論文 / 著書情報
Article / Book Information

論題(和文)	音声認識研究の今後の展開
Title	
著者(和文)	古井貞熙
Author	SADAOKI FURUI
出典(和文)	音声言語による人間・機械対話システムの高度化と新展開, , , pp. 93
Journal/Book name	, , , pp. 93
発行日 / Issue date	2002, 1

日本学術振興会未来開拓学術研究推進事業研究プロジェクト「音声言語による人間・機械対話システムの研究」公開シンポジウム“音声言語による人間・機械対話システムの高度化と新展開”、p.93 (Jan. 2002)

音声認識研究の今後の展開

古井 貞熙
東京工業大学大学院情報理工学研究科計算工学専攻
furui@cs.titech.ac.jp

音声認識技術の最近20年間の進歩は、図1に示すようにまとめることができる。この図からわかるように、孤立単語、読み上げ音声などの認識は、極めて大語いでも実用レベルに達しているが、自然な話し言葉(spontaneous speech)に関しては、まだ限られた性能しか得られていない。

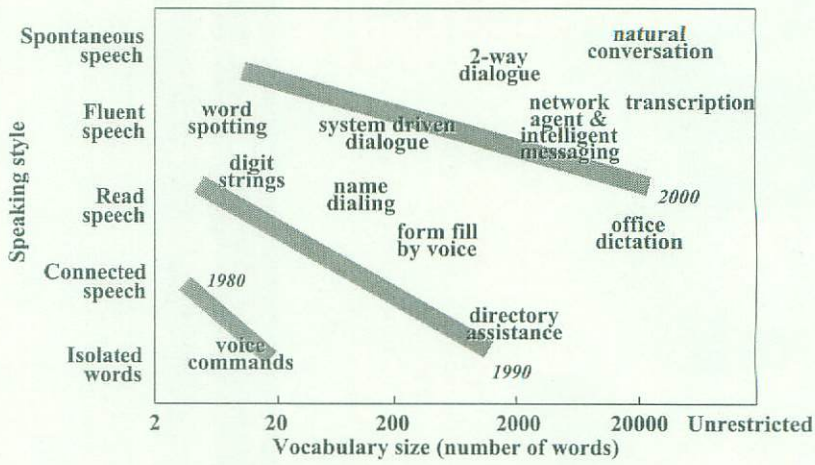

Fig.1: Progress of speech recognition technology along the dimensions of vocabulary size and speaking styles.

音声認識のタスクは、表1に示すように、対象音声に関する2つの規準、すなわち人がコンピュータに対して発声している音声か人に対して発声している音声か、対話(dialogue)か独話(monologue)かによって、4つのカテゴリーに分類することができる。カテゴリーI は、人と人との対話を対象とするもので、会議議事録の自動作成などを目的とした研究が始まっている。カテゴリーII は、人が人に聞いてもらうことを前提に一方向的に話している独話を対象とするもので、ニュース音声の自動字幕化、講演録や講義録の自動作成、音声メールの文字化などの研究が行われている。これらは、いたるところに存在する音声ドキュメントのアーカイブ化として、今後重要度が増してくると考えられる。カテゴリーIII は、人とコンピュータシステムとの対話を対象とするもので、情報検索、予約などを行う実用システムが、米国を中心にすでに多数利用されている。あらかじめ明確に定義された応用タスクを前提としてシステムを設計し、正しく動作すればよしとするのが普通で、この点で他のカテゴリーとアプローチが異なる。カテゴリーIV は、人がコンピュータに独話で話している音声を対象とするもので、ディクテーションがこれにあたる。

Table 1. Categorization of speech recognition tasks.

	Dialogue	Monologue
Human to human	(Category I) Switchboard, Call Home (Hub 5), meeting task	(Category II) Broadcasts news (Hub 4), lecture, presentation, voice mail
Human to machine	(Category III) ATIS, Communicator, information retrieval, reservation	(Category IV) Dictation

人がコンピュータを意識して発声しているカテゴリーIII やIV の音声や、原稿を読み上げた音声は、人と話しているカテゴリーI や II の音声と、音響的にも言語的にも大きく異なることがわかっている。図2に、10 人の学会講演音声に対して、種々の学会講演音声の書き起こしから作った言語モデル(SpnL)と、一般の Web 上にある書き言葉に整形された講演録から作成した言語モデル(WebL)による test-set perplexity と未知語率(OOV)の比較を示す。これからわかるように、話し言葉のモデル化のためには、そのコーパスを広範囲に作成し、そこから種々の現象を学習することが必須である。このような研究を、諸外国と連携して進めることが重要である。

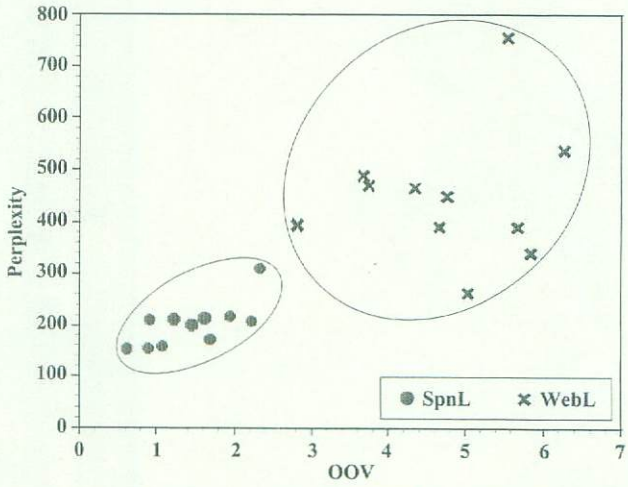

Fig.2: Test-set perplexity and OOV rate for the two language models.