

N-012

ウェーブレット変換を用いた音声のバイオメトリクス認証法

A Biometric Authentication Method for a Digital Sound File
Using a Discrete Wavelet Transform

吉富 康成† 浅田 太郎† 衣川 洋平† 田伏 正佳†

Yasunari Yoshitomi Taro Asada Yohei Kinugawa Masayoshi Tabuse

1. まえがき

インターネットの普及、パーソナルコンピュータ、携帯電話など情報端末の普及、音声、画像の圧縮技術の向上などにより、これまで、CD、DVDなどの媒体を用いて保管や配布をしていたデジタルコンテンツが、ネットワークを介して容易に配信できるようになった。

これに伴い近年、オンライン音楽・動画配信サービスを行う企業が出現した。他方、不正に音楽、動画等のコンテンツを配信、アップロードする Web サイトも多く存在し、Peer to Peer(P2P)のネットワーク形態を利用したファイル共有ソフトの悪用者も増加している。特に音楽は動画よりもデータ量が小さく、伝送が容易であるためネットワーク上に著作権を侵害している音楽コンテンツが大量に出回っており、大きな問題となっている。このため、音楽の著作権を保護する技術が種々開発されてきた。それらの技術の代表的なものの一つが電子透かし法[1-3]である。この方法では、対象となる音声データに著作権者に関する情報(以下、「電子透かし」と表記)を埋め込む。この電子透かし法への要求の一つは種々の信号処理に対する耐性である。ネットワーク上の大半の音楽ファイルは何らかの圧縮によって、ファイルサイズを小さくした状態で存在している。ネットワーク上の音楽データの適切な流通管理および不正配信の追跡に際し、電子透かしを証拠保全のために使用するためには、圧縮耐性を有している必要がある。しかし、電子透かし法の場合、圧縮耐性と音質がトレードオフの関係にあるため、圧縮耐性を高めると音質が劣下してしまう。

そこで、この課題を解決するため、本研究では、ウェーブレット変換を用いて、対象となる音声データが元来もっている情報を抽出し、認証に用いるバイオメトリクス認証法の開発を行った。そして、本法では、音質の劣化は皆無なので、各種圧縮に対する耐性を検討した。

なお、「バイオメトリクス認証」という表現は、圧縮などの信号処理を施されても、コンテンツがもつ属性だけを基にそのコンテンツをほぼ確実に特定できる、ということを強調するために用いている。また、本研究は、著者らの電子透かしの研究[4]から生まれたものである。音声コンテンツの、ここでいう「バイオメトリクス認証」に関して、著者らの知る限り、報告されている研究例はない。

2. ウェーブレット変換

自然画像にウェーブレット変換を施し、その多重解像度表現部のウェーブレット係数のヒストグラムをとると0を中心とした分布となる[5]。楽曲に対するウェーブレット係数についても同様の傾向を示す(図1)[4]。著者らは、この性質を利用した電子透かし法を開発した[4]。そして、本研究では、この性質を利用してバイオメトリクス認証法を開発した。以下に、本法で用いた離散ウェーブレット変換に

ついて述べる。まず、窓幅に区切った音声信号 $f(n)$ を、レベル0のスケーリング係数 $s_k^{(0)}$ とみなす。次に、レベル j のスケーリング係数 $s_k^{(j)}$ 、およびウェーブレット係数 $w_k^{(j)}$ を式(1)、(2)を使用して、 $s_k^{(0)}$ から、逐次レベル J まで求める。

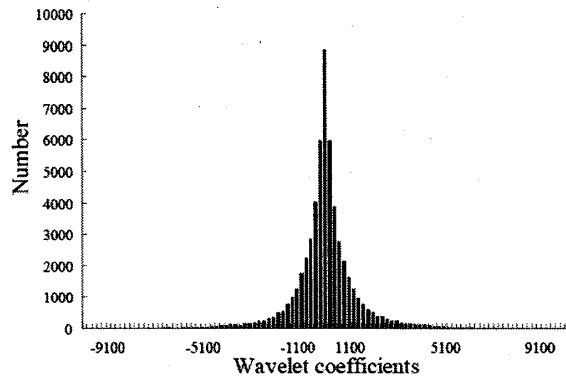


図1 音声(Jazz)のウェーブレット係数の分布[4]

