

調波音・打楽器音分離手法を用いたギター・ベースギターの自動採譜

Automatic Transcription of Guitar/Bass Guitar Part Using Harmonic/Percussive Sound Separation Method

木村 拓也

Takuya Kimura

法政大学情報科学部デジタルメディア学科

E-mail: 10k1015@stu.hosei.ac.jp

Abstract

This paper describe a system to support the work of transcription beginner instrument by output the TAB music and staff notation of guitar and bass guitar from music acoustic signal. First, the system separate Harmonic/Percussive sound to increase the accuracy of detection, such as chord, f0 and tempo. Next, the system detects chord using constrained by bass pitch and Chroma vector. Also, the system detects f0 of bass line and guitar solo using a filter which passes the f0 and harmonics. In this method, using constrained by detected chords and bass pitch distribution used each chords obtained by pre-learning. As a result, the accuracy chord and f0 of bass/guitar solo detection is improved. Finally, the system converts obtained chord and f0 of bass/guitar solo data to MIDI file, and the score is output by import MIDI file to free composer software. Songs of Research subject does not include change of rhythm and modulation, and the tempo is constant. The average recognition rate of the automatic transcription for six songs; bass was 80 %, chord was 94 %, guitar solo was 80 %. As a result of comparing the transcription accuracy with and without use of the present system are shown to demonstrate the effectiveness of this system.

1 まえがき

楽曲を練習する際には、楽譜を用いるのが一般的である。しかし、最新の楽曲やマイナーな楽曲は楽譜が出版されていないことがあり、そのような場合には楽曲の音源を聴いて採譜する。だが、この作業には相対音感やリズム感といった音楽的感覚や楽典などの知識が必要なため、楽器の未経験者や初心者には難しい。そこで、自動採譜を行うことで、楽器初心者の練習の補助を行うシステムを構築する。

自動採譜とは、音楽の音響信号波形から人間の代わりにコンピュータを使って楽譜を書き起こさせるプロセスである。楽曲の自動採譜は、和声認識 [1] や拍節認識 [2], リズム認識 [3] といった、分解された要素問題ごとに様々な研究がなされているが、解析から五線譜及び TAB 譜の出力までを全て行う研究はされていない。また、自動採譜のフリーソフトウェアも存在するが、全てのパートがひとつのピアノロールで表示される形式のものも多く、楽譜としては機能していない。

そこで、本稿では音楽音響信号から、楽器初心者でも理解しやすい五線譜及び TAB 譜の出力を行うシステムを提案する。研究対象とする楽曲は、楽曲全体を通してテンポが一定であり、

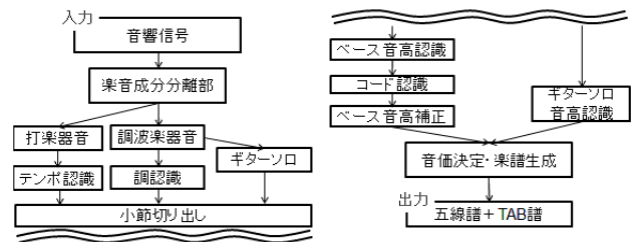


図 1. システム全体の処理のフローチャート

変拍子や転調を含まないものとする。

2 システム

2.1 概要

図 1 にシステム全体の処理のフローチャートを示す。前処理として音源分離部で、スペクトログラムの滑らかさの異方性に基ついた調波・打楽器成分分離手法 [4] で調波・打楽器成分を分離し、また音楽の繰り返し構造除去によるメロディ抽出手法 [5] でギターソロ抽出を音源分離部で行う。分離した調波・打楽器音をそれぞれ調認識・テンポ認識に利用し、求めたテンポを利用して小節切り出しを行う。

音高認識部では、各モジュール単体での解析では情報量が少ないという問題を解決するために、音楽知識を基にした補正をモジュール間で行うことで解析精度を向上させる。まず基本周波数と倍音を通過させるフィルタ (倍音フィルタ) を用いて認識したベース音高を行う。次に、クロマベクトルと求めたベース音高を利用したコード認識を行う。そして、各コードのベース音高分布を事前学習させ、求めたコードのベース音高分布を考慮し、ベース音高認識結果の補正を行う。ギターソロ認識は、ギターの音階に対応させた倍音フィルタを用いて行う。解析結果は、楽譜生成を行うために、楽譜へのアライメントを考慮して MIDI に変換する。それをフリーソフトにインポートして五線譜と TAB 譜を出力した。出力した楽譜は演奏者が練習に用いて修正可能である。

