

# ケプストラムとDTWを用いた歩行足音識別

板井 陽俊<sup>1</sup> 安川博<sup>1</sup>

<sup>1</sup> 愛知県立大学大学院情報科学研究科

近年、歩行足音による個人識別が可能であることが示され、歩行足音の解析や識別率の向上に関する研究が進められている。本研究では周波数領域における特徴量と動的計画問題を適用し、歩行足音の識別を行う。従来研究ではスペクトルグラムをDTWの入力としていたが、本稿では各種ケプストラムを入力とするDTWを適用した。最適な信号長、ケプストラム次数について検討を行い本人識別率を求めた結果、従来手法と同等以上の歩行足音識別率を実現した。

## Footstep Identification Using Cepstrum and DTW

Akitoshi ITAI<sup>1</sup> Hiroshi YASUKAWA<sup>1</sup>

<sup>1</sup>Graduate School of Information Science and Technology

The feasibility of a personal identification has been confirmed by using the feature parameter from walking footsteps. Nomura[1] has shown that DTW (:Dynamic Time Warp) and the spectrogram is effective for a footstep identification. We present the footstep identification method based on cepstrum parameters and DTW. In experimental results for the footstep classification, empirical parameters, i.e. the cepstrum order and a kind of cepstrum, are obtained by using DTW. The proposed method gives us the good accuracy as well as a conventional method.

### 1 はじめに

人の歩行には固有の特徴があり、足音から人物や歩行状態を特定できる可能性がある。足音から人物が特定出来れば、映像を利用した防犯・監視カメラ、個人の指紋を利用した指紋認証システムなどに代表されるセキュリティシステムの一貫として、防犯対策、早期犯人逮捕などへの応用が期待される。足音を用いた研究は過去に幾つか紹介されているが [2]-[3]、個人識別に関する報告は少なく、多角的な方法により検討することが望まれている。人の足音を特徴づける主な要素として履物の種類や、地面の素材、歩行者のくせが挙げられる。小路 [4]、高須賀 [5] は、人間の聴覚特性を考慮した特徴量を用いることで歩行足音による個人識別が可能であることを示した。しかし、被験者数が5人を越えると認識率が著しく低下するため、クラスタリングを用いた識別では、被験者数の増加にロバストな個人識別システムを構築するために多くの特徴量を併用することが必要であることが指摘されている。

一方、野村ら [1] はDTWを導入することにより歩行足音の識別が可能であることを示した。[1]では歩行足音のスペクトログラムをDTWの入力としているが、その他のパラメータについては検討されていない。本研究で

表 1: Recording conditions

Location	A passage in Aichi Prefectural University
Subject	Ten male students (22-24 years old) A-J
Footwear	A pair of slippers
Measurement	Ten times per subject
Sampling frequency	44100Hz

は音声認識などに適用されるケプストラム、MFCC(Mel Frequency Cepstrum Coefficient)とDTWを用いた歩行足音識別を行い、その有効性を示すとともに、最適な信号長やパラメータの次数についてを定める。

### 2 歩行足音の収録

本稿で用いる歩行足音の収録条件を表1に示す。収録の際、足音はマイクロフォン(ONO SOKKI LA-5120, 周波数範囲:20-12500Hz)からサウンドカード(Card-Deluxe, Digital AUDIO Labs, inc.)を介して、コンピュータに取り込んだ。足音測定の風景を図1に示す。

