

非定常ガボール変換を用いて打撃音劣化を低減する タイムストレッチングの実験的評価*

◎ 赤石夏輝 (農工大), Nicki Holighaus (ÖAW), 矢田部浩平 (農工大)

1 はじめに

タイムストレッチングは信号のピッチを変えずに時間スケールを伸長する処理である。位相ボコーダ(PV)に基づく手法が広く用いられるが、この手法には打撃音が劣化する課題がある。我々はこれまで、打撃音成分の時間スケールを保持するように時間周波数表現の振幅を時間方向に収縮する手法を提案した[1, 2]。特に非定常ガボール変換(NSDGT)を用いた手法[2]は、信号を完全再構成できるため、高品質なストレッチ信号が得られる。本稿では、この手法の有効性を客観評価と主観評価によって確認する。

2 打撃音を保持するタイムストレッチング

PVに基づくタイムストレッチングは、図-1の上段に示すように、離散ガボール変換(DGT)によって元信号(左)の時間周波数表現の振幅(中央)を得て、PVによってストレッチ後の位相を生成し、逆DGTによってストレッチ信号(右)を合成する3段階の処理を行う。位相生成においてPVは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信号は適切に得られる。一方で、パルス成分に対しては適切な位相を生成できず、打撃音が劣化する課題がある。

打撃音劣化の主な要因として、振幅と位相の不整合がある。理想的な位相生成が行われれば、原理的には信号を時間的に局在化させる位相が得られる。しかし、振幅自体は図-1の上段中央のように時間方向に広がる。そのため、位相生成の精度に関わらず、ストレッチ信号の打撃音成分が時間的に広がってしまう。

この振幅と位相の不整合を低減するために、我々はこれまで振幅(図-1中央)を処理する手法DIALGA[1]及びSELEBI[2]を提案した。DIALGAは、振幅のビンを打撃音の中心フレームに向けて移動させることで時間スケールを保持する。しかしその処理は経験的であり、ストレッチ信号の合成の安定性に課題があった。対してSELEBI[2]は信号の変換にNSDGTを導入した手法であり、窓関数を打撃音成分の周辺で適応的に変更することで、図-1中央下段のような、打撃音成分の時間スケールを保持した振幅を直接得ることができる。NSDGTを用いることによって、その性質に基づく信号再構成を行えるため、高品質なストレッチ信号が得られる。

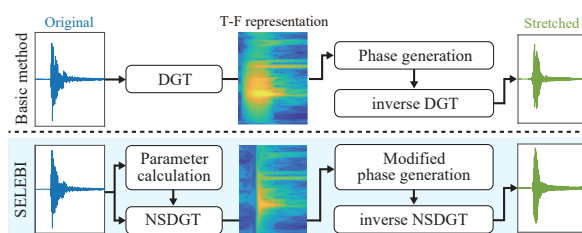


図-1 PVに基づく通常のタイムストレッチング(上段)とSELEBI(下段)のフローチャート。

3 合成信号を用いた客観評価実験

理想的なストレッチ信号(真の信号)を利用できる合成信号を用いた実験を行った。5つの特徴的な信号に対して、2倍及び4倍のストレッチングを行った。使用した信号は(A)インパルス、(B)1000 Hzの正弦波とインパルスを含む信号、(C)複数の正弦波(1000, 2000, 3000 Hz)とインパルスを含む信号、(D)50 Hzの正弦波を指数減衰させた打撃的信号、(E)その打撃的信号と1000 Hzの正弦波を含む信号である。

比較手法はPV, PVDR [3], PVHP [4], DIALGA [1], SELEBI [2]である。DIALGAとSELEBIでは、位相生成にPV及びPVDRを用いる場合を評価した。以降、用いたPVをDIALGA(PVDR)のように表記する。パラメータやサンプリング周波数は[2]と同じとした。評価には、真の信号のDGT係数とストレッチ信号のDGT係数の振幅の相対誤差[5]を用いた。

結果を表-1に示す。SELEBI(PVDR)は2倍及び4倍のストレッチの両方で、ほぼ全ての信号で最小またはそれに次ぐ小さな誤差を示した。特に(D)と(E)のより複雑な信号でSELEBI(PVDR)の誤差が他の手法よりも大幅に小さかった。よって、SELEBIは複雑な信号に対しての有効性が高いことが示された。

