サッカー・ラジオ実況音声のピッチ列パターン照合によるゴール時検出*

岩永羊平, ☆新井翔太, 堂元健太郎, 宇津呂武仁 (筑波大)

1 はじめに

近年の科学技術の発達により，音声認識技術が身近な存在になりつつある．最新の音声認識技術は，Siri やしゃべってコンシェル等のアプリ，あるいは，国会議事録の作成等において実利用されている．ここで，音声認識技術を利用することによって，人間の発話を文字に書き起こすことは実現されつつあるが，発話内容の理解や発話意図の把握，発話に含まれる感情の認識といった，計算機にとってのより発展的な課題の解決にはまだほど遠いのが現状である．これらの発展的課題の中でも，本研究においては特に，発話に含まれる感情を認識することを目的とする．

発話に感情が含まれる音声の事例としては様々な対象が存在するが，喜怒哀楽が容易に把握できる音声の事例の一つが，スポーツの実況や観戦の音声である．その中でも，喜怒哀楽の感情の識別が比較的容易である対象としてサッカーが挙げられる．サッカーにおいては，試合の局面ごとに攻撃をしているチームと守備をしているチームが明確に分かれており，バスケットボール等に比べて試合展開の時系列な遷移が緩やかであるため，状態遷移の同定が相対的に容易であると考えられる．そこで，本研究ではサッカー実況音声を対象として，応援チーム側から見て試合の状態遷移を同定するタスクを取り上げる．特に本研究では，視覚情報を伴わず，音響・音声情報のみを情報源として，試合の状態遷移の手がかりを得ることができる対象として，J リーグのラジオ中継を選択する．そして，Fig. 1 に示すように，刻一刻と変化する試合展開をとらえて，応援チーム側の立場で一喜一憂するロボットの実現を目的とする．

通常，J リーグのラジオ中継は，ホームチーム側のラジオ局が提供しており，ホームチームを応援する立場で実況中継される．アナウンサーの声のピッチ (声の高さ)・パワー (声の大きさ) はホームチームが攻撃している場面では相対的に高・大となる傾向があり，逆にアウェイチームが攻撃している場面では相対的に低・小となる傾向がある．この傾向が最も顕著に現れる場面がゴール時のアナウンサーの実況音声である．ゴール時にアナウンサーは,「ゴーーーール」という掛け声もしくはそれに類する実況音声を発する．ここで，ホームチームのゴール時には，この「ゴーーーール」という実況音声は，ピッチ・パワー・長さともに高・大・長となるが，アウェイチームのゴール時には，低・小・やや短となるという傾向が容易に観測される．これらの音響的手がかりは，ホームチームとアウェイチームのどちらが攻撃していてゴールを決めたのか，を識別する有力な手がかりとなる．

Table 1 ゴール時の実況音声のピッチ波形の統計

波形	ホームゴール	アウェイゴール
フラット	14	0
傾斜	9	7
その他	2	7
計	25	14

以上をふまえて，本論文では，音響情報の中でも特に音声のピッチに注目してホームチームのゴールシーンを同定することを目的とする．ホームチームのゴール時の特徴として，アナウンサーの「ゴーーーール」という音声のピッチ波形がフラットな直線となる傾向があることが分かった．そこで，本論文では，この音響的な特徴を手がかりとしてフラットな波形を探索することにより，ホームチームのゴールシーンを同定する方式を提案し，評価実験を行った結果を示す．

2 分析対象実況音声データ

本論文で用いる分析対象の実況音声データとしては，J リーグの試合のうち，2012 年の 2 試合，2013 年の 9 試合の計 11 試合の実況音声を対象とした．Table 1 のホームゴール数，アウェイゴール数に示すように，この 11 試合において，ホームチームのゴールは計 25 ゴール，アウェイチームのゴールは計 14 ゴールであった．なお，これらの 11 試合におけるホームチームの異なり数は 3 チームであり，実況を担当するアナウンサーの異なり数は 4 人であった．

3 ゴール時の実況音声のピッチ波形の分類

通常，J リーグのラジオ中継のアナウンサーは，ゴール時に,「ゴーーーール」という掛け声もしくはそれに類する実況音声を発することが多い．そこで，まず，前節で収集した実況音声を対象として，ゴール時における実況音声のピッチ波形の類型化を行った．その結果，ゴール時における実況音声のピッチ波形は以下の三種類に大別できた．

*Detection of Goal Scenes by Pitch Sequence Pattern Matching in Radio Football Live Speech, by IWANAGA, Yohei, ARAI, Shota, DOMOTO, Kentaro, UTSURO, Takehito (University of Tukuba)

