あなたの発言:

音声のデジタル処理についてお尋ねします。いつ頃現れたものでしょうか。代表的な技術について手短に説明して下さい。

8s考えました >

ChatGPT の発言:

音声のデジタル処理の歴史は、意外に古く、**基本的な考え方は1930年代に現れ、1960〜70年代に実用化が進み、1980年代に一般家庭に普及した**、とまとめられます。

興味深いことに、ここでも音楽の録音より、電話通信の技術が先行しています。

1. 主な技術の歴史

年代	技術・出来事	意義
1937年	PCM（パルス符号変調）の発明	音声を数値として表現する基本技術
1943年頃	デジタル音声による秘話通信	第二次世界大戦中の軍事利用
1940〜50年代	標本化理論の確立	連続信号を離散的な値から復元できる条件を明確化
1960年代	デジタル電話通信の実用化	多数の通話をデジタル伝送
1960年代末	デジタル録音機の開発	NHK技術研究所が先駆的な開発
1970年代	デジタル録音の商用化	日本コロムビア（DENON）などが先行
1982年	CD発売	デジタル音声が一般消費者に普及
1990年代	MP3などの音声圧縮	音楽ファイルの小型化、ネット配信への道
2000年代以降	高性能DSP、ソフトウェア処理	ノイズ除去、音声認識、音質補正などが普及

PCMは1937年にアレック・リーブスが考案しました。実用化には時間がかかりましたが、1940年代にはベル研究所が重要な役割を果たしています。 IEEE Comm... +1

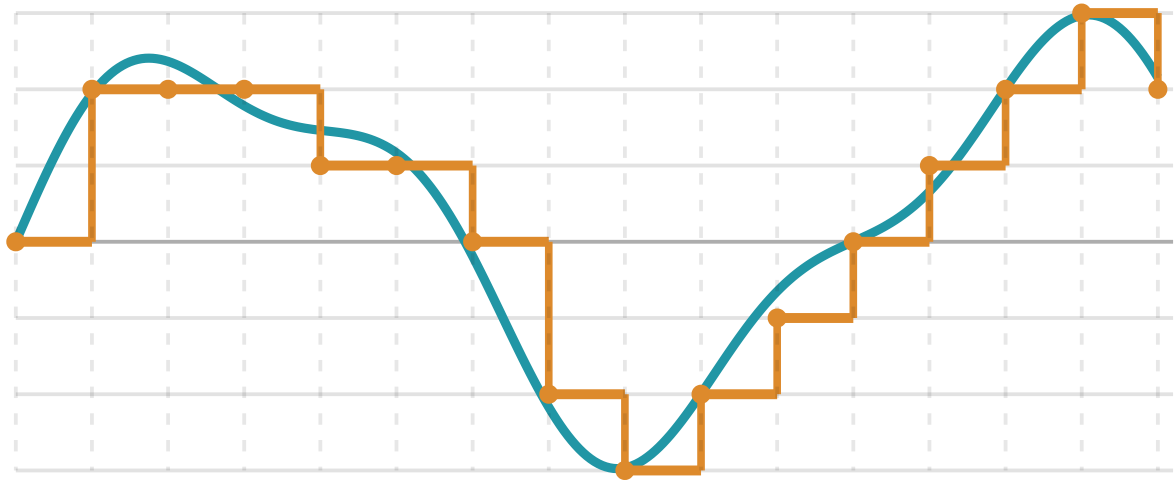
日本のNHKも先駆的で、1960年代末にはデジタル録音機を開発し、1971年にはその装置による商業録音が発売されました。 公益社団法... +1

