

# 文内の位置による発話速度の変化

正木 芽衣子<sup>†</sup> 柏岡 秀紀<sup>†‡</sup> ニック・キャンベル<sup>†‡§</sup>

<sup>†</sup> 奈良先端科学技術大学院大学 <sup>‡</sup>ATR 音声言語通信研究所 <sup>§</sup>CREST

## 1 はじめに

近年、音声言語処理技術の進展とともに、単に言語情報を伝達するだけでなく、話者の意図や感情などのパラ・非言語情報をも伝達可能な合成音声<sup>1</sup>が求められている。そのために、発話速度の制御も重要な課題となる。発話速度の制御を行うものとして、話速変換装置がある。話速変換装置を用いて発話速度をゆっくりとすると、聞き手の理解を助けるが明瞭度が損われ [1]、また、主観的なピッチがより低く認知され、認知的なバイアスが生じる [2] ことが報告されている。本稿では、音声合成における音韻継続長の予測モデルを、異なる発話速度に対応させることを目標として、発話速度の異なるデータ間の分析を試みる。

発話速度は、ポーズと密接な関係があることが知られている。ポーズが聞き手の理解に重要な役割を果たしており、人は呼吸と脳裏にある文法とを合わせながら話しているという [3]。また、日本語の音韻継続長において品詞の影響があり、自立語、特に情報を多く担っている箇所における母音長の伸張傾向が顕著であることが知られている [4]。以上のことから本稿では、品詞による継続時間長の伸縮率に注目し、品詞と発話速度に応じた音韻継続時間長の関係を明らかにするために、速い、普通、遅いの3つの発話速度のデータを収録し分析した結果について報告する。

## 2 データ収録

女性話者3名により、絵本『百万回生きたねこ』 [6] を3つの異なる発話速度で収録した。発話収録に用いたテキストは絵本を読みやすくするために漢字を用いて書き直したものである (2241文字)。収録前に練習をおこない、指定した時間で読めるようになってから収録を行った。時間指定なし (以後この発話を普通発話と呼ぶ)、遅い、速いの順に収録した。音読しやすいようテキストはA4横置き、縦書きとし、4ページに収まるよう1ページ26行×28文字 (728文字) で設定した。遅いときは原稿1ページあたり3分、速いときは原稿1ページあたり1分半を目安に発話するよう指示した。聞き手がわかりやすいように話すようにという指示も行った。読み間違った際は中断せずにそのまま読み続けた。そのため今回は1箇所でも誰かが誤った場合、話者間、発話速度間での比較が容易なように、同じ文を全話者、全発話速度の発話から削除した。分析対象は、その結果106文から82文 (1679モーラ) となった。

音声データはサンプリング周波数 48kHz, 16bit で

<sup>1</sup>Differences in speech timing according to position in the utterance

Meiko Masaki, Hideki Kashioka, Nick Campbell  
NAIST, ATR-SLT, CREST

表 1: ポーズ・誤り文除去後の各発話の継続時間長 [sec] ( ) は1秒あたりのモーラ数 [mora/sec]

	速い	普通	遅い
話者 A	182.91 (9.27)	220.74 (7.60)	271.16 (6.19)
話者 B	171.45 (9.79)	206.46 (8.13)	276.85 (6.06)
話者 C	195.22 (8.60)	218.06 (7.69)	279.51 (6.00)

DAT を用いて収録し、16kHz にダウンサンプリングした後、Julius Segmentation Kit を用いて音素アライメントを行った。ポーズを全て除去したあとの各発話の継続時間長は表1のようになった。

## 3 分析

分析に利用した品詞情報は、発話テキストの茶釜<sup>1</sup>の出力結果による。この品詞区分に基づいて分類した結果、937語、37種の品詞を得た。37種の品詞のうち出現数の多い上位6種の品詞について個数と出現頻度を表2に示す。

表 2: テキスト中における各品詞の個数と出現頻度

品詞	助動詞	名詞-一般	動詞-自立
個数	191	161	102
頻度%	20.3	17.1	10.8
品詞	助詞 係助詞	助詞 格助詞-一般	助詞 接続助詞
個数	81	75	45
頻度%	8.6	8.0	4.8

