ディジタル音情報ハイディングにおける情報改ざん検出の実現可能性

代表研究者	鵜木　祐史	北陸先端科学技術大学院大学　情報科学研究科　准教授
共同研究者	赤木　正人	北陸先端科学技術大学院大学　情報科学研究科　教授
共同研究者	宮内　良太	北陸先端科学技術大学院大学　情報科学研究科　助教

1 研究調査の要旨

近年，ディジタル音響信号・音声信号の高機能編集（修正・加工されたことを知覚できないような編集機能）が可能になりつつある．そのため，これらの技術を悪用した音情報の改ざんといった重要な社会問題が起こりはじめている．本調査研究では，情報通信技術で利用されるディジタル音情報における情報改ざんの実態・可能性を調査するとともに，情報改ざんを未然に防ぐための要素技術として，改ざん検出技術の実現可能性を検討した．研究代表者によって提案された蝸牛遅延に基づいた知覚不可能な電子音響透かし法（CD 法）を利用して，音声改ざん検出の検討を行った．この方法は代表的な音声分析合成による音声改ざん検出に対応可能であるが，一部の音声符号化処理に脆弱であることがわかった．そこで，この問題を解決するために，線形予測符号化に着目した音情報ハイディング法として，線スペクトル周波数を使用したフォルマント強調に基づく電子透かし法を提案した．総合評価の結果，提案法は，知覚不可能な電子透かし法であるとともに，リサンプリングや量子化操作，音声符号化・復号化といった通常の情報処理に対して頑健性を有することがわかった．更に，波形接続型音声合成を利用した音声内容の改ざんや，話者性を変える意味でのピッチ変換・話速変換，AD/DA 変換を介した音の再収録という意味での雑音残響付与に関して，提案法が脆弱性を有することも確認された．以上から，提案法がディジタル音情報の改ざん検出技術として十分応用可能であることがわかった．

2 研究背景

近年，音声分析合成技術（例えば，一種の VOCODER である STRAIGHT[1,2]）の急速な発展に伴い，音声の高品質な編集が可能に成りつつある．例えば，素片接続型音声合成[3, 4]や声質変換[5]，音声モーフィング[6]，感情音声合成[7]，歌声合成・変換[8,9,10]といったアプリケーションも登場している．

素片接続型音声合成は，テキスト読み上げ音声（text-to-speech）のように音声の内容（何といっているか）を合成・変換可能な技術である．声質変換や感情音声合成，歌声合成・変換は，それぞれ，言語情報を保存したまま，話者の個人性（誰がしゃべっているか），感情や歌声といった非言語情報（どのようにしゃべっているか）を合成・変換可能な技術である．これらの技術の進展は音声の高機能編集（修正や加工されたことを知覚できないような編集機能）を可能としている一方で，音声の情報改ざんの可能性も著しく高めている．そのため，音声コンテンツそのものを悪意ある情報処理（モラルのない間違った使い方）から守るために，オリジナルの音声コンテンツの真正性を見極める技術も同時に確立する必要がある．

音メディア認証（Sound media authentication, SMA）技術では，ある時刻やある場所で起こったイベントの「音響信号表現」として，収録した音の真正性を見極める技術を確立することを狙いとしている．もっとも重要なことは，それがオリジナルの収録であるかどうか，あるいは音声符号化・復号化を経由して別のデータフォーマットへコピーされたかどうかを認証することである．特に，音声符号の自動識別法は，音声収録後の改ざんを検出するために提案されたもの[11]であり，観測された信号から音声符号の種類を自動的に検出し，検出された符号の性質を SMA 技術で利用している．SMA 技術では，CELP（Code Excitation Linear Prediction）のような音声符号だけでなく，様々な符号化音声や自然音声に対しても，オリジナルデータの真正性を見極められるように取り組まれている．

現状をみると，このアプローチは，音声符号化のようなパラメトリック分析合成技術に対して非常に有効な手法であるが，ノンパラメトリック分析合成技術（例えば波形編集型音声合成や STRAIGHT といった合成音の自然性が非常に高いもの）による合成音声に対しては，音声合成法（音声符号化）の自動識別が可能であるか疑問が残る．そのため，汎用的な音声の情報改ざん検出として SMA 技術を確立するにはまだ時間がかか

るものと予想される．

真正性を見極めるためのもう一つのアプローチは，フラジャイル電子音響透かし技術（例えば，Wu & Kuo の方法[12]など）を利用することである．一般に， 電子音響透かし技術[14]は，三つの条件(a) 知覚不可能性，(b) 頑健性，(c) 秘匿性のすべてを満たさなければならないが，フラジャイル電子音響透かしは，条件(b)に関して，状況とともに頑健性と脆弱性のバランスを取る必要がある．つまり，悪意あるすべての攻撃に対して脆弱であり，悪意のない情報処理に対しては頑健である必要がある．しかし，実際の方法が悪意あるすべての攻撃に対して本当に脆弱であるかどうか，リサンプリングや音声符号化・復号化，情報圧縮といった悪意のない情報処理に対して頑健であるかは，明らかになっていない．また，実際の方法が情報改ざんを受けた位置の特定やその様態を正確に検出できるかどうかも定かではない．

研究代表者らは，これまでに，蝸牛遅延（cochlear delay，CD）に基づいた電子音響透かし法（以後，CD 法と呼ぶ）を提案してきた[15,16]．この方法では，オリジナルの音信号に知覚不可能な透かし情報を埋め込み，オリジナルの音信号を利用してもしなくても（ブラインド／ノンブラインド），正確にかつ頑健に透かしの埋め込みデータから透かし情報を検出できる．CD 法は，リサンプリングや量子化，情報圧縮（MP3）といった情報処理に対して頑健であり，音声の情報改ざんという意味で悪意ある攻撃（情報操作）に対して脆弱であるため，改ざん検出法として期待がもたれている．しかし，一部の音声符号化（G.729）に対して頑健でないため，改良が求められていた[17]．

本報告では，蝸牛遅延に基づいた電子音響透かしを利用した音声改ざんの検出法を概説するとともに，上記の問題点を解決した新しい電子音響透かし法とそれを利用した音声改ざん検出法を提案する．

