

音声の生成過程に着目した電子透かし

畑田 充弘[†] 酒井 俊之[†] 小松 尚久[†] 山崎 恭^{††}

[†] 早稲田大学理工学部 電子・情報通信学科

^{††} 北九州市立大学国際環境工学部 情報メディア工学科

E-mail: †{mitsu,sakai,komatsu}@kom.comm.waseda.ac.jp, ††yamazaki@env.kitakyu-u.ac.jp

あらまし 音声の生成モデルは音源の生成, 調音, 放射の三段階に分けられる. 従来の音声の電子透かし手法は量子化値や符号化パラメータに着目したものであるといえる. 本稿では音声の生成過程において調音を操作することで音声に対する新たな電子透かし手法を提案し, 透かし入り声道モデルを適用することで提案手法を実現する. 提案手法では透かし入りの声道モデルを LSP によって作成されるコードブックによって表現され, 音声の電子透かしは, 1) 音声から LSP を抽出する, 2) 抽出された LSP をコードブックの LSP ベクトルと置き換える, 3) 置き換えた LSP を使って音声を合成する, という手順となる. 上記の過程において, 電子透かしは間接的に音声に埋め込まれる. 本稿では提案手法の精度と音声品質に関する評価について述べる.

キーワード 著作権保護, 電子透かし, 音声生成モデル, 調音, 線スペクトル対, コードブック

A Study on Digital Watermarking Based on Process of Speech Production

Mitsuhiro HATADA[†], Toshiyuki SAKAI[†], Naohisa KOMATSU[†],
and Yasushi YAMAZAKI^{††}

[†] Dept. of Electronics, Information and Communication Eng., WASEDA Univ.

^{††} Dept. of Information and Media Sciences, The Univ. of Kitakyushu

E-mail: †{mitsu,sakai,komatsu}@kom.comm.waseda.ac.jp, ††yamazaki@env.kitakyu-u.ac.jp

Abstract A speech production model can be divided into three parts, which are glottal source, articulation and radiation respectively. Some digital watermarks for speech that have been proposed are based on changing quantized values or parameters of a coding scheme. In this paper, we propose a new watermarking method for speech by manipulating articulation in process of speech production. That is, we realize the proposed method by applying watermarked vocal tract model. In the proposed method, the watermarked vocal tract model is expressed by codebook made by LSP(Line Spectrum Pair) parameters. The procedures of watermark for speech are as follows; 1) LSPs are extracted from speech. 2) Some of the extracted LSPs are replaced with the codebook vectors. 3) Speech is synthesized using replaced LSPs. In above process, watermark are embedded indirectly to speech. Evaluation tests on speech quality and accuracy of the proposed method will be discussed with simulation results.

Key words Copyright Protection, Digital Watermark, Speech Production Model, Articulation, LSP, Codebook

1. はじめに

デジタル技術の急速な発展とインターネットの普及に伴い、マルチメディアデータが簡単に複写、編集できるようになった。このような特性は非常に便利で魅力的である反面、著作権侵害に対する著作権保護の必要性が高まり、法制度の強化により不正行為者が厳しく罰せられる環境の整備が求められている [1]。この問題を解決する技術的手段の一つとして電子透かしが注目を集めており、種々のアルゴリズムが提案されている。

電子透かしとは人間の知覚（視覚、聴覚）特性を利用し、デジタルコンテンツに対してコンテンツ自体とは別の情報を、人間に知覚できないように埋め込む技術のことである。人間が知覚できないという点と秘密にされた情報を知ることができないという点では、暗号技術に類似性もあるが、暗号の場合コンテンツの暗号化を人間は認識できるが、コンテンツ自体が秘密情報となっているのに対して、電子透かしの場合は、コンテンツ自体は認識できるが、埋め込まれている情報が秘密となっている点で本質的に異なっている。

