INCLUDEの第五回
公開日:04/7/8

第五回:音声圧縮！

皆さん、こんにちは。木星まで、あと2ヶ月の距離まで近づいた「INCLUDE」第五回です。

今回は、INCLUDE初の試み、「INCLUDE HYPER」をお届けします。といっても、何をするのかは単純。ハード開発局の不定期連載（第四回:予定）での解説の一部を、ここで説明するだけ。連動していると言いましょうか。

ハード連載の第四回で、携帯用の音声再生機を作ります（予定）。イヤホンから、CDの曲などが流れるものです。今人気のICプレーヤーです。（暗がり用のランプ機能付き！４８００円！嘘…。）しかし、いかんせん、記憶容量が少ないです。32Kバイトしかありません（シリアルEEPROM使用）。2HDフロッピーの約40分の一 ・・・

何ができるのでしょうか～ぁ！！！！！！！！！！！！！

これは、つまり２５６KBit。つまり、8KBitのモノラル音声なら、３２秒の音声を収納できるわけです。これはこれで、まあまあです。英会話の発音とか、音質にこだわらないものなら実用範囲でしょう。しかし、音質を維持しながら記録時間もかせぎたい場合、ただボーっと記録していては、ものの４・５秒で容量がいっぱいになってしまいます。

ただ記録するなら、音声はPCM音声なので、サンプリング周波数が物を言うわけです。高いに越したことはありません。高いほどいい・・・。高いほどいい・・・。高いほど・・・

どうすればいいんでしょうか～ぁ！！！！！

ならば、圧縮するまでです。単純ですね・・・。GUSHAーーーーーーすればいいんですよ。

かのMP3は、１２８KBitの音声でCDぐらいの音質を実現できます。しかも、エンコーダ（圧縮アルゴリズム）によって音質に選択性を持たせることができるので、すばらしい規格です。見習いたいものです。

また、ATRACは音質もさることながら、アルゴリズムが他よりも簡単「らしい」ので、導入にもってこいです。

一方、MICROSOFTのWMAは、音質と低ビットレートの両立を成し遂げた、優良規格です。さすがですね！

最近は、Oggなるものも登場しました。これについてはまだ未知数です。

個人的には、WMAが気に入ってます。シャカシャカ音になりやすいMP3よりも、聞きやすいですね。一方、MP3はその圧倒的な知名度と、確保済みの市場での独占的な商売が売りです。最近は、なにかごたごたがあるようですが・・・。

しかし、単なる「先人の知恵」とは違い、これらは、特許で保護された売り物ですので、そうやすやすと、他人の規格などを公の場でご紹介するのは厳しいものです。

そこで！例にもれず、OIDUSオリジナルで行きましょう。名づけて

「OCA」！ （Oidus Compressed Audio）

かっこいいですね～・・・いや、やっぱりかっこぉぉ悪いかも・・・。イインデスッテ！こういうこと一度でいいからやってみたかったんですから・・・。

なにをするアルゴリズムかと言うと、単純に、音声圧縮アルゴリズムです。名前の通りですね。そして、音質は当然、PCMそのものよりも「良くなくては」いけません。でないと、再生時に余計な負荷をかけてまで、聞きにくくなった圧縮音声を扱う意味がありません。

これはつまり、元ファイルに対する、圧縮後の音声の誤差が少ないことです。誤差は少ないほどいいです。特に、周波数方向の誤差と、振幅方向の誤差率（特に上方向への、つまり、音が元よりもうるさくなる方向）は、極力避けなければいけません。

また、20MHzのCPUくらいで実用なアルゴリズムでないといけません（今回ハードのほうで使用したCPUは20MHzのRISCCPUでしたので。実行に4クロックかかるので、実質5MHz駆動です。しかも、掛け算命令なし！もちろん、割り算命令なし！しかも、SIN命令なし！当然、COS命令なんてなし！）。しかも、ずぅっっとループですから、そのうちの44.1KHz分は、スピーカへの出力の負荷に費やされます。SINを実現する場合、大体500ステップくらい必要ですかね。そういうレベルなんです！

