スマートスピーカー
音響測定

APPLICATION NOTE

はじめに

スマートスピーカーは、オーディオパフォーマンスのテストを困難にする独特の特性を備えた比較的新しいクラスの民生用オーディオデバイスです。このアプリケーション ノートでは、デバイスのオーディオ品質の最も重要な客観的測定である周波数応答に焦点を当てた、スマートスピーカーの音響測定の概要を説明します。

背景

スマート スピーカーは、音声を使用してインテリジェント仮想アシスタント (IVA) と対話できるようにする、インターネットに接続された (通常はワイヤレス) マイク内蔵のパワード スピーカーです。コマンド・音声のみを使用して、インターネットまたは接続されたデバイスからオーディオ コンテンツ (ニュース、音楽、ポッドキャストなど) を再生したり、ホーム オートメーション デバイスを制御したり、接続されたオンラインショッピング サービスから商品を注文したりするなどのタスクを実行するように指示できます。Amazon は、Alexa として知られる IVA を搭載した Echo と呼ばれるスマート スピーカーをリリースした最初の大手企業であり、依然として圧倒的な市場シェアを保っています。この分野への他の主要な参入企業には、Alphabet (Google アシスタント⇔ Google Home ⇔ スマート スピーカー)、Apple (HomePod mini スピーカーの Siri)、Microsoft (サードパーティ スピーカーの Cortana)、Samsung (Samsung Galaxy Home スピーカーの Bixby) およびいくつかの企業が含まれます。中国、日本、韓国からの参加者もいます。

スマートスピーカーは、最初に導入された当初はささやかな始まりでしたが、すぐに人気が急上昇しました。初代 Amazon Echo のリリース直後の 2014 年 11 月、人気のあるテクノロジーレビュー サイトはこのイベントについて次のように述べました。「音声を認識して質問に答える Siri のような機能を備えたワイヤレススピーカーです。」しかしそれ以来、スマートスピーカーの普及は爆発的に進み、力強い成長は少なくとも今後 5 年間は続くと予測されています。2014 年の導入から 3 年以内に、米国だけで家庭に 6,700 万台のスマートスピーカーが設置され、その数は 2018 年には 78% 増加して 1 億 1,800 万台に達しました。世界のスマートスピーカー市場は 2017 年に 44 億米ドルと評価され、2025 年までに 233 億米ドルに達すると予測されています。IVA を持つ数社が他のメーカーにこの技術をライセンス供与しています。たとえば、Bose と SONOS は、Alexa 対応および Google アシスタント対応のスマート スピーカーを提供しています。そして、「スマート」になっているのはスピーカーだけではありません。IVA テクノロジーは、それをサポートするマイクとスピーカーを備え、冷蔵庫、電子レンジ、セットトップ ボックスなどのあらゆる種類のデバイスに追加され、これらのデバイスの音声制御を可能にします。さらに、ほとんどのスマートフォンはスマート スピーカーとしても機能します。

スマートスピーカーIVA

スマート スピーカーとの対話は、特定の「ウェイク ワード」またはフレーズ（たとえば、Amazon の場合は「Alexa」、Apple の場合は「Hey Siri」など）で始まり、その後にコマンドが続きます。通常の動作モードでは、スマート スピーカーは半休止状態にありますが、常にウェイク ワードを「リッスン」しており、音声コマンドの取得と処理をトリガーします。

音声認識に関しては、スマート スピーカー自体はウェイク ワード（またはフレーズ）のみを認識できます。より計算量の多い音声認識とその後の処理は、接続されたサーバー上の IVA によって実行されます。IVA はユーザーの音声をテキストに変換し、コマンドの解釈を試みます。要求された応答をデバイスから呼び出すには、音声コマンドに IVA が認識できる一連のキーワードが含まれている必要があります。インタラクションが成功すると、要求されたアクションが IVA（例:「タイマーを 10 分間セットする」）または接続された Web サービス（例:「インターネット ラジオ局を再生する」）によって実行される場合があります。

オーディオサブシステム

スマートスピーカーは以下の複数の個別のオーディオサブシステムを搭載しています。

- マイクロホンアレイ
- アクティブパワードラウドスピーカーシステム
- ビームフォーミング、音響エコーキャンセル、ノイズ抑制などのタスクのためのフロントエンド信号処理アルゴリズム

単一のマイクの代わりにマイクのアレイが使用され、デバイスがビームフォーミングを利用できるようになります。ビームフォーミングは、処理のために IVA に送信される音声信号の信号対雑音比を効果的に高めることができる信号処理技術です。アレイ内の異なるマイクで受信した信号間の相関に基づいて、ビームフォーミング アルゴリズムは室内で話者の最も可能性の高い方向を検出し、信号を減衰させる方法でさまざまなマイク信号を組み合わせることで、ある意味その方向に焦点を合わせることができます。他の方向から来ています。これにより、IVA に送信される音声信号内の周囲の騒音と室内の残響のレベルを効果的に低減できます。ノイズ抑制は、音声以外の信号のレベルを低減するために使用することもできます。

理想的には、スマート スピーカーは、部屋で音楽やスピーチなどのオーディオ コンテンツを再生しているときでも、(最初にウェイク ワードを認識することによって) 音声コマンドに応答できるようになります。音響エコー キャンセル (AEC) は、このタスクのためにスピーカー出力がマイク入力を完全にマスクしてしまうのを防ぐために不可欠です。ラウドスピーカー システムで再生されている信号は、AEC アルゴリズムの基準信号として使用でき、再生中のコンテンツを無視してウェイク ワードを認識できるようになります。通常、コマンド認識を向上させるために、ウェイク ワードが検出された後、再生は一時停止されます。

オーディオシグナルパス

スマートスピーカーの主なオーディオパスは、デバイスと IVA、または Wi-Fi または有線接続でインターネットを使用するネットワークサーバーの間です。入力側では、音声コマンドを含む音声信号がデバイスのマイクアレイで感知され、デジタル化された後、信号処理とコマンド解釈のために IVA にアップロードされます。出力側では、デジタルオーディオコンテンツが Web サーバーからデバイスに送信され、そこでデジタルからアナログに変換され、最終的にデバイスのスピーカーシステムで再生されるときに音響信号に変換されます。

I 上記の 2 つの主要なパスに加えて、スマートスピーカーには次のような他のいくつかのオーディオパスがある場合があります。

- 外部パワードスピーカーシステムに接続するためのアナログ出力ジャック
- スマートスピーカーを簡易パワードスピーカーとして使用するためのアナログ入力端子
- Bluetooth 接続により、外部 Bluetooth スピーカーでオーディオ コンテンツを再生したり、スマートフォンやタブレットから音楽ソースとしてコンテンツをストリーミングしたり、場合によっては通話用のハンズフリー デバイスとして機能したりすることができます。
- 他のスマートスピーカーへのネットワーク接続によるマルチルームミュージック、ステレオペアリング、またはインターコム機能
- ホームオートメーションデバイスへの接続 (セキュリティデバイスへの双方向インターコム接続や可聴ステータスメッセージなど)