2.2 楽音成分分離部

2.2.1 メディアンフィルタリングによる調波・打楽器成分の分離

■手法の概要 音高認識部の精度向上のため、FitzGerald の手法を用いて調波・打楽器成分の分離を行った。この手法では、調波楽器、打楽器の成分がそれぞれ時間方向、周波数方向に滑らかである性質を利用し、混合音のスペクトログラムを各方向に細かく分割したフレームに対しメディアンフィルタ処理を行うことで各楽器成分を強調する。メディアンフィルタ処理が行われたスペクトログラムの 2 つの結果は、元のスペクトログラムに適用するウィナーマスクを生成するために使用される。本稿では、スペクトログラムをサンプリングレート 8000Hz, フ

レーム幅 512 点, シフト幅 5 点で求めた.

■**メディアンフィルタ処理** 混合音のスペクトログラムを各方向に分割すると, 周波数方向では調波楽器の倍音構造が, 時間方向では打楽器のオンセットがそれぞれピークとして表れる. こららのピークを外れ値と見なし, 式 (1) に定義したメディアンフィルタを掛けて平滑化することにより, 打楽器と調波楽器の成分を抽出する. ここでは, $n = 17$ で計算を行った.

$$y(n) = \text{Median}\{x(n-k : n+k), k = (l-1)/2\} \quad (1)$$

(1) 式のメディアンフィルタを各方向に分割したスペクトルに対し適用することで, 調波・打楽器成分をそれぞれ強調する. ここで, S_i, S_h, M はそれぞれ i 番目の周波数方向のスペクトル, h 番目の時間方向のスペクトル, メディアンフィルタ処理である.

$$P_i = \mathcal{M}\{S_i\} \quad (2)$$

$$H_i = \mathcal{M}\{S_h\} \quad (3)$$

■**ウィナーフィルタに基づくマスクの生成** ノイズが加算された信号に対してウィナーフィルタを用いることで元の信号を推定することができる. 元の信号のパワースペクトルを H , ノイズのパワースペクトルを P としたとき, ウィナーフィルタは次のように設計される.

$$WF = \frac{H}{H+P} \quad (4)$$

Derry はウィナーフィルタの各値を二乗することでフィルタの効果を高めている. そのフィルタを元のスペクトログラムの各要素と乗算することにより調波・打楽器成分の推定値を得る.

$$\hat{H} = \frac{H^2}{H^2+P^2} \otimes S, \quad \hat{P} = \frac{P^2}{P^2+H^2} \otimes S \quad (5)$$

2.2.2 楽音の繰り返し構造部除去手法を利用したギターソロ抽出

Pardo らが開発した音楽の基本的な特徴である繰り返しの着目し, 音楽と声に分離する手法を用いてギターソロ抽出を行う. この手法は, 音楽の伴奏はある程度の間隔の繰り返し構造を持つが, メロディーはその特性上繰り返し構造を持たないという点を利用した分離方法である. まず, 音響信号スペクトラム V の自己相関で周期を推定する. $B(i, j)$ は V^2 の i 番目の周波数成分を j 単位分ずらしたときの周波数成分である. b を式 (6) に示す.

$$B(i, j) = \frac{1}{m-j+1} \sum_{k=1}^{m-i+j} V(i, k)^2 V(i, k+j-1)^2$$

$$b(j) = \frac{1}{n} \sum_{i=1}^n B(i, j) \quad (6)$$

次に伴奏のベースとなるスペクトログラムを推定する. b の最初の極大値までの点数から求めた音響信号の周期を p とし, 長さ p のセグメントでスペクトログラムを切り出し, その幾何平均を取る.

$$\bar{V}(i, l) = \left(\prod_{k=1}^r V(i, l + (k-1)p) \right)^{\frac{1}{r}} \quad (7)$$

式 (7) を用いて伴奏のスペクトログラムを推定する. 式 (8) は i 番目の周波数成分のスペクトログラム V が \bar{V} の何倍であるかを計算したものである.

$$\tilde{V}(i, l + (k-1)p) = \left| \log \left(\frac{V(i, l + (k-1)p)}{\bar{V}(i, l)} \right) \right| \quad (8)$$