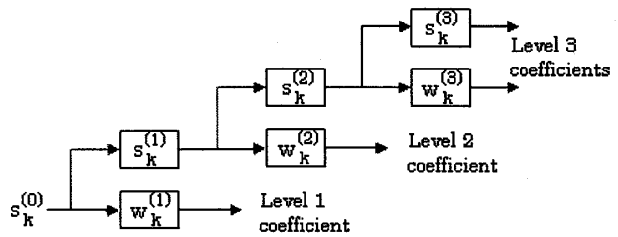


図2 多重解像度解析

$$s_k^{(j)} = \sum_n \overline{p_{n-2k}} s_n^{(j-1)} \quad (1)$$

$$w_k^{(j)} = \sum_n \overline{q_{n-2k}} s_n^{(j-1)} \quad (2)$$

図2にスケーリング係数が次々にスケーリング係数とウェーブレット係数とに分解される様子を示した。

本研究では、 p_k としてドベシーのスケーリング関数を表す数列を用いた。 q_k は式(3)から求められる。

$$q_k = (-1)^k p_{1-k} \quad (3)$$

3. 方法

3.1 コーディング

音声データのウェーブレット係数 V のヒストグラムの概形を図3に示す。画像に対する電子透かし法[5]、音声に対する電子透かし法[4]と同様にして、コーディングに用いるパラメータの設定を行う。まず、ヒストグラムで $V > 0$ の部分の全面積を S_p 、 $V \leq 0$ の部分の全面積を S_m とし、

†京都府立大学大学院, Graduate School of Kyoto Prefectural University

それぞれの面積を等分するよう $Th(plus)$, $Th(minus)$ を算出しパラメータとして設定する。次に、コーディングに用いるパラメータである $T1, T2, T3, T4$ を求める。 $T1, T2, T3, T4$ に要求される条件は以下ようになる。

$$(T1 < Th(minus) < T2 < 0 < T3 < Th(plus) < T4)$$

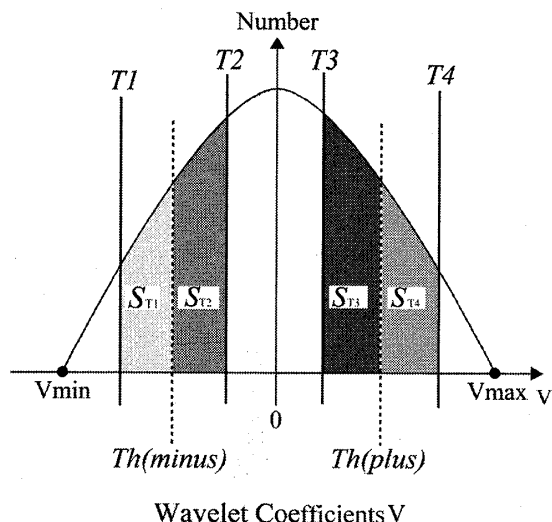


図3 多重解像度表現部ウェーブレット係数のヒストグラムの概形

- $V = T1$ と $V = Th(minus)$ で囲む面積 S_{T1} と、 $V = Th(minus)$ と $V = T2$ で囲む面積 S_{T2} が等しい
- $V = T3$ と $V = Th(plus)$ で囲む面積 S_{T3} と、 $V = Th(plus)$ と $V = T4$ で囲む面積 S_{T4} が等しい
- $V = T1$ と $V = Th(minus)$ で囲む面積 S_{T1} と S_m の比と、 $V = T3$ と $V = Th(plus)$ で囲む面積 S_{T3} と S_p の比が等しい

本論文では $S_{T1}/S_m = S_{T3}/S_p$ として 0.3 を用いた。

本研究で新たに考案したコーディングに際し、ウェーブレット係数 V について、以下のように領域を A~E に分類する(図4)。

- A: $V \leq T1$
- B: $T1 < V < T2$
- C: $T2 \leq V \leq T3$
- D: $T3 < V < T4$
- E: $T4 \leq V$

