

波形符号化技術

内容

[概要](#)

[前提条件](#)

[要件](#)

[使用するコンポーネント](#)

[表記法](#)

[パルス符号変調](#)

[フィルタリング](#)

[サンプリング](#)

[音声のデジタル化](#)

[量子化とコーディング](#)

[コンパANDING](#)

[A-law および u-law コンパANDING](#)

[差分パルス符号変調](#)

[適応型 DPCM](#)

[32 KB/秒の特定手順](#)

[関連情報](#)

概要

人間はアナログ通信に向いていますが、アナログ伝送は効率が良くありません。伝送損失が原因でアナログ信号が弱くなると、ランダム伝送ノイズの構造から複雑なアナログ構造を分離することが困難になります。アナログ信号を増幅すると、同時にノイズも増幅されるため、アナログ接続でのノイズは最終的に、使用に耐えないほど増大してしまいます。「1ビット」状態と「0ビット」状態しかないデジタル信号は、より簡単にノイズを分離できます。デジタル信号は破損なしで増幅できます。デジタルコーディングは、長距離接続でのノイズ破損の影響をあまり受けません。また、世界の通信システムは、パルス符号変調 (PCM) と呼ばれるデジタル伝送形式に変換されています。PCM は、元の音声波形を符号化することから「波形」符号化と呼ばれている符号化の一種です。このドキュメントでは、アナログ音声信号をデジタル信号に変換するプロセスについて詳細に説明します。

前提条件

要件

このドキュメントに特有の要件はありません。

[使用するコンポーネント](#)

このドキュメントの内容は、特定のソフトウェアやハードウェアのバージョンに限定されるものではありません。

表記法

ドキュメント表記の詳細は、「[シスコテクニカルティップスの表記法](#)」を参照してください。

パルス符号変調

PCM は、ITU-T G.711 仕様で定義された波形符号化方式です。

フィルタリング

アナログ信号をデジタル信号に変換する最初の手順は、信号内の高周波数成分を除去することです。こうすることで、下流側ではこの信号を容易に変換できるようになります。会話での音声出力エネルギーは通常、200 または 300 ヘルツから、約 2700 または 2800 ヘルツまでの範囲となります。標準的な会話、および標準的な音声通信に対しては、およそ 3000 ヘルツの帯域幅が確立されています。したがって、精密なフィルタ（非常に高価）を用意する必要はありません。機器の仕様の観点からは、4000 ヘルツの帯域幅が設けられています。このように帯域幅を制限するフィルタは、エイリアシングを防止するために使用されます（アンチエイリアシング）。エイリアシングは、入力したアナログ音声信号が、ナイキスト基準 $F_s < 2(BW)$ によって定義された周波数以下の周波数でサンプリングされる場合に生じる現象です。これは、サンプリング周波数が、入力アナログ信号の最大周波数より低い状態です。この結果、サンプルの周波数スペクトルと入力アナログ信号の間で折り返しが生じます。元の入力信号を再構築するために使用されるローパス出力フィルタは、このような折り返し信号を検出できるほど高性能ではありません。したがって、元の信号には由来しない、新たな信号が生成されることとなります。このように、サンプリング時に偽信号が生成されることを、エイリアシングと呼びます。

サンプリング

アナログ音声信号をデジタル音声信号に変換するための 2 番目の手順は、フィルタ処理された入力信号を、一定のサンプリング周波数でサンプリングすることです。これには、パルス振幅変調（PAM）と呼ばれるプロセスを使用します。このステップでは、元のアナログ信号を使用して、振幅と周波数が一定のパルス列の振幅を変調します。（図 2 を参照）。