オーディオテスト

スマートスピーカーのオーディオサブシステムには、マイクとマイクアレイ、A/D および D/A コンバーター、パワーアンプ、スピーカードライバー、デジタルシグナルプロセッサ、オーディオコーデックなど、全体的なパフォーマンスと音質に貢献する多数のコンポーネントが含まれています。さらに、ビームフォーミング、エコーキャンセル、ウェイクワード認識などのいくつかのシステムレベルの機能が全体の品質に貢献します。ある段階で、これらのコンポーネントとシステムをそれぞれテストする必要があります。スマート スピーカーシステム全体のエンドツーエンドのパフォーマンスをテストすることも望ましいです。

研究開発、検証、製造テスト、品質保証といったテストのコンテキストが異なれば、目標も異なり、サブシステムやコンポーネントへのアクセスのレベルも異なります。たとえば、製品設計中に、R&D エンジニアはチップ レベルの接続を物理的に利用することで、システム オン チップ (SOC) のアクティブ クロスオーバー機能を分離できる可能性があります (また、結果として生じる信号)。同様に、製造テストの場合、メーカーは、既製のユニットでは利用できない機能テストを可能にするために、特別なテスト固有のファームウェアをデバイスに一時的にロードするオプションを持っています。たとえば、ノイズ リダクションを無効にして、マイク入力システムを音声の代わりにサイン波信号でテストできるようにすることができます。

スマートスピーカーのプライマリ入力および出力オーディオパスの全体的なエンドツーエンドパフォーマンスをテストすることは、次の理由により非常に困難な場合があります。

1. スマートスピーカーへの入力とスマートスピーカーからの出力はとも音響的なものであり、音響テストはその性質上、電子 (アナログまたはデジタル) オーディオテストよりも複雑です。音響テストには、校正済みのマイク (通常は無響性テスト チャンバー) と、DUT マイクを刺激するための高品質のスピーカーシステムが必要です。

2. スマートスピーカーは本質的にオープンループデバイスです。入力側では、信号 (通常は音声) がキャプチャされてデジタル化され、デジタルオーディオファイルとしてどこかのサーバーに送信されます。入力パスのパフォーマンスを評価するには、オーディオファイルをサーバーから取得し、最初に生成された信号と比較して分析する必要があります。出力側では、サーバー上のオーディオファイルとして生成されたオーディオコンテンツがデバイスにストリーミングされ、そこで変換されます。アナログであり、デバイスのスピーカーシステムで再生されます。出力パスのパフォーマンスを評価するには、デバイスのスピーカー出力を測定マイクで測定し、サーバーからの元の信号と比較する必要があります。元の信号はエンコードされたオーディオ信号 (MP3 や AAC など) の形式であることが多く、分析前にデコードする必要があります。

3. デバイス内の A/D コンバータおよび D/A コンバータには、常に オーディオアナライザーとはサンプルレートが異なるため、分析中に何らかの形での補正が必要になります。

周波数応答測定

オーディオデバイスのパフォーマンスの最も重要な側面は、その周波数応答です。周波数応答は「伝達関数」測定の一種です。被試験デバイス (DUT) の場合、単位入力あたりの DUT からの出力の大きさと位相を周波数の関数として表します。多くの場合、デバイスは周波数応答曲線の「形状」の観点から比較されます。これは通常、振幅応答のみ (位相ではなく) を参照し、さらに振幅を基準値に正規化します。たとえば、応答の大きさは、正規化された曲線が 1 kHz で 0 dB を通過するように、ある基準周波数 (たとえば 1 kHz) での値に正規化される場合があります。通常、オーディオシステムでは、ソース素材がスペクトル色付けなしに忠実に記録および再生されることを保証するために、平坦な周波数応答 (周波数に対する応答の大きさが一定) であることが望ましいです。平坦な周波数応答は、電子オーディオ システムではかなり実現可能ですが、音響機器、特にラウドスピーカーでははるかに困難です。

スピーカーの場合、周波数応答は、もう 1 つの重要な指標である感度を導き出すための基礎でもあります。感度は、単位入力あたりのデバイスからの出力の尺度です。たとえば、「94 dBSPL、1 W/1 m」という仕様のスピーカーシステムは、1 W の入力に対して、1 m の距離で 94 dB の音圧レベルを出力することを示します。スピーカーの場合、通常、スピーカーは平坦な周波数応答を持たないため、感度は周波数範囲にわたる周波数応答の大きさを平均することによって計算されます。

周波数応答は非常に重要なオーディオ品質の指標であるため、オーディオアナライザーには周波数応答を測定するためのいくつかの異なる方法があります。たとえば、スマートスピーカーのテストに考慮できる、Audio Precision オーディオアナライザーで利用可能な測定値には、次のようなものがあります。

- Stepped sine sweep
- Logarithmically-swept sine (chirp)
- Multitone
- Transfer function

スマート スピーカーの主な入力パスと出力パスの測定に関して、各測定手法には特定の利点と欠点があります。

Stepped Sine Sweep

Stepped Sine Sweep は、オーディオアナライザーが存在する前でさえ、オーディオテストの初期の時代から使用されてきた周波数応答を測定する古典的な手段です。これには、対象の周波数範囲（通常は 20 Hz ～ 20 kHz と考えられる可聴周波数範囲の一部）内の一連の離散周波数ステップを「スイープ」することによって DUT をテストすることが含まれます。各周波数ステップで、デバイスはサイン波で刺激され、その出力が分析されてレベルや高調波歪みなどの指標が決定されます。

一般に、他の技術に対する Stepped Sine テストの利点の 1 つは、その長い使用の歴史により、測定の完全性に疑問がある場合に他の技術と比較できる一種の「ゴールドスタンダード」とみなされることが多いことです。。もう 1 つの利点は、全高調波歪み（THD）やチャンネル間位相などの他のオーディオ品質指標を周波数応答と同時に測定できることです。

Stepped Sine テストの一般的な欠点はテスト時間です。多くの場合、刺激周波数が急激に変化するとデバイスは過渡応答を起こし、デバイスが定常状態の動作に落ち着くまでの各ステップでより長い時間が必要になります。関連するもう 1 つの欠点は、周波数分解能が低いことです。周波数ステップは離散的であるため、解像度が高くなるとより多くのステップが必要となり、テスト時間が増加します。

