

音声検索

第10回音声言語シンポジウム発表資料

秋葉友良

(豊橋技術科学大学)

背景

- 音声ドキュメント処理
 - 音声データを文書として扱うための技術の総称
- 鍵となる技術
 - 検索
 - 要約
 - マイニング
 - 構造化
 - コンテンツ生成
 - ...

音声言語シンポジウムにおける 音声ドキュメント処理(1)

- 第1回音声言語シンポジウム(1999)

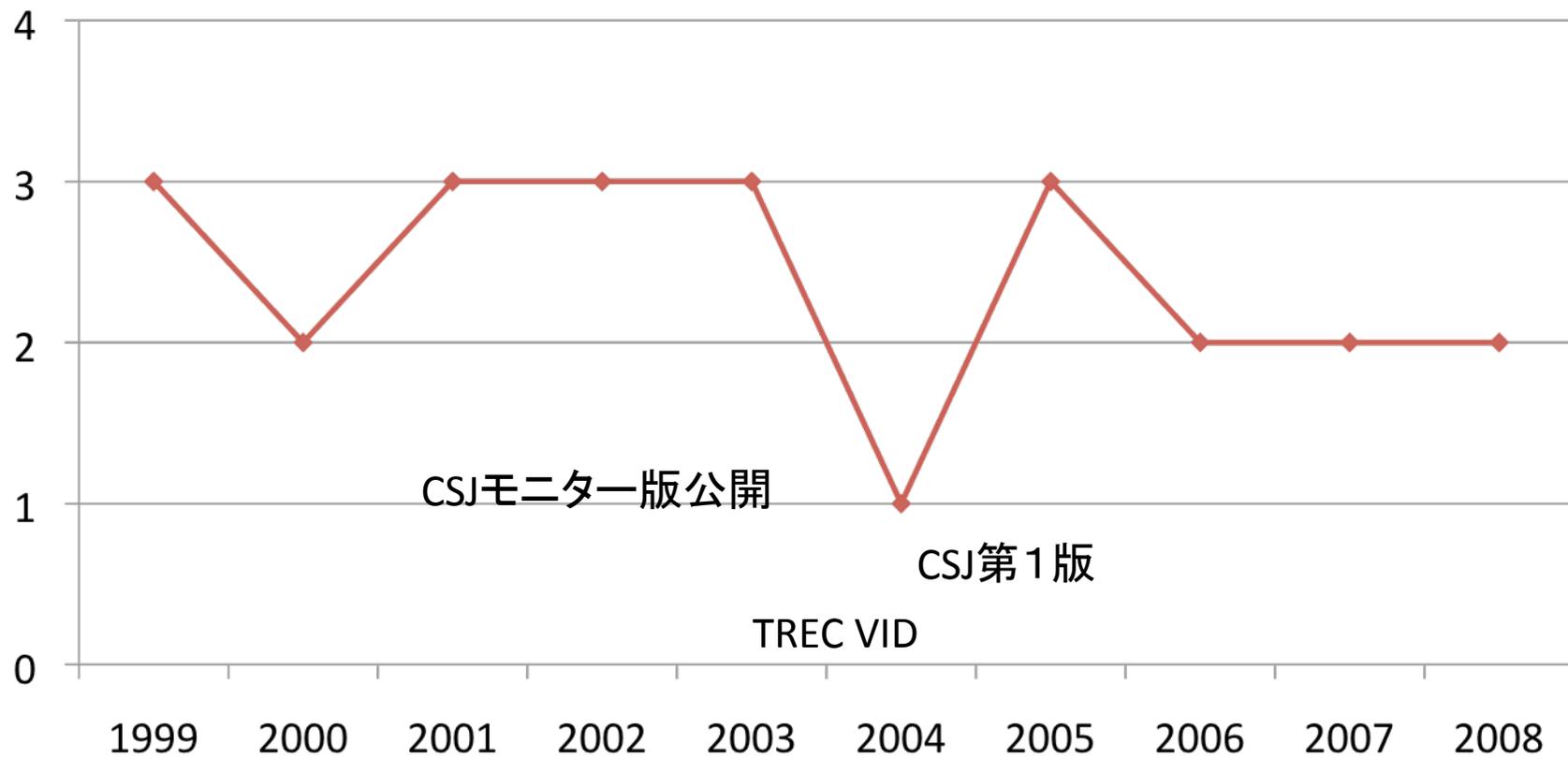
- 音声ドキュメント処理に関する発表は3件

- 「キーワードの音声入力によるニュース音声の検索法」, 西崎博光, 中川聖一(豊橋技科大)
 - 「ニュース音声に対する検索方法の比較」, 鷹尾 誠一, 緒方 淳, 有木康雄 (神戸大)
 - 「話題語と言語モデルを用いた音声自動要約法の検討」, 堀 智織, 古井 貞熙 (東工大)

