情報処理学会論文誌 Vol. 50 No. 2 563–574 (Feb. 2009)

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

佐古 淳[†1] 滝口哲也[†2] 有木康雄[†2]

近年，多くのマルチメディア・コンテンツの所有が可能となってきた．大量のコンテンツの中から欲しい情報を得るためには，検索のためのメタ情報を付与しておく必要がある．本研究では，マルチメディア・コンテンツの一例としてスポーツ実況中継，特に野球実況中継に注目し，実況中継音声から音声認識を用いてメタ情報を抽出することを目的としている．野球のメタ情報としては，今何が起こっているかを表すイベントと，その積み重ねである状況が存在すると考えられる．まず，現実にイベントや状況が存在し，これを基にアナウンサは実況を行う．本研究では，実況音声から単語列だけを推定する音声認識を拡張し，実況音声から単語列・イベント系列・状況系列すべてを同時に推定する音声認識手法を提案する．定式化により，イベント依存音響モデル，状況遷移モデル，イベント推定モデル，状況依存言語モデルを得る．これらを確率の枠組みで統合的に用いることで，単語列とメタ情報の同時推定を行う．実験により，イベント検出 F 値 0.87，イベント正解率 0.86，状況正解率 0.77 を得た．その他，各モデルの「メタ情報付与性能」への寄与や，音声認識率と「メタ情報付与性能」との関係について考察を行う．

Extracting Meta-information for Sports Live Games Based on Speech and Situation Recognition

Atsushi Sako,[†1] Tetsuya Takiguchi[†2] and Yasuo Ariki[†2]

Recently a large quantity of multimedia contents are broadcast and accessed. In order to retrieve exactly what we want to know from multimedia database, automatic extraction of meta-information is required. We focused on live speeches, especially baseball commentary speeches as a kind of multimedia contents. The purpose of this study is to provide meta-information based on speech recognition techniques. Events and situations are defined as meta-information. First of all, an event is occured or a situation is changed, then an announcer speaks based on an event and a situation. In this paper, we propose a extended speech recognition technique that estimates not only a word sequence but also a event sequence and a situation sequence concurrently. As a result of formulation, event dependent acoustic model, situation transition model, event estimation model and situation dependent language model are derived. A word sequence and meta-information are estimated based on these models. The experimental results showed that the proposed method provided meta-information with a high degree of accuracy.

1. はじめに

近年，デジタルテレビや WWW などの発展により，映像や音声など，多くのマルチメディア・コンテンツを所有することが可能となってきた．このような大量のコンテンツに対しては，ユーザが欲しい情報を検索できる必要がある．また，すべてのコンテンツを視聴するには時間がかかりすぎるため，要点だけ，または好みのシーンだけを抜き出して視聴したいという要求も存在する．一方で，放送局の立場としては，できるだけコストをかけずに新たなコンテンツを生み出したいという要求が存在する．ユーザが望むシーンのみを提供可能になれば，新たなコンテンツ・ビジネスとなる可能性がある．

このような要求を満たすためには，映像や音声などのコンテンツに対してメタ情報を作成しておく必要があると考えた．メタ情報を利用することにより，自動的にハイライトシーンを提供したり，ユーザの望むシーンを検索したりすることが可能になると期待できる．また，計算機により自動的にメタ情報の付与ができれば，人手で付与する場合に比べ，コストと労力の削減が可能となる．

本研究では，マルチメディア・コンテンツの 1 つとして，スポーツ実況中継，特に野球実況中継を対象としてメタ情報の作成を目的としている．メタ情報として，イベントの種別（投球・投球結果・ヒット・アウト・ホームラン・タイムリーなど）と状況（イニング・表裏・アウトカウント・出塁状況・ボールカウント）を定義する．メタ情報の詳細については 2 章で述べる．

野球中継に対してメタ情報を作成する研究としては，カメラワークを抽出して映像を構造化する手法[1)] や，映像中のテロップを解析する手法[2)]，クローズドキャプションを用いる手

†1 神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University
†2 神戸大学自然科学系先端融合研究環
Organization of Advanced Science and Technology, Kobe University

法[3] などが提案されている．しかし，野球の実況中継に対して正確に映像認識を行うことは難しい．そこで，本研究ではアナウンサの実況中継音声を音声認識し，この言語情報をもとにメタ情報を作成するアプローチを採用した．これは，放送局の立場として，コンテンツの成立に最低限必要な実況中継音声のみを用いてメタ情報を作成可能であるという利点がある．また，学術的観点からは，音声のみからどのように状況を認識しているかという理解の助けになるものと期待できる．本研究では，テレビではなく，ラジオの実況中継音声を用いた．これは，映像がないために，音声のみで状況が理解できるように実況中継がなされることによる．そのため，ラジオの実況中継音声には，テレビのものより情報が多く含まれる．

音声認識を用いてメタ情報を作成するにあたって，誤認識が問題となる．野球中継における実況放送音声の認識性能向上については，音響モデル適応・言語モデル適応を用いた手法が提案されている[4]．本研究も，同じく，音響モデル適応・言語モデル適応を行う．しかし，それでも単語正解精度は 7 割程度であり，依然として音声認識誤りが含まれている．したがって，認識誤りに対して頑健なメタ情報推定方法が必要となる．本研究では，実況音声がイベント・状況といったメタ情報に依存して生成されると仮定し，実況音声から単語列だけでなく，イベント・状況まで同時に推定する音声認識を提案する．これにより，音声認識器が最終的に出力する 1-best の認識結果だけでなく，はっきりと単語が確定していない認識仮説（本研究ではワードグラフを用いた）まで利用した統合的なメタ情報の付与が可能となる．

以下，次章で野球中継に関するメタ情報について述べる．3 章で提案手法である音声認識と統合されたメタ情報推定手法について述べ，4 章で実験について述べる．最後に 5 章でまとめる．

2. 野球中継に関するメタ情報

本章では，本研究で作成すべきメタ情報について述べる．本研究では，野球の実況中継音声を対象としている．野球自体のルールについては参考文献 18)，19) を参照されたい．

野球の実況中継音声は，アナウンサが野球の試合進行に基づいて，これを伝えるために発話を行ったものである．ただし，試合進行とは直接関係しない，選手の情報や球場の様子，解説者との会話などの発話も行う．本研究では，アナウンサの発話の内容を端的に表したものを「イベント」と呼ぶこととした．具体的なイベントとして，**表 1** のうち「重要なイベント」「その他のイベント」の内容を設定した．重要なイベントは，投球やストライク，ボール，ヒット，アウトなど，試合が進行し，試合の状況が変化する際になされる発話に対して付与される内容を設定した．また，その他のイベントは，「解説者との会話」や，選手の説明，現在の試合状況の説明などの「実況一般」，打者がボールを打ち，試合の状況がどう展開するかはっきり決まらない段階を表す「守備の実況」を設定した．これらは，試合の状態が変化しない場合になされる発話に対して付与される内容である．検索の際に用いられるイベントは，試合進行に関係する前者のものが多いだろうという考えから，前者のイベントを「重要なイベント」とした．

表 1　メタ情報の定義
Table 1　The specification of meta-information.