音響デバイスの場合、Stepped Sine テストの最大の欠点の 1 つは、反射のないテスト環境（通常は無響室）が必要であることです。反射面のある通常の部屋は使用できません。反射された音波が対象の直接音波と（建設的または破壊的に）結合し、重大な測定誤差を引き起こすためです。

スマート スピーカーの入力パスをテストする場合、Stepped Sine テストにはさらに 2 つの欠点があります。

1. スマートスピーカーは、音声信号をキャプチャして処理するように設計されています。したがって、サイン波信号を減衰するように設計されたデジタル信号処理（DSP）が搭載される可能性が非常に高くなります。この機能をテストのために無効にできない限り、階段状サイン波信号を使用してこのパスを効果的にテストすることは難しい場合があります。
2. エンドツーエンド テストを実施する場合、音声コマンドの処理を試行するときにデバイスが記録する信号の長さには制限があります（たとえば、Alexa IVA サービスの場合、キーワードを含めて約 7 秒）。この場合、Stepped Sine 刺激は約 6 秒未満である必要があります。
   ステップサインテストは、対象の周波数範囲内でテスト環境が無響であれば、スマート スピーカーの出力パスのテストに適しています。

Logarithmically Swept Sine
対数掃引サイン技術（チャープまたは連続掃引とも呼ばれる）は 2000 年に導入されました。これは、周波数が短時間（数分の 1 秒から数秒）で低周波から高周波まで対数的に連続的に掃引される一種のサイン波刺激を使用します。対数掃引サイン波テストは、短いテスト時間で THD や高調波歪みの個々の成分など、多くのオーディオ品質指標を導き出すことができるため、オーディオ テストで非常に一般的です。

Logarithmically Swept Sine テストには、音響測定に関してさらにいくつかの利点があります。

1. 周波数応答はインパルス応答から得られるため、解析では「時間選択的」または「準無響」技術として知られる技術を利用できます。周波数応答を計算する前に、室内反射の影響を除去するためにインパルス応答の終わりにウィンドウ処理が行われます。これにより、低周波数での測定の精度が低下しますが、慎重に使用すれば、高価な電波暗室を必要とせずに、音響デバイスの中高周波数範囲の応答を正確に測定できるようになります。
2. Audio Precision の APx500 オーディオ アナライザーでは、音響テスト用に設計された Logarithmically Swept Sine 測定（「音響応答」として知られています）を使用して、スピーカーの Rub & Buzz の欠陥を検出できます。

対数掃引サイン波技術を使用して、スマートスピーカーの入力パスをテストできます。ただし、信号は依然としてサイン波の一種であるため、純音をブロックする DSP を無効にできないデバイスをテストする場合は注意が必要です。たとえば、短い掃引を使用する必要がある場合、システムにはサイン波信号が存在するという事実を認識して反応する時間が十分にない場合があります。

準無響機能と、Rub & Buzz を検出するオプションの機能を備えた対数掃引サイン波は、スマート スピーカーの出力パスをテストするための理想的な刺激となります。

Multitone
名前が示すように、Multitone 信号は、個別の周波数の一連のサイン波（トーン）が結合されて構成されます。各トーンの位相は、多くの場合、信号のクレストファクター（ピークと RMS レベルの比）を低減するために調整されます。周波数応答を測定するには、通常、トーンの周波数が対数的に配置されます。AP オーディオ アナライザーでは、DSP 技術を使用したマルチトーン分析測定により、周波数応答と同時に歪みとノイズを測定できます。Multitone 解析のユニークな機能の 1 つは、信号の存在下でノイズを測定できることです。

Multitone 解析のメリットは以下の通りです。

1. Multitone テストは高速であり、1 回の測定で多数のオーディオ品質メトリクスを提供します。
2. 各 Multitone には固有の「シグネチャ」があり、ノイズやその他の信号が存在する場合でも、アナライザーが信号を正確にトリガーできるようになります。
3. 適切な数のトーンが含まれる場合、信号は純粋なトーンとは十分に異なるため、ノイズ抑制アルゴリズムによってブロックされる可能性は低くなります。

Multitone 解析には何点か欠点もあります。

1. すべての周波数が同時に刺激されるため、各周波数での信号エネルギーレベルは、純粋なサイン信号またはチャープ信号の場合よりも低くなります。言い換えれば、信号の波高率はサイン波よりもはるかに高いということです。
2. ステップサインと同様に、Multitone の周波数分解能は通常限られており、音響テストには無響室が必要です。

無響室などの自由音場環境では、Multitone 解析は、スマートスピーカーのメイン入力パスと出力パスの両方をテストするのに適しています。1 つの注意点は、ノイズ キャンセルアルゴリズムが信号をブロックしないように、入力側で十分なトーンを使用する必要があることです。出力側では、IVA の観点からは、Multitone 信号はデバイスが再生する単なる音楽トラックにすぎません。

Transfer Function

Transfer function 解析は、「デュアル チャネル FFT 解析」または「動的信号解析」とも呼ばれ、広帯域* 信号（ノイズ、音楽、音声など）で DUT を刺激し、DUT から出力を取得することを伴います。次に、複素離散フーリエ変換（DFT）として知られる数学的手法を使用して、入力信号と出力信号から伝達関数（周波数応答）が導出されます。この文脈における「複素数」という用語は、周波数応答に振幅と位相の両方が含まれるという事実を指します。分析の副産物は、コヒーレンス関数として知られる結果です。この関数は、出力信号が入力とどの程度コヒーレントであるか（つまり、入力に関連している、または入力によって引き起こされる）、各周波数で 0 から 1 までの値を持つ関数です。

ほとんどのアナライザーでは、アナライザー入力チャネルの 1 つを使用して、DUT への入力とその出力を測定する必要があります。AP の APx500 オーディオ アナライザーでは、ジェネレータ信号またはディスク上のファイルから刺激信号を取得できるため、追加の測定のために入力チャンネルが解放されます。APx 測定には、オープンループ測定に特に役立つもう 1 つの機能があります。それは、相互相関に基づく信号マッチング アルゴリズムを使用して、対象の信号でトリガーする機能です。

他の周波数応答測定手法に対する Transfer Function 解析の主な利点は、あらゆる広帯域信号を使用できることです。これにより、音声または音声に似た信号を刺激として使用できるため、スマートスピーカーの入力パスをテストするのに最適な選択肢となり、信号がノイズ キャンセリングアルゴリズムによって変更されることなく取得および処理されることが実質的に保証されます。Transfer Function 解析は、音楽やノイズを刺激としてスマートスピーカーの出力経路にも利用できます。スマート スピーカーの入力パスと出力パスの両方をテストする場合も、反射を避けるために無響室を使用する必要があります。

