

平成 12 年度

学士学位論文

インターネット電話における連続通話の
保証

An assurance of a serial call in internet telephon

1010462 山本達也

指導教員 福本昌弘

2001 年 2 月 5 日

高知工科大学 情報システム工学科

要旨

インターネット電話における連続通話の保証

山本達也

時代が進むにつれ、情報の記録、伝達が可能になり、約 100 年前に電話が発明され、約 50 年前に放送が始まると、大量の情報が迅速に伝達されるようになった。

近年、デジタル技術が情報の記録、伝達に大きな地位を占め始めている。デジタル技術はコンピュータと結合し、情報を伝達、処理、記録する技術で、我々の周りに浸透している。この中で、情報圧縮技術が重要な役割を果たしている。その好例として、デジタル携帯電話やミニディスクがある。情報圧縮符号化に要求される機能や性能はめまぐるしく変化してきている。その市場規模は拡大していくことは間違いない、情報圧縮への期待はますます増大するだろう。

そして現在、インターネット電話の使用者が増加している。インターネット電話とは、インターネット網を通じて行うリアルタイムの音声通話である。この増加の要因として、インターネットの普及や通話料金の安さなどがあるであろうが、その一方で、音飛びや、話がよく聞こえなくなるトラブルもある。これは、インターネット網の混雑によるものである。デジタル携帯電話やミニディスクのようにもとの音声の、人間には聞きとることのできない部分を取り除くことで、できるだけ忠実に、少ない情報量で表現する。こうすることで、音声の途切れの直接の原因であると思われる、回線の混雑を軽減できると考えた。

キーワード

Abstract

An assurance of a serial call in internet telephon

Tatsuya Yamamoto

As a time progresses, informational record and transfer are attained, and a lot of information came to be transmitted quickly when a telephone was invented about 100 years ago and broadcast started about 50 years ago.

In recent years, the digital technology is begining to occupy the big status to informational record and transfer. The digital technology is combined with a computer, and it has permeated with the technology which transmits, processes and records information our surroundings. In this, the role with important information compression technology is played. There are a digital cellular phone and mini disk as the good example. The function and performance which are required of information compression coding have changed quickly. The market scale of expanding is infallible, the expectation for information compression will increase increasingly.

And now, the user of an Internet telephone is increasing. An internet telephone is a voice telephone call carried out with real time through the internet network. As the factor of this increase, on the other hand it come out the spread of the internet and the low price of the phonecall charges, but they are in trouble with the sound jump and the trouble they can't hear the talk. This is cause of the internet network. I express the original sound as faithful and few information as like a digital cellular phone or mini disk by removing part of the original sound which the human being can't be caught. By carrying out like this, I thought that was mitigable confusion of a circuit it is guessed

that it is the direct cause of an audio way piece.

key words

目次

第 1 章 序論	1
1.1 本研究の背景と目的	1
1.2 本論文の概要	1
第 2 章 インターネット電話	3
2.1 まえがき	3
2.2 インターネット電話	3
2.2.1 一般電話とインターネット電話	3
2.2.2 インターネット電話	4
PC-PC タイプ	4
PC-電話機 タイプ	5
電話機-電話機 タイプ	5
2.2.3 利点	6
2.2.4 問題点	6
第 3 章 音声の符号化	8
3.1 まえがき	8
3.2 音声の符号化	8
3.2.1 周波数領域での符号化	9
帯域分割符号化	9
直交変換符号化	9
適応予測符号化 (APC)	10
適応ビット割り当て適応予測符号化 (APC-AB)	10
ノイズシェイピング	11

目次

重み付きベクトル量子化による変換符号化	11
3.3 離散フーリエ変換	12
3.3.1 フーリエ変換	12
3.3.2 離散フーリエ変換(DFT)	13
3.4 Karhunen-loeve Transform	16
3.4.1 Karhunen-loeve Transform(KLT)	16
3.4.2 問題点	21
3.5 離散コサイン変換	21
3.5.1 利点 1	23
3.5.2 利点 2	24
3.5.3 利点 3	25
第 4 章 評価	27
4.1 評価	27
4.2 結論	27
謝辞	30
参考文献	31
付録 A フーリエ級数	32
付録 B 分散・共分散	34
B.1 分散	34
B.2 共分散	34
付録 C 離散コサイン変換	36

図目次

2.1	PC-PC タイプ	4
2.2	PC-電話機タイプ	5
2.3	電話機-電話機タイプ	6
3.1	連続信号および離散信号の時間領域・周波数領域でのスペクトル	14
3.2	DCT のために用いる信号の対称化	22
3.3	$N=8$ の場合の基底関数	22
3.4	8 次の固有ベクトル	26
4.1	入力信号	28
4.2	DCT した信号	28

第1章

序論

1.1 本研究の背景と目的

そして現在、電話機が家庭に1台ではなく、1人1台は電話を持つようになり、遠く離れた家族や友人等に電話をするときには時間を気にすることがたびたびある。そして最近、普通の電話回線より料金が安いということで、インターネット電話の使用者が増えてきている。インターネット電話の特徴として、先に述べた、通話料金が安いというメリットがある一方で、通信回線の状態によって通話品質が劣化するというデメリットもある。この原因として、インターネット利用者の増大に伴い、回線の混雑による伝送速度の低下、情報の欠落などが挙げられる。インターネットの利用者が増える一方である今日、回線は今以上に混雑することが予想される。そこで、人間の聴覚特性を利用し、人が聞きとることのできない部分を省き、音声ができるだけ忠実に、できるだけ少ない情報量で表現するとともに、帯域に応じて通信品質を変えるという解決策を用いて、音声が途切れることなく通話できる方式を考える。

1.2 本論文の概要

本論文の概要は以下の通りである。

まず第2章においてインターネット電話の問題点を挙げ、本研究の目的であるインターネット電話の連続通話の妨げになっている音声の途切れがなぜ起こるのかを指摘する。

次に第3章で音声の符号化方法を考える。そして、どの領域で符号化するとよいか考え、

1.2 本論文の概要

適当な符号化方法を選択する。また、その符号化方法を考えたときに、どの変換方法が適当かをそれぞれの変換方法の特徴を述べ、最適な変換方法を探し出す。

そして最後に、本研究で得られた結果をもとに、結論を記した。

第 2 章

インターネット電話

2.1 まえがき

まず、インターネット電話がどういったシステムか長所、短所を調べることで理解を深め、音声の途切れがどういった原因で起こっているかを調べる。音声ができるだけ少ない情報量で表現するために、音声信号を周波数領域で符号化することを提案し、さまざまな変換法の中から適当なものを選ぶ。

2.2 インターネット電話

2.2.1 一般電話とインターネット電話

電話の音声はアナログ信号で伝送するが、インターネット電話は、まず音声をデジタル信号に置き換え、これをインターネットプロトコルと呼ばれるインターネットの伝送方式で送る。そのため、電話の通信網ではなく、インターネット網か、ネット電話会社が独自に持つデータ通信網を使う必要がある。

私達が一般の電話を利用するときは、相手先電話番号をダイヤルすると、まず最寄りの電話局の交換機に接続される。その後、いくつかの交換局を経由して、相手先最寄りの交換機から相手電話機に接続され、通話が開始される。インターネット電話の場合も考え方はほとんど同じである。違いは途中のいくつかの交換局の代わりにインターネットに接続するサーバーを使うといったところである。

2.2 インターネット電話

2.2.2 インターネット電話

インターネット電話は、音声をデジタル信号にしてインターネット網を電話回線として使用した通話システムである。インターネット電話は、それを利用する機器によって3つのタイプが存在する。

- PC(パソコン)-PC タイプ^{*}
- PC-電話機タイプ
- 電話機-電話機タイプ

PC-PC タイプ^{*}

現在インターネットと呼ばれるもののほとんどが、このタイプに属している。これは利用者がインターネットに接続していないとできないもので、専用のインターネット電話アプリケーションを立ち上げ、パソコンに接続されたマイクとスピーカーで会話をする。相手を探す方法として、相手のIPアドレスを直接指定する方法とチャットシステムのように決められた専用サーバーにお互いが接続して見つける方法がある。IPアドレスを指定する場合、相手がダイヤルアップ接続の場合は接続の度にIPアドレスが変わってしまうという欠点がある。

