

MIDIデータのベロシティを 異なる音源に適応させる試み

北原 鉄朗 片寄 晴弘
(JST CrestMuse / 関西学院大学)

研究の背景

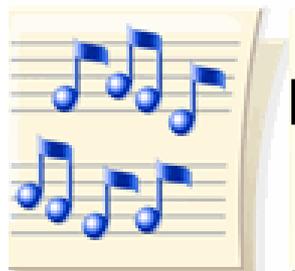
MIDI (Musical Instrument Digital Interface)

- 電子楽器間で演奏データを転送するためのインターフェース
- 音楽データのフォーマットとしても重宝
 - ファイルサイズが小さい
 - 編集・加工などの二次利用が容易
- 問題点
 - 音源が変わると音のニュアンスが変わる

音源が変わると音のニュアンスが変わる

Prog#1

Vel.=70



MIDIファイル

音源A

音源B



X社のピアノ



Y社のピアノ



音源が変わると音のニュアンスが変わる

- 音色（サンプリング元の音源）が違う

⇒ やむを得ない

- 音の大きさが違う

⇒ 別の音源で演奏のニュアンスを再現するには、
ベロシティを微調整する必要がある



この(半)自動化が目的

スペクトル誤差最小化基準に基づく ベロシティの自動調整

音源A用に作成されたMIDIファイルの
ベロシティを、音源B用に自動調整する

- MIDIファイル中のベロシティ: $v = (v_1, \dots, v_k)$
- 音源Aで再生したときの音響特徴: $x_A(v)$
- 音源Bで再生したときの音響特徴: $x_B(v)$
- 元のベロシティ v_0 に対して、 $x_A(v_0)$ と $x_B(v)$ が
できるだけ近くなるように v を決定

$$\hat{v} = \operatorname{argmin}_v d(x_A(v_0), x_B(v))$$

音響特徴

- パワースペクトル(窓幅4096点、ハミング窓)
- ログパワースペクトル
- 各倍音のパワー
- 各倍音のログパワー
- MFCC(12次まで)

いずれもフレーム(10ms)毎に算出して平均化

距離尺度

- 平均二乗誤差
- コサイン距離

最適化手法

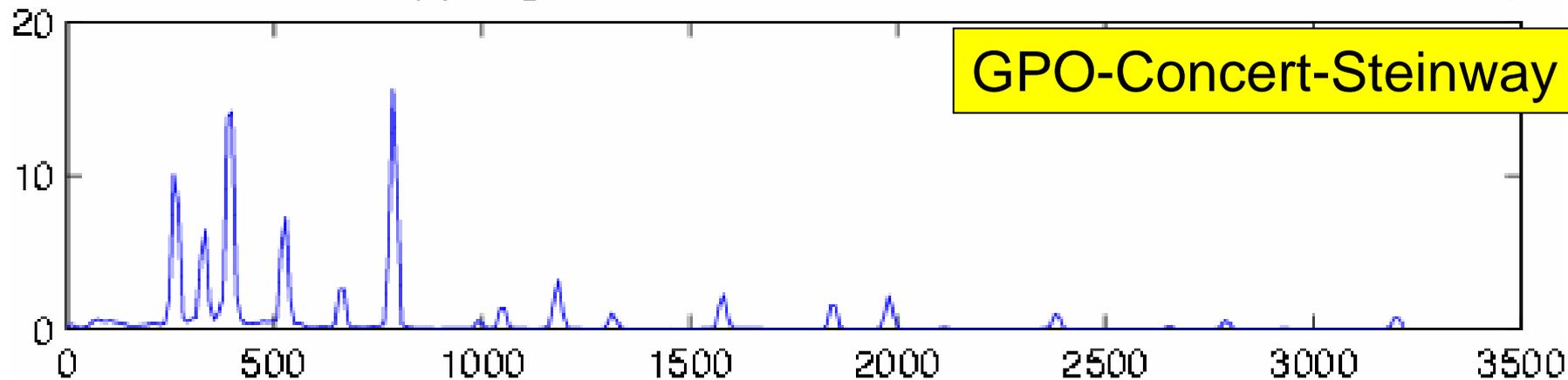
- シミュレーテッドアニーリング
 1. 元のベロシティ v_0 を v の初期値に設定
 2. $d(x_A(v_0), x_B(v))$ を算出し、エネルギー E とする
 3. v を更新し、 v' とする
 4. $d(x_A(v_0), x_B(v'))$ を算出し、エネルギー E' とする
 5. If $E' < E$ then v' を採用
else 確率 $P = \exp((E - E')/T)$ で v' を採用
 6. 温度パラメータ T を更新 ($\alpha < 1$ を乗算) し、
 $T = 0$ になるまで繰り返す

