

POLQA[®]による次世代の (3G/4G) 音声品質試験 ホワイトペーパー

POLQA[®] (Perceptual Objective Listening Quality Analysis : 受聴品質の知覚的客観的分析) は、ITU-T 勧告 P.863 に基づく次世代のモバイル音声品質試験標準であり、HD 音声、3G、VoLTE (4G)、VoHSPA、および VoIP のスーパー・ワイドバンド要件に対応するために特に開発されたものです。このホワイトペーパーでは、R&S[®]UPV オーディオ・アナライザに実装された POLQA[®] アルゴリズムについて説明するとともに、特定の規格に依存せずにオーディオ測定を行うためのハードウェア・セットアップの例を示します。

*POLQA[®] および PESQ[®] は、OPTICOM Dipl.-Ing.
M. Keyhl GmbH, Germany および Psytechnics
Ltd., UK の登録商標です。*

目次

1	はじめに.....	3
2	概要.....	4
2.1	POLQAアルゴリズム.....	5
2.1.1	技術的概要.....	7
2.1.1.1	一時アライメント.....	8
2.1.1.2	サンプル・レートの推定.....	11
2.1.2	知覚モデル.....	12
2.1.2.1	定数設定値の事前計算.....	15
2.1.2.2	ピッチ強度密度.....	15
2.1.2.3	通話を含むフレーム、無音フレーム、および スーパーサイレント・フレームの計算.....	16
2.1.2.4	周波数、雑音、および反響インジケータの計算.....	16
2.1.2.5	参照信号のスケーリング.....	16
2.1.2.6	線形周波数応答歪みに対応する元のピッチ強度密度の部分的補正.....	17
2.1.2.7	マスキング効果のモデル化、ピッチ・ラウドネス密度の計算.....	17
2.1.2.8	参照信号および劣化信号における雑音の補正.....	17
2.1.2.9	最終的な雑音密度の計算.....	18
2.1.2.10	最終的なMOS-LQO POLQA計算.....	18
3	PESQからPOLQAへ.....	20
3.1	POLQAの拡張機能.....	20
3.2	POLQAはPESQを置き換えるものになり得るか?.....	20
4	テスト・ソリューション.....	21
4.1	ダウンリンクのPOLQA測定.....	22
4.2	アップリンクのPOLQA測定.....	23
5	参考文献.....	24
6	追加情報.....	25
7	略語.....	26

1 はじめに

POLQA は、ITU-T 勧告 P.863^[2]に基づく次世代のモバイル音声品質試験標準です。これは HD 音声、3G、VoLTE（4G）、VoHSPA、および VoIP（Voice over Internet Protocol）のスーパー・ワイドバンド（SWB）要件に対応するために特に開発されたものです。このホワイトペーパーでは、R&S[®]UPV オーディオ・アナライザに実装された POLQA[®]アルゴリズムについて説明し、PESQ^[2]（Perceptual Evaluation of Speech Quality：通話品質の知覚的評価）と比較してその改善点を指摘するとともに、通話品質試験のハードウェア・セットアップの例を紹介します。

2 概要

PESQ ITU-T P.862 勧告では、現在のネットワークおよび新 ネットワークで必要とされるすべての条件が考慮されているわけではありません。このため、POLQA への移行が必要とされています。POLQA の性能強化により、以下のことが実現されました。

• 3G/4G/LTE およびオーディオ・コーデックで使用する新しいタイプの通話コーデック（例：AAC および MP3）。
• 非線形処理を使用した音声機能強化（VQE/VED）システム。
• SBR（Spectral Band Replication：スペクトラム帯域複製）など、オーディオ帯域幅を変更するコーデック。
• 背景雑音レベルが極めて高い状況下での信号の測定。
• 音圧の変化から生じる影響の正確なモデル化。
• NB（狭帯域、300 から 3400 Hz）および SWB（スーパー・ワイドバンド、50 から 14000 Hz）モードのサポート。
• VoIP および 3G パケット・オーディオで見られるような時間のスケーリングおよびワープの処理。
• 音響インタフェースを通じて録音された信号の評価。
• 反響、線形、および非線形フィルタリングの正確な重みづけ。
• AMR（GSM/UMTS）コードと EVRC（CDMA2000）コードによる送信結果の直接比較。

POLQA の適用対象としては、以下の分野が考えられます。

• コーデックの評価。
• 送信および受信方向での端末試験（音響経路およびマイクの影響がある場合、またはない場合）。
• 帯域幅の拡張。
• ネットワークへのデジタル接続またはアナログ接続を使用したライブ・ネットワーク試験。
• エミュレートしたネットワーク、およびプロトタイプ・ネットワークの試験。
• UMTS、CDMA2000、GSM、TETRA、WB-DECT、VoIP、POTS、Video、電話通話、Bluetooth。
• Voice Activity Detection（VAD：無音圧縮）、Automatic Gain Control（AGC：自動利得制御）。
• Voice Enhancement Devices（VED：音声強調処理装置）、Noise Reduction（NR：ノイズ・リダクション（低減））。
• Discontinuous Transmission（DTX：不連続送信）、快適雑音挿入。

2.1 POLQAアルゴリズム

POLQA アルゴリズムは、参照信号 $X(t)$ を、コーディング、デコーディング、LAN、および RF コンポーネントなどを含む通信システムを経由することで劣化した信号 $Y(t)$ と比較します。このアルゴリズムの出力は、人間による主観的な聴き取りテストを実施した場合に $Y(t)$ に関して判定されることが予測される知覚品質です。

最初のステップとして、参照信号および劣化信号はフレームと呼ばれる非常に小さい時間スライスに分割されます。次に、参照信号の各フレームの遅延が、関連する劣化信号のフレームと比較して計算されます。続いて、劣化信号のサンプル・レートが推定が行われます。推定されたサンプル・レートが参照信号のサンプル・レートと大幅に異なる場合は、高いサンプル・レートの信号をダウン・サンプリングして、遅延が再判定されます。

POLQA は、図 1 に示すように、検出された遅延セットに基づいて知覚モデルを使用し、参照（入力）を、これと整合する位置にある SUT（System Under Test：被試験システム）の劣化した（出力）信号と比較します。

このプロセスの主要な目的は、知覚されたピッチ（パーク）とラウドネス（ソーン）を考慮して、この両方の信号を、人間の聴覚システム内の心理的／物理的表現に類似した内部処理に変換することです。これは次のような一連のステージを経て実現されます。

• 時間アライメント
• 校正済み受聴レベルとのレベルのアライメント
• 時間周波数マッピング
• 周波数ワープ
• 圧縮可能なラウドネス・スケーリング

POLQA による知覚品質の予測では、SWB モードの場合は再生レベルが考慮されます。NB モードの場合、通話品質は一定の受聴レベルに基づいて判断されます。代表値で処理を行うことで、ローカル（即時）利得の変化および線形フィルタリング効果を反映することができます。

POLQA はまた、参照信号内の低レベルの雑音の排除、および劣化した出力信号内の雑音の部分的な抑制も行います。参照信号および劣化信号の特性を変更する操作は、理想化処理で 사용됩니다。このような主観的試験は、参照信号との直接の比較を行わずに実施されます（絶対カテゴリ・レーティング）。POLQA は、認知モデル内で計算される 6 つの品質インジケータを提供しています。

• 周波数応答インジケータ（FREQ）
• 雑音インジケータ（NOISE）
• 室内反響インジケータ（REVERB）
• 時間-ピッチ-ラウドネス・ドメインにおける内部相違を記述する 3 つのインジケータ

これらのインジケータを組み合わせることで、客観的な受聴品質である MOS が得られます。POLQA では、参照信号は常にクリーンである（雑音がない）ものと想定されています。