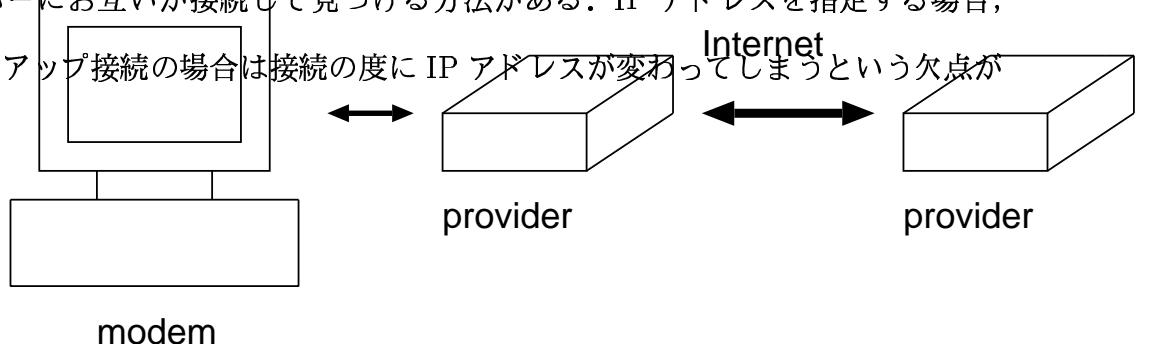


図 2.1 PC-PC タイプ

2.2 インターネット電話

PC-電話機タイプ

このタイプのインターネット電話は、利用者が呼び出したい相手側のコンピュータではなく、家の電話を直接呼び出すことができる。したがって呼び出された側は、インターネットを経由して電話をしていることを意識することなく、通話ができる。呼び出しは相手先に一番近いサービス提供会社の電話網ゲートウェイ^{*1}サーバーが行う。このタイプは、通常の電話をかける動作をインターネット上につながったパソコン上で行うというのが基本的な考え方である。

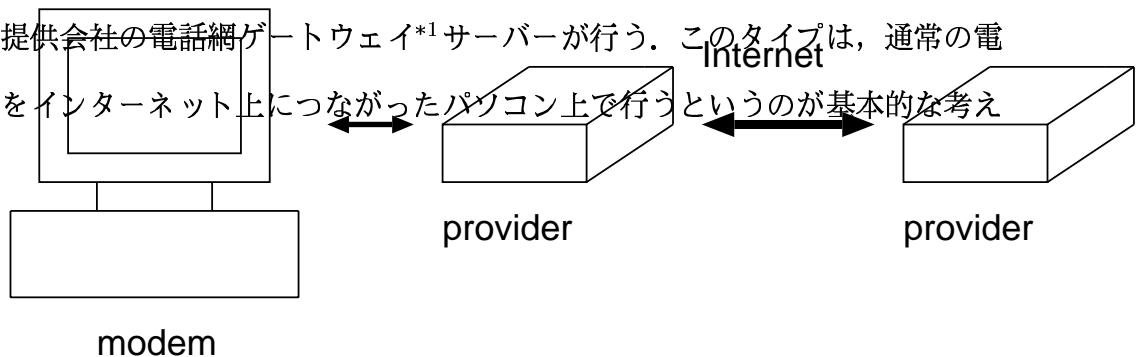


図 2.2 PC-電話機タイプ

電話機-電話機タイプ

このタイプは、インターネット電話に必要とされるコンピュータ、サウンドカード、マイク等を必要としない。まず国内の電話回線を通り、そこからゲートウェイに行き、プロバイダーを通し、相手のゲートウェイ、国内回線を伝い、相手先の電話機へかかるという仕組みである。これは、コンピュータを必要とせずに、家にある互いの電話機だけでやり取りができるということになる。このタイプは、今までインターネットとは無縁であった層のインターネット利用の拡大につながると思われる。

^{*1} 使用しているネットワークを規格の異なったネットワーク等に接続する場合などに使われるハードやソフトの総称。双方のネットワーク間のプロトコルの違いなどを調整して、他のネットワークとの接続を可能にする。ゲートウェイは専用の機器であったり、その役割を持たせたネットワーク上のコンピュータであったりする。

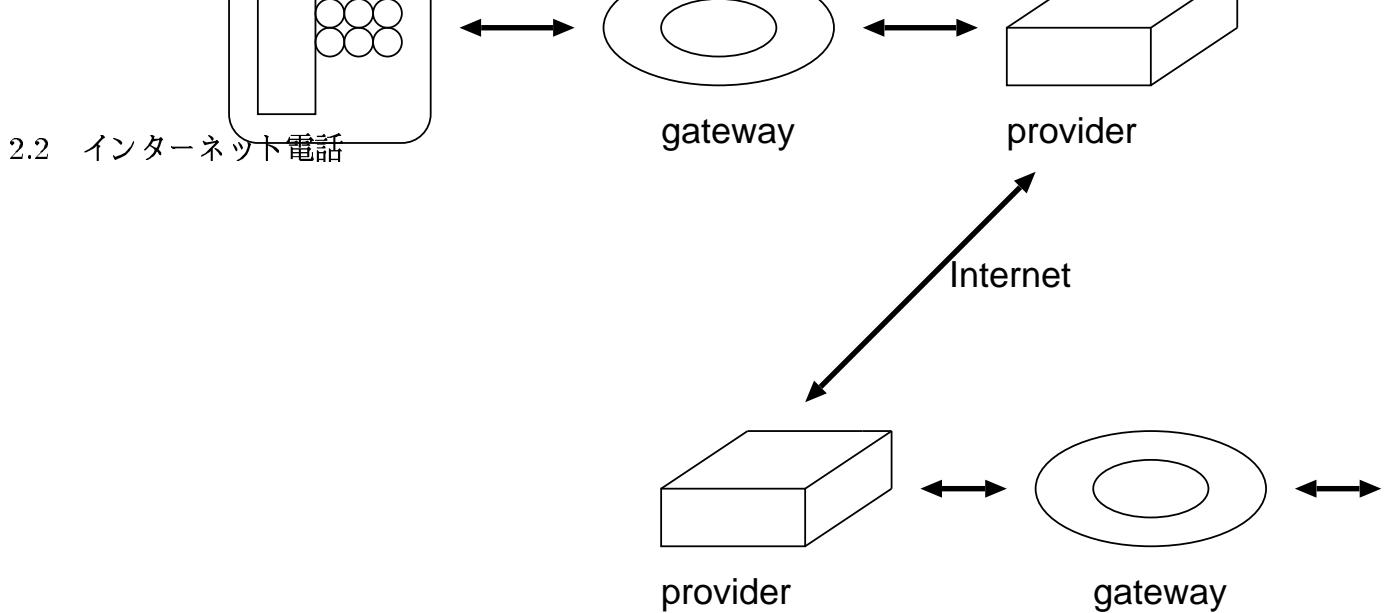


図 2.3 電話機-電話機タイプ

2.2.3 利点

インターネット電話の利点として、通常の電話と比較して、通話料金が安いことが挙げられる。つまり、距離(国、地域)を気にしなくてよいということである。安い料金で海外の人とやり取りができるというのは魅力的である。

2.2.4 問題点

問題点として、

- 手続きが長い
- 通話品質が悪い
- 互換性の問題

がある。

インターネット電話の問題点の一つに、電話をかけるまでの手続きが長いことが挙げられる。特にダイヤルアップ同士でつなぐ場合、PC-PC タイプは相手もパソコンを所有し、インターネットに接続していることが条件になる。電話機-電話機タイプでも、ID やパスワードを入力する手間は、どうしても避けることができない。受話器を上げるとどこでも呼び出しができる通常の電話と比べると、相手と通話できるまでの時間が長くかかってしまう。

2.2 インターネット電話

二つ目に、通話品質が悪いということである。インターネットは利用者が多くなればなるほどルートが塞がれ、データの流れが悪くなる性質がある。音声が途中で途切れたり、相手の声が聞きとれなかったりするのはネットワークの状態が一定でないからである。これは、インターネット電話が回線を『占有』するのではなく『共有』する仕組みから起こる問題である。

三つ目に、互換性の問題がある。PC-PC タイプではそれぞれのインターネット電話ソフトの通信手順や音声符号化形式が違うことがよくある。この場合、同じソフトを持つ者同志でしか会話ができないことになる。これは標準的なプロトコルがないためで、多く乃ソフトウェアは独自の方式を採用している。

第3章

音声の符号化

3.1 まえがき

第2章でインターネット電話での通話中に音声が途切れるのは、回線の混雑による伝送速度の低下、情報の欠落などであることを示した。これを、デジタル携帯電話やミニディスクに活用されている情報圧縮技術を用いて、音声信号をできるだけ少ない情報量で表現し、回線の混雑を軽減する。

3.2 音声の符号化

音声信号にはかなりの冗長性が含まれていて、これを利用すると音声を伝送する際に、それの持つ情報を完全に送受しなくても品質の良い音声を再現できる[19]。

音声のデジタル伝送などに不可欠な技術である音声の符号化は、音声信号から冗長性(なくても済む情報)を取り除くことにより、音声を少ない情報量で表現することである。音声信号には、声道による音声生成の物理的メカニズムと言語構造の性質から、大きな冗長性がある。また、聴覚器官の性質から、人間が聞くことができる音の帯域幅には制限がある。効率良く音声を符号化するためには、音声のもつ性質、人間の聴覚の特性^{*1}をうまく利用することが重要である。人間の聴覚特性などは周波数領域で考えることが多く、聴覚に合わせ

