

## サッカー・ラジオ実況音声におけるゴール時検出 およびホーム・アウェイ識別\*

岩永羊平, ☆堂元健太郎, 宇津呂武仁 (筑波大)

### 1 はじめに

本研究においては、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ゴールを決めたのか、を識別する有力な手がかりとなる。

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ゴールシーンの同定、およびホームチームのゴールとアウェイチームのゴールの

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	19	0
傾斜	12	6
その他	15	20
計	46	26

識別を目的とする。また、ゴール時におけるピッチ波形は、フラットな直線、右下がりの傾斜を示す直線、その他の波形の3種類に分類できることが分かった。さらに、フラットなピッチ波形はホームチームのゴールにしか現れない事と、右下がりの傾斜を示すピッチ波形はホームチーム・アウェイチームの両方に現れるが、基本周波数の最小値がホームチームのゴール時の方が高い傾向にある事が分かった。そこで、本論文では、この音響的な特徴を手がかりとしてホームチーム、アウェイ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。以上のうち、文献 [1] においては、フラット波形検出手法およびその有効性を示した。本稿では、傾斜波形の検出手法およびホームチームとアウェイチームの識別手法の有効性を示す。

### 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20試合の計22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45ゴール、アウェイチームのゴールは計26ゴールであった。なお、これらの22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

\*Detection of Goal Scenes in Radio Football Live Speech and Discrimination of Scenes of Goals of Home/Away Teams, by IWANAGA, Yohei, DOMOTO, Kentaro, UTSURO, Takehito (University of Tsukuba)



Fig. 1 サッカー実況音声聞いて得点シーンにおいて一喜一憂するロボット

おける実況音声のピッチ波形の類型化を行った。直線波形の定義を

ピッチ波形の直線箇所が 0.85 秒以上継続する波形 (1)

とした時、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。

**フラット波形** 定義 (1) を満たす直線波形のうち、ピッチの値の変動がほとんどない直線的に平坦 (フラット) な波形 (Fig. 2)

**傾斜波形** 定義 (1) を満たす直線波形のうち、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)

**その他の波形** 上記の「フラット波形」、「傾斜波形」のいずれにも属さない波形

次に、ホームチームのゴール計 46、および、アウェイチームのゴール計 26 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。このことから、ゴール発話がフラット波形ならばホームチームのゴールであると言える。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。

## 4 傾斜波形の検出

### 4.1 手順

傾斜波形検出手法の模式図を Fig. 4 に示す。傾斜波形の検出においては、まず、サッカー実況音声の全区間を  $[t_0, \dots, t_N]$  (ただし、 $\Delta t = t_{i+1} - t_i = 0.01(\text{s})$ ) 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  $(t_i, p_i), \dots, (t_{i+m}, p_{i+m})$  (ただし、 $m = \frac{d}{\Delta t}$ )

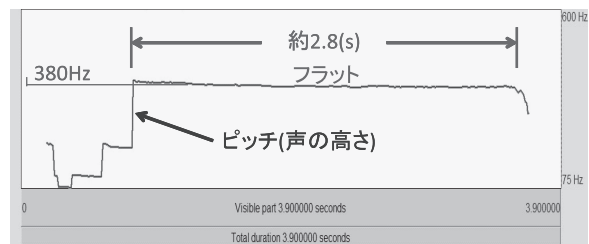


Fig. 2 フラット波形

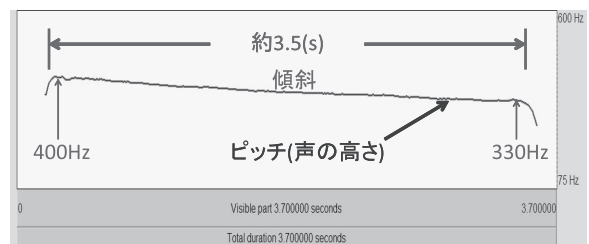


Fig. 3 傾斜波形

に対して、以下の式より傾き  $a$  と切片  $b$  を求める。

$$a = \frac{(m+1) \sum_{j=i}^{i+m} t_j p_j - \sum_{j=i}^{i+m} t_j \sum_{j=i}^{i+m} p_j}{(m+1) \sum_{j=i}^{i+m} (t_j)^2 - (\sum_{j=i}^{i+m} t_j)^2}$$

$$b = \frac{\sum_{j=i}^{i+m} (t_j)^2 \sum_{j=i}^{i+m} p_j - \sum_{j=i}^{i+m} t_j p_j \sum_{j=i}^{i+m} t_j}{(m+1) \sum_{j=i}^{i+m} (t_j)^2 - (\sum_{j=i}^{i+m} t_j)^2}$$

また、近似直線とピッチ点の誤差平均  $e$  を次式とする。

$$e = \frac{\sum_{j=i}^{i+m} |p_j - at_j - b|}{m+1}$$

この平均誤差  $e$  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  $[t_0, \dots, t_N]$  に対して行い、誤差平均  $e \leq u$  を満たす時間長  $d$  の区間のうち、連続する区間

