信号包絡線を用いた打撃音の客観評価指標の提案とタイムストレッチング性能の評価 *

◎赤石夏輝, 山田宏樹, 矢田部浩平 (農工大)

1 まえがき

タイムストレッチング手法に対して，聴感上の評価と対応する適切な客観評価を行うことは難しい。これは，ストレッチ信号には真の信号が存在せず，信号の波形に基づく性能評価ができないためである。これまで，位相の整合性で評価する指標が提案されている。しかし，従来手法では打撃音劣化に対して適切な評価ができないという課題がある。そこで本稿では，特に打撃音に対するストレッチ性能を評価するための，信号包絡線を用いた客観指標を提案する。

2 タイムストレッチングと打撃音劣化

タイムストレッチングは，信号の音高を変えずに時間スケールを伸長する処理である。特に，位相ボコーダ（PV）に基づく手法 [1] が広く用いられている。

PV に基づくタイムストレッチング手法は，離散ガボール変換（DGT）を応用して信号を伸長する。DGT によって得られた元信号の複素スペクトログラムを $\mathbf{X} \in \mathbb{C}^{M\times N}$ とする。ただし，M は周波数ビン数，N は時間フレーム数である。ここで，ストレッチ信号のスペクトログラム $\mathbf{Y} \in \mathbb{C}^{M\times N}$ を $\mathbf{Y} = |\mathbf{X}| \odot \exp(\mathrm{i}\mathbf{\Phi}_\mathrm{s})$ のように与える。ただし，$|\cdot|$ は要素ごとの絶対値，$\odot$ は要素積，$\mathrm{i} = \sqrt{-1}$ である。また，$\mathbf{\Phi}_\mathrm{s} \in \mathbb{R}^{M\times N}$ はストレッチ信号の位相であり，これは PV によって生成される。最後に，伸長したシフト幅を用いた逆 DGT で $\mathbf{Y} \in \mathbb{C}^{M\times N}$ を時間信号に変換することでストレッチ信号が得られる。

PV は位相を生成する際に信号に正弦波モデルを仮定しているため，正弦波成分のストレッチに適している。しかし，正弦波モデルではパルス成分の位相の構造を考慮できないため，パルス成分に対して適切な位相を生成できない。そのため，PV は打撃音を劣化させてしまうという性質がある。この打撃音劣化を防ぐために，いくつかの手法が提案されている [2–4]。

3 従来の評価指標

タイムストレッチング手法の性能評価では，生成されたストレッチ信号がどれだけ妥当なものであるかを評価する必要がある。しかし，ストレッチ信号には聴感上最も理想的な信号となる真の信号が存在しない。そのため，真の信号の波形からの差を計算するような客観評価を行うことができない。

従来のタイムストレッチング性能の評価指標には，生成された位相の整合性で評価するものがある [1]。位相の整合性は，複素スペクトログラムを逆 DGT し，再び DGT したときに生じる誤差によって確かめられる。一般に，PV が整合性のある位相を与えられていない場合，生成された信号の音は不自然になる。

ストレッチ信号の複素スペクトログラム $\mathbf{Y}$ を逆 DGT し，再び DGT したものを $\mathbf{Z} \in \mathbb{C}^{M\times N}$ とおく。このとき，位相の整合性を評価する指標 D は

$$D = \frac{\sum_{n=C}^{C+P-1}\sum_{m=0}^{M-1}\left(|Y[m,n]| - |Z[m,n]|\right)^2}{\sum_{n=C}^{C+P-1}\sum_{m=0}^{M-1}|Y[m,n]|^2} \quad (1)$$

のように計算される。ただし，C は評価する区間の開始位置の時間インデックス，P は評価する区間の時間フレーム数である。

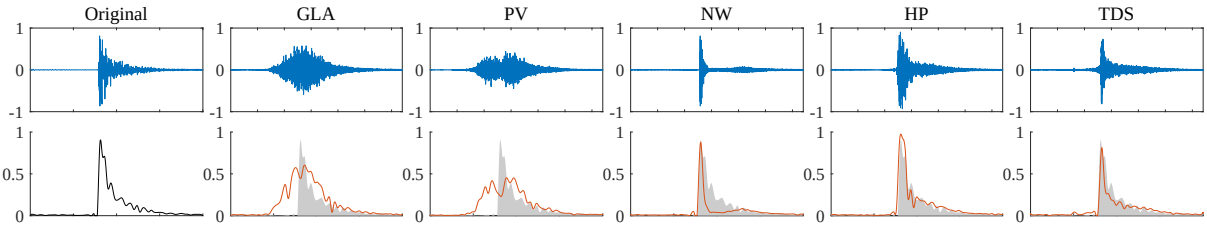
しかしながら，位相の整合性が保たれていても実際の信号の音質が良いとは限らない。この例に当てはまるのが，タイムストレッチングで劣化した打撃音である。打撃音のストレッチにおいて，整合性はあるが適切ではない位相が PV によって与えられてしまう。このとき，位相の整合性で評価する従来の手法では，実際の聴感と対応した評価ができない。したがって，従来手法は特に打撃音劣化を防ぐタイムストレッチング手法を適切に評価できないという課題がある。

4 提案手法

提案手法では，時間波形の概形を用いた評価を行う。これは，打撃音として適切な位相が与えられている場合には時間信号の波形にアタックや減衰などの打撃音らしい特徴が現れることに基づいている。理想的なタイムストレッチングでは，元信号とストレッチ信号の間で時間スケールだけが異なり，波形の概形は一致する。そのため，ストレッチされた信号の概形と元信号の概形の近さを調べることで，タイムストレッチング手法の打撃音劣化を防ぐ性能を評価することができると考えられる。

提案手法では，信号の概形をとるために信号包絡線を用いる。これは，元信号とストレッチ信号の間で時間スケールが異なっていても概形をうまく抽出でき

* Singal-envelope-based objective index for evaluation of time stretching methods. By Natsuki AKAISHI, Koki YAMADA, and Kohei YATABE (Tokyo University of Agriculture and Technology).