重要なイベント
投球　ストライク　ボール　ファール
フォアボール　三振　牽制球　盗塁
ヒット　ツーベース　スリーベース
ランニングホームラン※　ホームラン　得点
アウト　ダブルプレー　トリプルプレー※
ボーク※　デッドボール
その他のイベント
解説者との会話　実況一般　守備の実況
状況
イニング　表裏　得点　ストライクカウント
ボールカウント　アウトカウント　出塁状況

※についてはコーパス中には現れなかった．

本研究ではさらに，イベントの積み重ねとして，「状況」を定義した．状況は試合進行の状況と一致するように，イニング，および表裏，得点，アウトカウント，ストライクカウント，ボールカウント，出塁状況を定義した．本研究では，「イベント」「状況」をあわせて，メタ情報と定義した．**表 2** に実際の実況の書き起こし例と，人手で付与したメタ情報を記す．このような情報を自動的に付与することが本研究の目的である．ただし，1 つのイベントが必ず 1 つの発話で実況されるわけではない．この際，両方に同じラベルを付与すると状況の遷移がおかしくなるため，どちらかの発話を選んでラベルを付与するようにした．このとき，どちらを選択すべきか迷うような発話が，全体の 0.5%程度存在した．

3. 提 案 手 法

本章では，音声認識とメタ情報の付与を統合し，これらを同時に行う手法について述べる．以下，発話数を T，1 発話内のフレーム数を F_t，単語数を N_t とし，観測音声特徴の系列を $\mathbf{U} = \{\mathbf{O}_1, \cdots, \mathbf{O}_T\}$，$\mathbf{O}_t = \{o_1, \cdots, o_{F_t}\}$ とする．1 発話の長さは，おおよそ書き起こしテキ

表 2　実況中継に対するイベント・状況の具体例
Table 2　An example of events and situations for a commentary speech.

発話	イベント	状況
ほんとうにこの人は鉄人ですね，ええ．	会話	1 回裏 0 対 0 1S 2B 2O 1 塁
全イニング出場，全試合全イニング出場しています，バッターボックスの○○．	実況一般	1 回裏 0 対 0 1S 2B 2O 1 塁
ボールカウントワンストライクトゥーボール，投球四球目．	実況一般	1 回裏 0 対 0 1S 2B 2O 1 塁
ピッチャーの××，足が上がって第四球を投げました．	投球	1 回裏 0 対 0 1S 2B 2O 1 塁
高目，ストライク決まりました．	ストライク	1 回裏 0 対 0 1S 2B 2O 1 塁
シュートぎみ，アウトコースいっぱいに決まっています．	実況一般	1 回裏 0 対 0 2S 2B 2O 1 塁
ボールカウントトゥーストライクトゥーボール．	実況一般	1 回裏 0 対 0 2S 2B 2O 1 塁

ストの句点から句点までの長さである．ただし，書き起こしテキストを作成する際，句点の挿入に明確な基準があったわけではない．このため，句点は恣意的に挿入されたものである．また，発話の系列を $\mathbf{D} = \{\mathbf{W}_1, \cdots, \mathbf{W}_T\}$，1 発話内の単語の系列を $\mathbf{W}_t = \{w_1, \cdots, w_{N_t}\}$ とする．イベントと状況のメタ情報は 1 発話ごとに付与し，イベントの系列 $\mathbf{E} = \{e_1, \cdots, e_T\}$，状況の系列 $\mathbf{S} = \{s_1, \cdots, s_T\}$ とする．本研究においては，観測音声特徴の系列 $\mathbf{U}$ から，発話系列 $\mathbf{D}$，イベント系列 $\mathbf{E}$，状況系列 $\mathbf{S}$ を同時に推定することが目的となる．まず，次節において，本研究で仮定した実況中継音声の生成モデルについて述べる．

3.1　実況中継音声の生成モデル

本研究で仮定する実況中継音声の生成モデルを**図 1** に示す．実況中継音声は，試合の状況や起こったイベントをアナウンサが “実況” したものである．このため，発話される単語は，試合の状況とイベントに依存して生成されると考えた．また，起こったイベントの種類によっては，アナウンサは興奮を交えて実況する場合がある．そこで，音声は，単語とイベントに依存して生成されるものと考えた．さらに，イベントは状況に依存して生成されると考えた．これは，状況に応じて起こるイベントと起こらないイベントが存在するためである．最後に，状況は，以前の状況と以前のイベントに依存して生成されると考えた．以前の状況においてイベントが発生し，それにより変化した状況が現在の状況であることを意味している．これらをふまえ，次節において提案手法の定式化について述べる．

3.2　定　式　化

通常の音声認識では，発話全体の観測音声特徴の系列 $\mathbf{U} = \{\mathbf{O}_1, \cdots, \mathbf{O}_T\}$ から，発話の系列 $\mathbf{D} = \{\mathbf{W}_1, \cdots, \mathbf{W}_T\}$ を推定する．すなわち，観測特徴系列 $\mathbf{U}$ が既知の条件で，発話の系列 $\mathbf{D}$ が生成される確率 $P(\mathbf{D}|\mathbf{U})$ が最も高くなるような発話系列 $\hat{\mathbf{D}}$ を推定する．

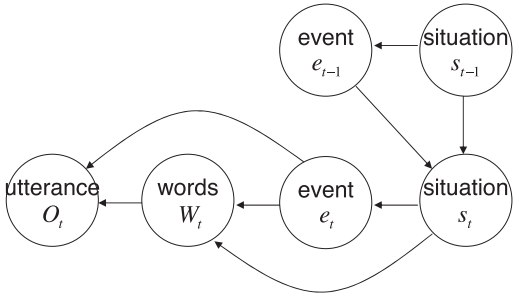

図 1　実況中継音声の生成モデル
Fig. 1　Generation model of commentary speeches.

$$
\begin{aligned}
\hat{\mathbf{D}} &= \operatorname*{argmax}_{\mathbf{D}} P(\mathbf{D}|\mathbf{U}) \\
&= \operatorname*{argmax}_{\mathbf{D}} P(\mathbf{W}_1, \cdots, \mathbf{W}_T|\mathbf{O}_1, \cdots, \mathbf{O}_T) \\
&= \operatorname*{argmax}_{\mathbf{D}} P(\mathbf{U})^{-1} P(\mathbf{W}_1) P(\mathbf{O}_1|\mathbf{W}_1) \\
&\quad \times \prod_{t=2}^{T} P(\mathbf{W}_t|\mathbf{W}_1^{t-1}, \mathbf{O}_1^{t-1}) P(\mathbf{O}_t|\mathbf{W}_1^{t}, \mathbf{O}_1^{t-1}).
\end{aligned} \tag{1}
$$

ベイズの定理により，式 (1) が導かれる．ここで，

- 発話 $\mathbf{W}_t$ は他の情報に依存しない
- 観測音声 $\mathbf{O}_t$ は発話 $\mathbf{W}_t$ のにみ依存する

との仮定をおくと，式 (1) は，

$$\hat{\mathbf{D}} = \underset{\mathbf{D}}{\operatorname{argmax}} P(\mathbf{W}_1)P(\mathbf{O}_1|\mathbf{W}_1)\prod_{t=2}^{T} P(\mathbf{W}_t)P(\mathbf{O}_t|\mathbf{W}_t) \tag{2}$$