3 音声改ざん検出の概要

図１に，情報改ざん検出の概略[16]を示す．ここでは，オリジナルの音信号の真正性を見極める方法を考える．ただし，本稿で取り扱う音信号の真正性とは，収録した音信号が流通し後で，その音信号が情報改ざん（あるいは何らかの改変）を受けたか受けなかったかを指すものである．収録側では，オリジナルの音信号と秘密データ（透かし情報）をペアとして保存しておき，情報ハイディング技術（ここでは知覚不可能な電子音響透かし法）を利用して，秘密情報（透かし情報）を音信号に埋め込み，公開するものとする．一般ユーザーは，許可を得て流通したオリジナルデータを取り扱うが，その中に透かし情報が埋め込まれているかどうかを知っていてもいいし，知らなくてもよい．ここで，悪意のある第３者が音声分析合成技術などを利用してオリジナルの音信号の内容そのものを改変し，更にそれをオリジナルのものだと偽って公開したものとする．さて，どのようにしてオリジナルの音信号の真正性を見極めるとよいだろうか．

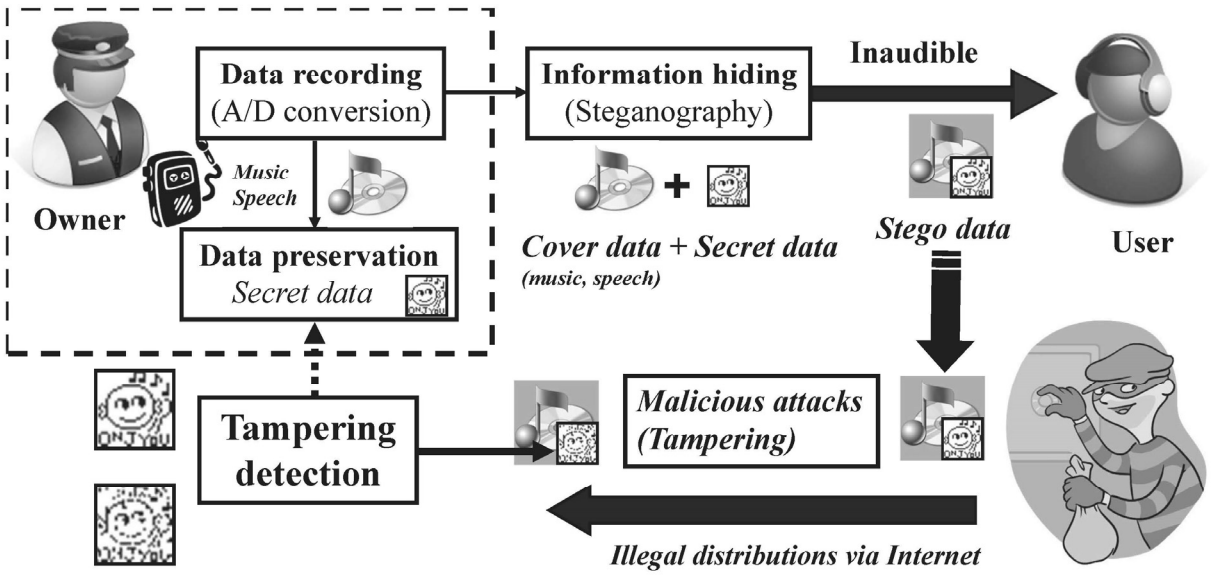


図１　音情報の改ざん検出の概略

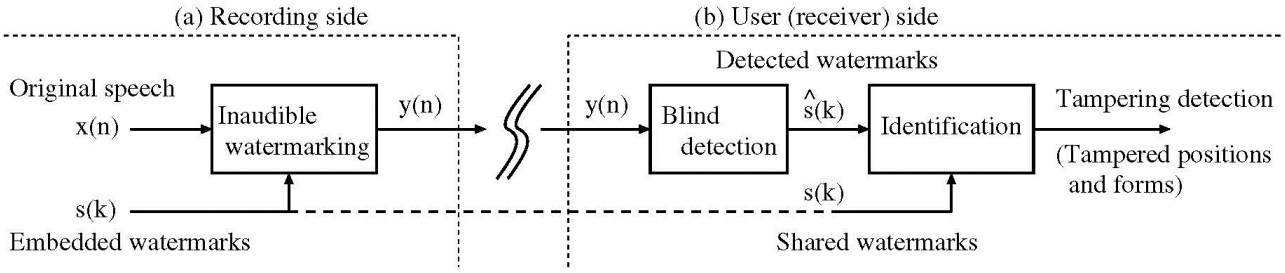


図２　知覚不可能な電子音響透かしを利用した改ざん検出のブロックダイアグラム

仮にフラジャイル電子音響透かし法を利用して，秘密データを音信号に埋め込んだ場合であれば，あらゆる情報処理によって埋め込まれた秘密データはもろくも壊れた状態で検出されるだろう．ここで，もし埋め込まれた秘密データが悪意のある改ざんに対してのみ脆弱で，音声符号化・復号化や情報圧縮，リサンプリングといったディジタル音データのフォーマット変換など（いわゆる悪意のない情報処理）には頑健であるとすれば，当初の狙いのとおり，検出された秘密データの状態から，情報改ざんの有無を検出できるものと考えられる．本研究では，このような処理を次のように実現する．

図２に，電子音響透かしを利用する改ざん検出法のブロックダイアグラムを示す[16]．この方法は，知覚不可能な電子音響透かし法による透かしデータ（秘密データ）の埋め込み部，検出部，識別部の三つの処理ブロックから成る．まず，収録側では，電子透かし法を利用して，改ざん検出に利用する透かし情報（秘密データ）s(k)をオリジナルの音声信号 x(n)に埋め込み，透かし入り信号 y(n)を生成する．次に，ユーザー（受信）側は，改ざん検出に利用する透かし情報 s(k)を収録側から受け取り，共有できる状態とする．検出処理は，受信した y(n)から透かし情報 s(k)をブラインドで検出する．最後に，識別部では，透かし情報 s(k)とユーザー側で検出された透かし情報 s(k)が同じものであるか識別し，改ざんを受けた位置を特定する．可能であれば，改ざんの様態まで特定する．次節では，研究代表者らによって既に提案された CD 法による音声改ざん検出法について概説する．