収録時、被験者にはマイクロフォンに向かって5mの距離を10回ずつ歩いてもらい、マイクロフォンの横を

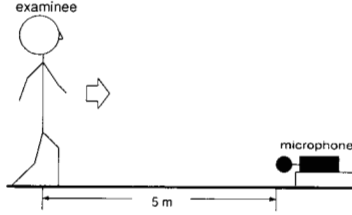


図 1: Recording environment

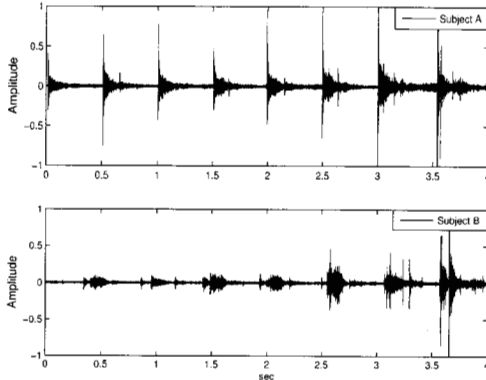


図 2: Footstep waveforms during 4 seconds

通り過ぎる直前までの足音を録音する。距離を 5m とした理由は、各被験者には 10 回連続で歩行してもらうため、疲労による歩行のばらつきが生じないように考慮したためである。実際に足音を収録した廊下では、空調、着衣の摩擦音、壁や障害物に音が反射して生じる反響音など、足音以外にも様々な雑音も同時に収録されてしまい、足音が雑音に埋もれてしまうことがある。歩行足音に含まれる定常的な雑音の除去法としてスペクトルサブトラクション法が用いられているが、本稿では屋内空間における特徴量抽出と足音識別を目的としているため雑音除去等の前処理を行わない。

## 2.1 歩行足音の特性

歩行足音波形の例を図 2 に示す。図は上から被験者 A, B における歩行足音であり、マイクロフォンの横を通過する瞬間を終端とし、その直前 4 秒間分の振幅値をプロットしたものである。

歩行足音に共通する特徴として、約 0.5 秒ごとに発生する振幅値の変動が挙げられる。この振幅は履物が床に衝突した瞬間に発生する音（以下、インパクト音）を表しており、被験者がマイクロフォンに向かって歩く現行の収録環境では時間経過とともにインパクト音が大きくなる。本研究では通過する瞬間に被験者が停止を意識する影響を考慮し、マイクロフォンに最も近

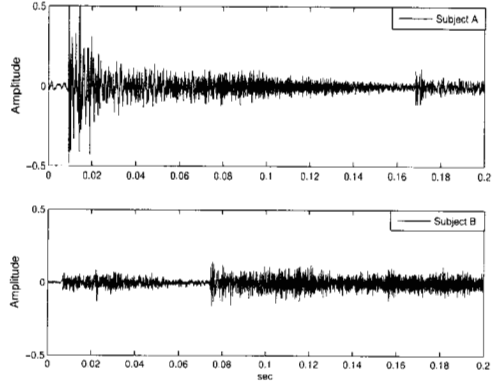


図 3: Footstep waveforms during 0.2 seconds

い足音より一步前に収録された歩行足音を自然な歩行足音として解析対象とする。

次に、マイクロフォンの横を通過する 4 歩手前の足音について、発生から 0.2 秒間の波形を図 3 に示す。図 3 より、被験者 A には歩行足音による明確な振幅のピークが見られるが被験者 B は大きなピークを含まないが 0.075 秒に 2 回目の振幅増減が見られる。このように歩行足音には個人的な特徴が見られ、メルケプストラムやスペクトル包絡線、ウェーブレットなどを用いた個人識別が行われている。

## 3 DTW

DTW は特徴ベクトルからなる時系列に対して非線型時間伸縮の存在を仮定し、入力信号と参照信号間における距離の総和が最小となる、つまり類似度が最大となる参照信号を求めるものである [6]。本研究では特徴ベクトル間の距離としてコサイン距離を用いる。時刻  $i$  における入力特徴ベクトルを  $\mathbf{A}_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$ 、時刻  $j$  の参照信号を  $\mathbf{B}_j = [b_{j1}, b_{j2}, \dots, b_{jm}]^T$  とすると、両信号間のコサイン距離  $D(i, j)$  ( $1 \leq i, j \leq L$ ) は

$$D(i, j) = \frac{|\mathbf{A}_i^T \mathbf{B}_j|}{\sqrt{\sum_{l=1}^n a_{il}^2} \sqrt{\sum_{l=1}^m b_{jl}^2}} \quad (1)$$

のように表される。ここで、 $n, m$  はそれぞれ入力および参照信号の長さ、 $T$  は転置を表し、 $|\cdot|$  は一次元ベクトルの各要素の絶対値を出力する演算子である。さらに時刻  $i$  の入力ベクトルと時刻  $j$  の参照信号における累積距離  $M(i, j)$  は以下の漸化式を用いて計算する。

$$M(i, j) = \min \begin{cases} M(i-1, j) + D(i-1, j) \\ M(i-1, j-1) + D(i-1, j-1) \\ M(i, j-1) + D(i, j-1) \end{cases} \quad (2)$$

但し、 $M(0, 0) = 0$  とする。DTW の処理には Dynamic Time Warp (DTW) in Matlab[7] を用いた。

表 2: Parameters

パラメータ	変数	値
信号長	(none)	20000-
窓長	2N	512 to 4096
シフト幅	(none)	100
フレーム数	L	194
高域強調フィルタ	(none)	[1 -0.97]

## 4 ケプストラム

本研究では特徴ベクトルとしてケプストラム, LPC ケプストラムおよび MFCC[8] を用いる.

