一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IN2015-95 (2016-01)

一球速報と実況音声認識を用いた野球映像の自動タギング

荒澤　孔明†　　服部　峻††

†,†† 室蘭工業大学 ウェブ知能時空間研究室　〒 050–8585 北海道室蘭市水元町 27–1
E-mail: †12024006@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

あらまし　我々が野球映像から見たいシーンだけを自由に選抜し，パーソナライズされたハイライト映像を制作できるようにするためには，野球映像を自動的に打席シーン毎に分割し，その打席シーンに関するタグ情報も付加する野球映像の自動タギングシステムが必要である．そこで本稿では，一球速報の Web テキスト，及び，実況音声認識を用いた複数のタギングアルゴリズムを提案し，それらを比較するため，再現率や適合率，F 値に関して評価実験を行う．
キーワード　タギング，実況音声，一球速報，Web テキスト抽出，野球映像

Automatic Baseball Video Tagging Using Ball-by-Ball Textual Report and Voice Recognition

Komei ARASAWA† and Shun HATTORI††

†,†† Web Intelligence Time-Space (WITS) Laboratory, Muroran Institute of Technology
27–1 Mizumoto-cho, Muroran, Hokkaido 050–8585, Japan
E-mail: †12024006@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

Abstract　To enable us to select the only scenes that we want to watch in a baseball video and personalize its highlights sub-video, we require an Automatic Baseball Video Tagging system that divides a baseball video into multiple sub-videos per at-bat scene automatically and also appends tag information relevant to at-bat scenes. This paper proposes several Tagging algorithms using ball-by-ball textual report and voice recognition, and performs evaluation experiments to compare them with regard to their recall, precision, and F-measure.
Key words　Tagging, Ball-by-Ball Voice, Ball-by-Ball Textual Report, Web Text Extraction, Baseball Video

1. ま え が き

スポーツ報道番組等ではハイライト映像がよく使用される．その映像の多くは番組側で制作されているため，同じ試合のハイライト映像でも番組によって異なったシーンがハイライト映像として使用される．ハイライトシーンはその試合の見どころでもあることから，試合を全て見られなかった人達に対して，短時間で試合を楽しんでもらうのがハイライト映像を制作する目的とも言えるだろう．すなわち，番組側は多数の視聴者が見たいと思うシーンを推測しながらハイライト映像を制作する．しかしながら個々の視聴者が見たいと思うシーンはそれぞれで異なり，それら全てのシーンをハイライト映像に含めるわけにはいかない．例えば，ある視聴者が選手 A を応援していて，その選手 A のプレーシーンをハイライト映像として視聴したいと思っていたとする．しかし，仮にその試合で選手 A にあまり目立った活躍が無かった場合，番組側が制作するハイライト映像で選手 A のプレーシーンが使用されることはおそらく無いだろう．つまり，このハイライト映像はその視聴者のニーズには応えられなかったということになる．従って，番組側が制作するような多数の視聴者の一般的なニーズ予測に基づいたハイライト映像では，個々の視聴者のニーズに全て応えるということは不可能であることが言える．

そこで，視聴者自身でハイライト映像を制作することができれば，前述した，番組側が制作するハイライト映像の課題について解決できるのではないかと考えた．視聴者自身がハイライト映像を制作する手段の一つとして，「視聴者が予めスポーツ中継番組を録画し，自分が見たいシーンだけを集めて編集する」等が挙げられる．しかしながら，これは映像全体に対して，見たいシーンが存在した場合そのシーンを編集し，それ以外のシーンは早送りにする等の作業を行うというのが一般的であるため非常に手間がかかる．

では，もし，その試合映像がシーン毎にチャプター分割されていたらどうだろうか．チャプターとは，試合映像の各シーンに見出しを付け，見たいシーンへ直接移動できる機能のことである．前述した録画した映像から視聴者自身が見たいシーンを探すという作業も，既にシーン毎にチャプター分割されている

ため，映像を全て見る必要が無くなり，視聴者はシーンを選抜するだけで良いことになる，つまり余計な手間をかけずに済む．すなわち，予めチャプター分割されたスポーツ映像を視聴者に提供することができれば，各々が見たいシーンを容易に視聴することが可能になるのではないかと考えた．そこでスポーツ映像を自動でシーン毎にチャプター分割するシステムを提案する．映像をチャプター分割するためには「その映像内でどのような出来事が起こったか」，及び，「その出来事はいつ起こったか」という二つの情報が必要になる．本稿では野球映像の自動チャプター分割を目指し，一球速報の Web テキストと実況音声認識を用いて，「その映像内でどのような出来事が起こったか」，及び，「その出来事はいつ起こったか」という二つの情報の取得手法を提案する．また，本稿で目指すチャプター分割は**タギング（タグ付け）**[1] と呼ばれる分割手法で，これは，単に見出しを付加したシーン毎に分割するのでなはく，タグ情報と呼ばれる，そのシーンで起こった詳細な情報も付加したシーン毎に分割する手法のことを言う．本提案システムは，このタグ情報をシステム利用者に提供することで，それらのタグ情報に基づいたシーンの選抜を可能にする．