そして、対象となるウェーブレット係数 V が属す領域に応じて、以下のようにコーディングを行う。

- Cの時→0
- A, Eの時→1
- B, Dの時→対象外

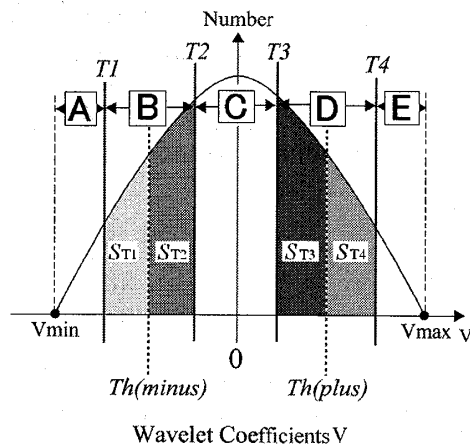


図4 コーディング時の原音のウェーブレット係数の選定のための領域分け

得られた 0,1 データのビット列(コード)が認証に用いられる。ウェーブレット係数に一連の番号を付し、さらに、認証対象となる A, C, E の領域のウェーブレット係数にフラグ 1 を与え、B, D の領域のウェーブレット係数にフラグ 0 を与えて区別する。B, D の領域のウェーブレット係数をコーディング対象外としたことが本法の特徴である。

3.2 認証

認証の対象となる音楽は、原音、またはそれに圧縮などの信号処理を施した音声データである。その音声データに、コーディングのために施したのと同じ条件でウェーブレット変換を施し、ウェーブレット係数 V' のヒストグラムを得る。そして、ヒストグラムで $V' > 0$ の部分の全面積を S'_p , $V' \leq 0$ の部分の全面積を S'_m とし、それぞれの面積を等分するよう $Th'(plus)$, $Th'(minus)$ を算出する。以下での説明の都合上、ウェーブレット係数 V' について、次のように領域を F~H に分類する(図5)。

- F: $V' < Th'(minus)$
- G: $Th'(minus) \leq V' \leq Th'(plus)$
- H: $Th'(plus) < V'$

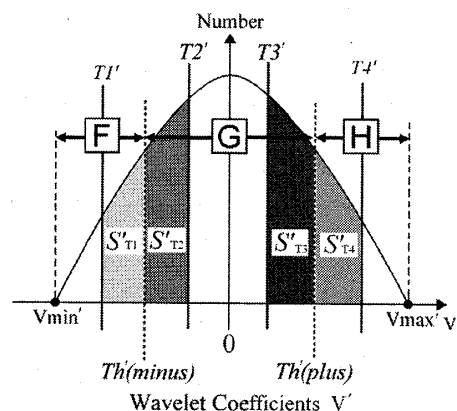


図5 認証時のウェーブレット係数の領域分け

次に、コーディング時に選択されたウェーブレット係数について、認証時に属する領域に応じて、

- ・Gの時→0
- ・F, Hの時→1

とし、0,1 データのビット列を作成する。そして、予め得ておいた原音でのビット列と、認証対象でのビット列との、各ビットの一致率(=認証率)を求める。この認証方法は、著者らの電子透かしの研究[4]で、透かしの検出に用いた方法と0,1の決め方の部分だけは同じである。

4. 評価実験

本章では提案手法の性能評価のために行った計算機実験について述べる。

4.1 方法

実験には、パーソナルコンピュータとして、Dell Dimension 8300 (CPU : Pentium IV 3.2GHz, メモリ : 2.0GB), OS に Windows XP, 開発言語として Visual C++6.0 を使用した。

実験には、RWC 研究用音楽データベース(音楽ジャンル [6])から、各ジャンル (Classic, Jazz, Popular, Rock, HipHop) の先頭の音声 5 ファイルを選択し、デジタルの保存形式には WAVE 形式を用い、44.1kHz, 16 ビット, モノラルに変換して 1 分から 1 分 10 秒までの 10 秒間分を切り取り実験に用いた。なお、音声ファイルにおける認証箇所の影響を調査した実験では、音声ファイルのいくつかの時間帯で各々10秒間分を順次切り出して使用した。

MP3,AAC,WMA の 3 種類の圧縮形式で、128kbps, 96kbps, 64kbps のすべての条件について、認証に対する耐性を調査した。その際、まず、原音を WAVE 形式でコーディングし、次いで、各ファイル形式に変換し、WAVE 形式に逆変換して認証を行った。