2. 代表的な技術

(1) PCM — 音を数値にする

最も基本的な技術です。

音声を時間間隔 Δt ごとに測定し、その振幅を有限個の数値の一つとして記録します。



元の連続信号　　標本化・量子化された信号

模式図。実際のPCMでは各標本の振幅を2進数で記録します。

CDの場合には、

- 標本化周波数：44,100 Hz
- 量子化：16ビット（65,536段階）
- チャンネル数：2（ステレオ）

という規格が採用されました。

標本化定理によれば、周波数帯域が $f_{\max}$ 以下の信号は、理想的には $2f_{\max}$ より高い周波数で標本化すれば復元できます。

(2) デジタルフィルタ ― 音質を調整する

音声を数列 x_n として扱えば、例えば

$$y_n = \frac{x_{n-1} + x_n + x_{n+1}}{3}$$

のような演算が可能です。

これは簡単な平滑化フィルタで、高周波成分を弱める作用があります。

より高度なフィルタを用いれば、低音・高音の調整、雑音の抑制などができます。

前に話題になった**プリエンファシスとディエンファシスも、デジタルフィルタで実現できます。**

(3) FFT ― 周波数成分を調べる

高速フーリエ変換（FFT）は、離散フーリエ変換を高速に計算するアルゴリズムです。

特に1965年のクーリーとテューキーの論文を契機として広く普及しました。

音声を周波数成分に分けて分析できるため、雑音除去、音声認識、音響測定などの基礎になります。

(4) デジタル音声圧縮 ― データ量を減らす

PCMでは大量のデータが必要です。

そこで、音声の特徴や人間の聴覚特性を利用して、記録する情報量を減らします。

代表例が1990年代に普及したMP3です。

人間には聞こえにくい成分などを省略することで、聴感上の品質をなるべく保ちながらデータ量を大幅に減らします。

なお、電話音声の圧縮にはもっと古い歴史があり、1930年代のベル研究所におけるホーマー・ダドリーのボコーダー研究にまで遡ります。 AES E-Libra… +1

3. デジタル処理は何を変えたのか

私が特に重要だと思うのは、次の区別です。

アナログ技術でも可能だったこと

周波数特性の調整、増幅、フィルタリング、音の合成など。

デジタル技術によって飛躍的に容易になったこと

複雑なフィルタの正確な実現、録音の編集、劣化の少ない複製、誤り訂正、効率的な圧縮など。

つまり、デジタル化によって初めて音の加工が可能になったわけではありません。**アナログ回路で実現していた処理を数値演算に置き換え、さらに複雑で柔軟な処理を可能にした**ことが大きな変化です。

そして歴史を振り返ると、エジソン以来の録音技術と、ベル以来の電話通信技術という二つの流れが、20世紀後半にデジタル音声技術として合流した、と見ることができそうです。

あなたの発言:

CDの場合は、開発したメーカー (フィリップスとSONY?) が規格を定めたのだと思いますが、デジタル録音の場合はファイル・フォーマットの規格は誰がどのように定めたのでしょうか。WAVE ファイルが良く知られていますが、それが初めてというわけでもないのでしょうか。

36s考えました >

ChatGPT の発言:

はい。ご理解のとおり、CDの規格はソニーとフィリップスが共同開発したものですが、**デジタル録音のファイル形式については、CDのような統一規格が最初からあったわけではありません。**

また、**WAVE（WAV）は1991年にMicrosoftとIBMが規格化したもので、それ以前から別の音声ファイル形式が存在していました。**

この歴史を理解するには、次の三つを区別するのが重要です。

1. **音声の数値表現**：PCMなど。音をどう数値に変換するか。
2. **録音媒体への記録方式**：磁気テープやCDなどに、数値をどのように記録するか。
3. **コンピューターのファイル形式**：音声データをどのような構造のファイルに格納するか。

この三つは別々に発展してきました。

1. 1970年代：デジタル録音機の時代

CDが登場する前から、放送局やレコード会社ではデジタル録音が行われていました。



例えば、1978年に発売されたソニーのPCM-1600は、U-matic方式のビデオテープレコーダーを使って音声をデジタル記録する装置でした。

当時は、音声データを映像信号に似た形に変換し、ビデオテープに記録する方法が広く使われました。 chapter 2 +1

この段階では、録音機の製造会社が独自の記録方式を採用することが珍しくありませんでした。

ただし、機器間の互換性を確保するための規格化も進み、日本ではEIAJ（日本電子機械工業会）によるPCMアダプターの規格が1970年代末に成立しています。 AES +1

重要なのは、**これらは主として磁気テープ上の記録方式の規格であり、現在のようなコンピューターのファイル形式ではなかった**ということです。

2. 1980年代：コンピューター用の音声ファイル

パーソナルコンピューターで音を取り扱うようになると、ファイルとして保存し、ソフトウェア間で交換するための規格が必要になりました。

代表的なものを年代順に示します。

年	ファイル形式	開発・規格化した主体
1980年代半ば	IFF 8SVX	Electronic Arts（Amiga向けなど）
1988年	AIFF	Apple Computer
1980年代末頃	AU / SND	Sun Microsystems、NeXT
1991年	WAVE（WAV）	Microsoft、IBM
1990年代	Broadcast Wave	欧州放送連合（EBU）

