1—Q—14

局所的話速変化を表現する概形近似モデルの種々の音声への適用と検討*

◎高丸 圭一　広重 真人　荒木 健治　栃内 香次（北大院・工）

1　はじめに

自由会話において、話速の大きな変化は話者による"聞き手の注意をひく"表現の一つであると考えることができる。筆者らは話者の意図的な制御による話速変化を捉えることを目的として、観測される話速変化から話者による話速制御成分を抽出する回帰直線とcos関数を用いた話速概形近似モデルを提案した[2]。本稿では、本モデルを自由会話音声に適用し、モデル化の前処理であるMDAFのモデルへの影響について検討を行う。

2　話速概形近似モデル

2.1　モデルの概要

音声において観察されるモーラ持続長変化(話速変化)は、音素の性質などの話者の制御によらないもの、リズムの意図的変形などの話者の制御によるものなど、種々の要因による持続長変化をすべて重畳したものと考えられる。筆者らが提案する話速概形近似モデルはパラ言語情報の認識・理解に役立てることを指向し、観察される話速変化の中から話者による制御に起因する成分を抽出し、表現するものである。本モデルでは、意味内容を保持する最小の文法的な単位である文節程度の長さ(「フレーズ」と称する)を単位として、「話速の大局的な変化」と、話者のフレーズに対する制御による「フレーズ内での時間構造の非線形な変形」の2つを表現する。

2.2　処理の流れ

図1にモデル化の処理の流れを示す。

2.2.1　モーラセグメンテーション・モーラ持続長算出

モーラ持続長を算出するために手作業でモーラセグメンテーションを行う。長音・多重母音・促音などのモーラ、著しい調音結合が起きている箇所などモーラ境界が明確に定められない箇所では複数モーラを1つのセグメントとし、セグメント内の平均モーラ長を算出する。

2.2.2　"Mora Duration Adjusting Factor" の適用

モーラに含まれる音韻の特徴によるモーラ持続長の大きな変化は我々が抽出を指向する話者のコントロールによる話速変化とは異なるものであるが、モデルによる持続長変化の概形推定に影響を与える可能性がある。そこでモデル化の前処理として、持続長が周囲のモーラの平均値と大きく異なる幾つかのモーラ種に対して、その伸縮率に応じた係数MDAF(mora duration adjusting factor)を適用する[1]。この処理によって、著しい短縮を示すモーラ種の持続長概形への影響は取り除かれ、他の要因による持続長変化が表出しやすくなると考えられる。MDAFの適用は以下の式に従う。

$$\text{換算モーラ持続長} = \frac{\text{観察されるモーラ持続長}}{\text{MDAF}}$$

本稿では、長音、多重母音、撥音、促音、母音が無声化しているモーラにMDAFを適用した。

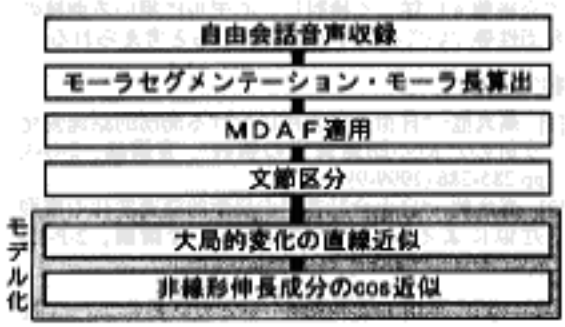

図1: 話速概形近似モデルの処理の流れ

2.2.3　大局的変化の直線による近似

各フレーズの話速の大局的変化を表すためにフレーズごとに回帰直線を求める。この直線がフレーズ内の平均的な速さと、当該フレーズが加速/減速傾向を表現する。ただし、フレーズ末モーラに伸びが生じている場合、フレーズ末モーラを含むラベルを回帰直線計算から除外する。

