Podspotter: 音リアクションイベント検出に基づくポッドキャストブラウザ

Podspotter: Podcast Browser Based on Acoustic Event Detection

須見 康平　河原 達也　緒方 淳　後藤 真孝*

Summary. 本稿ではポッドキャストを対象として，音リアクションイベント，特に笑い声とあいづちの検出に基づくホットスポットの提示を行うインタフェース Podspotter を提案する．Podspotter は一覧性の低い音声コンテンツであるポッドキャストに対して，発話内容の音声認識ではなく，対話の中の有用な非言語情報を検出し，その先行発話を提示する．これにより，ユーザが興味を示すような区間や，ユーザにとって有益な情報を含んでいる区間を抽出できると期待される．また音響イベントの色分け表示とそれに基づく再生機能を提供することで，コンテンツ全体の音響的な流れを見ながら選択的に視聴することが可能である．

1　はじめに

ポッドキャストなどの音声・音からなるコンテンツは，一度すべてを聴かなければ内容や情報の所在を把握することができないため，オンデマンドな検索や閲覧が非常に困難であるという問題がある．これに対して，音声認識を用いたテキスト化によって，検索・閲覧を可能にするサービス（Google Audio Indexing[1]，PodCastle[2, 3] など）が実現されている．しかし，多くの音声メディアは純粋な音声だけでなく，音楽や音響効果，環境音，背景雑音などの多くの要素を含み，現状の技術によって，内容を完全に把握するのは困難である．また自由なスタイルの発話が多く，話し言葉特有の言い回しや多人数での同時発話などもあるため，音声認識は容易ではない．

そこで我々は，笑い声やあいづちなどの人のリアクションによって生じる音響的な非言語情報（音リアクションイベント）に着目し，その検出を行うことで，視聴者が興味を持ちそうな箇所（＝ホットスポット）を特定することができるのではないかと考える．笑い声はおもしろいと思わせる発話が出現した直後に起こることが多く，あいづちは聞き手の関心の度合いを表す機能をもつため [4]，これらのイベントの直前にホットスポットの候補が存在する可能性が高い（図 1）．本研究では，音リアクションイベント検出手法 [5] によって得られる音響イベント系列をもとに，効率的な視聴を可能にするポッドキャスト視聴インタフェース **Podspotter** を提案する．

2　Podspotter の特長

本研究で提案する Podspotter は，音リアクションイベント検出に基づく 2 種類のホットスポットの提示機能と各音響イベントの視覚化機能を持つ．これらの機能により，ポッドキャストの音声データを効率的に部分抽出して視聴することができる．

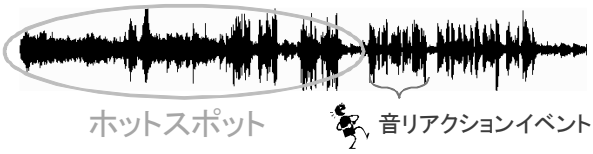

図 1. 音リアクションイベントとホットスポット

Podspotter を用いてポッドキャストを視聴するためには，ポッドキャストの音声データ（MP3 形式）とタイムスタンプが付与されたイベントタグが必要である．音声データはポッドキャストの各配信先からダウンロードすることが可能であり，イベントタグはそれらのデータに対して音響的な解析を行うことにより得られる．我々が提案した音リアクションイベント検出手法では，8 つの音響イベント（男性音声，女性音声，音楽，男性音声と音楽の混合，女性音声と音楽の混合，笑い声，あいづち，無音）を検出対象とし，実際のポッドキャストデータに対して，音声・音楽・混合に関して 85%，笑い声 74%，あいづち 64%の精度で自動検出することが可能となっている [5].

2.1　2 種類のホットスポッティング

本研究では「おもしろスポット」と「なるほどスポット」と名づけた 2 種類のホットスポットに着目し，部分抽出して提示を行う．それぞれのスポットを以下のように定義する．

- **おもしろスポット**
  笑い声の直前に焦点をあてた切り出し．
- **なるほどスポット**
  「あー」,「へー」,「ふーん」といった関心を示すあいづち [4] に基づく切り出し．

各ホットスポット区間を，分割セグメント数と時間長の制約，さらにイベント境界が存在するか否か

* Kouhei Sumi and Tatsuya Kawahara, 京都大学大学院 情報学研究科 知能情報学専攻, Jun Ogata and Masataka Goto, 産業技術総合研究所