- ニュース音声を対象

音声言語シンポジウムにおける 音声ドキュメント処理(2)

発表件数



処理対象の音声ドキュメントは多様化

音声検索

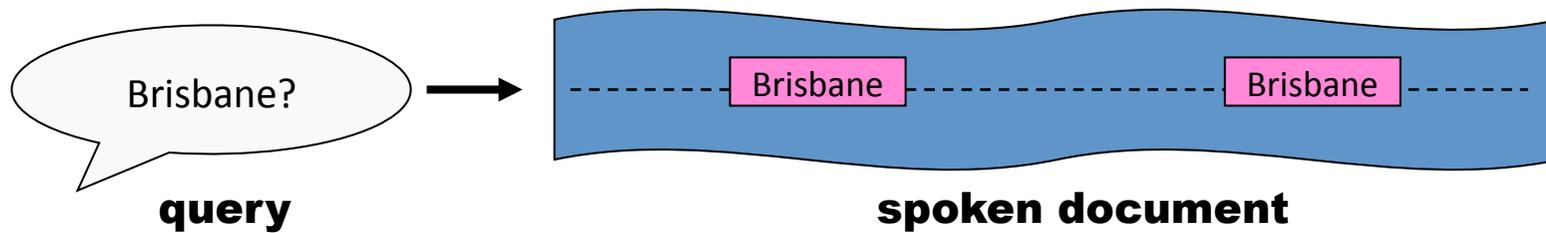
- 音声入力検索(Speech-driven Information Retrieval)
 - 情報検索システム(必ずしも音声ドキュメントではない)の入力に音声を用いる
- 音声ドキュメント検索(Spoken Document Retrieval)
 - 音声ドキュメントを対象とした情報検索

TREC SDR Track (1997-2000)

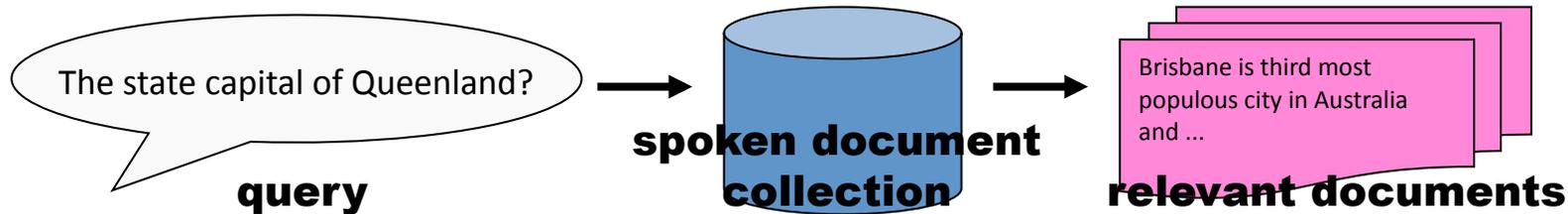
- 米国NIST主催の評価型ワークショップTREC (Text Retrieval Evaluation Conference) における音声ドキュメント検索タスクの評価
 - ニュース音声を対象
 - タスク
 - 1年目: Known Item Retrieval (キーワード検索)
 - 2年目以降: Ad hoc Retrieval (内容検索)
 - 自動書き起こし(WER 74%程度)を用いた索引付けで、人手書き起こしに匹敵する検索精度(約45%)を達成
 - 解決済(Solved)と判定

音声ドキュメント検索

- キーワード検索(Known Item Retrieval)



- 内容検索(Ad hoc Retrieval)



TREC SDR Track (1997-2000)