さて、音声圧縮においては、いろいろなアルゴリズムがありますが、どれも、基本は人間に聞き取れない部分の削除が基本になっています。これは、第一回で同じ事を言いました。

では、なぜそうする意味があるのでしょうか？

人間は、まず耳を通じて音声を取り込みます。これは、空気の振動です。これを、耳の中の鼓膜で増幅し、蝸牛で周波数を解析し、末梢神経に送ることで、脳へ電気信号として送る「らしい」です。詳しくは、他のページの解説なりをご覧下さいね。

さて、肝心の説明。

人間は、20KHz以上の音声を聞き取ることはできません。理由は僕は知りません。なるものはなるんですから、「へぇ」で済ましてしまいましょう。また、20Hz以下の音声も聞き取れません。こちらは、超低周波音というらしいです。これ以下も聞き取れません。（個人的に、秒間20回しか振動しない音まで聞き取れるなんて、すごいと思いましたね～。20KHzに比べたら、ほぼ極限じゃぁ無いですか！と言うか、1Hzの音ってどんな音なんでしょうね。聞いてみたいなぁ。ちなみに、超低周波音は、ものすごく波長が短く、物体を大きく揺らすみたいで、人間はこの音が嫌い「らしい」です。これまた、調べてみるといいでしょう。）

30年前のCD時代の幕開けの時から、つい5・6年前まで、音声記録の際には、この20KHzの音以上はカットされてきました。なぜなら、無駄だから。単純です。つまり、聞こえないわけ。聞けないものを記録したって、何の意味もありません。（まあ、最近ではこの解釈は明らかな、知識不足ということになります。長年の研究で、20KHz以上の音が、人間がリラックスする要素を含んでいるらしいことが分かってきましたから。DVD AUDIOや、SUPER AUDIOなどが、96KHzで記録したりしています。）

しかし、CDの記録方法では、携帯ICプレーヤの時代、700MBという容量なので、収まりきりません。現在、シリコンオーディオ（登録商標だったかも・・・。）と呼ばれる、ICに音声を記録した個型プレイヤーが巷ではやってます。上記の解説通り、記録できるICチップに、音声データを記録するもので、音とびが無い、省電力、ランダムアクセスが得意、などの理由で、いまや携帯電話の一部や、PDAにも搭載されています。（どちらかと言うと、後者はソフトで処理することが多いですけど。）

余談ですが、現在最高の記録用ICの容量は8GBitで、CDのおよそ1．3倍くらいです（FlashROMです）。これなら、CDだって丸々収まりますね。

しかし！これはでたばかり。べらぼうに高いです。これを除く、本体すべての製造費より、このチップのほうが高い場合なんて、ざらにあります。それくらい高価なものなのです。

また、少々安いものでは、容量が4分の一くらいのものになってしまいます。これでは、CDなんて到底収まりきりません。

インターネットでの音楽配信の問題もあります。いくらブロードバンド時代とはいえ、生のCD音質のデータを配信するのは、無駄もありますし、同時アクセスを困難にしますので、この分野の発展にとって、大きな障害となることでしょう。

こう言った、容量の問題などで、CD音質の音声よりも、さらに音質を下げなくてはならなくなってきました。（CD基準になることが多いのは、おそらく、CDの音質のよさと言うブランドと、長年の蓄積、記録媒体がCDとDVDとMDくらい、ということによるものでしょうね。）

CDの音声で、「ある程度」無駄な部分は、20KHz付近の音声です。人間にとって、聞き取りにくい音は、記録しても聞き取りにくいのですからちょん切っちゃいます。チョッキン！

こうすることにより、たいてい20KHzの音が密集する楽曲などの音声（つまり、携帯プレーヤで皆さんが主に聞く音声）は、音質をある程度維持しつつ、容量を思いっきり減らすことができるのです。