となる．ただし，$P(\mathbf{U})^{-1}$ は $\mathbf{D}$ によらないため無視した．$P(\mathbf{O}_t|\mathbf{W}_t)$ は音響モデル，$P(\mathbf{W}_t)$ は言語モデルである．

これに対し，本研究では，発話全体の観測音声特徴の系列 $\mathbf{U} = \{\mathbf{O}_1, \cdots, \mathbf{O}_T\}$ から，尤もらしい発話系列 $\mathbf{D} = \{\mathbf{W}_1, \cdots, \mathbf{W}_T\}$，イベント系列 $\mathbf{E} = \{e_1, \cdots, e_T\}$，状況系列 $\mathbf{S} = \{s_1, \cdots, s_T\}$ を同時に推定する．

$$\begin{aligned}(\hat{\mathbf{D}}, \hat{\mathbf{E}}, \hat{\mathbf{S}}) &= \underset{(\mathbf{D},\mathbf{E},\mathbf{S})}{\operatorname{argmax}} P(\mathbf{D}, \mathbf{E}, \mathbf{S}|\mathbf{U}).\\ &= \underset{(\mathbf{D},\mathbf{E},\mathbf{S})}{\operatorname{argmax}} P(\mathbf{W}_1, \cdots, \mathbf{W}_T, e_1, \cdots, e_T, s_1, \cdots, s_T|\mathbf{O}_1, \cdots, \mathbf{O}_T)\\ &= \underset{(\mathbf{D},\mathbf{E},\mathbf{S})}{\operatorname{argmax}} P(s_1)P(e_1|s_1)P(\mathbf{W}_1|e_1, s_1)P(\mathbf{O}_1|\mathbf{W}_1, e_1, s_1)\\ &\quad \times \prod_{t=2}^{T} P(s_t|\mathbf{O}_1^{t-1}, \mathbf{W}_1^{t-1}, e_1^{t-1}, s_1^{t-1})P(e_t|\mathbf{O}_1^{t-1}, \mathbf{W}_1^{t-1}, e_1^{t-1}, s_1^{t})\\ &\quad P(\mathbf{W}_t|\mathbf{O}_1^{t-1}, \mathbf{W}_1^{t-1}, e_1^{t}, s_1^{t})P(\mathbf{O}_t|\mathbf{O}_1^{t-1}, \mathbf{W}_1^{t}, e_1^{t}, s_1^{t}).\end{aligned} \tag{3}$$

ここで，3.1 節に留意し，以下の仮定を置く．

- 状況 s_t は直前の状況 s_{t-1} と直前のイベント e_{t-1} にのみ依存する．
- イベント e_t は状況 s_t にのみ依存する．
- 単語列 $\mathbf{W}_t$ はイベント e_t と状況 s_t にのみ依存する．
- 観測特徴 $\mathbf{O}_t$ は，単語列 $\mathbf{W}_t$ とイベント e_t にのみ依存する．

以上の仮定により次式が導かれる．

$$\begin{aligned}(\hat{\mathbf{D}}, \hat{\mathbf{E}}, \hat{\mathbf{S}}) &= \underset{(\mathbf{D},\mathbf{E},\mathbf{S})}{\operatorname{argmax}} P(s_1)P(e_1|s_1)P(\mathbf{W}_1|e_1, s_1)P(\mathbf{O}_1|\mathbf{W}_1, e_1)\\ &\quad \times \prod_{t=2}^{T} P(s_t|e_{t-1}, s_{t-1})P(e_t|s_t)P(\mathbf{W}_t|e_t, s_t)P(\mathbf{O}_t|\mathbf{W}_t, e_t).\end{aligned} \tag{4}$$

$P(s_t|e_{t-1}, s_{t-1})$ は状況遷移モデル，$P(e_t|s_t)$ はイベント生成確率，$P(\mathbf{W}_t|e_t, s_t)$ はイベント・状況依存言語モデル，$P(\mathbf{O}_t|\mathbf{W}_t, e_t)$ はイベントに依存した音響モデルである．ここで，イベント・状況依存言語モデルは，各イベント・各状況ごとに作成した trigram を用いた．また，イベント依存音響モデルには，各イベントごとに作成した HMM を用いた．本手法では，イベント・状況に依存して単語列を生成するモデル $P(\mathbf{W}_t|e_t, s_t)$（実際にはイベント・状況依存 trigram）がイベントおよび状況の推定に大きな役割を果たす．

一方，$P(\mathbf{W}_t|e_t, s_t)$ を

$$P(\mathbf{W}_t|e_t, s_t) = \frac{P(\mathbf{W}_t|s_t)P(e_t|\mathbf{W}_t, s_t)}{P(e_t|s_t)}. \tag{5}$$

のように変形すると，言語モデルは状況のみに依存するようになり，認識仮説からイベントを推定するモデル $P(e_t|\mathbf{W}_t, s_t)$ が現れる．これを式 (4) に代入すると次式が導かれる．

$$\begin{aligned}(\hat{\mathbf{D}}, \hat{\mathbf{E}}, \hat{\mathbf{S}}) &= \underset{(\mathbf{D},\mathbf{E},\mathbf{S})}{\operatorname{argmax}} P(s_1)P(\mathbf{W}_1|s_1)P(e_1|\mathbf{W}_1, s_1)P(\mathbf{O}_1|\mathbf{W}_1, e_1)\\ &\quad \times \prod_{t=2}^{T} P(s_t|e_{t-1}, s_{t-1})P(\mathbf{W}_t|s_t)P(e_t|\mathbf{W}_t, s_t)P(\mathbf{O}_t|\mathbf{W}_t, e_t).\end{aligned} \tag{6}$$

状況遷移モデル，およびイベント依存音響モデルについては式 (4) と同様である．一方，イベント・状況依存言語モデル，およびイベント生成確率は，状況依存言語モデルとイベント推定モデルへ変化している．また，通常の音声認識である式 (2) と比較すると，言語モデル・音響モデルがそれぞれ状況依存・イベント依存に変化し，新しく状況遷移モデル・イベント推定モデルが追加されている．本手法では，イベント推定モデル $P(e_t|\mathbf{W}_t, s_t)$ を用いて，イベント e_t の推定を認識仮説 $\mathbf{W}_t$ から識別的に行う．この点が，イベント及び状況を生成モデルによって推定する式 (4) の手法と大きく異なっている．

以下，3.3 節から 3.6 節において，各モデルの詳細について述べる．

3.3　状況依存言語モデル

本節では，状況に依存した言語モデルについて述べる．また，比較のため，イベント・状況の両方に依存した言語モデルについても述べる．

本研究で用いる言語モデルは trigram をベースとする．学習データ量の関係から，同じ状況の発話だけを集め，そこから trigram を構築することは困難である．そこで，言語モデル適応を用いて，状況依存の言語モデルを構築する．言語モデル適応には，MAP 推定によるもの[5] や N-gram 出現回数の重み付き混合によるもの[6] が報告されている．本研究では後者の手法を用いて適応を行う．ただし，2 章で定義した状況のうち，イニング・表裏・得点については無視し，アウトカウント，ボールカウント，出塁状況のみに依存した言語モデル適応を行った．これは，イニング・表裏・得点まで限定してしまうと，そのような状況がコーパス中にほぼ 1 回しか出現しないためである（たとえば 9 回裏 1 対 1 といった状況）．本研究で用いたコーパスは，ラジオ放送の録音ではなく，放送局の提供によりアナウンサの接話マイクにより収録したものであった．そのため，攻守交代中は実況が行われておらず，

