

# 20kHzより彼方にある世界

原典 David E Blackmer  
抄訳 株式会社メディア・インテグレーション | 発行 2013.05

このドキュメントでは、Earthworks の創業者 David E Blackmer が、人間の聴覚メカニズムに関する研究を基礎に、ハイディフィニション・オーディオについての考察と議論を展開します。

我々はどうして超高品質な音響再生に接するべきか、そこには多くの議論が存在します。コンパクトディスク (CD) 規格においては、一般的に人間の可聴帯域を超える 20kHz 以上には有益な情報は無いものと想定し、ブリックウォール・フィルターで 20kHz 以上の帯域をカットしています。しかし現実には、多くの人が周波数帯域の上限が 20kHz に設定されたものと、20kHz 以上も含むワイドバンドのオーディオ信号を実際に聴いたとき、大きな違いを感じます。多くのデジタル・オーディオ機器がサンプル・レートは 24bit/96kHz (もしくはそれ以上) に対応していることは皆さんもご存じでしょう。

多くのエンジニアは、人間の聴覚は 20kHz 以上の周波数帯から意味のある情報を得られない、と教えられてきました。「20kHz 以上のシグナルに使い道などない」とが不要な情報を捉えること自体誤った判断を招く、馬鹿らしい、時間の無駄だ」など、エンジニアからの怒りに満ちた投書を貰ったことさえあります。「20kHz の壁」が何十年の間、絶対的であったことは、良識あるオーディオ・エンジニア全員にとっての常識です。40kHz までの帯域に意味のある重要なオーディオ情報が存在すると感じる我々のような人間は、ただの思い違いと考えられていました。



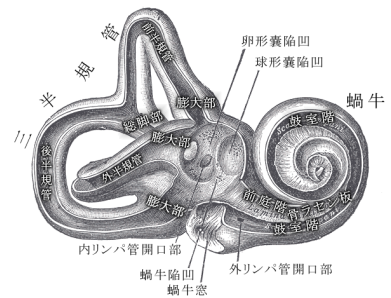
ワイドバンド・オーディオの理解には、人間の聴覚に関わるメカニズムに注目する必要があります。これらの理解を深めることで、聴力の分析・変換の素地となるモデルを作り、より進化したオーディオシステム・デザインの基準に繋がっていくのです。

そもそも、私が 20kHz を超える人間の聴覚能力とその理解についての研究を始めたのは、80 年代後半に起きたある出来事がきっかけになっています。当時、私は手に入れたばかりの MLSSA システムを使い、複数のドーム・ツイーターの音響特性を比較していました。中でも選りすぐりのユニットのグループは、現実的に言って 20kHz 付近までまったく同じ周波数特性を備えていたにも関わらず、そのサウンドは非常に異なっていました。そこで、20kHz より上の特性を計測結果のビジュアルで詳しく見てみると、そこに明らかな違いを確認することができました。メタルドーム・ツイーターには 20kHz 以上の帯域に連なる杭のような不規則なピークと凹みがありました。シルクドーム・ツイーターでは、20kHz を超えてからの特性はなめらかに下がっていたのです。

20kHz 以上は聴こえないはずなのに、メタルドームはシルクドームよりも尖ったサウンドに聞こえる。聴覚上の違いは大きく、明らかでした。自分がかっきりと自覚したその音の違いについての理由を考え始めたのです。