1) ケプストラム: 音波形を  $s_t$ , そのパワースペクトルを  $|S_\omega|^2$  とすると, ケプストラム係数  $C_l$  は

$$C_l = \frac{1}{2N} \sum_{\omega=-(N-1)}^{N-1} \log |S_\omega|^2 \cos\left(\frac{2\pi\omega n}{N}\right) \quad (3)$$

と表される. 低次のケプストラムはスペクトル包絡の情報を含んでおり, 歩行足音識別におけるスペクトル包絡の有効性は [4] において示されている.

2) MFCC: 音声認識で一般的に用いられているケプストラム領域の特徴量であり, 以下の処理により得られる.

1. 信号切り出し
2. 高域強調フィルタ
3. ハミング窓
4. FFT(振幅スペクトル)
5. メル対数フィルタバンク
6. DCT

3) LPC ケプストラム: 線形予測係数から求められるケプストラム係数であり, 低次の係数より LPC スペクトル包絡が得られる.

## 5 DTW と歩行足音による個人識別

本研究で用いる足音データは 10 人の被験者から抽出した 10 回分の歩行足音を用いる. 各歩行足音の識別・評価には leave-one-out を用いた. つまり被験者 A の歩行足音を識別する場合, 残りの 99 データを参照データとして DTW を適用し, 入力データは最短パスのコストが最小となる参照データが属するクラスに分けられる. [1] は任意の歩行足音開始時刻から数歩分の足音を含む信号を抽出しているが, 本稿ではマイクロホン直前の 1 から 4 歩目までの足音を別々に抽出し, それぞれについて DTW を適用し個人識別を行う. また, 1

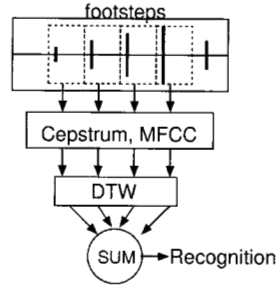


図 4: Recognition process

から 4 歩目までのコスト関数の総和をとり識別を行った (図 4). 処理の流れは以下の通りである.

1. 入力データにおける 1 から 4 歩目の特徴量抽出
2. DTW による最小コスト  $M_{min}(L, L)$  の計算
3. 1 から 4 歩目のコストの総和
4. 最小コストとなり参照データに識別

本研究では入力データとして表 2 のパラメータに基づいて抽出された MFCC, ケプストラムを適用する. ここで, 信号長は歩行足音の発生時刻から特徴抽出に用いるために切り出す信号の長さを表しており, 4 種類の切り出し窓長に対してフレーム数が 194 となるように信号長を設定した. また, 長は 1 フレーム分のケプストラムと MFCC を計算するハミング窓長を表す.

## 識別結果

本研究では各パラメータを求める窓長を  $L = 512, 1024, 2048, 4096$  (12ms-93ms) の 4 種類とした. MFCC の次数は 4 から 4 きざみで 40 まで変化させ, ケプストラム, LPC ケプストラムの次数をそれぞれ 10 から 10 きざみで 200 まで変化させた. 4 歩分の歩行足音における累積距離の総和をとり, 最小の累積距離を返すクラスにより識別を行う. MFCC を用いた認識率を図 5, ケプストラム, LPC ケプストラムを特徴量として用いた結果を図 6, 図 4 に示す. 横軸は特徴量の次数, 縦軸は本人認識率 (正しく識別されたデータ数/総入力データ数\*100) を表す. 実線は  $L=512$ , 破線は  $L=1024$ , 点線は  $L=2048$ , 一点鎖線は  $L=4096$  とした場合の識別率である.

図 5 より, MFCC では窓長が小さくなるにつれ本人認識率が低くなり,  $L = 512, 1024$  では次数の増加にともない識別率が低下している. また,  $L = 2048, 4096$  では次数 12 において識別率が最大となり, 以降次数の増加にともない識別率が低くなっている様子が伺える.