本研究では、予備実験をもとに決定したレベル 8 でウェーブレット変換を施した。そして、認証率に対する変換レベルの影響も調査した。

4.2 結果と考察

4.2.1 認証過程

表 1 に、認証過程の例を示す。ウェーブレット係数で抜けている番号 (例えば,573~578) の場合、該当するウェーブレット係数が B または D の領域に属する値であったためにコーディングの対象外となり、認証にも用いられていない。原音と認証対象のコンテンツの符号の一致率として、認証率を算出する。表 1 の例では、No.579 のみ不一致となっている。

表 1 認証過程の例(Hiphop)

ウェーブレット係数 No.	original		MP3,128kbps		符号の一致
	領域	符号	領域	符号	
572	C	0	G	0	○
579	C	0	F	1	×
580	C	0	G	0	○
584	C	0	G	0	○
588	E	1	H	1	○
589	C	0	G	0	○
590	C	0	G	0	○

4.2.2 各種圧縮処理に対する耐性

原音を対象とした場合は、どの楽曲でも 100% の認証率がえられた。各種圧縮処理を施しても、100% または 100% に近い認証率がえられた(表 2)。

表 2 各種圧縮処理に対する耐性(認証率)

信号処理方法	条件 (kbps)	Classic	Jazz	Popular	Rock	Hiphop
		(%)				
MP3	128	99.86	99.86	99.86	100	99.57
	96	99.86	100	99.86	100	99.86
	64	99.72	99.72	99.86	100	98.28
AAC	128	100	100	100	100	97.55
	96	99.86	100	99.86	100	100
	64	100	100	99.14	99.72	99.00
WMA	128	100	100	100	100	100
	96	100	100	100	100	100
	64	100	99.86	100	100	100

4.2.3 コーディングした対象以外の認証結果

著作権を主張するために認証を行うことを想定している。このため、コーディングした対象以外を認証した場合、低い認証率がえられる必要がある。そこで、今回対象とした 5 曲について、コーディングと認証のすべての組合せ(25 通り)について、認証率を求め、表 3 に示した。コーディング時に選定したウェーブレット係数を用いてその楽曲を認証した場合の認証率は 100% であるのに対して、他の楽曲を認証した場合は、50% 程度(44.09~55.62%)の認証率となった。

表 3 コーディングと認証の組合せと認証率の関係

		コーディング				
		Classic	Jazz	Popular	Rock	Hiphop
認証	Classic	100	44.22	47.98	52.89	53.61
	Jazz	55.62	100	49.86	51.01	46.54
	Popular	44.09	53.89	100	49.57	52.88
	Rock	47.69	50.86	49.42	100	48.13
	Hiphop	48.41	50.43	47.55	51.44	100

(%)

次に、同一楽曲において、コーディング対象とした 10 秒間と異なる箇所をも認証対象とした認証率を、図 6 に示す。横軸は、楽曲開始から切り出し開始点までの時間を示す。認証対象の切り出し開始点を 1 秒ずつシフトさせた。横軸の値 60 秒の点が、コーディング対象の区間を認証した場合に相当し、いずれの楽曲でも認証率は 100% となっている。コーディング対象と異なる区間の認証率の時系列変化は、楽曲によって異なった。Jazz, Popular, Rock については、40~60% の値が多く得られたのに対して、Classic, Hiphop では、認証率が周期的に変化した。そして、コーディング対象と異なる区間での最高認証率は、93.95%(Hiphop)であった。このため、著作権を主張するためには、例えば 95% 以上の認証率をうる必要があると考えられる。

また、図 6 より、同一曲内でも、コーディング対象と異なる区間では、95% 以上の認証率を得られていないので、コーディング対象以外の楽曲で、95% 以上の認証率を得る可能性は大変低いと考えられる。つまり、入手したコンテ