これにより、先ほどご紹介したICプレーヤなどで、実用的な製品などが作れるようになるのです。無論、音質は小型プレーヤレベルですけど。（というか、配信と持ち運びと、音楽ライブラリ作り、DVDビデオの音声・・・など音声圧縮を利用するものは、たいてい音質が犠牲になってる感じです。）

今回のハードは、それの1000000万分の一くらいに感じる、32KBという「極」低容量ですので、圧縮は必須です。圧縮することにより、ただサンプリング周波数を下げて再生時間をかせぐよりも格段に音質を上げられることは、MP3などの例もあり、明らかでしょう。

ただ、ICチップ自体は、音質と引き換えに、100円近くで買えるものなので、まあ、良しとしましょうか。

さて、それではそれでは。実際のアルゴリズムに入りましょう・・・ちょっとまった。すごく大切な解説があるんですよぉ。僕自身、難しくて理解しきってない上に、今回のような音声圧縮には欠かせない、たちの悪い（分かる人にはものすごくたちがいいんでしょうが・・・）数学を・・・理解せなばイカンザキ！

音声データも、元をたどれば単なる1次元数列です。これは、メモリアドレス上に存在する単なるデータに代わりありません。

しかし、たちの悪いことに、音声は周波数成分を含みますので、ただA/Dコンバータを通過した、数列データとして保持すると、周波数に関係なく記録してしまうので、圧縮に応用するのが困難になります。目的は高周波成分のカットですから。単純に、ローパスフィルタを通して、通過できたときだけ保持するのなら、容量は下がるでしょうね（コンデンサなどには交流信号の強い味方？敵？なリアクタンスと言うものがありますから）。しかし、そうすると、時間方向の情報が失われてしまいます。

つまり、数列に過ぎないデータを、意味のある（元来の）周波数データに置き換えなくてはなりません。この操作が、音声圧縮において「必須アミノ酸」なのです。レベルは・・・イソロイシンくらいでしょうか・・・（笑）バリンロイシンイソロイシン！…。

これを実現するのが、「フーリエ」さんの考えた数学的手法、「フーリエ変換」です。フーリエ変換とは、直交変換と言うものの一つです。

直交という状態は、数学ではつまり、「見方によらない」と言う意味でしょうかね。ある事象を、違う側面で見ても、お互いの情報を含んでいれば、もうかたっぽから相手の状態を「自分の状態」から説明できます。

つまり、お互いは、ある事象のそれぞれの側面でしかないわけです。(素粒子物理学なんかでも、「側面」という言葉は良く使われます。理解しておきましょう！)

今回の周波数への置き換えは、時間信号である音声を、周波数信号という側面で見ることに他なりません。それを実現するのが、「フーリエ変換」であるのです。(これしか方法が無いわけではありませんが、これが一番文献とかも多いですし、一般的で、かつ、この分野の常識ですので。)

公式は・・・

$$F(\omega) = \int_{-\infty}^{\infty} X(t) \times e^{j\omega t} dt$$

です。・・・。HTMLでの上手い改行の仕方を知らないので、これでご勘弁を！ちなみに、見にくいでしょうが、jωtは、eの指数です(そうは見えない！)。

F(ω)は、周波数信号です。つまり、変換後の信号関数。ωは角速度です。また、X(t)は時間信号、すなわち、変換前の数列信号です。角速度とは、単振動の弧を移動する速度のことです。つまり、1秒間に進める角度のこと。または、弧の長さを進むその距離を単位時間で割ったものとしても同じことです。角度は弧との変換ができますから。また、jとは、虚数単位のことです。普通の数学では「i」と表記しますが、交流電流の単位表記と間違いやすいので(と言うか、文字は同じ)、「j」と表記します。

さて・・・意味がわかりませんねぇ～ぇ。さっぱりです。わからんのです。と言うか、数学で万年赤点の僕にとって、初の試みでしょう。微積すら満足にわからないのに、敷居が高すぎます。しかし、やるしかないので、やります！