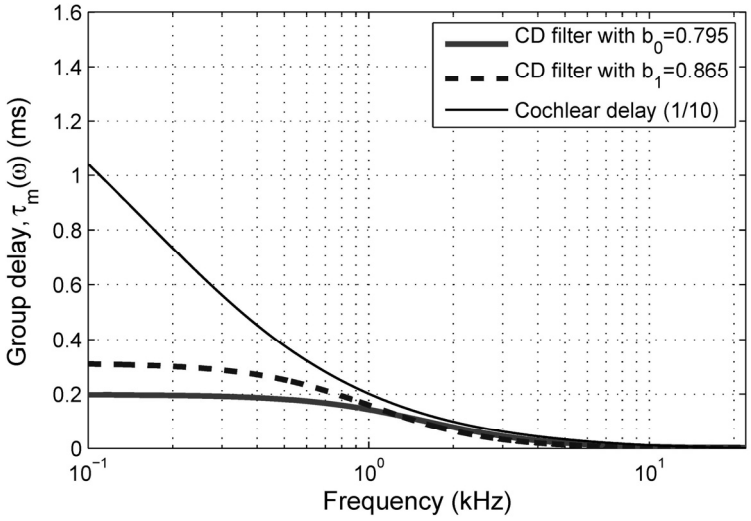


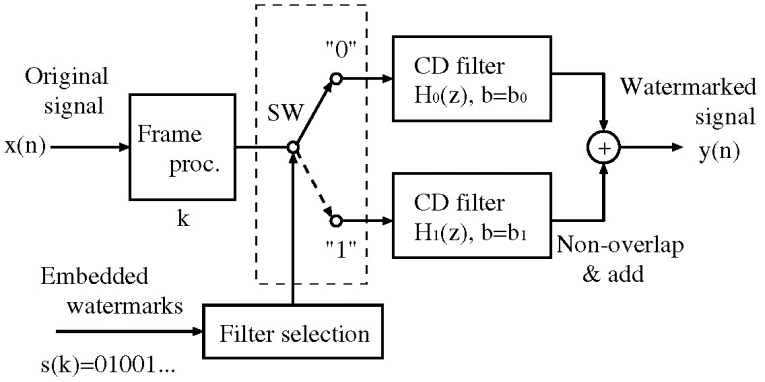
図３　蝸牛遅延特性と IIR オールパスフィルタの群遅延特性

4 蝸牛遅延に基づく電子音響透かしを利用した改ざん検出法

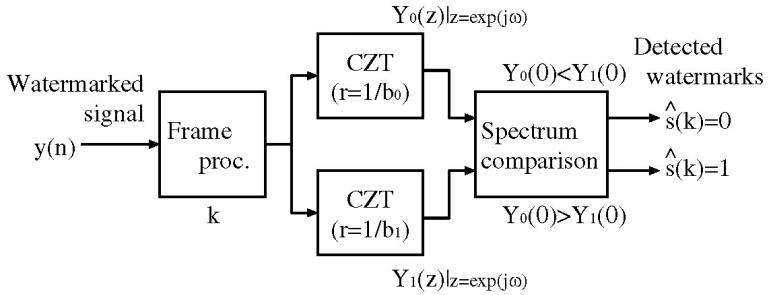
4-1 埋め込み部

蝸牛遅延とは，信号の周波数成分に依存した蝸牛内の基底膜振動で見られる進行波の伝搬の遅延（図３参照）のことである．この蝸牛遅延と音の同時性判断に関する心理物理学的な検討結果から，蝸牛遅延に似た遅延パターンをもった音と原音の弁別が非常に難しく，我々人間の聴覚系はこういった遅延に対して鈍感な

システムであることが示唆されている[15]．研究代表者らは，この特性に着目し，図３に示すように，電子透かし情報の二値データ（0 と 1）に対応する二種類の異なる蝸牛遅延に沿った遅延特性を原信号に付与することで，電子音響透かしを実現した．



(a) 透かし情報の埋め込み方法



(b) 透かしの検出方法

図４　蝸牛遅延に基づいた電子音響透かしのブロックダイアグラム：(a) 埋め込み法と(b) 検出法

図４は，上述のアイディアに基づいた電子音響透かしのブロックダイアグラムを示す．まず，蝸牛遅延を模擬した，二つの IIR 全域通過フィルタ（以後，蝸牛遅延フィルタと呼ぶ）$H_0(z)$ と $H_1(z)$ を異なるパラメータ値 b_0 と b_1 を用いて設計する（ここでは，b_0=0.795 と b_1=0.865 である）．次に，図４(a)に示すように，例えば，オリジナルの信号をフレーム分割（K 個）し，１フレームに 1 bit 割り当てる形で透かし情報 s(k)を埋め込む．ここでは，s(k)=01001...とする．最後に，各フレームで，透かし情報 s(k)（0 あるいは 1）のビット値に対応した蝸牛遅延フィルタ $H_m(z)$（$H_0(z)$ あるいは $H_1(z)$）を選択し，オリジナルの音信号 x(n)をフィルタリングする．ここでは，$y(n)=-b_m x(n)+x(n-1)+b_m y(n-1)$ として x(n)へのフィルタリングを行う．ここで，サンプル値 n の範囲は$(k-1)\Delta W \leqq k\Delta W$ であり，k はフレーム番号，$\Delta W=f_s/N_{bit}$ は，フレーム長である．f_s はサンプリング周波数であり，N_{bit} は透かしの埋め込み時のビットレート（bps）である．また，隣り合った二つのフレーム間での透かし入り信号 y(n)（$y(k\Delta W-1)$ と $y(k\Delta W)$）で不連続性が生じないようにするために，各フレーム k における y(n)の最後の数サンプル（0.5-ms のサンプル点）を Spline 補間によりスムージング処理を施してある．

4-2 検出部

先行研究では，オリジナルの音信号 x(n)を利用してノンブライドで透かし情報 s(k)を検出する方法[15]と，オリジナルの音信号 x(n)を利用せずに透かし入り信号 y(n)からブラインドで透かし情報 s(k)を検出する方法[16]の二種類がある．前者の場合，y(n)から s(k)を正確に検出するために，電子音響透かしで利用した蝸牛遅延フィルタ $H_m(z)$ の群遅延に一致する位相差を直接利用できる．後者の場合，ブラインド検出の手がかりとして，蝸牛遅延フィルタ $H_m(z)$ の極と零点の関係を利用できる．音声情報の改ざん検出を実現するという意味では，ブラインド法が適切であるため，提案法の透かし情報の検出には，透かし入り信号 y(n)のみが利用可能であるとした．

図４(b)は，透かし情報 s(k)のブラインド検出のブロックダイアグラムを示す．透かし情報 s(k)は，次の

三つの手順で検出される[16].(1) 零点の情報として,$r=1/b_0$と$r=1/b_1$の条件での二つのチャープz変換(CZT)を利用して，透かし入り信号y(n)を分析する．(2) 透かし情報s(k)を見つけるために，CZTスペクトル$Y_0(0)$と$Y_1(0)$の相対的に最小となるスペクトル値を比較する．(3) フレーム番号kが全フレーム数に到達するまで，これらの処理を繰り返す.

