CSM 音声合成とは？

CSM 音声合成は音声符号化の一種で、元の音声を複数のサイン波（＝正弦波）の組み合わせで再現する手法です。CSM は Composite Sinusoidal Modeling の略で「複合正弦波モデル化」という意味です。

「どんな連続波形も三角関数の足し算で表せる」という理論があります。この理論を元に、瞬間的な音声波形をサイン波に分解します（周波数解析）。強く現れている音程のサイン波を適切な音量で複数同時に再生すれば、元の波形と似た波形を再現できます。

この音声合成法に必要な要件は「任意の音程・音量のサイン波をできるだけ多く出力できること」だけです。FM 音源はサイン波のカタマリのような音源ですから、この音声合成法と、とても相性が良いと言えるでしょう。事実、YAMAHA の FM 音源チップには CSM 音声合成をサポートするような機能が存在しています。

CSM 音声合成モード

ほとんどの YAMAHA 製の FM 音源には、内蔵のタイマがオーバーフローしたときに全チャンネル（もしくは一部チャンネル）を即座にキーオン / キーオフする機能があります。YAMAHA はこの機能を「CSM モード」もしくは「CSM 音声合成モード (CSM Speech Synthesis Mode)」と呼んでいます。

開発者は「音の出力を定期的に確実に更新するから、適切に出音の各パラメータを設定してね！じゃ、後はよろしく！」とでも言いたかったのでしょう。どちらかと言えば、これは CSM 音声合成そのものではなく、CSM 音声合成時に必要な待ち時間を管理するための機能と思ったほうが良いでしょう。機能としてはそれだけなので、「CSM 音声合成モードを使えば誰でも音声合成ができる！」というわけではありませんし、「CSM 音声合成モードがないから、CSM 音声合成はできない」ということもありません。

いずれにしても音声合成用のデータは自前で用意する必要があります。また、自前で時間管理を行い適切にサイン波が発音できれば、この方式による音声合成自体は実現可能ですから、CSM 音声合成モードを使わなければならない必然性もありません。事実、本稿の解説も CSM モードは使用していません。NRTDRV が管理している時間単位で処理をしています。

X1 には FM 音源タイマ割り込みがないため、CSM モードを単純利用できません。プログラムを書いて、ムリヤリ使用することは可能だと思われますが、現時点ではそのような仕組みは NRTDRV 上に実装されていません。

> ※ややこしいですが、本稿では「CSM 音声合成モード＝ FM 音源上に搭載されている機能」「CSM 音声合成＝複数のサイン波を同時に出力して元波形を再現する方法」を指しています。明確に使い分けていますが、ココを混同するとワケが分からなくなると思いますので、どうかそのようにお読みください ^^;

■（参考資料）各 FM 音源の CSM モードレジスタ

YM2151(OPM)

14H bit7

YM2203(OPN), YM2608(OPNA)

27H bit7/bit6(00= 通常 /01= 効果音モード /10=CSM 音声合成モード)

YM3812(OPL2), Y8950(MSX-AUDIO)

08H bit7

YM2413B(OPLL)

確認できず

使用例

最も著名なのは、やはりなんといっても、一連のゲームアーツ制作の PC-8801 用ゲームソフトでしょう。

シルフィード (1986 年) にはじまり、ぎゅわんぶらあ自己中心派 (1987 年)・ゼリアード (1987 年)・ヴェイグス (1988 年) といったゲームで、喋りまくっています。後期は、PC-8801 以外の機種でも FM 音源搭載機種ならば、積極的に CSM 音声合成を使用しています（X1turbo 版のゼリアード (1988 年) や MSX 版のぎゅあんぶらあ自己中心派（1988 年・要 FM-PAC）など）。

もっとも、『喋りまくっている』と言うと語弊があるかもしれませんが・・・「ザカリテ」や「しめさば」がキーワードでしょうか（笑） もし、ご存じない方は、YouTube 等に動画がありますので、一度ご覧になってみてください。残念ながら、筆者は当時あまり耳にする機会がなかったのですが、それでも店頭デモ等でオープニングに必ず流れる「Presented by GAMEARTS」の音声合成はよく聞きましたし、強いインパクトを受けたものでした。今聞いても、その独特の質感は面白いです。

これらの音声合成を実現していたのは、ゲームアーツ社内で三橋正邦氏が開発した「CSM トーキングシステム」です。ゲームアーツ製のゲーム以外ではほとんどお目にかからなかった手法でしたが、三橋氏の手腕の賜物だったわけですね。

OPM での使用例

実はあまり確認できません。X1turbo 版の「ゼリアード」（ゲームアーツ）くらいでしょうか。

OPM が標準で搭載されていたパソコンは事実上、X1 と X68000 の二機種しかないといって良いでしょう。X1 においては前述の通り、OPM のタイマ割り込みがないため、CSM 音声合成モードは非常に扱いづらいものになっています。X68000 においては ADPCM が搭載されていたため、単にしゃべらせるだけなら ADPCM を利用した方が断然楽でした。

マイナーなところでは MSX(SFG-01) や PC-8801(響) への搭載例もあります。MSX ではシステムに喋らせていたという情報もありますが、残念ながら筆者は未聴です。また、アーケードゲームでの使用実績はかなりの数に登りますが、X68000 と同様の理由で、CSM 音声合成はおそらくほとんど使われてなかったのではないかと考えられます。

利点と欠点

低リソース・低負荷で再生が可能

当時の 8bit パソコンの性能では PCM の処理は容量的にも処理能力的にも厳しいですが、CSM 音声合成は余裕を持って処理できます。ただし補足すると、これは PCM データと比較した場合の話で、単純な演奏データとして見るとそれなりにメモリ喰らいです。「もってけ〜」の場合、容量的に TV サイズ版しか作れませんでした。おおよその内訳は、システム及びドライバ部 (拡張用予約を含む) が 16kbyte、バックトラック演奏用データが 8kbyte 弱、CSM 用データが 40kbyte ほどです。

データ作成に時間がかかる

離散フーリエ変換、およびデータの評価はかなり時間のかかる処理です。「シルフィード 100 の秘密」で語られていますが、当時は PC-9801(10MHz) で一秒解析するのに数時間かかったそうです。現在は非力なパソコンでも Excel で数十秒の解析にものの数分〜数十分で済みます。良い時代になったものです。

元になる音声波形データが必要

ゼロから作り上げることはできません。特殊な形式ですが「変換」の一種なので、元ネタが無いとなにもできないという、PCM と同じ欠点は引き継いでしまいます。ただし、周波数解析自体が PCM の各種エフェクトに使われる技術でもあります。さまざまな加工ができる余地があり、いろんな可能性を秘めています。

変換元のサンプルによってはまともな音声にならない

人の声なら何でも良いというわけでもなく、大勢の人間がワイワイ騒いでいるようなものや、深いリバーブがかかったもの、ヴォコーダーボイスなどは再現性が非常に悪く、はっきり言ってなにがなんだかわからないものになります。複雑な波形の再現には向かない方法だと言えるでしょう。

原著論文

古くからある技術ですが、一体いつごろ確立されたものなのか疑問に思ったので調べてみました。日本国内ではどうやらこの論文が原点のようです。ちなみに筆者は未読です。機会があったら読んでみたいですが、理解できるのかしら。。。

嵯峨山茂樹・板倉文忠 (1979 年) 「複合正弦波による音声合成」
http://hil.t.u-tokyo.ac.jp/~sagayama/