2—10—17

物語朗読における異なる話速と発話スタイル間の発話時間長制御について*

◎正木 芽衣子（奈良先端大） △柏岡 秀紀（奈良先端大／ATR）
ニック・キャンベル（奈良先端大／ATR／CREST）

1 はじめに

近年，音声言語処理技術の進展にともない，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，様々な発話スタイルにも対応可能な合成音声が求められている．発話スタイルが形成される要因には多くのものが考えられるが時間要因はどのスタイルにも深く関わっており，意図や感情を示す際にも，発話速度の制御が欠かせない．談話行為の自動判別でも，音韻継続長の貢献が大きく，F0は音韻継続長と独立には貢献しないことが報告されている [1]．本研究では，様々な発話スタイルでの音声合成を可能にする継続時間長の予測モデル構築を目標とし，発話速度及び対象となる聞き手が異なるデータ間の分析を試みる．ポーズは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言われている [2]．また，ポーズ挿入傾向・ポーズ長分布は話者により相違はあるが，性質の異なる長短2種類のポーズが存在することが定量的に明らかになっている [3]．以上のことから本稿では，速い，普通，遅いの3つの発話速度で読み上げたデータと，聞き手に幼児を想定して絵本を用いて読み聞かせたデータを用いて，ポーズについて分析を行った．

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』[5] を3つの異なる発話速度で収録した．音読しやすいように，原文を漢字に書き直し，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，聞き手がわかりやすいように話すよう指示した．指定した時間で読めるようになるまで収録前に練習を行い，各話者とも時間指定なし（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．文内のポーズだけでなく，文間のポーズも分析対象とするため，段落や一文ごとの収録は行わず，物語を初めから終わりまで通して収録した．読み間違った際は中断せず，読み飛ばしや言い直しは各自の判断に委ねた．そのためモーラ数は各発話によって異なる．

その後，実際に絵本を使って収録を行った（以後この発話を読み聞かせ発話と呼ぶ）．読み手の傍らで幼児が聞く姿を想定させ，その幼児のために絵本を読むよう指示した．前述の収録と同様に通して収録を行った．ページをめくる時間や幼児が絵を眺めるための余韻なども全てポーズに含まれる．

音声データはサンプリング周波数48kHz，16bitでDATを用いて収録した．16kHzにダウンサンプリングした後，音素セグメンテーションキットJulius Segmentation Kitを用いて音素アライメントを行い，単語間の無音区間がポーズとして付与された．語の定義は発話テキストを日本語形態素解析システム茶筌[1]で解析した出力結果による．各発話の音韻継続長，モーラ数，平均モーラ長を表1に，ポーズ長，ポーズ境界頻度の合計を表2に示す．

表 1: 各発話の平均モーラ長 [ms]

話速	話者	音韻継続長	モーラ数	平均モーラ長
速い	FME	251830	2391	105
	FMI	234410	2388	98
	FKM	268150	2386	112
普通	FME	306720	2409	127
	FMI	283960	2389	119
	FKM	298330	2387	125
遅い	FME	369770	2384	155
	FMI	376740	2382	158
	FKM	382110	2383	160
読み聞かせ	FME	349190	2380	147
	FMI	288640	2358	122
	FKM	331680	2378	139

表 2: 全音韻継続長，全ポーズ長，ポーズ境界頻度 [s]

話速	話者	音韻継続長	ポーズ長	ポーズ境界頻度
速い	FME	251.83	63.04	156
	FMI	234.41	48.22	145
	FKM	268.15	62.66	167
普通	FME	306.72	174.94	221
	FMI	283.96	126.29	216
	FKM	298.33	90.99	238
遅い	FME	369.77	265.11	283
	FMI	376.74	196.87	289
	FKM	382.11	201.30	393
読み聞かせ	FME	349.19	262.30	308
	FMI	288.64	143.86	213
	FKM	331.68	177.66	301

3 分析

発話速度及び発話スタイルの違いによるポーズ長分布の変化を調べるため，各発話のポーズ長と発話境界頻度の関係に着目する．文内のポーズは句と句の関係を，文間のポーズは文と文の関係を反映していると考えられ，文内と文間のポーズでは大きく役割が異なる。また、一般的に文内のポーズより文間のポーズの方が長いことが多い。そこで、ポーズを文内のポーズと文間のポーズにわけて分析を行った。どの発話速度においても，文内と文間のポーズ長の平均，標準偏差は大きく異なった（表3）．

*Modeling the characteristics of different speaking styles, Meiko Masaki(NAIST), Hideki Kashioka(NAIST, ATR), Nick Campbell(NAIST, ATR, CREST)