この際の長い無音区間によってイニング・表裏の推定が可能であった．ただし，イニング・表裏の情報は，提案手法の 4 つのモデルのいずれでも用いていないため，これによってその他の実験結果が影響を受けることはない．

また，比較のためのイベント・状況依存言語モデルについては，まず状況に応じた適応を行った後で，イベントに応じた適応を再度行うことにより構築した．状況については，状況依存言語モデルと同様，イニング・表裏・得点を無視した．イベントについては，すべてのイベントを考慮し，適応を行った．

3.4 イベント推定モデル

本節では，イベント推定モデルについて述べる．イベント推定モデル $P(e_t|\mathbf{W}_t, s_t)$ は，認識仮説の単語列からイベントを推定するモデルである．認識仮説の単語列は 10 単語程度になることもあり，数式通りに確率を計算すると相当にスパースなモデルとなってしまう．そこで，本研究では，識別的な手法を用いて認識仮説の単語集合からイベントを推定し，その後に確率化を行う．識別的な手法として AdaBoost を用い，その後に sigmoid 関数を利用して確率化する．以下，次項で AdaBoost について説明し，3.4.2 項において確率化手法について述べる．

3.4.1 AdaBoost

AdaBoost は，いくつもの識別器を組み合わせて 1 つの高度な識別器を構成する *ensemble learning method* の 1 つである．Schapire ら[7] が提案している学習のアルゴリズムを**図 2** に示す．図中，I は，$I(true)$ ならば 1，$I(false)$ ならば −1 となる．ϵ_t が 0.5 未満の弱学習器を見つけ続けることができれば，学習誤差 0 の最終学習仮説を生成できる．また，未知のサンプルに対する汎化誤差も小さくできることが実験的に報告されている[8),9)]．一方，雑音を有するサンプルの場合，過学習を起こすことが報告されている．これに対しては，AdaBoost の学習過程をマージン最大化ととらえ，SVM における Soft Margins の概念を導入した手法も提案されている[10),11)]．本研究では，認識結果を扱うため，サンプルには多くの雑音が乗っているものと考えられる．このことから，通常の AdaBoost ではなく，Soft Margins 付きの AdaBoost を用いることとした．

AdaBoost を用いたテキスト分類手法としては，文献 7)，12) などが提案されている．これらの文献では，テキスト分類のための弱学習器として，Decision Stumps が用いられている．Decision Stumps とは，ある素性の有無に基づいて分類を行う単純な手法である．素性には，単語 unigram や単語 bigram，ラベル付き順序木などが用いられる．本研究では，単語 unigram と単語 bigram 両方を素性の候補として用いた．学習時には，学習サンプルを最もうまく分類するような “素性” を選択し，その際の重みを得る．識別時には，式 (7) に従い，学習によって得られたすべての素性について，サンプル中にその素性があれば，重み α をスコアに加算，なければ減算することを繰り返し，最終的なスコアの正負によりクラスを識別する．

Input: n examples $Z = \{\mathbf{z}_1 = (\mathbf{x}_1, y_1), \cdots, \mathbf{z}_n = (\mathbf{x}_n, y_n)\}$

Initialize: $w_1(\mathbf{z}_i) = 1/n$ for all $i = 1 \cdots n$

Do for $t = 1, \cdots, T$,

(1) Train a base learner with respect to weighted example distribution w_t and obtain hypothesis $h_t : \mathbf{x} \mapsto \{-1, 1\}$

(2) Calculate the training error ϵ_t of h_t:

$$\epsilon_t = \sum_{i=1}^{n} w_t(\mathbf{z}_i) \frac{I(h_t(\mathbf{x}_i) \neq y_i) + 1}{2}.$$

(3) Set

$$\alpha_t = \log \frac{\epsilon_t}{1 - \epsilon_t}$$

(4) Update example distribution w_t:

$$w_{t+1}(\mathbf{z}_i) = \frac{w_t(\mathbf{z}_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}{\sum_{j=1}^{n} w_t(\mathbf{z}_j) \exp\{\alpha_t I(h_t(\mathbf{x}_j) \neq y_j)\}}.$$

Output: final hypothesis:

$$f(\mathbf{x}) = \frac{1}{||\alpha||_1} \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}).$$

図 2　AdaBoost のアルゴリズム
Fig. 2　AdaBoost algorithm.

$$f(\mathbf{x}) = \frac{1}{||\alpha||_1} \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}). \tag{7}$$

例として，ストライクとその他の発話を識別するための学習を行う場合を考える．仮に，「ストライク」という単語があればストライク，なければその他という分類が最も正解率が高い場合，弱識別器 $h_1(\mathbf{x})$ のための “素性” は「ストライク」という単語となる．この条件に従って分類を行うと，たとえば「ボールカウントトゥーストライク」という発話は，「ストライク」という単語があるにもかかわらずストライクではないため，識別誤り例となるであろう．AdaBoost では，このような識別誤り例の重みを大きくし，次回の弱識別器を選択す

る際に重視するようになる．たとえば，次の弱識別器 $h_2(\mathbf{x})$ は,「ボールカウント」という単語がなければストライクである，というように構成される．ここでは「ボールカウント」という単語が素性となる．このように，素性の選択と，学習サンプルの重み更新を繰り返しながら次々と弱識別器 $h_t(\mathbf{x})$ を構築していく手法が AdaBoost である．

3.4.2　イベント推定の疑似確率

AdaBoost は確率に基づく手法ではないため，出力されるスコアを式 (6) の中で用いるためには，AdaBoost の出力結果を確率化する必要がある．また，AdaBoost は 2 値判別手法であるため，多種のイベントを識別するためには one-vs-rest 法などを用いた拡張が必要である[13]．one-vs-rest 法では，あるクラス k に着目し，クラス k とそれ以外のクラスを識別する強識別器 $f_k(\mathbf{W}_t)$ を全クラス数分作成する．本研究では，sigmoid 関数と one-vs-rest 法を組み合わせ，式 (8) に従い，疑似確率 $P(e_t = k|\mathbf{W}_t, s_t)$ を求めた．

$$P(e_t = k|\mathbf{W}_t, s_t) = \frac{sigmoid(f_k(\mathbf{W}_t))}{\sum_l sigmoid(f_l(\mathbf{W}_t))}, \tag{8}$$

$$sigmoid(x) = \frac{1}{1 + exp(-w_1 x - w_0)}. \tag{9}$$

ここで，w_1 および w_0 は sigmoid 関数における重み係数であり学習により推定する．

AdaBoost の出力スコアは，識別境界面からの距離と捉えられることが報告されている[11]．すなわち，境界に近いほど確率 0.5 に近く，境界から遠いほど確率 0，もしくは確率 1 に近づくと考えられる．本研究では，完全な確率とはいえないものの，式 (9) の sigmoid 関数を利用し，AdaBoost の出力スコアを疑似確率化する．ただし，イベントは多クラスであることから，$\sum_l P(e_t = l|\mathbf{W}_t, s_t) = 1$ の制約を満たすために式 (8) のように正規化して用いる．sigmoid 関数は**図 3** に示すような関数であり，最小値 0，最大値 1 となる関数である．また，識別境界付近を詳細にモデル化できる特徴を持つ．