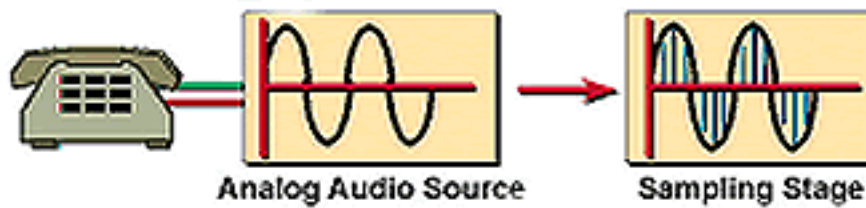
パルス列は一定の周波数で推移しますが、これをサンプリング周波数と呼びます。アナログ音声信号を 1 秒間に 100 万回サンプリングすることも、1 秒間に 2、3 回サンプリングすることも可能です。サンプリング周波数をどのように決定すればいいのでしょうか。ハリー・ナイキストという科学者は、十分な量のサンプルが抽出されていれば、元の音声信号を再構築できることを発見しました。ナイキストは、サンプリング周波数が、入力した元のアナログ音声信号の最高周波数の 2 倍以上であれば、受信先でローパスフィルタによってこの信号を再構築できると定義しています。ナイキスト基準は、次のように表されます。

$$F_s > 2(BW)$$

F_s = Sampling frequency

BW = Bandwidth of original analog voice signal

図 1：アナログ サンプリング



音声のデジタル化

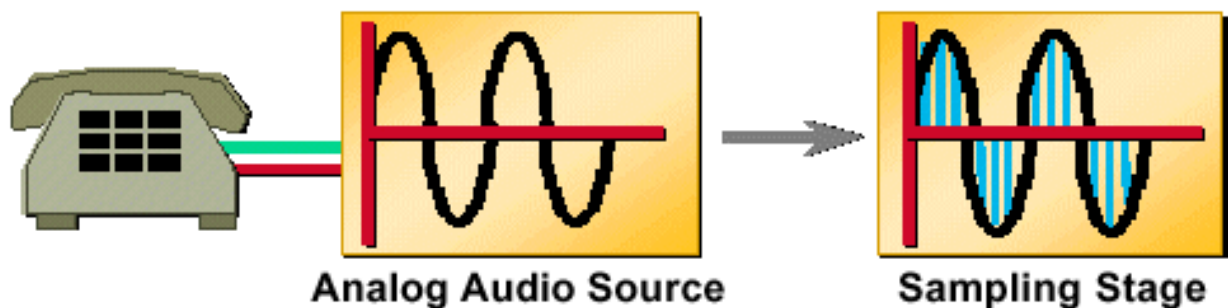
入力アナログ音声信号をフィルタ処理およびサンプリング (PAM を使用) した後は、これらのサンプルを、テレフォニー ネットワーク上で伝送できるようにデジタル化する必要があります。アナログ音声信号をデジタル化するプロセスを、PCM と呼びます。PAM と PCM の唯一の違いは、PCM はさらに一歩進んだプロセスとなっていることです。PCM では、2 進数のコードワードを使用して、個々のアナログ サンプルをデコードします。PCM には、送信元側にアナログからデジタルへのコンバータがあり、宛先側にデジタルからアナログへのコンバータがあります。PCM では量子化と呼ばれる手法を使用して、これらのサンプルをエンコードします。