進化的な観点に立つと、人間の聴覚が現在のようになったのはそれが生存に欠かせない道具だったからだと考えられます。音響に対する人間の感覚は驚異的です。人間は周囲の世界について、ほんの僅かなディテールまで感じることができます。その能力により、我々の祖先は危険を察知し、食料を見つけ、コミュニケーションを図り、自然界の音を楽しみ、そして現在音楽と呼ばれる芸術を賛美するまでに至ったのです。そもそも人間の聴覚が、一般に周波数分析の仕組みだと見なされること自体が、大変な誤解だと私は考えます。広く知られる人間の聴覚モデルでは、まず聴覚というダイナミックレンジが広いコム・フィルター周波数分析システムが、各周波数コンポーネントの強弱の情報を検知し、脳に出力します。これを脳が解釈したものが音響の認知と呼ばれています。このコム・フィルターは、我々の周波数分析において間違いなく重要なパートを占めています。そしてこのフィルターの素晴らしさは筆舌に尽くしがたいものです。各周波数に対応するゾーンは、機械的なマイナスイ抵抗システムと合わせて非常に鋭くチューニングされています。しかも各周波数フィルターのチューニング Q は、脳幹付近に並んだ分析準備センター (蝸牛神経核) から蝸牛へ送られる命令に沿って調整されているのです。蝸牛神経核から伸びる毛のような細胞の出力は、驚くべき転送速度を誇る無数の神経繊維に繋がれています。人間の持つ周波数情報の処理能力は間違いなく素晴らしい。しかしながら、そこには音の高低を聞き分けるという能力だけでは説明できないものが確実にあります。

内耳は非常に微細な構造を持った複雑な器官です。空気圧の変化による音波は、内耳において神経パルスに変換されます。ここで大きな役割を担うのが蝸牛で、液体で満たされたカタツムリにも似たチューブ形状をし



ています。音波はまず鼓膜で機械的な振動に変換され、次に卵円窓という器官に送られます。最後に蝸牛を経由して鼓膜基板へ到達します。この鼓膜基板は、音響的に例えると、アクティブ・トランスミッターの役割を果たしています。鼓膜基板には内毛細胞と外毛細胞、2種類の有毛細胞が列をなしています。

内毛細胞は前述した周波数解析システムと明確に結びついています。波形フィルターからの出力を周波数情報に変換する作業に関わるのは、鼓膜基板にある 15000 の有毛細胞のうちわずか 3000 ほどです。では外毛細胞は一体何をしているのでしょうか？

12000 ある外毛細胞は、3 ないし 4 列で配置されています。その数は内毛細胞の実に 4 倍もあります！しかし、そのうち脳の神経路に接続されているのは全体のわずか 20% にすぎないのです。外毛細胞は、分散ネットワークのように神経線維で相互接続されており、この配列は

波形分析と低周波数変換の役割を果たしています。さらに、高速な反応で波形を増幅、鮮明化する機能を持つ筋繊維（アクチン）に対する司令塔の役目も担っているのです。こうした処理を経て、波形はコム・フィルター効果を生み出す鼓膜基板へ送られ、外有毛細胞がさらに情報を抽出し、分析センターであるオリブ複合体へ送信されます。この情報は最終的に脳の大脳皮質へ達し、ここで初めて音響パターンの意識的な認識が発生するのです。外有毛細胞から送られる情報は周波数よりもむしろ波形に依っており、脳内で周波数や他の情報と相互に関連しあいながら、聴覚上の認知を生み出しています。

我々が持つ聴覚分析の仕組みは、境界（著しい変化またはイベントの開始・終了）に対して途方もなく敏感に働きます。この検知プロセスの一つの例が、人間は初期音に対して大きな反応を示すのに対し、初期音に続く複雑な反響についての反応は弱くなるということです。音という複雑な信号に含まれる情報や意味、周波数バランスなどを我々が認識する際、初期音の要素が占める割合は非常に大きく、人間の聴覚システムは、ある音色に埋め込まれた突発的な情報（インパルス）に対し、明らかに敏感に反応します。これがハイエンドの世界で一般的に「空気感」と呼ばれる、あの感覚を表しているのではないかと推測されます。「質感」や「音色」もこれに近いものでしょう。いずれも、それぞれのサウンドに際立った固有のキャラクターを与える要素ですが、なんと呼ぶにせよ、インパルス情報は人間の聴覚にとって重要なパートを占めていると言えるのではないのでしょうか。

