



講義ビデオの活用に向けた 講義音声の発話特徴分析

CLE研究会(2011/5/13 長岡技術科学大学)
○永井孝幸(熊本大学),
豊田寿行,名古屋孝幸,西澤弘毅,今井正和
(鳥取環境大学)

講義ビデオの活用

- 学科開講科目の収録(2003年～)
 - － 毎週20科目以上(鳥取環境大学)
 - － 収録自動化(2010年後期～)
- 講義ビデオの活用
 - － 復習(学生): LMS経由で配信
 - － 講義改善(教員): 話し方、時間配分の見直し



ビデオを全て見直すのは時間的に困難



講義の実施状況を短時間で把握できないか？



関連研究・事例

- 既存研究
 - 連続音声認識(講義音声コーパスなど)
 - 話速・声量分析(話し方改善システム)
- 商用システム
 - クラウド型音声認識サービス
英語: Echo360, Yap 日本語: Google音声認識API
- 本研究
 - 学科全体の講義を対象に発話特徴を分析
 - 特徴的な場面・講義を検出



今回の分析内容

- 発話区間の分析
 - 出現位置、発話長の集計
 - 講義全体の傾向抽出
 - 全体の傾向から外れた場面・講義を抽出
- 音声認識ソフト(Julius)の適用
 - 自然発話の認識率は低いと言われるが・・・？

- ビデオをどう分析すればよいか？
- ビデオを集積すると何が分かるか？



分析対象の講義音声

- 収録期間: 2010年度後期(2010/9~2011/1)
 - 21科目, 18講師, 258コマ(約400時間)
 - 無編集状態で分析

収録地点	収録方法	対象	収録回数	収録時間(分)
鳥取	自動	対面講義(17科目)	193	18,240
熊本	自動	対面講義(2科目)	59	5,285
	手動	対面講義(2科目)	6	345
合計			258	23,870



- 音声収録方法

- 講師マイク/ピンマイクを音声ミキサ経由で収録
- Bluetoothワイヤレスマイク(手動収録時)



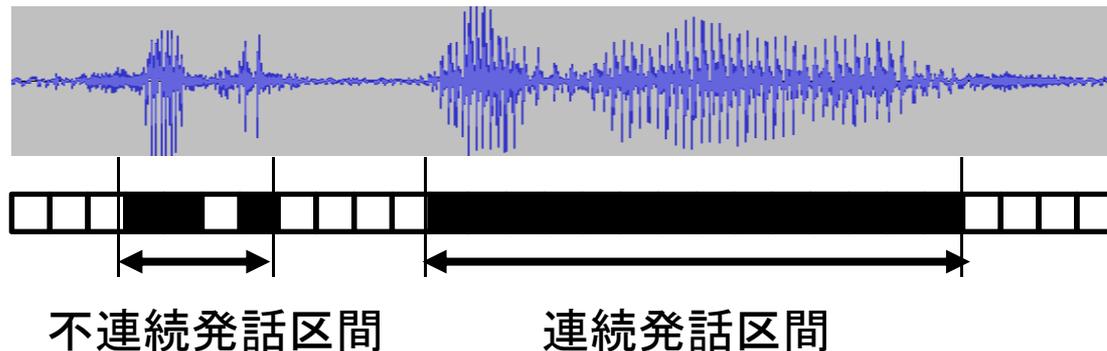
講義音声の実際

- 講師が発話している時はよい
- 無発話時の雑音
 - 定常ノイズ
 - 板書時のペンの音、紙をめくる音、衣擦れ
 - 教室前方の学生が出す声・音
- マイク音声が取れないケース
 - 音響機器・マイクの電源入れ忘れ
 - オフレコ発言時に意図的に電源OFF
 - そもそもマイクを使わない講師



発話状況分析の前処理

- 0.1秒単位で有音/無音区間を判定
 - 平均音量以上なら有音区間と判定
- 無音区間の長さの分布
 - 0.1秒以下:31%,0.3秒以下:56%,0.6秒以下:71%