エンドツーエンドのスマートスピーカー測定に関する考慮事項

次に、スマートスピーカーのエンドツーエンドテストの実際的な側面をいくつか検討します。2 つの主要なオーディオパスはまったく異なるため、個別に説明します。

オーディオ テストを実施する場合、一般に業界標準に準拠するか、少なくともガイドラインとして使用することをお勧めします。スマート スピーカーは非常に新しいため、現在、スマート スピーカーをテストするための業界標準はありません。ただし、形状と機能の点では、スマート スピーカーはスピーカーフォンに似ており、IEEE 1329 など、スピーカーフォンのテストに焦点を当てた国内および国際規格がいくつかあります。これらの規格は、スマート スピーカーをテストする際のガイドラインとして使用できます。

スマートスピーカー向け入力パス

スマート スピーカーの主な入力パスでは、音声 (または音声に似た) 信号がデバイスのマイク アレイで感知され、デジタル化されて IVA にアップロードされ、信号処理とコマンド解釈が行われます。このパスのオーディオ テストには、デバイスを「トリック」してテスト信号を取得して保存し、記録された信号をバックエンド サーバーから取得して、元のテスト信号と比較することが含まれます。

物理的なテスト設定

図 1 は、IEEE 標準 1329 に基づくテスト設定を示しています。この規格では、スピーカーフォンを無響室または模擬自由音場(時間選択測定技術を使用)内で幅約 1 m、長さ 1 m (39 インチ x 39 インチ) の卓上でテストすることが指定されています。送信方向 (スマート スピーカーの入力パスに相当) については、デバイスはテーブルの端から 40 cm (15.7 インチ) の位置に配置され、マウスシミュレーターはテーブルの端と 30 cm (11.8 インチ) テーブルの上に配置されます。図 1 では、マウスシミュレーターはテーブルの表面ではなく、スマート スピーカーの上部から 30 cm (11.8 インチ) の位置にあります。これは、スマートスピーカーのマイクアレイは、通常、スピーカーフォンとは対照的に、デバイスの上面に配置されているためです。通常、デバイスのベースにマイクが配置されています。

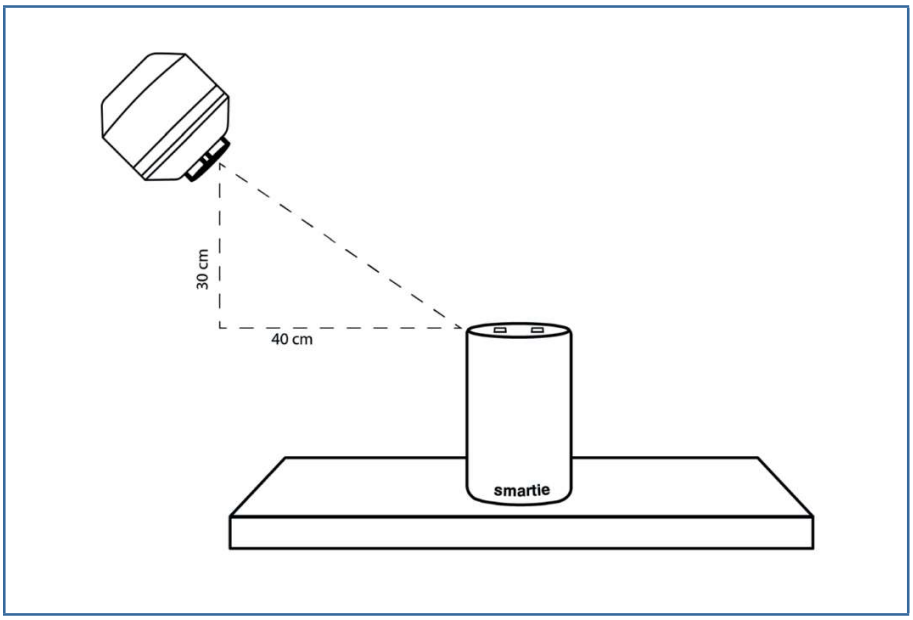

図 1.
スマートスピーカー入力パステストの設定

マウスシミュレーターの校正
他のスピーカーと同様に、マウスシミュレーターの周波数応答は平坦ではありません。一般的な口シミュレータは、100 Hz ～ 8 kHz の周波数範囲（男性の音声の周波数範囲）内で ±10 dB 以上の平坦性からの偏差があります。口シミュレータが音声などのオーディオ信号を忠実に再現するには、この範囲内でフラットな応答を持つようにイコライズする必要があります。マウスシミュレーターをイコライズ（およびキャリブレーション）するための測定は、通常、口の開口部の中心でリップリングの 25 mm（1 インチ）前に位置する口基準点（MRP）と呼ばれる点で 1/4 インチの測定マイクを使用して行われます。。マウスシミュレーターをイコライズするには、その周波数応答を測定し、反転して EQ カーブとして適用します（オーディオ アナライザーに組み込まれた機能）。これにより、口シミュレータの周波数応答の大きさは、指定された許容誤差内で平坦になります（例: 100 Hz ～ 8 kHz で ±0.5 dB）。

テスト信号も既知のレベルで提示される必要があります。IEEE 1329 では、MRP での −5 dBPa（89 dBSPL）のレベルを指定しています。これは、音声の通常または公称レベルに対応します。音声信号が使用される場合、レベルは「アクティブ音声レベル」を使用して設定する必要があります。これは、音声フレーズと文の間の無音期間を無視するレベルメトリックです。スマートスピーカーの入力パスには、自動ゲイン制御（AGC）などの非線形処理が行われる可能性があります。したがって、いくつかの異なるレベルでテストを繰り返すことが賢明かもしれません。スピーカーフォンの音量評価については、IEEE 1329 では、MRP で 79 ～ 99 dBSPL のレベルで 5 dB ステップでのテストを指定しています。

テスト信号サンプルレート
分析フェーズでは、スマート スピーカーによって取得され、IVA にアップロードされた信号が刺激信号と比較されます。これには、2 つの信号のオーディオサンプルレートが同じである必要があります。スマート スピーカーは通常、入力パスとして 16 kHz のサンプル レートを使用し、8 kHz 未満の帯域幅を実現します。このサンプルレートは、Voice over Internet Protocol (VoIP) や新しいバージョンの Bluetooth Hands-free Profile (HFP) などの「広帯域音声」アプリケーションで使用されます。通常のデジタル電話回線 (4 kHz 弱) の 2 倍の帯域幅があり、より明瞭で自然な音声を可能にするため、広帯域と呼ばれます。