^{*1} 音声を符号化する際に用いる聴覚特性として、マスキング効果と可聴範囲がある。

マスキング効果は、小さな音が大きな音の存在により聞こえなくなる現象である。

可聴範囲は、人間が音として知覚できる音の周波数および強さの範囲のこと、個人差があるが、正常な耳で聞こえる周波数範囲は 20(Hz)～20(kHz)である[2]。

3.2 音声の符号化

た処理は、周波数領域でそのまま容易に実行できる [4]。このことから、周波数領域での符号化を行い、聴覚に大きく影響する成分を優先的に伝送できるようにする。

3.2.1 周波数領域での符号化

帯域分割符号化

帯域分割符号化 (SBC:Sub-Band Coding) は、周波数方向で複数の帯域に分割して符号化する方式である。帯域ごとにダウンサンプルして符号化するので、量子化サンプル数は帯域分割しない場合と同じであるが、分割帯域内ではスペクトルのダイナミックレンジ^{*2}が小さく、予測などの処理が楽になる。また、量子化雑音は帯域外に出ることなく、符号器あるいは復号器で必要な帯域の選択が容易にできる。例えば、情報が少ないときには、全帯域で量子化ひずみを出すのではなく、高域を送らず低域だけの量子化ひずみを小さくするとか、伝送路が混んでいるときには復号器側で高域の情報を捨てるといった処理が可能である。

この方式は、人間の聴覚特性に対応してノイズシェイピング（後述）などを適用することが容易にでき、量子化雑音の聴覚的妨害の程度を小さくして、全体としてビット数が低減できる。量子化雑音が分割された帯域のみに発生し、他の帯域に影響を与えることなくという長所がある [1][2]。

直交変換符号化

聴覚の冗長性に基づき、音声がほぼ定常と考えられるフレーム（ブロック）を、離散コサイン変換 (DCT:Discrete Cosine Transform) 等で周波数領域に直交変換して符号化する方法である。復号化では、逆変換ブロック波形の総和として音声波形を再生する。

直交変換符号化は、信号を周波数領域で分割する点は帯域分割符号化と同じであるが、それよりもはるかに自由度が大きい。

^{*2} 信号の再現能力を表す数値。最小値と最大値の比率であり、dB の単位で表す。アナログデータをデジタル化する場合は量子化ビット数が大きいほど、ダイナミックレンジが広がる。

3.2 音声の符号化

直交変換には、離散フーリエ変換 (DFT:Discrete Fourier Transform), カルーネン・レーベ変換 (KLT:Karhunen-Loeve Transform) なども用いられる。離散コサイン変換 (DCT) を用い、適応的にビット割り当てを行う方式を適応変換符号化 (ATC:Adaptive Transform Coding) と呼ぶ。ATC は、ブロック符号化の代表例の一つで、ブロック単位の波形を周波数領域に変化して量子化する。スペクトル包絡の形状を対数スペクトルの折れ線近似で補助情報として伝送し、フレームごとに各 DCT 成分に対する最適量子化ビット数を計算して、スカラ量子化^{*3}する。音声の大部分は周波数領域でエネルギーが低域に偏っている。この偏りに合わせて適応的にひずみを制御すれば全体の量子化ひずみを削減できる [1][2]。

適応予測符号化 (APC)

適応予測符号化 (APC) は、ブロック符号化法の一つで、時間領域で過去のサンプルからの予測値との誤差を量子化することでひずみを小さくする。量子化器の構成、量子化符号の構成、量子化符号の決定法、量子化雑音の制御法など数多くの変形や改善がある [5]。

適応ビット割り当て適応予測符号化 (APC-AB)

適応ビット割り当て適応予測符号化 (APC-AB:Adaptive Predictive Coding with Adaptive Bit allocation) は、帯域分割符号化 (SBC) と適応予測符号化 (APC) を組み合わせて情報圧縮を行う方式で、音声信号を帯域分割し、各帯域信号をベースバンドに変換したのち、 p 次の線形予測分析を行って、予測符号化する。適応予測符号化 (APC) では、短時間予測と長時間予測 (ピッチ予測) の両方を行う。残差波形に対して、残差パワーをもとに帯域ごとにビット割り当てを行うとともに、1ピッチ^{*4}区間をさらに時間分割して、その区間内の残差パワーに応じて、波形歪みの平均を最小にするように適応ビット割り当てを行う。

^{*3} 1係数に対する値を量子化すること。

ベクトル量子化は、複数のデータに対する値を同時に量子化すること。

^{*4} 音声のピッチとは、声の高低感のことをさす。しかし、この高低感は声帯の振動数である基本周波数と強く関係しているので、有聲音波形の基本周期 (あるいは周波数) のことをピッチと呼ぶこともある。

3.2 音声の符号化

補助情報としては、予測係数、残差電力、分割区間の設定パラメータ（区間の長さ）が伝送される。信号の再生は、符号化と逆の手続きで行われる [1][2].

ノイズシェイピング

マスキング効果を利用して、各周波数の雑音スペクトルがしきい値以下となるように、量子化雑音スペクトルを整形することをノイズシェイピングと呼ぶ。量子化雑音に微分特性を乗じ、雑音分布を変える技術である。低周波領域ほど量子化雑音が小さくなる。これにより、同じ信号対量子化雑音比でも、聴覚的な雑音を減らすことができる [2][5].

重み付きベクトル量子化による変換符号化

低ビット符号化のため、ベクトル量子化を組み合わせたものに重み付きベクトル量子化による変換符号化 (TCWVQ: Transform Coding with Weighted Vector Quantization) がある。この方式は、線形予測残差信号を DCT 領域でベクトル量子化するが、適応ビット割り当てを行わず、DCT 成分をインタリープする形で並び換えてベクトルに分割し、スペクトル包絡の重み付き距離でベクトル量子化を行う方法である。

以上の変換方式の中から、周波数領域を符号化するにあたって比較的自由度の大きい直交変換に着目する。それに使用される離散フーリエ変換、カルーネン・レーベ変換、離散コサイン変換の 3 つの変換方式が、音声を符号化するのにどれが適当であるか比較検討する。

3.3 離散フーリエ変換

3.3.1 フーリエ変換

波形を有限の時間範囲内で定義したときには、その周波数スペクトルは、直流、および $1/T$ (Hz)の周波数とその高調波とからなること、及び、その波形から求められるフーリエ係数で規定される振幅と位相の周波数成分を全部加え合わせると、もとの波形が形成される。では波形を定義する時間の区間長を無限に長くしたどうなるか考える。

フーリエ級数

$$\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} (a_n - j b_n) \exp(j \frac{2\pi n t}{T}) \\ &= \sum_{n=-\infty}^{\infty} c_n \exp(j \frac{2\pi n t}{T}) \end{aligned} \quad (3.1)$$

$$\begin{aligned} c_n &= \frac{1}{T} \int_{-T/2}^{T/2} x(t) \exp(-j \frac{2\pi n t}{T}) dt \\ &= \frac{1}{T} \int_0^T x(t) \exp(-j \frac{2\pi n t}{T}) dt \\ &= a_n - j b_n \end{aligned} \quad (3.2)$$

で $T \rightarrow \infty$ とする。式(3.1),(3.2)の c_n の中の $1/T$ を級数の法に移す。すなわち、 $T c_n$ を X_n と書き直し、

$$X_n = \int_0^N x(t) \exp(-j \frac{2\pi n t}{T}) dt \quad (3.3)$$

$$x(t) = \sum_{n=-\infty}^{\infty} \frac{1}{T} X_n \exp(j \frac{2\pi n t}{T}) \quad (3.4)$$

と書き換えられる。ここで、さらに、

$$\begin{aligned} \frac{2\pi n}{T} &= 2\pi f \\ \frac{1}{T} &= \Delta f \end{aligned}$$

3.3 離散フーリエ変換

において、 $T \rightarrow \infty$ の極限をとると、式(3.3),(3.4)は次のようになる。

$$X(f) = F(x) = \int_0^\infty x(t) \exp(-j2\pi ft) dt \quad (3.5)$$

$$x(t) = F^{-1}(x) = \int_{-\infty}^\infty X(f) \exp(j2\pi ft) df \quad (3.6)$$

この式は、時間 t の関数としての波形 $x(t)$ が周波数 f の関数として定まる $X(f)$ で表される複素振幅の複素指数関数 $\exp(j2\pi ft)$ を積分して作られることを示す。式(3.5)がフーリエ変換で、時間関数からそれに対応する周波数関数を求める式であり、式(3.6)がフーリエ逆変換で、周波数関数から時間関数を求める式である。 $F(x)$ は x のフーリエ変換を、 $F^{-1}(X)$ は X のフーリエ逆変換を表す。

式(3.5),(3.6)は、その積分が収束することを条件として成立するものであり、無限の時間一定の振幅をとり続けたり、振幅が無限に大きくなるような波形は扱うことができない。

フーリエ変換は、時間領域から周波数領域への信号の変換であり、パワースペクトルとして信号の特徴を解析するのに有効である。

3.3.2 離散フーリエ変換 (DFT)

時間領域と周波数領域との間に次のような関係がある。非周期信号はフーリエ変換によって周波数領域で対応付けられ、周波数スペクトルも非周期信号である。時間軸上の周期信号はフーリエ級数によって、周波数領域に対応付けられ、そのときの周波数スペクトルは離散的なものとなる。時間軸上で離散的な非周期信号のフーリエ変換は周波数軸上で周期的な信号となる。そして、時間軸上で離散的な周期信号の周波数軸上での表現は離散的な周期信号になる。この関係を図(3.1)に示す。(c)の関係が離散フーリエ変換である。

