

連続単語認識における単語間調音結合補正法

7N-2

°大槻朋子

大河内正明

日本アイ・ビー・エム株式会社

サイエンス・インスティテュート

1. はじめに

連続単語音声認識においては、認識率向上のための単語間調音結合への対処が大きな課題である。この解決法として従来から、標準パツの重ね合わせ引き離し¹⁾、半単語対方式²⁾などが提案されている。前者は、孤立発声された標準パツのみを用い、その重ね合せ及び引き離しを行うことにより調音結合に対処しようというものであるが、このような方法では十分に調音結合に対処できないのみならず、コンテキストを考慮していないため、登録語彙によっては、かえって誤認識を起こすことがある。これに対し、後者は調音結合を受けた部分を含む標準パツを用いるというものであり、調音結合の部分の情報をより積極的に利用しようとするものであるが、標準パツの数は単語数の2乗に比例して増加する。

筆者らは、連続単語認識において、標準パツの増加を最小限におさえ、調音結合に効果的に対処するため、

- 方針1: 音韻に応じた標準パツの切り出し
- 方針2: VCV標準パツによる単語接合部補正
- 方針3: 複数尺度によるマッチング結果の評価

の3つの方針に基づく「複合パツ認識法」を提案し、これを文節認識に適用して評価実験を行ったので、報告する。

2-1. 音韻に応じた標準パツの切り出し(方針1)

孤立発声された単語を標準パツとして、これらの連結により連続発声された入力パツとマッチングを行う場合、標準パツの語頭音韻が破裂音である場合、連結された標準パツでは語頭破裂音の前の無音部は考慮されない。また、語頭音韻が母音の場合、連結された標準パツには入力パツには現れない母音前半部が付加されていると考えられる。これらに対処するため語頭音韻に応じて標準パツを切り出した。すなわち、破裂音に対しては無音部の付加を行ない、母音に対しては前半部を削除した。これは、文献(1)の考え方で基本的には同じであるが、音韻の情報を積極的に用いることにより不必要な標準パツの重ね合せ、引き離しを無くしようとするものである。

2-2. VCV標準パツを用いた単語接合部補正(方針2)

連続発声において、調音結合によるパツ変形を受けた単語接合部は孤立発声により作成した標準パツでは十分に

対処することができない。これを解決するため、単語接合部に調音結合の影響を受けたVCVパツを標準パツとして用いた。文献(2)においても同様の考え方に基づき、調音結合の影響を受けた部分を含むパツ(半単語対パツ)を標準パツとして用いる方式を提案しているが、この場合、必要とされる標準パツの数は登録単語数を N とすると、 N^2 必要となる。これに対し本方式では最大でも登録単語数+全VCVパツ数($N+600$ 程度)でよいことになる。また、文献(2)で用いられている半単語対標準パツの場合には調音結合の程度は、パツ登録時の発声状態に大きく依存する。しかし、VCVパツを用いれば、比較的調音結合の強い標準パツを安定して得ることが可能である。

2-3. 複合パツ認識法

調音結合の度合は、結合する単語の言語的な結合の強さ、および発声者の態度によって異なる。このような調音結合の揺らぎに対処するため、調音結合の程度を考慮した複数の評価法を組合せること(方針3)が有効である。具体的には、方針1と方針2を組合せた図1のような2種類の標準パツを用いて認識を行ない両者のうち、より入力パツに近い方の結果を採用している。図1は入力音声「谷も」の場合の例で、(b)の標準パツは方針1により切り出した単語パツ「谷」、「も」である。また、(c)は方針2により単語接合