2.2.4　非線形伸長成分のcos曲線による近似

フレーズ内の各モーラは話者の話速制御により、均等には伸長せず、伸長しやすい箇所を中心として非線形に変形すると考えられる。そこで、フレーズ内のこの非線形な変形を表現するために、上に凸な曲線をあてはまる。モーラ持続長と回帰直線との差分値を上に凸な箇所を1つだけ持つ関数で近似する。本稿では以下に示すcos関数を用いる。

$$\hat{y} = -A\cos\left(\frac{2\pi}{T+|\Delta T|}(t+\varphi)\right)$$

ただし、$\varphi = -\Delta T\ (\Delta T \geq 0),\ 0\ (\Delta T < 0)$、$T$はフレーズ長(ただし、フレーズ末モーラに伸長が起きている場合はフレーズ長からフレーズ末モーラを除いた長さ)を、$A\ (A \geq 0)$は振幅を表す。各フレーズにおいて、回帰直線との差分値との2乗誤差が最小になるAと凸の頂点の位置$\tau = (T-\Delta T)/2$を求める。Aの大きさがフレーズ内での局所的話速変化の大きさを表し、τがフレーズ内での制御の中心位置を表していると考えられる。

3　自由会話音声へのモデルの適用

前節で述べた話速概形近似モデルを男性話者2名による自由会話音声33文に適用した。これらの音声は男子大学生により発話された自由会話音声である。図2に例を示す。各フレーズにおけるAの値とフレーズ長で正規化したτの値の分布を図3に示す。

3.1　考察

τの値が0または1付近でAの値が小さいフレーズ、すなわち非線形成分にcos曲線が当てはまっていないフレーズが多く見られた。すべてのフレーズに意図的伸長成分があるわけではないため、これらのフレーズの多くは意図的伸長が存在しないフレーズであると考えられる。しかしながら、次のような場合において、意図的伸長があるにもかかわらず近似

* 'A Study on the Application of Speech Rate Model to Several Spontaneous Conversational Speech', by Keiichi TAKAMARU, Makoto HIROSHIGE, Kenji ARAKI and Koji TOCHINAI (Graduate School of Engineering, Hokkaido University)

できていない箇所が観察された。

(1) 1つのフレーズが長い場合: このようなフレーズでは1フレーズ内に複数の意図的伸長の制御が働いている可能性があり、フレーズの分け方に問題があると考えらる。

(2) 調音結合などのために複数モーラを1まとめにしてセグメンテーションを行った箇所を含むフレーズにおいて、適切にcosによる近似が行えない例が見られた。

このような箇所でのフレーズ区分の方法は今後の検討を要する。τがフレーズ中心付近に存在した場合はcos曲線はフレーズに含まれる局所的伸長を適切に表現していると考えられる。

4 MDAFのモデルへの影響

MDAFがモデルに適切に作用しているかどうかを確認するために、MDAFの対象となるモーラを含み、

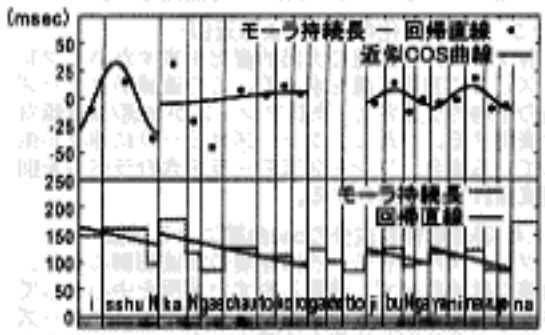

図2: 話速概形近似モデルの適用例

(一瞬／考えちゃうところが／ちょっと／自分が／やになるよな)