まず、周期が T である周期信号 $x(t)$ を考え、この信号をフーリエ級数で表現すると、

$$x(t) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_0 t) \quad (3.7)$$

となる。ただし、 $f_0 = 1/T$ である。したがって、 $x(t)$ を周期 T_s で標本化した信号は、

3.3 離散フーリエ変換

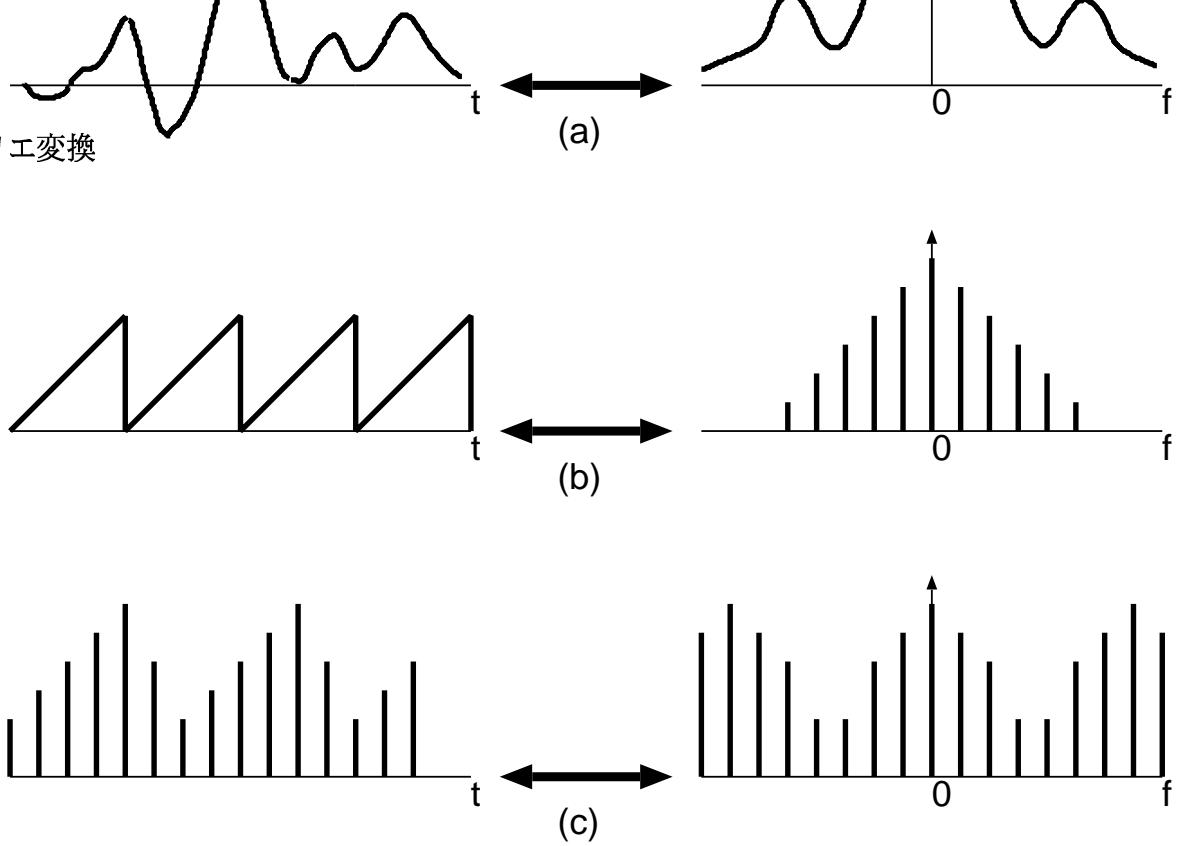


図 3.1 連続信号および離散信号の時間領域・周波数領域でのスペクトル

$$x(kT_s) = \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi n f_0 k T_s) \quad (3.8)$$

となる。ここで、 $x(t)$ の周期 T と標本化周期 T_s との間に、

$$T = N T_s \quad (3.9)$$

の関係があるとすると、 $f_0 T_s = 1/N$ であるから、

$$\begin{aligned} x(kT_s) &= \sum_{n=-\infty}^{\infty} c_n \exp(j2\pi \frac{nk}{N}) \\ &= \sum_{n=-\infty}^{\infty} c_n \left\{ \exp\left(\frac{j2\pi}{N}\right) \right\}^{nk} \\ &= \sum_{n=-\infty}^{\infty} c_n W^{nk} \end{aligned} \quad (3.10)$$

となる。ただし、 $W = \exp(j2\pi/N)$ であり、1 の N 乗根の一つである。ここで nk は整数値をとるので、1 の N 乗根の性質から、 W^{nk} のとり得る値は以下の N 通りに限られる。

3.3 離散フーリエ変換

$$\{W^0, W^1, \dots, W^{N-1}\}$$

したがって、級数の順序が入れ替え可能なものとして同じ値を持つ W^{nk} をまとめると、式 (3.10) は次のような形で表すことができる。

$$x(kT_s) = \sum_{n=0}^{N-1} c'_n W^{nk} \quad (3.11)$$

ここで、簡略化のために $x(kT_s)$ を x_k , c'_n を $(1/N)X_n$ とおくと、以下の関係式が得られる。

$$x_k = \frac{1}{N} \sum_{n=0}^{N-1} X_n W^{nk} \quad (3.12)$$

以上の結果より、式 (3.10) の無限級数が有限の項の和で表されることが示された。

では、逆に、 X_k ではどのような形で表すことができるか考える。式 (3.12) の両辺に W^{-mk} を掛け合わせ、 k について 0 から $N-1$ までの和をとると、

$$\begin{aligned} \sum_{k=0}^{N-1} x_k W^{-mk} &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} X_n W^{nk} W^{-mk} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} X_n \sum_{k=0}^{N-1} (W^{n-m})^k \end{aligned} \quad (3.13)$$

ここで、

$$\sum_{k=0}^{N-1} (W^{n-m})^k = \begin{cases} 0 & (n \neq m) \\ N & (n = m) \end{cases} \quad (3.14)$$

であるから、

$$\sum_{k=0}^{N-1} x_k W^{-mk} = X_m \quad (3.15)$$

したがって、 X_n は次の式によって求められる。

$$X_n = \sum_{k=0}^{N-1} x_k W^{-nk} \quad (3.16)$$

3.4 Karhunen-loeve Transform

$W = \exp(j2\pi/N)$ であるので、これを式 (3.16) に代入すると、

$$X_n = \sum_{k=0}^{N-1} x_k \exp(-j \frac{2\pi n k}{N}) \quad (3.17)$$

明らかに $X_{n+N} = X_n$ であるので X_n も周期系列となっている。 X_n は周波数 W^{nk} に対応したスペクトルであり、周波数軸上で周期関数になる。この関係を離散フーリエ変換 (DFT:Discrete Fourier Transform) と呼ぶ。これに対応して式 (3.12) の関係は逆離散フーリエ変換 (IDFT:Inverse DFT) と呼ばれる。 x_k , X_n はともに周期が N の系列であるから、通常は $0 \leq k, n \leq N - 1$ の範囲で使用される。離散フーリエ変換は有限個数の信号系列 x_k に対して定義されている。この変換では、切り出した信号の外側にも同じ信号が繰り返されている（周期信号）と仮定している。この仮定が実際の処理において、スペクトルの誤差を生む原因になっている。

離散フーリエ変換の利点として、

1. 周波数領域への変換であるため、量子化雑音の周波数領域での制御が容易である。
2. 高速求解アルゴリズムである高速フーリエ変換 (FFT:Fast Fourier Transform) を適応できる。FFT によって次数 N の変換に対して積和演算量が N^2 回から $N \log_2 N$ 回に減少する。音声では、入力の虚数部が 0 であるので、更に乗算回数を半分程度に削減できる [4]。

が挙げられる。

3.4 Karhunen-loeve Transform

3.4.1 Karhunen-loeve Transform(KLT)

帯域分割符号化は音声信号成分の周波数軸上での偏りを利用して量子化雑音を低減する手法であり、周波数軸上での情報を利用してはいるが、符号化する信号は帯域分割された時間領域の波形であった。帯域分割符号化において帯域分割数を非常に大きくすれば、周波数軸

3.4 Karhunen-loeve Transform

上への直接的変換であるフーリエ変換を行って周波数成分を符号化することに等しくなる。このように、例えば周波数領域に変換して音声信号の成分が低域に偏る性質を利用するものが変換符号化である。変換符号化には次の利点がある。

特徴抽出 例えば周波数領域などの、他の座標系に写像した方が信号の特徴を把握するのに便利な場合がある。

情報圧縮 情報量が特定の座標軸に集中するために次元数が削減でき、限られた量子化ビット数の範囲内で量子化雑音の低減を図れる。

周波数軸上への変換にはフーリエ変換が用いられるが、以下では、 N 次元の信号が与えられたときに、これを別の一般的な N 次元の座標系に変換し、歪み最小となる量子化を行って得られる S/N 向上量の理論的な最大値と、その最大値を与える最適直交変換法について考えてみる。