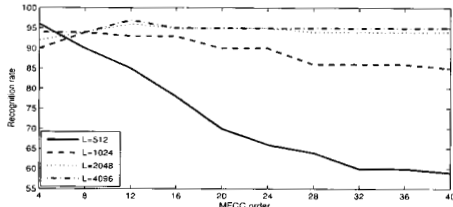


図 5: Recognition rate for MFCC

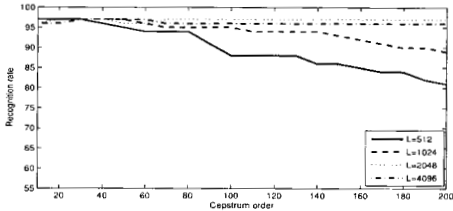


図 6: Recognition rate for cepstrum

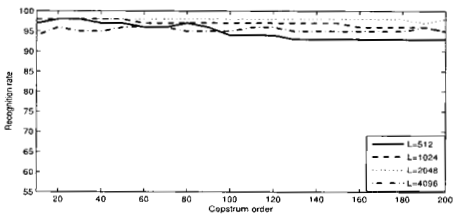


図 7: Recognition rate for LPC cepstrum

もっとも高い認識率は  $L = 4096$ , 次数 12 のパラメータにより得られた 97%であった。

図 6 より, ケプストラムは MFCC に比べ識別率が高いことがわかる。図 6 では図 5 と同様に次数の増加にともない識別率が低下する傾向があるが,  $L = 512$  では 10-30 次, その他の  $L$  では 30-50 次に最大の識別率が現れている。特に 30 次のときは  $L$  に関わらず 97% の識別率となった。また, LPC ケプストラムはケプストラムより次数の増加による認識率の落ちこみが小さく, 低次のケプストラムを用いることにより 98% の足音が正しく識別された(図 7)。一方,  $L = 4096$  の場合は認識率が低下する傾向にあり, ケプストラム, LPC ケプストラム共に,  $L = 2048$  における大部分の次数で最大の識別率となった。

以上より MFCC, ケプストラムと DTW を用いた識別では各特徴量の次数の増加にともない本人認識率が低下する傾向にあることがわかった。これは次数を増加させることで不要な情報が含まれていることが原因であると考えられ, 次数が増加することにより周波数成分の微細な特徴が含まれるようになるため, 歩行足音は周波数成分の微細な変動に有益な情報が含まれないと推測される。

## 6 まとめ

本稿では歩行足音による個人識別法として, ケプストラムと DTW を用いた識別手法について検討を行った。10 人の歩行足音を用いて識別を行った結果, MFCC, ケプストラムともに特徴量とする次数が大きくなるほど識別率が低下する傾向にあることがわかった。また, MFCC では窓長を長くするほど認識率が向上し, ケプストラムでは  $L = 2048$  で最大の認識率が得られた。

今後の課題として, 参照データ数と識別率の関連を求める, DTW に閾値を設けることにより他人を棄却するシステムの構築などが挙げられる。

## 参考文献

- [1] 野村祐基, 磯村太郎, 板井陽俊, 安川博: 歩行足音のコサイン距離を用いた DP マッチングによる個人識別, 信学技報 SIS2006-72, pp.7-12, 2007.
- [2] M.Tanaka and H.Inoue: "A Study on Walk Recognition by Frequency Analysis of Footsteps," Trans. IEE of Japan, 119-C, 6, pp.762-763. 1999.
- [3] Bin She: "Framework of footstep detection in indoor environment", ICA2004. Mo.P2. Engineering acoustics (I)(EAP01), pp.715-718, 2004.
- [4] Y.Shoji, et.al, "Personal Identification Using Footstep Detection in In-door Environment," IEICE Trans. Fundamentals, Vol.E89-A, NO.8, 2005.
- [5] T.Takasuka, Y.Shoji, H.Yasukawa: "On footstep Identification Using Parametric Analysis Based on Psycho-Acoustics," (in Japanese), IEICE Technical Report on CAS2004-58, pp.13-18, Nov. 2004.
- [6] D. Sankoff and J. B. Kruskal: Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Addison-Wesley, Reading, MA, 1983.
- [7] D. Ellis, Dynamic Time Warp (DTW) in Matlab, 2003. Web resource, available: <http://www.ee-columbia.edu/~dpwe/resources/matlab/dtw/>.
- [8] John Griffin, "Discrete Time Processing of Speech Signals," 1993.