今後、音楽や画像、動画だけではなく様々なコンテンツ形態が考えられ、人間の音声を対象としたコンテンツの流通が想定される。例えば有名人の声から好きなフレーズを作成し、目覚し時計や携帯電話の呼出音などに組み込むサービスや、講演会等の音声に電子透かしを埋め込むことで信頼性を持たせるといった利用形態が考えられる。電子透かしは埋め込むコンテンツの特性を最大限に利用し、コンテンツに応じた透かしの手法がとられるべきであり、音声の電子透かしにおいても知覚的に重要な領域に埋め込む必要がある [2]。従来の音声に対する電子透かし手法は音声の波形に対して、量子化や周波数変換などの処理を加え、その性質を利用した手法、または符号化の特性を利用した手法であった [3]。著者らはこれら従来の手法とは異なり、音声としての特徴を考慮し、音声の生成過程に着目した電子透かしを実現する一手法を提案 [4] しており、本稿では透かし情報の生成による透かしの精度および音声品質の評価を行う。

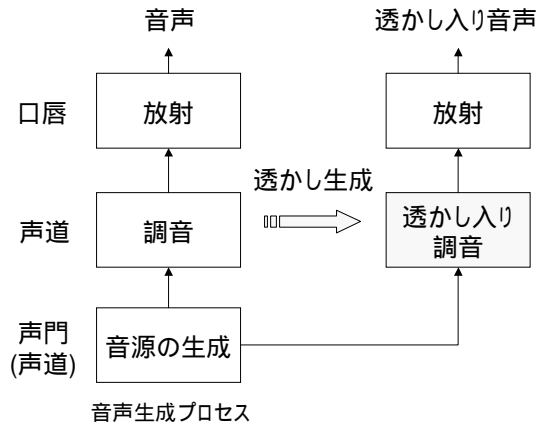


図 1 提案手法の概念

2. 声道情報への電子透かし

音声の生成過程は 1) 声帯の周期的な開閉運動により生じる、規則的な空気の振動による音源の生成、2) 様々な言語音を発生するために声道の形を調整する調音、3) 口唇からの放射の三段階で表される。従来の手法は、主に放射後の音声に対して透かしを埋め込む手法であり、我々が提案する手法は図 1 に示すように、透かし入りの調音として疑似的に個人の声道モデルを作成し、この透かし入りの調音を通して放射される音波には間接的に透かし情報が拡散されるものと考えられる。そのため、透かし位置の特定が困難となり、また符号化に特化した手法ではないため多くの符号化にも耐性が保てることが期待される。

本提案手法を実現する一手法として、音源を表すパラメータとしてピッチ、調音を表すパラメータとして LSP (線スペクトル対: Line Spectrum Pair) に着目し、LSP 分析合成系 [5] に提案手法を適用する。LSP は音素を特徴づける優性な周波数成分であるホルマントの両側に対になって立つ性質を持ち、この LSP が表す周波数を一部変更することは声道の形を疑似的に変えることにほかならない。

3. 透かしの埋め込み/抽出

電子透かし操作は、透かし情報生成、透かし埋め込み、透かし抽出の三つの独立段階で構成され、各段階について提案手法のアルゴリズムを以下に示す。

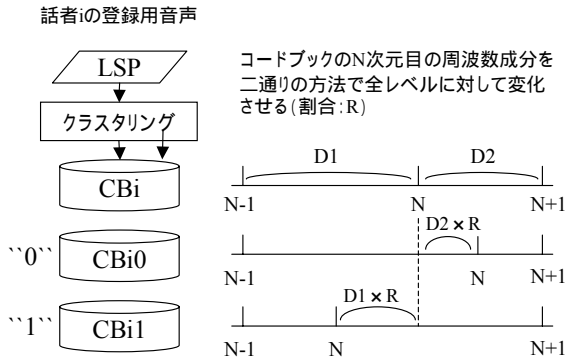


図2 透かし情報生成

3.1 透かし情報生成

図2に示すように、話者*i*の学習用音声からLSPを16次元で算出し、LBG+Splittingアルゴリズムを用いてクラスタリングを行い256レベルのコードブック(以下CB*i*)を作成する。このCB*i*は話者*i*の声道情報の特徴を反映したものであり、このCB*i*のLSPの周波数成分を二通りの方法で変化させ、それぞれに透かし情報0と1を割り当てて透かし入りのコードブック(CB*i*0, CB*i*1)を作成する。具体的な変化のさせ方としては様々な方法が考えられるが、ここではCB*i*のN次元目の周波数成分を全レベルに対して変化させる。N次元目とN-1, N+1次元目のそれぞれの距離をD1, D2とし変化させる割合をRとすると、CB*i*のN次元目の値にD2 × Rを加算したものをCB*i*0のN次元目とし、D1 × Rを減算したものをCB*i*1のN次元目とする。また、ベクトル量子化の際に用いる距離尺度としてLSPのユークリッド距離に、スペクトルのピーク周波数付近を重視する重み係数をかけたLSPの重み付きユークリッド距離[6]を用いた。