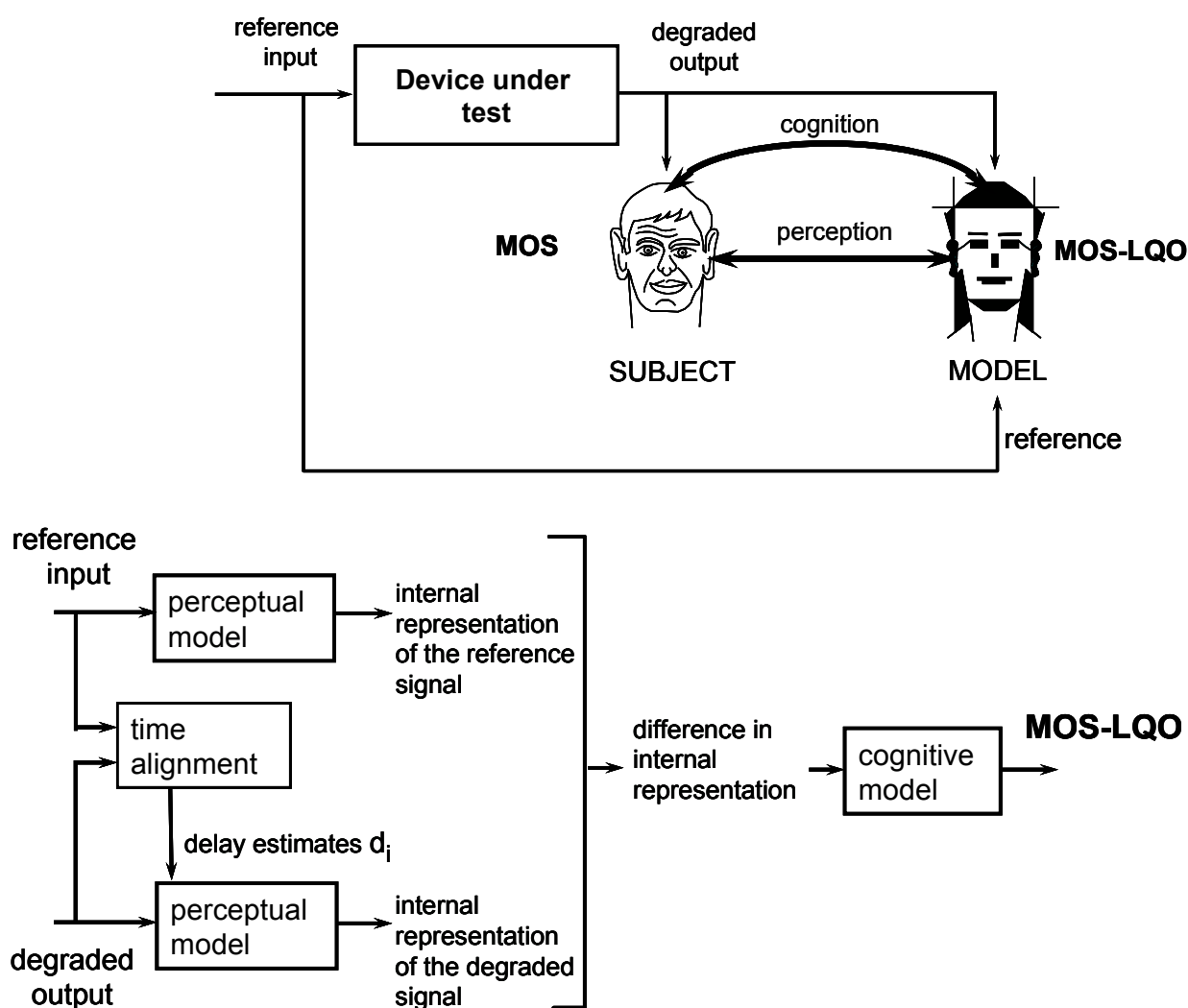


図1：POLQA の基本原理

主観的受聴品質スコアと客観的受聴品質スコアの相違として、主観的スコアは受聴者グループや試験の設計によって左右されます。客観的測定の場合、試験の前後関係や受聴者パネルの個人的な行動の影響を受けることはありません。この測定には、「平均的な受聴者グループによる平均的な試験スコア」が反映されます。客観的モデルは、特定の実験から得られた絶対スコアを正確に再現するものではなく、相対的な品質ランキングを再現します。客観的品質測定を適切に行うことで、多数の異なる主観的実験結果との高い相関性が得られるはずです。

POLQAスコアは、膨大な量の個別データ・セットの平均値を反映したMOSスケールにあらかじめマップされているため、日常処理として客観的スコアへの特別なマッピングを適用する必要はありません。

SWBモードの場合、POLQAは常に、50から14000 Hzのバンドパス・フィルタで事前フィルタリングされた、たとえば48 kHzのサンプリング・レートモノラル信号を必要とします。この信号はNBモードでも使用できます。またこれを16または8 kHzでダウン・サンプリングすることもできます。

POLQA評価フェーズで使用される通話ファイルの属性は以下のとおりです。

<ul style="list-style-type: none"> 各参照通話ファイルは、少なくとも1秒の、しかし2秒を超えないギャップで分割された複数の文から構成されている必要があります。
<ul style="list-style-type: none"> 各ファイル内の通話は3秒（最小値）以上でなければなりません。
<ul style="list-style-type: none"> 参照通話ファイルは、先頭および末尾に、通話信号のクリッピングを防止するだけの十分な無音時間（たとえば、それぞれ200ミリ秒）を含む必要があります。
<ul style="list-style-type: none"> SWB参照通話サンプルの場合、参照ファイルの先頭、末尾、および文の間のギャップ部分のノイズフロアが、-84 dBov(A)¹⁾ を超えてはなりません。
<ul style="list-style-type: none"> 被測定物の録音に使用する部屋では、200 Hz超の反響時間が300ミリ秒より短くなければなりません（たとえば、無響室）。

SUT を通過した劣化信号は、電気インタフェースまたは音響インタフェースのいずれかで捕捉されます。

MOS スケールは 1 から 5 の範囲ですが、飽和により、予測スコアの最大値は SWB の場合は MOS-LQO = 4.75 に、NB の場合は MOS-LQO = 4.5 になります（2.1.2.10 を参照）。

2.1.1 技術的概要

図 2 は、POLQA アルゴリズムの概要を示しています。

入力には 2 つの 16 ビットの波形が使用されます。最初の波形には参照信号（歪みがない信号）が含まれ、2 番目の波形には劣化信号が含まれています。

POLQA アルゴリズムは、入力信号のサンプル・レートの相違の補正に使用されるサンプル・レート・コンバータに加えて、一時アライメント・ブロック、サンプル・レート・エスティメータ、そして MOS 計算を実行する実際のコア・モデルから構成されます。

最初のステップとして 2 つの入力信号の間の遅延が判別され、この 2 つの信号のサンプル・レートが相対的に推定されます。サンプル・レートの推定は、一時アライメントによって算出された遅延情報に基づきます。

サンプル・レートの相違幅がおおよそ 1%を超える場合は、より高いサンプル・レートを持つ信号でダウン・サンプリングが行われます。ステップが 1 つ終わるたびに、その結果が平均遅延信頼性インジケータ（遅延推定の品質を測る基準）とともに保管されます。再サンプリング・ステップの結果の中から、全体的信頼性が最も高いものが最終的に選択されます。

遅延が正しく判別されてサンプル・レートの相違が補正された後、信号および遅延情報は知覚モデルに渡されます。知覚モデルは、知覚性や歪みの不快感を計算して、これらを MOS スケールにマッピングします。

¹⁾ dBov (= オーバーロード) は、クリッピングを起こさずにデバイスで処理可能な最大値と比較した、信号（通常はオーディオ信号）の振幅です。これは dBFS と似ていますが、アナログ・システムにも適用可能です。dB(A) フィルタは広く使用されています。dB(A)は、人間の耳に感知できるおおよそ 1 kHz における 40 dB など等ラウドネス曲線の逆数に対応します。dB(A) フィルタを使用する騒音計では、非常に高い周波数または低い周波数に対する感度が低くなります。