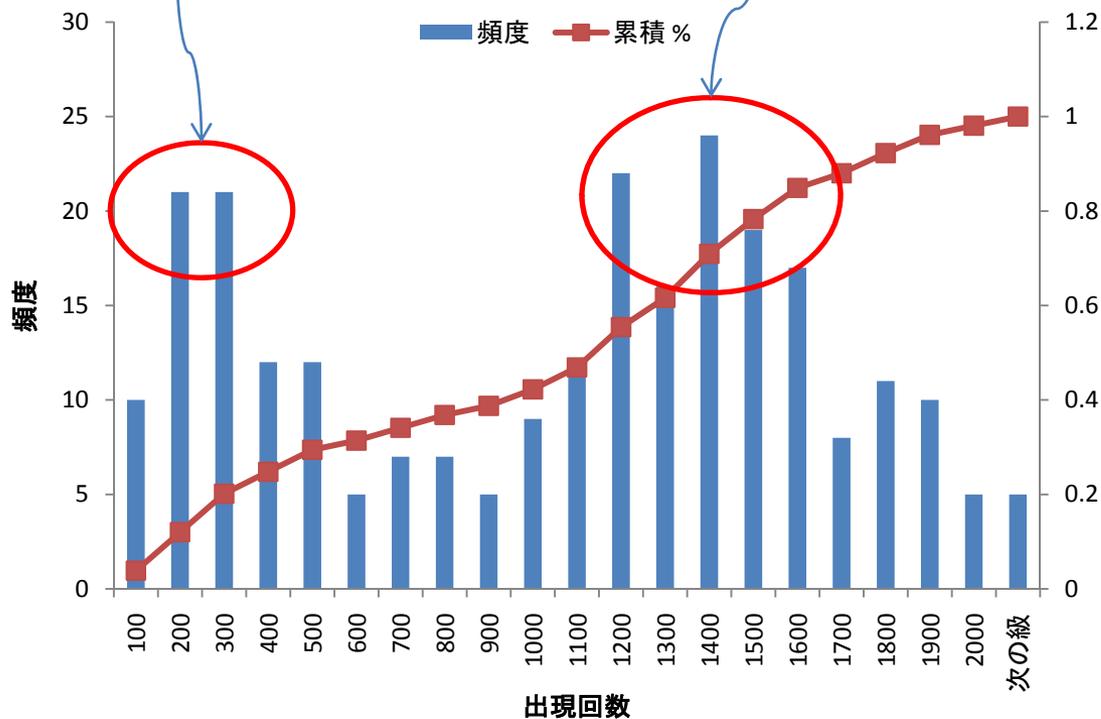
- 0.6秒以上の無音区間で発話区間に分割

不連続発話区間の出現回数

- 2コブに分布(演習主体科目と講義主体科目)

時間内演習主体科目
(数学、プログラミング)

講義主体科目



講義主体でも
10分は無音



講義発話の開始/終了位置

- 調べたいこと
 - 講義を順調に開始/終了できているか？
 - 講義が本格的に始まるタイミングはいつか？
- 最初の発話区間 \neq 講義の始まり
 - 「あ、あ。」など講義の開始ではない発話もある
- 講義の開始/終了時刻
 - 仮定: ある程度まとまった長さの発話を伴う
 - 「1秒以上の不連続/連続発話区間」を検出・集計