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

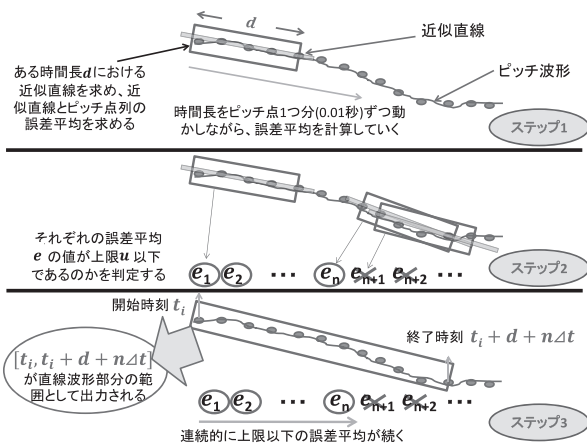


Fig. 4 傾斜波形の検出手法

$[t_i + n\Delta t, t_i + n\Delta t + dt]$  ( $= [t_i, t_i + n\Delta t + dt]$ )  
を検出し、傾斜波形候補区間とする。

## 4.2 評価

### 4.2.1 評価手順

傾斜波形検出手法の評価においては、まず、Table 1におけるフラット波形および傾斜波形の計 37 ゴール分における実況音声のピッチ点列を参照用ピッチ点列とする。そして、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、提案手法によって検出したピッチ点列と参照用ピッチ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傾斜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いければ、参照用ピッチ点列の傾斜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評価尺度とする。

$$\begin{aligned} \text{再現率} &= \frac{\text{検出が成功したピッチ点列の数}}{\text{参照用ピッチ点列の数}} \\ \text{適合率} &= \frac{\text{検出が成功したピッチ点列の数}}{\text{検出されたピッチ点列の数}} \end{aligned}$$

そして、ピッチ数値列の誤差平均  $e$  の上限  $u$ 、および、誤差平均算出区間の時間長  $d$  の組み合わせを網羅した上で、前節の傾斜波形検出を行い、再現率・適合率の値が最適となる組み合わせを求める。ここで、ピッチ数値列の誤差平均  $e$  の上限  $u$  の値は、1~3 の間の値を 0.05 間隔で変化させた 41 通りとし、誤差平均算出区間の時間長  $d$  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 451 通りの  $u$  と  $d$  の各組に対して、前節の傾斜波形検出手法を適用し、再現率・適合率の値が最適となる組み合わせを求める。

### 4.2.2 評価結果

ピッチ数値列の誤差平均  $e$  の上限  $u$ 、および、誤差平均算出区間の時間長  $d$  の組み合わせの合計 451 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、(a)  $u = 1.25$ ,  $d = 1.5$ , (b)  $u = 1.4$ ,  $d = 1.5$ , (c)  $u = 2.0$ ,  $d = 1.4$ , (d)  $u = 2.05$ ,  $d = 1.3$ , (e)  $u = 2.7$ ,  $d = 1.0$ , の 5

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いる点のうちの一つである。これらは、Fig. 5 のプロットにおいて最良点の候補となる。また、これらの 5 点においては、条件 (a) での再現率 37.8%, 適合率 100% から、条件 (e) での再現率 100%, 適合率 19.1% へと、傾斜波形検出の条件が緩まる方向へと遷移しており、より多くの参照用傾斜波形を検出するものの、誤検出となるピッチ点列も合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「まあー」や「うーん」という発話、および、観客の「おー」という歓声が挙げられる。さらに、傾斜波形検出のみで検出された例として、「うーん」等の発話を語尾が弱くなるように発話したピッチ波形で、右下がりな直線となる波形が検出された。このように、ピッチ波形が右下がりとなる発話は、直線であったとしても、ピッチの値が同じ値を保ち続けないため、フラット波形としては検出されない。他には、「非常にあのー」(発話時間が 1.12(s)) や「もったいないよなー」(発話時間が 1.08(s)) 等、同じ声の高さで一定時間発話している波形で、フラット波形検出において誤検出してもおかしくないものが検出された。フラット波形検出では、全体の傾きが 0 (=フラット) でも、局所的に凹凸しているものは検出しない。一方、傾斜波形検出では全体で直線ならば、局所的に凹凸していても検出する。そのため、「非常にあのー」等の波形を傾斜波形検出のみで誤検出したと考えられる。これらの「うーん」や「非常にあのー」、「もったいないよなー」といった発話は、声のパワーの平均値がゴール時の発話と比べて比較的低い傾向がある。例えば、上で挙げた最良点候補のうち、最も傾斜波形取得の条件が緩い条件 (e) で取得した波形においては、ゴール時の発話のうち 75% 以上においてパワーの平均値が 78(dB) 以上となり、一方、誤検出された波形のうちの 75% 以上においてパワーの平均値が 78(dB) 未満となった。したがって、声のパワーの平均値を用いることで、これら誤検出した発話とゴール時の発話を識別することはある程度可能であると考えられる。