蝸牛からの出力される全ての信号は、パルス・レートおよびパルス・ポジション・モジュレート信号として神経線維上で伝えられます。これらの信号は周波数や強さ、波形、時間軸上の変化といった情報を取り出すために使われます。低域の周波数から神経パルスへの変換については、聴覚システム上で驚くべき処理方法が用いられます。有毛細胞から出力される低周波数は、そのほとんどがプラスの音響波に対応するグループ状のパルスとして伝達されます。マイナスの音波については、あっても僅かに含まれるのみなのです。さらに効果を上げるためなのか、実際の神経線維が伝達するのもプラス波形情報だけなのです。この半分の波形ピークのみが聴覚神経上に載ってくるという状況は、およそ1kHzまで続き、少なくとも5kHz付近までビジュアルで確認できます。各プラス音波パルスのグループは、波形のおよそ中心軸付近に明確な開始と終了の境界となって表れ、この鋭い境界をともなうパルス・グループの変換は、人間の聴覚の時間分解能において、大きな割合を占める非常に重要なメカニズムです。1929年、Von Bekesyは音の場所に対する人間の感受性を測定し、時間分解能は両耳間で10 $\mu$ sよりも優れている、という結果を発表しました。またNordmarkは1976年の論説において、両耳間の分解能は2 $\mu$ sを超えとも結論づけています。つまり、250Hzにおける両耳間の時間分解能はおおよそ10 $\mu$ s、1°の位相よりも細かいということになるのです！

人間の聴覚システムは、周波数とともに波形も信号分析に使用しています。高周波数帯の限界まで正確な波形、そして5 $\mu$ s～10 $\mu$ s単位まで正確なディテールを保持することは重要で、低周波数帯のディテールも同じように正確でなくてはなりません。ドラムのような低域を主にす

るサウンドは、その波形が全くそのままに再現されたとき、我々、聴く者の心を動かす大きな力を持っています。Dead Can DanceのInto the Labyrinthに収録された比類のないドラムサウンドを注意して聴いてください。このドラムサウンドは非常に低い基音、おそらく20Hz付近に成り立っているはずですが。これをサンプリングしてみると、最初のプラス波形の長さだけが、続く40Hzの波形の2倍の周期となっていました。明らかに、20Hzのうち半分の周期だけでも、サウンド全体が20Hzの基音を持つと見せかけるには十分であるということがわかります。

人間の聴覚システムは、内・外有毛細胞をフルに活用し何百もの音の要素をほぼ同時に解析します。それぞれの要素には音の発生場所、周波数、時間、強度、トランジェントといったイベント情報が含まれ、これらを基に聴覚は認識される全てのサウンドについて精細な空間マップを作り上げるのです。さらに各ソースのポジション、キャラクター、音色、音の大きさ、その他、認識される様々なラベルを貼られた音響ソースとイベント音のクオリティを判断するための情報には、音の波形、その身分証明書として埋め込まれたトランジェント情報、そしてこれは私の私見ですが、少なくとも40kHzまでの高周波数を持つ固有の要素が内包されているのでしよう。たとえその帯域のみを切り離して「聴く」ことができないとしても。

人間の聴覚認識が要求する条件を完全に満たすには、15Hzから少なくとも40kHz(80kHzという識者もいる)にわたる周波数帯域を、120dB以上のダイナミックレンジで再生できるサウンドシステムが必要になるはずですが。これはトランジェントのピークを正しく制御するために必須の条件です。さらに高周波数帯における数ミリ秒のトランジェント精度、30Hzまでの低域で1-2°という位相の精度も要求されます。どれも今日のシステムでは実現が難しいものばかりですが、我々が認識するサウンド・クオリティーの劣化とは、現在のサウンドを提供するシステムの妥協の産物だ、という認識が必要かもしれません。もちろん信号の変換器が明らかな問題の一つではありますが、ストレージやその他の電子的部品と内部結線もまた重要な要素です。