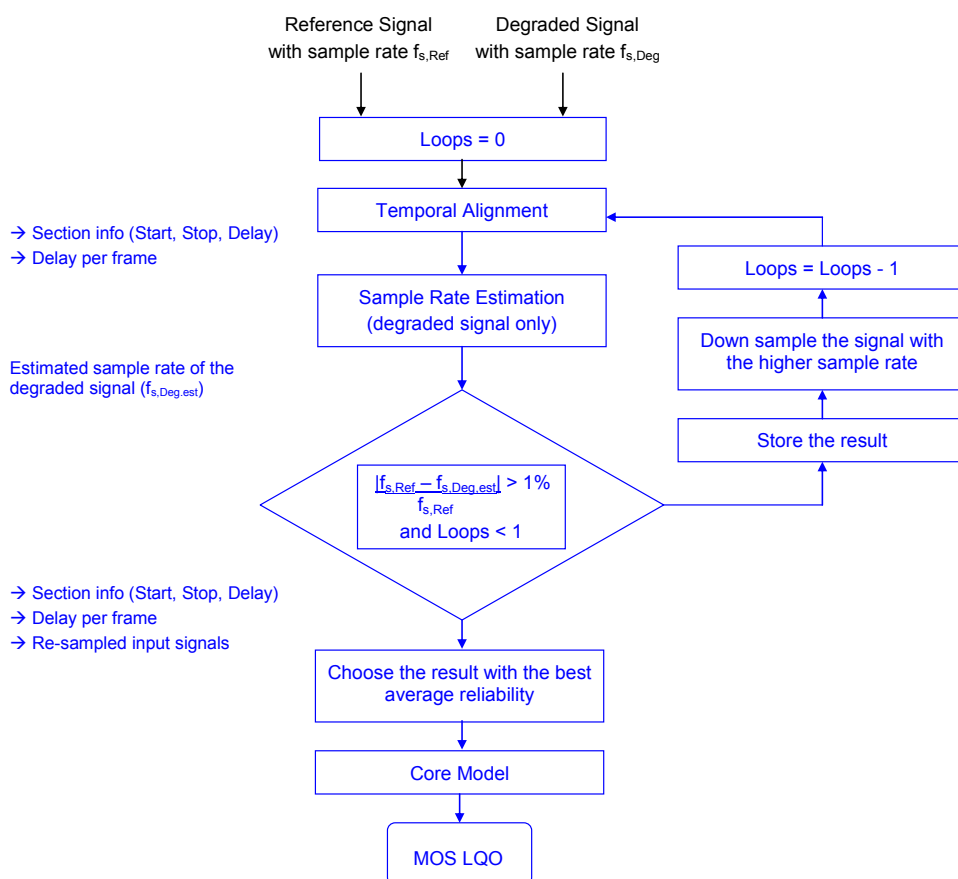


図2：POLQA の概要

2.1.1.1 一時アライメント

一時アライメントは：

- | |
|---|
| <ul style="list-style-type: none"> 等距離のフレーム・ペアに信号を分割し、フレーム・ペアごとに遅延を計算します。 可能であれば、参照信号内で、劣化信号内の対応するセクションと一致するものを検索します（その逆はありません）。 検索範囲が広がり、大量の計算能力が必要になること（これは、時間スケールの入力信号の場合は特に深刻です）を避けるため、ステップごとにフレーム単位の遅延を調整します。 |
|---|

一時アライメントは以下の主要ブロックから構成されます。

- | |
|---|
| <ul style="list-style-type: none"> フィルタリング 事前アライメント (Pre-Alignment) 粗調整アライメント (Coarse Alignment) 微調整アライメント (Fine Alignment) セクション組み合わせ |
|---|

入力信号は、等距離のマクロ・フレームに分割されますが、その長さは、入力サンプル・レートによって異なります。遅延は「マクロ・フレーム」単位で判別されます。計算上の遅延とは、参照信号を基準とした劣化信号の遅延です。

事前アライメントでは、信号内の通話を含むセクションを判別し、マクロ・フレームごとに初期の推定遅延を計算し、各マクロ・フレームの遅延に関して必要となる推定検索範囲を計算します。

粗調整アライメントでは、多次元検索と、ビタビ（Viterbi）に似たバックトラッキング・アルゴリズムを使用して、検出された遅延のフィルタリングを行い、フレーム単位の遅延の調整処理を繰り返し実行します。必要となる相関の長さおよび検索範囲を絞るために、粗調整アライメントでは、ステップを重ねるたびに精度を上げていきます。

微調整アライメントでは、各フレームの正確なサンプル遅延が、最大限の精度で、入力信号で直接に判別されます。このステップにおける検索範囲は、粗調整アライメントによる最後の反復時の正確性によって決定します。最終ステップでは、ほとんど同じ遅延を持つすべてのセクションが組み合わされて、「セクション情報」と呼ばれるものが構成されます。

この一時アライメント・プロシージャの特性は以下のとおりです。

• 静的遅延に関するハードリミットはありません。
• 静的遅延を中心として前後 300 ミリ秒の可変の遅延に対応するように設計されていますが、ハードリミットはありません。
• 遅延はフレームごとに異なる可能性があります。
• サンプル・レートの多少の相違（約 2%未満）は対応可能です。より大きな相違が検出されると、一時アライメントの外部で補正されます。
• ピッチ修正を含む（または含まない）、時間的に引き延ばされたか一時的に圧縮された信号は正しく処理されます。
• SNR が 0 dB を下回る、非常に雑音の多い条件下でも、アライメントは正しく機能します。
• 信号レベルの変化に伴う問題は認識されていません。

一般的な遅延検索方式

一時アライメントに関連する多くのモジュールでは、2 つの信号間の遅延の検出について、同じ方式が使用されています。この方式は、以下の方法で作成されるヒストグラムに基づきます。

• 2 つの信号間の相関性を計算する。
• 検出されたピーク値をヒストグラムの中心に配置する。
• 両方の信号をわずかにシフトする。
• このステップを繰り返す。

ヒストグラムに十分な値が入力されると、値のフィルタリングが行われ、ピークが判別されます。ヒストグラム内のピークの位置は、2 つの信号間の遅延オフセットに相当します。

一般的な遅延信頼性測定

一時アライメントの多くのステップでは、2 つの信号間で検出された遅延の信頼性測定に、単純なピアソン相関が使用されます。

バンドパス・フィルタ

次に処理に進む前に、両方の入力信号のバンドパス・フィルタリングが実行されます。フィルタの形状はモデルの動作モード（SWB または NB）によって異なります。

SWB モードの場合、320 Hz から最大 3400 Hz で信号のバンドパス・フィルタリングが行われます。NB モードの場合は、290 Hz から最大 3300 Hz でバンドパス・フィルタリングが行われます。

これらのフィルタ済み信号は一時アライメントにのみ使用されるものであることにご注意下さい。知覚モデルでは別の方法でフィルタリングされた信号が使用されます。

事前アライメント

事前アライメントでは、最初に、劣化信号内のリパース・ポイントが特定されます。リパース・ポイントは、信号が通話休止状態から通話を含む状態に推移する位置です。

リパース・ポイントは、通話を含むセクションの開始位置をマークし、リパース・セクションは、リパース・ポイントから開始する通話を含むセグメント全体を示します。

リパース・ポイントごとにリパース・セクション情報が計算されます。セクション情報には、セクションの開始位置と終了位置、初期の遅延値、検出された遅延の信頼性標識、およびその正確性（つまり、遅延が正確とされるための上限と下限）が保管されています。

粗調整アライメント

粗調整アライメントでは、フレームごとに遅延を段階的に調整します。ここで各信号は小さなサブセクション（フィーチャ・フレーム）に分割され、各サブセクションの特性値（フィーチャ）が計算されます。

結果として得られたベクトルはフィーチャ・ベクトルと呼ばれます。フィーチャ・フレームは等距離であり、処理を反復するたびにフレームの長さが短縮されます。この長さは、マクロ・フレームの長さとは関係ありません。長さを繰り返し短縮していくことで、反復のたびに推定遅延の正確性が増し、それと同時に検索範囲も絞られます。

複数のフィーチャ・ベクトルが計算され、マクロ・フレームごとに、最も適したフィーチャを使用して、現行フレームの最終的な遅延値が判別されます。

粗調整アライメントの結果として、サンプルで表した各マクロ・フレームの遅延のベクトルが得られます。その正確性は、最後の反復時のフィーチャ・フレームの長さに依存します。

微調整アライメント

微調整アライメントでは、参照信号と劣化信号を可能な限り高い精度で処理し、サンプルで表した各フレームの正確な遅延を判別します。必要となる検索範囲は、これに先立つアライメント・ステップによって大幅に減少しています。このため、非常に短い相関性を使用しながらも、正確性を維持して正確な遅延値を予測できます。微調整アライメントの結果として、各マクロ・フレームのサンプルの正確な遅延値を入手できます。

遅延値が同じセクションの結合

このステップでは、同じ遅延を持つすべてのセクションを組み合わせます。つまり、このセクション全体に関して一組の情報（遅延、信頼性、開始、停止、通話状態）が保管されることになります。

2 番目のステップでは、以下の条件が当てはまる場合、セクション $n+1$ とセクション n が結合されます。

セクション $n+1$ に通話状態が含まれていて、かつ、両方のセクションの遅延の相違が 0.3 ミリ秒よりも小さい場合、または
セクション $n+1$ が通話休止から構成されていて、かつ、両方のセクションの遅延の相違が 15 ミリ秒よりも小さい場合。

この結果として作成されたセクション情報は、心理音響的モデルに渡されます。

2.1.1.2 サンプル・レートの推定

ここで重要なことは、サンプル・レートの相違に起因する遅延の変化と、パケット損失による歪みや揺らぎ吸収バッファの調整に起因する変化とを区別することです。POLQA では、サンプル・レートの相違が原因と考えられるすべての遅延変化のヒストグラムを計算することで、これに対応しています。

参照信号と劣化信号の両方の再生速度の相違のうちで、知覚的に無意味なものを補正するためには、**サンプル・レート比率の検出**が必要になります。このような相違には、さまざまな原因が考えられますが、これには意図的なものもあれば偶発的なものもあります。