N 次元の入力信号ベクトル $\mathbf{x}_n = (x_{1n}, \dots, x_{Nn})^T$ を N 行 N 列の変換行列 \mathbf{A} (行列成分は a_{ij}) によって出力信号ベクトル $\mathbf{y}_n = (y_{1n}, \dots, y_{Nn})^T$ に変換して符号化する。すなわち、

$$\mathbf{y}_n = \mathbf{A} \mathbf{x}_n \quad (3.18)$$

である。ただし、次元 i の入出力信号 x_{in} , y_{in} はそれぞれ $n = 1, 1, \dots, M$ に対する M 個の時系列信号である。出力信号ベクトル \mathbf{y}_n を量子化する際に、平均 R ビットの量子化ビット数が与えられるとし、これを次元 i に R_i ビットずつ配分して量子化する。すなわち、次式の量子化ビット数を与える。

$$R = \frac{1}{N} \sum_{i=1}^N R_i \quad (3.19)$$

このとき信号 y_{in} の $n = 1, 2, \dots, M$ における分散を σ_{yi}^2 とすると、次元 i で生ずる量子化誤差の分散 σ_{ei}^2 は、

3.4 Karhunen-loeve Transform

$$\sigma_{ei}^2 = C \cdot 2^{-2R_i} \sigma_{yi}^2 \quad (3.20)$$

で与えられる。C は比例定数である。変換された出力信号ベクトル y_n の量子化誤差分散 σ_{ei}^2 の全次元にわたっての総数 E は、式 (3.20) より、

$$E = C \sum_{i=1}^N 2^{-2R_i} \sigma_{yi}^2 \quad (3.21)$$

である。変換信号 y_{in} を最適ビット配分を行い量子化することによって得られる信号対量子化雑音比の利得 G(S/N 向上量) の結果をしめすと、

$$G = \frac{(1/N) \sum_{i=1}^N \sigma_{yi}^2}{(\prod_{i=1}^N \sigma_{yi}^2)^{1/N}} \quad (3.22)$$

となる。変換利得(変換によるひずみ削減量)G は単に座標軸の変換のみでなく、歪み最小となる最適ビット配分も同時に行われた場合の S/N 向上量を表している。

これまでの説明では最適変換が行われた場合の変換利得の上限について述べたが、最大の変換利得を与える具体的な最適変換の方法を明らかにする必要がある。式 (3.22) の結果から、変換利得 G を最大とするためには変換後の信号の分散の大小差が次元間で最大(相乗平均が最小)となればよい。次にこの変換法について考える。

式 (3.18) で定義した変換を次元 i, 時刻 k の出力信号 y_{in} について書き直せば、

$$y_{in} = \sum_{j=1}^N a_{ij} x_{jn} \quad (3.23)$$

のように変換係数 a_{ij} と入力信号 x_{jn} の線形結合として表される。そこで、 y_{in} の n についての分散 σ_{yi}^2 を式 (3.24) で示される変換係数の正規化条件のもとで最大化(最小化)することを考える。この条件は変換前後の信号の総分散値が一致するための条件である。

$$\sum_{j=1}^N a_{ij}^2 = 1 \quad (3.24)$$

変換信号 y_{in} の k に対する分散 σ_{yi}^2 は、式 (3.23) より、

3.4 Karhunen-loeve Transform

$$\begin{aligned}\sigma_{y_i}^2 &= \frac{1}{M} \sum_{n=1}^M \left(\sum_{j=1}^N a_{ij} x_{jn} - \sum_{j=1}^N a_{ij} \bar{x}_j \right)^2 \\ &= \sum_{j=1}^N \sum_{m=1}^N a_{ij} a_{im} c_{jm}\end{aligned}\quad (3.25)$$

となる。ここで、 \bar{x}_j は x_{jn} の k に対する平均値である。また、 c_{jm} は、次元 j と m の入力信号の共分散であり、

$$c_{jm} = \frac{1}{M} \sum_{n=1}^M x_{jn} x_{mk} - \bar{x}_j \bar{x}_m \quad (3.26)$$

で与えられる。 $\sigma_{y_i}^2$ の極値を求めるためにラグランジュの未定乗数法を用いる。この場合の拘束条件は、式 (3.24) の正規化条件であり、未定乗数を λ とすると $\sigma_{y_i}^2$ は式 (3.27) が満たされるときに極値をとる。

$$\sum_{m=1}^N a_{im} c_{jm} - \lambda a_{ij} = 0 \quad (3.27)$$

これは、 N 個の変換係数 a_{ij} ($j = 1, 2, \dots, N$) についての連立一次方程式であり、これが自明な解 $a_{i1} = a_{i2} = \dots = a_{iN} = 0$ を持たないためには、係数の行列式が 0 でなければならぬ。したがって、分散・共分散行列を \mathbf{C} (行列の成分は c_{ij})、単位行列を \mathbf{I} とすると、

$$|\mathbf{C} - \lambda \mathbf{I}| = 0 \quad (3.28)$$

が成り立つので、ラグランジュの未定乗数法における λ は、分散・共分散行列 \mathbf{C} の固有値である。また、式 (3.28) は分散・共分散行列の特性方程式と呼ばれる。この固有値 λ を式 (3.27) の a_{ij} に関する連立一次方程式に代入すると、 N 個の式のうち少なくとも一つは 1 次独立ではなくなり、残りの $N-1$ 個の式を式 (3.24) の正規化条件と共に解けば、正規直交係數 a_{ij} が求まる。この正規直交係數の組を固有値 λ に対する固有ベクトルと呼ぶ。式 (3.28) は λ の N 次方程式であるので、 λ の解は N 個 $(\lambda_1, \dots, \lambda_N)$ 存在し、それに対応して正規直交係數 a_{ij} も N 組 $(i = 1, 2, \dots, N)$ 存在する。次元 i の変換信号 y_{in} の分散 $\sigma_{y_i}^2$ は、式

3.4 Karhunen-loeve Transform

(3.26) と式 (3.27) より,

$$\sigma_{y_i}^2 = \sum_{j=1}^N \sum_{m=1}^N a_{ij} a_{im} c_{jm} = \lambda_i \quad (3.29)$$

となる。これより、 $\sigma_{y_i}^2$ は i 番目の固有値 λ_i に等しく、かつ分散は非負の量であるので $\lambda_i \geq 0$ である。

さらに、固有値は以下の性質を有する。

$$|\mathbf{C}| = \prod_{i=1}^N \lambda_i \quad (3.30)$$

$$\text{tr } \mathbf{C} = \sum_{i=1}^N \sigma_{x_i}^2 = \sum_{i=1}^N \lambda_i \quad (3.31)$$

ただし、 $\text{tr } \mathbf{C} = c_{11} + c_{22} + \dots + c_{NN}$ 。したがって、 N 個の個有値の総数は N 次元の入力信号の分散の総和に等しい。以上述べた手続きによって最適直交変換された出力信号は次元間で互いに無相関であり、入力信号に含まれていた次元間の相関が除去され、符号化利得が最大となる。

N 次元の変換信号を最適ビット割り当て量子化を行うことによって得られる最適直交変換利得(式 (3.22))の結果によれば、変換信号の分散 $\sigma_{y_i}^2$ の相乗平均ができる限り小さくすれば変換利得が最大化される。これは、 $\sigma_{y_i}^2$ 間の大小差が大きい程、最適直交変換利得 G が高いことを表している。しかし、 $\sigma_{y_i}^2$ の最大値(最小値)は、入力信号の分散・共分散行列の最大固有値(最小固有値)に等しいから、変換利得の上限値は入力信号の次元間の相関から一意に定まる。すなわち、最適直交変換利得 G を、分散・共分散行列 \mathbf{C} またはその固有値 λ_i の表現として書き改めれば、

$$G = \frac{(1/N) \sum_{i=1}^N \lambda_i}{(\prod_{i=1}^N \lambda_i)^{1/N}} = (1/N) \frac{\text{tr } \mathbf{C}}{|\mathbf{C}|^{1/N}} \quad (3.32)$$

である。式 (3.32) は直交変換符号化による S/N 向上量の上限値であり、いかなる直交変換もこの値を上回る特性を得ることはできない。

以上述べたように、一定区間内の次元間の統計量(分散・共分散)に基づいて変換信号の大小差を最大とする最適直交変換をカルーネン・レーベ変換(KLT:Karhunen-Loeve

3.5 離散コサイン変換

Transform) と呼ぶ。KLT は統計学においては主成分分析法と呼ばれており、互いに相関のある多種類の特性値の持つ情報を互いに無相関な少数の総合特性値に、情報の損失が最小になるように要約して表現する際の手段として用いられる。

