

7M-3 文法情報を用いた音声認識装置の言語処理方式

ローケン・キム キュンホ 木村 晋太 奈良 泰弘
 (株) 富士通研究所

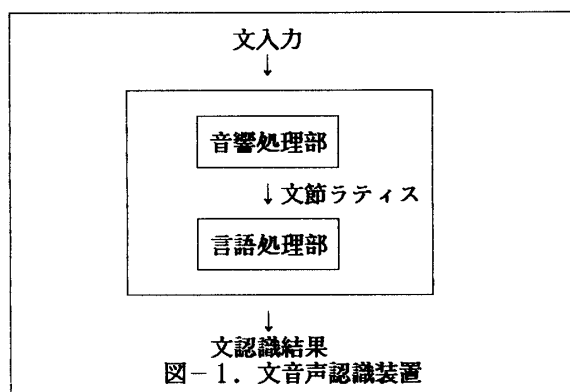
1. はじめに

筆者らは、音声による文章入力をめざし、音響処理の後処理としての文レベルの言語処理方式を開発している〔1〕。本稿では、評価実験結果について報告する。

評価実験では、まず、音声データに雑音を加えることにより数種類の認識率の文節ラティスを大語彙音声認識方式〔2〕を用いて作成した。次に、各認識率の文節ラティスを言語処理部に入力し、文認識率を測定した。

2. 言語処理部の特長

音声認識装置に音声で文を入力する際、音響処理だけでは認識率の限界があるため、文法や意味情報などの知識を利用し、文法的、意味的に正しい文を選ぶことによって認識率を上げる必要がある。筆者らは、第一段階として、文法情報だけを含んだ言語処理部を音声認識装置の一部として開発した(図-1)。言語処理部では、まず、音響処理部から認識結果として生成された文節ラティス(図-2)を、認識候補間の認識距離の差の大きさによっていくつかのグループに分ける。次に、グループごとに探索優先指数を計算し、グループ間では最良優先探索方式、グループ内では縦型探索方式で候補文を生成する(図-3)。選ばれた候補文の文法的妥当性はBUP

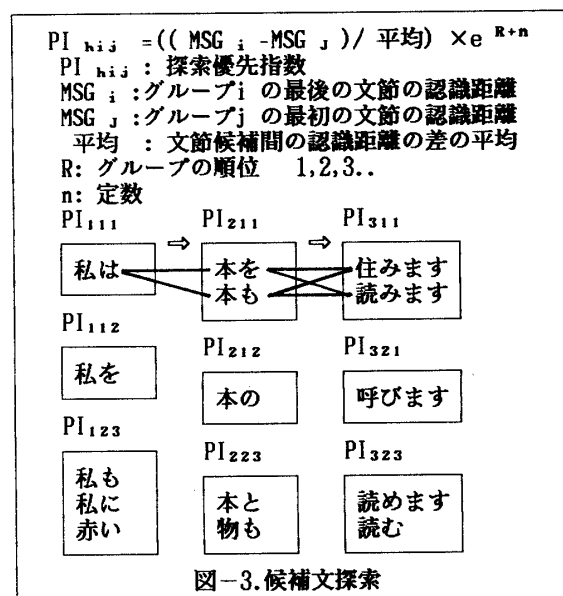


(bottom-up parser)とLFG(lexical functional grammar)方式による表層格検査で確認される(詳細は参考文献3を参照)。

入力文	私は	本を	読みます
第一候補	私は(326)	本を(318)	住みます(282)
第二候補	私を(366)	本も(320)	読みます(303)
第三候補	私も(405)	本の(351)	呼びます(317)
第四候補	私に(406)	本と(398)	読めます(362)
第五候補	赤い(414)	物も(404)	読む(370)