3.2 透かし埋め込み

透かし埋め込みの概要を図3に示す。透かし埋め込みの対象となる話者*i*の音声からLSPを算出し、CB*i*を用いて最も距離の小さいインデックスを算出する。ピッチとともに透かしを埋め込むフレームの選択を行う。ここで透かしを埋め込むフレームの選択が重要となり、これまでの検討により、埋め込むフレームによって透かしの抽出率が大きく変化することがわかっている。これはベクトル量子化

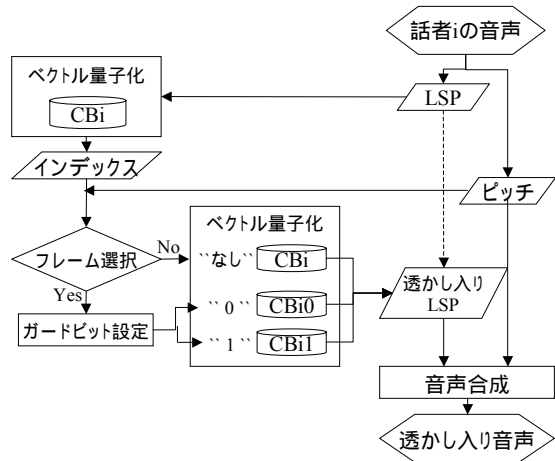


図3 透かし埋め込み

を含めたLSP分析合成系の前後において、LSPに誤差が生じるためであると考えられ、この誤差が大きいフレームほど抽出率が悪いことを実験により確認した。そこで、この誤差の小さいフレームを選択するに当たって、ピッチおよびインデックスが安定、つまり連続したフレームにおいて、それぞれが同じ値をとっていることを選択条件とする。また、音声合成時および透かし抽出時の線形予測の影響により、単一フレームに透かしを埋め込むと透かし情報が大幅に消失してしまうため、埋め込む情報を保護するためのガードビットとして前後のフレームも透かし埋め込みの対象フレームとする。透かし埋め込みの対象となるフレームに対しては、埋め込む透かし情報に合わせてそれぞれCB*i*0, CB*i*1から対応するインデックスのLSPをベクトル量子化により置き換える。透かし埋め込み対象外のフレームについてはCB*i*を用いてベクトル量子化を行う。これより透かし入りのLSPが生成され、ピッチとの合成により透かし入り音声生成される。

3.3 透かし抽出

透かし情報の抽出は、検査対象となる話者*i*の音声からLSPとピッチを算出し、CB*i*, CB*i*0, CB*i*1を用いて抽出を行う。透かしを埋め込むフレームは3.2で述べたピッチとインデックスが安定したフレームであり、音源の安定している有声音源であると考えられる。そこで、ピッチが0であるフレームは、音源が安定していない無声音源であると考え、

抽出の対象外とする。そして、フレーム毎のLSPと最も近いコードブックのレベルを算出するに当たって、透かしを埋め込んだN次元目のLSPを除いて三つのコードブックと距離比較を行い、それぞれのコードブックの中で最も近いレベルが同じレベルとなったら当該レベルとみなす。次に、透かしの有無を判定するために、透かしの生成時に設定した割合Rをもとに、前述のLSPの誤差を考慮して閾値を設定し、透かしの有無、あるいはCBi0,CBi1との距離比較により透かし情報の抽出を行う。ガードビットに応じた連続するフレームを対象として、透かし情報が抽出される頻度に基づき、埋め込んだ透かし情報と1フレーム毎の対応で透かしの抽出率を算出する。

4. 提案手法の評価

本節では3.1において,CBi0,CBi1を作成する際に、第三者による透かし位置の特定を困難にするためにも透かしを埋め込む次元は固定しないほうがよいと考えられる。また3.2で述べた透かし埋め込み前後のLSPの誤差の影響によって、透かし埋め込み次元によっても抽出率の変化が予測される。また、透かし情報生成の際に設定する,CBiのN次元目を変化させる割合Rが、透かしを埋め込んだ音声の品質と抽出率に大きく影響することが考えられ、本節では提案手法の評価として上記二点について各実験結果から考察を行う。