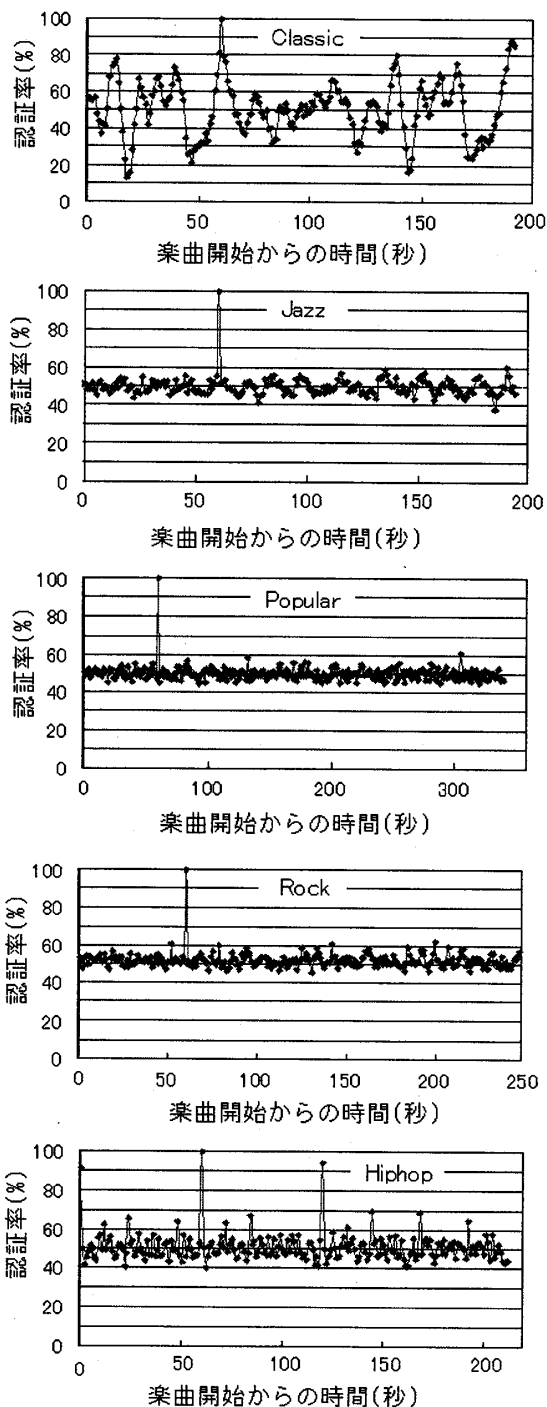


図6 同一楽曲内における認証結果

ンツで95%以上の認証率が得られれば、ほとんどの場合、特定の区間で予めコーディングを行って登録しておいた楽曲と判定できると考えられる。

4.2.4 ウェーブレット変換レベルの認証率への影響

上記の結果は、いずれもウェーブレット変換のレベルを8で行ったものである。認証における圧縮処理への耐性に対するウェーブレット変換のレベルの影響を表4~6に示す。

表4 ウェーブレット変換レベルと認証率の関係(128kbps)

(1)Classic

レベル	信号処理		
	MP3	AAC	WMA
2	99.07	99.41	99.99
3	99.67	99.69	100
4	99.99	99.9	100
5	99.98	99.98	100
6	100	100	100
7	100	100	100
8	99.86	100	100

(2)Jazz

レベル	信号処理		
	MP3	AAC	WMA
2	98.63	99.33	99.95
3	99.64	99.8	100
4	99.97	99.99	100
5	100	100	99.98
6	100	100	100
7	100	100	99.93
8	99.86	100	100

(3)Popular

レベル	信号処理		
	MP3	AAC	WMA
2	92.78	95.08	99.95
3	95.05	98.5	100
4	97.37	99.7	99.99
5	98.53	99.98	100
6	99.71	100	100
7	100	100	100
8	99.86	100	100

(4)Rock

レベル	信号処理		
	MP3	AAC	WMA
2	94.2	95.88	99.81
3	96.72	98.79	99.98
4	98.89	99.85	100
5	99.64	99.98	100
6	100	100	99.96
7	100	100	100
8	100	100	100

(5)Hip-hop

レベル	信号処理		
	MP3	AAC	WMA
2	95.65	61.31	99.6
3	96.27	72.89	99.69
4	96.85	84.3	99.69
5	98.19	91.64	99.89
6	99.67	94.57	100
7	99.93	96.67	100
8	99.57	97.55	100

(%)

表5 ウェーブレット変換レベルと認証率の関係(96kbps)

(1)Classic

レベル	信号処理		
	MP3	AAC	WMA
2	95.95	98.76	99.96
3	96.92	99.08	99.99
4	97.89	99.48	100
5	98.53	99.64	100
6	99.53	99.78	100
7	99.71	100	100
8	99.71	99.86	100