テスト スティミュラスがまだ存在しない場合は、DUT 出力と同じサンプルレート (通常は 16 kHz) でテストスティミュラスを作成するのが合理的です。オーディオアナライザーは通常、さまざまなサンプル レートでスティミュラス信号を作成できます。スティミュラス信号がすでに存在するものの、異なるサンプルレートである場合は、ある時点で DUT 出力のサンプルレートに変換する必要があります。これは、Audacity (オープンソース) や GoldWave (安価) などのオーディオ波形編集ソフトウェアパッケージを使用して簡単に実現できます。

ウエイクワードの利用
入力パスをテストするときは、ウェイク ワードを使用してスマート スピーカーをアクティブにし、音声コマンドを予測して数秒間の音声を録音する必要があります。1 つのオプションは、テスト オペレーターがデバイスにコマンドを与えるかのようにウェイク ワードを発声し、すぐにマウスシミュレーターからテスト信号を生成することです。これは機能しますが、ウェイク ワードとテスト信号の間のタイミングはテストごとに異なる場合があります。

もう 1 つのオプションは、人のウェイク ワードの発言を録音し、音声波形エディタを使用してそれを刺激信号の先頭に追加する (つまり、刺激を含む音声ファイルの先頭に挿入する) ことです。この場合、テストを実施するには、刺激を伴うウェイクワードを必要なレベルでマウスシミュレーターを通して単に再生するだけです。
この方法では、ウェイクワードと刺激の間の時間はテストごとに一定です。
このオプションは、DUT がテスト チャンバー内にあり、オペレーターとオーディオアナライザーがテスト チャンバーの外にある場合にも機能します。

DUT 出力信号の取得と解析
ウェイクワードとそれに続く音声刺激信号でスマート スピーカーを起動するプロセスにより、DUT が数秒間の音声を録音し、音声認識処理のために IVA にアップロードします。 分析を完了するには、この録音されたオーディオ ファイルをバックエンドサーバーから取得し、それを .WAV 形式に変換し、オーディオアナライザー制御ソフトウェアで分析する必要があります。 このプロセスの詳細は、使用される IVA によって異なります。 たとえば、人気のある Alexa サービスには、ユーザーが自分のアカウントにログインして、接続されたスマート デバイスとのやり取りを管理できる Web ポータルがあります。

[履歴] の下の [設定] メニューには、日付/時刻スタンプ、解釈されたコマンドの転写、PC のスピーカーで録音されたオーディオを再生するツールなど、接続されたスマートデバイスとの各対話の記録があります。Web ブラウザに組み込まれている Web 開発者モード機能を使用すると、記録された .WAV ファイルを直接ダウンロードできるハイパーリンクを取得できます。その後、オーディオアナライザーソフトウェアで開いて分析できます。

記録された .WAV ファイルを IVA のサーバーから取得し、刺激信号に関して分析することで得られた周波数応答の大きさを図 3 に示します。この場合、マウスシミュレーターは、MRP で測定された rms レベルが 89 dBSPL となるようなレベルで駆動されました。

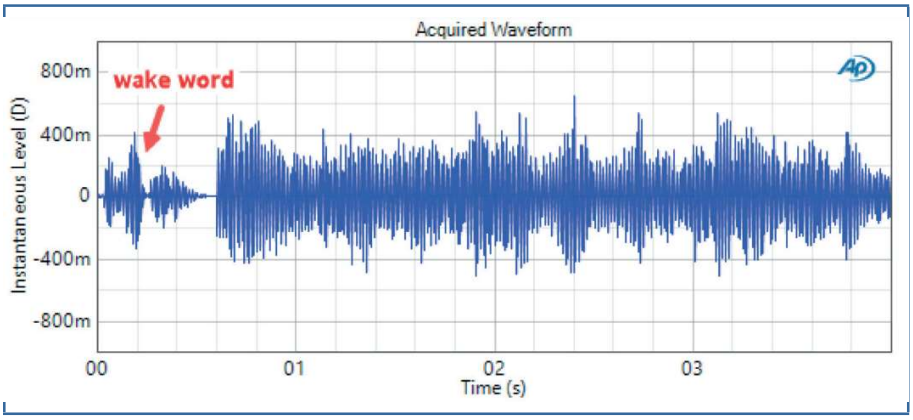

図 2.
刺激波形の最初の部分 (ウェイクワードが先頭に追加された会話の音声)。

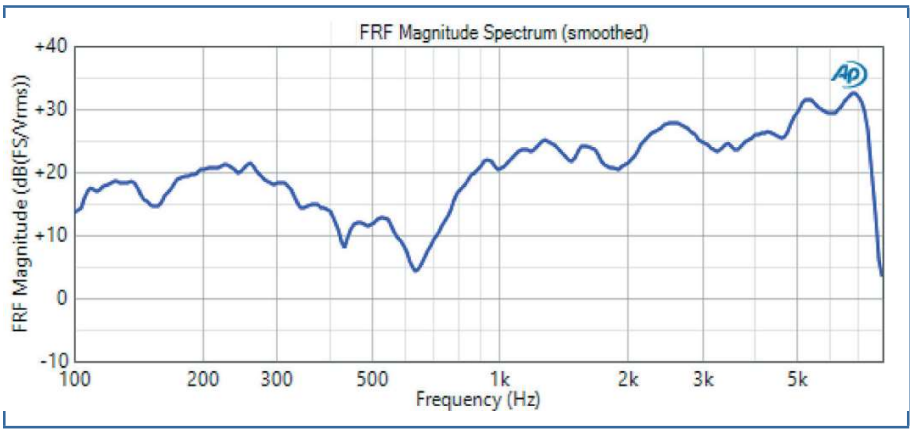

図 3.
図2の刺激信号を使ってスマートスピーカーの周波数応答振幅の測定データ

例 – スマートスピーカー入力パスの対数掃引サイン波測定 この例では、対数掃引サイン波刺激を使用した同じスマートスピーカーの入力パスのテストを取り上げます。図 4 は、パイロット トーンとウェイク ワードがチャープ信号の前に挿入された、50 Hz から 8 kHz までの 0.35 秒の長さのスイープのスティミュラス波形を示しています。アナライザーはパイロット トーンでトリガし、パイロット トーン周波数を使用して DUT とアナライザーの間のわずかなサンプル クロックの差を補正します。この場合、ウェイク ワードとパイロット トーンは周波数応答の分析に含まれません。