4-3 識別部

識別部では，検出された透かし情報s(k)のすべてが，収録側の透かし情報s(k)に一致するかどうかを調査する．もし，検出された透かし情報s(k)にて，不一致な点が一つもなければ，これは，ユーザー側で観測されたy(n)がオリジナルとして収録されたものであり，情報改ざんを受けていないことを意味する．もし，検出された透かし情報s(k)にて，少なくとも一つ以上の不一致な点があれば，フレーム番号kに該当する，透かし入り音声y(n)のフレーム内で,改ざんを受けた可能性を示し,収録側の透かし情報s(k)と不一致なs(k)の観測パターンが，その改ざんの様態を示すことになる．この不一致に関して，どのように識別するかについては，評価結果のところで具体的に説明する.

4-4 問題点

CD法の具体的な評価結果に関しては，6節で詳細に述べるが，ここでは主要な問題点を簡便に述べる．CD法を埋め込みビットレート 4 bps として音声信号に適用した場合，埋め込みによる音声の品質は PESQ（Perceptual Evaluation of Speech Quality）[18]で約4.0 (ODG)，LSD (Log-Spectrum Distortion) [19]で約0.5 dBであり，劣化がほとんど気にならない程度であることがわかっている[17]．また，このときのビット検出率は100%であり，サンプリング周波数や量子化ビットの変更にもまったく影響を受けないことがわかっている[17]．更に，雑音付与，残響付与，素片接続型音声合成を想定した改変に関してはビット検出率が約50%台になり，脆弱性を有していることもわかっている[17]．しかし，改ざんには該当しない処理として，PCM符号系（G.711とG.726）ならびにCELP符号系（G.729）といった音声符号化を利用した場合，埋め込みした透かしのビット検出率がG.711で100%，G.726で約80%，G.729で約50%まで低下することがわかった．これらから，CD法は音声改ざん検出として利用可能性が非常に高いものであるが，音声符号化処理に対する頑健性を増す必要がある（特にG.729に対して耐性を持たなければならない）ことがわかった.

次節では，この問題点を解決するために，フォルマント強調という新しい考え方に基づいた電子音響透かし法とそれを利用した音声改ざん検出法を提案する.

5 フォルマント強調に基づく電子音響透かしを利用した改ざん検出

5-1 音声符号化と音源フィルタ理論

CELP 系を代表とする多くの音声符号化手法では，線形予測符号化（Linear Predicted Coding）が利用されている．線形予測分析では，相関性のある時系列信号を線形予測フィルタ（LP係数）とその残差信号（LP残差）のフィルタ処理として表現できる．音声の音源フィルタ理論と合わせて考えると，線形予測フィルタは声道特性を表し，残差信号は音源特性を表すことになる．上述したようにG.711やG.726は主にPCM符号系であるため，波形への位相情報として埋め込んだ透かし情報はそのまま波形ベースで符号化され伝送されることになる．しかし，G.729 のように線形予測符号化ベースの符号化では，波形への位相情報として埋め込んだ情報は主に音源情報に含まれるため，ほとんど伝送されないことになる．これは，CD法が音声符号化処理に頑健でない主要な理由であった.

本研究では，この問題に対応するために，音声符号化に特化して電子音響透かし法の実現を検討する．そこで，線形予測符号化にも対応できる埋め込み法を拡張するため，線形予測フィルタで表現されるフォルマント情報（スペクトル包絡線情報）に積極的に知覚不可能な形で透かし情報を埋め込む方法を考える.

5-2 フォルマント強調

音声分析合成系やHMM音声合成を利用した音声合成技術では，音声の品質をいかに高めるか（合成音の自然性をいかに高めるか）が主要な課題になっている．音質を向上させる代表的な方法として，フォルマント

強調という技法が知られている．フォルマントは声道の共振現象に対応しており，音声のスペクトル包絡上の特定の周波数ピーク（エネルギーの集中しているところ）を指す．一般に，第 1・第 2 フォルマント（低い周波数から順番に数え上げたときの第 1，第 2 番目）は，音韻性に強い関わりを持っており，特に母音弁別で利用される重要な特徴である．そのため，フォルマント自体を操作する場合，フォルマントピークの移動は音韻性を加工することにつながるため避けなければならない処理である．また，スペクトル包絡におけるすべてのフォルマントを大幅に加工することは音色操作につながるため同様に避けなければならない．しかしながら、各フォルマントの共振性，つまりフィルタでみたときの Q 値を加工することは音韻性に影響を与えず，自然性・明瞭性を向上させる意味でよい音色操作になることが知られている．これがフォルマント強調処理である．

図５に線形予測分析により得られた１つのフォルマントの例を示す．ここで，フォルマント形状を帯域通過フィルタに見立ててその Q 値（中心周波数／帯域幅）を操作すると，図中の黒の点線を青の実線のようにピーク形状を盛り上げながら急峻にすることができる.線形予測符号化では LP 係数を直接得ることができるが，この係数の感度は非常に高いため安易にかつ直接的に操作できない．そこで，LP 係数を安定に操作するために線スペクトル周波数（LSF）対を利用する．ここでは，図５に示すように中心周波数 fc に対して対称になるように LSF の対を中心周波数に向けて移動操作することで Q 値を高める（フォルマントを強調する）処理を行う.