この中で特に注目したいのは、WAVEより早く登場したAppleのAIFFです。

AIFF（1988年）

AIFFは *Audio Interchange File Format* の略です。

Appleが1987〜88年に音楽ソフトウェアの開発者たちと協議して策定した規格で、最初の仕様書は1988年1月21日付です。

基盤になったのは、ゲームソフト会社Electronic Artsが1985年に開発したIFF（Interchange File Format）でした。

つまり、意外なことに、**初期の標準的な音声ファイル形式の発展には、コンピューターゲーム業界も大きく関わっていた**わけです。 The Zamzar… +1

WAVE（1991年）

WAVEはMicrosoftとIBMが1991年8月に公開した、RIFF（Resource Interchange File Format）という汎用ファイル構造に基づく音声形式です。

Windows 3.1のマルチメディア機能とともに広まり、現在まで広く使われています。

The Library … +1

AIFFもWAVEも、特定の録音機を動かすための形式ではなく、様々なプログラムが音声データを共通して扱えるようにするための形式でした。

3. WAVEファイルはどのように作られているか

WAVEの基本的な考え方は、非常に単純です。

RIFFヘッダー
このファイルがRIFF形式のWAVEファイルであることを示す。

fmt チャンク
PCMかどうか、標本化周波数（例：44,100 Hz）、量子化ビット数（例：16ビット）、チャンネル数（例：2）など。

data チャンク
実際の音声データ。PCMなら標本値を時系列に並べたもの。

単純化したPCM形式のWAVEファイルの構造

この方式なら、ファイルを受け取ったプログラムは、標本化周波数などを事前に知らなくても、ヘッダーを読んで適切に再生できます。

なお、WAVEは入れ物の形式であり、PCM以外の圧縮音声などを格納することもできます。

4. CDとWAVEの関係

CDの音声データは、44.1 kHz、16ビット、ステレオという意味では、一般的なWAVEファイルと同じPCM形式です。

しかし、音楽CDそのものに `.wav` というファイルが格納されているわけではありません。

CDには、音声の標本値のほか、誤り訂正や同期などのための情報が、CD-DA（Red Book）の規格に従って記録されています。

この音声をコンピューターに取り込み、WAVE用のヘッダーを付ければ、通常のWAVファイルとして扱えます。

5. 誰が規格を決めたのか

歴史を整理すると、次のようになります。

- **1970年代の録音機**：各メーカーが独自に開発し、業界団体も相互接続のための規格を策定した。
- **1980年代のコンピューター**：Appleなどが、音声データをソフトウェア間で交換するための形式を策定した。
- **1990年代以降**：Microsoftなどの規格が普及し、放送業界などでも共通形式への標準化が進んだ。

したがって、WAVEは最初の音声ファイル形式ではありません。また、国際標準化機関が最初から定めた規格でもなく、企業が公開した仕様が広く採用された、いわゆる事実上の標準（de facto standard）です。

ここで一つ、興味深い点があります。

デジタル録音の歴史は1960年代末から始まっていますが、現在私たちが使っているWAVEやAIFFのような音声ファイル形式の歴史は、それより20年ほど新しいのです。

その間にあったのは、「コンピューターのファイルとして音を保存する」というより、「専用の録音機を使い、磁気テープにデジタル信号を記録する」という発想でした。

この違いは、デジタル録音の歴史を理解するうえで、かなり重要だと思います。

参考として、米国議会図書館のWAVE形式解説 ↗ と、Appleの初期AIFF規格書 ↗ を挙げておきます。後者は当時の開発思想も読み取れる興味深い資料です。

あなたの発言:

> 考えてみると、デジタル録音はパソコンやインターネットよりも歴史が古いわけですから当然ですね。

ChatGPT の発言:

そうですね。特に面白いのは、**デジタル化とコンピューター化は、必ずしも同じことではない**という点だと思います。

初期のデジタル録音機は、音声信号を数値に変換し、その数値を磁気テープに記録する専用装置でした。汎用コンピューターもファイルシステムも必要ありません。

また、デジタル録音が最初から音質向上だけを目指したわけではなく、電話通信や秘話通信などの要求から発展したことも興味深いところです。

もう一つ付け加えるなら、デジタル録音自体はパソコンより古いのですが、**デジタル信号処理の数学的な基礎はさらに古い**のです。標本化の考え方は1920年代に遡りますし、フーリエ解析はもちろん19世紀以前からあります。

こうして振り返ると、音声技術は、物理学、数学、電気工学、通信工学、計算機科学が、時代をずらしながら合流してきた分野なのですね。

あなたの発言:

サンプリング周波数についてお尋ねします。音楽CDの場合44100Hz が標準的、それ以外で 48000 Hz というのを良く目にします。

前者の場合、素因数分解が印象的で、平方根も整数値になる。何か狙ってそうしたわけでしょうか。

一方の48000の方は誰が決めたのでしょうか。

16s考えました >

ChatGPT の発言:

ご指摘のとおり、44,100はなかなか面白い数です。

$$44100 = 210^2 = 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^2$$

確かに平方数です。しかし、調べてみると、**44,100 Hzが選ばれた理由は数論的な性質ではなく、当時のテレビの走査方式との関係**でした。

一方、48,000 Hzは放送・業務用音声の標準化の過程で選ばれたものです。こちらは誰が決めたかも、かなり明確です。

1. 44,100 Hzはなぜ選ばれたか

1970年代、デジタル音声を記録するためにビデオテープレコーダー（VTR）が利用されていました。

PCMで数値化した音声をテレビの映像信号に変換し、VTRに記録していたのです。

当時、テレビには大きく分けて二つの走査方式がありました。

- 欧州など：625本、毎秒50フィールド
- 日本・米国など：525本、毎秒60フィールド（白黒方式の場合）

そこで、映像信号の1走査線に音声の標本を3個記録する方式を考えます。

同期信号などに使う走査線を除外すると、次の計算になります。

625本・50フィールド方式

$$294 \times 3 \times 50 = \boxed{44100}$$

294本／フィールド × 3標本／本 × 50フィールド／秒

525本・60フィールド方式

$$245 \times 3 \times 60 = \boxed{44100}$$

245本／フィールド × 3標本／本 × 60フィールド／秒

驚くべきことに、両者が一致します。

もちろん、これは適当な本数を選んで偶然一致したということではなく、両方式で利用可能な走査線数を踏まえて設計されたものです。

ただし、実際のNTSCカラーテレビは毎秒約59.94フィールドなので、対応する値は約44,056 Hzになります。

ソニーが開発したPCM録音装置の技術的条件が、CDの44.1 kHzという標本化周波数につながりました。 Stason +2

したがって、平方数であることについては、意図的にそうしたという史料は見当たりません。**テレビの同期方式に適合するよう選んだ結果、平方数にもなった**と考えるのが妥当でしょう。

2. 48,000 Hzは誰が決めたか

こちらはCDとは異なり、特定のメーカーが決めたというより、放送・音響技術の標準化団体が協議して決めたものです。

1970年代には、業務用デジタル録音機の標本化周波数が統一されていませんでした。

32 kHz、44.1 kHz、50 kHzなどが使われ、機器間の接続に問題が生じていました。

1981年頃には、次の団体などで48 kHzを共通規格として採用する方向に議論が進んでいました。

- AES（Audio Engineering Society：音響技術者の学会）
- EBU（European Broadcasting Union：欧州放送連合）
- SMPTE（映画・テレビ技術者の団体）
- CCIR（国際無線通信諮問委員会、現在のITU-Rに相当）

1982年のSMPTEの論文に、まさにこの標準化の経緯が述べられています。 SMPTEジャーナ...