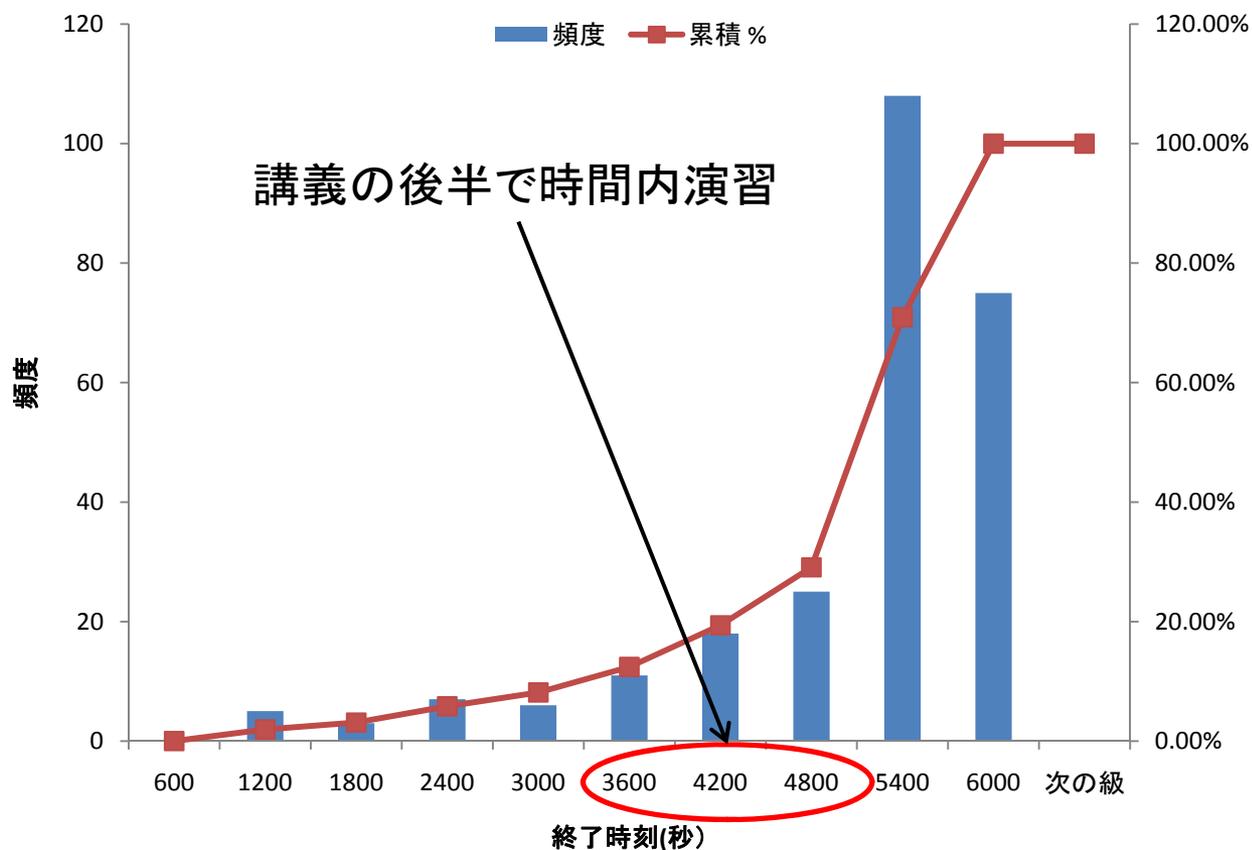
講義開始時点の発話区間

- 発話開始時刻の平均
 - 不連続発話: 142秒 (講義開始から2分)
 - 連続発話: 184秒 (講義開始から3分)
- 不連続発話と連続発話の時刻差
 - 10秒程度: 「さてはじめましょうか。……えと、」
 - 数分以上:
 - マイクをONにしてから資料配付・レポート返却
 - マイクをONにしてからPC・プロジェクター起動
 - 講義を始めたがプロジェクター映らず



講義終了時点の不連続発話区間

- 7割以上の講義で90分目以降に最後の発話



発話の開始位置が特異な講義の例

- 発話開始位置が10分目以降(258件中14件)

講義番号	先頭発話区間検出時刻(秒)		講義状況
	不連続発話	連続発話	
講義1	687.1	963.2	マイクの付け忘れ→途中でON→無言で演習問題の解説を板書
講義2	837.7	851.7	演習見回り→「あ、あ」→身支度→「えっと、それでは・・・」
講義3	844.5	846.7	演習見回り→世間話→「はい、では・・・」
講義4	854.9	854.9	演習見回り→あいさつ→講義開始
講義5	996.7	1024.9	マイク完全に入れ忘れ
講義6	1033.5	1113.2	マイク完全に入れ忘れ
講義7	1192.7	1192.7	マイク入れ忘れ。途中からON。
講義8	1229.9	1229.9	冒頭小テスト→マイクON→講義開始
講義9	1964.6	1975.6	マイク入れ忘れ。途中からON。
講義10	2123.4	2123.4	マイク入れ忘れ。途中からON。
講義11	2204.7	2207	マイク入れ忘れ。途中からON。
講義12	2326.6	2329.5	マイク入れ忘れ。途中からON。
講義13	3955.2	3957.7	マイク入れ忘れ。途中からON。
講義14	4394.2	4401.1	マイク入れ忘れ。途中からON。

BRD方式の講義

マイク入れ忘れ



発話の終了位置が特異な講義の例

- 発話終了位置が25分目以前(258件中6件)