2. 提 案 手 法

2.1 システム概要

映像をシーン毎にタギングするには基本的に「その映像内でどのような出来事が起こったか」，及び，「その出来事はいつ起こったか」という二つの情報が必要になる．野球映像のタギングでも同様にこれら二つの情報が必要になるが，それらの情報の細かさ，つまり，タギングをする際の分割の細かさも考慮する必要がある．野球映像をタギングする場合，一球毎に「いつからいつまでが○○投手が何球目を投げたシーン」等のようにタギングするケースや，イニング毎に「いつからいつまでが先攻チームの攻撃シーン」等のようにタギングするケースといった様々なタギングケースが考えられるが，本稿ではタギングをする際の分割の細かさを打者毎に定め，「いつからいつまでが打者○○の打席で，どのような結果になったシーン」のようなタギングを目指す．ここで，映像のタギングに必要な二つの情報のうち，「その映像内でどのような出来事が起こったか」というフィールドを**イベント**，「その出来事はいつ起こったか」というフィールドを**イベント時間**と言う．

イベントの取得には，一球速報 [2] と呼ばれる Web テキストを利用する．本稿で目指すタギングは打席シーン毎の分割であるため，イベント E_i $(i = 1, 2, \cdots, N)$ も打席シーン毎に取得する．これを図 1 の Step 1 に示す．また，あるイベント（打席シーン）について，そのイベントの打者が打席に入った時の状況とその打者名，及び，その打席シーンの結果をそのイベントの内部要素とする．イベントの取得手順としては，タギング対象の試合が終了した後，一球速報を用いて打者が変わる毎，つまり打席シーン毎に取得する．取得するイベント E_i の例に「2 アウト 1 塁の状況で，第△打席目に○○選手がホームランを打った」を挙げる．但し，同じ選手が複数回打者として登場したシーンであっても違うイベントとして扱う．

次に，これらのイベントにイベント時間を付加していくのだが，ここではイベント時間について，まず，全イベント E_i に対してイベント開始時間 T_i $(i = 1, 2, \cdots, N)$ を付加する．これを図 1 の Step 2 に示す．次に，イベント E_i に対して，次のイベント E_{i+1} のイベント開始時間 T_{i+1} をイベント E_i のイベント終了時間とすることで，イベント時間が算出される．イベント開始時間の算出方法として，一般的な野球中継番組で行われている実況放送を利用できないかと検討した．野球中継番組では一般的に，実況者，及び，解説者の音声も放送される．その会話内容には選手の特徴，プレーの実況と解説，次の展開の予測などが含まれる．その中で，ある打者が打席に入っていることを示す実況に着目した．ある打者が打席に入っていることを示す実況とは，例えば「ここでバッター○○です」のような実況である．そこでプロ野球中継番組から 2 試合を参考にして，登場した打者延べ 139 人に対し，打席シーンの中で，その打者が打席に入っていることを示す実況が少なくとも一つ以上存在したか否かという調査を行った．ある打者の打席シーンとは，その打者が打席に立った瞬間から，出塁した，またはアウトを宣告された瞬間までのことを言う．調査結果より，ある打者の打席シーン中，その打者が打席に入っていることを示す実況が少なくとも一つ以上存在した確率は 83% であった．

そこで，このようなある打者が打席に入っていることを示す実況を利用して，各イベントにイベント開始時間，及び，イベント終了時間を付加する，以下の手法を提案する．例えば，1 番打者 A，2 番打者 B に対して，実況者は A が打席に入る瞬間に「1 番の A です」，B が打席に入る瞬間に「次は B です」と実況したとする．このようなケースで「1 番の A です」と実況された瞬間を A の打席シーン，つまり，A のイベントの開始時間，また，「次は B です」と実況された瞬間を B の打席シーン，つまり，B のイベントの開始時間，及び，A のイベントの終了時間とすることでイベント時間を算出する．このプロセスでイベント時間を付加した複数のイベントを生成することで，タグ情報を含む打席シーン毎のタギングを実現する．