量子化とコーディング

図 2 : パルス符号変調 : ナイキストの定理

Pulse Code Modulation—Nyquist Theorem

Voice Bandwidth =
200 Hz to 3400 Hz



Codec Technique

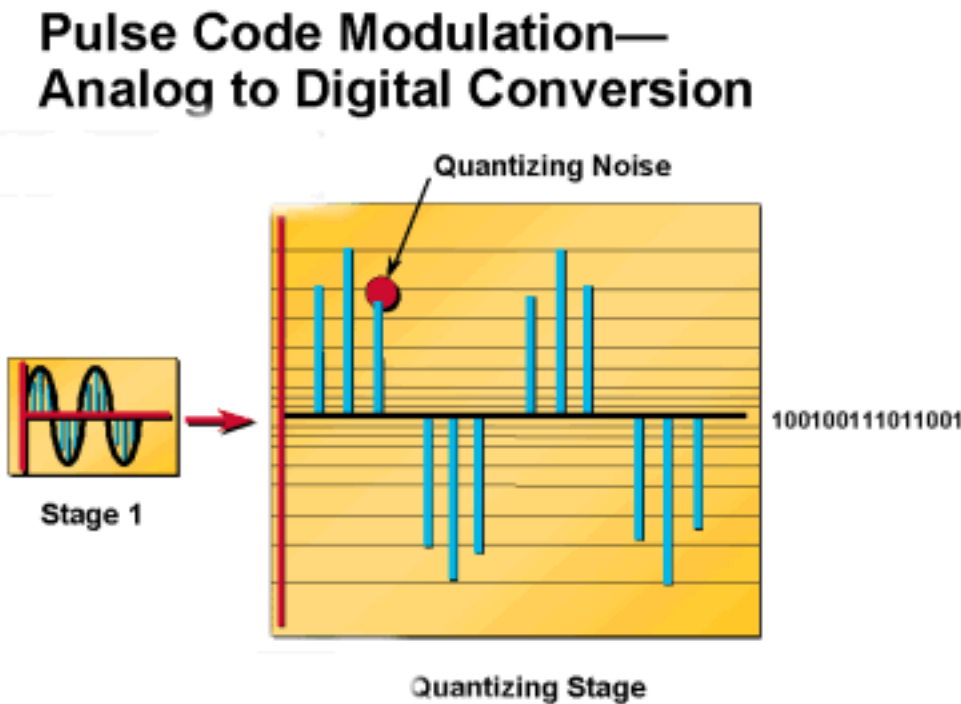
量子化とは、アナログ サンプルの個々の値を、一意のデジタル コードワードを割り当て可能な

離散値に変換するプロセスです。

量子化フェーズに入った入力信号サンプルは、量子化間隔に割り当てられます。すべての量子化間隔は、入力アナログ信号のダイナミックレンジ全体に均等に配置されます（均一量子化）。個々の量子化間隔には、バイナリコードワードという形式の離散値が割り当てられます。標準的に使用されるワードのサイズは8ビットです。入力アナログ信号が1秒間に8,000回サンプリングされた場合、個々のサンプルに8ビット長のコードワードを割り当てるとすると、PCMを使用するテレフォニーシステムでの最大伝送ビットレートは、1秒あたり64,000ビットとなります。図2は、PCMシステムのビットレートがどのように算出されるかを示します。

個々の入力サンプルには、その振幅の高さに最も近い量子化間隔が割り当てられます。実際の高さに適合する量子化間隔が割り当てられていない入力サンプルが存在すると、PCMプロセスでエラーが発生します。このエラーを、量子化ノイズと呼びます。量子化ノイズは、音声信号の信号対雑音比（SNR）に影響を及ぼすランダムノイズと同義です。SNRは、背景雑音と比較した信号強度の尺度です。通常はデシベル単位（dB）で測定されます。入力信号の強度（マイクロボルト単位）を V_s とし、ノイズレベル（マイクロボルト単位）を V_n とすると、信号対雑音比 S/N （デシベル単位）は $S/N = 20 \log_{10}(V_s/V_n)$ という数式で表すことができます。SNRはデシベル単位（dB）で測定されます。SNRが高いほど、音声品質も高くなります。量子化ノイズが生じることで、信号のSNRが低下します。したがって、量子化ノイズが増加すると、音声信号の品質が低下することになります。図3は、量子化ノイズが生じるしくみを示します。コーディングでは、 N ビットワードに対して $2N$ の量子化ラベルが生成されます。

図3：アナログ/デジタル変換



量子化ノイズを削減する方法の1つは、量子化間隔を狭めることです。量子化間隔を狭めることで、入力信号の振幅の高さと量子化間隔との差が縮小します（間隔を狭めれば量子化ノイズが減少する）。ただし、量子化間隔を狭めればそれだけ、必要となるコードワードの量も増えます。その結果、より多くのコードワードを処理するためにPCMシステムのキャパシティを考慮しなければならないという新たな問題が発生します。

SNR（量子化ノイズを含む）は、均一量子化における音声品質に影響を与える最も重要な要因で

す。均一量子化では、入力アナログ信号のダイナミックレンジ全体において、等しい量子化レベルが使用されます。したがって、低信号では SNR が小さくなり (低信号レベルの音声品質)、高信号では SNR が大きくなります (高信号レベルの音声品質)。生成されるほとんどの音声信号は低信号に分類されるため、音声信号のデジタル化においては、高い信号レベルの音声品質を向上させることはきわめて非効率的です。低信号の音声品質を高めるため、均一量子化 (均一 PCM) は不均一量子化プロセスに置換されますが、このプロセスをコンパANDINGと呼びます。

