

スペクトログラムの局所的な相関に基づく音響信号の評価*

☆ 岡崎泰地, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

音響信号処理手法や音響デバイスの改良を図る上では、音響信号の品質を定量的に評価することが重要であり、様々な評価指標が提案されている。しかし、それらの多くは信号全体の品質を単一の値で表すものであり、手法や伝達系が信号に及ぼす影響が、どの時刻・周波数に顕著に表れているかを具体的に把握することは難しい。このような影響を時刻・周波数ごとに定量的に把握できれば、手法やデバイスの改良に有用な指標になると考えられる。そこで本稿では、各時間周波数ビンに対し、振幅の相関と位相のコヒーレンスに基づく局所的な評価値を与える手法を提案し、その性質を実験により確認する。

2 音響信号の評価

音響信号の代表的な評価指標である SNR は、評価対象となる信号とリファレンス信号の差異を時間領域で評価する。また、周波数領域での評価指標としては、LSD (Log-Spectral Distance) がスペクトル形状の差異を評価する指標として用いられる。また、時間周波数領域での評価指標としては、PESQ や ViSQOL が人間の聴覚特性やスペクトル構造を考慮して知覚品質を評価する指標として用いられる。これらの指標は、それぞれの目的に応じた評価を行う一方で、最終的には信号全体を単一の値として表すため、差異が時間周波数平面上のどこに現れたかを把握することができない。加えて、これらの指標は主として振幅に基づく評価を行っており、局所的な波形構造や時間遅延関係を反映する重要な情報である位相を考慮できていない。

3 提案手法

本稿では、時間周波数ビンごとにリファレンス信号との差異を定量的に評価することを目的として、振幅分布の相関と位相差の一貫性に基づく評価指標を提案する。図-1 に提案手法の概要を示す。入力信号 $x(n)$, $y(n)$ に短時間フーリエ変換 (STFT) を適用し、得られた複素時間周波数表現から各時間周波数ビンの評価値を算出する。具体的には、各時間周波数ビンを中心とした領域 (局所領域) を取り出し、その内部で、振幅の相対的な分布パターンを評価するための相

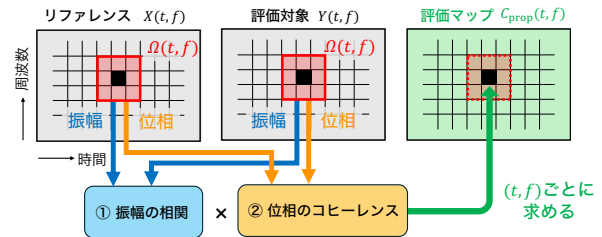


図-1 提案手法の概要。

関 (3.1 節) と、周期量である位相差の一貫性を評価するためのコヒーレンス (3.2 節) をそれぞれ計算する。その後、振幅分布と位相差の双方が一致した領域を高く評価するため、両者の積を計算する。これを時間周波数平面全体について求めることで、局所的な差異を表す評価マップを得る (3.3 節)。

3.1 振幅相関

振幅スペクトログラムから局所領域内の平均や分散を算出すると、振幅値の大きい成分の影響を強く受けるため、小さな成分の差異が反映されにくい。一方、対数圧縮では、振幅値が 0 に近づく値が発散するため、数値的に扱いにくい。そこで本稿では、音圧と知覚ラウドネスの関係がべき乗則で近似されることを踏まえ [1]、振幅スペクトログラムを $|X(t, f)|^\alpha$ によって変換する。ここで、 $\alpha \in [0, 1]$ である。続いて、局所領域 $\Omega(t, f)$ における振幅の平均値を $\mu_X(t, f)$, $\mu_Y(t, f)$, 標準偏差を $\sigma_X(t, f)$, $\sigma_Y(t, f)$, 共分散を $\sigma_{XY}(t, f)$ とする。これらを用いて局所振幅相関を

$$C_{amp}(t, f) = \frac{\sigma_{XY}(t, f)}{\sigma_X(t, f)\sigma_Y(t, f) + \varepsilon} \quad (1)$$

と定義する。ここで、 $\varepsilon > 0$ はゼロ除算を防ぐための微小定数である。 $C_{amp}(t, f) \in [-1, 1]$ は、値が大きいほど、局所領域内における振幅分布の相関が高い。

3.2 位相コヒーレンス

局所領域内で位相差が一定である場合、局所的な波形構造や時間遅延関係が保存されていると考えられる。そこで、本稿では位相差の一貫性を評価するため、局所位相コヒーレンスを

$$C_{phase}(t, f) = \left| \frac{1}{|\Omega|} \sum_{(i, k) \in \Omega} e^{j(\angle X(i, k) - \angle Y(i, k))} \right| \quad (2)$$

と定義する。 $C_{phase}(t, f) \in [0, 1]$ は、局所領域内で位相差が一定であるほど高い値を示す。

* Acoustic signal evaluation based on local correlation of spectrograms. By Taiji OKAZAKI, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

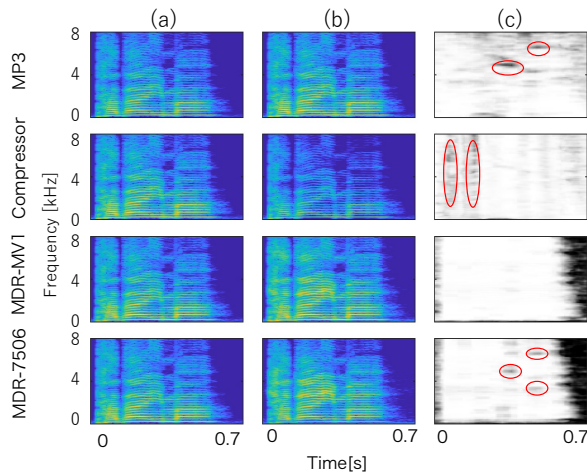


図-2 各条件における局所評価結果. 各行は MP3 圧縮, コンプレッサ処理, MDR-MV1 及び MDR-7506 を示す. 各列は (a) 原音, (b) 処理後信号の振幅スペクトログラム, (c) 局所評価マップである. (c) では白が一致度 1, 黒が 0 を表す.