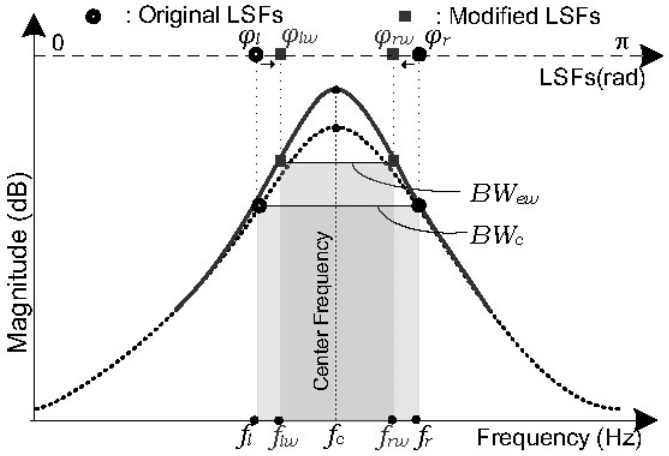


図 5　LSF 制御によるフォルマント強調

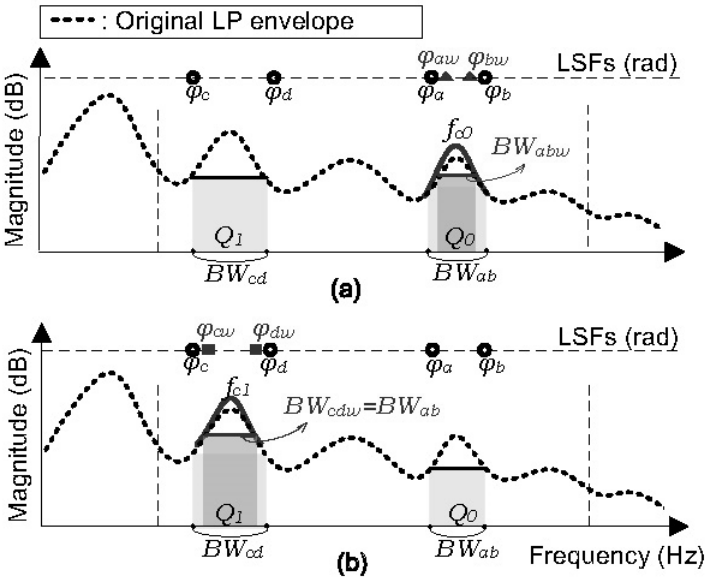


図 6　フォルマント強調による情報埋め込みの概要

5-3 フォルマント強調を利用した透かしの埋め込み法と検出法

フォルマント強調は，上述したように音声の音質・自然性を高めるためによく利用される技術であるが，本研究の独創的なアイディアは，特定のフォルマントの Q 値を操作して強調しても，音質・自然性が向上するだけで何も違和感を生じないことから，これを逆手にとって透かし情報を知覚されないようにスペクトル包絡にフォルマント強調の度合として隠すということである．

このときの透かし情報の埋め込み法の概要を図６に示す．まず，透かし情報としてビット情報である 0 あるいは 1 を埋め込むことを考える．ここではそれぞれに対応するフォルマント位置として Q_0，Q_1 を定め，埋め込み情報に対応して，Q_0 あるいは Q_1 のフォルマントを強調する．提案法では，2 次以上(p/2)次未満のフォルマントを対象に Q_0 と Q_1 を定めている．ただし p は LP 次数である．ここでは，0 を埋め込むときは，図６(a)のように，1 を埋め込むときは図６(b)のようにフォルマントを強調する．

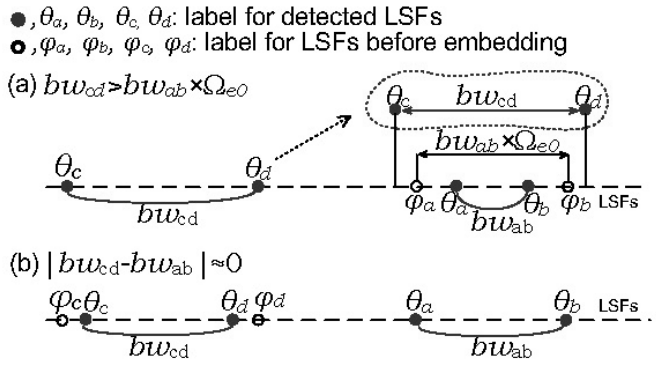


図７　フォルマント強調により埋め込んだ情報の検出の概要

次に，図６により埋め込んだ透かしを検出するために，図７に示す方法をとる．ここでは Q_0 と Q_1 のフォルマントからそれぞれ帯域幅を求め，どちらの帯域幅が狭いか（強調されているか）を探ることで埋め込みした透かし情報を得る．例えば，図 7(a)のように Q_0 の帯域幅が Q_1 の帯域幅より狭いため，フォルマントが強調されたことがわかるため，透かし情報として 0 が検出される．反対に，両方の帯域幅がほぼ等しいときは Q_1 が強調されたことを意味するため，図 7(b)のように透かし情報として 1 が検出される．これらの情報は，すべて観測した LP 係数（フォルマント情報）から得られるため，ブラインド処理として実現できる．

5-4 改ざん検出法

図２の改ざん検出法を実装するために，ここでは図６と図７の処理を利用して図８のように実装した．

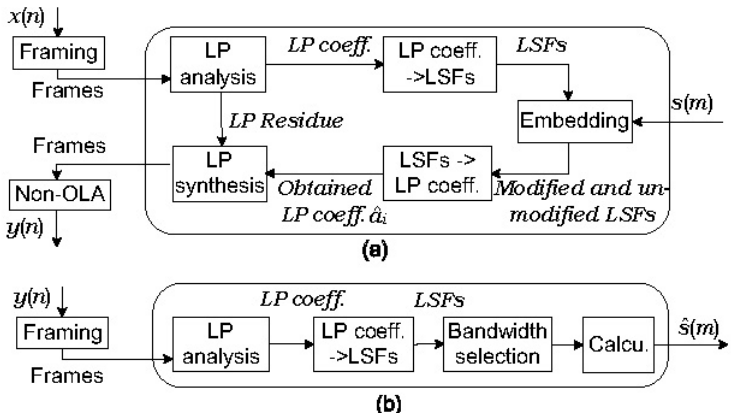


図８　フォルマント強調に基づいた電子音響透かし：(a) 埋め込み法と(b) 検出法

図 2 の三つの処理ブロックに関しては，図 8(a)を知覚不可能な電子音響透かしとして，図 8(b)をブラインド検出として組込み，識別処理は 4.3 節と同様の処理で実現した．次節では，CD 法ならびに提案法の有効を確認するために評価を行う．

6 評価

6-1 知覚不可能性と検出率に関する評価

はじめに，二つの客観評価尺度を利用して，知覚不可能性を評価した．ここでは，PESQ（perceptual evaluation of speech quality）[18]と対数スペクトル歪み（log-spectrum distortion, LSD）[19]を調べた．また，提案法により正しく埋め込んだ情報を検出できるかどうか判断するために，ビット検出率（BDR）を調べた．