コンパANDING

コンパANDINGとは、まず送信元でアナログ信号を圧縮し、宛先に到達した時点で、この信号を元のサイズに伸張するプロセスです。コンパANDINGという用語は、「compressing (圧縮)」、「expanding (伸張)」の2つの単語を1つに組み合わせて作成された造語です。コンパANDINGプロセスでは、入力アナログ信号のサンプルは対数セグメントに圧縮されます。これらの各セグメントは、均一量子化によって量子化および符号化されます。圧縮プロセスは対数的です。サンプル信号が増えると、圧縮の量も増えます。つまり、少量のサンプル信号より、多量であればあるほどサンプル信号は多く圧縮されます。この結果、サンプル信号が増えるほど、量子化ノイズも増加します。入力サンプル信号のダイナミックレンジ全体にわたって量子化ノイズが対数的に増加することで、このダイナミックレンジ全体の SNR が一定に保たれます。コンパANDINGの ITU-T 標準規格は、A-law および u-law と呼ばれます。

A-law および u-law コンパANDING

A-law および u-law は、Consultative Committee for International Telephony And Telegraphy (CCITT; 国際電信電話諮問委員会) の G.711 で定義されている音声圧縮方式 (コーデック) であり、16 ビットの線形 PCM データを 8 ビットの対数データに圧縮します。

A-law コンパAND

A-law 圧縮では、線形サンプル値を 12 ビットの大きさに制限します。これは、次の式で定義されます。A は圧縮パラメータ (ヨーロッパでは $A = 87.7$)、x は圧縮対象となる正規化された整数です。

$$F(x) = \begin{cases} \frac{A * |x|}{1 + \ln(A)} & 0 \leq |x| < \frac{1}{A} \\ \frac{\text{sgn}(x) * (1 + \ln(A|x|))}{1 + \ln(A)} & \frac{1}{A} \leq |x| \leq 1 \end{cases}$$

u-law コンパAND

u-law (このドキュメントでは u-law と Mu-law を同じ意味で使用しています) 圧縮では、線形サンプル値を 13 ビットの大きさに制限します。これは、次の式で定義されます。m は圧縮パラメータ (米国および日本では $m = 255$)、x は圧縮対象となる正規化された整数です。

$$F(x) = \frac{\text{sgn}(x) * \ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad 0 \leq |x| \leq 1$$

A-law 標準は、主にヨーロッパおよびその他の地域で使用されています。u-law は北米および日本で使用されています。

A-law と u-law の類似点

- いずれも、対数的な入力と出力の関係を線形近似します。
- いずれも 8 ビットのコードワードで実装されます (量子化間隔ごとに 1 レベル、合計 256 レベル)。8 ビットのコードワードで、64 キロビット/秒 (kbps) のビットレートに対応できます。この値は、サンプリングレート (入力周波数の 2 倍) にコードワードのサイズを乗算することで得られます ($2 \times 4 \text{ kHz} \times 8 \text{ ビット} = 64 \text{ Kbps}$)。
- いずれも、ダイナミックレンジを 16 のセグメントに分割します。正セグメント 8 個、負セグメント 8 個各セグメントの長さは直前のセグメントの 2 倍各セグメント内では均一量子化を使用
- いずれも、類似した手法で 8 ビットワードを符号化します。ビット 1 (MSB) が極性を識別ビット 2、3、4 がセグメントを識別最後の 4 ビットでセグメントを量子化 (u-law は A-law より信号レベルが低い)

A-law と u-law の相違点

- 線形近似が異なるため、長さおよび傾きが異なります。
- セグメントに対する 8 ビットコードワードのビット位置の数値割り当て、およびセグメント内の量子化レベルが異なります。
- A-law は u-law よりもダイナミックレンジが広くなります。
- u-law は A-law よりも低レベル信号における信号対歪み性能が高くなります。
- A-law では均一 PCM 相当に 13 ビットが必要です。u-law では均一 PCM 相当に 14 ビットが必要です。
- 国際接続では A-law を使用する必要があり、u から A への変換は u-law 使用国が行います。