具体的には、**1984年にAESがAES5として48 kHzを推奨し、1986年にはCCIRが勧告646で放送スタジオに48 kHzを採用しました。** AES +1

3. なぜ48,000だったのか

大きな理由は三つあります。

第一に、可聴帯域20 kHzを十分にカバーできること。

標本化定理では40 kHzを超えればよいことになりますが、実際のアンチエイリアシングフィルタには遷移帯域が必要です。48 kHzならナイキスト周波数が24 kHzとなり、余裕が生じます。

第二に、既存の32 kHzとの関係が簡単なこと。

$$\frac{48000}{32000} = \frac{3}{2}$$

当時、放送音声の伝送には32 kHzも利用されていました。簡単な整数比は標本化周波数の変換に有利です。

これは1986年のCCIR勧告でも、48 kHzを選ぶ理由として明記されています。 ITU

第三に、映画・テレビのフレーム周波数との関係がよいこと。

フレーム周波数	1フレーム当たりの標本数
24 Hz（映画）	2,000
25 Hz（欧州テレビ）	1,920
30 Hz	1,600

これらはすべて整数です。

ただし、29.97 Hzのカラーテレビでは整数にならず、同期に工夫が必要でした。

The Library of ...

4. 二つの周波数の関係

数論的な面白さという点では、こんな違いがあります。

$$44100 = 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^2$$

$$48000 = 2^7 \cdot 3 \cdot 5^3$$

最大公約数は300なので、

$$\boxed{\frac{44100}{48000} = \frac{147}{160}}$$

となります。

したがって48 kHzから44.1 kHzへの標本化周波数変換は、理論的には147倍のアップサンプリングと160分の1のダウンサンプリングを組み合わせて実現できます。実際にはデジタルフィルタを用いて効率よく処理します。

ちなみに、現在でも44.1 kHzと48 kHzの系列は並存しており、88.2 kHz、96 kHz、176.4 kHz、192 kHzといった周波数にも、その名残が認められます。

歴史的には、**44.1 kHzは初期のPCM録音におけるVTRとの互換性から生まれ、48 kHzは放送・業務用途での国際的な標準化によって定着した**、という違いがあります。

両方とも約40年前の技術的事情を引き継いで、現在まで使われているわけです。

個人的には、44,100という一見不可解な数字が、テレビの走査線数からきれいに導かれるという話は、技術史の中でも印象的な例だと思います。

あなたの発言:

なるほど。離散Fourier変換の式を扱っていて、時々 \sqrt{N} が出て来て、N=44100のときキリが良くて便利だと思っていたのですが、そういう事情でしたか。そうか、SONYは当時ビデオで頑張っていたのでしたね。

映画・テレビのフレーム周波数との関係、ということは逆にSONY が VTR の規格を作るときに、そういうことは考えなかったということなのでしょうか。

9s考えました >

ChatGPT の発言:

いえ、むしろ逆で、**ソニーはVTRとの同期を非常に強く意識していました。** ただし、48 kHzを選んだ人たちとは、少し異なる考え方をしていたと言えます。

ここは興味深いところなので、少し詳しく説明します。

1. ソニーが考えたのは「1走査線当たり何標本か」

ソニーの初期のPCM録音機は、VTRの映像信号の走査線を利用して音声データを記録しました。

そこで重要になるのは、映像の1フレーム当たり何個の標本を記録できるか、ということです。

例えば、毎秒30フレームの映像では、

$$44100/30 = 1470$$

となります。1フレームにつき1470標本、つまり1フィールドにつき735標本です。

この735を245本の有効走査線に振り分ければ、1走査線当たり3標本になります。

欧州の毎秒25フレームでも、

$$44100/25 = 1764$$

で、やはり1走査線当たり3標本とすることができます。

つまり、**44.1 kHzは、映像の走査線を音声データの記録媒体として使うのに都合がよかった**のです。

ただし、実際のNTSCカラーテレビは約29.97フレーム／秒なので、初期の機器には44.056 kHzを採用したものもありました。 World Radio… +1

2. 48 kHzの場合は、少し発想が違う

48 kHzの標準化では、音声データを映画やテレビの映像と同期させ、編集できることが重視されました。

例えば、

$$48000/24 = 2000, \qquad 48000/25 = 1920$$

ですから、映画の24フレーム／秒でも、欧州テレビの25フレーム／秒でも、1フレームの音声が常に整数個の標本で構成されます。

つまり、

- 44.1 kHz：VTRの走査線に音声データを記録しやすい。
- 48 kHz：映像と音声を時間的に対応させて編集しやすい。

という、異なる要請があったわけです。

ただし、48 kHzでも約29.97フレーム／秒では1フレーム当たりの標本数が整数になりません。この問題は当時の標準化でも検討され、5フレーム当たり8008標本とする方法が採用されました。1982年のSMPTEの論文には、この問題が詳しく述べられています。 SMPTEジャーナ...