これらの評価は，音声データを対象とした点を除けば，Unoki & Hamada [15]と Unoki & Miyauchi [16]で利用したものと全く同じ評価である．ODG（objective difference grade）は，-0.5 から 4.5 の 5 段階で評定付けされ，その数値は，MOS 値で 1（非常に悪い；劣化が非常に気になる）から 5（非常によい；劣化がまったく気にならない）に対応するものである．PESQ の ODG 値で 3 以上と LSD で 1 dB 以下の評価基準を知覚不可能性の判断基準とした．90%のビット検出率も検出能力の判断基準として利用した．本節では，提案法の有効性を確認するために，大まかに三つのステップで評価した．ここでは，ATR 音声データベース（B セット）[20]にある 12 個の音声刺激すべて（男性／女性の日本語文章）を利用した．オリジナルの刺激は 20-kHz サンプリング周波数，16-bit 量子化で収録され，その長さは 10-sec であった．実験で利用したビットレート N_{bit} は，4 bps であった．埋め込みには，提案法の他に比較対象として CD 法と，代表的なものである LSB（leastsignificant bit-replacement）法[21]を利用した．

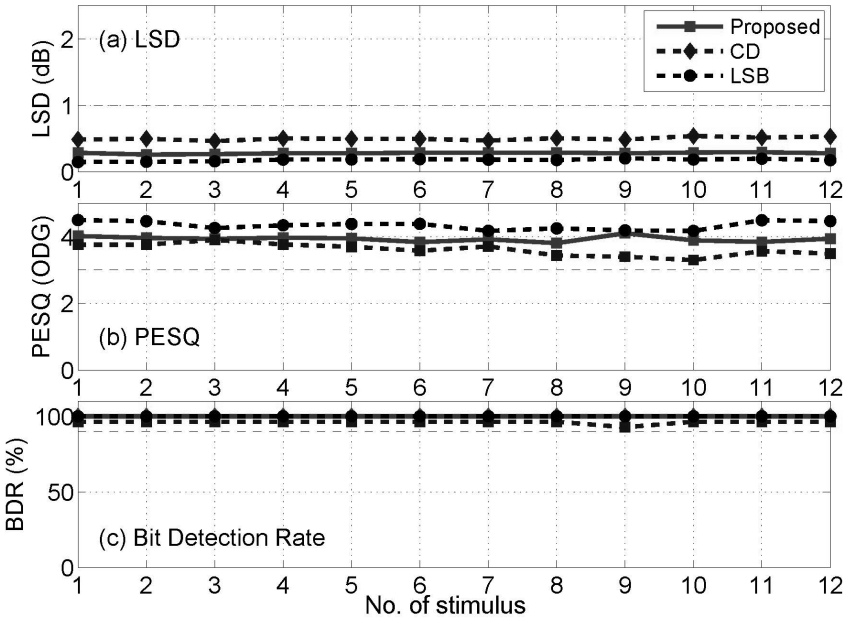


図 9　客観評価実験の結果

図 9 に LSD，PESQ，ビット検出率（BDR）の結果を示す．図中の横軸は刺激番号を指す．CD 法，提案法ならびに代表的な方法（LSB 法）いずれとも，三つの評価尺度で設けた基準値を満たしており，知覚不可能な形で透かしを埋め込み，それを正しく検出できることがわかる．

6-2 頑健性に関する評価

提案法での検出処理の頑健性を評価した．PCM（pulse code modulation）と適応差分 PCM（adaptive

differential PCM, ADPCM) といった様々な音声符号化・復号化 (G.711 & G.726 [22]) は，インターネットや電話会議システムなどディジタル音声を符号化するために，現在の音声工学で幅広く利用されている．これらの処理は悪意のないものであるため、検出処理ではユーザー側で観測された符号化・復号化後の音声信号 y(n)から正確に透かし情報 s(k)を検出できなければならない．

図 10 は、提案法と CD 法，LSB 法に関する頑健性の評価結果を示す．前述したように，CD 法（青の点線）では，PCM 符号系 (G.711 & G.726) に関する結果から十分に基準を満たすものとなっているが，CELP 符号系 (G.729) では検出率が約 50%であることがわかる．これに対し，フォルマント強調を利用した電子透かしでは，いずれの符号化でも基準値を上回っており，十分に頑健性を満たすことがわかる．参考として挙げた LSB 法では，いずれの符号化処理にも脆弱であり，基準を満たさないことがわかる．