差分パルス符号変調

PCM プロセスの段階では、入力サンプル信号間の差は最小限となります。差分

PCM (DPCM) は、この差を計算し、入力サンプル信号全体ではなく、この小さな差分信号のみを送信します。入力サンプル間の差は入力サンプル全体より小さいため、伝送に必要なビット数も少なくなります。したがって、音声信号を送信するために必要となるスループットを低減できます。DPCM を使用することで、音声伝送のビットレートを 48 Kbps にまで低減できます。

DPCM では、現在のサンプル信号と直前のサンプル信号の差はどのようにして計算されるのでしょうか。DPCM の前半部分の機能は、PCM とまったく同じです。「差分 PCM」と呼ばれるのもこのためです。入力信号が一定のサンプリング周波数 (入力周波数の 2 倍) で抽出されます。次に、これらのサンプルは PAM プロセスによって変調されます。ここで、DPCM が処理を引き継ぎます。サンプリングされた入力信号は、プレディクタ (予測器) と呼ばれる場所に格納されます。プレディクタは格納されたサンプル信号を、差分演算器に渡します。差分演算器は、直前のサンプル信号と現在のサンプル信号とを比較し、その差分を PCM の量子化および符号化フェーズに送ります。このフェーズでは、均一量子化、あるいは A-law または u-law によるコンパニングが使用されます。量子化および符号化された差分信号は、最終的な宛先に伝送されます。ネットワークの受信側では、すべての処理が逆に行われます。まず、差分信号が逆量子化されます。次に、プレディクタに格納されたサンプル信号にこの差分信号が追加され、ローパスフィルタに送られます。ここで、元の入力信号が再構築されます。

DPCM は、音声伝送のビット レートを抑制する優れた手法です。しかし一方で、音声品質に関する別の問題も伴います。DPCM では、直前のサンプル入力信号と現在のサンプル入力信号との差分が量子化およびエンコードされます。DPCM では、差分信号の量子化に均一量子化が使用されます。均一量子化により、小さな入力サンプル信号では小さな SNR が、大きな入力サンプル信号では大きな SNR が生成されます。このため、高い信号ほど音声品質も高くなります。人間の発する音声によって生成される信号は、ほとんどが小さいため、このシナリオはきわめて非効率的です。音声品質に関しては、小さな信号を重視する必要があります。この問題を解決するため、適応型 DPCM が開発されました。

適応型 DPCM

適応型 DPCM (ADPCM) は、ITU-T G.726仕様で定義された波形符号化方式です。

ADPCM では、DPCM プロセスで生成された差分信号に合わせて量子化レベルを適応的に変化させます。ADPCM では、量子化レベルをどのように適応制御するのでしょうか。ADPCM では、差分信号が低い場合は、量子化レベルのサイズが大きくなります。差分信号が高ければ、量子化レベルのサイズは小さくなります。このように、ADPCM では、入力差分信号のサイズに合わせて量子化レベルが適応的に変化します。この結果、差分信号のダイナミックレンジ全体において SNR が均一になります。ADPCM を使用することで、音声伝送のビット レートを 32 kbps まで下げることができ、これは A-law または u-law PCM のビット レートの半分に相当します。ADPCM では A-law または u-law PCM 同様、公衆電話交換網レベルの通話品質が実現されます。符号化機構にはフィードバックループが必要で、エンコーダ出力ビットを使用して量子化機構が再調整されます。

32 KB/秒の特定手順

ITU 標準 G.726 として適用できます。

- A-law または u-law PCM サンプルが線形 PCM サンプルに変換されます。
- 次のサンプルの予測値が計算されます。
- 実際のサンプルと予測値の差分が計算されます。
- 差分は 4 ビットとして符号化され、これらのビットが送信されます。
- プレディクタに 4 ビットをフィードバックします。
- 量子化機構に 4 ビットをフィードバックします。

関連情報

- [音声に関する技術サポート](#)
- [音声とユニファイド コミュニケーションに関する製品サポート](#)
- [Cisco IP Telephony のトラブルシューティング](#)
- [テクニカルサポート - Cisco Systems](#)