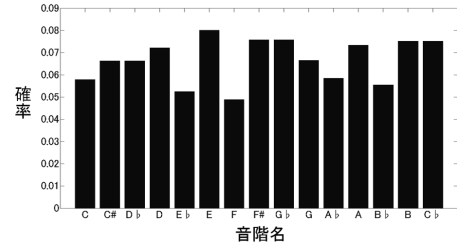


図 2. 調の候補の例 (調が E の場合)

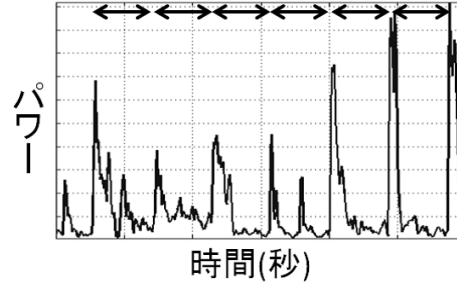


図 3. サンプル音源のパワーとビート間隔

この比率がある程度以下の周波数成分は伴奏部分で, それ以外はメロディーやギターソロが含まれている. この閾値を 0.8 とし, パイナリマスクを作成し, それを元のスペクトログラムに掛けることでギターソロを抽出した.

2.3 調認識

クロマベクトルを用いた調認識を行う. 音階名で表される基本情報 (例えば「C3」) はオクターブ (「3」) と半音階名 (「C」) を指定することで一意に定めることができる. このうちオクターブ違いの同じ音階の成分を全て足しあわせて, 1 オクターブ内の半音階の 12 音の強さを表したものがクロマベクトルである.

i 番目の音階名の j 番目のオクターブのパワーを $C(i, j)$ とすると, クロマベクトル $v = [v_1, v_2, \dots, v_{12}]$ は式 (9) のように計算できる.

$$v_i = \sum_{j=1}^7 C(i, j) \quad (9)$$

2.2 で分離した楽曲全体の調波楽器音に対し, サンプリングレート, フレーム幅, シフト幅をそれぞれ 8000Hz, 512 点, 5 点でスペクトログラムを計算し, 100 セントごとに各音階を区切り, 式 (9) に則りクロマベクトルを求めた. 次に, 求めたクロマベクトルからパワーの大きい 7 音を抽出する. この 7 音と, 各調の構成音 7 音との相関が最も大きかった調を, 楽曲の調と推定する. また, 推定結果は図 5 のように視覚化されるので, ユーザーが結果が誤りだと判断したら, これを頼りに調を再選択する.

2.4 テンポ認識・小節切り出し

楽譜生成には音価と小節の区切りを決定するために曲のテンポを認識することが必要不可欠である. そのため, 分離した打楽器成分のパワーのピークに対し自己相関を用いて音量変化の周期を検出することによって, ビート間隔の候補を推定しテンポを決定する. テンポの認識は楽器初心者でも比較的簡単に行えるため, 検出されたテンポが誤っていると判断した場合には候補の中から正しいテンポを選択し結果を修正する.

また, 小節の開始点を探すために, 認識したテンポからビート間隔でフィルタを作成し, そのフィルタを 1 点ずつずらしながらパワーとの内積を求めた. 開始点の候補を複数検出し, 実際にユーザーに切り出した音声聞かせ, 開始点を選択させる.