3.4.2 問題点

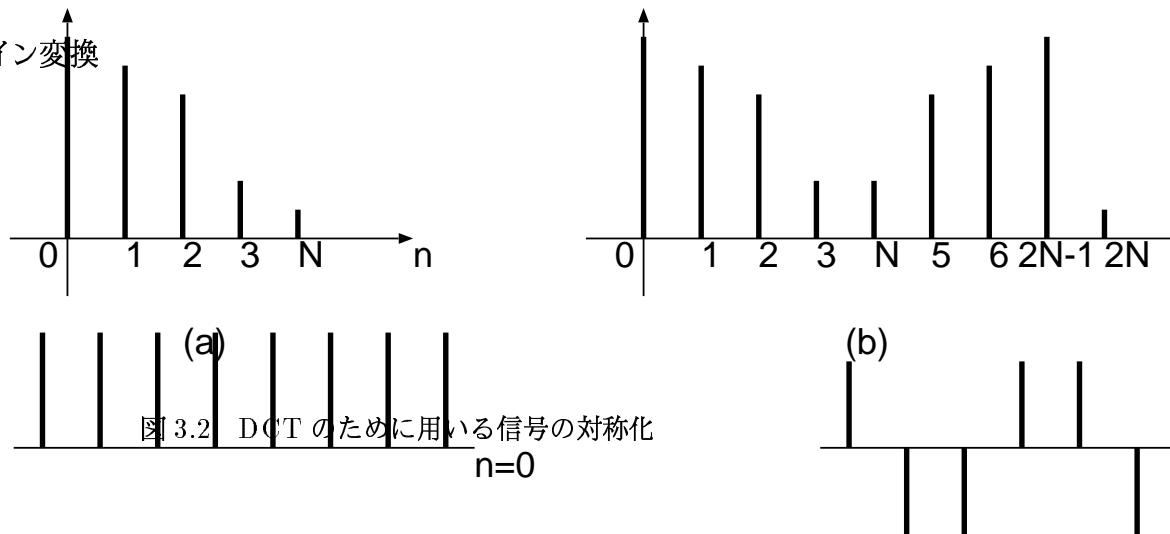
KLT は、実用的には多くの問題がある。まず、変換に付属する演算量が非常に大きい。変換長を N とすると、まず、変換と逆変換のためにそれぞ N^2 回の乗算が必要である。また、音声信号では相関の性質、すなわちスペクトル包絡の形状がフレームごとに変化する。このため、フレームごとに固有値と固有ベクトルを計算する必要がある。また、固有値や固有ベクトルをそのまま伝送すると、伝送情報も莫大となる。固有値の計算は N^3 のオーダの演算量を必要とし、 N が 128 以上となると DFT や DCT の 1,000 倍程度の演算が必要となり、実用的なシステムへの適用は非現実的である [4]。

3.5 離散コサイン変換

離散コサイン変換 (DCT:Discrete Cosine Transform) は信号系列を偶関数化することにより、高周波成分を抑えると考えればよい。長さ N の DFT は、 N の周期で無限に拡張した関数のフーリエ変換に相当する。すなわち、長さ N のデータの DFT を計算する場合、 N の周期で無限に繰り返す関数の周波数を計算していることになる。それに対して、長さ N の DCT は N を鏡の面とみなして対称に折り返したデータの DFT に相当する。したがって、DCT は N の周期で折り返した関数の周波数を計算することになる。図 (3.2(a)) に示す信号をそのまま DFT 分析すると、かなりの高周波成分が検出されてしまうが、図 (3.2(b)) のように元の波形を逆にたどるように延長させると滑らかな波形が得られ、高調波成分の軽減が行われる。つまり DCT は N の周期で折り返した関数の周波数を計算することになる。

DCT とその逆変換は、以下のようになる。

3.5 離散コサイン変換



$$X(k) = \sqrt{\frac{2}{N}} \alpha(k) \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \quad (3.33)$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} \alpha(k) X(k) \cos \frac{(2n+1)k\pi}{2N}$$

ただし, $\alpha(0) = 1/\sqrt{2}$, $\alpha(k) = 1(k \neq 0)$. $N=8$ の場合の基底関数は図 (4.2) のようになる.

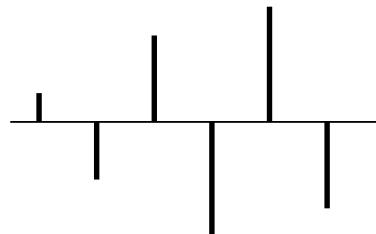
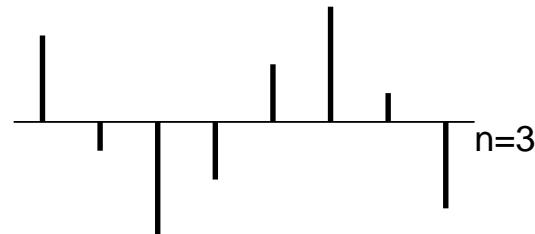
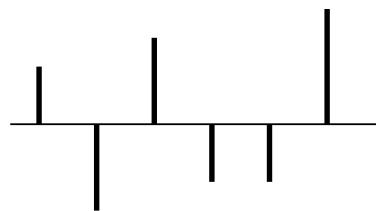
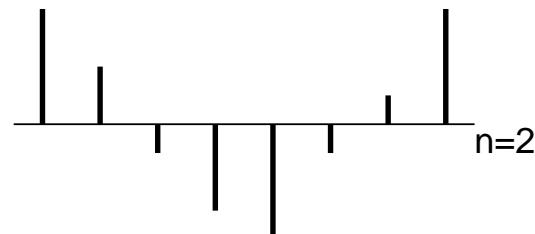


図 3.3 $N=8$ の場合の基底関数

DCT の利点をあげると,

1. DCT と DFT 変換係数間の関係から, DCT 係数は FFT を利用して計算できるため,

3.5 離散コサイン変換

高速変換が可能である。

2. 変換ブロックの接合点における端効果の影響が少ない。
3. 相関の大きい信号に対しては最適直交変換であるカルーネン・レーベ変換 (KLT:Karhunen-Loeve Transform) に極めて近い変換利得が得られる。

が挙げられる [9]。これらの利点を順を追って説明する。

3.5.1 利点 1

DFT 係数を用いて DCT 係数を求めるには、次に示す手続きによる。N サンプルの信号 $x(n)$ から、式 (3.35) に示されるように対称な 2N サンプルの信号 $y(n)$ を作る。

$$y(n) = \begin{cases} x(n) & (n = 0, 1, \dots, N-1) \\ x(2N - 1 - n) & (n = N, N+1, \dots, 2N-1) \end{cases} \quad (3.35)$$

ここで、 $y(n)$ の DFT 係数 $Y(k)$ をもとめると、

$$\begin{aligned} Y(k) &= \frac{1}{\sqrt{2N}} \sum_{n=0}^{2N-1} y(n) \exp(-j \frac{2\pi kn}{2N}) \\ &= \sqrt{\frac{2}{N}} \exp(j \frac{k\pi}{2N}) \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \end{aligned} \quad (3.36)$$

したがって、 $x(n)$ の DCT 係数 $X(k)$ と、 $y(n)$ の DFT 係数 $Y(k)$ との間には、式 (3.33) と式 (3.36) から、

$$X(k) = \alpha(k) \exp(-j \frac{k\pi}{2N}) Y(k) \quad (3.37)$$

の関係が成り立つ。これより、N 個の $x(n)$ から 2N 個の $y(n)$ を作り、 $y(n)$ を FFT することによって DCT 係数 $X(k)$ を求めることができる。

3.5 離散コサイン変換

3.5.2 利点 2

次に、DCT が符号化処理の結果生ずるブロック端の不連続性による劣化が少ないことを示す。入力信号 $x(n)$ の DCT 係数を $X(k)$ とすると、 $X(k)$ を量子化することにより量子化雑音 $E(k)$ が加わる。したがって、復号器で得られる DCT 係数 $X'(k)$ は、

$$X'(k) = X(k) + E(k) \quad (3.38)$$

である。これを DCT 逆変換すると出力信号 $x'(n)$ は、

$$x'(n) = x(n) + e(n) \quad (3.39)$$

となり、単に入力信号 $x(n)$ に量子化雑音 $e(n)$ が加わった形となる。量子化ビット数が十分大きければ式 (3.38) のように入力信号と無相関な量子化雑音を発生させることが可能であるが、伝送速度が非常に低い場合には必ずしもこのような線形性が満たされるとは限らない。すなわち、重要な低域成分のみを残し、高域を欠落させるような非線形な変換が行われる可能性がある。この変換を $H(k)$ で表せば、DCT 係数 $X'(k)$ は、

$$X'(k) = H(k)X(k) + E(k) \quad (3.40)$$

となる。 $H(k)$ は、例えば低域通過特性をもつ変換である。これを復号器で逆変換すれば、

$$x'(n) = h(n) * x(n) + e(n) \quad (3.41)$$

となる。ここで、 $*$ は周期的畳み込みと呼ばれ、

$$h(n) * x(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (3.42)$$

で定義される。したがって出力信号にはブロックの左端と右端で隣接するブロックからの信号の漏れ込みがあり、時間領域での折り返し歪みが加わることになる。これをブロックの端効果と呼ぶ。しかし、先に述べた方法で DCT 係数を計算する場合にはブロックの左端の