我々Earthworksが目指すゴールとは、従来の機材よりもはるかに精密なオーディオ・ツールを作りだすことです。そして常に限界を更に上へと広げようとしています。例えば我々のLAB102プリアンプの周波数特性は2Hz～100kHzものレンジを±0.1dB以内の誤差でとらえる仕様となっています。これは、興味のない人にとってはまったく必要とされないレンジの広さかもしれませんが、そのサウンドはまさに真実のままの再生と言える精度さを備えているのです。事実LABプリアンプで1dB下がるポイントは0.4Hzと1.3MHzですが、これはその精確性においてキーとなる部分ではありません。矩形波の立ち上がり要する時間が1/4ミリ秒であることを考慮すると、このインパルス・レスポンスは実質的に完璧といえます。

マイクロフォンは、オーディオ・チェーンの出発点です。空気の振動をとらえ、電気信号へ変換する。現在の多くのマイクが、この点で精度さを欠いています。私が忠実なサウンドを再現するために必要と信じている、15Hz～40kHzの全レンジで優れた周波数特性を示すことができるマ

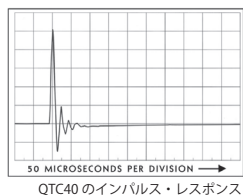
イクはごく僅かにすぎません。ほとんどのマイクにおいて、実際の動作に関連する音響デバイスは、音響波を受け取るダイアフラムですが、ドラムの打面のように、ダイアフラムは空気の振動を受けるとそれ自身も響いてしまいます。更に良くないことに、ピックアップ・カプセルは、多くの内部共鳴と反射を持ったケースに収められているため、これがさらに音への色付けの原因になります。単一指向性のマイクは、複数ポイントのサウンドを同時にサンプリングすることで指向性を得ているために、そもそもの性質として無指向性（オムニ）のマイクよりもサウンドの精確さでは劣ってしまいます。ダイアフラムの鳴り、反射、そして複数の検出経路、これらが積み重なり許容できるフェーズを超えてしまうのです。こうしたマイクは、タイム・ドメインにおいてシグナルを汚してしまっているのです。

我々は果てしない数の測定と注意深いリスニングテストを繰り返し、マイクのインパルス・レスポンスこそ、周波数や音量のレスポンス以上に音のクオリティを判断する材料として優れていることを学んできました。長く、非対称なインパルス特性を持つマイクは、より短いインパルス・テイルのマイクよりも色付けが顕著になります。これらのポイントをより明確に描くために、我々はQTC1(現行モデルのQTC40に相当)とある知名度の高い二つのオムニ・マイクモデルを用いて、慎重に様々な素材を録音しました。どちらも40kHzまで-1dB以内というフラットな周波数特



性を示しましたが、ハイクオリティなスピーカースystemでこの二つのマイクの収録音を再生したところ、それぞれのマイクは非常に違って聞こえました。さらに我々の研究室にある、

ほぼ完全なステップ/インパルス・レスポンスを備えたスピーカで再生すると、違いはより明らかになりました。ただこれらのマイクを判別するにあたって、明確に指摘できる相違点といえば、それぞれのインパルス・レスポンスだけだったのです。



QTC40のインパルス・レスポンス

我々は、こうしたインパルス・レスポンスからマイクの周波数特性を導き出すシステムを開発しました。このインパルス変換の手法で得られた多くの結果と、より一般的な手法での測定で得られたデータを比較し、我々のインパルス変換が一次標準として有効であることを確信するに至ったのです。

波形をインパルス・レスポンスとして捉えることは、高周波数帯の情報を解析する方法として、より優れています。低域の周波数については、ステップファンクション・レスポンス（インパルス・レスポンスの積分）を調べることと比較的容易に理解することができます。どちらのカーブもタイム・ウィンドウ、サンプリング・プロセスおよびノイズにより規定された、周波数と時間軸のレスポンスに関する全ての情報を内包しています。