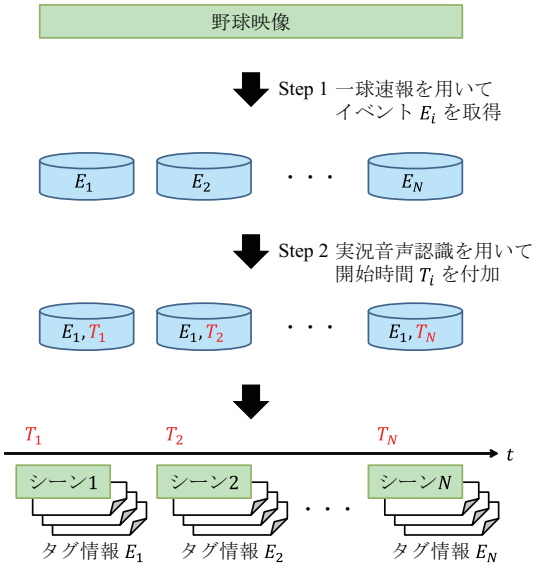

図 1 システム概要図

2.2 一球速報を用いたイベントの取得

本システムでは，Yahoo! JAPAN スポーツナビが提供する Web サイトである一球速報 [2] を用いてイベントを取得する．一球速報からはプロ野球を中心とした野球の試合のリアルタイム速報，及び，試合後の情報等を取得することができる．本稿では，タギング対象の試合が終了した後に，この一球速報の Web テキストを用いて打者が切り替わる毎にイベントを取得し，これらのイベントをタグ情報とする．

一般的なハイライト映像には含まれないようなシーンを視聴したいと思うユーザに対して，そのシーンを提供することが本システムの目的の一つであるため，シーンを選抜させる際にはより詳細なタグ情報を付加したシーンを提供する必要がある．例えば，あるユーザが選手 A のプレーした全てのシーンを視聴したいと思っている場合，ユーザはタグ情報に選手名「A」が含まれるシーンを選抜することになる．この場合，システムは守備や走塁等の選手 A の打席以外のシーン，また，選手 A の関与が極僅かであったシーンも，そのシーンに選手 A がどのように関与したかを示すタグ情報をユーザに提供しなければならない．以上のことから，一球速報の Web テキストを用いて，イベントとしてある打者が打席に入った時の状況とその打者名，及び，その打席シーンの結果を取得するのに加え，その打者の打席シーン中に，走塁，守備のいずれかに関与した選手，また，関与した内容も取得する．イベントの例として,「ある打者○○が 2 アウト 2 塁の状況で第 3 打席に立ち，サードゴロを打ちアウトを宣告された，そのシーンには，投手として選手□□，1 塁走者として盗塁を決めた選手××，打球処理をした三塁手として選手△△が関与した」を挙げる．

2.3 実況音声認識

各イベントにイベント開始時間を付加するために，野球中継番組の実況音声認識を利用する．実況音声を認識するために音声認識ソフトとして AmiVoice [3] を使用する．AmiVoice には，マイクまたは音声ファイルから音声認識した結果をテキストとして出力する機能，AmiVoice 搭載の辞書にユーザ指定の単語を新語として追加登録する機能がある．本稿では，音声認識する際のノイズをより少なくするため，一試合の実況音声を音声ファイルに変換して認識する．また，音声認識する際の辞書には AmiVoice 搭載の標準大汎用音響モデルを使用する．

野球中継における，周囲の歓声や場内アナウンスは目立ったノイズとなり，イベント開始時間を付加するための「ここでバッター○○です」のような，ある打者が打席に入っていることを示す実況を正確に認識することは困難であった．そこで「ここで○○です」や「○○が打席に入りました」等のある打者が打席に入っていることを示す実況に含まれる選手名に着目した．選手名の認識は，ある打者が打席に入っていることを示す一文の認識よりも認識率は高くなることが期待される．

本来，選手 A の打席シーンについて「ここで打席に A が入ります」のような一文を試合全体の実況音声認識結果から取得し，その実況時間を選手 A のイベント開始時間とする手法を検討していた．しかし，その一文を正確に認識することができなかったため，試合全体の実況音声認識結果から選手名「A」のみの取得に変更したところ,「次のバッターは A です」や「A が守備でファインプレーをしました」等あらゆる場面で実況された選手名「A」が取得された．このことから，試合全体の実況音声認識結果から選手名「A」を取得することで，選手名「A」が複数出現することを想定し，それら複数の「A」の中で，本来認識することを目指していた「ここで打席に A が入ります」のような打者「A」が打席に入っていることを示す実況に含まれる選手名「A」を探索する手法を提案する．