3.5　状況遷移モデル

本節では，状況遷移モデルについて述べる．状況遷移モデルは，試合進行の状況がイベントによって変化していくことを表すルールモデルとなる．モデルの一部を**図 4** に示す．このモデルにより，ストライクカウントは 2 まで，1 度に 1 ずつしか増えない，フォアボールイベントが生じるのはボールカウントが 3 のときのみ，といった野球のルールを表現することができる．また,「ストライクイベントが起きた際にはストライクカウントが 1 増える」といったルールが表現されていることも重要な点である．状況は，状況依存言語モデルからも推定可能である.「ボールカウント，ワンエンドツー．投げた．ストライク．ボールカウント，ツーエンドツーになりました」と実況している場合は，状況依存言語モデルからだけでも，どの点で状況が変化したか推定することができる．しかし,「ボールカウント，ワンエンドツー．投げた．ストライク」とだけ実況している場合は，状況依存言語モデルからでは変化をとらえることができない．本モデルを用いて，“以前の状況”，および “ストライクイベントの発生” から現在の状況を推定する必要がある．状況遷移モデル $P(s_t|e_{t-1}, s_{t-1})$ は，人手によってメタ情報が付与された学習コーパスから直接計算することができる．

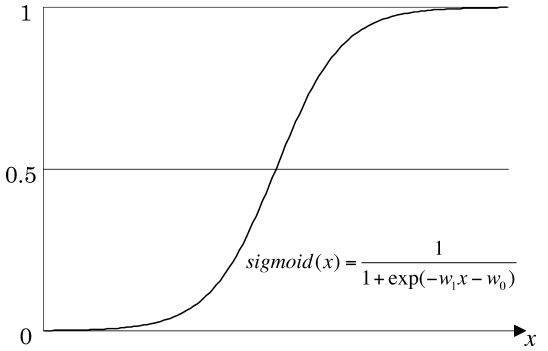

図 3　sigmoid 関数
Fig. 3　sigmoid function.

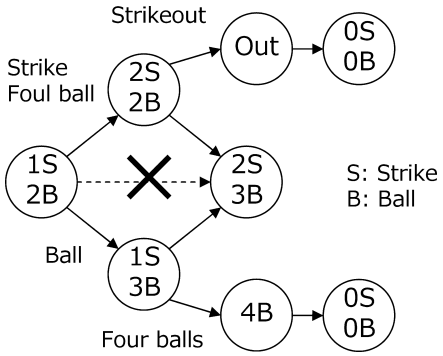

図 4　状況遷移モデル
Fig. 4　Situation transition model.

3.6　イベント依存音響モデル

本節では，イベント依存音響モデルについて述べる．本研究で取り扱うラジオの実況中継音声は，講演音声と比較しても発話速度が速く（講演音声 7.31 mora/s に対し，実況中継音声 8.51 mora/s），雑音レベルが強い，感情の起伏も激しいといった特徴がある[4]．本研究では認識性能向上のため，文献 4) と同じ手法を用いて教師ありの音響モデル適応を行う．適応のベースラインとなる音響モデルは，比較的「話し言葉」に近い特徴を持つ日本語話し言葉コーパス（CSJ: Corpus of Spontaneous Japanese）モニタ版[14] から作成した．また，適応手法として，MLLR[15] や MAP 推定[16] を単独で用いるよりも高精度に適応できるとされる MLLR+MAP[17] を用いた．これは，MLLR によってモデルパラメータの変換を行い，それを事前知識として MAP 推定を行う手法である．

これに加え，イベントごとの音響モデル適応も行う．まず，上記の手法で CSJ コーパスから作成したモデルに，すべてのイベントを含む実況中継音声を用いて適応を行う．こうして得られたモデルをイベントの数だけ複製し，それぞれにイベントごとの実況中継音声を用いて再度適応を行う．これによりイベントに依存した音響モデルを得る．守備の実況やホームランイベントなどにおいて，臨場感を伝えるための興奮した音声に適応されたモデルが得られるものと考えられる．

4. 実　　験

本章では，本研究で行った実験について述べる．実験の目的は，提案手法によってどの程度の精度でメタ情報を付与することができるか確かめることと，提案手法の中の 4 つのモデルそれぞれが，どの程度精度に寄与しているか確認することである．ここで，精度の指標として，以下の 3 点を用いた．

- 重要なイベント検出の F 値
- 重要なイベントの正解率（推定されたイベントと正解ラベルが一致しているか）
- 投球ごとの状況正解率

1 つ目の重要なイベント検出の F 値は，実況一般や解説者との会話といった検索の条件になりにくいイベントの中から，検索の条件になるホームランや三振などの重要なイベントが正しく検出できているかを調べるための指標である．なお，重要なイベントは 2 章で定義したものを用いた．2 つ目の重要なイベントの正解率は，重要なイベントが正しく検出されたうえでそのラベルが正解ラベルと等しい割合である．イベントの内容まで正しく推定できているかを調べるための指標である．3 つ目の投球ごとの状況正解率は状況を正しく推定できているかを調べるための指標である．発話ごとでなく投球ごとである理由は，本来，投球ごとにしか試合の状況が変化しないこと，実際はアナウンサの言い直しなどにより投球以外の場合でも状況が変化する場合があることによる．以上の指標を用いて，提案手法の評価実験を行った．まず，人手による書き起こしテキストを用いたメタ情報付与実験を行い，システムの上限値を調べた．その後，実況中継音声に対して，提案手法を用いてメタ情報付与実験を行った．

ただし，実験に際しては，式 (6) の各モデルに，重みを乗じて用いた．これは音声認識における言語重みと同様の調整パラメータである．今回の実験では，網羅的な重みの探索は行っておらず，言語重み，イベント推定モデル重み，状況遷移モデル重みの順で，重みの探索を行った．具体的には，

- まず，通常の音響モデル・言語モデルのみを用いて，単語正解精度を指標に最適な言語重みを探索し，この言語重みを状況言語モデルの重みにそのまま用いる，
- 次に，通常の音響モデル・言語モデル（状況依存でないもの）とイベント推定モデルを用いて，イベント検出 F 値を指標にイベント推定モデルの重みを探索する．以後，イベント推定モデルの重みにはこのときの重みを用いる，
- すべてのモデルを用い，状況正解率を指標に，状況遷移モデルの重みを探索する，

という方法で重みを決定した．

以下，次節において，学習コーパスの仕様について述べる．

4.1　学習コーパスの仕様

本節では，実験で用いた学習コーパスについて述べる．実況中継音声は，ラジオの音声を用いた．これは，映像がないため，テレビの実況中継より音声の情報量が多いためである．使用した音声データは，**図 5** のとおりラジオ放送される前の，アナウンサのみの音声を収録したものである．球場の歓声などはノイズとして重畳はしているものの，音声認識に大きく影響をあたえるほどではない．また，解説者との会話は存在するが，解説者の声は含まれていない．発話速度が速い，言い間違いが多い，発音がなまけているなどの特徴があり，音声認識にとっては困難なタスクとなっている．