卓越したクオリティを備えるサウンドシステムのエレクトロニクスは、同

様に優れたものでなければなりません。歪みやトランジェントの混合変調は、各増幅ステージで100万に数個程度のパーツに抑えなくてはなりません。各チェーンに多数の増幅器を持つようなシステムではなおさらです。そして非常に重要なことですが、オーディオ・アンプの内部回路のデザインにおいて、各ステージにおける信号のリファレンスポイントは、著しくノンリニアな振る舞いを持つ電源のリターン電流から切り離しておくことが重要です。アンプ内の各ステージに設けられた差動入力回路は先行するステージの信号を真の形で抽出する必要があります。アースやクロストークなど、混合的な信号の汚染を防ぐために、回路上のフィードバックはすべて出力端子を基準にし、入力端子と直接比較するべきでしょう。こうしたルールを遵守しなければ、耳障りな「トランジスタの音」に繋がります。しかし、トランジスタを適正に扱うことで、歪みや相互変調、電源ノイズなど各種エラーを極小に抑え、認知しうる限りにおいて完璧なオーディオ信号の増幅を実現できます（「認知しうる限り」とは、つまりシステム内の各要素が限りなく優秀であれば、人間の聴覚を最大限に発揮しても認識可能なエラーは発生しないということです）。現在私が目指すアンプシステムのデザイン上の目標とは、19kHzと20kHz、2波間の混変調を含む全ての高調波歪みが100万分の1パート以下、かつ正弦波におけるAウェイトド・ノイズの発生が、最大出力時の少なくとも130dB以下であること。システム内の増幅ステージをいくつ通過しても、検知可能なシグナル品質の劣化を排除することは可能ではありません。

多くの音声信号には、著しく高いトランジェント・ピークが備わっています。音量のインジケータのレベル表示をゆうに20dBほどオーバーすることも度々あります。こうしたピークを測定し正しく処理されているか判断を下すためには、オーディオ増幅システムでは適切な計測ツールを用意することが重要です。多くのピークメーターは、完全な瞬間的ピークレベルを表示するわけではありません。およそ300マイクロ秒から1ミリ秒ほどの間における、ピーク近似値の平均を表示しているに過ぎないのです。パワーアンプからスピーカに至る、システム内のあらゆるコンポーネントは元のピークを限りなく正確に再現できるようデザインされるべきです。レコーディングにおいては、そのシステムの能力を越え、扱うことのできないピークは丸め込まれます。アナログテープ・レコーダーはピークをスムーズにコンプレッションすることで、サウンドへの負の影響が少ないと見られてきました。

多くのレコーディング・エンジニアは、むしろピークのクリップを好み、意図的に使うことさえあります。デジタル・レコーディングにおいては、ブリックウォール・エフェクトを用いていかなるピークの超過をも刈り取ることで、スピーカのツイーターとリスナーに甚大な被害を及ぼしていると言えるかもしれません。コンプレッサーやリミッターが、システム上扱えないレベルのピークをなめらかに抑えるため、頻繁に使われています。これらの機



器においてはRMSレベル検知の方が、ピーク近似値を用いるよりもサウンドはなめらかになります。さらに歪みの少ないシグナル・プロセッサを慎重に選ぶことが重要です。優れたデザインがなされたユニットでは、ゲインに変化がない状態であれば歪みは非常に少なくなります。コンプレッション時に起こる歪みのほとんどは第三次倍音によるものであり、耳で検知するのはそれほど容易ではありません。気づいたとしても耳に許容できるものとして聞こえるでしょう。

たとえば、フィードバックを排し、超ハイエンドなランクとそれに見合った高評価を得ている真空管プリアンプのスペックを見てみると、どの程度の歪みが許容範囲とされ、むしろその歪みが、オーディオに過剰なほど投資する裕福なマニアに望まれているかがわかるでしょう。