図 2. Podspotter の概観

に応じて決定する．セグメント数 N_{max} 以下かつ時間長 D_{max} 秒以下を満たし，イベント境界を含む場合は，セグメント数と時間長をできるだけ大きくするような境界までをホットスポットして切り出す．現在の実装では，$N_{max} = 10$，D_{max} は笑い声の場合は 20 秒，あいづちの場合は 25 秒と設定した．

2.2 イベントブラウジング機能

8 つの音響イベントを各セグメントごとに色分けして提示することで，音響的な情報を視覚化することが可能となる．例えば，男性話者から女性話者への切り替わり点や笑い声やあいづちの出現箇所を知ることができ，効率的にポッドキャストを視聴できる．

3 Podspotter のインタフェース

Podspotter は，各機能を実現する複数のパネルとコントロール部から構成される（図 2）．このインタフェースを，Adobe Flex 及び ActionScript を用いて実装した．これにより，Flash 環境に対応したウェブブラウザ上で動作させることが可能であるため，OS に関わらず利用することができる．以下では Podspotter の主要な機能を担う全体表示パネルとホットスポットパネルについて説明する．

3.1 全体表示パネル

各音響イベントを色分けして表示することで，イベントブラウジング機能を実現する．男性音声（青），女性音声（赤），音楽（緑），笑い声（黄），あいづち（青緑）を各ブロックによって表現する．音楽と音声の混合区間については，半分ずつ色分けされたブロックで表し，無音区間はブロックを表示しない．

ブロック列の下線はタイムラインを表し，左端から右端までの時間単位を 10 秒から 180 秒の範囲で変化させることができる．ブロックと下線はどちらもクリックされた箇所からの再生が可能で，再生時には再生箇所のブロックのハイライトと，下線上に再生ポインタが表示される．

3.2 ホットスポットパネル

抽出した「おもしろスポット」と「なるほどスポット」を時間順で提示する．各ホットスポットの提示は，アイコンとセグメント列を表すブロック群から構成される．「おもしろスポット」を表すアイコンと「なるほどスポット」を表すアイコンの違いによってそれぞれ区別し，ブロックの色分けは前述の全体表示パネルと同様である．各アイコン・ブロックはクリック可能であり，アイコンがクリックされた場合はそのスポットの先頭から再生され，ブロックがクリックされた場合はそのブロックの先頭から再生される仕組みとなっている．

コントロール部右側のアイコンを伴う 2 つのボタンはスポットジャンプボタンである．再生中の箇所より後ろで最も近いホットスポットにジャンプすることができる．

4 まとめ

本稿では，高精度な音リアクションイベント検出に基づくポッドキャストブラウザ Podspotter を提案した．ポッドキャスト中の検出した 8 つの音響イベントを色分けして提示することで，全体の流れを把握しやすくし，笑い声とあいづちに基づく 2 種類のスポッティング機能を提供することで，効率的な視聴環境を構築した．

今後，ホットスポットの抽出について，ヒューリスティックな切り出しではなく，統計的な抽出方法を検討する．また話者認識の導入により，性別だけでなく個別の話者の切り替わり点を表示できるようにし，より対話の流れを掴みやすくすることを目指す．さらにクライアントサーバモデルで実装を行うことで，ウェブサーバ上のデータをブラウザ上の Podspotter（クライアント）が取得することにより，広く利用することが可能になると考えられる．

参考文献

[1] C. Alberti, et al. An Audio Indexing System for Election Video Material. In *Proc. ICASSP*, pp. 4873–4876, 2009.

[2] 後藤, 緒方, 江渡. PodCastle の提案: 音声認識研究 2.0 を目指して. 情処研報, SLP-65-7, pp. 35–40, 2007.

[3] 緒方, 後藤, 江渡. PodCastle の実現: Web2.0 に基づく音声認識性能の向上について. 情処研報, SLP-65-8, pp. 41–46, 2007.

[4] 常, 高梨, 河原. ポスター会話におけるあいづちの形態的・韻律的な特徴分析と会話モード間との相関の分析. 人工知能研資, SIG-SLUD-A802, pp. 7–13, 2008.

[5] 須見, 河原, 緒方, 後藤. ポッドキャストを対象とした音リアクションイベント検出. 情処研報, SLP-77-24, 2009.