(D)の4倍ストレッチ信号と、振幅の各ビンごとの相対誤差を図-2に示す。SELEBIは他の手法に対して誤差が小さく、打撃音が発生直前の部分に広がったり、発生時間が大きくずれたりしなかった。よって、SELEBIはより正確に打撃音を保持できると言える。

4 実信号を用いた主観評価実験

TSM toolboxの7つの音源からそれぞれ1sを切り出したサンプルに対し、2倍及び4倍のタイムストレッチングを行った。比較手法はPV, PVDR, DIALGA(PVDR), SELEBI(PVDR)とした。

*Experimental evaluation of percussion-preserving time stretching using non-stationary Gabor transform.
By Natsuki AKAISHI (Tokyo University of Agriculture and Technology), Nicki HOLIGHAUS (Austrian Academy of Sciences) and Kohei YATABE (Tokyo University of Agriculture and Technology).

表-1 合成信号を用いた実験における客観評価指標の比較。最高スコアを太字に下線, 2 番目に良いスコアを太字で示した。

		PV	PVDR	PVHP	DIALGA (PV)	DIALGA (PVDR)	SELEBI (PV)	SELEBI (PVDR)
2 倍	(A) Impulse	0.2357	0.2402	0.6993	0.1009	0.1110	0.2326	0.2326
	(B) Sinusoid + Impulse	0.0113	0.0091	0.0130	0.0090	0.0092	0.0106	0.0051
	(C) Harmonic Sinusoid + Impulse	0.0129	0.0107	0.0139	0.0136	0.0107	0.0128	0.0098
	(D) Transient	0.5806	0.5666	0.5924	0.5659	0.5647	0.5791	0.5552
	(E) Sinusoid + Transient	0.2506	0.2446	0.4244	0.2455	0.2445	0.2521	0.2414
4 倍	(A) Impulse	0.3080	0.3080	1.3679	0.0559	0.0559	0.3069	0.3069
	(B) Sinusoid + Impulse	0.0220	0.0149	0.0242	0.0267	0.0168	0.0206	0.0151
	(C) Harmonic Sinusoid + Impulse	0.0298	0.0264	0.0307	0.0315	0.0270	0.0334	0.0265
	(D) Transient	1.3835	1.3153	1.3838	1.4153	1.3097	1.3812	1.2266
	(E) Sinusoid + Transient	0.4746	0.4468	0.6175	0.4585	0.4407	0.4719	0.4316

表-2 実信号を用いた主観評価実験における 3 つの指標の平均スコアの比較。最高スコアを太字に下線で示した。

		総合的な品質				アタックの鋭さ				調波成分の明瞭さ			
		PV	PVDR	DIALGA	SELEBI	PV	PVDR	DIALGA	SELEBI	PV	PVDR	DIALGA	SELEBI
2 倍	Bongo	14.98	82.48	73.28	94.18	19.62	85.92	72.88	92.31	15.71	77.22	69.11	85.72
	CastanetsViolin	28.77	72.73	80.69	78.78	10.01	69.17	71.35	75.38	50.33	72.27	81.55	78.32
	DrumSolo	17.43	56.48	58.07	64.49	10.56	54.93	54.49	59.88	22.74	57.35	51.17	57.69
	Glockenspiel	33.57	67.24	63.93	74.21	29.48	63.26	63.73	73.26	40.28	67.59	66.47	74.66
	Jazz	33.57	68.26	63.71	68.58	28.56	62.05	61.80	68.65	33.92	67.63	64.59	62.09
	Pop	29.93	78.76	76.29	74.61	21.30	73.34	74.17	69.54	35.32	77.95	74.62	76.45
	Stepdad	30.78	72.09	68.63	68.25	26.71	71.79	69.92	63.98	30.71	67.04	67.83	61.62
Average		27.00	71.15	69.23	74.73	20.89	68.64	66.91	71.86	32.71	69.58	67.91	70.93
4 倍	Bongo	7.19	67.83	69.07	87.59	5.89	73.46	74.68	88.47	13.14	61.25	67.37	80.75
	CastanetsViolin	25.23	71.86	67.95	74.03	9.56	69.66	66.41	73.51	41.17	78.10	72.79	74.43
	DrumSolo	5.29	42.33	44.11	57.56	3.61	40.08	39.60	58.01	11.75	44.03	48.53	53.67
	Glockenspiel	26.96	70.55	63.28	79.05	20.24	65.85	60.62	79.68	36.01	67.56	66.00	77.46
	Jazz	23.67	58.26	56.04	78.64	16.57	58.55	52.70	79.35	27.98	57.49	57.60	73.16
	Pop	21.01	57.78	60.79	76.13	15.34	54.14	58.68	79.50	23.15	57.82	62.00	73.99
	Stepdad	17.44	66.55	64.32	63.97	9.72	63.11	63.76	61.62	26.23	68.00	62.07	62.23
Average		18.11	62.17	60.79	73.85	11.56	60.69	59.49	74.31	25.63	62.04	62.34	70.81