Fig. 1　サッカー実況音声を聞いて得点シーンにおいて一喜一憂するロボット

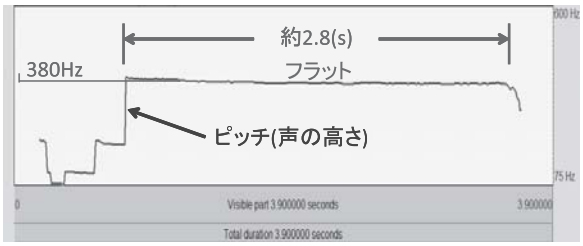

Fig. 2　フラット波形

Fig. 3　傾斜波形

- ピッチの値の変動がほとんどない平坦 (フラット) な波形 (Fig. 2)
- ピッチの値が直線的に減少し右下がりの傾斜を示す波形 (Fig. 3)
- ピッチの値の変化が直線以外となる波形

次に，ホームチームのゴール計 25，および，アウェイチームのゴール計 14 に対して，ゴール時の実況音声のピッチ波形をこれらのいずれかに分類した．この分類結果を Table 1 に示す．

この結果から分かるように，ピッチ波形が平坦 (フラット) となるのはホームチームのゴール時のみとなった．また，右下がりの傾斜となるピッチ波形については，ホームチーム・アウェイチームどちらのゴール時についてもほぼ均等に観測された．一方，直線以外の波形については，アウェイチームのゴール時の方がやや多いという結果となった．

以上の結果に基づき，本論文では，ホームチームのゴール時の実況音声を同定することを目的としてまず最初に取り組む課題として，ピッチ波形が平坦 (フラット) となる区間を検出する手法を提案し，その評価結果を示す．

4　フラット波形の探索

サッカー実況音声において，フラット波形の区間を探索するためには，まず実況音声からピッチ列を抽出する必要がある．本論文では，音声分析用フリーソフトウェア praat[1]を用いて，実況音声のピッチ抽出を行う．

フラット波形探索手法の模式図を Fig. 4 に示す．フラット波形の探索においては，まず，サッカー実況音声の全区間を $[t_0, \ldots, t_N]$ (ただし，$\Delta t = t_{i+1} - t_i = 0.01(\mathrm{s})$) とする．そして，時間長 d の区間 $[t_i, t_i + d]$ におけるピッチの数値列に対して，数値列の分散 σ^2 を算出し，この分散が上限 u 以下か否かの判定結果を求める．この操作をサッカー実況音声の全区間 $[t_0, \ldots, t_N]$ に対して行い，分散 $\sigma^2 \leq u$ を満たす時間長 d の区間のうち，連続する区間