電気系統における様々なパーツ同士の接続は、ノイズやシグナルエラーを排除するよう完全にデザインされなければなりません。電源やアースに起因する電流、ACからの電磁波、無線やラジオ、クロストーク、絶縁に使われる被膜などが、こうしたノイズに直接関わってきます。

電気的な信号を再度空気振動に変換するラウドスピーカーは、オーディオシステムの終端にあります。多くの場合、スピーカーの精度はマイクのそれに比べても劣ることが多いと言えます。これまで挙げてきた要件を満たすスピーカーを制作するには、様々な問題を解決する必要があります。理想的なスピーカーとは点音源です。今のところ、15Hzから40kHzの全レンジを精確に再現できるシングルドライバは存在せず、複数のドライバをつかうスピーカーシステムでは、何らかのトレードオフや妥協が必要なのが現実です。

我々も、自らEarthworksマイクロフォンと同様のタイム・ドメインの原則を適用し、いくつかの実験的なスピーカーシステムを製作したことがあり、結果も非常に期待できるものでした。インパルスやステップ・ファンクション・レスポンスが完璧に近づいていくにつれ、魔法のようなことが起きたのです。現実と見紛うようなサウンド・クオリティーが得られました。SR71マイクとこれらのスピーカーシステムで、ジャズ・ライブの演奏のSRを行ったところ、音量が増幅されても質感は全く変化せず、オーディエンス側で聴いても、アンプを通した音には感じられませんでした。もちろん音が大きくなっていることは間違いありません。さらに、相当ゲインを上げてみても、それはスピーカーから出ている音には感じられなかったのです。

QTC1(現行モデルのQTC40に相当)を使って96kHzでレコーディング

されたバッハの合唱曲を、我々のエンジニアリングによるスピーカーを通して聴いてもらうことができれば、ディテールとイメージ感の凄まじさを体験することができるでしょう。左右だけではなく、前後そして上下まで、まるで演奏者と同じ空間にいると錯覚するほどです。ゴールに向かってここまで突き進むことができた、それ自体がエキサイティングな発見でした。

1920年代にあったVictor Talking Machine Company(ビクター)の広告について、聞いたことがあります。Enrico Carusoが蓄音機を評して、あまりに音がよく自分が実際に歌う声と区別ができない、と。Acoustic Researchは70年代に、より説得力はあったにせよ、弦楽四重奏のライブ演奏とレコード素材で、似たような広告を打ちました。果てしなく長い道のりの末、我々は完璧な認知を手に入れたのでしょうか?基準をはっきりさせるためには、優れたインパルス/ステップ・レスポンスを持ったマイクとスピーカーを用いたテストを行うべきでしょう。ほとんど完璧な周波数特性と、歪みの極めて少ない増幅ステージ。音楽やボーカルなど様々な素材で、SRからスタジオモニターのシステムでテストすれば、演奏者も聴者も、その結果に驚くだろうことは間違いありません。完璧なシステムは存在しない、そんなテストは不可能だと言われるかもしれませんが、そんなことはありません、我々はすでに何度も実行してきたのです。

最後に、こうした事についてより理解を進めたいという人へ、私がオーディオに深く関わるすべての人に、何度でも読んで欲しいと思う参考文献を紹介して、本論文を結びたいと思います。

An Introduction to the Physiology of Hearing, 2nd edition, James O. Pickles, Academic Press 1988 ISBN 0-12-554753-6 or ISBN 0-12-554754-4 pbk.

Special Hearing, revised edition, Jen Blauert, MIT Press 1997 ISBN 0-262-02413-6

Experiments in Hearing, Georg von Békésy, Acoustical Society of America ISBN 0-88318-630-6

Hearing, Gulick et al, Oxford University Press, 1989 ISBN 0-19-50307-3