[1]ChaSen version 2.1 for Windows

表 3: 文間と文内のポーズ長の平均 [ms] と標準偏差

話速	話者	文内		文間	
		平均	SD	平均	SD
速い	FME	193	115	501	244
	FMI	143	91	400	313
	FKM	151	103	501	254
普通	FME	269	183	1348	412
	FMI	188	110	989	428
	FKM	163	130	651	191
遅い	FME	377	218	1858	621
	FMI	256	144	1404	491
	FKM	268	224	1166	341
読み聞かせ	FME	369	185	1758	1392
	FMI	272	320	1075	797
	FKM	224	243	1255	706

文内のポーズについてポーズ長を x 軸にとり，ポーズ境界頻度の 3 話者合計を y 軸にとり，ポーズの分布を見た（図 1）．多くのものに 100ms 以下のピークが見られるが，これはポーズを自動付与したために，無声化した語境界などの無音区間がポーズと付与されたためだと思われる．それ以外では，速い発話では 200ms，普通発話では 300ms，遅い発話，読み聞かせ発話では 400ms 付近に，それぞれピークが見られ，発話速度に応じた長さを示している．そこで発話速度の影響を抑えるため，ポーズ長を平均モーラ長（表 1）で除算し，正規化を行った．平均モーラ長は，全音韻継続長を全モーラ数で除算したものである．文内のポーズについて，正規化したものを図 2 に示す．ポーズ長では発話速度に応じたピークを示しているが，モーラ数に換算すると 2 モーラ付近にピークが集中している．

文間のポーズ長とポーズ境界頻度の分布では平均モーラ長で正規化を施しても，文内のポーズのように特定のモーラにピークが集まることはなかった．

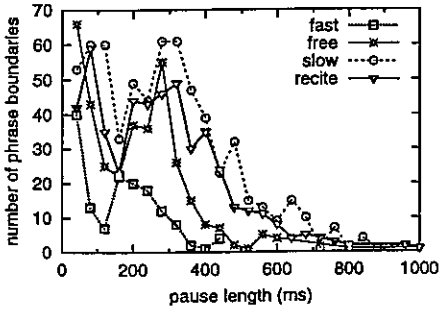

図 1: 文内における速い・普通・遅い・読み聞かせ発話間のポーズ長とポーズ境界頻度

4 考察

文内のポーズにおいては，発話スタイル，発話速度，話者を問わず，ポーズ長をモーラ数に換算すると，2 モーラ付近に集中している．このことは，文内のポーズがその発話における平均モーラ長の 2 拍分の長さをとる傾向があることを示している．つまり，文内のポーズ長は話すスピードや話者，発話スタイルで実時間には差があるものの，誰もが話速を問わず，2 モーラ分の長さを文内のポーズ長に割り当てていることになる．

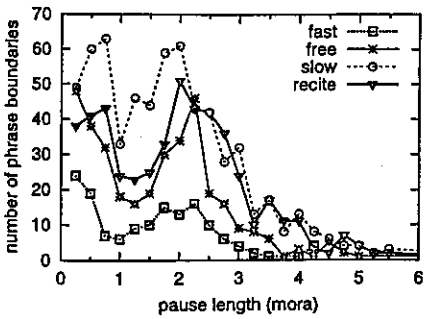

図 2: 文内における速い・普通・遅い・読み聞かせ発話間のポーズ長とポーズ境界頻度（正規化後）

このことから，文内のポーズは発話速度に比例して，ポーズ長が伸縮していることがわかる．しかし，発話速度に応じてポーズ境界の位置やポーズ境界頻度も変化するため（表 2），普通発話のポーズ長を一定比率伸縮させるだけでは不適切であり，異なる発話速度での音声を合成するには，特徴が異なる発話スタイルごとにモデルを準備する必要がある．

また，文間のポーズにおいては，平均モーラ長で正規化を行っても，明確な特徴は見受けられなかった．このことは，文と文の関係の解釈は句と句の関係に比べ個人差が大きいことや、段落間のポーズも文間のポーズとして扱ったことが影響していると思われる。

今後，談話構造も考慮し，文間のポーズについてさらに分類し分析する必要がある．

5 まとめ

文内のポーズ長は話すスピードや話者，発話スタイルで実時間では差があるものの，発話スタイル，発話速度，話者を問わず，2 モーラ分の長さを文内のポーズに割り当てており，このことは異なる発話スタイルに対応した継続時間長の予測モデルを構築する上で有益である．

今後，さらにデータを増やし，詳細な分析を行い異なる発話スタイル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] Shriberg,E. et al, Language and Speech, 41, pp.439-487(1999).

[2] 杉藤美代子，日本人の声－日本語音声の研究 1，和泉書院，pp.263-277(1994).

[3] 海木延佳，匂坂芳典，信学論 (D)，J79-D，9，pp.1455-1463(1996-09).

[4] 正木芽衣子 他，情報処理学会第 63 回全国大会講演論文集 (2)，pp.169-170(2001).

[5] 佐野洋子:百万回生きたねこ，講談社，(1977).