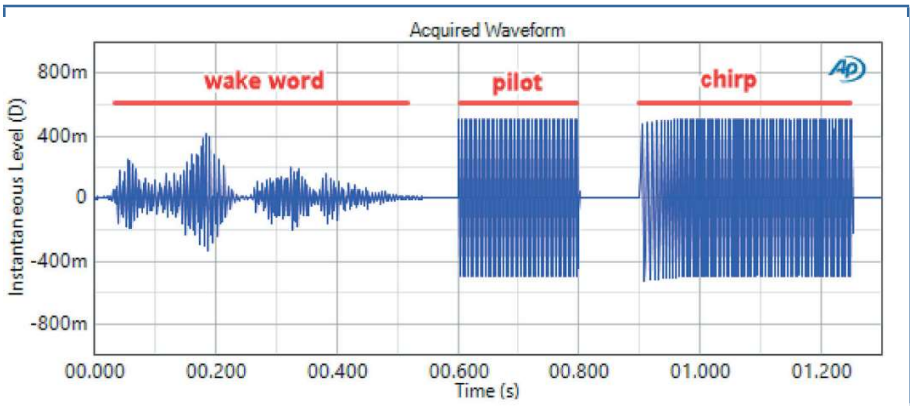

図 4.
スマートスピーカーの対数掃行サイン波テストの刺激信号波形（ウエイクワード、200ms パイロットトーン、350ms チャープ掃行、周波数 50Hz–8kHz）

図 5 は、0.35、1.0、および 4.0 秒のスイープ長に対するスマート スピーカーの入力パスの音響レベル応答を示しています。

1. T 上記のチャープ解析と音声信号の伝達関数測定からの応答曲線は、形状が非常に似ています。
2. チャープ刺激に対する応答は、スイープの長さによってあまり変化しません。

私たちがテストした 2 番目のスマートスピーカーは動作が異なります。音声信号で測定された応答は、チャープで測定された場合とはまったく異なり、チャープ応答は掃引長によって大幅に変化します。これは、2 番目のスピーカーがこの例でテストしたものよりも積極的なノイズ抑制アルゴリズムを備えていることが原因である可能性が高くなります。

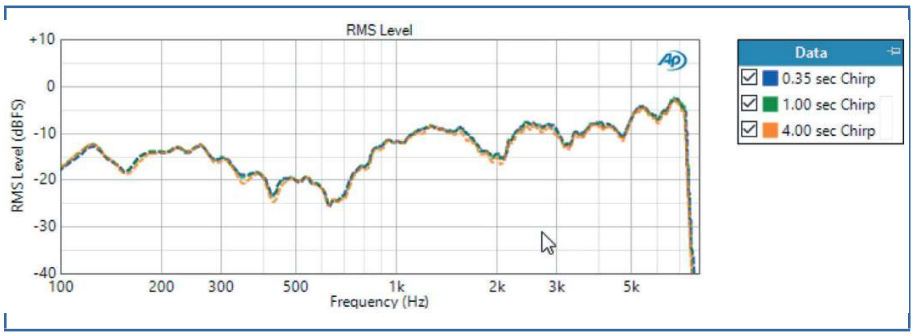

図 5.
0.35 秒から 4.0 秒まで長さが変化するチャープ信号によるスマート スピーカーの入力パス レベル応答。

例 – スマート スピーカー入力パスのマルチトーン測定

上記の最初のスマート スピーカーの入力パスは、100 Hz から 8 kHz までの標準的な 1/6 オキャットブ周波数で対数的に間隔をあけたトーンで構成されるマルチトーンで再度測定されました。信号全体の長さ（最低周波数トーンの周期）は 200 ミリ秒です。スマートスピーカーを起動するために、ウェイクワードが再び刺激信号の前に付加されました。マルチトーン測定の場合、測定の高度に選択的なトリガーメカニズムを使用して信号自体をトリガーし、ウェイクワードは分析から除外されます。

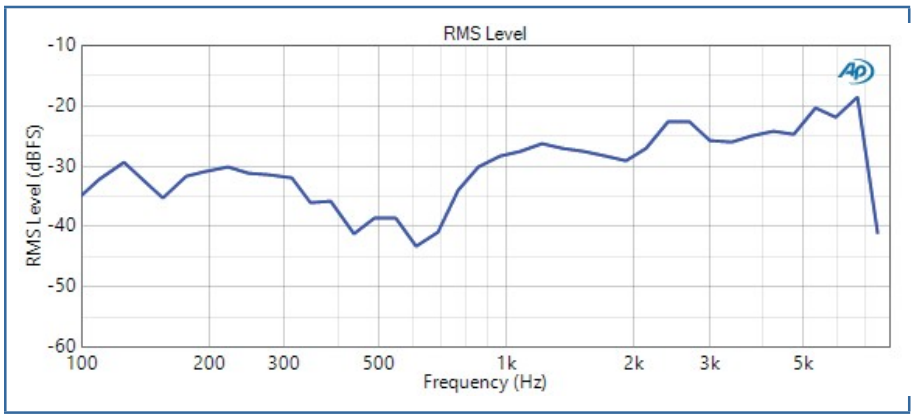

図 6.
マルチトーンで測定されたスマート スピーカーの入力パスのレベル応答

図 6 は、マルチトーン信号に対するスマートスピーカー入力システムの応答を示しています。解像度は低いにもかかわらず、曲線の形状は伝達関数およびチャープ測定からの曲線とよく一致しています。ただし、絶対レベルはチャープ応答のレベルより約 17 dB 低いことに注意してください。これは、マルチトーン信号の波高率がチャープ信号よりもはるかに低いためです。

スマートスピーカー出力パス

スマート スピーカーのプライマリ出力パスには、Web サーバーからデバイスに送信されるデジタル オーディオ コンテンツが含まれます。そこでコンテンツはデジタルからアナログに変換され、最終的にデバイスのスピーカー システムで再生されるときに音響信号に変換されます。このパスのオーディオ テストでは、オーディオアナライザーの入力に接続されたマイクを使用してデバイスからの音響出力信号をキャプチャし、これを変更されていない元のオーディオ ファイルと比較する必要があります。

物理的テスト設定

図 2 は、スマート スピーカー出力パスのセットアップを示しています。これも、ガイドラインとしてスピーカーフォン用の IEEE 1329 標準の使用に基づいています。この場合、ロシミュレータは測定マイクに置き換えられます。このテストも、準無響または(時間選択的)測定技術が使用されない限り、無響室内で実施する必要があります。