- 実況ポイント ： 実況音声認識で打者名が出現した瞬間のことを言う．また，E_i の打者名の実況ポイントを出現した順に $P(i, 1)$, $P(i, 2)$，…, $P(i, j)$ とする．但し，その打者名が打者以外のシーンで実況された場合も $P(i, j)$ に含まれる．
- 打席実況ポイント ： 上記の実況ポイントのうち，打席に入っていることを示す実況に含まれる実況ポイントのことを言う．但し，各イベントの打席シーン中で 1 回のみの出現とは限らない．

選手名が正確に音声認識されるかどうかは，認識する際の辞書に影響されると考えた．AmiVoice 搭載の辞書には，佐藤，高橋など一般的な苗字は登録されているのに対し，珍しい苗字や，外国人の名前は登録されていないと推測した．ここで推測したと述べたのは，AmiVoice 搭載の辞書における登録単語の詳細については確認できなかったためである．従って，AmiVoice 搭載の辞書へ選手名を新語として追加登録することを検討した．単語登録方法としては，一球速報を用いて取得したイベントに基づき，その試合でプレーした選手のみを抜粋し，それらの選手名を記述文法を用いた新語に変換して辞書へ追加登録する．ここで記述文法とは「実況者が選手名を実況する際によく使う言い回し」のことを言う．例えば，ある選手名「A」を音声認識させる際，単に「A」という選手名を辞書へ追加登録するだけでは，野球映像に含まれる周囲のノイズによって高い認識率は期待されない．そこで，実況者が選手名を実況する際によく使う言い回しである,「○番の A」「ピッチャーの A」等のような複数の単語の集合を一つのフレーズとして辞書へ追加登録する．この記述文法を用いた辞書登録によって選手名の高い認識率が期待される．但し，記述文法で用いる打順，及び，ポジションも一球速報を用いて取得する．以上のように選手毎に記述文法が用いられた新語を生成し，それらの新語を AmiVoice 搭載の標準大汎用音響モデルに追加登録する．本稿では，それらの新語を追加登録した辞書を用いて実況音声を認識する．

2.4 タギングアルゴリズム

2.4.1 イベント開始時間の算出

イベント開始時間の算出には，野球中継の実況音声認識を用いた．本稿では，実況音声を認識し,「ここで○○です」のようなある打者が打席に入っていることを示す実況が存在した時，その実況時間がその打者のイベント開始時間となるのではないかという仮説を立てた．この仮説に基づき，複数の実況ポイントの中から「ここで○○です」のようなある打者が打席に入っていることを示す実況を推測し，その推測に基づいてイベント開始時間を算出する手法を提案する．

まず，イベント E_1 については例外的にイベント開始時間 T_1 が算出される．本提案システムはテレビレコーダへの搭載を想定しているため，番組録画した時刻を認識し，また，一球速報を用いて試合開始時刻を取得することで，録画した映像の中での試合開始時間を算出し，その試合開始時間を T_1 とする．

続いて，イベント E_2 以降のイベント開始時間の算出方法を述べる．イベント E_i について，複数の実況ポイント $P(i,j)$ が存在する時，その中からイベント E_i の打者が打席に入っていることを示す打席実況ポイントを探索する必要がある．本稿では，まず，イベント E_i が映像内のどの辺りで起こるのかを推定イベント開始時間 $\hat{T}_i$ $(i=1,2,\cdots,N)$ として算出し，この推定イベント開始時間 $\hat{T}_i$ に基づきイベント E_i の打者が打席に入っていることを示す打席実況ポイントを探索する手法を提案する．ここでイベント E_i の推定イベント開始時間 $\hat{T}_i$ の算出方法として，以下の 3 種類を挙げる．

(i) 単位イベント当たりの平均時間を用いた推定

まず，一球速報を用いて取得した試合時間 T に基づき，単位イベント（打席）当たりの平均時間 A を次の式より算出する．

$$A = \frac{(\text{試合時間})}{(\text{総イベント数})} = \frac{T}{N} \tag{1}$$

次に，この平均時間 A を用いて，イベント E_i の推定イベント開始時間 $\hat{T}_i$ を次の式より算出する．

$$\hat{T}_i = \hat{T}_{i-1} + A = T_1 + A \times (i-1) \tag{2}$$

(ii) 単位一球当たりの平均時間を用いた推定

まず，一球速報を用いて取得した各イベント E_i に要した球数 β_i $(i=1,2,\cdots,N)$ に基づき，それぞれのイベントに要した球数を合計し，試合全体の総球数を求める．この総球数を用いて単位一球当たりの平均時間 B を次の式より算出する．

$$B = \frac{(\text{試合時間})}{(\text{総球数})} = \frac{T}{\sum_{i=1}^{N} \beta_i} \tag{3}$$