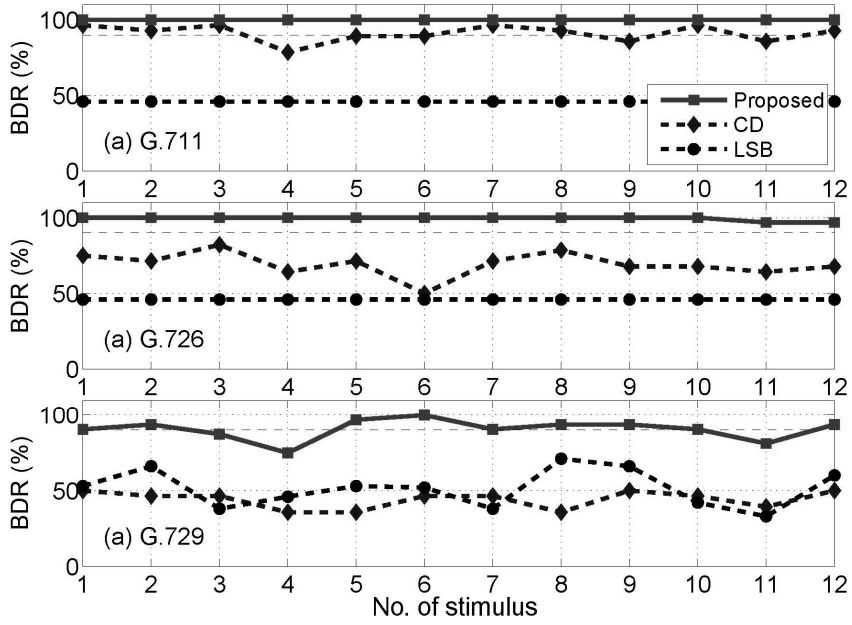


図 10　音声符号化に対する頑健性の評価結果

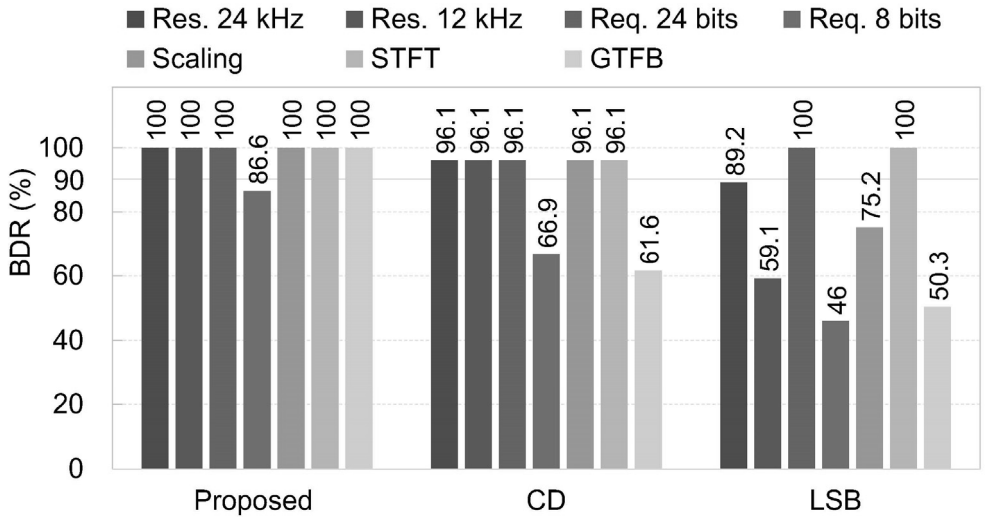


図 11　主要な信号処理に対する頑健性の評価結果

次に，音声符号化以外の一般的な信号処理として，再サンプリング（アップサンプリング 16 kHz→24 kHz，ダウンサンプリング 16 kHz→12 kHz)，量子化ビットの変更 (16 bits→24 bits，16 bits→8 bits)，スケール変換，代表的な分析合成系として STFT (Short time Fourier Transform)や GTFB (Gammatone filterbank) を単なる分析・合成処理のみに利用して合成音を利用した場合での透かし情報の検出結果を調べた．このと

きの結果を図 11 に示す．提案法はダウンサンプリングのときを除けばすべて 100%の検出率を得た．ダウンサンプリングのときは約 87%まで低下したがそれほど大きな問題ではない．これは，ダウンサンプリングにより 4〜8 kHz のフォルマント強調処理が影響を受けたことによる当然の結果であり，事前に低いサンプリング周波数で埋め込み処理を行っておけば回避できる．CD 法も同様の結果となっているが，GTFB は一種の帯域制限をする処理になっているため低域に埋め込まれた透かし情報が欠落したことが原因である．LSB 法は一般的に頑健性が低い処理であり，その結果が同様に観測される．

表 1　脆弱性の結果（提案法）

改ざんの種類	ビット検出率（%）	改ざんの種類	ビット検出率（%）
白色雑音の付与	45.19	音声の切り貼り	42.86
残響の付与	68.80	話速アップ（+4%）	71.56
低域通過フィルタリング	41.98	話速ダウン（-4%）	79.51
高域通過フィルタリング	49.85	ピッチシフト	68.12

6-3 脆弱性に関する評価

提案法の脆弱性を，先に検討した項目[17]に沿って評価した．CD 法ならびに LSB 法については本報告では割愛する．ここでは，図 12(a)に示すようなビットマップイメージを透かし情報 s(k)として，提案法を利用してオリジナルの音声信号 x(n)に埋め込んだ（4 bps）．その後で，次のような編集により，透かし入りの音声信号 y(n)を改ざんした．

(c) 雑音の付与
(d) 残響の付与
(e) 波形編集型音声合成による編集（音声の切り貼り）
(f) 低域通過フィルタリング
(g) 高域通過フィルタリング
(h) 話速アップ（+4%）
(i) 話速ダウン（-4%）
(j) ピッチシフト（-4%）

ここで，雑音・残響の付与ならびに低域・高域通過フィルタリングは，一度ディジタル信号を DA 変換してスピーカー等で出力したものを再び AD 変換してディジタル信号に戻したこと(AD/DA 変換を介した音の再収録)を想定したものであり，信号の真正性を疑う処理として想定した．次に，音声の切り貼りは，波形編集型音声合成そのものにより音声改ざんを意味している．最後に，話速度の変換やピッチシフトは話者性などを変換することを意図したものである．

表 1 にこれらの攻撃に対する脆弱性の結果を，図 12 に具体的な透かしの検出例を示す．図 12 から，いずれも改ざんの種類によって透かし情報がパターン化した形で破損していることがわかる．このような壊れ方のパターンは提案法ならびに CD 法で若干異なるが，おおよそ類似した傾向にあった．特に，一番重要な音声改ざん (e) では，透かし情報がすべて欠損していることから切り貼り等の攻撃を受けたことが容易に予想できる．今後は詳細な追加検討が必要であるが，提案法や CD 法は改ざんを受けた位置だけでなく，改ざんの様態についても検出できることがわかる．

7 まとめ

本稿では，フォルマント強調に基づいた電子音響透かし法とこれを利用した改ざん検出法を提案した．三つの客観評価試験（PESQ，LSD，ビット検出率），音声符号化・復号化に対する頑健性試験，悪意ある攻撃に対する脆弱性試験を行うことで提案法を評価した．その結果をまとめると，1 番目と 2 番目の評価結果から，提案法が知覚不可能性と悪意のない情報処理に関して頑健性を有することがわかった．最後の評価結果から，提案法が音声コンテンツの改ざんといった悪意ある編集に対して脆弱であることがわかった．また，提案法が改ざんを受けた位置だけでなく，改ざんの様態についても検出できることを明らかにした．