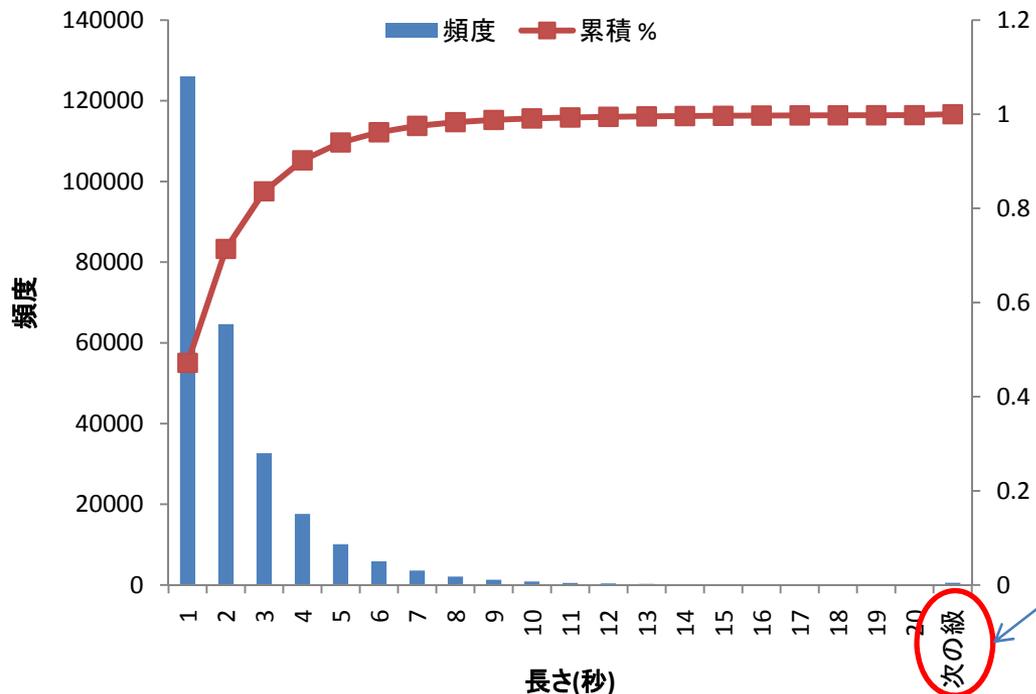
講義番号	最終発話区間検出時刻(秒)		講義状況
	不連続発話	連続発話	
講義15	711.2	711.2	演習内容をまとめて解説した後、 マイクOFFで演習
講義16	738	900.1	演習内容をまとめて解説した後、 マイクOFFで演習
講義17	976.2	976.2	途中で マイク電池切れ のために音声途切れ途切れ
講義18	976.2	1002.7	演習内容をまとめて解説した後、 マイクOFFで演習
講義19	1189.4	1189.4	講義中に マイク電池切れ
講義20	1477.6	1477.6	演習内容をまとめて解説した後、 マイクOFFで演習

- マイクOFFで演習(4件)
- マイク電池切れ(2件)



不連続発話区間の出現状況

- 長い発話＝「話が途切れない」(好ましくない)
- 不連続発話区間の長さ分布



全267,137区間のうち

- 71%の区間が2秒以下
- 99%の区間が10秒以下

20秒より長いのは539件

次の綴

3分以上の不連続発話区間

- 全267,137区間のうち69件が該当

– 長時間マイクをOFFにした時のノイズor **長台詞**

講義	不連続発話区間	開始位置 (分)	終了位置 (分)	区間長 (秒)	講義状況
講義1	区間1-1	0	6.1	363	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-2	6.1	10.3	248.7	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-3	10.3	14.3	244.4	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-4	16.2	21.3	303.5	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-5	25.1	34.1	535.3	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-6	34.1	46.9	768.8	板書・スライドで講義。話が区切れそうで区切れない。
	区間1-7	47.5	53.7	373.8	板書・スライドで講義。話が区切れそうで区切れない。
講義2	区間2-1	33.1	36.9	231.3	板書とスライドで計算過程を解説
	区間2-2	37	41.9	298.5	板書とスライドで計算過程を解説
	区間2-3	41.9	45.2	192.9	板書とスライドで計算過程を解説
	区間2-4	45.2	55	588	板書とスライドで計算過程を解説
	区間2-5	60.9	65.5	270.7	板書とスライドで計算過程を解説
	区間2-6	66.8	71.7	290.7	板書とスライドで計算過程を解説。67:30秒から無音で演習開始。黒板掃除の雑音が入る。
	区間2-7	71.7	78.7	424.3	73:40までは無音で演習見回り中。その後、演習問題の解説。
	区間2-8	81.3	94.9	810.9	途中からマイクON。終盤10分マイクOFF。
講義3	区間3-1	12.6	17	266.4	板書・スライドで講義。話が区切れそうで区切れない。
	区間3-2	20.8	25.1	258.6	板書・スライドで講義。話が区切れそうで区切れない。
	区間3-3	60.3	76	938.9	冒頭解説の後はマイクOFFのまま演習
	区間3-4	76.3	80.7	265.3	演習中。マイクOFFのまま