次に，この平均時間 B を用いて，球数が β_i であるイベント E_i の推定イベント開始時間 $\hat{T}_i$ を次の式より算出する．

$$\hat{T}_i = \hat{T}_{i-1} + B \times \beta_{i-1} = T_1 + B \times \sum_{k=1}^{i-1} \beta_k \tag{4}$$

(iii) 推定方法 (ii) に攻守交替の時間を考慮した推定

推定方法 (ii) と同様のアプローチで推定イベント開始時間の算出をするが，攻守交替直後のイベントに対してのみ，攻守交替時間をパラメータ Δt_0（秒）として推定イベント開始時間に加算する．但し，イベント E_i が攻守交替直後のイベントか否かについても一球速報を用いて取得し，次の関数を用いて表す．

$$\text{change}(E_i) = \begin{cases} 1 & (E_i \text{が攻守交替直後のイベント}) \\ 0 & (otherwise) \end{cases} \tag{5}$$

まず，式 (3) と式 (5) に基づき，攻守交替を考慮した単位一球当たりの平均時間 B' を次の式より算出する．

$$B' = \frac{T - \Delta t_0 \times \sum_{i=1}^{N} \text{change}(E_i)}{\sum_{i=1}^{N} \beta_i} \tag{6}$$

次に，この平均時間 B' を用いて，イベント E_i の推定イベント開始時間 $\hat{T}_i$ を次の式より算出する．

$$\hat{T}_i = \hat{T}_{i-1} + B' \times \beta_{i-1} + \Delta t_0 \times \text{change}(E_i) \tag{7}$$

これら 3 種類のいずれかの推定方法で算出されたイベント E_i の推定イベント開始時間 $\hat{T}_i$ を用いて，実況ポイント $P(i,j)$ の中からイベント開始時間 T_i となる打席実況ポイントを探索する．本稿では，イベント E_i のイベント開始時間 T_i となる打席実況ポイントは，推定イベント開始時間 $\hat{T}_i$ の付近にあると仮定した．そこで，パラメータ Δt_1（分）と Δt_2（分）を用い，$\hat{T}_i + \Delta t_1$ から $\hat{T}_i + \Delta t_2$ までの範囲を探索し，その領域内で最初に出現した実況ポイント $P(i,j)$ を打席実況ポイントと推測することで，イベント開始時間 T_i とする．但し，領域内に実況ポイント $P(i,j)$ が存在しなかった場合は，推定イベント開始時間 $\hat{T}_i$ をそのままイベント開始時間 T_i とする．図 2 にはイベント E_i の打者の実況ポイントを「●」とした時，また，その試合に出場した他の選手名の実況ポイントを「▲」とした時のイベント開始時間 T_i の算出方法を示す．

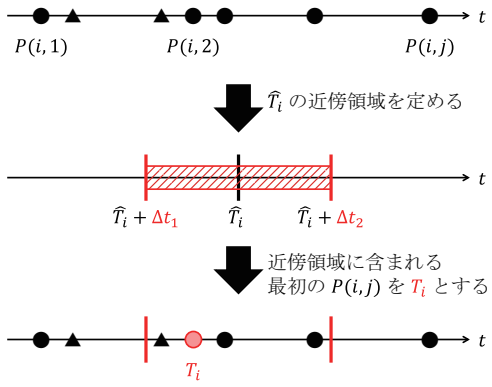

図 2　イベント E_i のイベント開始時間 T_i の算出方法

2.4.2　イベント終了時間の算出

全てのイベント E_i にイベント開始時間 T_i を付加した後，それらのイベント開始時間 T_i に基づき，各イベント E_i にイベント終了時間 T_i' $(i=1,2,\cdots,N)$ を付加する．イベント E_i の終了時間 T_i' には次のイベント E_{i+1} のイベント開始時間 T_{i+1} が付加される．但し，イベント開始時間 T_i を算出する際は，イベントの前後関係を考慮せずに，各イベント E_i で独立したイベント開始時間 T_i の算出をしたため，$T_i \geqq T_{i+1}$ となる場合がある．このような場合には，推定方法 (i)，(ii)，(iii) においてそれぞれ，A，$B \times \beta_i$，$B' \times \beta_i$ をイベント開始時間 T_i に加算した値をイベント終了時間 T_i' とする．また，最後のイベント E_N には次のイベント E_{N+1} のイベント開始時間 T_{N+1} が存在しないため前ケースと同様に，推定方法 (i)，(ii)，(iii) においてそれぞれ，A，$B \times \beta_N$，$B' \times \beta_N$ をイベント開始時間 T_N に加算した値をイベント E_N の終了時間 T_N' とする．