いずれの場合も、その結果は同じであり、2つの信号のサンプル・レートのわずかに数パーセントの範囲内の相違として記述されます。これは、公称レートではなく他の信号と比較した実効サンプル・レートを意味します。

POLQA では、フレーム単位の遅延ベクトルと、一時アライメントによって判別された、信号内で検出された音声信号を含むセクションに基づいて、これらの状況を検出します。このアルゴリズムは、サンプル・レートの相違によって遅延に変化が生じること、そして、その変化の程度が実効サンプル・レート比率に対応するという理論に基づいています。サンプル・レートの相違に起因する遅延の変化は、大幅なものであるよりも小幅なものであることが多いため、比較的小規模の変化のみが受け入れられます。

計算上のヒストグラムは、フレームごとの遅延分布として表現されます。つまり、各検出された遅延のバリエーションは、遅延の変化がない状態のその前のセクションの到達時間で除算されます。

信頼できないピークをフィルタリングしてヒストグラムから外した後のピーク値の位置は、サンプル・レート比率を示すことになります。正確な値を計算するために、ヒストグラムに保管されているサンプル数 *NumAvg* がカウントされ、すべての値の加重平均 (*AvgBin*) が計算されて、この値からサンプル・レート比率 *SRRatio* が導出されます。

検出されたサンプル・レート比率が 0.01 よりも大きい場合は、これより大きいサンプル・レートの信号でダウン・サンプリングが行われ、処理全体が初めから実行されます。信頼に足るサンプル・レート比率を信号で判別できないために過剰なループが発生することを防ぐため、この処理が最大で 1 回実行されます。

サンプル・レートを完全には判別できない場合（たとえば、追加で発生する様々な遅延を伴う信号の場合）でも、検出されたサンプル・レート比率は、信号を一時アライメントで確実に処理するために必要とされるだけの十分な正確性を備えています。

2.1.2 知覚モデル

図3は、内部処理の計算に使用される知覚モデルの簡略化したブロック図を示しています。

参照信号および劣化信号のピッチ強度密度（時間と周波数の関数としての強度）は、時間および周波数で整列した時間信号から導出されます。続いて、これらの密度を使用して、POLQA の最初の 3 つの品質インジケータ、周波数応答歪み（FREQ）、付加雑音（NOISE）、および室内反響（REVERB）が導出されます。

いくつかのステップを経て、ピッチ強度密度から、参照信号および劣化信号の内部処理が導出されます。これらの密度に関連して、1 つはメイン・ブランチを表すもの、1 つは大規模な歪みに関するメイン・ブランチ、1 つは追加の歪みに関するもの、そしてもう 1 つは追加の大規模な歪みに関するものという、4 種類の異なる計算が実行されます。

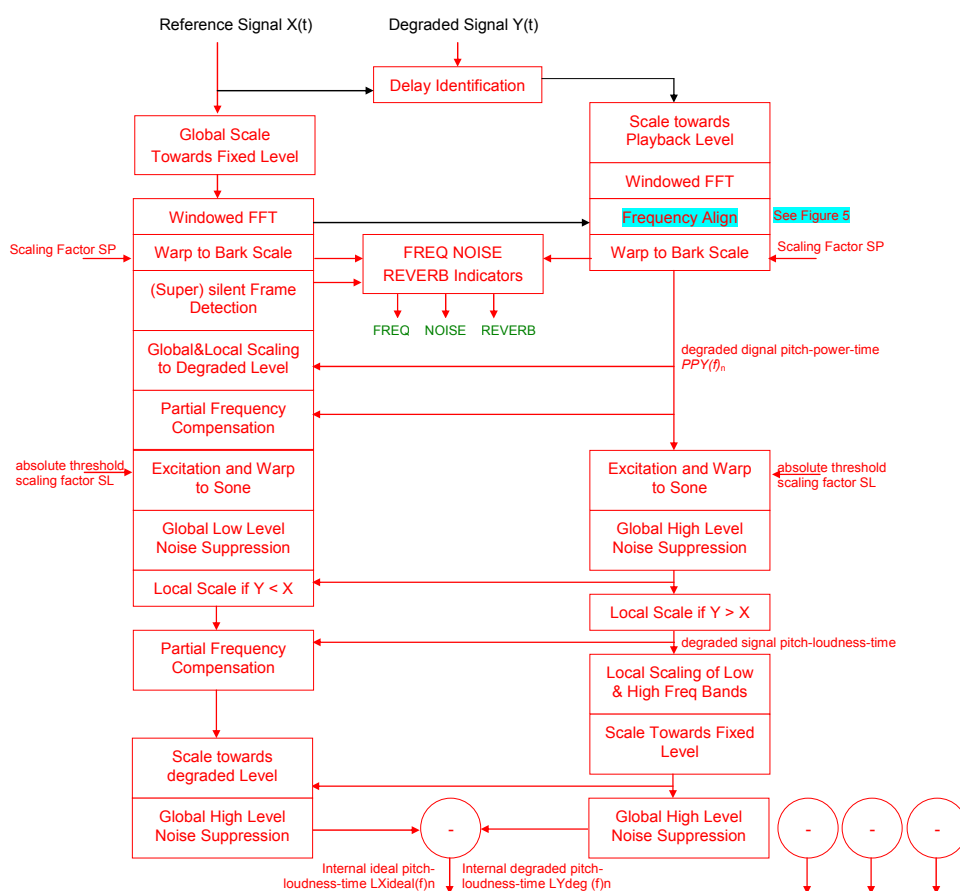


図3：POLQA 知覚モデルの最初の部分の概要

バーク尺度

バーク尺度は、最初の 24 の聴覚臨界帯域に対応する、1 から 24 の範囲を持つ音響心理学的尺度です。Hz で示した帯域分けは、20、100、200、300、400、510、630、770、920、1080、1270、1480、1720、2000、2320、2700、3150、3700、4400、5300、6400、7700、9500、12000、および 15500 です。

$$Z / \text{Bark} = 13 * \arctan(0.00076 * f / \text{Hz}) + 3.5 * \arctan\left(\left(\frac{f}{7.5\text{kHz}}\right)^2\right)$$

励振

ラウドネス密度は、励振レベルから計算されます。これは、ある周波数グループ内のレベルと、この周波数グループにおける受聴の絶対しきい値との相違です。

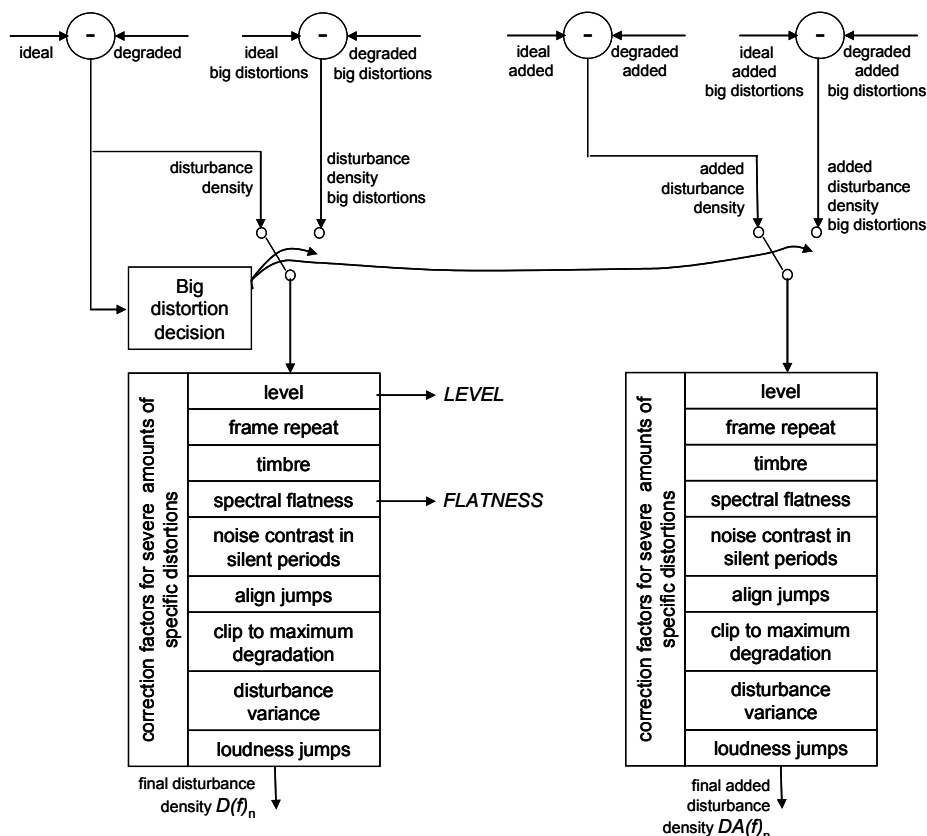


図4：POLQA 知覚モデルの2番目の部分の概要

図4では、4種類の異なる内部処理から、最終的な雑音密度を計算しています。

レベルの変化

デジタル送信におけるレベルの変化とは、通常は、S/N 比（SNR）が雑音によって変わることを意味します。DUT 内の自動利得制御（AGC）回路は、測定強度の合計値に応じて、ボリュームの上げ下げを行います。