(2)Jazz

レベル	信号処理		
	MP3	AAC	WMA
2	96.48	98.62	99.97
3	98.42	99.39	100
4	99.79	99.95	100
5	100	99.98	99.98
6	100	100	100
7	100	100	99.93
8	100	99.86	100

(3)Popular

レベル	信号処理		
	MP3	AAC	WMA
2	85.61	89.7	99.62
3	94.74	94.7	99.9
4	97.43	98.4	99.85
5	98.79	99.42	99.98
6	99.89	100	100
7	99.86	100	100
8	99.86	100	100

(4)Rock

レベル	信号処理		
	MP3	AAC	WMA
2	89.79	92.56	99.33
3	95.37	96.3	99.78
4	98.58	98.94	99.98
5	99.69	99.8	100
6	99.96	99.89	99.96
7	100	100	100
8	100	100	100

(5)Hiphop

レベル	信号処理		
	MP3	AAC	WMA
2	92.76	94.86	99.46
3	94.13	96.16	99.54
4	95.38	97.43	99.6
5	97.01	98.84	99.84
6	98.91	99.67	100
7	99.57	99.93	100
8	99.86	100	100

(%)

表6 ウェーブレット変換レベルと認証率の関係(64kbps)

(1)Classic

レベル	信号処理		
	MP3	AAC	WMA
2	98.18	98.08	99.88
3	98.87	98.51	99.96
4	99.54	99	99.98
5	99.91	98.73	100
6	99.96	99.13	100
7	99.93	99.42	100
8	99.86	100	100

(2)Jazz

レベル	信号処理		
	MP3	AAC	WMA
2	92.41	97.23	99.85
3	95.45	98.45	100
4	98.39	99.54	100
5	99.93	99.95	99.98
6	100	99.89	100
7	100	99.86	99.93
8	99.71	100	100

(3)Popular

レベル	信号処理		
	MP3	AAC	WMA
2	73.87	81.86	96.73
3	87.94	90.43	99.44
4	95.21	96.02	99.42
5	98.26	97.77	99.84
6	99.78	98.8	100
7	100	99.57	100
8	99.86	99.14	100

(4)Rock

レベル	信号処理		
	MP3	AAC	WMA
2	82.67	88.43	96.12
3	91.55	93.67	99.14
4	97.09	97.56	99.69
5	99.4	98.91	99.95
6	99.93	99.17	99.96
7	100	99.35	100
8	100	99.71	100

(5)Hiphop

レベル	信号処理		
	MP3	AAC	WMA
2	82.13	92.95	97.86
3	86	94.35	98.85
4	89.31	95.82	98.78
5	92.62	96.7	99.31
6	97.43	97.9	99.86
7	98.48	98.26	99.86
8	98.27	98.99	99.86

(%)

128kbps(表 4), 96kbps(表 5), 64kbps(表 6)の3条件で、認証率に明確な違いは出なかった。レベルを上げると認証率が幾分高くなる傾向があった。本実験では、レベル7以上で、いずれの信号処理の条件でも95%以上の認証率が得られた。なお、レベル6の条件での最低認証率は、94.57%(Hiphop, AAC, 128kbpsの場合)であった。

なお、原音(データ数 441000)をレベル0としてウェーブレット変換しており、コーディング条件としたレベルにおいて40%のウェーブレット変換係数がコーディング対象となる。また、ウェーブレット変換のレベルを1つあげると、コーディング対象のウェーブレット変換係数の数が半分になる。

4.2.5 電子透かし法との比較

電子透かし法の場合、圧縮耐性と音質がトレードオフの関係にあるため、圧縮耐性を高めると音質が低下してしまう。著者らは、圧縮耐性と音質を両立させるため、埋め込み条件を最適化させる研究[4],[7]を行ってきたが、処理時間がかかるという課題が残されている。