$$[t_i,\ t_i + d],\ [t_i + \Delta t,\ t_i + \Delta + dt],$$

$$\ldots,\ [t_i + n\Delta t,\ t_i + n\Delta + dt]$$

$$\Big(= [t_i,\ t_i + n\Delta + dt] \Big)$$

を検出し，フラット波形候補区間とする．

[1] http://www.fon.hum.uva.nl/praat/

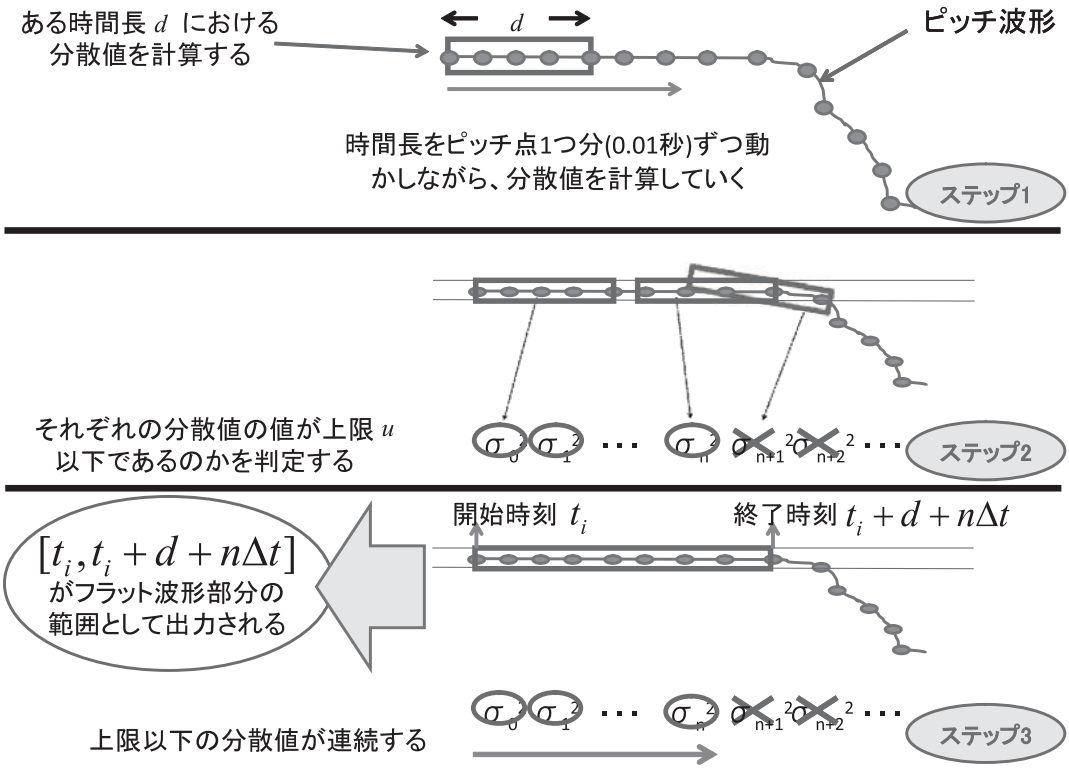

Fig. 4　フラット波形の探索手法

5　評価

5.1　評価手順

前節のフラット波形探索手法において最終的な評価結果を得るためには，ピッチ数値列の分散 σ^2 の上限 u，および，分散算出区間の時間長 d の組合わせを網羅した上で，前節のフラット波形探索を行い，再現率・適合率の値が最適となる組合わせを求める．ここで，ピッチ数値列の分散 σ^2 の上限 u の値は，1~15 の間の整数値 15 通りとし，分散算出区間の時間長 d の値は，0.5~1.5(s) の間の値を 0.1(s) 間隔で変化させた 11 通りとして，全組み合わせの合計 165 通りの u と d の組を評価対象とする．そして，u と d の各組に対して，前節のフラット波形探索手法を適用し，再現率・適合率の値が最適となる組合わせを求める．

フラット波形探索手法の評価においては，まず，提案手法によって検出したピッチ点列と参照用ピッチ点列を比較する．ここで，提案手法よって検出したピッチ点列の区間と参照用点列の区間が一箇所でも重複した場合に，参照用ピッチ点列のフラット波形の検出に成功したと判定し，点列区間が一箇所も重ならなければ，参照用ピッチ点列のフラット波形の検出に失敗したと判定する．そして，次式の再現率と適合率を

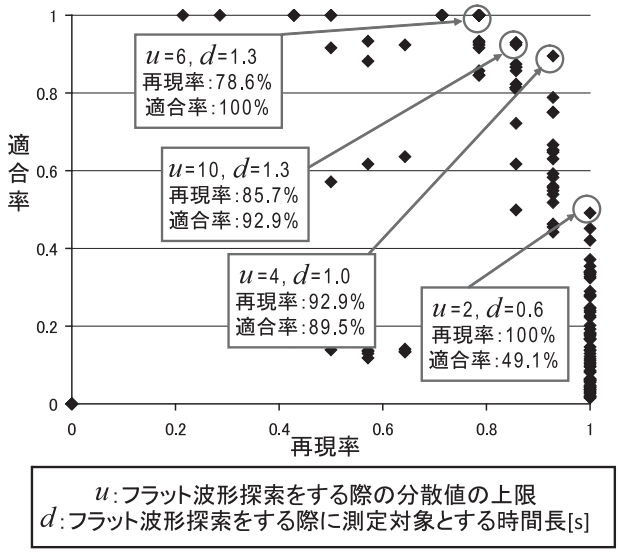

Fig. 5　評価結果

算出し，これを評価尺度とする．

$$\text{再現率} = \frac{\text{検出が成功したピッチ点列の数}}{\text{参照用ピッチ点列の数}}$$

$$\text{適合率} = \frac{\text{検出が成功したピッチ点列の数}}{\text{検出されたピッチ点列の数}}$$

5.2　評価結果

ピッチ数値列の分散 σ^2 の上限 u，および，分散算出区間の時間長 d の組合わせの合計 165 通りについ

て，再現率，適合率をプロットした結果を Fig. 5 に示す．この結果から分かるように，(1) $u = 6$，$d = 1.3$，(2) $u = 10$，$d = 1.3$，(3) $u = 4$，$d = 1.0$，(4) $u = 2$，$d = 0.6$，の 4 通りの u および d の組み合わせの各々においては，他のいずれの u，d の組み合わせに対しても，再現率・適合率のいずれか一方もしくはその両方において上回っており，Fig. 5 のプロットにおいて最良点の集合となった．また，これらの 4 点においては，条件 (1) での再現率 78.6%，適合率 100%から，条件 (4) での再現率 100%，適合率 49.1%へと，フラット波形検出の条件が緩まる方向へと遷移しており，より多くの参照用フラット波形を検出するものの，誤検出となるピッチ点列もあわせて増加するという結果となった．