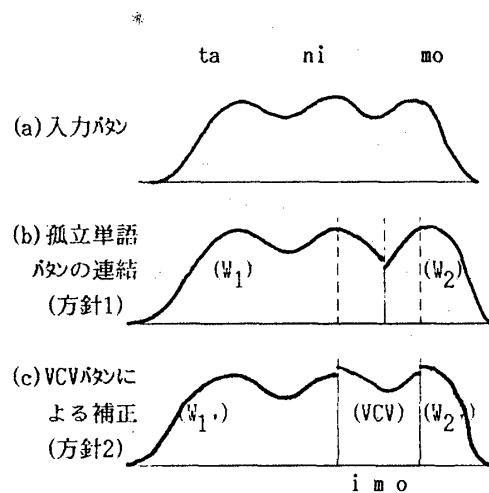


図1 マッチングに用いる2種類の標準パツ

部をVCVパターン「imo」で補正したものである。この場合、各単語パターンとVCVパターンの接合部は、語頭、語尾の母音部中央とし、前処理により単語パターンの特徴ベクトルの時間変化パターンから求める(図1-(b)破線部)。そして、 $(W_1) + (W_2)$ と $(W_1) + (VCV) + (W_2)$ の2種類を標準パターンとしてそれぞれDPマッチングを行い累積距離D, D'を求め、 $\min(D, D')$ をパターン評価に用いる。図2にこの方式を逆DPマッチングに適用した例を示す(一般的な2段DPマッチングにも適用可能)。パターン (W_1) , (W_2) に対する部分のDPマッチングは、パターン (W_1) , (W_2) に対するDPマッチングの際に線分(a), (b)上で中間結果として求めておく。これを■で示した領域のVCVパターンによる入力側両端点間-DPマッチングの両端で加算することにより、パターン $(W_1) + (VCV) + (W_2)$ の距離が求まる。

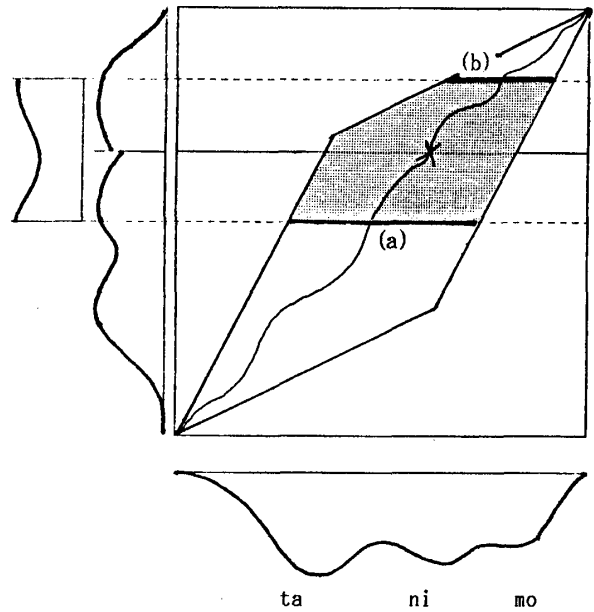


図2 逆DPマッチング

3. 実験結果

実験では調音結合の影響が大きいと考えられる自立語+付属語の文節音声を対象とした。用いた単語は表1のように自立語23語、付属語13語である。自立語としては、語尾に/N/の付くか否かだけの対立を示すもの(tani(no)「谷の」taniN(o)「他人を」等)、語尾が/yo/, /i/の対立を示すもの(meiy(o)「名誉を」meii(o)「名医を」等)を選んだ。これらはいずれも付属語が結合することにより、調音結合のため誤認識を起こしやすいと考えられる対である。付属語としては、調音結合部の認識率を調べるため、語頭子音のみ異なる1モーラの助詞を選んだ(ga ka da wa ya ra 等)。入力文節は、これら自立語+付属語の組合せのうち誤認識を起こす可能性の高い112文節(例:進化の, 新刊を; 剰余を, 上位を)を選んだ。入力および登録用の音声は女性話者1名について発声速度を変えた2種類のものを用いた。認識方法として次の4つの方法を比較した。

- ・従来方法: 単に単語パターンを結合してマッチング
- ・方法1 : 方針1を適用
- ・方法2 : 方針2を適用
- ・方法3 : 方針1, 2, 3を組合せて適用

実験結果を表2に示す。方法3が最も認識率が高く、従来方式に比べ平均誤認識率が44.9%から26.8%と39.6%減少した。また、方法1に比べて、方法2は平均3.2%改善されたのに対し、方法3ではさらに7.1%改善された。これは、調音結合の揺らぎが大きい場合に、方法3(複合パターン認識法)が有効であることを示している。

4. むすび

連続単語音声認識における調音結合に対処する方式として、「複合パターン認識法」を提案し、その有効性を確かめた。今後より一般的な対象語彙についても更に実験を進める必要がある。

表1 対象単語

自立語	蚊 伍 顔 谷 他人 たにし 進化 新刊 深海 潮影 塩加減 寄与 貴意 名誉 名医 貸与 大意 剰余 上位 容疑 容姿 用事 当人
付属語	の も を が か だ は や ら で と へ に

表2 実験結果 (文節誤認識率)

登録音声	入力音声	従来方式	方法1	方法2	方法3
ゆっくり	ゆっくり	36.6	31.3	31.3	21.4
速く	速く	52.7	38.4	35.7	25.0
ゆっくり	速く	43.8	39.3	35.7	36.6
速く	ゆっくり	46.4	39.3	33.0	24.1
平均		44.9	37.1	33.9	26.8

参考文献

- 1) 田嶋, 古村, "標準パターンの重ね合せ引き離しによる連続単語音声認識法," 音声研究会資料.
- 2) 藤井他, "単語対標準パターンを用いた連続数字音声認識," 音声研究会資料.