音声認識ソフトの利用

- 調べたいこと
 - 対面講義の自由発話でどのくらい有効か？
 - 字幕書き起こし・キーワード抽出に使えるか？
- Juliusディクテーションキットを利用
 - 音響・言語モデルは書き言葉用(キット付属)
- 分析手順
 - 講義音声の不連続発話区間で分割
 - 各音声ファイルをJuliusに入力



音声認識結果の例

- 一部の講師についてはそれなりの結果
 – 音響が悪い？ 方言も影響？

発話	原文	認識結果
発話1	ちょっと待ってね。電気消しとこう。	となつてね！ 元気です。
発話2	はい、前がこんだけ空いてんねん。	マイナスとなつて・。
発話3	えと、今日の資料はですね、えー、	四仕様で、寝て。
発話4	えーと、この科目は、えー、三年生、を中心に	で探せ、五、中心に。
発話5	情報の価値	じゃこの感じ。
発話6	(スー)皆さんは	Nさんは？
発話7	できた人から前に提出して終わりにしましょう。	形ですから、満員でして怖い。
発話8	通信プロトコルを理解するっていうのが目的になる。	大変 感覚を 理解 する っ て の が 目的 になる。
発話9	えー、後期が始まって今日が最初の授業か と思います。(スー)えーとー…、ま、三年、	データを基が、始まって、今日が最 初の、事業か と思います。デート、 馬鹿な
発話10	今日の議題に入りますけど、	今日 の 時代 に 入 り ます けど 。
発話11	えー、分担します。	で、分担します。

←←← ???

←←← 惜しい？

←←← 惜しい



認識結果が全て正しかった単語

- 認識結果と元の音声を聞き比べ
 - － 自動認識にひっかからないケースの率は未調査

講師	検出語(カッコ内は検出回数)
講師2	情報セキュリティ(3), 考えて(3), コンピュータ(3), この(3), お金(3), そして(3), という(3), 対策(6), それ(7)
講師3	今(4), データベース(11)
講師4	はじめ(3), モデル(3), 最初(3), 1980(3), 正確に(3), 大量のデータ(3), いうことで(3), こんな(4), なに(4), 今(5), いうの(5), パソコン(6), コンピュータ(8)

- 講師12名中、3名は認識結果良好
 - － 3名とも話し方が明瞭。キーワード検出も可能。



頻出発話冒頭語の分析

- ビデオ編集や発話回数の分析に使えるかも
- 頻出する語の正解率が高いとは限らない
 - 言語モデルに無理に当てはめている模様

講師	検出語	検出回数	適合結果			適合率(%)		一部正解の例
			正解	一部正解	不正解	正解	正解+一部正解	
講師1	no	27	0	0	27	0	0	
	kono	24	2	5	17	9	30	konpaira
	de	18	4	0	14	23	23	
講師2	de	27	3	19	5	12	82	e
	kono	15	12	0	3	80	80	
	jo:ho:	13	10	0	3	77	77	
講師4	de	46	7	30	9	16	81	e:(28件)
	kono	28	14	10	4	50	86	konna(5件)
	yuu	26	25	0	1	97	97	

de(で)
de:(で一)
e:(え一)
は頻出



現場の講師からのコメント

- 「自動的にキーワードを抽出し映像にラベル付けができれば、復習の際の効率が上がるのではないか」
- 「ミニ演習に割く時間など、授業時間の使い方を反省する際材料になるかもしれない」
- 「音声認識の結果を参考にすることで、聞きやすい講義にする材料になるのではないか」



まとめ

- 発話区間の分析
 - 特異な講義の検出(時間内演習,長講釈)
 - 音響不具合の検出
 - 音響ノイズ誤認識への対策が必要
- 自動音声認識
 - 発話明瞭な講師については適用可能
 - 発話明瞭度の指標に使えるかもしれない
 - 「de/de:/e:」の検出は比較的精度がよい

音声情報だけでも講義実施状況の把握に役立つ