実験はATR標準音声を用いて行った。透かし埋め込みには男性話者50文章合計222秒、女性話者50文章合計233秒を使用し、透かしの埋め込み情報量としては約4.3[bit/s]となっている。3.2で述べた透かしを埋め込むフレームの選択条件としてピッチが連続3フレーム、インデックスが連続5フレーム同じ値をとるフレームを透かし埋め込みの対象とし、ピッチの抽出は困難を極めるため、理想的な環境でピッチが抽出できるものとし、ATR音声データベースに付属されているピッチを使用した。また、ガードビットとして透かしを埋め込むフレームの前後2フレームにガードビットとして透かしを埋め込む。本実験では3.3に示す透かし抽出の際に設定する閾値を、透かし情報生成時の変化させる割合Rの0.5倍とした。その他、実験に関する諸元を

表1に示す。

表1 諸元

登録用音声	男女話者(MYI,FYM)共に 5秒程度のATR標準音声50文章
透かし埋め込み音声	男女話者(MYI,FYM)共に 5秒程度のATR標準音声50文章
サンプリング周波数	12(KHz)
量子化ビット	16ビット
フレーム長	384ポイント(32msec)
フレーム周期	60ポイント(5.0msec)
LSP分析回数	16次元
コードブックレベル数	256レベル
窓掛け	ブラックマンウインドウ

4.1 透かし埋め込み次元と抽出率

LSP分析回数16次元それぞれの次元に対して、R=0.5として実験を行った結果を表2に示す。BER(Bit Error Rate)は透かしを埋め込んでいないフレームから0または1の情報が抽出された率を示す。全体としては男性話者で抽出率が約90%、女性話者で約87%、BERは男女話者共に約3%となっているが、透かしの埋め込み次元と埋め込む情報によって精度のばらつきが大きい。透かし情報0,1に対応する透かし埋め込み用CBi0,CBi1の生成方法が異なるため抽出率が異なっている。これは

表2 透かし埋め込み次元毎の抽出結果

次元	男性話者(MYI)			女性話者(FYM)		
	透かし抽出率		BER	透かし抽出率		BER
	情報"0"	情報"1"		情報"0"	情報"1"	
0	0.77	0.97	0.03	0.46	0.99	0.03
1	0.92	0.74	0.03	0.95	0.43	0.05
2	0.94	0.81	0.03	0.96	0.69	0.02
3	0.87	0.97	0.02	0.94	0.90	0.02
4	0.93	0.88	0.02	0.84	0.98	0.03
5	0.94	0.84	0.02	0.96	0.82	0.03
6	0.92	0.96	0.02	0.90	0.91	0.02
7	0.94	0.95	0.02	0.93	0.93	0.02
8	0.95	0.92	0.02	0.85	0.97	0.02
9	0.92	0.96	0.02	0.99	0.78	0.02
10	0.94	0.94	0.02	0.93	0.96	0.02
11	0.95	0.90	0.02	0.94	0.97	0.02
12	0.94	0.92	0.02	0.91	0.92	0.02
13	0.95	0.95	0.02	0.97	0.92	0.02
14	0.93	0.96	0.02	0.97	0.94	0.02
15	0.97	0.27	0.11	0.99	0.31	0.10

3.2 で述べた透かし埋め込み前後の LSP の誤差が LSP の次元によって平均的に高周波方向または低周波方向へ誤差が含まれるためあると考えられる。透かしとして変化させた周波数方向が、この誤差を多く含む周波数方向であった場合、その LSP の次元の周波数値は拡大され、逆の場合縮小される。

これらの誤差を統計的に判断し、抽出率の高い透かし埋め込み次元と透かし情報を見極めることで、透かしの精度の向上が期待できる。また、透かしの埋め込み可能情報量の改善策として複数の次元に対して CBI の LSP 周波数成分を変化させることが考えられ、複数の次元を選ぶ際にも有効に活用できると考えられる。