実験

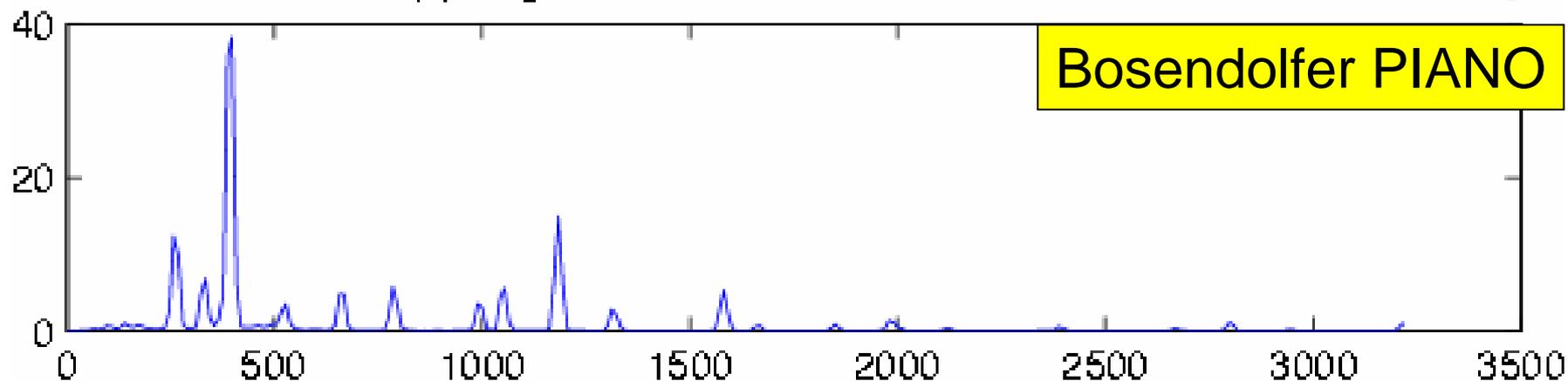
- 使用MIDIデータ
 - C4 - E4 - G4 の3和音 (ベロシティ: 70 - 50 - 100)
- 使用音源 (ピアノのサウンドフォント)
 - 音源A: GPO-Concert-Steinway-ver2.sf2
 - 音源B: Bosendorfer PIANO / GIGA (Crypton)