無音期間中のノイズ・コントラスト

ノイズ・コントラストとは、雑音音質の突然の変化、たとえば、DTX（Discontinuous Transmission：不連続送信）中にシミュレートされた背景騒音（快適騒音）がオンになったことを意味します。

アライメント・ジャンプ

元のブロックと送信されたブロックの間の遅延は、IP ベースの送信時のブロック損失による再送信の必要性や複数回出現するブロックによって変わる可能性があります。パケット遅延が変化する場合は、このような影響が発生することがあります。

ラウドネス（音量）

ラウドネスとは、主として振幅の心理的知覚に関わる音の品質です。ラウドネスは主観的な測定基準であり、音圧、音圧レベル（dB 単位）、音の強さ、または音響出力など、音の強度に関する客観的測定基準としばしば混同されています。

A-特性（A-重みづけ特性）などのフィルタは、大多数の人の音量の知覚を反映するように、音の測定を調整することを目的としています。ただし、音量の感じ方は A-特性よりもはるかに複雑なプロセスです。さらに、音量の感じ方は人によって異なるため、単一のメトリックを使用して汎用の測定値を得ることはできません。また、音量は周波数、帯域幅、および持続時間を含めて、音圧以外のパラメータの影響も受けます。

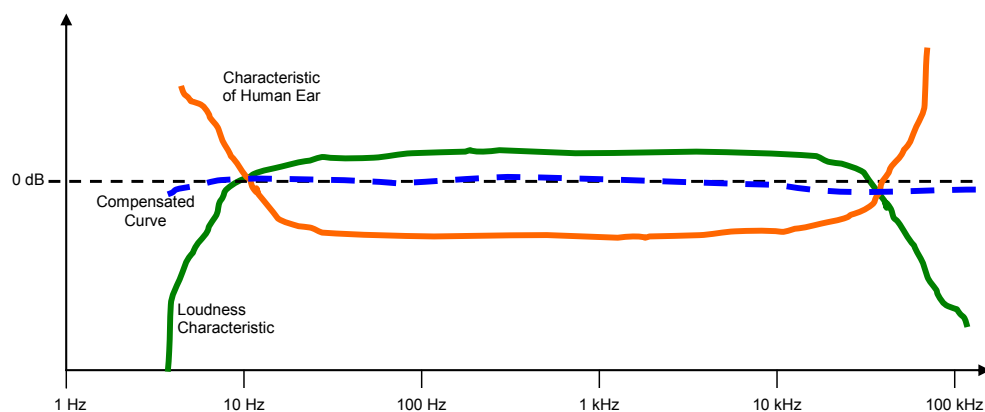


図 5：音量の補正

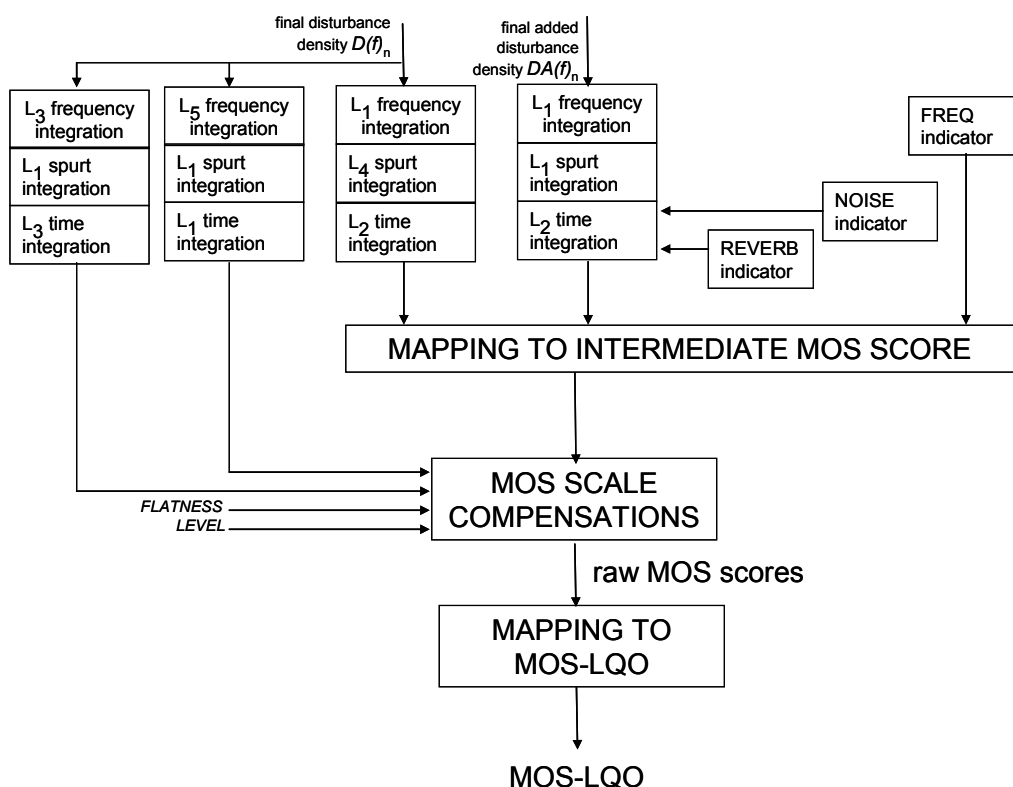


図 6：POLQA 知覚モデルの 3 番目の部分の概要

図 6 では、最終的な雑音密度から MOS-LQO を計算しています。

2.1.2.1 定数設定値の事前計算

FFTウィンドウ・サイズはサンプル周波数に依存する

ウィンドウ・サイズ W は、サンプリング周波数 f_s によって異なります。

$$0 < f_s \leq 9\text{kHz} \rightarrow W = 256$$

$$9 < f_s \leq 18\text{kHz} \rightarrow W = 512$$

$$18 < f_s \leq 36\text{kHz} \rightarrow W = 1024$$

$$36 < f_s \leq 72\text{kHz} \rightarrow W = 2048$$

POLQA は、8、16、および 48 kHz のサンプリング・レートを使用してテスト済みです。再サンプリングを行った場合、主観的試験の場合のように、正確に同じ MOS スコアが再現されることはありません（特に、再サンプリングでの偏差が 2 倍以上になる場合）。

2.1.2.2 ピッチ強度密度

劣化信号 $Y(t)$ を校正係数 C で乗算し（ $C = 2.8 * 10^{(-26 - dBov)/20} * 10^{(73 - dB(A))/20}$ ）、周波数ドメインに変換します（FFT フレームのオーバーラップ 50%）。参照信号は、一定の最適レベルに合わせてスケールリングされます。

ファイルの FFT フレームに対して、周波数ドメイン内でデワープが実行されます（このファイルでは、周波数軸が、参照信号と比較してワープされます）。

最初に、一連の計算の際の、非常に狭い周波数応答の歪みおよび全体的なスペクトラム形状の相違の影響を低減するため、参照信号および劣化信号の FFT 強度スペクトラムの事前処理が行われます。この事前処理には、両方の強度スペクトラムに関して 100 Hz 相当のスライディング・ウィンドウ平均を計算し、これを対数で表現し、218.75 Hz 相当のウィンドウ長を使用してスライディング・ウィンドウ正規化を実行することが含まれます。

現在のフレームについて、劣化信号に対する参照信号のピッチ比率が計算されて、ワープ係数の検索範囲の判別に使用されます（これは、1 から、前述のピッチ比率の間の値になります）。可能な場合は、この前後のフレームから成る 1 組のフレームで検出された最小および最大ピッチ比率に基づき、検索範囲が拡張されます。

この関数が、検索範囲全体を通じて反復されます。現在の反復時のワープ係数を使用して、劣化信号の強度スペクトラムがワープされ、前述の方法で、ワープされた強度スペクトラムの処理が行われます。

処理された参照信号およびワープされた劣化信号のスペクトラムの相関性が計算され、共通の周波数下限値と 1500 Hz の間でビンニングされます。反復が完了すると、「最良の」（関連性が最も高い）ワープ係数が取得されます。次に、処理済みの参照信号とワープした最良の劣化信号のスペクトラムの相関性と、元の参照信号と劣化信号のスペクトラムの相関性とが比較されます。

