にいる時

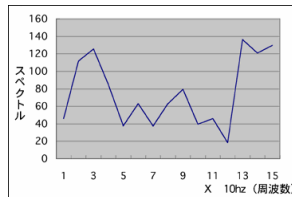


図 3.3 屋外に出た時

図 3.2 と図 3.3 は、狭い屋内から外に出た場合の音から抽出されたデータである。それぞれのグラフは「狭い屋内にいる時」と「屋外に出た時」であり、そのスペクトルの大きさに顕著な差が現れた。

以上から、「その時点の生活音のデータ値の差」を取ることで、特徴点が現れたと考える。

尚、一定の値を「しきい値」とし、その値を越えたら特徴点とする方法も考えられたが、全体的にスペクトル値が高い場合において、特徴点が抽出出来ないため、一秒ごとの、0 ~ 80Hz の音声のパワーの合計、つまり「生活音のデータ値」の相対的な変化から判断を行う。

ここで「バスに乗った時」と「走っている時」の音を比較分析し、特徴点を判断する為の「しきい値」を決定した。

3. 評価実験

本実験は、以下の四種類の生活音のデータを用意し、それぞれから特徴点を抽出できるかどうか考案システムの評価をしている。

「実験1:屋外から車の中へ移動した時」(表1)

「実験2:部屋の中から廊下へ移動した時」
 「実験3:屋外から狭い屋内に移動した時」
 「実験4:時速 40Km で走る車のエンジン音」(表2)

しきい値を「300」に設定した場合、評価実験 1 ~ 3 では、行動や状況の変化による、特徴点を抽出する事が出来た。

表1 「実験1:屋外から車の中へ移動した時」

	生活音のデータ値	相対的な差	特徴点	
時間 (秒)	1	51.40		
	2	55.79	4.39	0
	3	363.83	308.04	3
	4	99.33	264.5	0
	5	568.92	469.59	5
	6	161.66	407.26	6
	7	26.01	135.65	0
	8	30.47	4.46	0

実験1では、3秒目にドアを開け、5~6秒で席に座ると言う動作を行っているが、それがきちんと抽出されているのがわかる。

これらから考えると、低周波帯に注目することにより、評価実験 1、2 のようなドアの開け閉めと言った単発衝撃音を含んだ行動の変化を認識することが出来た。また、評価実験 3 のような空間の大きさの変化のみであっても特徴点が抽出されているが、低周波帯と何らかの相関関係があると考えられる。

評価実験 4 では、行動や状況の変化がないのに、特徴点が抽出された。原因として、低周波帯の周波数特性があるからだと思われる。特に、車のエンジン音は低周波を多く含んでいる為、「生活音のデータ値」そのものが大きく、その為、相対的な差を取った場合にしきい値を越える数値が現れたのだと考えられる。

表2 「実験4:時速 40Km で走る車のエンジン音」

	生活音のデータ値	相対的な差	特徴点	
時間 (秒)	1	584.13		
	2	1226.03	641.9	2
	3	3150.92	1924.89	3
	4	3231.48	80.56	0
	5	5121.77	1890.29	5
	6	3220.67	1901.1	6
	7	4669.29	1448.62	7
	8	4947.03	277.74	0

このような低周波を多く発する物体が近くにあった場合を考えると、しきい値に上限を設けるか、特定時間内に特徴点が一定以上抽出された場合には無視するなどの方法が考えられる。また、相対比較する前の「生活音のデータ値」そのものに上限を加える手法も考えられるが、それは今後さらにこの周波数帯を分析して手法を決定するしかないと思う。

4. システムの応用

「生活音から行動や状況の変化に伴う特徴点を抽出するシステム」を使い、実際に「ユーザーの生活行動の推定」する例を考えてみる。

まず最初に必要となるのは、「ユーザーが一体何をしようとしているのか?」という生活行動の情報である。

この情報を先に用意しておく事によって、予定されている行動や状況の変化と、抽出された特徴点とをマッチングさせて、その行動を推定できると考えられる。この生活行動の情報がなければ、センサからの情報を取得してもシステムは全く見当違いの支援を行う事になる。例えば、ユーザーが駅までの道順を探しているのに、ユーザーの身体状況についての情報を返してしまう、といった間違いが考えられる。