単語の継続時間長は構成音素等により異なる。そのため、発話速度の違いによる単語の継続時間長変化を調べるために伸縮率を求めた。文内での位置が等しく発話速度が異なるある語 a 同士の伸縮率 SR は以下のように定義される。

$$SR(a) = \frac{\text{遅い・速い発話内の語 a の継続時間長}}{\text{普通発話における語 a の継続時間長}}$$

各語で計算した伸縮率を各品詞ごとに比較したところ、遅い発話において、3話者に共通な品詞間での有意な差 (図1) が見られた。箱の中央の値がその品詞の

<sup>1</sup>ChaSen version 2.1 for Windows

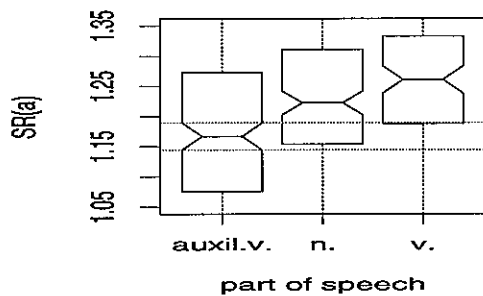


図 1: 話者 A による遅い発話と普通発話の伸縮率 左から助動詞・名詞-一般・動詞-自立

伸縮率の中央値をあらわしている。各箱中央付近のくびれた部分はその分布の 95 パーセント信頼区間を表す。x 軸に平行な直線は助動詞の分布の 95 パーセント信頼区間の境界を示す。動詞や名詞といった内容語は、もともと付属語などの機能語に比べ継続時間長が長い [4]、図 1 から、発話速度を遅くしたときに、助動詞よりもさらに継続時間長の伸縮率も高いことがわかる。これに対して、発話速度を速くしたときには、内容語と機能語の有意な差はみられなかった (図 2)。

また同じモーラで品詞間に差があるどうかをみるため「と」について分析した。各人において速く発話する際に有意な差があったが、3 人に共通した有意な差はみられなかった。

#### 4 考察

遅い発話に対して内容語の伸縮率が増加したが、速い発話では有意な差は見られなかった。これは速く発話するには限界があり、普通、発話において既に縮んでいる機能語をさらに縮めることは難しいと考えられる。普通発話で伸びている内容語は物理的には縮めやすいはずである。にもかかわらず、伸縮率がほとんど変わらないということは、速い発話においても内容語が伸びる傾向が保持されているともいえる。内容語や機能語を発話する際に人はほとんど意識していない。しかし、このような傾向が見られることは、発話速度の制御において無意識のうちに文法的な役割を使い分けしているといえる。それゆえ、より自然な音声を作成するためには文法的な働きを考慮すべきである。

「と」の伸縮率における差が話者内ではあるが速い発話においてみられたことは、品詞や速い発話をわかりやすく話すための技術力の違いや、話者の個性があらわれている可能性がある。今後データ数を増やし、各音韻ごとの比較が必要である。また、今回ポーズについては除去する以外特に考慮しなかった。ポーズによる音韻的な影響のため品詞による違いが出てくることも考えられる。今後ポーズを含めて分析を行う予定である。

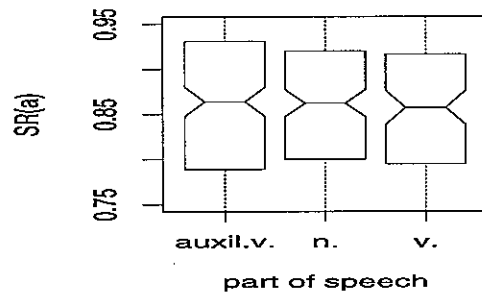


図 2: 話者 A による速い発話と普通発話の伸縮率 左から助動詞・名詞-一般・動詞-自立

#### 5 まとめ

動詞や名詞といった内容語は、もともと付属語などの機能語に比べ継続時間長が大きいですが、発話速度を遅くしたときに、助動詞よりも伸縮率が高く、より伸びることがわかった。発話速度を速くしたときには内容語と機能語の有意な差はみられなかった。

今後、ポーズや音韻による影響も考慮した、異なる発話速度に対応した継続時間長の予測モデルを構築したい。

#### 謝辞

本研究の一部は科学技術振興事業団戦略的基礎研究推進事業 (CREST/JST) の援助により行われた。

#### 参考文献

- [1] 宮坂栄一: 話すスピードと明瞭度, 言語, 大修館書店, No.9, pp.36-39(1999).
- [2] 内田照久: 音声の発話速度の制御がピッチ感及び話者の性格印象に与える影響, 日本音響学会誌, 56 巻 6 号, pp.396-405(2000).
- [3] 杉藤美代子: 談話分析・発話とポーズ, 日本人の声-日本語音声の研究 1, 和泉書院, pp.263-277(1994).
- [4] 海木延佳, 武田一哉, 匂坂芳典: 言語情報を利用した継続時間長の制御, 電子情報通信学会論文誌 (A), J75-A, 3, pp.467-473(1992-03).
- [5] 峯松信明, 片岡嘉孝, 中川聖一: 講演詞の話し言葉に対する分析, 情報処理学会研究報告, SLP-8-7, pp.39-46(1995).
- [6] 佐野洋子: 百万回生きたねこ, 講談社, (1977).