図–1　各時間信号の波形（上段）とその絶対値の包絡線（下段）。それぞれのストレッチ信号の絶対値の包絡線の図には，元信号の絶対値の包絡線（左下段）が囲む範囲を灰色で示している。

るからである。ここで，対象となる打撃音は変化が急峻な信号である。そこで，変化が急峻であっても滑らかな概形を抽出できるピーク包絡線を用いる。ピーク包絡線は，信号を短く区切り，各区間の 0 以外の極値をスプライン関数によって結ぶことによって与えられる包絡線である。そして，包絡線によって得られた波形の概形を比べるためにスケール不変な SNR を計算する。これによって，処理の途中で信号の振幅のスケールが変わっても適切に信号を比べられる。

提案する評価指標 B は，ストレッチ信号の絶対値の包絡線と元信号の絶対値の包絡線をそれぞれ $\mathbf{v}$，$\hat{\mathbf{v}}$，また $\alpha = \hat{\mathbf{v}}^{\mathsf{T}}\mathbf{v}/\|\mathbf{v}\|^2$ として以下のように計算する。

$$B = 10\log_{10}\left(\frac{\|\alpha\mathbf{v}\|^2}{\|\alpha\mathbf{v}-\hat{\mathbf{v}}\|^2}\right) \tag{2}$$

ここで，タイムストレッチングを行うと信号を伸長した分だけサンプル数が増えるため，そのままだと $\mathbf{v}$ と $\hat{\mathbf{v}}$ の長さが異なり SNR を計算できない。そこで，信号の特徴を保持したまま長さを揃えるために，ストレッチ信号の絶対値の包絡線をリサンプリングする。

5　実験

提案手法の妥当性を確かめるために，ストレッチした打撃音の客観評価の比較を行った。タイムストレッチングを行うアルゴリズムとして，位相復元手法の Griffin-Lim Algorithm（GLA），広く使われている手法の PV [1]，そして打撃音劣化を防ぐ手法の NW [2]，HP [3]，TDS [4] を用いた。評価のための単純な打撃音として，TSM toolbox から 22 050 Hz でサンプルされたボンゴの音源を用いた。評価に用いた音源のサンプル数は 4000 サンプル（181 ms）である。評価は 3.2 倍にストレッチした信号に対して行った。

図–1 に元信号とストレッチ信号の時間信号の波形と包絡線を示す。GLA および PV では打撃音劣化が起こっており，信号のアタックの部分が鈍くなっていることが見てとれる。一方で，NW，HP，TDS では打撃音劣化が緩和されており，信号のアタックが立っている。また，NW は HP，TDS と比べて信号の減衰部分が保持できていない。これらの特徴は包絡線にも現れていることが見て取れる。

表–1　各手法に対する客観指標の値の比較。D は値が小さいほど良く，B は値が大きいほど良い。各評価指標で最も良いスコアが太字で示されている。

	GLA	PV[1]	NW[2]	HP[3]	TDS[4]
D (↓)	**0.06**	0.19	0.49	0.95	0.33
B (↑)	1.23	-0.74	3.77	**10.7**	10.4

これらの事実を踏まえて，表–1 に示される客観評価の結果を見ていく。なお [4] の主観評価実験では，GLA を除く 4 つの手法は評価が高い方から TDS, HP, NW，PV の順になっていた。従来の指標は GLA を特に良く，打撃音劣化を防ぐアルゴリズムを悪く評価している。対して，提案する指標は HP と TDS を良く，打撃音劣化が起こっている GLA と PV を悪く評価している。ここから，従来法と比べて提案手法が打撃音を妥当に評価できているといえる。さらに打撃音劣化を防ぐ手法での結果を見ると，提案する指標では HP，TDS より NW を悪く評価している。これは [4] による主観評価実験と対応しており，提案手法は打撃音の特徴をより保持できるものを評価できる手法であるといえる。

6　むすび

本稿ではタイムストレッチング手法の打撃音劣化に対する頑健性を評価する客観指標を提案した。提案手法では，元信号とストレッチ信号の包絡線を抽出し，スケール不変な SNR を計算することで包絡線を比較する評価を行った。実験では，提案手法による評価の妥当性が示された。今後はタイムストレッチング以外の評価に提案手法を適用する。

参考文献

[1] J. Laroche and M. Dolson, “Improved phase vocoder time-scale modification of audio,” *IEEE Trans. Speech Audio Process.*, **7**(3), 323–332, (1999).
[2] F. Nagel and A. Walther, “A novel transient handling scheme for time stretching algorithms,” *J. Audio Eng. Soc.*, (2009).
[3] J. Driedger, M. Müller, and S. Ewert, “Improving time-scale modification of music signals using harmonic-percussive separation,” *IEEE Signal Proc. Lett.*, **21**(1), 105–109, (2014).
[4] N. Akaishi, K. Yatabe and Y. Oikawa, “Improving phase-vocoder-based time stretching by time-directional spectrogram squeezing,” in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 1–5, (2023).