元のベロシティでのスペクトル

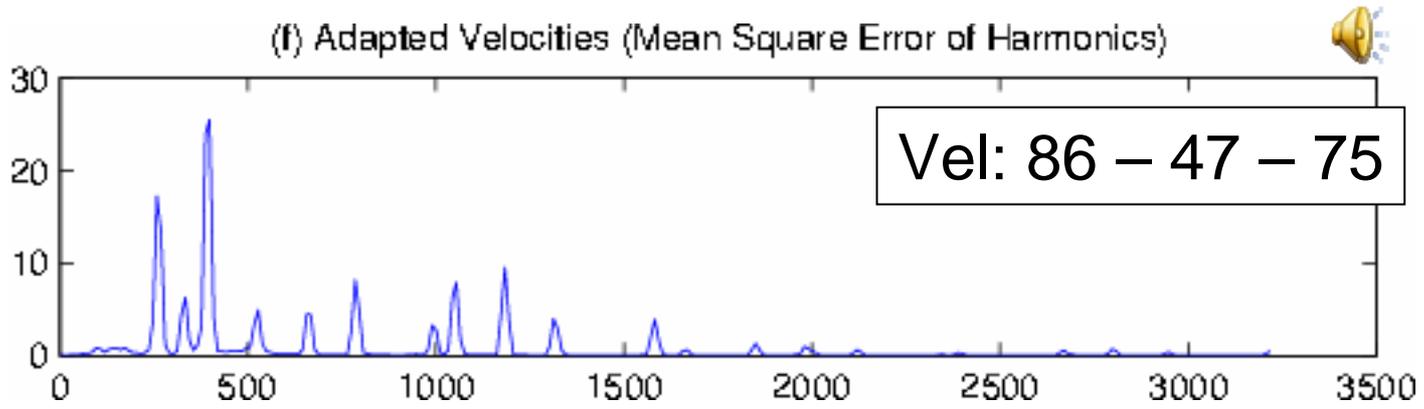
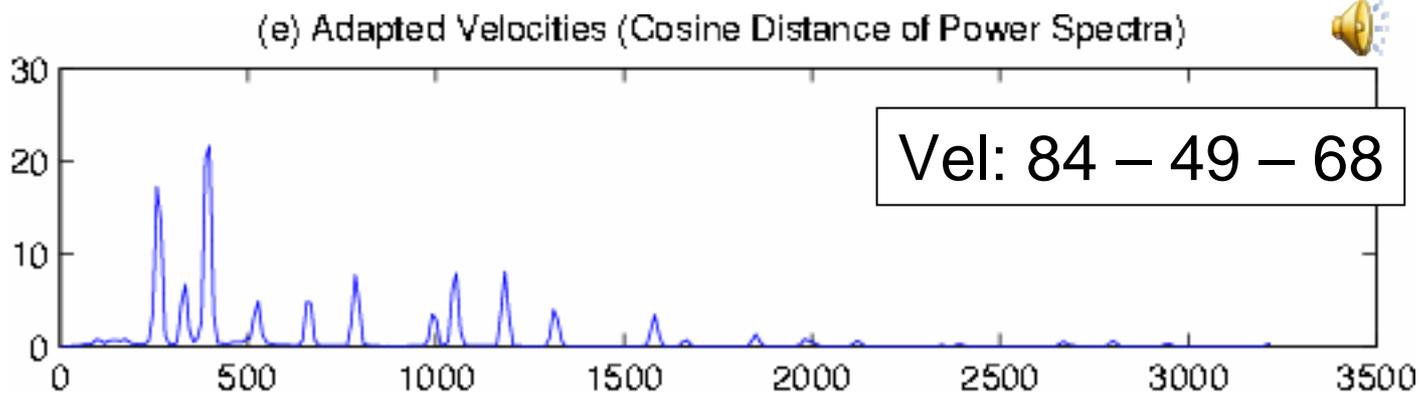
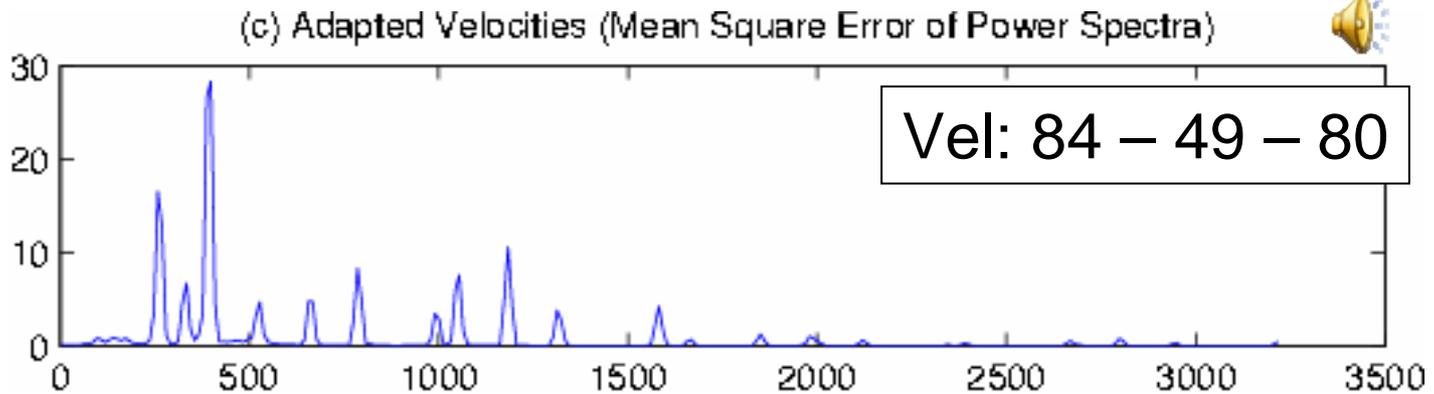
(a) Original Velocities with Tone Generator A