4.2 透かし情報生成手法による抽出率と音声品質

透かしの埋め込み次元を 0 次元目とし、3.1 において CBI の N 次元目を変化させる割合 R によって、透かし 0,1 のそれぞれの抽出率と音声品質の客観的評価として SNR を求めた結果を図 4 に示す。図中の "wm:0", "wm:1" はそれぞれ透かし 0,1 の抽出率に対応している。また、SNR について、LSP 分析合成系とベクトル量子化による音声品質の劣化が激しいため、透かし埋め込みによる音声品質の劣化を評価するために、LSP 分析合成系において LSP を CBI でベクトル量子化して合成した音声を基準となる音声とした。図 4 からわかるように、透かし情報として変化させる割合 R が大きくなると、透かし 0,1 とともに抽出率は向上する反面、音声品質は劣化する。また、R が小さくなると、抽出率は低下し、BER が上昇する。R の変化によって音声品質の

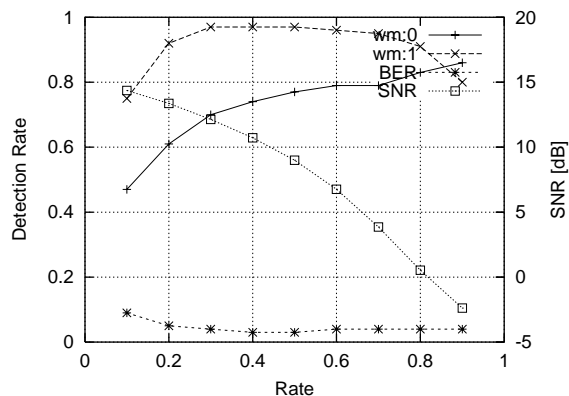


図 4 変化割合による抽出率と SNR (透かし次元 : 0)

劣化が著しく、0 次元目は透かしを埋め込む次元として難しいと考えられる。

高次元の周波数領域は音質にほとんど影響を与えない冗長部分であるといえ、符号化などの処理を加えると透かし情報が消失してしまうことが考えられ、ここでは男性話者について透かしを埋め込む次元を 1~5 次元とした結果を図 5~図 9 に示す。抽出率、音声品質については同様な結果が得られたが、低次元は音声品質に与える影響が大きく、人間の聴覚はスペクトル包絡の谷の部分より極の部分に敏感であることから [6], 透かしを埋め込む次元として谷に相当する 3,4 次元目は、極に相当する 1,2 次元目に比べて音声品質の劣化も少ない。

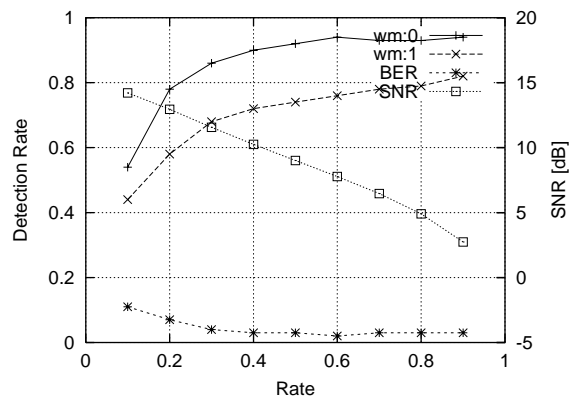


図 5 変化割合による抽出率と SNR (透かし次元 : 1)

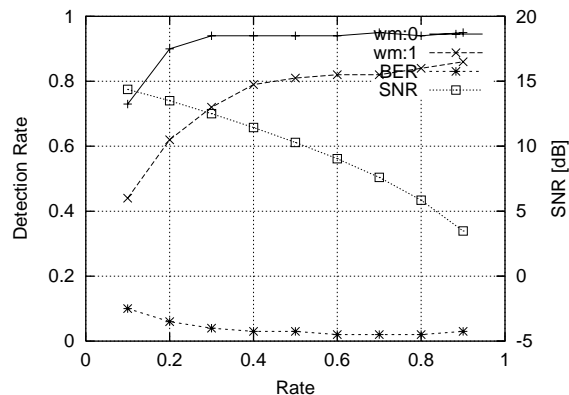


図 6 変化割合による抽出率と SNR (透かし次元 : 2)