設定されたしきい値に従って相関性が増加する場合、「最良の」ワープ係数は保持されます。必要であれば、前のフレーム・ペアによって確認された値からの相対変化の最大値を使用してワープ係数を制限できます。

デワープの後、Hz で示される周波数スケールが、人間の聴覚は高周波数よりも低周波数に対する感性が高いことを反映した、バークで示されるピッチ・スケールに合わせてワープされます。これは、FFT 帯域のビンニングと、信号波の集合を正規化して FFT 帯域の対応強度を合計することで実現

されます。この結果として得られた参照信号と劣化信号は、ピッチ強度密度 $PPX(f)_n$ および $PPY(f)_n$ で表されます。この f はバークでの周波数で、添え字 n はフレーム添え字を表します。

2.1.2.3 通話を含むフレーム、無音フレーム、およびスーパーサイレント・フレームの計算

POLQA には、3 つのフレーム・クラスがあります。

• 通話を含むフレーム：	参照信号フレーム・レベル > 平均レベル - 20 dB
• 無音フレーム：	参照信号フレーム・レベル < 平均レベル - 20 dB
• スーパーサイレント・フレーム：	参照信号フレーム・レベル < 平均レベル - 35 dB

2.1.2.4 周波数、雑音、および反響インジケータの計算

周波数応答歪みの全体的影響を示すため、参照信号と劣化信号の平均スペクトラムから 1 つのインジケータが計算されます。付加雑音を考慮せずに周波数応答歪みの影響を推定するために、参照信号の無音フレームに対する劣化信号の平均雑音スペクトラム密度が、劣化信号のピッチ・ラウドネス密度から差し引かれます。

参照ファイルおよび劣化ファイルの通話を含む全フレームのバーク帯域ごとに、結果として得られた劣化信号と参照信号のピッチ・ラウドネス密度の平均が計算されます。次に、これら 2 つの密度の間のピッチ・ラウドネス密度の相違がピッチで統合されて、平均周波数応答相違インジケータが導出されます。このインジケータを、連続するバーク・ピッチ帯域における変化率と組み合わせることで、周波数応答歪み (FREQ) の影響を定量化するインジケータが得られます。

付加雑音の影響を示すため、参照信号の無音フレームに対する劣化信号の平均スペクトラムから 1 つのインジケータが算出されます。劣化信号の平均ピッチ・ラウドネス密度と、無音フレームおよび参照ピッチ・ラウドネス密度ゼロとの間の相違により、付加雑音を定量化する、雑音ラウドネス密度関数が決定します。この雑音ラウドネス密度関数をピッチで統合することで、平均雑音影響インジケータ (NOISE) が導出されます。

室内反響の影響については、参照信号および劣化信号の時系列から、時間とエネルギーの関数であるエネルギー時間曲線 (ETC : Energy Time Curve) が計算されます。ETC は、インパルス応答エンベロープを表します。最初に、直接音 (60 ミリ秒以内に到達する音) の後の ETC 曲線の最大値を判別することで、最大の反響が計算されます。次に、ある期間内で、直接音および最大の反響から 100 ミリ秒以内に到達した反響を除いて、2 番目に大きい反響が判別されます。次に、ある期間内で、直接音および最大の反響音と 2 番目に大きい反響から 100 ミリ秒以内に到達した反響を除いて、3 番目に大きい反響が判別されます。その後、3 つの最も大きい反響のエネルギーを組み合わせ、1 つの反響インジケータ (REVERB) が構成されます。

2.1.2.5 参照信号のスケーリング

現在、参照信号は理想的なレベルにあります。劣化信号は再生レベルと同じレベルで表現されています。

参照信号と劣化信号を比較する前に、品質計算の準備として、全体レベルおよびローカル・レベルの小規模な変化の補正が行われます。参照信号と劣化信号の通話帯域の平均強度に基づき、全体のレベルで同等化が行われます (400 から 3500 Hz)。

参照信号が劣化信号に合わせてスケーリングされ、これによりレベルの相違の影響が補正されます。ゆっくりと変化する利得歪みを正しくモデル化できるように、ローカル・スケーリングにより、最大約 3 dB のレベル変更が実行されます。

2.1.2.6 線形周波数応答歪みに対応する元のピッチ強度密度の部分的補正

SUT による不可聴の線形周波数応答歪みを除外するため、SUT の送信特性に関して、参照信号の部分的フィルタリングが行われます。このためには、通話を含むすべてのフレームにわたり、元のピッチ強度密度および劣化したピッチ強度密度の平均強度スペクトラムが計算されます。

劣化したスペクトラムと元のスペクトラムの比率から、パーク・ビン単位の部分的補正係数が算出されます。

2.1.2.7 マスキング効果のモデル化、ピッチ・ラウドネス密度の計算

マスキングをモデル化するには、ピッチ強度密度のスミアリング表現を計算します。時間ドメインと周波数ドメインの両方のスミアリングが考慮されます。

時間-周波数ドメインのスミアリングには、畳み込み方式が使用されます。スミアリングされたこの表現から、時間-周波数プレーンに隣接する音量が大きい機器によって部分的にマスキングされている低振幅の時間-周波数コンポーネントを抑制（除外）し、参照信号および劣化信号のピッチ強度密度表現が再計算されます。この抑制（除外）は、スミアリングされた表現をスミアリングされていない表現から差し引くか、スミアリングされていない表現をスミアリングされた表現で除算するという、2つの異なる方法のいずれかによって実現されます。

この結果として得られたピッチ強度密度表現が、ツヴィツカーによる強度の法則の改良版を使用して、ピッチ・ラウドネス密度表現に変換されます。

2.1.2.8 参照信号および劣化信号における雑音の補正

SUT（つまり、透過的システム）に由来しない、参照信号の低レベルの雑音は、SUT に起因するものとして、計算から除外する必要があります。

これを行うには、スーパーサイレント・フレームに対する参照信号の平均定常状態雑音ラウドネス密度 $LX(f)_n$ をピッチの関数として計算します。

次に、この平均雑音ラウドネス密度が、参照信号のすべてのピッチ・ラウドネス密度フレームから差し引かれます。この結果として、参照信号の理想化された内部処理が得られます。

劣化信号内の可聴の定常状態雑音の影響は、非定常状態雑音の影響よりも小さくなります。これは、すべての雑音レベルに当てはまります。劣化信号から定常状態騒音の一部を取り除くことで、この効果の影響をモデル化できます。これを行うには、対応する参照信号のフレームがスーパーサイレントとして分類されている、劣化信号 $LY(f)_n$ のフレームの平均定常状態雑音ラウドネス密度を、ピッチ関数として計算します。

その後、平均雑音ラウドネス密度が、劣化信号のすべてのピッチ・ラウドネス密度フレームから差し引かれます。低レベルの雑音と高レベルの雑音では、部分的補正の戦略が異なります。低レベルの雑音の場合、補正はごくわずかですが、ラウドな付加雑音については、より積極的な抑制が行われます。

この結果、受聴試験で認められた主観的な影響に相当する付加雑音を加えた、劣化信号の内部処理が作成されます。

2.1.2.9 最終的な雑音密度の計算

2 つの最終的な雑音密度が計算されます。最初のものは、理想的なピッチ-ラウドネス-時間と劣化したピッチ-ラウドネス-時間の相違の関数から導出されます。2 番目のものは、理想的なピッチ-ラウドネス-時間と劣化したピッチ-ラウドネス-時間の関数から導出されます。この結果の雑音密度は、付加密度と呼ばれます。

さまざまな規模の歪みに対応するため、2 種類の密度が計算されます。1 つは $LX_{ideal}(f)_n$ と $LY_{deg}(f)_n$ の相違から導出されるもので、中小規模の歪みを対象とする知覚モデルを使用して計算され、もう 1 つは $LX_{ideal}(f)_n$ と $LY_{deg}(f)_n$ の相違から導出されるもので、中規模から大規模の歪みを対象とする知覚モデルを使用して計算されます。この 2 つの方法の切り替えは、中小規模の歪みによる雑音の最初の推測を行うことで実行されます。

次のステップでは、大規模な雑音に対応して、雑音および付加雑音密度の最終的補正を行います。最適受聴レベルからの大幅の偏差は、劣化信号レベルから導出されたインジケータによって定量化されます。

この全体の LEVEL インジケータは、MOS-LQO の計算にも使用されます。フレームの反復から生じた深刻な歪みが、参照信号の連続フレームと劣化信号の連続フレームの相関性の比較から導出されたインジケータによって定量化されます。

最適音質からの深刻な偏差は、下位周波数帯に対する上位周波数帯の雑音の比率から導出されたインジケータによって定量化されます。フレーム単位の補正が全体のレベルで実行されます。