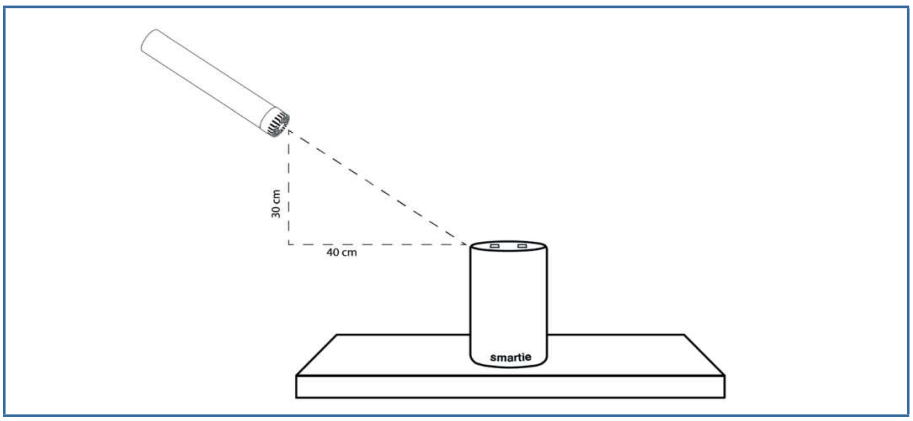

図 7.
スマートスピーカー出力パスの設定条件

テスト信号の準備

スマートスピーカーは通常、オンライン音楽サービスやインターネット ラジオ局、または「クラウド」またはユーザーのデバイスにあるユーザー独自の音楽リポジトリからオーディオ コンテンツを再生できます。通常、サードパーティのストリーミングサービスからストリーミングされたオーディオ ファイルの正確なコピーを取得することはできません。したがって、最良のアプローチは、刺激信号を含むオーディオファイルを準備し、音楽リポジトリにアップロードすることです。人気の Alexa サービスの場合、Amazon は当初、ユーザーがスマートデバイスにストリーミングするために独自のコンテンツをアップロードすることを許可していましたが、後にポリシーを変更してこれを禁止しました。効果的な回避策は、Alexa の「スキル」があるサードパーティのストリーミング サービス (Plex Media Server など) を使用することです。このアプローチを使用すると、自分の PC 上にメディア サーバーを作成し、サーバーから接続されたスマートスピーカーにファイルをストリーミングするように IVA に指示することができます。

ここでも、分析では、オーディオ アナライザーによって取得された信号のサンプル レートが元の信号のサンプル レートと一致する必要があります。ただし、ファイルのストリーミング元の音楽サービスには、ファイルをストリーミングする前に満たさなければならない独自の制約がある可能性があります。たとえば、一部のサービスでは、サンプル レート 44.1 kHz で .MP3 形式にエンコードし、アーティスト、アルバム、トラックなどのマルチメディア タグを割り当てる必要があります。刺激信号を含む音声ファイルを準備したら、サーバーにアップロードする必要があります。

DUT 出力のキャプチャーと解析
刺激信号を含むファイルが準備され、接続されている音楽サービスにアップロードされると、テストを実行できます。スマート スピーカーにテスト信号を再生させるには、通常、音声コマンドが必要です。たとえば、ウェイク ワードが「Smartie」であるスマート スピーカーの場合、コマンドは「Smartie: Audio Precision の Song Test Signal を再生してください」となる可能性があります。IVA がコマンドを正常に解釈し、音楽トラックを見つけた場合、数秒後に、「わかりました、曲を再生します。オーディオ精度によるテスト信号」などの「音声」応答で応答し、その後にテスト信号が続きます。テスト信号をキャプチャするには、テスト信号が再生される前にオーディオアナライザーでの測定を開始する必要があります。

(IVA の音声応答ではなく) テスト信号が確実に分析されるようにするには、オーディオ アナライザーに目的の信号をトリガーする手段が必要です。トリガーメカニズムは、測定/刺激の種類によって異なる場合があります。たとえば、AP オーディオ アナライザーでは、次のトリガー メカニズムが使用されます。

- マルチトーン信号には独自のシグネチャがあり、これにより信号自体で測定を高い確率でトリガーできるようになります。
- オープンループ測定の場合、ログスイープサイン(またはチャープ)測定とステップサイン測定では、サイン波パイロットトーンが使用されます。アナライザーは、周波数選択的なしきい値トリガーを使用して、パイロット トーンとテスト信号を特定します。
- 伝達関数測定には、テスト信号自体でトリガできる「一致」機能があります。繰り返しのフレーズを含む音楽などの信号の場合、ユーザーは、トリガーを改善するために、最大長シーケンス (MLS) などの固有の信号の短いセクションを刺激の先頭に追加できます。

例 – スマート スピーカー出力パスの対数掃引サイン波測定 図 8 は、2.0 秒の掃引長で 50 Hz ～ 20 kHz の対数掃引サイン波測定を使用したスマート スピーカーテストからの刺激波形と応答波形を示しています。0.2 秒のパイロット トーンが波形の先頭に表示されます。

パイロットトーンは、トリガーだけでなく、オーディオアナライザーと DUT の間のサンプル レートの差を測定するためにも使用されます。上で述べたように、DUT のサンプルクロックはオーディオ アナライザーのサンプルクロックと正確に一致することはありません。たとえば、元の信号は 48 kHz のサンプルレートで作成された可能性がありますが、DUT のサンプルレートはわずかに異なります (たとえば 47.990 kHz)。2 つの信号のサンプルレートは同一である必要があります。同一でない場合、結果にアーティファクトが生じます。チャープ測定が APx500 オーディオ アナライザーのオープンループ構成で実行される場合、システムはパイロットトーンの周波数を分析することによってサンプルレートの差を測定します。次に、刺激と応答が同一のサンプルクロックを持つように刺激信号を再作成します。

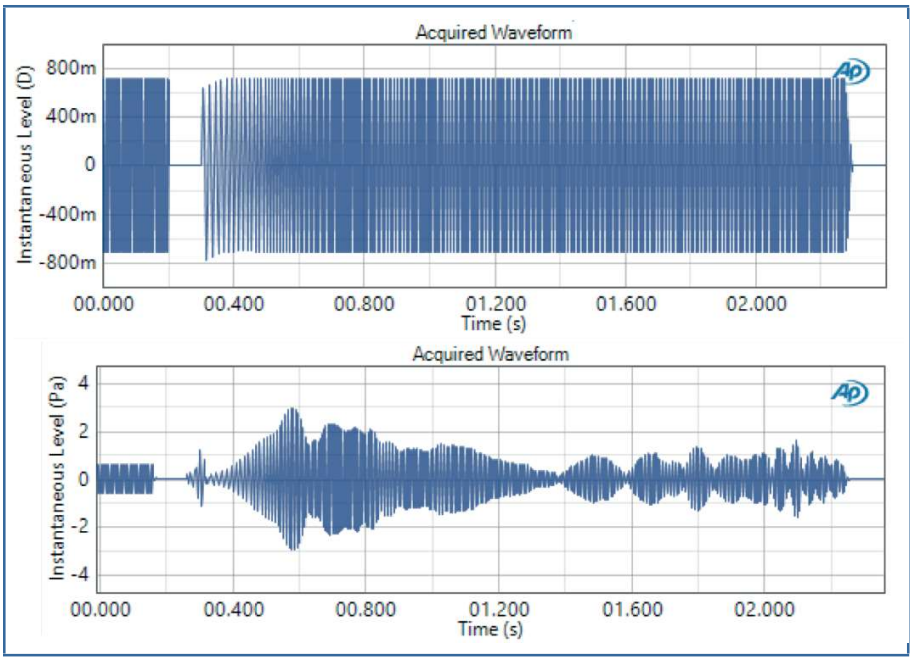

図 8.
スマートスピーカーテストの刺激波形（上）と音圧応答波形（下）。200 ms のパイロット トーンを使用して、2.0 秒で 50 Hz から 20 kHz までの対数掃引サイン波チャープ。