3. システム評価と今後の課題

本稿では，打者毎のイベントを取得した後，それらにイベント時間を付加するという野球映像のタギングアルゴリズムを提案した．イベントの取得に関しては全て一球速報の Web テキストに依存しているため，各イベントのタグ情報は全て正確であることを前提とし，イベントの取得に関する評価は行わない．従って本章ではイベントに付加したイベント時間についてのみ評価を行う．本稿では，2015 年 10 月 24 日に行われた「ヤクルト対ソフトバンク」の試合を実験対象の野球映像とした．

3.1 実況音声に関する評価と課題

評価実験に用いた試合の総イベント数（総打席数）は 71 個であり，そのうち打者が打席に入っていることを示す打席実況ポイントが存在したイベント数は 63 個であったことから，ある打者の打席シーン中に，実況者は 88% の確率でその打者名を実況したことが分かる．また，その 63 個のイベントについて，打席実況ポイントの延べ総数は 125 個であり，そのうち正確に音声認識されていた打席実況ポイント数は 80 個であったことから，打席実況ポイントの認識率は 64% であったことが分かる．本稿のタギングアルゴリズムは「ここで○○です」のような打席実況ポイントを試合全体の実況音声認識結果から探索するということが本質となるため，より正確な実況音声認識が必要となる．評価実験に用いた野球映像においても，本来ならば「ここで○○です」のような打席実況ポイントが存在したにも関わらず，正確に認識されずに打席実況ポイントが無しと認識されたイベントが 12 個あった．今後，この 12 個のイベントで正しく認識されるはずであった打席実況ポイントを正確に認識する手法を検討することで，より高いタギング精度を目指す．

あるイベントに対しての打席実況ポイントの有無は試合や実況者への依存性が強いため，本提案アルゴリズムでは妥協せざるを得ないが，少なくとも，64% の音声認識率に関しては今後改善できる可能性がある．一つは音声認識ソフトの見直しである．本稿では音声認識ソフトとして AmiVoice を使用したが，オープンソースソフトウェアではないため，AmiVoice の辞書，及び，音響モデルに触れることは出来なかった．今後はより柔軟に扱うことのできる音声認識ソフトを検討することで，実況音声の更なる認識率の向上が期待できる．もう一つは記述文法の見直しである．2.3 節で前述したように，AmiVoice 搭載の辞書に選手名を登録する際，実況者がよく使う言い回しである記述文法を用いて辞書へ追加登録した．例えば，より言い回されているフレーズを調査したり，実況者によって登録する言い回しを変更したりする等の適切な選手名の辞書登録方法を検討することで，実況音声認識の向上を目指す．

3.2 イベント時間の算出に関する評価と課題

イベント時間の評価を行うためには，実際のイベントの開始時間と終了時間を定義する必要がある．実際のイベント開始時間は，各打者が打席に入った時に，その打者の名前やシーズン成績等がディスプレイに表示された瞬間と定義する．実際のイベント終了時間は基本的に，次のイベントの実際のイベント開始時間と定義するが，そのイベントが攻守交替する直前のイベントであった場合のみ，実際のイベント終了時間は，攻守交替時にイニング表がディスプレイに表示された瞬間と定義する．以上の実際のイベント開始時間と終了時間の定義に基づき，本タギングアルゴリズムのイベント時間を再現率と適合率，及び，F 値を用いて評価する．但し，システム全体のタギング精度には F 値を用いる．また，推定方法 (i) から (iii) の全てに用いられたパラメータ Δt_1（分）と Δt_2（分），及び，推定方法 (iii) のみで用いられたパラメータ Δt_0（秒）は以下の範囲とする．

- $-30 \leqq \Delta t_1 \leqq 30$（1 分刻み）
- $\Delta t_1 \leqq \Delta t_2 \leqq 30$（1 分刻み）
- $0 \leqq \Delta t_0 \leqq 600$（10 秒刻み）