さて、公式だけをただ書き連ねていくのもどうかと思うので、必死に説明しましょう。

まず、あらゆる波信号は、フーリエ級数展開と言うものであらわすことができるのです。これに関しては、異才フーリエさんの、長年の研究の賜物として、ありがたく利用させていただきましょう。

波はつまり、周期があります。それは、三角関数で表記できます。なぜ、波が三角関数で表記できるのかと言うと、波は、単振動が起こした初期振動が、媒質を伝わっていく軌跡をあらわすからであって、つまり、三角関数の定義である「単位円」での上下振動が、三角関数で表記できるからによるものです。

これにより、たとえば、周波数1Hzの信号は、振幅をAとすると、Asinθと表記できるわけです。これをθと角速度の間で変換をすると、AsinωTとなります。どう導き出すかと言うと、⊿t秒間に弧を動くとすると、そのときに進む角度を⊿θとしたとき、ω＝θ ÷ t で角速度を定義するため、ωにtをかけたものがθになります。

さて、これを踏まえて、波の合成を考えてみましょう。

たとえば海岸に打ち寄せる波が、跳ね返った波と打ち消しあうのを見たことがあるでしょう。あれが波の合成です。あるいは、沖合いからの波がだんだん高くなっていくのもそうです。

これらのように、波は重ね合わせることができるのです。

また、波の独立性というものがありますので、重なった波を単純に複数の波の和とすることができます(一見しただけでは判断がつきませんが。可逆変換なのです)。詳しくは物理を学んでください。

これらのことから、複雑な波を複数の波の和と捕らえることができ、それはつまり、式で表す事が出来るというわけです。

また、sinは偶関数ですので、0から始まる波以外を表す事ができません。よって、cosも考えてみます。ちなみに、cosは奇関数です。

さて、ここで、フーリエ級数と言うものが出てきます。

波は、ある基本周波数を考えたときに、その基本周波数の整数倍の波の和として表す事ができます。早い話、これがフーリエ級数です。これは波の重ね合わせの一般式みたいなものです。

波の要素である振幅(音の大きさ)・振動数(周波数=周期の逆数)を1つの式であらわすと、

$$波＝A\sin\omega nt$$

とあらわせます。Aが振幅です。$\omega＝2\pi \div T$、$f＝1 \div T$ですので、つまり、$\omega \div 2\pi$が周波数。tは時間軸上のsinの位置を表すのに必要な変数です。そして、nが肝です。nは基本周波数の整数倍の波であることを表す変数です。

そして、これらの波を重ね合わせることを、一つの式で表します。つまり、波の一般式です。

振動の中心をa0と置き、先ほどの波の式と、「基本周波数の整数倍の波の和」と言うことを考えて、以下のように置きます。ちなみに、an=振幅の係数、$\omega \times n$=周波数です。

$$X(t)=a0 + \sum_{n=1}^{\infty} (an \times \cos\ \omega \times n \times t + bn \times \sin\ \omega \times n \times t\ )$$

これで、波の一般式が完成しました。これが、先ほどのフーリエ変換の公式のX(t)です。これは、音声信号の全てを表す一般式と考えていいですね。つまり、音声の式。

次に、フーリエ変換の式のeごたごたの部分について説明します。

eの指数に虚数単位がありますが・・・とりあえず、オイラーの公式と言うものを知っておく必要があります。

オイラーの公式とは、

$$\pm e\hat{}j\omega t＝\cos\omega t \pm \sin\omega t$$

です。一見、意味のわからない代物ですが、これを利用すると、ある1つの複素数で2つの値をあらわせるのです。複素数とベクトルが密接な関係にあることを考えると、これは式の変数を減らす工夫と考えてよろしいのではないでしょうか。