図 9 は、図 8 の刺激波形と応答波形から導出された音響レベル応答を示しています。約 1.5 kHz から始まる一定の周波数間隔で発生する曲線のくぼみは、スピーカーからマイクに直接伝播する音波と破壊的に結合したテーブルからの音響反射によるものと考えられます。

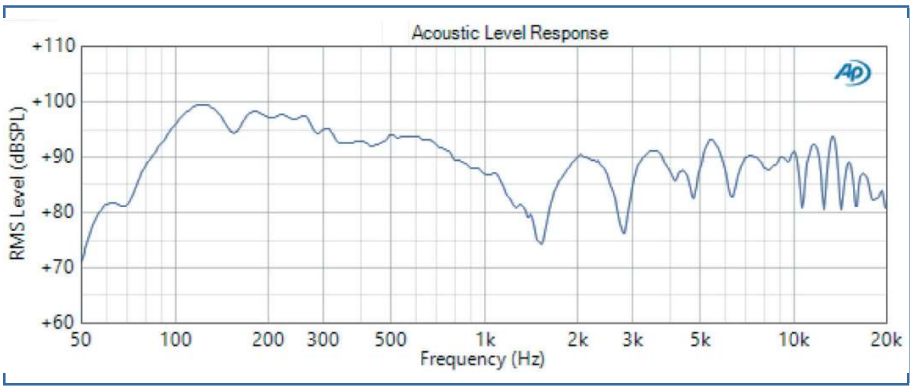

図 9.
スマート スピーカーの周波数応答レベルは、図 3 に示す波形を使用した対数掃引サイン波テストによって測定されます。

例 – スマート スピーカーの出力パスの伝達関数測定 図 10 は、同じスマート スピーカーの出力パスをテストするために使用される音楽信号の最初の数秒を示しています。波形の最初の短い 0.1 秒のバーストは、トリガ用の信号マッチングを向上させるために音楽信号の先頭に付加された短い MLS 信号です。

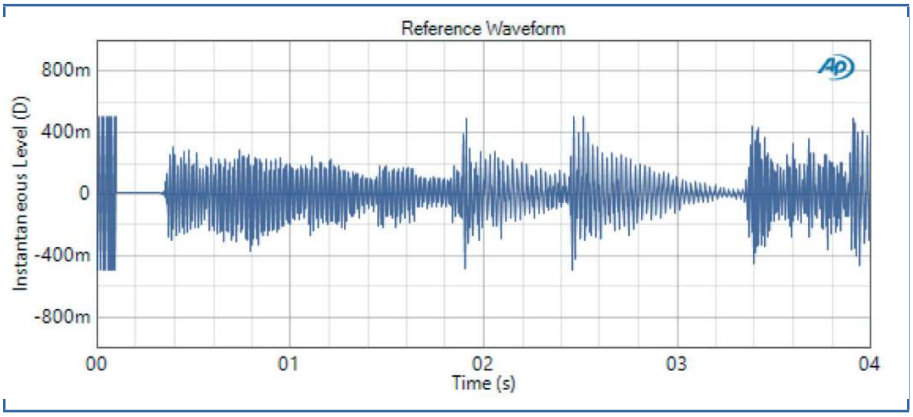

図 10.
スマート スピーカーを刺激するために使用される音楽信号の最初の部分。

20 秒間の音楽信号の分析から計算された周波数応答の大きさを図 11 に示します。この結果は、スマート スピーカーの感度対周波数を dB (Pa/FS) 単位で直接測定したものであることに注意してください。

図 9 と図 11 の類似点に注目してください。2 つの曲線間の形状のわずかな違いは、デバイスが特定の周波数でのチャープ信号とは異なる音楽信号に応答するためである可能性があります。

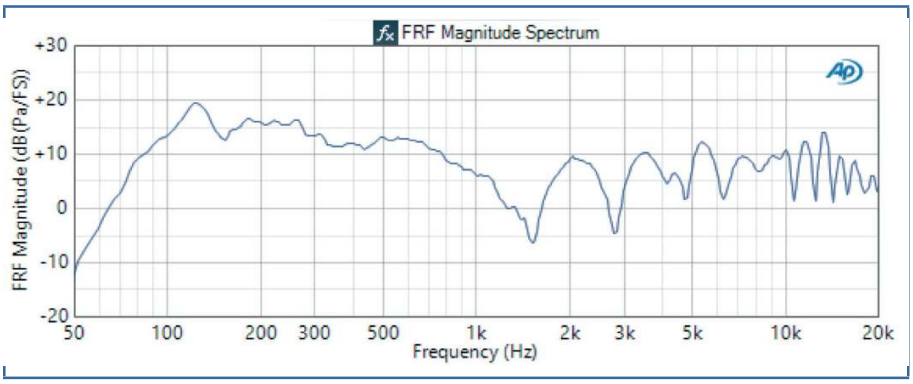

図 11.
音楽信号の伝達関数測定を使用して測定されたスマートスピーカーの周波数応答の大きさ。

最後に

これで、スマート スピーカーの音響測定の概要を終わります。スマート スピーカーのテストに伴う固有の課題のいくつかについて説明しました。スマート スピーカーの入力と出力は両方とも音響的であり、音声コマンドによって制御する必要があり、テストには IVA サービスおよびバックエンド サーバーとの対話が含まれます。これらの課題にもかかわらず、優れたオーディオ アナライザーといくつかの音響アクセサリ、さらに障害のいくつかを克服する少しの創意工夫があれば、スマート スピーカー デバイスの有意義な音質測定を行うことは比較的簡単です。

詳細は以下のサイトを参照願います。
https://www.ap.com/download/appnote-smart-speaker-acoustic-measurements/?wpdmdl=9081&ind=DQpBcHBOb3RlX1NtYXJ0X1NwZWFrZXJfQWNvdXN0aWNfTWVhc3VyZW1lbnRzX0FQJkdSQVNfRklOQUwyLnBkZg