3.5 離散コサイン変換

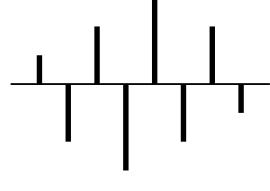
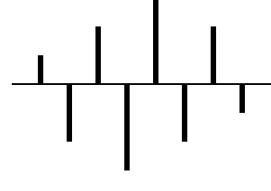
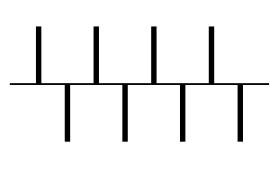
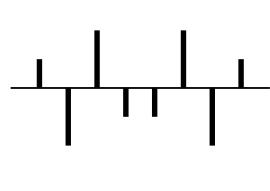
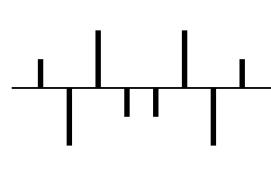
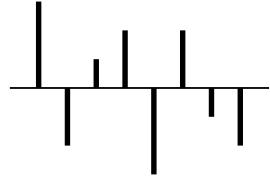
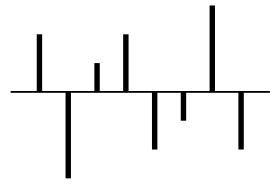
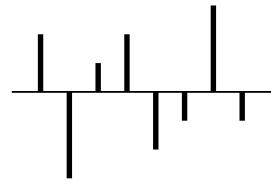
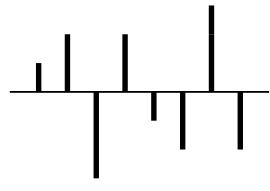
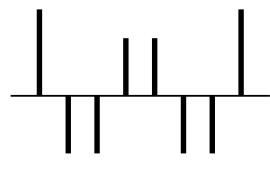
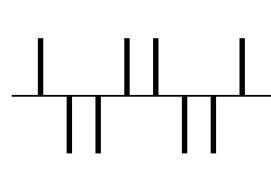
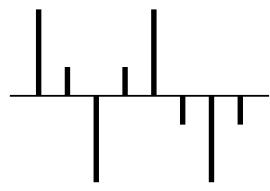
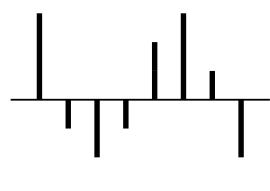
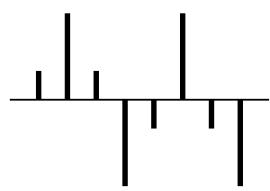
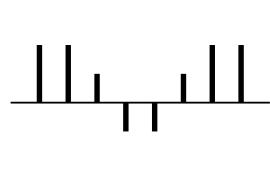
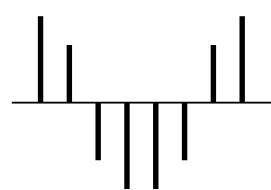
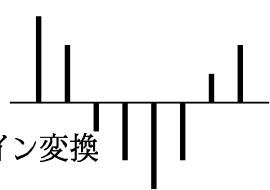
信号と右端の信号とは等しいので、このような時間領域での折り返し歪みの影響が小さい。DCT は変換後のデータの虚部が消えて実数の変換になる。DFT では一般につなぎ目で不連続に拡張されるが、DCT は連続に拡張される。DFT は不連続な点に含まれる高周波のノイズを取り除くために窓関数^{*5}をかけたりするが、DCT は連続になるため、窓関数がなくとも高周波成分はある程度小さくなる。

3.5.3 利点 3

次に、DCT の固有ベクトルを他の直交変換の固有ベクトルと比較してみる。図(3.4)は DFT と DCT、KLT について次元数 $N=8$ に対しての固有ベクトルをしめしたものである。これにより、DCT の固有ベクトルはほぼ KLT の固有ベクトルに等しいことが分かる。KLT は最適直交変換であるが、信号の統計的性質に依存することから、符号化への適用は現実的ではない。そうなると、KLT の固有ベクトルに近い DCT が最も適当であるということになってくる。

^{*5} 時間窓による誤差を軽減するための処理

3.5 離散コサイン変換



(a)DFT

(b)DCT

(c)KLT

図 3.4 8 次の固有ベクトル

第 4 章

評価

4.1 評価

音声の周波数領域を符号化するにあたって用いられる DFT, KLT, DCT の特徴を調べることで, DCT が最も適当であることが分かった。最適直交変換である KLT は DCT に極めて近い特性を示す [22] が, 一般に高速アルゴリズムが存在しないため, 理論解析などに用いられるだけで, 応用例が限定されてしまう。DFT は高速アルゴリズムである高速フーリエ変換 (FFT) を適応でき, 演算量が減少するなどの利点もあるが, 切り出した信号の外側にも同じ信号が繰り返されていると仮定して演算するので, 実際に処理するときにスペクトルの誤差を生んでしまうという欠点もある。この 3 つを比較して, DCT が最適であると判断したが, 低ビットレートの符号化においてブロック歪みが目立つという欠点がある。

図 (4.1) はある入力信号を表していて, 図 (4.2) はそれを DCT した信号である。

4.2 結論

本研究は, まず, インターネット電話の音声の途切れの原因を究明し, それが回線の混雑による伝送速度の低下, 情報の欠落であることを示した。そして, その解決策として, 人間の聴覚特性を利用することで, 人の聞きとることのできない部分を省き, できるだけ少ない情報量で表現することを提案した。これは, ディジタル携帯電話やミニディスクなどにも活用されていて, 音声信号に含まれる冗長性を取り除く, すなわち音声を符号化することで音声を表現することである。また, 人間の聴覚特性などは周波数領域で考えることが多く, 聽

4.2 結論

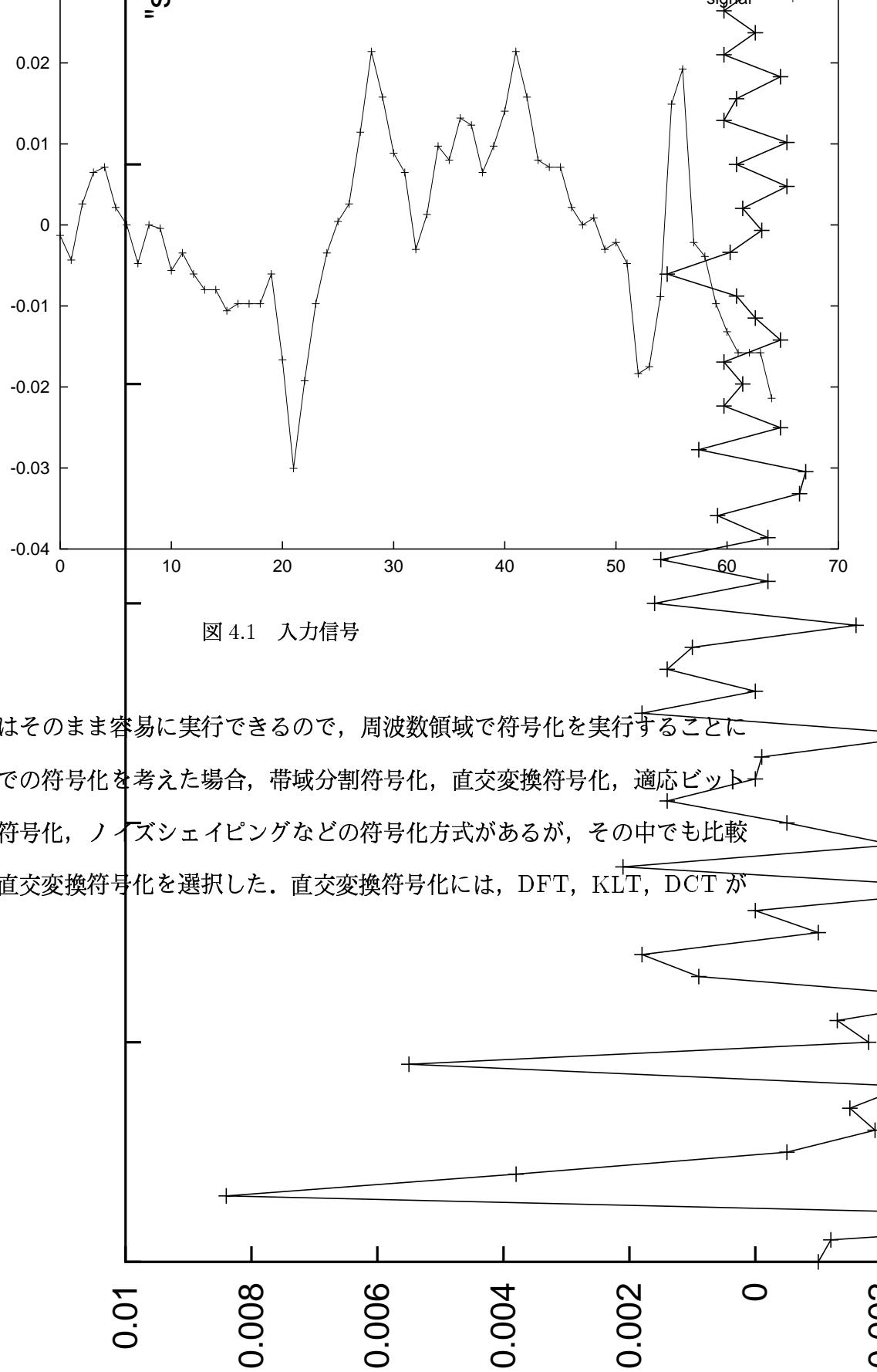


図 4.1 入力信号

覚に合わせた処理はそのまま容易に実行できるので、周波数領域で符号化を実行することにした。周波数領域での符号化を考えた場合、帯域分割符号化、直交変換符号化、適応ビット割り当て適応予測符号化、ノイズシェイピングなどの符号化方式があるが、その中でも比較的自由度の大きい直交変換符号化を選択した。直交変換符号化には、DFT、KLT、DCT が