発話の単位は，人手による書き起こしテキストにおいて句点で区切られた単位とした．音声認識，およびメタ情報の付与はともに，1 発話ごとに行った．メタ情報付与のための学習データは，発話ごとにメタ情報ラベルを人手で付与し，作成した．学習データの分量を**表 3** に示す．全 4 試合で，1 試合あたり約 2,000 発話であった．総発話数は約 9,000，単語数は 80 K，異なり単語数（辞書サイズ）は約 3,000 語であった．

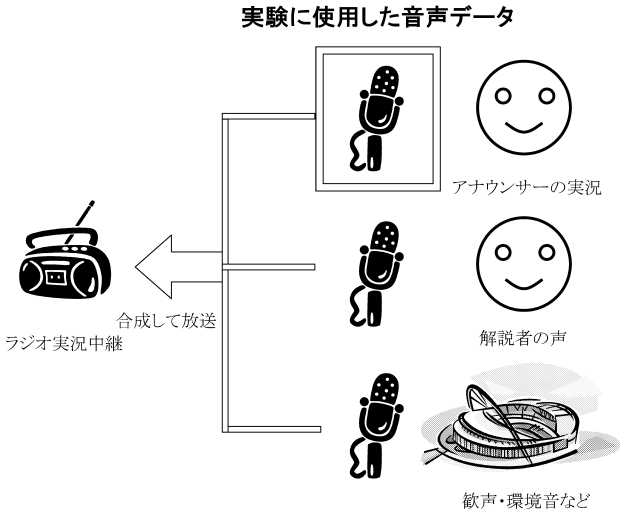

図 5　実験で使用した音声データ
Fig. 5　Recording environment of speech corpus.

表 3　学習コーパスの仕様
Table 3　The specification of our corpus.

日時	話者	時間	発話数	単語数
2003/09/05	A	1.73	2,232	21 K
2003/09/06	B	1.81	2,210	22 K
2003/09/15	B	1.76	2,320	21 K
2003/09/16	A	1.61	2,010	20 K

4.2　人手による書き起こしテキストを用いたメタ情報の付与

システムの上限値を調査するため，人手によって書き起こされたクリーンな（認識誤りの含まれていない）テキストに対するメタ情報付与実験を行った．書き起こしテキストを用いるため，提案手法のうちイベント依存音響モデルは利用せず，状況遷移モデル・イベント推定モデル・状況依存言語モデルのみを用いて実験を行った．

実験は，4 fold のクロスバリデーション法により行った．実験結果を**表 4** に示す．

イベント検出 F 値，イベント正解率はともに 9 割以上の精度で識別ができている．状況正解率が他に比べて低い値なのは，1 度間違えると状況依存言語モデルなどで修正されるまで誤りが継続してしまうためと考えられる．そのほか，識別に失敗しているものは，出現頻度の低い表現を用いた発話や，人手でもラベルの付与に迷うような発話などであった．

表 4　クリーンな書き起こしテキストによるメタ情報付与結果
Table 4　Results of extracting meta-information using clean transcription.

イベント検出 F 値	0.93
（再現率/適合率）	(0.94/0.92)
イベント正解率	0.92
状況正解率	0.87

表 5　音響分析条件と HMM の仕様
Table 5　Condition of acoustic analysis and HMM specification.

音響分析	サンプリング周波数	16 kHz
	特徴パラメータ	MFCC（25 次元）
	フレーム長	20 ms
	フレーム周期	10 ms
	窓タイプ	ハミング窓
H	タイプ	244 音節
	混合数	32 混合
M	母音 (V)	5 状態 3 ループ
M	子音+母音 (CV)	7 状態 5 ループ

4.3　実況中継音声を用いたメタ情報の付与

次に，実況中継音声から提案手法を用いてメタ情報を付与する実験を行った．本節では，実況中継音声からのメタ情報付与精度，音声認識とメタ情報付与の統合の効果，各モデルの精度に対する寄与，音声認識率とメタ情報付与精度の関係の 4 点について明らかにする．まず，音声認識の条件と結果について述べる．

4.3.1　音声認識条件と結果

ベースラインの音響モデルは，日本語話し言葉コーパス（CSJ: Corpus of Spontaneous Japanese）モニタ版[14] のうち，男性話者 200 名の講演音声を用いて作成した．音響分析条件と HMM の仕様を**表 5** に示す．これらの条件で音響モデルを作成し，さらに，MLLR+MAP [17] により音響モデル適応を行った．音響モデル適応は，同一話者の別の日時の実況中継音声を用いて行った．適応データの分量は，約 1 時間半であった．

言語モデルは，野球実況中継音声の書き起こしテキストから trigram モデルを作成した．異なり単語数は約 3,000，コーパスサイズは約 8 万形態素であった．

4 つの試合の音声認識結果の単語正解精度の平均，およびキーワードの F 値の平均を**表 6**

表 6　音声認識結果の単語正解精度
Table 6　Word accuracy of the speech recognition results.

単語正解精度	キーワード F 値（再現率/適合率）
63.8%	0.76 (0.74/0.77)

表 7　AdaBoost によって選択された素性語の例
Table 7　An example of features selected by AdaBoost.

メタ情報識別素性例
あたり あんまり きのう ほんと よく アウトー
インサイド ストライク スリー ツーアウト バッター ボール
ワン 一塁 回っ 外れ 甘い 監督 詰まっ 球 空振り 牽制
三振 始まり 送球 打ち 直球 変化球 方向 etc.

に示す．ここでのキーワードは AdaBoost によって学習された素性を用いた（**表 7**）．以後，これらの認識結果を用いて実験を行った．

4.3.2　音声認識結果に対するメタ情報の付与

実況中継音声に対し，提案手法によるメタ情報付与実験を行った．学習とテストはクリーンな書き起こしテキストの場合と同様に 4 fold のクロス・バリデーション法により行った．音響モデル・言語モデルともにオープンの条件において，提案手法を用いた場合のメタ情報付与実験結果を**表 8** に示す．ここで，“1-best” の結果は，提案手法ではなく，音声認識とメタ情報付与を統合しない場合の結果である．すなわち，まず通常の音声認識結果を出力し，その 1 通りの結果に対して提案手法の各モデルを用いてメタ情報付与を行ったものである．認識と統合せず，認識仮説を用いないところが提案手法と異なる点である．音響モデルについても，イベント依存のものではなく，通常の音響モデルを用いた．また，“SE trigram（Situation and Event dependent trigram）” は，式 (6) の提案手法ではなく，イベント・状況依存言語モデルを用いる式 (4) に基づく手法である．

提案手法における識別性能について，クリーンテキストの場合に比べて精度は低下するものの，高精度な識別ができている．ただし，イベントの正解率を詳細に見てみると，投球・投球の結果については比較的高い精度を保っているものの，ヒットやアウトといった試合が大きく動くイベントの正解率に低下が見られた．このようなイベントの際には，臨場感を伝えるためにアナウンサが興奮して発話を行う場合があり音声認識率が低下する．認識率の低下がイベントの正解率に影響を与えているものと考えられる．提案手法では，イベント依存音響モデルの効果により興奮した音声の音声認識率改善を期待していたが，効果は限定的であった．イベント依存音響モデルの効果については 4.3.4 項で述べる．

表 8　提案手法によるメタ情報付与実験結果
Table 8　Results of extracting meta-information using recognized transcription.