ここで，誤検出となったピッチ点列としては，アナウンサーの「まぁー」(フラット波形部分が 0.97(s)) や「うーん」(フラット波形部分が 1.25(s)) という発話，および，観客の「おー」(フラット波形部分が 1.20(s)) という歓声などが挙げられる．このうち，アナウンサーの「まぁー」や「うーん」という発話の場合はピッチが 80∼100Hz 程度であり，このピッチの値の絶対値を参照することにより，通常，300∼400Hz 程度のピッチの値を示す「ゴーーーール」という発話との間の識別は容易であると考えられる．また，観客の「おー」という歓声の場合は，パワーが約 75dB 程度であり，このパワーの値の絶対値を参照することにより，通常，85dB 程度のパワーの値を示す「ゴーーーール」という発話との間の識別は容易であると考えられる．

6　関連研究

スポーツ実況音声に関する関連研究として，スポーツ実況音声の音響分析に関する研究 [1]，および，音声認識に関する研究 [2–5] が挙げられる．文献 [2–4] は，野球の実況音声に対して音声認識技術，および，イベント推定手法を併用し，試合へのメタ情報を付与する手法を提案している．一方，文献 [6] は，文献 [5] の手法によってサッカー実況音声を音声認識した結果 (論文中では，実際には人手による書き起こしテキストを利用している) に対して SVM を適用し，試合中のコメントを「試合記述文」と「解説文」に分類することによって，試合中のイベント情報を時間ごとに説明するセグメントデータを付与する手法を提案している．また，文献 [1] においては，サッカーの実況音声中において，歓声等を対象とした音響分析を行い，音声のパワーを用いて観客の盛り上がりを同定し，試合中のイベントを抽出する手法を提案している．

以上の関連研究のうち，特に文献 [1] は，サッカー実況音声の音響情報に着目し試合中のイベントを抽出する点において，本研究とその目的が類似している．しかし，文献 [1] が観客の歓声を対象として音響分析を行い，音声のパワーを手がかりとして試合中のイベント抽出を行うのに対して，本論文では，アナウンサーの実況音声を対象として音響分析を行い，音声のピッチを手がかりとして試合中のイベント抽出を行う点において，両者は大きく異なっている．

その他，本論文のピッチ列分析手法に関連して，文献 [7] においては，対話音声における「うんうん」と「うーん」の識別において，F0 の軌跡の違いを分析した結果を報告している．

7　おわりに

本論文では，J リーグの実況音声における音響情報のうち，特に音声のピッチに注目してホームチームのゴールシーンを同定することを目的とした．ホームチームのゴール時の特徴として，アナウンサーの「ゴーーーール」という音声のピッチ波形がフラットな直線となる傾向があることを示した．そして，この音響的な特徴を手がかりとしてフラットな波形を探索することにより，ホームチームのゴールシーンを同定する方式を提案し，評価実験を行った結果を示した．

今後は，ゴール時点における他のピッチ波形として，直線的右下がりの傾斜波形，および，直線以外の波形の検出方式について研究を進める．そして，それらの波形検出方式を統合することにより，ホームチームのゴール時点とアウェイチームのゴール時点を識別する方式を確立する．

参考文献

[1] 塩崎他：“音響信号処理に基づくサッカー映像のインデクシング手法”, 第 3 回 FIT 一般講演論文集, 第 3 巻, pp. 107–108 (2004).

[2] 有木他：“音響・言語モデルの適応処理によるスポーツ実況中継の音声認識”, 信学論, **J87–D–II**, 6, pp. 1208–1215 (2004).

[3] 佐古他：“音声・状況の同時認識に基づくスポーツ実況中継へのメタ情報付与”, 情報処理学会論文誌, **50**, 2, pp. 563–574 (2009).

[4] 佐古他：“音声・状況の同時認識に基づく野球実況中継へのメタ情報付与”, 第 3 回音声ドキュメント処理ワークショップ講演論文集, pp. 59–64 (2009).

[5] 佐藤他：“実況・対談における発声変形を考慮した音響モデルの検討”, 情報処理学会研究報告, **2005–SLP–59**, pp. 31–36 (2005).

[6] 山田他：“アナウンサーと解説者のコメントを利用したサッカー番組セグメントメタデータ自動生成”, 信学論, **J89–D–II**, 10, pp. 1428–1440 (2004).

[7] 石井他：“「うんうん」と「うーん」の識別における音響特徴の分析”, 音講論 (秋), pp. 265–266 (2013).