3.3 評価マップ

振幅相関と位相コヒーレンスの両方が高い領域のみを高く評価するため, それらを統合した評価値を

$$C_{\text{prop}}(t, f) = C_{\text{amp}}(t, f) \cdot C_{\text{phase}}(t, f) \quad (3)$$

と定義する. $C_{\text{prop}}(t, f) \in [-1, 1]$ は, 振幅分布の相関が高く, 位相差が一定であるほど高い値を示す. 各時間周波数ビンについてこの値を算出することで, この評価値を時間周波数平面上に表す評価マップを得る.

4 実験

音響信号処理後及び音響伝達後の信号に対する提案手法の挙動を確認するため, MP3 圧縮, コンプレッサ処理, 及びヘッドホン伝達特性が信号に及ぼす影響を評価した. 評価信号にはサンプリング周波数 44.1 kHz の音声信号を用いた. STFT では窓長及び FFT 長を 2048 点, シフト長を 256 点とし, Blackman 窓を用いた. 局所領域は, 統計量を安定して算出しつつ局所性を保つため, 5×5 ビンとした. べき乗圧縮係数は, Stevens のべき法則における指数を参考に, $\alpha = 0.6$ とした [1]. 処理及び収録による遅延は相互相関に基づいて補正した. 結果を図-2 に示す.

MP3 圧縮では, 入力音声 128 kbps で圧縮し, 圧縮前後を比較した (図-2 第 1 行). 評価マップ (c) では, 全体として高い評価値を保つ一方, 高域を中心とした赤丸で示す領域で評価値が低下した. MP3 圧縮においては, 高域成分が優先的に劣化するため, その影響が現れたと考えられる.

コンプレッサ処理では, 処理前後の信号を比較した (図-2 第 2 行). 評価マップ (c) では, 大部分が高い評価値を維持し, 赤丸で示した音の立ち上がり

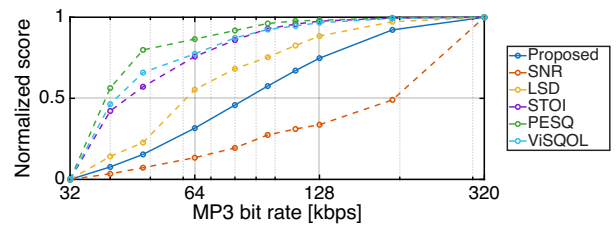


図-3 MP3 ビットレートに対する各評価指標の正規化値. 32-320 kbps における各指標の最小値を 0, 最大値を 1 とした. LSD は値が小さいほど原音に近い, 正規化値の向きを反転した.

付近で評価値が低下した. これは, コンプレッサの非線形処理によって音の立ち上がり付近に歪みが生じ, その影響が現れたためと考えられる.

ヘッドホン伝達特性の評価では, XS LAV Mobile を取り付けた 3D プリンタ製ダミーヘッドを用いて, MDR-MV1 及び MDR-7506 の再生音を収録し, 入力音声と比較した (図-2 第 3, 第 4 行). 評価マップ (c) では, MDR-MV1 は全体に高い評価値を保つ一方, MDR-7506 では赤丸で示した複数の局所領域で評価値が低下した. この結果から, 提案手法はヘッドホン間の伝達特性の違いを局所的な評価値として可視化できることが分かる. また, MDR-7506 では評価値の低下が時間方向にも局在していることから, 周波数応答だけでは説明しにくい伝達特性の違いが反映されている可能性がある.

さらに, MP3 圧縮音声 (32-320 kbps) について評価マップ全体の平均値を評価スコアとし, SNR, LSD, STOI, PESQ および ViSQOL と比較した (図-3). 提案手法の評価スコアはビットレートとともに単調増加し, 他の指標と同様に, 圧縮率の低下に伴って原音との一致度が高くなる傾向を示した. また, 広いビットレート範囲で評価値が段階的に変化し, 圧縮条件の違いを識別できることが確認された.

5 むすび

本稿では, スペクトログラムの局所的な振幅相関と位相コヒーレンスに基づく音響信号の評価指標を提案した. MP3 圧縮, コンプレッサ処理及びヘッドホン伝達を対象とした実験により, 各処理によって生じる局所的な差異が評価マップに反映されることを確認した. 今後は, 振幅相関と位相コヒーレンスの寄与, 局所領域サイズ及びべき乗圧縮係数が評価値に与える影響を検討する.

参考文献

- [1] M. Tsuruta-Hamamura, H. Hasegawa and S. Iwamiya, "Gender differences in judgment criteria using verbal expression might be a principal factor causing gender differences in perceived loudness," *Acoust. Sci. & Tech.*, **47**(4), 334-343 (2026).