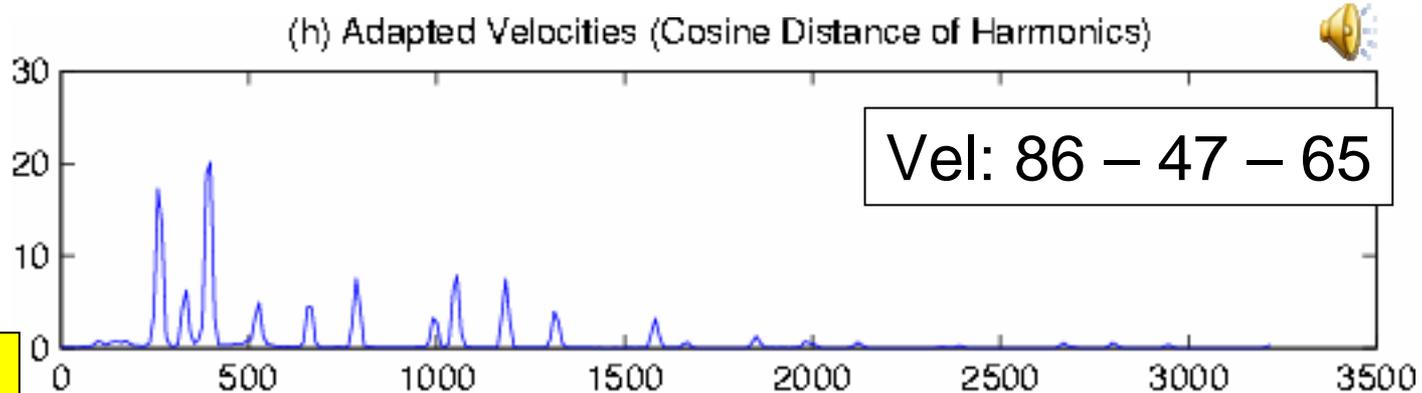
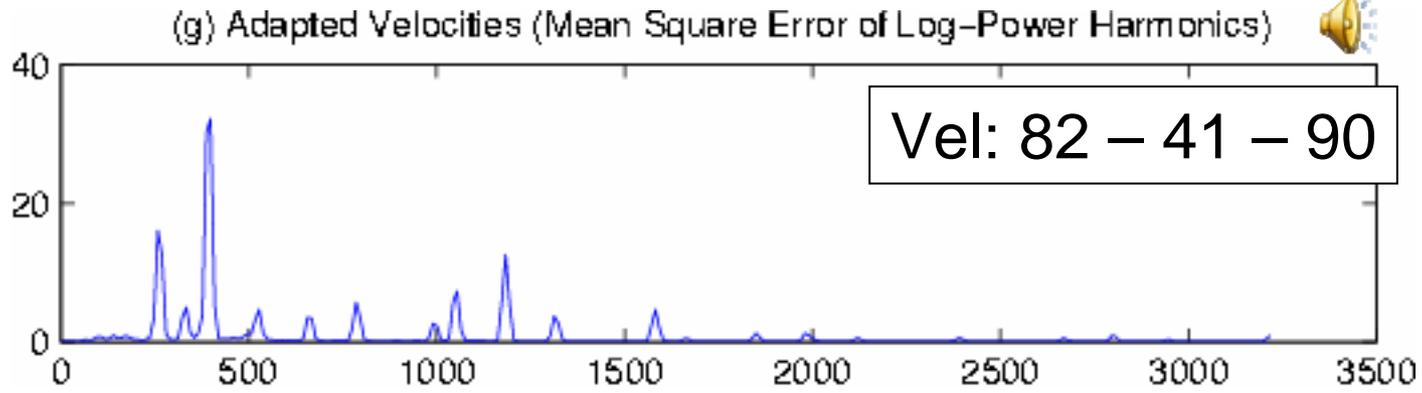
(b) Original Velocities with Tone Generator B