表 1 には各推定方法に基づいたタギングで，最も F 値が高かった時のパラメータ Δt_1 と Δt_2，及び，その F 値を示す．また本稿では，実況音声を用いたタギングの有効性を検証するため，各イベント E_i において，推定方法 (iii) に基づいた推定イベント開始時間 $\hat{T}_i$ をイベント開始時間 T_i とし，次のイベントの推定イベント開始時間 $\hat{T}_{i+1}$ をイベント終了時間 T_i' とする，すなわち，実況音声認識を用いずに，イベント開始時間の推定のみに基づいたタギング (iii)*をベースラインとする．

表 1　各推定方法に基づいたタギングの評価

推定方法	Δt_1	Δt_2	Δt_0	再現率	適合率	F 値
(iii)*	−	−	−	0.140	0.112	**0.125**
(i)	−9	4	−	0.646	0.343	**0.448**
(ii)	−8	6	−	0.589	0.311	**0.407**
(iii)	−8	5	280	0.588	0.352	**0.440**

表 1 より，(i) から (iii) のいずれの推定方法に基づいたタギングもベースライン (iii)*と比較し，F 値が高くなっていることが分かる．しかし，推定方法 (i) から (iii) に基づいたタギング精度を比較すると，推定方法 (i) が最高 F 値となったものの，推定方法 (iii) との差はほとんど見られなかった．また，推定方法 (ii) に関しては，推定方法 (i)，または，(iii) と比較して F 値が低いことが分かるが大きな差ではなかった．この結果を野球映像のモデリングの観点から考察する．本稿ではイベントを，そのイベントの打者が打席に入ってから次のイベントの打者が打席に入るまでと定義し，野球映像全体において，各イベント同士が一続きになるようなタギングアルゴリズムを提案した．ここで我々はイベントを「打者が打席に入り，投手が投球するシーン」と「選手の交代，及び，攻守交替のシーン」の二つのモデルに構造化した．このイベントのモデル構造に基づいて各推定方法を考察すると，推定方法 (ii) では，試合時間を総球数で割って求めた単位一球当たりの平均時間を用いて推定イベント開始時間を算出しているため，投手が投球していないシーン，例えば，選手の交代シーンや攻守交替のシーンも投球しているのと同じ扱いとしての算出式になる．従って，推定イベント開始時間が適確に算出されなかったことが推定方法 (ii) に基づいたタギングが推定方法 (i) に基づいたタギングよりも高い精度にならなかった要因と考えられる．推定方法 (iii) でも推定方法 (ii) と同様の課題を残したが，推定方法 (iii) の推定イベント

開始時間の算出方法は，攻守交替のシーンは投手が投球していない扱いとしての算出式になる．従って，推定方法 (ii) と比較して推定イベント開始時間が適確に算出されたことが推定方法 (iii) に基づいたタギングが推定方法 (ii) に基づいたタギングよりも F 値が高くなった要因と考えられる．

本稿の評価実験では，推定方法 (iii) に基づいたタギングにおける最高 F 値が推定方法 (i) に基づいたタギングにおける最高 F 値を超えることはなかったが，我々は，単位イベント当たりの平均時間を用いてイベント開始時間を推定するよりも，単位一球当たりの平均時間を用いてイベント開始時間を推定する方がより適確なイベント開始時間の推定ができると考えている．本稿では，推定イベント開始時間 $\hat{T}_i$ の算出の際，イベントのモデル構造を考慮し切れなかったため期待通りの精度は得られなかったが，今後はイベントのモデル構造を厳密化して推定イベント開始時間を算出することでより高い精度を目指す．

3.3 パラメータに関する評価と課題

図 3 には推定方法 (i) に基づいたタギングについて，Δt_1 と Δt_2 の変化に伴う F 値を示す．この Δt_1 と Δt_2 は，「ここで○○です」のような打者が打席に入っていることを示す実況を探索するためのパラメータであるため，各試合，特に各実況者による依存性が非常に高いと考えられる．また，図 3 からは Δt_2 の値が一定値を超えたところで F 値に大きな変化が見られなくなったことが分かる．一方で Δt_1 の値は F 値に大きく影響していることが分かるが，これは評価実験に用いた試合のみに限らず，その他の試合でも同様に見られる傾向であると考える．本稿のタギングアルゴリズムでは，イベント開始時間を算出するために推定イベント開始時間を求め，その近傍領域内に存在する最初の実況ポイントをイベント開始時間とする手法を提案した．このことから，領域の上限を定めるパラメータ Δt_1 に関しては，値をある程度高く設定することで実況ポイントが出現した場合，その値より高い値に設定してもタギング精度に影響しないことが分かる．一方，領域の下限を定めるパラメータ Δt_2 に関しては，下限を高く設定しすぎると，探索のターゲットとしていた「ここで○○です」のような打席実況ポイントが仮にその領域よりも前に存在していた場合，見逃してしまうことも考えられ，逆に低く設定し過ぎると，探索のターゲットとしていた打席実況ポイントとは異なる，例えば，前の打者の打席シーンで実況された「次は○○です」のような実況ポイントを探索してしまうことも考えられる．