- 米国NIST主催の評価型ワークショップTREC (Text Retrieval Evaluation Conference) における音声ドキュメント検索タスクの評価
 - ニュース音声を対象
 - タスク
 - 1年目: Known Item Retrieval (キーワード検索)
 - 2年目以降: Ad hoc Retrieval (内容検索)
 - 自動書き起こし(認識精度74%程度)を用いた索引付けで、人手書き起こしに匹敵する検索精度(約45%)を達成
 - 解決済(Solved)と判定

TREC SDR以降の音声ドキュメント検索 研究

- 検索対象の多様化
 - ニュース音声, テレビ番組, 講義, 講演, 会議, UGM (User Generated Media)
 - 低い認識精度(30-50%)
- 研究の中心は**キーワード検索**(Known Item Retrieval)
 - 文書検索の場合でも、「検索クエリが含まれる文書」を探すタスクを対象としている。

音声ドキュメントのキーワード検索

- クエリとして与えられた語を、音声ドキュメントから「素早く」「正確に」見つけることを目的とする
 - 2段階の検索手法
 - 自動書き起こしを用いた索引付け、候補抽出
 - 候補区間に対する厳密な照合
 - NIST Spoken Term Detection Evaluation (2006, 2009?)
- 研究対象と手法
 - 挿入、脱落、置換誤りへの対処
 - 複数候補の利用
 - N-best候補やラティス表現を利用した索引付け
 - より効率的な索引付け手法へ
 - PSPL(Position Specific Posterior Lattice): 位置情報を利用した圧縮
 - Confusion Network: 第1候補による圧縮
 - 認識語彙外語への対応
 - サブワード(音素,音節)認識・索引の併用

音声ドキュメントの内容検索

- 研究は少ない
 - 評価のコストが高い(人間による正解判定が必要)
- 既存のテストコレクション
 - TREC SDR (1997-2000)
 - ニュース音声
 - CLEF Cross-Language SDR (2003-2004)
 - ニュース音声(TREC SDRと同じ)
 - TRECVID (2003-2008)
 - ビデオ(ニュース、ドキュメンタリー、教育番組、など)
 - **プーリングによる正解判定**
 - 参加者が提出した検索結果に対してのみ正解判定を行う
 - **英語、ニュース音声**が中心

日本語話し言葉コーパス(CSJ)を対象 としたテストコレクション

- 情報処理学会SLP音声ドキュメント処理WGで構築中
- 構成要素
 - 日本語の講演音声(CSJの学会講演および模擬講演)を対象文書
 - 講演の一部(約5発話程度)の可変長区間が適合となるような39の検索質問と適合性判定結果
 - 検索対象文書のリファレンス音声認識結果
- 参考文献
 - Akiba et al., “Test Collections for Spoken Document Retrieval from Lecture Audio Data”, In Proc. of LREC, 2008.
 - 秋葉, 他: “SLP音声ドキュメント処理ワーキンググループ活動報告”, 第10回音声言語シンポジウム, (2008) (明日発表予定)

CSJテストコレクションと TREC SDRの比較

	CSJ SDR	TREC9 SDR
言語	日本語	英語
対象文書	講演音声	ニュース音声
時間	623.6時間	557時間
文書数	2,702(30,762*)	21,754
単語/文書	2,324.9(204.2*)	169
検索クエリ	39	50
書き起こし	高品質	低品質(WER 10.3%)
音声認識WER	21.4%	26.7%

* 30発話単位を文書と見なした場合

音声検索: まとめ

- TREC SDR Track以降は、キーワード検索の研究が中心。内容検索の研究は少ない。
 - テストコレクション構築コストが高いため。
- しかし、ニュース音声以外の音声ドキュメントを対象としたテストコレクションも構築されつつある。

音声検索：今後の展望

- 音声ドキュメントの増大・多様化が加速
 - TVコンテンツのオンライン配信
 - UGM(User Generated Media)の増大
- 内容検索の必要性が増大
 - 「既に知っていることを問う」(ナビゲーション的な質問)から「知らないことを問う」(インフォメーション的な質問)へ
- 評価用テストコレクションの構築
 - プーリング: 大規模な検索対象への正解判定
 - 研究機関間での協力が必要
- 音声入力情報検索の可能性
 - 情報要求の表現手段としての音声の利用