実験結果1



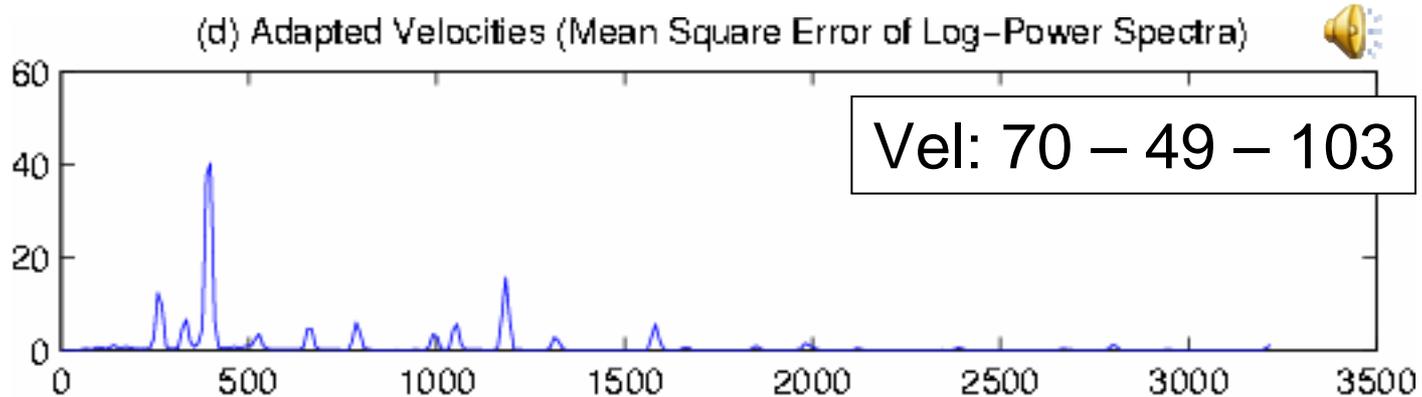
実験結果2



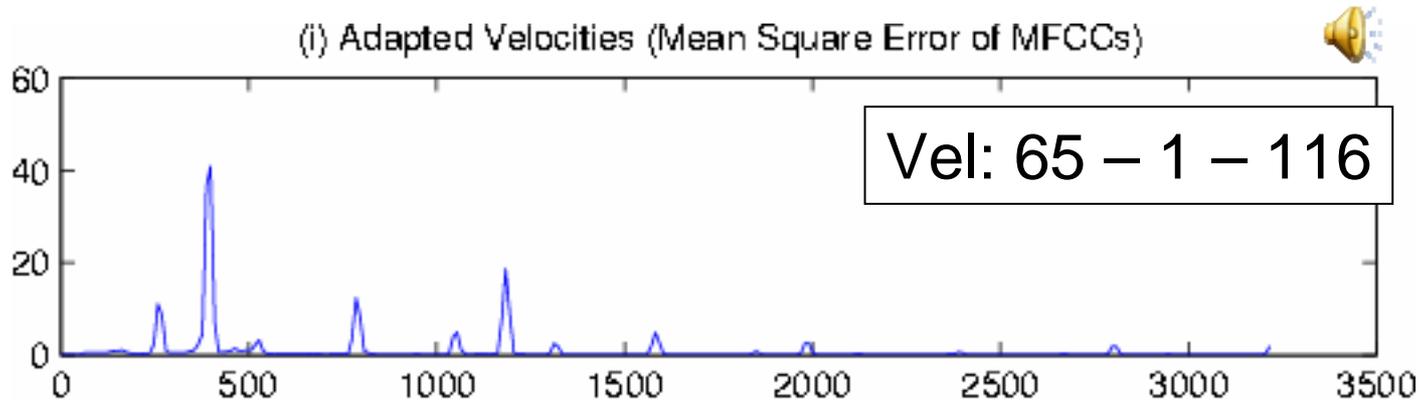
共通点

- G4のベロシティが減少
 - 音源BのG4の基音成分が非常に大きいため
- C4のベロシティが増加
 - 音源BのG4の第2倍音の小ささをC4の第3倍音でカバー?

実験結果3



- 初期値からほとんど変化なし
 - LogPowerは雑音の影響を受けやすいから？



- E4がほとんど無音に
 - MFCCは粗すぎて適切な評価ができない？

まとめ

【目的】音源A用に作成されたMIDIファイルをできるだけよく再現するようにベロシティを自動調整

【手法】パワースペクトルなどの音響特徴が最も近くなるベロシティをシミュレーテッドアニーリングで探索

【結果】パワースペクトルを近づけることはできたが、聴感上似ているとは言えない

【今後の課題】

- 聴感上の類似に対応する音響特徴の探究
- 聴取実験など