

打撃音の瞬時性を保つ高倍率タイムストレッチングによる スローモーション再生*

☆ 三浦旺太, 赤石夏輝, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

スポーツ映像の観戦やフォーム解析のためにスローモーション再生が用いられる。このとき、映像に合わせる音響信号もスロー再生される。ただし、音のスロー再生では音高が下がり、印象が変化するため、実際の現象と音からの感覚に違いが生じる場合がある [1]。音高低下を防ぐ方法として、音高を保ったまま音響信号を伸ばす処理であるタイムストレッチングが有効である。しかし、通常的手法による高倍率の時間伸長では打撃音の瞬時性が損なわれ、不自然な音が発生される。スポーツ競技で発生する音には、ボールの衝突音、足音、着地音等の打撃音が多く含まれており、これらは動作の強度やタイミングに関する情報を含むため、その聴感を損なわない時間伸長が求められる。そこで本稿では、スポーツ映像のスロー再生に向けた、打撃音の瞬時性を保つ高倍率タイムストレッチングを提案し、主観評価実験によって性能を評価する。

2 スローモーション再生における映像と音

映像のスローモーション再生は、撮影時と再生時のフレームレートを変えることで実現できる (図-1)。例えば、1 秒あたり 60 枚の画像を撮影した映像を、1 秒あたり 30 枚ずつ再生すると、映像は半分の速度で再生される。このとき、各フレームの画像情報 (被写体の見た目や質感) は変化せず、動作の時間経過のみを引き伸ばして観察できる。

一方で音響信号は、再生速度を単純に落とすと音高が下がってしまう。例えば、図-1 下部の赤線のように、1 秒あたりの再生サンプル数を半分にすると、周波数が元信号の半分になるため音が低くなる。このように、映像では被写体の見た目が維持されるのに対し、音では聴感に変化する。この音高の変化を抑え、図-1 下部の青線のような理想的な伸長信号の生成を目指す手法がタイムストレッチングである。

3 タイムストレッチング

タイムストレッチングは、音響信号の音高を維持したまま時間長を伸長する処理である。一般的な手法の概念図を図-2 左に示す。まず、図上部の入力信号を一定長の区間で重複させながら切り出す。次に、

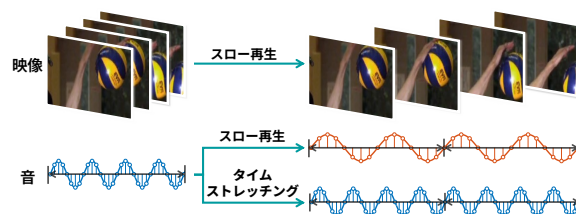


図-1 単純なスロー再生とタイムストレッチングの比較。

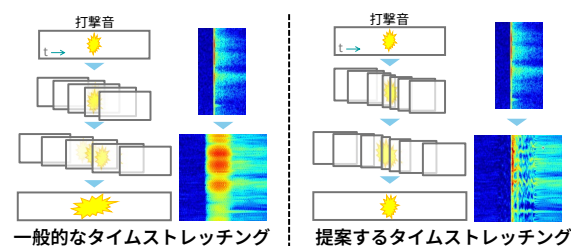


図-2 一般的なタイムストレッチング (左) と打撃音付近で切り出し区間を短くしたタイムストレッチング (右) の模式図と振幅スペクトログラム。

中段のように、切り出した区間の間隔を広げて再配置し、重なり合う部分が滑らかに繋がるように信号を操作しながら加算する。このとき、区間を配置する間隔の広がり時間が時間伸長率に対応する。以上の操作において、各区間での周波数分布は変わらないため、音高を維持したままの時間伸長が可能となる。従来のタイムストレッチング手法は、音楽や音声のような正弦波成分を多く含む信号を主な対象として設計されている。そのような信号では、切り出し区間をある程度長く設定することで良好なストレッチ信号が得られることが知られている [2]。しかし、打撃音に対して同様の処理を適用すると、図-2 左下段のように打撃音が時間方向に広がり、時間的な局在性が失われる。

4 打撃音の劣化を防ぐタイムストレッチ

従来のタイムストレッチングでは、区間の再配置時の広がりや区間同士を滑らかに繋ぐための操作 (位相復元) による劣化により打撃音が瞬時性を失う (図-3)。本稿ではこれらに対し、打撃音周辺において切り出し方を変える適応切り出し (図-2 右) と、正弦波モデルを仮定しない位相復元手法である Griffin-Lim アルゴリズム (GLA) [3] を導入した手法を提案する。適応切り出しで再配置時の広がりや劣化範囲を抑制し、GLA で区間接続時の打撃音劣化を低減する。

*Slow-motion audio playback with percussive-sound-preserving time stretching. By Outa MIURA, Natsuki AKAISHI, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