図 4.2 DCT した信号

4.2 結論

用いられるが、この3つを比較することで、DCTが適当であると判断した。DCTで音声信号を符号化することで、回線の混雑は軽減されるはずである。

謝辞

本研究を行うにあたり、御指導、ご助言を頂いた福本昌弘講師に深く感謝いたします。

また、本研究の審査をして頂いた島村和典教授、菊地豊助教授にも深く感謝いたします。

最後に、日頃から多面にわたって御協力を頂いた福本研究室の諸氏に感謝いたします。

この場を借りて、皆様に感謝の意を表します。

参考文献

- [1] 古井貞熙, 音声情報処理, 森北 (1998)
- [2] 古井貞熙, 音響・音声工学, 近代科学社 (1992)
- [3] ディジタル信号処理, 電子情報通信学会 (1988)
- [4] 守谷健弘, 音声符号化, 電子情報通信学会 (1998)
- [5] 電子情報通信ハンドブック, 電子情報通信学会 (1998)
- [6] 羽鳥光俊, ディジタル信号処理, 丸善 (1994)
- [7] 樋口龍雄, ディジタル信号処理の基礎, 昭晃堂 (1991)
- [8] 笠原正雄, 田中初一, ディジタル通信工学, 昭晃堂 (1992)
- [9] 鎌田一雄, 辻井重男, ディジタル信号処理, 昭晃堂 (1994)
- [10] 貴家仁志, ディジタル信号処理, 昭晃堂 (1997)
- [11] 佐川雅彦, 貴家仁志, 高速フーリエ変換とその応用, 昭晃堂 (1992)
- [12] 山下まこと, やさしいディジタル伝送, 電気通信協会 (1984)
- [13] 日井支朗, 信号解析, オーム社 (1998)
- [14] 前田渡, ディジタル信号処理の基礎, オーム社 (1980)
- [15] 浜田望, よくわかる信号処理, オーム社 (1995)
- [16] Hwei P. Hsu, 村崎憲雄, 間多均, 鮑本一裕, マグロウヒル大学演習 信号処理 (II), オーム社 (1998)
- [17] 谷萩隆嗣, マルチメディアとディジタル信号処理, コロナ社 (1997)
- [18] 森下巖, 小畠秀文, 信号処理, コロナ社 (1998)
- [19] 江原義郎, ユーザーズディジタル信号処理, 東京電機大学 (1991)
- [20] Y.Thomas, 中村尚五, プラクティスデジタル信号処理, 東京電機大学 (1995)
- [21] 今井聖, 音声認識, 共立 (1995)
- [22] 藤原洋, 安田浩, デジタル放送・インターネットのための情報圧縮技術, 共立 (1998)

付録 A

フーリエ級数

信号を表現する一つの代表的方法にフーリエ級数展開がある。これは信号を時間パラメータ t の周期関数と仮定し、三角関数の和で表現する方法である。また、ここでは信号は実数値をとるものとする。

周期 T をもつ信号を $x(t)$ と表す。すなわち、すべての t で

$$x(t) = x(t + T)$$

が成り立つとする。このとき、 $x(t)$ は適当な係数 $\{a_n\}$, $\{b_n\}$ を用いて、

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^N \{a_n \cos(\frac{2\pi nt}{T}) + b_n \sin(\frac{2\pi nt}{T})\} \quad (\text{A.1})$$

と表される。ここで、式 (A.1) をフーリエ級数展開、 $\{a_n\}$, $\{b_n\}$ を $x(t)$ のフーリエ係数と呼ぶ。 T のもとを解くに基本周期と呼ぶ。

波形がどのような値の $\{a_n\}$, $\{b_n\}$ で表されるかをしるには、

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(\frac{2\pi nt}{T}) dt \quad (\text{A.2})$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(\frac{2\pi nt}{T}) dt \quad (\text{A.3})$$

の関係を利用する。式 (A.1) ~ (A.3) は任意の自然数 N は ∞ に置き換えてよい。

信号が実数値をとる場合、正弦波を複素指数関数で表現すると、便宜上、 $n < 0$ に対しても式 (A.2), (A.3) をそのまま拡張して表現することができ、

$$a_n = a_{-n}, b_n = -b_{-n} \quad (\text{A.4})$$

と表すことができる。このとき、式(A.1)を変形して、

$$\begin{aligned}
 x(t) &= \frac{1}{2} \left\{ a_0 + \sum_{n=-\infty}^{-1} (a_n - jb_n) \exp(j \frac{2\pi nt}{T}) + \sum_{n=1}^{\infty} (a_n - jb_n) \exp(j \frac{2\pi nt}{T}) \right\} \\
 &= \sum_{m=-\infty}^{\infty} c_n \exp(j \frac{2\pi nt}{T})
 \end{aligned} \tag{A.5}$$

と書ける。ただし、

$$\begin{aligned}
 c_n &= \frac{a_n - jb_n}{2} \\
 &= \frac{1}{T} \int_{-T/2}^{T/2} x(t) \exp(-j \frac{2\pi nt}{T}) dt
 \end{aligned} \tag{A.6}$$

である。式(A.5)を複素フーリエ級数展開、式(A.6)の c_n を複素フーリエ係数という。

付録 B

分散・共分散

B.1 分散

確率変数 X の関数 $g(X - \mu)^2$ の平均

$$\begin{aligned}\sigma^2 &= E[g(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x) \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx\end{aligned}\tag{B.1}$$

を分散という。ただし、 μ は X の平均である。また、 σ を標準偏差という。一般にチエビシェフの不等式

$$P\left(\frac{|X - \mu|}{\sigma} \geq c\right) \leq \frac{1}{c^2} \quad (c > 0)\tag{B.2}$$

が成立する。つまり、確率分布が平均 $\mu \pm c\sigma$ を超えて分布する確率は c とともに小さくなる。したがって、分散は分布の平均値回りの集中度を表す。

B.2 共分散

二つの確率変数 X と Y の一次的関係を示す量として共分散

$$Cov[X, Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy\tag{B.3}$$

がある。ただし、 μ_x, μ_y はそれぞれ X と Y の平均であり、 $f(x, y)$ は結合確率密度である。

B.2 共分散

$$\rho = \frac{Cov[X, Y]}{\sigma_x \sigma_y} \quad (\text{B.4})$$

を相関係数という。ただし、 σ_x , σ_y は X と Y のそれぞれの標準偏差、シュワルツの不等式から $|\rho| \leq 1$ が導かれる。 $\rho = 0$ のとき X と Y は無相関であるという。互いに独立な X と Y は無相関であるが、逆は必ずしも成立しない。また、 $Y = \alpha X + \beta$ なる線形関係が成立するとき、 α の正負に応じて $\rho = \pm 1$ となる。

付録 C

離散コサイン変換

```
functions used:  
    dct1  
    idct1  
  
#include <stdio.h>  
#include <math.h>  
#include "dct1.h"  
  
  
void dct1(),idct1();  
  
void main(){  
  
    double s0[L];  
    double s0r[L];  
    double xk[L];  
  
    int i;  
    int s_len = L;  
  
    FILE *fp;  
    fp=fopen("sp2","r");  
    for (i = 0; i < L; i++){  
        fscanf(fp,"%lf",&s0[i]);  
    }  
    fclose(fp);  
  
    printf("入力信号 \n");  
    for(i = 0; i < L; i++){  
        printf("%7.4f ",s0[i]);  
    }  
}
```

```

printf("\n");

dct1(s0,L,xk);

printf("DCT した信号\n");
for(i = 0; i < L; i++){
    printf("%7.4f ",xk[i]);
}
printf("\n");

idct1(s0r,L,xk);

printf("IDCT した信号\n");
for(i = 0; i < L; i++){
    printf("%7.4f ",s0r[i]);
}
printf("\n");

}

void dct1 (
    double s0[],
    unsigned int s_len,
    double xk[] )
{
    unsigned int n, k;
    double ck,x;

    for (k = 0; k < s_len; k++)
    {
        if (k == 0)
        ck = 1/sqrt(2);
        else
        ck = 1;

        for (n = 0,x = 0; n < s_len; n++)
        {
x += cos((2*n+1)*k*PI/(2*s_len))*s0[n] ;
xk[k] = 2*ck*x/s_len;
        }
    }
}

void idct1 (
    double s0r[],
```

```

    unsigned int s_len,
    double xk[])
{
    unsigned int n, k;
    double ck;

    for (n = 0; n < s_len; n++)
        s0r[n] = 0;

    for (k = 0; k < s_len; k++)
    {
        if (k == 0)
            ck = 1/sqrt(2);
        else
            ck = 1;

        for (n = 0; n < s_len; n++)
        {

            s0r[n] += cos((2*n+1)*k*PI/(2*s_len))*ck*xk[k];
        }
    }
}

```