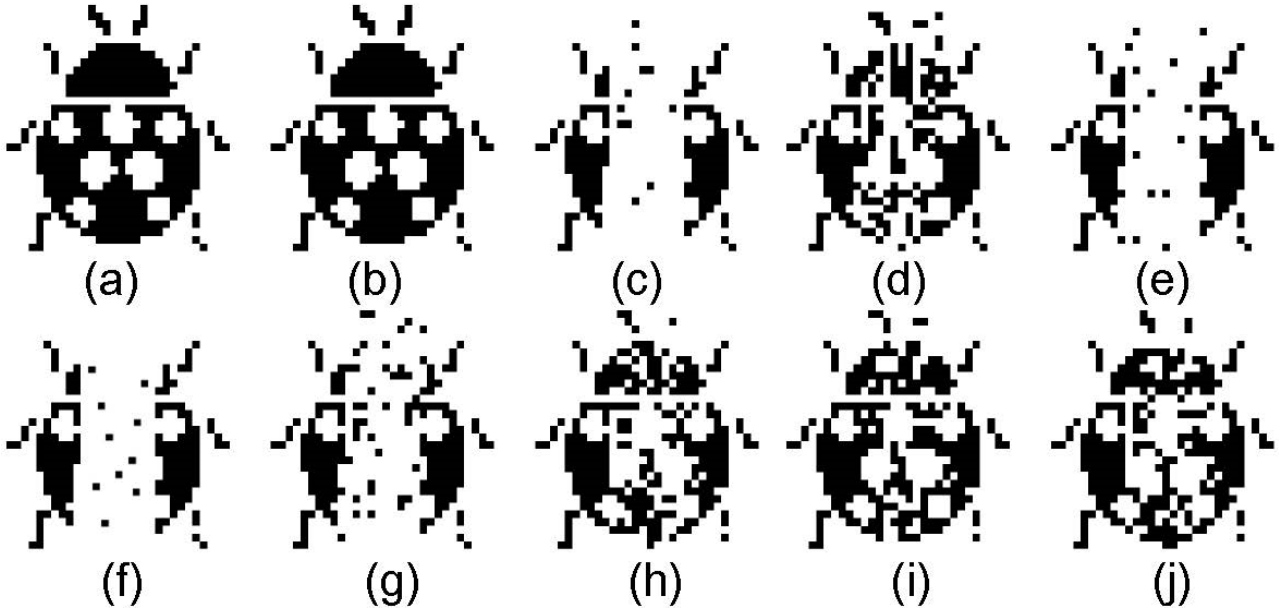


図 12　改ざんに対する脆弱性の評価結果：(a) オリジナルの透かし情報，(b) 通常時に検出された透かし情報，(c) 白色雑音の付与，(d) 残響付与，(e) 音声の切り貼り，(f) 低域通過フィルタリング，(g) 高域通過フィルタリング，(h) 話速アップ，(i) 話速ダウン，(j) ピッチシフト後に検出された透かし．

【参考文献】

1) 河原英紀:音声分析合成技術の動向, 日本音響学会誌, Vol. 67, No. 1, pp. 40-42, 2011.
2) Kawahara, H., Morise, M., Takahashi, T., Nishimura, R., Irino, T., and Banno, H.: Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, Proc. ICASSP2008, pp. 3933-3936, 2008.
3) 清山信正: 番組制作に利用される音声合成技術とその研究動向, NHK 技研 R&D, Vol. 131, pp. 14-19, 2012.
4) Toda,T., Kawai, H., Tsuzaki, M., and Shikano, K.: An evaluation of cost functions sensitively capturing local degradation of naturalness for segment selection in concatenative speech synthesis, Speech Communication, Vol. 48, No. 1, pp. 45-56, 2006.
5) Toda, T., Black, A. W., and Tokuda, K.: Voice conversion based on maximum likelihood estimation of spectral parameter trajectory, IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, No. 8, pp. 2222-2235, 2007.
6) H. Kawahara, H., Banno, H., Irino, T., and Zolfaghari, P.: ALGORITHM AMALGAM: Morphing waveform based methods, sinusoidal models and STRAIGHT, Proc. ICASSP2004, pp. 13-16, 2004.
7) Huang, C.-F. and Akagi, M.: A three-layered model for expressive speech perception, Speech Communication, Vol. 50, No. 10, pp. 810-828, 2008.
8) Moriyama, T., Mori, S., and Ozawa, S.: A synthesis method of emotional speech using subspace sonstraints in prosody, Trans. IPS Japan, Vol. 50, No. 3, pp. 1181-1191, 2009.
9) Saitou, T., Unoki, M., and Akagi, M.: Development of an F0 Control model based on F0 dynamic characteristics for singing-voice synthesis, Speech Communication, Vol. 46, pp. 405-417, 2005.
10) Saitou, T., Goto, M., Unoki, M., and Akagi, M.: Vocal Conversion from Speaking Voice to Singing Voice Using STRAIGHT, Proc. Synthesis of Singing Challenge, Special Session at INTERSPEECH 2007, Antwerp, Belgium, 2007.
http://www.interspeech2007.org/Technical/synthesis_of_singing_challenge.php

11) Zhou, J., Garcia-Romero, D., and Espy-Wilson, C.: Automatic Speech Codec Identification with Applications to Tampering Detection of Speech Recordings, Proc. Interspeech2011, pp. 2533-2536, 2011.
12) Wu. C.-P. and Kuo, C.-C. J.: Fragile speech watermarking based on exponential scale quantization for tamper detection, Proc. ICASSP2002, IV, 3305--3308, 2002.
13) Cvejic, N. and Seppanen, T.: Digital audio watermarking techniques and technologies, Idea Group Inc. (IGI), 2007.
14) Unoki, M. and Hamada, D.: Method of digital-audio watermarking based on cochlear delay characteristics, Int. J. Inn. Com. Inf., and Cont., Vol. 6, No. 3(B), pp. 1325-1346, 2010.
15) Unoki, M. and Miyauchi, R.: Reversible Watermarking for Digital Audio Based on Cochlear Delay Characteristics, Proc. IIHMSP2011, pp. 314-317, 2011.
16) Unoki, M. and Miyauchi, R.: Detection of Tampering in Speech Signals with Inaudible Watermarking Technique, Proc. IIHMSP2012, pp. 118-121, Greece, July 2012.
17) Yi, H. and Philipos, C. L.: Evaluation of objective measures for speech enhancement, Interspeech2006, pp. 1447-1450, 2006.
18) Lin, Y. and Abdulla, W. H.: Perceptual evaluation of audio watermarking using objective quality measure, Proc. ICASSP2008, pp. 1745-1748, 2008.
19) Takeda, K. et al.: Speech Database User's Manual, ATR Technical Report TR-I-0028, 1988.
20) Bassia, P. and Pitas, I. P.: Robust audio watermarking in the time domain, Proc. EUSIPCO1998, pp. 25-28, 1998.
21) http://www.itu.int/rec/T-REC/en

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Formant Enhancement based Speech Watermarking for Tampering Detection	Proc. Interspeech2014 (Accepted)	2014 年 9 月
Watermarking of Speech Signals Based on Formant Enhancement	Proc. EUSIPCO2014 (Accepted)	2014 年 9 月
Hybrid Speech Watermarking based on Formant Enhancement and Cochlear Delay	Proc. IIHMSP2014 (Accepted)	2014 年 8 月
Watermarking method for speech signals based on modifications to LSFs	Proc. IIHMSP2013 (pp. 263-266)	2013 年 10 月