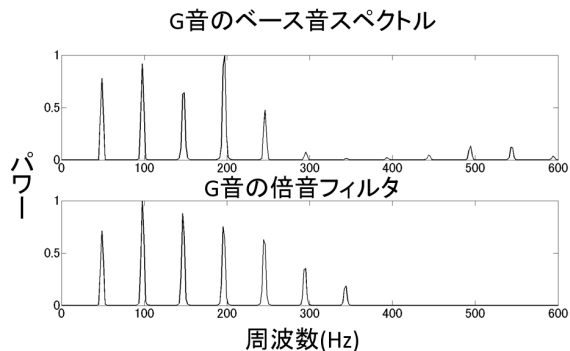


図 4. ベースのスペクトルと倍音フィルタ

2.5 音高認識モジュール

2.5.1 問題の定式化

n 個の小節に分割した信号に対して、各小節で 1 つのコードと、小節をさらに m 個に分割した各時刻でのベース音高を認識する。その際のコード系列を $C = [c_1, c_2, \dots, c_n]$ とし、各 C_n に対するベース音高系列を $B_n = [b_1, b_2, \dots, b_m]$ とする。 B_n を求める際には、各フレームでスペクトル特徴 s と、事前学習させた各コードで使用されるベース音高分布 L の 2 つを用いる。これらから、ベース音高系列に関する事後確率 $P(B|s, L)$ を考え、与えられた s 及び L に対する最大事後確率 (MAP) 推定問題としてベース音高系列認識を定義する。すなわち、 $\operatorname{argmax}_B P(B|s, L)$ が求めるべきベース音高系列である。

2.5.2 各コードで使用されるベース音高分布の学習

各コードで使用されるベースの音高は、音楽理論的観点からある程度定めることができる。よって、各コードで使用されるベース音高分布を事前学習で求め、ベース音高系列認識に利用する。まず、コードが一つのみ使われる解析区間を設定し、各解析区間でのコードの正解データを既存の楽譜から作成した。ベース音高については、TAB 譜共有サイト UltimateGuitar で信頼度の高いもののベースパートの MIDI データや、既存の楽譜から手入力した MIDI データを利用した。学習させる際には、調によって各コードの位置付けが変わってしまう点を解消するために、各調のトニック音を基準にし、そこから何音ずれるかを計算する。そして、各コードの区間でベースの音高と音価を計算し、音名を表す 12 次元の配列に音価を足していくことで、12 次元のヒストグラムからなる分布を作成する。学習させたベース音高分布 L はメジャー、マイナーの 24 種である。学習に用いた楽曲は、洋・邦楽の rock, pops のジャンルから抜粋した 14 曲である。

2.5.3 倍音フィルタを用いたベースライン推定

ベースラインを推定する場合、自己相関や通常の FFT のスペクトルから求める方法では、周波数帯域が重なりやすいバスドラムなどの影響を大きく受けてしまう。そこで、下図のように、ベースで使用される基本周波数とその 7 次までの倍音からなるフィルタ HF を $C0$ から $B2$ までの 36 種生成する。倍音の次数は、実際のベース音のスペクトルを観察して決定した。

ベース音高候補系列 B は倍音フィルタ HF を周波数スペクトル s に適用した際に、最も値が大きくなるものを探せば良いので、 $P(s|HF)$ が最大になる HF を求める。ここで、 $P(s|HF)$ は多項分布だとすると、その確率は式 (10) で表される。ここで、 ω は各周波数ビンを表し、倍音の次数は実験結果から一番精度の良かったものを使用した。

$$P(s|HF) = P(s|B) = \prod_{\omega} (HF_{\omega})^{s_{\omega}} \quad (10)$$

2.5.4 コード認識

コード認識を行うために、ベース音高とクロマベクトルを用いた手法を提案する。2.5 で切り出した区間クロマベクトルを

2.4 と同様求める。そこからパワーの強い 3 音を求めることで、コードの候補を推定する。さらに、ベースはコードが遷移する際、一音目にコードのルート音を弾くという音楽的知識に基づき、小節の頭の 8 分音符の区間で求めたベース音高と楽曲の調から定まるコードを、クロマベクトルに 1/区間数の重みを足すことによってコードを推定した。

2.5.5 最大事後確率の計算

これらから、 $\operatorname{argmax}_B P(B|s, L)$ を計算する。

$$P(B|s, L) = P(B|s)P(B|L) \quad (11)$$

ここで、 $P(B|s)$ についてベイズの定理から次式が成り立つ。

$$P(B|s) = \frac{P(s|B)P(B)}{P(s)} \quad (12)$$

この時、 $P(s)$ は定数、 $P(B)$ は一様分布で、ベース音高系列 B は倍音フィルタ HF と等価であるので、次式が成り立つ。

$$P(B|s) \propto P(s|HF(B)) \quad (13)$$

式 2 の右辺の対数を取り、式 (10)~(13) を用いて B を求める。

$$\operatorname{argmax}_B \left(\sum_{\omega} s_{\omega} \log HF(B) + \log P(B|L) \right) \quad (14)$$

2.6 ギターソロ音高認識

2.2.2 で抽出したギターソロ音源に対し、300Hz のハイパスフィルタを掛ける。そして、2.5.3 の HF を $C1$ から $E5$ までの 52 種とした。