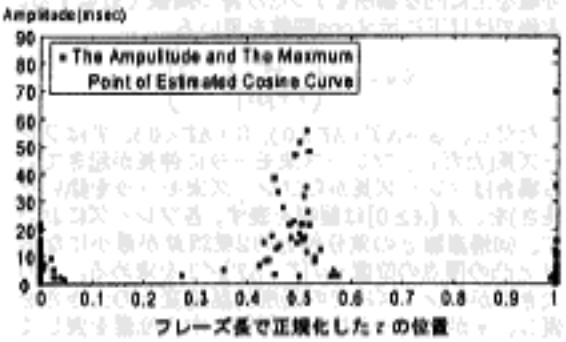

図3: 各フレーズにおけるAとτの分布

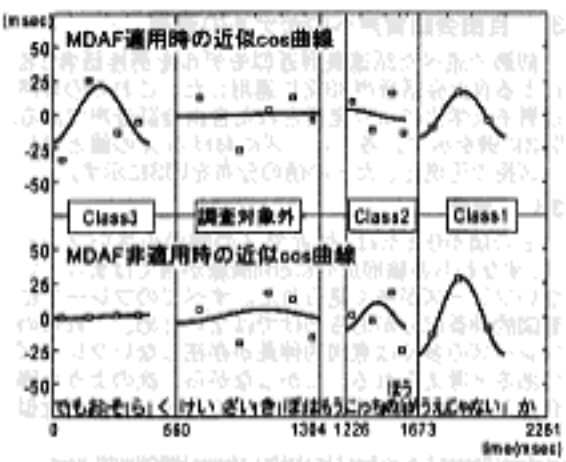

図4: MDAFの適用による近似cos曲線の変化の例

フレーズ内のセグメント数が3以上のフレーズについて、MDAF適用後にモデル化した場合と、MDAFを適用せずにモデル化した場合とのcosの振幅Aの差を調査した。ただし、MDAF適用時/非適用時いずれもAが5msec以下のフレーズはcosで近似すべき非線形制御は存在しないと判断し除外した。

MDAF適用時の振幅A_{adj}とMDAF非適用時の振幅A_{obs}の差分$Diff = A_{adj} - A_{obs}$の平均と標準偏差を表1に示す。ここで$Diff$によって各フレーズを

$$\begin{cases} Diff < \mathrm{ave} - 0.5\sigma & \text{振幅が減少(Class1)} \\ \mathrm{ave} - 0.5\sigma \le Diff < \mathrm{ave} + 0.5\sigma & \text{振幅の変化なし(Class2)} \\ \mathrm{ave} + 0.5\sigma \le Diff & \text{振幅が増加(Class3)} \end{cases}$$

のように分類した。各分類に属するフレーズ数を表2に示す。

4.1 考察

Class2に分類されるフレーズが最も多く、MDAFによって大きな影響を受けないフレーズが多いことが分かる。これは本モデルが、意図的伸長成分を頑健に抽出できることを示している考えることができる。また、Class1に含まれるフレーズ数とClass3に含まれるフレーズ数を比べると、MDAFの適用によって、cosで適切に近似されるフレーズは増加していると言える。MDAFは意図的伸長制御成分以外の成分を適切に抑制していると考えられる。しかしながら、Class1のフレーズで、MDAF非適用時には適切にcos近似されていたにもかかわらず、MDAF適用によって振幅がほとんど0になる例もあり、今後の検討を要する。MDAF適用時と非適用時の近似cos曲線の例を図4に示す。

5 まとめ

本稿では、筆者らが提案した話速概形近似モデルを自由会話音声に適用し、検討を行った。フレーズ長が長い場合やフレーズ内のセグメント数が少ない場合に、意図的非線形伸長成分を適切に近似できない例が見られた。近似が適切に行われている例では、τがフレーズ中心付近に存在する例が多く見られた。モデル化の前処理であるMDAFによってモデルに大きな影響が起きないフレーズが多く見られ、本モデルの意図的伸長成分抽出の頑健性を示していると考えられる。一方、MDAFの適用によって、振幅の値が増加しているフレーズが値が減少しているフレーズに比べ多数観察され、多くの箇所でMDAFは特定モーラ種の著しい短縮を適切に抑制していると考えられる。

Aやτの値と意図的伸長成分の存在の関係について今後個々に詳しく検討し、モデルに用いる曲線の妥当性等ついて検討する必要があると考えられる。

参考文献

[1] 高丸他: "自由会話音声における局所的話速変化分析のための話速表現の検討", 音講論, 2-Q-3, pp.285-286 (1999-9)

[2] 高丸他: "自由会話音声の局所的話速変化の概形近似によるモデル化の試み", 音講論, 2-P-12, pp.305-306 (2000-3)

表1: diffの平均・標準偏差

平均 ave	-0.512
標準偏差 σ	15.144

表2:分類毎のフレーズ数

Class 1	16
Class 2	41
Class 3	21