そして、いくら複素数があるとはいえ、利用したいのは右辺です。こちらは実数なので、普通に位相をあらわす式として使えます。

証明は、「マクローリン展開」というものを利用します。ただ・・・文章が長くなるし、証明をするのが目的ではないし、フーリエ変換では道具として使うだけなので、今回は「なし」のベクトルでいきたいと思います。

さて、これを縦方向に軸として取ると、これはつまり、波の軸になります。つまり、tにおける振幅の大きさです。

先ほどのX(t)は、このeを使ってあらわすとe^(jtω × 倍数)の和とあらわせます。これは、単なる式変換です(代入した結果)。

さてさて、先ほどのX(t)と、eごたごたとの積を考えてみますと、以下のようになります。

$$X(t) \times e^{j\omega t}$$

ですね。これをtで積分、つまり、時間軸に対して面積を求めると、

$$\int_{-\infty}^{\infty} X(t) \times e^{j\omega t} dt$$

となります。これは、つまり、フーリエ変換の右式です。左式のF(ω)は、ωしか値を与えません。つまり、ωの関数と言うこと。つまりは、任意のωに対する関数だと言うことです。ω＝２π ÷ T、f＝１ ÷ Tですので、つまり、ω÷２πが周波数になることは先ほど述べました。これを利用すると、任意の周波数での、信号の存在を確かめることができるわけです。

では、なぜそういえるのか、いよいよ大詰めです。

X(t)は波の一般式です。a0 + a1sinωt × b1cosωt + ・・・とあらわされます。

さて、問題は、ωが与えられたときに、X(t) × e ^jωt はどうなるのか・・・です。

たとえば、X(t)＝a1sinωt × b1cosωt + a2sin2ωt × b2cos2ωt 、つまり、

$$X(t) = e^{j\omega t} + e^{j2\omega t}$$

と置いてみましょう。

与えられたωが２ωとなるとき、つまり、基本周波の２倍とします。そのとき、X(t) × e ^jωt の式は、

$$e^{j2\omega t} (e^{j\omega t} + e^{j2\omega t})$$

となりますね。

このとき、e ^jωt の方は、与えられた周波数が違うので、積を取って、積分すると、結果が0になってしまいます。つまり、周波数のずれが起きて、面積が正の部分と負の部分とで相殺してしまったのです。

一方、同じ周波数を持つ e ^j2ωt は、ずれが起きないため、負の部分が掛け合わされて正

になり、つまり、正の部分が2つでき、結果0にならなかったわけです。

つまり、信号の、任意の周波数における波の強さを表す係数が求まるわけです。

これを、1から延々と求めていけば波の一般式が求まるわけです。すごく時間がかかりますけど。

だから、時間信号と周波数信号は直交なのです。つまり、双方向に可逆変換できると言うことです。

以上の説明から、任意の時間信号を周波数帯で表すことで、いらない部分をチョッキンしたりもできるわけです。

・・・・・・・・・・・・。ぜえぜえですねぇ。2時間もかかりました。でも、意外と言ってることは単純なんですね。驚きです！大発見です！

しかーし！これではおわりません！更なる理論が必要になってきます。

このフーリエ変換、すごく便利なんですけど、問題は計算数の多さです。これを各tに付いて行うと、掛け算やら複素数やらべき乗やらΣやら∫やらと、とてつもない計算を行わなくてはなりません。そうすると、ただでさえ軟弱な今回のCPUで、しかも掛け算命令もない中では、もうブラックホールの事象の地平面すれすれまで接近した状態なのです。どうしましょ！です。

そこで！コンピュータなどでの離散的な信号を扱う場面で有効な、「計算減らし」の方法があるのです。

その名も！「離散フーリエ変換」

続きは次回！「続！音声圧縮ー2」で！

おしまい

メール等の受付

当サイトの管理人は、MORIOです。

質問やご要望、ご感想、苦情などは、メールで受け付けております。以下のアドレス宛に送ってくださいませ。

master@morik.net

form 2006/1/9