3. ソニーにとって44.1 kHzを維持する意味

もう一つ重要なのは、CDの開発時点ですでにソニーが44.1 kHzの録音装置を実用化し、録音済みの音源も蓄積していたことです。

そのためCDの規格を決める際、44.1 kHzを採用すれば、既存のデジタル録音を標本化周波数変換なしでCDの製造に利用できました。

国立科学博物館の技術史調査報告書でも、**ソニーが当初提案していた50.4 kHzを取り下げ、既存のPCM-1600と同じ44.1 kHzを採用するに至った経緯**が説明されています。

産業技術史資料...

これは当時、かなり実際的な意味を持っていたと思います。

なお、ご指摘の離散フーリエ変換の話に戻ると、

$$\sqrt{44100} = 210$$

というのは確かに便利ですね。

ただ、DFTをFFTで計算する立場から見ると、44,100には7という素因数があり、48,000には7がありません。

$$44100 = 2^2 3^2 5^2 7^2, \qquad 48000 = 2^7 3 5^3$$

どちらも合成数なので高速なFFTを実現できますが、通常のradix-2 FFTにとってはどちらも2の冪ではありません。

このあたりも、**数学的な計算の都合より、当時の録音・放送機器の物理的な制約が優先された**ことをよく表しているように思います。

あなたの発言:

とても面白かったです。ありがとう。

ChatGPT の発言:

こちらこそ、ありがとうございました。

44,100が平方数であることから、CDの規格策定やVTRの歴史に話が広がったのは、私にも興味深い展開でした。

また気になることがあれば、ぜひお尋ねください。

あなたの発言:

そうだ。ずっと以前WAVEファイルを扱うプログラムを書いているときに、規格書が手に入らなくて、リバースエンジニアリングをしつつ、行き当たりばったりにやっていました。今 WAVE ファイルの規格を知ろうとする場合、何を見れば良いでしょうか。

22s考えました >

ChatGPT の発言:

現在なら、WAVEファイルの仕様書をかなり容易に入手できます。しかも、**1991年の原規格書までPDFで公開されています。**

昔、リバースエンジニアリングしながらプログラムを書かれたのであれば、まずその原規格書をご覧になるのが面白いと思います。

現在プログラムを書く目的なら、私のお薦めは次の順番です。

1. 最もお薦めする資料

(1) IBM・Microsoftの原規格書（1991年）

Multimedia Programming Interface and Data Specifications 1.0

1991年8月、IBMとMicrosoftが共同で発表した仕様書です。

原規格書 PDF（McGill大学の保存コピー） ↗

RIFFおよびWAVEの構造が、56〜65ページあたりに記されています。 RIFF1991

当時の仕様を知りたい場合には、これが最も直接的な資料です。

WAVEはRIFF（Resource Interchange File Format）という汎用のファイル構造の上に構築されています。したがって、WAVEだけでなくRIFFの基本構造も理解するのが近道です。

(2) McGill大学のWAVE File Specifications

Wave File Specifications — Peter Kabal, McGill University ↗

実際にプログラムを書くなら、こちらが一番便利かもしれません。

1991年の基本仕様だけでなく、1994年の拡張、マルチチャンネル音声なども整理されています。

特に便利なのは、`fmt` チャンクの構造、各フィールドのバイト数、PCMデータの表現などが表になっている点です。

最終更新は2022年9月27日で、かなり丁寧にまとめられています。 www-mmsp.ec...

(3) 米国議会図書館のWAVE形式解説

WAVE Audio File Format — Library of Congress ↗

こちらは保存・互換性の観点から書かれた資料です。

原規格書、拡張仕様、関連する標準化団体の文書へのリンクが充実しています。WAVEの仕様がどのように拡張されてきたかを調べる際に役立ちます。 The Library of ...