2.7 楽譜生成

楽譜生成を行うために、フリーソフトウェアにインポートすることで五線譜及び TAB 譜出力が容易に出来る・ソフトウェア上での譜面編集も可能な点から、MIDI-Tool*1 を使用して解析結果の MIDI 変換を行う。変換には、ノートナンバー・発音時刻・消音時刻を指定する必要がある。各区間で認識したコードは、指定した音価で出力する。ベース音については、フレームごとの解析結果を、指定した音価分の長さでモードを取り、音高を一意に定める。ギター・ソロについては、16 分音符単位でモードを取り音高を一意に定めたあと、音高の変化がない区間は、その区間で 1 つの音符が出力されるように音価を調整した。

3 評価実験

3.1 認識率調査

6 つの楽曲に対し、評価実験を行った。ベース・ギターソロの音高については、オクターブ検出ミスは許容し、楽譜の音階名のズレを誤検出とし、休符やスタッカートなどで音が鳴っていない区間は無視する。また、本稿で取り扱うコードは、メジャー、マイナーの 24 種類とし、その他のコードはこれらのコードの派生コードとし、sus4 や aug はメジャーに、m7 や dim はマイナーに分類する。認識率を式 (15) で求め、最小音符長は 16 分音符とする。

$$\text{正解率} = \frac{\text{正しく認識された音符数}}{\text{全音符数}} \quad (15)$$

ベース音、コード、ギターソロの平均正解率はそれぞれ 80%、94%、80% となった。コードが誤認識を起こす原因として、ベースがコードのルート音以外を弾く on コードの場合では、コード認識部で利用している音楽知識のみではカバーできずに、コードの誤認識を起こしてしまう。on コード以外の部分で

*1 <http://kenschutte.com/midi>

表 1. 実験に用いた楽曲

曲番号	曲名 (アーティスト)
1	To be With You(Mr.BIG)
2	衝動 (B'z)
3	罫 (THE BACK HORN)
4	everything back but you(Avril Lavigne)
5	バスロマンス (チャットモンチー)
6	DAYBREAK'S BELL(L'Arc en Ciel)

表 2. 正解率 (%)

曲番号	ベース	コード	ギターソロ
1	85	100	なし
2	76	100	なし
3	63	100	なし
4	87	88	61
5	84	100	90
6	84	75	90

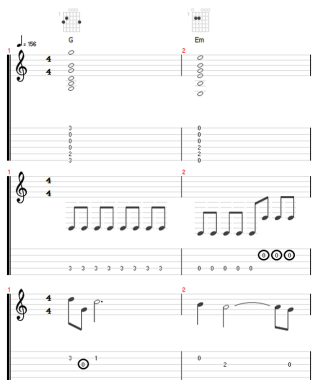


図 5. 出力された 6 番の楽曲の楽譜 (上から順にコード, ベース, ギターソロ, 丸で囲った音符は誤認識である)

はコードの誤認識はなかった。ベースの音高認識については、楽曲による差はあまり大きくなく、本稿の手法が対象となるポップスやロックのジャンルに対応できることがわかった。ギターソロについては、音源抽出の精度が大きく影響しており、特に 4 番の楽曲ではギターのバックアップパートの成分が除去できず、バックアップパートの音高を誤認識する問題があった。

3.2 有効性調査

本システムが楽器初心者の採譜支援に有効であるか確認するために、ギタリスト 4 人、ベーシスト 4 人の計 8 人の被験者を用いた評価実験を行った。評価には認識率調査を行った 4~6 番の楽曲を用い、各楽曲で制限時間を 15 分に定め、本システムの出力した楽譜を参照する X グループと何も参照せず自力のみで採譜する Y グループ、それぞれ 4 人ずつの間での平均エラー率を比較した結果を図 6 に示す。ただし、ここではコード認識については、派生コードをメジャー・マイナーとして認識することは不正解とし、楽譜生成の際の音価は 8 分音符に指定した。結果として、X グループは Y グループに比べて平均エラー率が 59 % 減少し、本システムが楽器初心者の採譜支援として有効であることが確認できた。また、被験者の採譜結果やコメントから、ベースについては、主にルート音を弾いている楽曲は 7 割程度の正解率であればシステムの出力譜面の誤認識部分を判断しやすいため、スムーズな採譜が行えた。しかし、音高が細かく変化するフレーズでは、誤認識部分の判断が難しく、効果的な支援には繋がらなかった。ギターコードに関しては、音色が大きく歪んでいると出力譜面との聴き比べが困難になる為、正解を出力していてもそれを誤認識と判断するケースがあった。これらは解析結果の信頼度が表示されないことが原因に挙げられる。

