

講義動画検索におけるユーザーの行動とその着目点の分析*

田中良典 宮澤幸希 菊池英明（早大人科）

1 研究の目的

インターネットを用いた講義配信を前提とした e-learning システムの本格運用が始まっている^[1]。そうした中で我々は音声認識を用いた講義動画検索システムの構築を目指している。講義音声は読み上げ音声に比べて自発性が高いために、現状では講義全体を精度良く認識して文字化することは困難である。本研究ではこの問題に対処するため、講義動画検索場面におけるユーザーの着目点を調査し、ユーザーが着目するシーンにおける音響的・言語的特徴を調べ、音声認識や検索支援に利用することを目指す。

2 ユーザーの検索行動

ユーザーが講義動画検索を行う状況には以下のようなものが考えられる。

- a) 特定の動画の中から、特定のシーンを見つけ出す。
- b) 不特定の動画の中から、不特定の該当するシーンを見つけ出す。

講義動画検索の利用状況として今回想定するのはユーザーが講義を一度受けた上で再びそのシーンを見たいと考える a) の状況とし、シーン探索時のユーザーの行動を調査する。

発話者の意図の指標となる強調発話に着目して検索に活かそうという先行研究^[2]に対して、本研究ではユーザーである聞き手の側に注目する。

聞き手の講義の理解に関しては、^[3]で示されるようにピッチや明瞭度といった音響的特徴や、自由発話に特有のフィラーといった情報が影響する。

よってそれらの特徴も含めて、講義動画からユーザーが着目点をどのように選択するかを調べ、あわせてテキスト・音声・動画といった提示されるメディアによる着目点の違いを調べる。

3 実験概要

ユーザーの着目点について、文字情報、音声情報、さらに講義動画特有のスライドなどからの情報による違いを調査するため、被験者を、動画を視聴する A 群、動画音声のみを聴く B 群、動画音声の書き起こしテキストを読む C 群、との 3 群にわけた。動画を見る際の影響について詳細化するため、A 群は 10 名、B、C 群はそれぞれ 4 名ずつとした。

実験内容として、被験者には一度、講義のコンテンツ（動画・音声・テキストのいずれか）を視聴してもらい、その後コンテンツを見ずに講義内容の要約を行なってもらった。その後、「要約の為に自分が見たいポイント」すなわち着目点を、動画・音声については時間を示して貰いテキストについては下線を引いてもらった。その際に検索システムを仮定した「検索キーワード」も記述してもらった。

実験素材としては、1 つの講義から複数の話題のまとまりを持つようにして 15 分を抜粋して A 群に用いた。B 群についてはその動画から音声を抽出した。C 群については、同じ部分の書き起こしテキストを印刷し、同じく 15 分間閲覧してもらった。

それぞれ視聴中のメモ等は認めず、一度視聴した直後に要約等に取り組んでもらった。

4 実験結果

4.1 3 群の行動比較

実験の結果、着目点として延べ 57 箇所、検索キーワードとして 46 種類が得られた。

元々理解のたやすい内容だったためか、タスクの要約文については 3 群の間で完成度に大きな差異は見られなかった。

着目点について、時間のズレが 3 秒以内のものは、同一点としてまとめた。その結果、A 群で複数人に着目された 8 箇所と、C 群で複数人に着目された 7 箇所との間で、4 箇所まで共通していたが、A 群の残りのうち 3 箇所は C 群では一度も着目点としてあげられな

* Analysis of actions and viewpoints of user in lecture video search, by TANAKA, Yoshinori, MIYAZAWA, Kouki, and KIKUCHI, Hideaki (Faculty of Human Sciences, Waseda University).

かった。このことから、要約という同じ目的を設定しても提示するメディアの違いにより着目点異なることが確かめられた。詳細に見ると、A群の着目点には図の映されたシーンが30箇所中7箇所あり、視覚情報の重要性が見られた。

検索キーワードについてみると、着目点として被験者が選択した箇所において、検索キーワードが発話されていないケースが、A,B,C群で偏りなくみられ、合わせて13箇所あった。このことは、講義動画検索システムの実用を目指す際に、検索キーワード以外の検索手段をシステムが用意する必要を示していると言える。

4.2 講義動画の着目点

4.2.1:言語的特徴

着目点について、指定された時間の冒頭から約3秒(ないし3文節)の範囲における言語的特徴を分析した。その結果、(1)指示語の存在、(2)談話標識の存在、(3)検索キーワードの存在、という3要素が見うけられた。また、これらを同時に満たす場合はあるが、どの特徴も持たないものはなかった。以下に着目点の中でのそれぞれの数と割合を示す。

- (1)指示語 : 4 (14%)
- (2)談話標識 : 9 (41%)
- (3)検索キーワード : 14 (64%)

4.2.2:音響的特徴

着目点内外の検索キーワードの音響的特徴を調べ、比較した。具体的には、該当箇所の音声からピッチ、パワーの測定を行ったところ、表1のようになった。

着目点内のキーワードに関しては、他の箇所でのキーワードよりもパワーの平均が有意に大きいこと(t検定で $p < 0.01$) が確かめられた。ピッチに関しては、有意な差が見られなかった。一般的には強調箇所でピッチが大きくなることがしばしば観察されるが、話者の特性を考慮する必要が示唆される。

Table 1: 着目点内外のキーワードの音響的分析

	パワー		ピッチ[Hz]	
	着目点	その他	着目点	その他
平均値	467.8	389.2	286.7	280.3
標準偏差	36.5	31.3	14.4	20.0

5 まとめ

講義動画を対象に、要約というタスクを与え、ユーザーがどのような点に着目して検索するのかを調べた。結果、文字情報を対象にしたものと、動画を対象としたものでは、その着目点異なることが確認できた。

ユーザーの着目点は、検索キーワードになり得る語が発話されるシーンの他に、指示語や談話指標などの言語的特徴と講義音声の音響的特徴の組み合わせによって選択される傾向が見られた。

今後は、講義音声の話者の特性や講義の難易度を考慮に入れて分析を続け、得られた結果を講義動画検索システムの実用性向上に活用する予定である。

謝辞

講義音声データをご提供いただいた早稲田大学人間科学学術院の教員の皆さんに感謝致します。

参考文献

- [1] 西村昭治, 浅田匡, 向後千春, 菊池英明, 金群, 松居辰則, 野嶋栄一郎, "「キャンパスモデル」e-Learningの実践: 早稲田大学人間科学部eスクールの取り組み," 日本教育工学会第20回全国大会講演論文集, pp.149-152, (2004).
- [2] 上野智子, 相川清明, "音響特徴を用いた音声ドキュメント中の強調発話の検出", 1回音声ドキュメント処理ワークショップ講演論文集, p.47-52, (2007).
- [3] 宗宮充宏, 小林健司, 西崎博光, "講義・講演音声自動評価のための音響・言語的分析", 第1回音声ドキュメント処理ワークショップ講演論文集, pp 9-16, (2007).