## 5 ホームゴール時とアウェイゴール時の識別

4 節では、ゴール時の傾斜波形の検出について述べたが、本節では、実況音声で傾斜波形となる場合について、そのゴールがホームチームのゴールもしくはアウェイチームのゴールのいずれであるかを識別する際

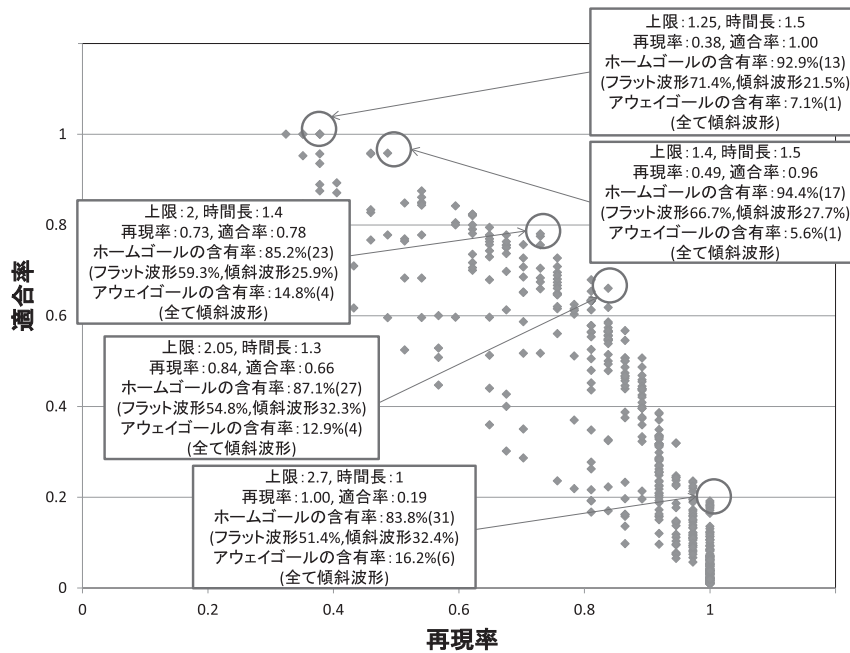


Fig. 5 傾斜波形検出実験の評価結果

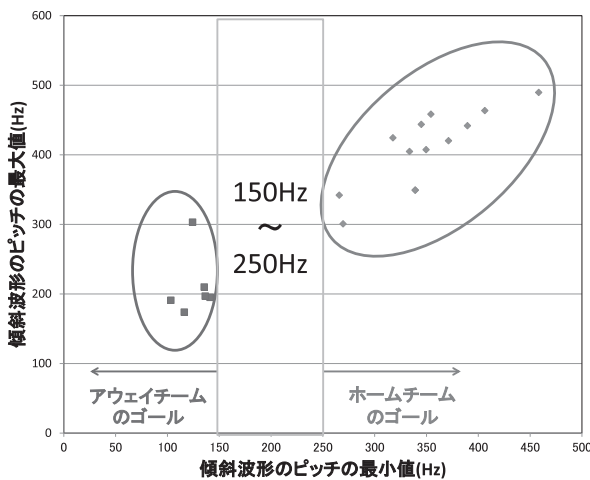


Fig. 6 実況音声で傾斜波形となるゴールにおけるピッチ最小値およびホームゴール・アウェイゴールの分布

の手がかりについて述べる。そのための準備として、まず、Table 1 のゴール時のピッチ波形の統計から、ピッチ波形がフラット波形であるゴール発話はホームチームのゴール時にしか現れないことが分かる。このことから、フラット波形検出手法によりフラット波形が検出された場合、その区間はホームチームのゴールシーンであるといえる。一方、ピッチ波形が傾斜波形であるゴール発話はホームチームのゴール時とアウェイチームのゴール時の両方に現れる。この場合、Fig. 6 に示すように、傾斜波形におけるピッチの最小値の境界領域を 150~250Hz とし、傾斜部分のピッチの最小値が 150Hz 以下ならばアウェイチーム

のゴール、傾斜部分のピッチの最小値が 250Hz 以上ならばホームチームのゴールとなることが分かった。

## 6 関連研究

本論文に関連して、文献 [2] においては、サッカーの実況音声において、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と同定し、試合中のイベントを抽出する手法を提案している。この研究は、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 [2] 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。

## 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、特にピッチに注目し、実況音声のピッチ列が右下がりの傾斜を示す波形を検出することにより、ホームチーム、アウェイ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。

## 参考文献

- [1] 岩永他：“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”，音講論（春），pp. 155-158（2014）。
- [2] 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）。