図 4 には，推定方法 (iii) で最も高い F 値が得られた時の $\Delta t_1 = -8$，$\Delta t_2 = 5$ を固定した場合の交替時間 Δt_0 の変化に伴う F 値を示す．図 4 より，$\Delta t_0 = 280$ の時の F 値をピークに F 値が減少していることが分かる．日本プロ野球においては，2 分 15 秒を目安に攻守交替を行うように推奨されており，本稿の評価実験で用いた試合でも攻守交替の平均時間は 2 分 36 秒と目安に近い時間であった．しかし最も F 値が高くなった時の Δt_0 は 280 秒（4 分 40 秒）であったため，実際の攻守交替の平均時間とは誤差が生じた．このような結果になったことも，前述した単位一球当たりの平均時間の算出の際のイベントのモデル構造を考慮し切れなかったことが要因であると考えられる．

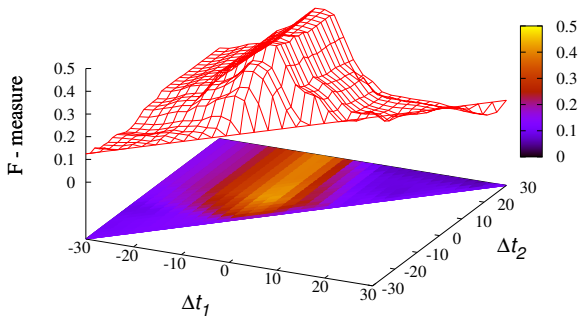

図 3　推定方法 (i) の Δt_1 と Δt_2 に依る F 値の変化

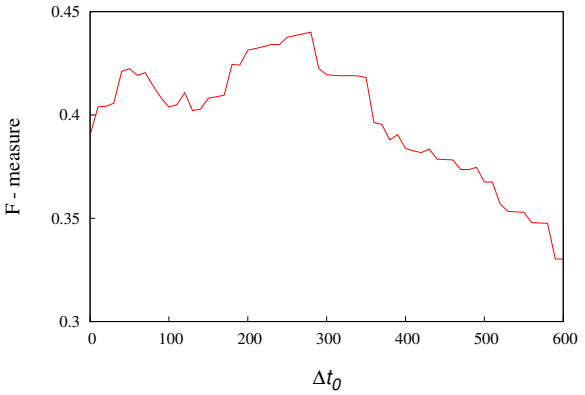

図 4　推定方法 (iii) の Δt_0 に依る F 値の変化

4. むすび

本稿では，野球映像を自動タギングするために，一球速報の Web テキストを用いてその試合で起こった打席毎のイベントを取得し，それらのイベントに基づき推定したイベント開始時間とその近傍領域，及び，実況音声認識を用いてイベント開始時間となる打席実況ポイントを探索するタギングアルゴリズムを提案した．評価実験では，再現率と適合率に関する評価を行い，最高タギング精度として F 値 0.448 を得た．また，実況音声認識を用いることの有効性，及び，推定イベント開始時間の近傍領域を定める際の最適なパラメータの傾向が判明した．一方で評価対象の映像が少なかったため，試合や実況者に依存するような評価尺度に関しての考察が不十分であったり，あらゆる試合でタギング精度を維持できるようなパラメータの最適化に関しての検証が行えなかったりといった課題も残った．今後は，複数の試合映像を用いて評価実験を行うことで，最適なパラメータについて検証し，タギング精度の更なる向上を目指す．

文　献

[1] 宮原 正典, 青木 政樹, 滝口 哲也, 有木 康雄, “顔表情からの関心度推定に基づく映像コンテンツへのタギング,” 情報処理学会論文誌, Vol.49, No.10, pp.3694–3702 (2008).

[2] Yahoo! JAPAN スポーツナビ – 一球速報 –, http://live.baseball.yahoo.co.jp/npb/game/2015102401/score.

[3] アドバンスト・メディア, 音声認識ソフト AmiVoice SP2, http://sp.advanced-media.co.jp/.