その為には、「ユーザーが一体何をしようとしているのか?」という情報が必要である。

ここでは、その情報を「生活マップ」(図5)と呼ぶ。

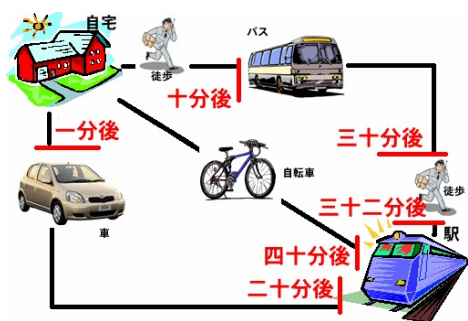


図5 生活マップ

例えば、「自宅から駅まで行く」と言う行動についての「生活マップ」があるとすると、音センサを使った場合に必要の特徴点は以下のように書かれている。

- ・ 自宅を出て徒歩で移動し、十分後にバスに乗り、三十分後にまた徒歩になり、三十二分後には駅に到着する(特徴点は三つ)
- ・ 自宅を出て自転車に乗り、四十分後に駅に到着する(特徴点は一つ)
- ・ 自宅を出て徒歩で移動し、一分後に車に乗り、二十分後に駅に到着(特徴点は二つ)

この生活マップとセンサから得られた情報を、例えば、生活音から十一分後、二十五分後、二十八分後になんらかの特徴点が見つかった場合、生活マップとマッチングさせた結果、最も近い、バスに乗って駅に向かったのではないかと、行動を推定を行うのである。

当然、GPS センサなどを併用することによって、より具体的な場所の情報から、生活マップにおける判断の選択肢がしぼりこまれる為、より精度の高い推定を行うことが可能となる。

このような「生活マップ」を使った行動推定のシステムとして、以下のような例が考えられる。

- ・ 外勤担当者が一日の営業記録をつける際の行動の確認
営業予定はあらかじめ決まっている場合、その業務日誌の作成に役立つ
- ・ 万歩計よりも詳細な行動情報の取得によるカロリー計算

揺れ具合などによって変化する万歩計では階段の昇降や車両に乗った際のカロリー計算に狂いが生じるが、行動そのものを推定できれば、計算が正確になる

- ・ 携帯電話などに組み込んだ場合、ユーザーの周囲の状況を判断し、その着信音のモードや大きさ、メロディを自動的に変化させ、ユーザーがマナーモードなどを設定する煩わしさをなくす

(現状では光センサを用いた、光の明暗差で音量を変化させたり、周囲の騒音のレベルで音量を変更する技術はあるが、周囲の状況を判断するシステムはない[4])

携帯型の端末において、一つ一つのセンサを利用したシステムは多いが、それぞれのセンサが連携する事によって高度な機能を実現するシステムは見られない。

このように、生活マップを用い、複数のセンサを利用することによっては、高度な行動推定に基づいたサービスが実現可能だと考えられる。

5. まとめ

本研究では、携帯電話や携帯情報端末に利用した「ユーザーの生活行動の推定技術」を開発するために、音響センサを使用した場合の有効性を挙げ、基礎技術として「生活音からユーザーの行動や状況の変化を抽出する」システムの開発を行った。

この基礎技術の課題としては、低周波を多く発する物体が近くにあった場合には、本研究で開発したシステムは誤った特徴点を抽出する事がある為に、しきい値に上限を設けるか、相对比较する前の「生活音のデータ値」そのものに上限を加えるなど、周波数帯を再分析して誤りをなくす事が挙げられる。

また、部屋から部屋への移動の際に単発衝撃音がないと状況の変化が抽出出来ない、などの問題点もある。

これらの問題点を解決し、特徴点の抽出精度を上げるのが今後の課題である。

また、表題の通り、それぞれのセンサを連携させて新しいシステムを構築するという考え方は、ユビキタス・コンピューティングやユビキタス・ネットワークに代表される、複数のコンピュータやマイクロコンピュータを組み合わせたシステムでは存在するが、携帯端末を積極的に使用した研究では見られない。
[sakamura02]

よって、今後は携帯端末での複数センサを連携させるアルゴリズムの確立し、実際にサービスが提供できるかが大きな課題である。

参考文献

- [yamada 01] 山田武志: “マイクロホンアレーを用いた環境音の認識に関する研究”, 電気通信普及財団研究調査報告書 第16号, pp.674-679, 2001
- [kojima 98] 小島篤博, 田原典枝, 田村武志, 福永邦雄, “動画像における人物行動の自然言語による説明の生成” 電子情報通信学会論文誌(D-II), vol.J81-D-II, no.8, pp.1867-1875, Aug. 1998.
- [kawakami 97] 川上俊史: “音声入力とブラウザ”, 工学図書株式会社, 1997
- [kawakami 99] 川上俊史: “Windows サウンドレコーダ活用法”, 工学図書株式会社, 1999
- [samamura 02] 坂村健: “ユビキタス・コンピュータ革命”, 角川書店, 2002