2. その後の拡張仕様

1991年の原規格だけでは、現在のすべてのWAVファイルを扱えるわけではありません。

仕様	内容	資料
1994年のRIFF拡張	WAVEFORMATEXなど	RIFFNEW.pdf ↗
WAVE_FORMAT_EXTENSIBLE	多チャンネル、24ビットPCMなどの拡張	McGillの解説 ↗
Broadcast Wave Format	放送用メタデータの追加	EBU Tech 3285 ↗
RF64	4 GBを超える音声ファイルなど	米国議会図書館の関連資料 ↗

特にBroadcast Wave Formatは、欧州放送連合が策定した正式な技術仕様です。 EBU Techno… +1

3. 自分で読み書きする場合の注意点

昔書かれたプログラムがどのようなものだったかは分かりませんが、WAVEファイルには、簡単そうに見えて落とし穴があります。

例えば、最も単純なPCM形式のWAVEは次の構造です。

```
RIFF
  file size
  WAVE
  fmt
    chunk size
    format code
    channels
    sample rate
    byte rate
    block align
    bits per sample
  data
    chunk size
    audio samples
```

しかし、実際には次の点が重要です。

- `fmt` の末尾には空白があり、4文字である。
- `fmt` チャンクの大きさは常に16バイトとは限らない。
- `fmt` と `data` の間に別のチャンクが入ることがある。
- チャンクのデータ長が奇数の場合、パディングが入る。
- 整数値は基本的にリトルエンディアンで格納される。
- PCMの8ビットは符号なし、16ビット以上は符号付きが基本である。

特に、**44バイト固定のヘッダーを仮定するプログラムは避けるべき**です。一般的なPCM WAVにはそれで対応できますが、拡張チャンクを含む正当なWAVEファイルを読み損ないます。

4. 私なら、こう読みます

まず1991年の原規格書を読み、RIFFの設計思想とチャンク構造を確認します。そのうえでMcGill大学の資料を参照し、1994年以降の拡張と実際のバイナリ構造を調べます。

単純なPCM音声を扱うプログラムなら、この二つでほぼ必要な情報は揃うはずです。

また、昔のプログラムと比較したいのであれば、Python標準ライブラリの `wave` モジュールや、Cで書かれた libsndfile ↗ の実装を参照するのも有益でしょう。

なお、昔の資料が入手しにくかった理由も少し分かる気がします。当初のWAVE仕様はMicrosoftとIBMのマルチメディア開発者向け文書の一部であり、独立した国際規格として出版されたものではありませんでした。

今なら、当時探しておられた情報の多くを、原典から直接確認できると思います。

あなたの発言:

> 今プログラムを書くならば、Python でプログラムを書くとして、パッケージに任せれば良いのかもしれません。
>
> 当時は、書籍に載っているプログラムもWAVEファイルによってはうまく読めなかったりして、かなりお粗末な状況でした。

ChatGPT の発言:

そうだったと思います。WAVEは一見単純な形式なので、書籍でも「ヘッダーを44バイト読み飛ばせば、あとはPCMデータ」というような説明がされがちでした。

しかし、WAVEは本来、**チャンクを順次読み取って解釈する形式**です。実際のファイルには余分なチャンクがあったり、 `fmt ` チャンクの長さが異なったりします。仕様に忠実でないプログラムが、特定のWAVEファイルしか読めないのも無理はありません。

現在のPythonなら、用途によって次のように使い分けられます。

- `wave` ：標準ライブラリ。非圧縮PCMを読み書きするなら十分なことが多いです。
- `scipy.io.wavfile` ：NumPy配列として読み込み、数値計算やFFTをする場合に便利です。
- `soundfile` ：libsndfileを利用しており、WAVE以外も含めて幅広い形式を扱えます。

例えば、現在なら次のように書くだけで、WAVEファイルを読み込んでフーリエ解析できます。

```
from scipy.io import wavfile
import numpy as np

fs, x = wavfile.read("sample.wav")

# ステレオなら片チャンネルを取り出す
if x.ndim == 2:
    x = x[:, 0]
```

```
x = x.astype(np.float64)

F = np.fft.rfft(x)
freq = np.fft.rfftfreq(len(x), d=1/fs)
```

もっとも、当時の経験には別の価値もありそうです。

WAVEのチャンク構造を理解し、バイト列を直接調べ、どこで読み込みに失敗するかを突き止めるという作業は、現在ならライブラリが隠してくれる部分です。

私が特に面白いと思うのは、**WAVEがもともと拡張可能な形式として設計されていたのに、初期の解説書やサンプルプログラムは、その拡張可能性を十分に考慮していなかった**という点です。

規格の設計思想と、それを利用するプログラムの品質は別問題だったわけですね。