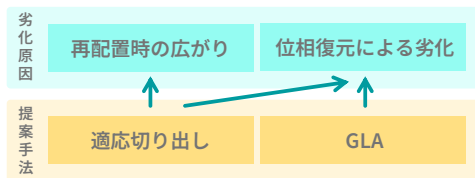


図-3 打撃音の劣化原因と提案手法の関係図。

4.1 適応切り出し

区間を重ねながら切り出すとき、時間的に局在する打撃音は、図-2 左のように複数の区間によって観測される。このとき、時間伸長に伴う区間の再配置によって、各区間に含まれる打撃音が時間的に分散して配置されるため、打撃音の時間的局在性が低下する。また区間長が長い場合、位相復元の影響により区間内で打撃音が時間方向に広がってしまう。

これらの問題に対処するために、打撃音周辺での区間の切り出し方を変更する。打撃音を含む区間を近接させて再配置し、さらに各区間内での信号の広がりの影響を抑えるために、図-2 右に示すように、適応切り出しでは打撃音周辺のみシフト幅と区間長を短くする。このとき、区間長の不連続な変化によるアーチファクトを防ぐために、区間長を滑らかに変化させる。これにより、図-2 右下部のような、打撃音の時間的局在性を保った信号が得られる。

4.2 タイムストレッチングのための位相復元手法

タイムストレッチングでは、切り出した各区間を周波数領域で表し、区間の配置に合わせて各成分の位相を操作し、滑らかに繋ぐ。代表的な手法の位相ボコーダ (PV) は、各成分を持続する正弦波とみなし、隣接区間で位相を伝播させる。一方、広帯域の成分が同時に生じる打撃音では帯域ごとの成分間の同期が崩れ、時間方向に成分が広がる。打撃音の瞬時性を保つ手法として提案された SELEBI [2] では、打撃音とそれ以外を選択的に扱い、打撃音の位相構造を考慮して位相を生成する。一方、特定の信号モデルを仮定しない GLA は、時間領域での信号生成と時間周波数領域への変換を反復し、各区間が整合する位相を推定する。本稿では、適応切り出しに GLA を組み合わせる。

5 主観評価実験

提案手法によって高倍率伸長したスポーツ動作時の音響信号の自然さを評価する実験を行った。実験には、撮影・録音した3種類の映像・音から作成した、4倍・8倍のスローモーション動画を用いた。撮影対象の一例として、各動画から切り出したフレームを図-4に示す。タイムストレッチングの比較手法には PV, DGT+GLA, SELEBI を用いた。DGT+GLA は一定切り出しと GLA による位相生成を用いた手法



図-4 撮影した動画から切り出したフレーム。

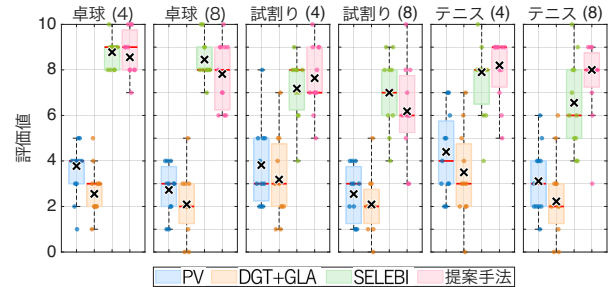


図-5 主観評価の結果。カッコ内の4・8は4倍・8倍ストレッチングを意味する。×は平均値を示す。

である。実験には正常な聴覚を有する20代男女11名が参加した。時間伸長時に生じる音高低下の参考例として、初めに元動画および単純なスロー再生動画を提示した。その後、各手法によって時間伸長した動画について、音高の保持や打撃音と映像との時間的な一貫性を含めた総合的な観点から、映像と音がどの程度自然に対応していると感じるかを、0-10点の整数値で評価するよう求めた。

結果を図-5に示す。適応切り出しで信号を扱う SELEBI や提案手法では全項目で高い評価を得た。項目によって SELEBI と提案手法の評価値の大小関係が異なる傾向が見られたことから、音源により適した位相生成方法が存在すると考えられる。また、DGT+GLAと比較して提案手法の評価値が大幅に高かったことから、位相推定手法が同一であっても、信号の切り出し方法が評価に大きく影響することが示された。

6 むすび

本稿では、適応切り出しと GLA によって打撃音の瞬時性を保持するストレッチ手法を提案し、スポーツ動作の音響信号を用いた主観評価によって有効性を確認した。今後は、適応切り出しに適した位相生成法を検討し、ストレッチ音の聴感品質の向上を目指す。

参考文献

- [1] R. L. Klatzky, D. K. Pai and E. P. Krotkov, "Perception of material from contact sounds," *Presence*, **9**(4), 399-410 (2000).
- [2] N. Akaishi, N. Holighaus and K. Yatabe, "SELEBI: Percussion-aware time stretching via selective magnitude spectrogram compression by nonstationary Gabor transform," *IEEE Trans. Audio Speech Lang. Process.*, **34**, 3754-3766 (2026).
- [3] Y. Masuyama, K. Yatabe and Y. Oikawa, "Griffin-Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Process. Lett.*, **26**(1), 184-188 (2019).