	提案手法	1-best	SE trigram
イベント検出 F 値	**0.87**	0.85	0.75
（再現率/適合率）	(0.88/0.85)	(0.85/0.84)	(0.73/0.76)
イベント正解率	**0.86**	0.83	0.78
状況正解率	0.77	0.74	0.67
単語正解制度	63.9%	63.8%	63.9%

次に，提案手法による結果を “1-best” と比較した場合について述べる．認識の結果，イベント検出 F 値およびイベント正解率については提案手法が高い性能を示している．これらについては，1-best の結果との比較で，二項分布の平均の差の検定により有意水準 5%で有意であった．提案手法では，状況やイベントまで考慮に入れたうえで認識仮説を選択することが可能である．これに対し，“1-best” では，音声認識誤りを修正する手段を持たない．提案手法では，たとえば具体例として，ボールカウントが 3 でない場合の「フォアボール」を正しい「ファールボール」に修正できている例や,「三振」と「阪神」の誤認識が修正されている例などが見られた．ただし，上記のようなキーワードについてはいくつか改善が見られたが，大部分の認識結果は共通しており，単語正解精度としてはほとんど変化がなかった．状況に依存する単語が，ルールに関連する用語などにある程度限定されているものと考えられる．認識結果が変わらない部分については提案手法と “1-best” は同じ結果となった．また，“SE trigram” では，精度の低下が見られた．AdaBoost を利用したイベント推定と異なり，イベント・状況依存言語モデルでは，識別性能が低下してしまうものと考えられる．状況正解率については，1-best との比較で，有意水準 5%では有意な差とならなかった．原因として，ヒットなどの認識性能が大きく低下する発話に対しては，提案手法でもほとんど改善が得られていないことが考えられる．さらには，1 度の誤りがしばらく連鎖して続くため，全体的に正解率が低下してしまうことが考えられる．

4.3.3　音声認識率とメタ情報正解率の関係

本項では，音声認識率とメタ情報正解率との関係について述べる．本来は多種多様なデータで実験を行い，関係を明らかにすることが望ましいが，データが限られているため，音声認識条件のうち，音響モデル・言語モデルのオープン・クローズドを変化させることで 4 通りの精度で実験を行った．ただし，実験は 2003/09/05 の実況中継音声のみを用いて行った．音響モデル適応は 4.3.1 項と同じ手法を用い，オープンの場合で約 1 時間半，クロー

表 9　音声認識結果の単語正解精度（2003/9/5 の場合）
Table 9　Word accuracy of the speech recognition results (2003/9/5).

音響モデル	言語モデル	単語正解精度	キーワード F 値
オープン	オープン	65.0%	0.80
クローズド	オープン	70.4%	0.84
オープン	クローズド	73.4%	0.85
クローズド	クローズド	78.3%	0.88

ズドの場合で約 3 時間のデータを用いて行った．言語モデル適応は，学習コーパス中にテストセットを含めたものをクローズド，含めなかったものをオープンとした．通常の音声認識の結果を**表 9** に示す．音響モデルについては，オープンであっても話者は同一であるため，言語モデルをクローズにする方が効果が大きくなっている．また，キーワードの F 値は，ほぼ単語正解精度に比例する結果が得られた．

また，状況推定モデルの AdaBoost の学習方法について，人手による書き起こしテキストから学習する方法と，誤りを含む認識結果から学習する手法が考えられる．ここでは，それぞれの場合における認識精度の変化について調べた．結果を**図 6** に示す．横軸が，音声認識条件を変化させたことによる単語正解精度，縦軸が重要なイベントの正解率である．

まず，結果より，認識結果を用いて識別素性の学習を行った方が高精度にメタ情報を識別できることが分かる．これは，AdaBoost によって認識誤りによらない頑健な素性が選択されるためと考えられる．また，人手による書き起こしテキストを用いて学習を行った場合，音声認識精度の低下とともに，イベント正解率も低下する傾向が見られる．これに対し，認識結果を用いて学習を行った場合，低下の傾向が緩やかであった．音声認識結果を用いて学習を行うことにより，認識性能に対して頑健なメタ情報の付与ができるものと考えられる．

4.3.4　提案手法における各モデルの寄与

本項では，提案手法における 4 つのモデルが，それぞれメタ情報付与精度にどの程度寄与しているかについて述べる．結果を**表 10** に示す．結果は，すべてを用いた場合から，各項目のモデルを排除した場合の結果である．E 音響はイベント依存音響モデル，E 推定はイベント推定モデルを表している．また，イベント依存音響モデルを排除することは，通常のイベント依存しない音響モデルを用いることを意味している．同様に，状況依存言語モデルを排除する際には，通常の言語モデルを用いた．

最も影響の大きなモデルは，イベント推定モデルであった．イベント推定モデルを除くと，イベント依存音響モデルのみを用いてイベント推定を行うことになる．しかしながら，イベント依存音響モデルによるイベント推定性能は非常に低く，特にイベントの内容までについては，ほぼ推定能力がないと考えられる．最も影響が少なかったのはイベント依存音響モデルであった．原因として，アナウンサが興奮した際には，興奮で適応した音響モデルであっても認識誤りが著しいこと，興奮していない際には，通常の音響モデルでもイベント依存音響モデルでも，精度に大きな差がなかったことがあげられる．そのほか，状況遷移モデル・状況依存言語モデルについては，状況正解率は低下したものの，イベント検出 F 値・イベント正解率は大きく低下しなかった．理由として，認識仮説が大きく変わらなかったこと，仮

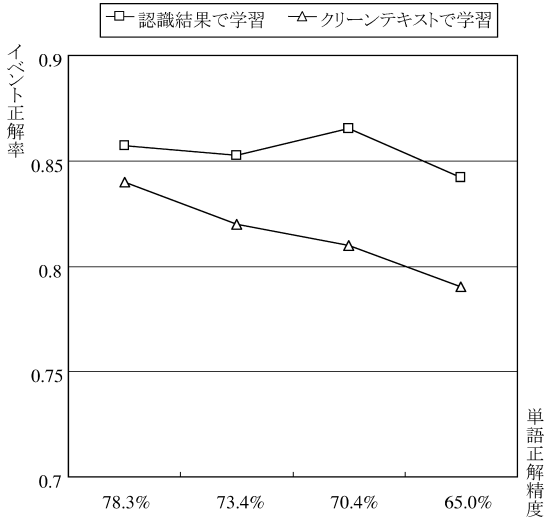

図 6　音声認識性能とメタ情報付与性能の関係（2003/9/5 の場合）
Fig. 6　Relations between ASR performance vs. extracting meta-information performance (2003/9/5).

表 10　提案手法から各モデルを排除した際のメタ情報付与精度
Table 10　Results of meta-information excepting models.