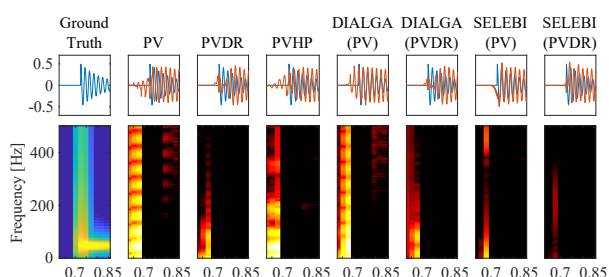


図-2 (D) の時間信号 (上段) と DGT 係数の振幅の各ビンごとの相対誤差 (下段)。横軸は時間 (s) である。上段は真の信号 (左) を青, ストレッチ信号を赤で示した。下段左には真の信号の振幅を示し, 黄色いほど振幅が大きい。下段左以外は暗いほど相対誤差が小さい。

実信号には真の信号が存在せず, スコアの基準となる参照信号を用意できない。そこで, そのような場合に適した ITU-R BS.2132-0 に基づく客観評価実験を行った。実験には正常な聴覚を有する 10 名が参加した。各手法の具体的な性質を調べるため, 「総合的な品質」, 「アタックの鋭さ」, 「調波成分の明瞭さ」の 3 つの項目で評価を行った。それぞれの項目を 0-100 ポイントまでの連続的なスコアで評価した。

結果を表-2 に示す。SELEBI は「総合的な品質」「アタックの鋭さ」において, ほぼ全ての条件で従来手法のスコアを大きく上回った。特に 4 倍のストレッチで

は, 「総合的な品質」で約 10 ポイント, 「アタックの鋭さ」で約 14 ポイント向上した。また, 「調波成分の明瞭さ」は 2 倍のストレッチでほぼ変わらず, 4 倍のストレッチで約 8 ポイント向上した。よって, SELEBI が調波成分の聴感上の品質を損なわずに打撃音成分の聴感上の品質を向上できることが示された。

5 むすび

本稿では, 我々の手法 SELEBI の有効性を客観評価と主観評価で確認した。SELEBI は実信号に対しても聴感上の品質が高いストレッチ信号が得られた。今後は実応用に向けて計算コスト削減に取り組む。

参考文献

- [1] N. Akaishi, K. Yatabe and Y. Oikawa, “Improving phase-vocoder-based time stretching by time-directional spectrogram squeezing,” *IEEE ICASSP*, pp. 1–5 (2023).
- [2] 赤石夏輝, N. Holighaus, 矢田部浩平, “非定常ガボール変換を用いて打撃音劣化を低減するタイムストレッチング,” 日本音響学会講演論文集, pp. 1465–1466 (2025.9).
- [3] N. Holighaus and Z. Průša, “Phase vocoder done right,” *EUSIPCO*, pp. 976–980 (2017).
- [4] J. Driedger, M. Müller and S. Ewert, “Improving time-scale modification of music signals using harmonic-percussive separation,” *IEEE SPL*, **21**(1), 105–109 (2013).
- [5] E. S. Ottosen and M. Dörfler, “A phase vocoder based on nonstationary Gabor frames,” *IEEE/ACM TASLP*, **25**(11), 2199–2208 (2017).