音質の偏差の全体のレベルは、MOS-LQO 計算にも使用される FLATNESS インジケータを使用して定量化されます。使用者の注意が雑音でそがれるような深刻な雑音レベルの変化は、参照信号の無音部分から導出されるノイズ・コントラスト・インジケータによって定量化されます。

最後に、雑音および付加雑音の密度が最大レベルでクリップされ、雑音およびラウドネスにおけるジャンプの変化を使用して、特定の雑音時間構造の補正が行われます。

2.1.2.10 最終的なMOS-LQO POLQA計算

4 つの異なる補正を使用して、MOS に似た中間的なインジケータから、POLQA の生スコアが導出されます。

- | |
|---|
| <ul style="list-style-type: none"> 雑音の特定の時間周波数特性に関する 2 つの補正。1 つは周波数、スパート、および時間に対する L_{511} ($L_5-L_1-L_1$) 集約によって、もう 1 つは周波数、スパート、および時間に対する L_{313} 集約によって計算される (図 6 を参照) |
| <ul style="list-style-type: none"> LEVEL インジケータを使用した、非常に低い音圧レベルについての 1 つの補正 |
| <ul style="list-style-type: none"> FLATNESS インジケータを使用した、音質の大幅な歪みに対する 1 つの補正 |

このマッピングには、POLQA ベンチマークには属していないものを含めて、多様な劣化要素のセットが反映されています。これらの MOS 生スコアでは、MOS に似た中間的なインジケータの計算に使用される三次多項マッピングを通じて、線形化がほぼ完了しています。

POLQA MOS 生スコアは、POLQA データベース・セット用に最適化された三次多項を使用して、最終的に MOS-LQO スコアにマッピングされます。

NB モードの場合、POLQA MOS-LQO スコアの最大値は 4.5 です。SWB モードの場合、最大スコアは 4.75 です。理想化プロセスの結果として、参照信号に雑音が含まれているか音質に著しい歪みがあるなど一部の状況下では、最大 MOS スコア（NB モードの場合 4.5、SWB モードの場合 4.75）が実現されないことがあります。

3 PESQ から POLQA へ

3.1 POLQAの拡張機能

• 背景雑音レベルが高い場合でも正確なスコア付けを維持。
• GSM/3G で使用される AMR (Adaptive MultiRate : 適応マルチレート) コーデックと CDMA2000 で使用される EVRC (Enhanced Variable Rate Codec : 拡張可変ビットレート・コーデック) の比較が可能。
• 参照信号の代表的スコア付け。
• サンプル内の通話レベルの影響。
• 50 Hz – 14 kHz の周波数範囲の SWB。
• 線形周波数歪みに対する感度。

NB の場合、POLQA 測定では測定の相対的不確実性が PESQ に比べて 27%低下します。

3.2 POLQAはPESQを置き換えるものになり得るか？

• 主要な通話コーデック (AMR、GSM) に対応する後方互換性のある MOS スケール (NB) です。PESQ を POLQA に容易に移行可能です (PESQ NB および POLQA NB の場合 1...4.5) 。
• SWB 用の拡張 MOS スケールでは HD-音声に対応可能です (POLQA-SWB の場合 1...4.75) 。
• すべてのサンプル周波数に関して 2 つの MOS スケールがあります。Fs = 8 kHz → MOS NB Fs = 16 kHz → MOS SWB

4 テスト・ソリューション

以下の図は、LTE ダウンリンクおよびアップリンクのフェージングを含む POLQA テスト・ソリューションの例を示しています。

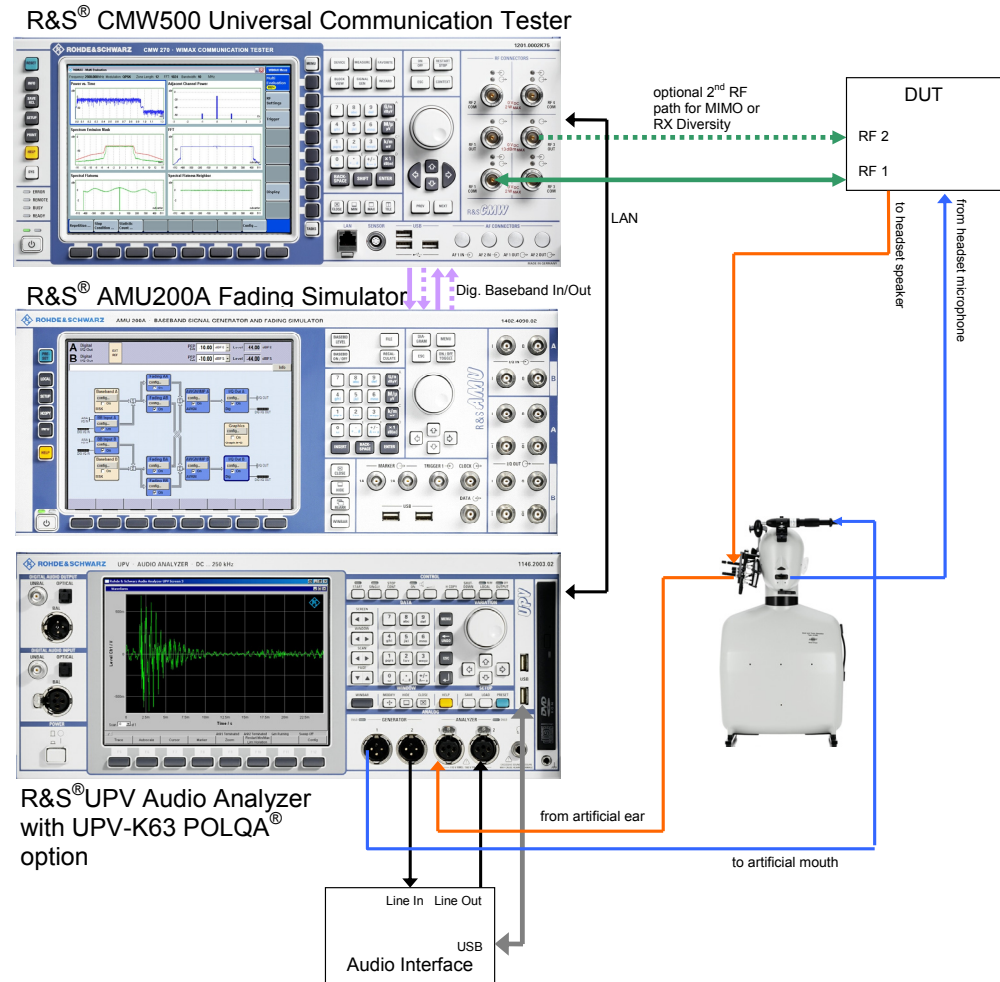


図7：POLQA 試験構成

試験構成は、基地局をシミュレートする R&S®CMW500A テスタ 1 台、オプションの R&S®AMU200A フェージング（チャネル）シミュレータ 1 台、および、POLQA 測定を実行する R&S®UPV オーディオ・アナライザ 1 台から構成されます。

オーディオ・アナライザまたは外部 PC のいずれかで実行される VoIP または IMS（Internet Media Service：インターネット・メディア・サービス）サーバに対してデジタル化したオーディオ・データを送信するために、外部のオーディオ・インタフェースが必要になります。

音響測定のため、人工的な耳と口を持つダミー・ヘッドを使用することができます。電気計測のため、DUT¹⁾ スピーカの出力をオーディオ・アナライザの入力に直接接続し、マイクロフォンの出力をオーディオ・アナライザの出力に直接接続します。

¹⁾ DUT がモバイル端末を表すのに対して、SUT はオーディオ信号の送信チェーン全体（オーディオ・アナライザの出力から入力まで）を意味します。

4.1 ダウンリンクのPOLQA測定

図 8 は、ダウンリンク POLQA 測定の際のアナログおよびデジタルの経路設定を単純化して示しています。

オーディオ・アナライザはアナログの試験信号（通話）を生成し、これがオーディオ・インタフェースの回線入力に送られます。アナログ信号はデジタル信号に変換されて、オーディオ・アナライザ自体か、または外部 PC 上で実行される VoIP または IMS サーバに送信されます。ここから LAN を経由して、

テストに信号が送信されます。

VoIP または IMS でコード化されたデータ・パッケージは、RF を通じてモバイル（DUT）に送信され、ここでデータがデコードされ、アナログ信号に変換されてイヤフォン・プラグに送られます。

この（劣化した）信号がオーディオ・アナライザの入力に渡されます。POLQA アルゴリズムによる MOS-LQO スコアの計算では、最初の（参照）信号とともに、この劣化信号が必要になります。