	E 音響	状況遷移	E 推定	状況言語
イベント検出 F 値	0.86	0.85	0.33	0.85
（再現率/適合率）	(0.87/0.85)	(0.86/0.84)	(0.35/0.32)	(0.86/0.84)
イベント正解率	0.85	0.83	0.04	0.83
状況正解率	0.77	0.58	0.58	0.45
単語正解制度	63.8%	63.9%	63.9%	63.8%

に変わっていても，AdaBoost で用いるキーワードには変化が少なかったこと（AdaBoost において，変化しにくい単語が素性として選択される傾向がある）があげられる．単語正解精度については，表 8 の結果と同様，ほとんど変化がなかった．

5. ま と め

本稿では，音声認識とイベント・状況推定を同時に行うことにより，メタ情報を付与する手法について述べた．実況音声が生成される過程をモデル化し，観測音声特徴から発話系列・イベント系列・状況系列を同時に推定するよう定式化を行った．これにより，イベント依存音響モデル・状況遷移モデル・イベント推定モデル・状況依存言語モデルを得た．特に，イベント推定モデルについては，識別的手法である AdaBoost を用いた．AdaBoost の出力スコアは確率ではないため，sigmoid 関数を利用し疑似確率化して用いた．

実験の結果，提案手法でイベント検出 F 値 0.87，イベント正解率 0.86，状況正解率 0.77 を得た．また，音声認識性能を変化させつつ実験したところ，性能に著しい差がなかったことから，音声認識性能に対して頑健にメタ情報の付与ができるものと考えられる．これは，AdaBoost の素性を学習する際，音声認識結果を学習データとすることにより，認識誤りに対して頑健な素性が選択されるためと考えられる．提案手法を用いることにより，実況音声が生成される背後にある知識が利用可能となり，これによって音声認識誤りが改善される例が見られ，メタ情報付与性能も向上した．

本研究では，状況やイベントが定義しやすい野球を対象として研究を行った．しかしながら，他のスポーツへ展開するためには，本研究で定義した状況は “堅すぎる” であろう．特に，球技では，ボールや選手の位置が状況として重要な意味を持つ可能性がある．音声からでは伝えにくいこのような状況を理解するためには，映像情報との統合も検討する必要がある．今後の課題として，その他のスポーツへのメタ情報の付与を行うため，映像の利用による位置情報の抽出，ボールや選手の位置といった，緩い状況を表現可能なモデルの提案などの研究が必要である．また，同時に，メタ情報には様々な粒度が存在する（単なるアウトか，フライでのアウトか，ファールフライかなど）．緩い状況の表現とともに，より詳細な状況やイベントを表現するためにも，階層的なイベント・状況モデルの研究が必要である．

参 考 文 献

1) 山本　拓，佐藤宏介，千原國宏：野球中継映像における各種プレイシーンの自動検索/編集システム，2000 信学総大，情報・システム 2，D12-77, p.247 (2000).
2) 館山公一，川嶋稔夫，青木由直：野球中継におけるシーン検索，第 3 回知能情報メディアシンポジウム論文集，pp.195–202 (1997).
3) 新田直子，馬場口登，北橋忠広：言語の画像の情報統合によるスポーツ映像からの人物・アクション・イベント抽出，信学技報，PRMU99-256 (2000).
4) 有木康雄，緒方　淳，藤本雅清，塚田清志：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（D-II），Vol.J87-D-II, No.6, pp.1208–1215 (2004).
5) 政瀧浩和，匂坂芳典，久木和也，河原達也：最大事後確率推定による N-gram 言語モデルのタスク適応，信学論（D-II），Vo.J81-D-II, No.11, pp.2519–2525 (1998).
6) 伊藤彰則，好田正紀：N-gram 出現回数の混合によるタスク適応の性能解析，信学論（D-II），Vol.J83-D-II, No.11, pp.2418–2427 (2000).
7) Schapire, R., Freund, Y., Bartlett, P. and Lee, W.: Boosting the margin: A new explanation for the effectiveness of voting mathods, *Annals of Statistics*, Vol.26, No.5, pp.1651–1686 (1998).
8) Freund, Y. and Schapire, R.: Experiments with a new Boosting algorithm, *Proc. 13th International Conference on Machine Learning Bari*, pp.148–146, Morgan Kaufmann, Italy (July 1996).
9) Schwenk, H. and Bengio, Y.: Adaboosting neural networks, *Proc. ICANN'97*, Vol.1327 of LNCS pp.967–972, Springer, Berlin (Oct. 1997).
10) Ratsch, G., Onoda, T. and Muller, K.-R.: Soft Margin for AdaBoost, *Machine Learning*, Vol.42, No.3, pp.287–320 (2001).
11) 小野田崇：Boosting の過学習とその回避，電子情報通信学会論文誌，Vol.J85-D2, No.5, pp.776–784 (2002).
12) 工藤　拓，松本裕治：半構造化テキストの分類のためのブースティングアルゴリズム，情報処理学会論文誌，Vol.45, No.9 (2004).
13) Alpaydin, E.: *Introduction To Machine Learning* (*Adaptive Computation and Machine Learning*), The MIT Press (2004).
14) 古井貞熙，前川喜久雄，伊佐原均：『話し言葉工学』プロジェクトのこれまでの成果と展望，第 2 回話し言葉の科学と工学ワークショップ，pp.1–6 (2002).
15) Leggetter, C.L. and Woodland, P.C.: Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models, *Comput. Speech Lang.*, Vol.9, pp.171–185 (1995).
16) Gauvain, J.L. and Lee, C.H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains, *IEEE Trans. Speech Ausio Process.*, Vol.2, No.2, pp.291–298 (1994).
17) 緒方　淳，有木康雄：音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応，信学技報，SP2001-105 (2001).
18) 日本プロフェッショナル野球組織：公認野球規則 2008，日本プロフェッショナル野球組織 (2008).

19) 粟村哲志：わかりやすい野球のルール，成美堂出版 (2008).

(平成 20 年 6 月 4 日受付)
(平成 20 年 11 月 5 日採録)

佐古　淳（学生会員）

2004 年龍谷大学理工学部電子情報学科卒業．2006 年神戸大学大学院自然科学研究科博士課程前期課程修了．現在，同大学院自然科学研究科博士課程後期課程在学中．音声情報処理に従事．日本音響学会，電子情報通信学会各会員．

滝口　哲也（正会員）

1999 年奈良先端科学技術大学院大学博士後期課程修了．1999 年日本アイ・ビー・エム東京基礎研究所．2004 年神戸大学工学部講師．博士（工学）．音情報処理，画像処理等の研究に従事．日本音響学会，電子情報通信学会，IEEE 各会員．

有木　康雄（正会員）

1974 年京都大学工学部情報工学科卒業．1976 年同大学大学院修士課程修了．1979 年同大学院博士課程修了．1980 京都大学工学部情報工学科助手．1990 年龍谷大学理工学部電子情報学科助教授，1992 年同教授．2003 年神戸大学工学部教授．工学博士．1987〜1990 年エディンバラ大学客員研究員．画像処理，音声情報処理に従事．電子情報通信学会，映像情報メディア学会，日本音響学会，人工知能学会，画像電子学会，IEEE 各会員．