数字は認識距離

図-2. 文節ラティスの例



3. 言語処理部の文回復能力評価

中学校の英語の教科書を参考し、認識対象の4000文節を選び、これを用いて入力用の100単文(日本語の文)を作成した(平均文長3.5文節)。次に、音声登録用の200文節と入力文を騒音レベルが約30dB(A)のスタジオで録音した後、入力用文データのみ別に録音した計算機室内の数種のレベルの環境雑音(主にファンの騒音)を加えることにより音声区間の平均S/N比(signal to noise ratio)が30dB, 20dB, 10dBとなる入力データを人工的に作成して、音響処理部に入力し、認識結果を得た(表-1)。

表-1. 音響処理部の認識結果 (%)

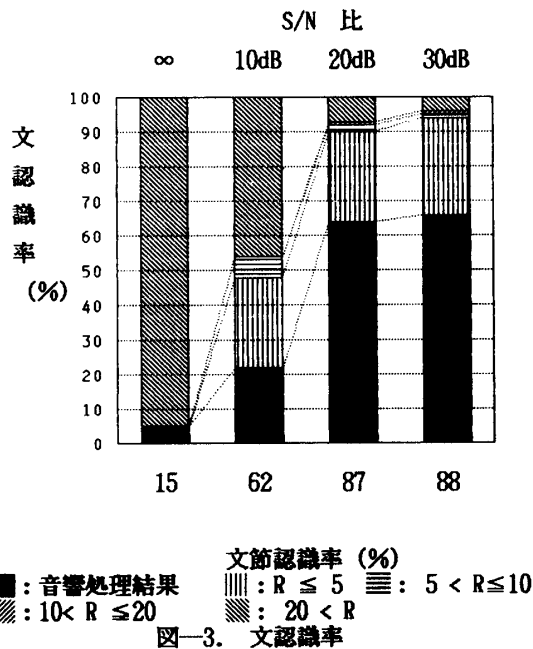
S/N 比	∞	30dB	20dB	10dB
文認識率	66	64	22	5
文節認識率	88	87	62	15

表-2 は、音響処理部の認識結果のうち誤認識したものを言語処理部で処理した結果である。この表で順位(R)は言語処理部が出力する文の順位、認識不可能は入力文節が10位までの認識候補に含まれていない文の数である。

表-2. 文認識順位 (文)

順位 (R)	∞	30dB	20dB	10dB
R ≤ 5	28	26	26	0
5 < R ≤ 10	1	2	6	0
10 < R ≤ 20	1	1	0	0
20 < R	4	7	16	1
認識不可能	0	0	30	94

(文節認識候補10位までを処理した例)



4. 考察

音声認識装置による実用的な文入力を実現するためには、入力文を少なくとも候補文順位の5位以内で認識することが求められる。本試験では、文法情報だけでも、S/N比が高い場合は(例, ∞と30dB) 候補文順位が5位以内で90%以上の文が認識できたが、S/N比が20dBの場合は5位以内の文認識率が50%を下回る。これは、各入力文節に対する認識候補を10位まで言語処理した結果であるので、文節候補の処理範囲を広げると(例えば20位まで)、文認識率がある程度上がる可能性がある。この場合は、意味情報を導入し、文法的には正しいが意味が通らない候補文を制限することによって、入力文がなるべく上位の候補文の中に入るようにする必要がある。S/N比が10dBの場合は今後検討する必要がある。

5. おわりに

本稿では、音声認識装置の音響処理部に人工的に雑音を加えた音声データを入力し、その認識結果を、文法知識を含んだ言語処理部に入力して、文認識能力を確認した結果を報告した。

6. 参考文献

- (1) Loken-Kim, K.H., A Postprocessor for a Large Vocabulary Japanese Speech Recognition System, EURO_SPEECH'89, Vol.2, pp.1-4
- (2) Kimura, S., 100,000 Word Recognition using Acoustic Segment Networks, ICASSP'90, CH2847-2/90/0000-0061
- (3) Loken-Kim, K.H., et al., A Large Vocabulary Japanese Language Speech Recognition System, Speech Technology, APR./MAY. 1990