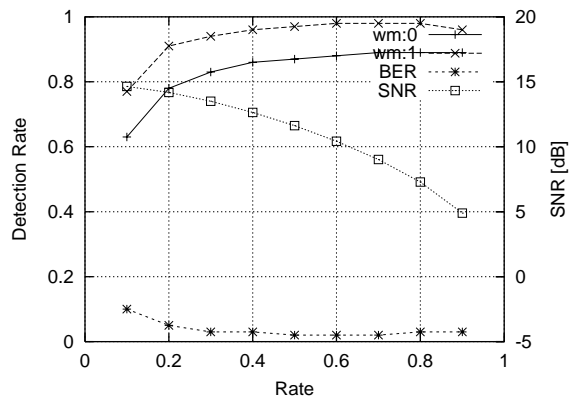


図 7 変化割合による抽出率と SNR (透かし次元 : 3)

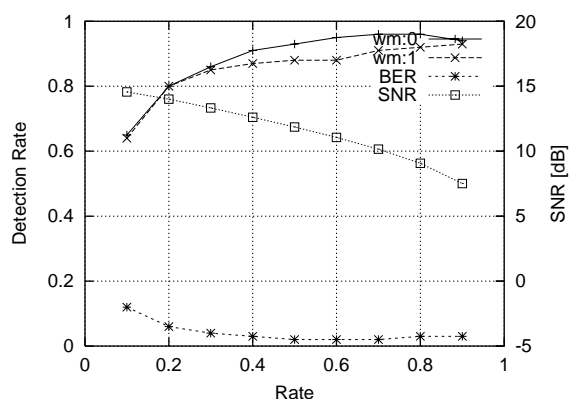


図 8 変化割合による抽出率と SNR (透かし次元 : 4)

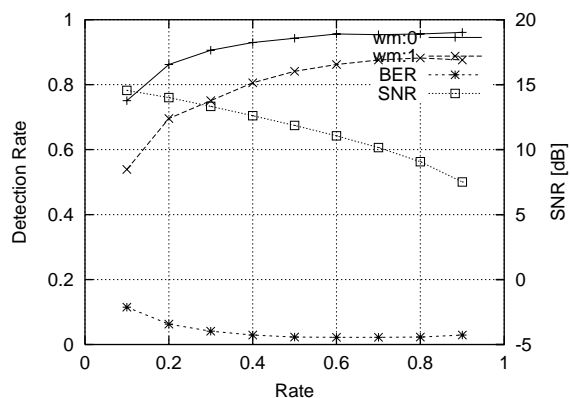


図 9 変化割合による抽出率と SNR (透かし次元 : 5)

5. む す び

本稿では、音声の生成過程に着目することで実現される電子透かし手法の一例について、基本的なアルゴリズムを示し、提案する電子透かし手法において透かし情報の生成と透かしの精度および音声品

質の評価を行った。抽出誤りに対する課題として 4.1 で述べた精度の向上に加え、誤り訂正符号の適用も有効であると考えられる。

現段階では透かし埋め込みプロセス以外でのベクトル量子化および合成段階での音声品質の劣化が激しく、早急な課題であり、この課題を改善した上で、主観評価も含めた音声品質の評価を行う。現在、透かしの埋め込みフレームの選択条件を暫定的に決めているが、埋め込み可能なフレームが増えれば、埋め込める情報量の改善が期待され、今後マスキング等を利用して音声品質を考慮した上で、埋め込み可能な透かしの情報量に関する評価と、電子透かしの安全性と符号化耐性を併せて検討する予定である。また、提案手法を実現する一例として、ITU-T 勧告 G.729(CS-A CELP) への適用について検討する予定である。

参 考 文 献

- [1] 松井甲子雄. “電子透かしの基礎”. 森北出版株式会社, 1998.
- [2] Paraskevi Bassia, Ioannis Pitas and Nikos Nikolaidis. “Robust Audio Watermarking in the Time Domain”. IEEE TRANSACTIONS ON MULTIMEDIA, VOL.3, NO.2, JUNE 2001.
- [3] “電子透かし技術に関する調査報告書”. 社団法人電子情報技術産業協会, 2001.
- [4] 畑田充弘, 小松尚久 他. “発声源に着目した電子透かしの生成手法に関する検討”. 電子情報通信学会総合大会講演論文集 A-7-15, 2001.3.
- [5] 古井貞照. “デジタル音声処理”. 東海大学出版会, 1995.
- [6] 保谷早苗, 板倉文忠. “木探索を用いた LSP パラメータの多段ベクトル量子化”. 信学技報. pp39-46, 1993.