3.3 あとがき

本研究では、楽器初心者の練習支援のためのギター・ベースギターの自動採譜システムを構築した。コード認識やテンポ抽出、音高推定の性能向上のために調波音・打楽器音成分分離、ギ

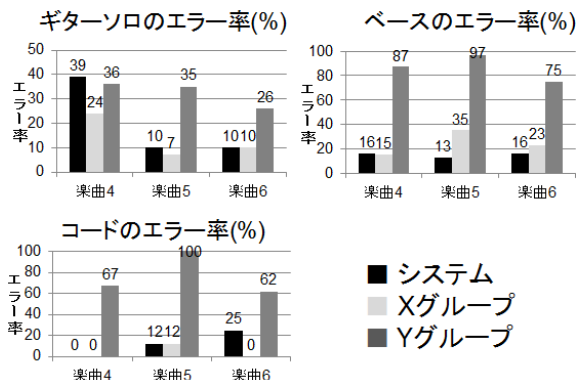


図 6. 本システムの利用の有無による採譜エラー率の比較

ターソロ抽出を行った。調波楽器成分に対し、クロマベクトルを用いた調認識を行った。打楽器成分に対してパワーのピークに対し自己相関を用いてテンポの認識を行った。音高認識部では、音楽知識に基づく制約をモジュール間で利して、倍音フィルタを用いたベース・ギターソロ音高認識、クロマベクトルと求めたベース音高を音楽知識に基づく制約に利用したコード認識、事前学習させた各コードのベース音高分布を用いたベース音高補正を行った。これらの情報を MIDI に変換し、フリーソフトにインポートして五線譜と TAB 譜を生成した。6 曲の楽曲に対し自動採譜を行ったところ、コード、ベース、ギターソロの平均認識率はそれぞれ 94 %, 80 %, 80 % であった。また、本システムの出力した楽譜を参照する被験者 4 人と何も参照にせず自力のみで採譜する被験者 4 人で採譜を行ったところ、本システムを利用した際には採譜のエラー率が 59 % 減少し、本システムが楽器初心者の採譜支援に有効であることを確認した。

今後、本システムの精度及び採譜支援効果を高めるアイデアとして次のものが挙げられる。今回事前学習で用いた楽曲は、研究対象となる楽曲の条件を満たしているものをランダムに選んだが、各ジャンルやアーティストによって使用されるベースの音高は特徴がある。特にアーティストによるフレーズの違いは弾き手に依存するものなので、採譜したい楽曲のアーティストの過去の曲があればそれを事前学習に用いることで、そのアーティストに特化した採譜が行える。コード認識モジュールでは、on コードや 1 音目にベースがルート音以外を弾くフレーズに対応できるより一般的な音楽知識を制約条件にすることで、より多くのフレーズに対応した採譜ができる。楽譜生成の際に信頼度を表示することによって、ユーザーはシステムが出力した楽譜の誤認識部分を判断しやすくなるため、効果的な採譜支援が行える。

参考文献

- [1] 蔵内, 他, "周波数スペクトルの谷状点に基づくコード推定", FIT2008, pp.5-8, August 2008.
- [2] 後藤, 他, "音響信号を対象としたリアルタイムビートトラッキングシステム: コード変化検出による打楽器音を含まない音楽への対応", 信学論, D-II, 情報・システム, II-情報処理 J81-D-2(2), 227-237, 1998-02-25
- [3] 大槻, 他, "隠れマルコフモデルによる音楽リズムの認識", 情処学論, pp.245-255, February 2002.
- [4] Derry FitsGerald, "HARMONIC/PERCUSSIVE SEPARATION USING MEDIAN FILTERING", DAFX10, Graz, Austria, September 6-10, 2010.
- [5] Rafii, Z., Pardo, B., "A simple music/voice separation method based on the extraction of the repeating musical structure", ICASSP, pp.221-224, 2011