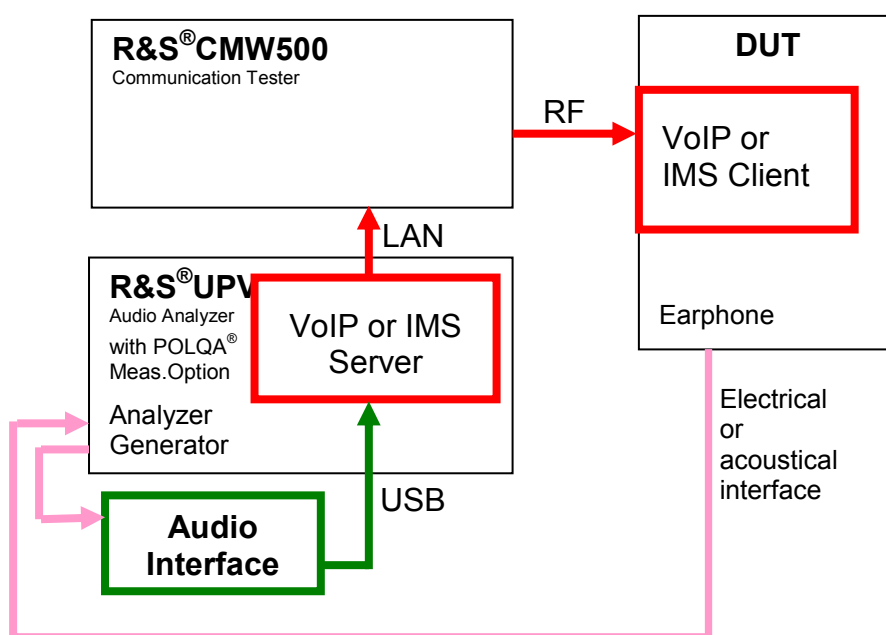


図 8 : ダウンリンクの POLQA 測定

4.2 アップリンクのPOLQA測定

図 9 は、アップリンク POLQA 測定の際のアナログおよびデジタルの経路設定を単純化して示しています。

参照信号（通話）がオーディオ・アナライザのジェネレータからモバイル（DUT）マイクロフォンの入力に送られます。この信号が VoIP または IMS パケットにコード化され、RF 発信用に変調されます。

IMS データ・パケットはワイドバンド無線機テスト内で復調され、LAN を通じてオーディオ・アナライザまたは外部 PC に渡され、そこで VoIP または IMS サーバによってデコードされます。

オーディオ・インタフェースはデジタルの通話データをオーディオ信号に変換します。この信号が、POLQA 測定用にオーディオ・アナライザの入力に送られます。

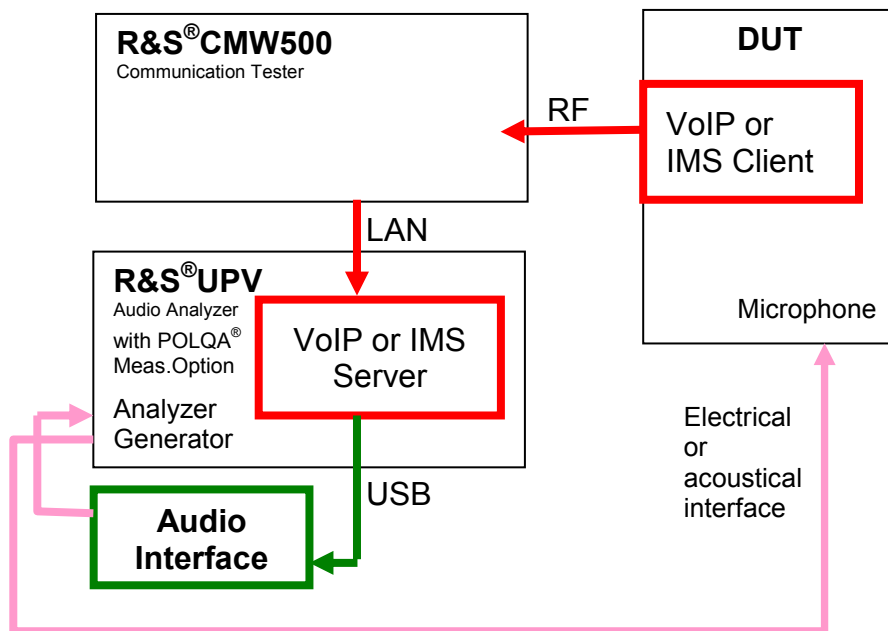


図 9 : アップリンクの POLQA 測定

5 参考文献

- [1] POLQA® Introduction – Jochim Pomy, OPTICOM GmbH
- [2] Draft New Recommendation ITU-T P.863
- [3] Recommendation ITU-T P.862
- [4] United States Patent US 8,032,364 B1
- [5] Application Note 1MA149 – “VoIP Measurements for WiMAX” - Ottmar Gerlach, Rohde&Schwarz GmbH & Co KG
- [6] Application Note 1MA164 – “VoIP PESQ® Measurements for WiMAX with R&S® CMWrun” – Ottmar Gerlach, Rohde&Schwarz GmbH & Co KG
- [7] Psychoacoustics – Facts and Models, E.Zwicker and H.Fastl, Springer Verlag 1990

6 追加情報

ご意見およびご要望は、TM-Applications@Rohde-Schwarz.com までお寄せください。

7 略語

3G	- 3 rd Mobile Generation (第3世代モバイル)
4G	- 4 th Mobile Generation (第4世代モバイル)
AMR	- Adaptive Multirate Codec (適応マルチレート・コーデック)
AMR-NB	- Adaptive Multirate Codec – Narrow Band (適応マルチレート・コーデック – 狭帯域)
AMR-WB	- Adaptive Multirate Codec – Wide Band (適応マルチレート・コーデック – 広帯域)
FFT	- Fast Fourier Transformation (高速フーリエ変換)
IMS	- Internet Media Service protocol used in LTE (LTE で使用されるインターネット・メディア・サービス・プロトコル)
LQO	- Listening Quality, Objective (受聴品質、客観的)
LTE	- Long Term Evolution
MOS	- Mean Opinion Score (平均オピニオン評点)
NB	- Narrowband (狭帯域)
PESQ	- Perceptual Evaluation of Speech Quality (通話品質の知覚的評価)
POLQA	- Perceptual Objective Listening Quality Analysis (受聴品質の知覚的客観的分析)
RMSE	- Root Mean Square Error (二乗平均平方根誤差)
SNR	- Signal to Noise Ratio (信号対雑音比)
SWB	- Super Wideband (スーパー・ワイドバンド)
UMTS	- Universal Mobile Telecommunications System
VoIP	- Voice over Internet Protocol
VoHSPA	- Voice over High Speed Packet Access
VoLTE	- Voice over Long Term Evolution

ローデ・シュワルツについて

ローデ・シュワルツ・グループ（本社：ドイツ・ミュンヘン）は、エレクトロニクス分野に特化し、電子計測、放送、無線通信の監視・探知および高品質な通信システムなどで世界をリードしています。

75 年以上前に創業し、世界 70 カ国以上で販売と保守・修理を展開している会社です。

ローデ・シュワルツ・ジャパン株式会社

本社／東京オフィス

〒160-0023 東京都新宿区西新宿 7-20-1

住友不動産西新宿ビル 27 階

TEL:03-5925-1288/1287 FAX:03-5925-1290/1285

神奈川オフィス

〒222-0033 神奈川県横浜市港北区新横浜 2-8-12

Attend on Tower 16 階

TEL:045-477-3570 (代) FAX:045-471-7678

大阪オフィス

〒564-0063 大阪府吹田市江坂町 1-23-20

TEK 第 2 ビル 8 階

TEL:06-6310-9651 (代) FAX:06-6330-9651

サービスセンター

〒330-0075 埼玉県さいたま市浦和区針ヶ谷 4-2-11

さくら浦和ビル 4 階

TEL:048-829-8061 FAX:048-822-3156

E-mail: info.rsjp@rohde-schwarz.com

<http://www.rohde-schwarz.co.jp/>

Certified Quality System
ISO 9001
DQS REG. NO 1954 QM

Certified Environmental System
ISO 14001
DQS REG. NO 1954 UM

このアプリケーションノートと付属のプログラムは、ローデ・シュワルツのウェブサイトのダウンロード・エリアに記載されている諸条件に従ってのみ使用することができます。

掲載されている記事・図表などの無断転載を禁止します。

おことわりなしに掲載内容の一部を変更させていただくことがあります。あらかじめご了承ください。

ローデ・シュワルツ・ジャパン株式会社

〒160-0023 東京都新宿区西新宿 7-20-1 住友不動産西新宿ビル 27 階

TEL:03-5925-1288/1287 FAX:03-5925-1290/1285

www.rohde-schwarz.co.jp