図7に、本研究で用いた音声データ Jazz を用いた際の、電子透かし法[4]における音質と MP3 圧縮(96kbps)後の透かし検出率の関係を示す。高い透かし検出率と良好な音質を実現するために、遺伝的アルゴリズム(GA)を用いて、透かし埋め込み条件に関する最適化問題を近似的に解いている。この埋め込みには、原問題(Original problem)、部分問題(Partial problem)で、各々、 2.52×10^4 s, 5.12×10^2 s を要している[9]。なお、開発環境として、Dell Dimension DXC051(CPU: Pentium(R) 3.00GHz, メモリ: 1GB), OSに WindowsXP, 開発言語として Visual C++6.0 を使用していた[4]。図7と同じ音声データ Jazz, MP3 圧縮(96kbps)の条件での本法での認証率は100%で(表2), 音質の劣化はない(図7の横軸の値は無限大)。また、本法でのコーディング、認証、の所要時間は、各々、 2.03×10^{-1} s, 2.19×10^{-1} s, であった。開発環境が著者らの先行研究[4]と本研究で異なるので正確な比較はできないが、電子透かしの埋め込みとコーディングに必要な時間を比較すると、本法の方が、3~5桁少なくなっている。かつ、音質についても本法の方が優れている。

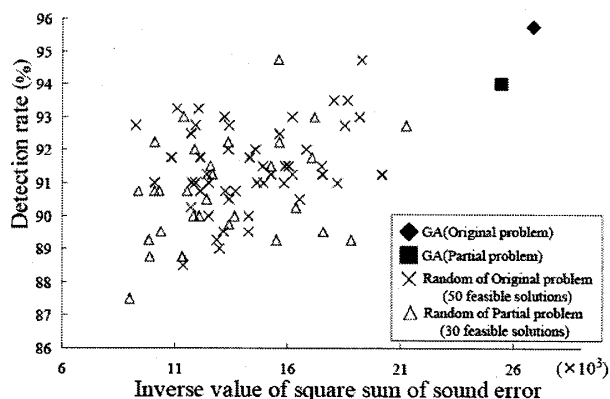


図7 電子透かし法における音質と透かし検出率の関係[4]

本法では、音声データに情報を付加することなく認証しているため、音質の劣化はない。また、表2に示したように、本法の圧縮耐性は大変優れている。また、本実験で標準としたレベル8では、コーディングに平均 2.16×10^{-1} s, 認証に平均 2.22×10^{-1} s の時間しかかからないため、処理時間は問題とならなかった。

著者らは、画像を対象として圧縮耐性と画質を両立させるため、埋め込み条件を最適化させる研究[8]も行ってきたが、処理時間がかかるという課題が残されている。このため、本法の画像への適用を進めている。

5. まとめ

ウェーブレット変換を用いた音声のバイオメトリクス認証法を開発し、MP3, AAC, WMA のいずれに対しても圧縮耐性が優れていることを実験で示した。認証に用いる情報は、コーディング時に得られる、①対象となるウェーブレット係数の識別フラグ、②コード、である。本法では、音声データに情報を付加することがないため、音質の劣化はない。また、計算の負荷も問題とならない。

参考文献

- [1]松井 甲子, “電子透かしの基礎”, 森北出版, 1998.
- [2]小野 東, “電子透かしとコンテンツ保護”, オーム社, 2001.
- [3]画像電子学会 編, “電子透かし技術”, 東京電機大学出版局, 2004.
- [4]村田 真一, 吉富 康成, 石井 博昭, “ウェーブレット変換を用いた音声電子透かし埋め込み位置の最適化”, 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, pp.210-211, 2007.
- [5]篠 政義, 崔 潤基, 相澤 清晴, “ウェーブレット係数の可変閾値決定に基づく電子透かし法”, 信学技報, SP2000-52, pp.29-34, 2000.
- [6]後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, “RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース”, 情報処理学会論文誌, vol.45, no.3, pp.728-738, 2004.
- [7]田仲 雅和, 吉富 康成, “対数振幅を用いた MP3 圧縮に耐性を有する音声電子透かしの埋め込み位置決定問題”, 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, pp.70-71, 2006.
- [8]D.Inoue and Y.Yoshitomi, “Watermarking using wavelet transform and genetic algorithm for realizing high tolerance to image compression”, 画像電子学会誌, vol.38, no.2, pp.136-144, 2009.
- [9]村田 真一, “ウェーブレット変換を用いた音声電子透かし埋め込み位置の最適化”, 大阪大学大学院情報科学研究科情報数理学専攻修士論文, pp.53, 2006.