

# 1-Q-12 口唇形状変化の特徴を利用した 話速変化情報抽出に関する基礎的検討\*

◎角丸貴洋 広重真人 荒木健治 柳内香次 (北大工)

## 1 はじめに

日本語音声の音韻的特徴のひとつとして、話速がほぼ一定であることがあげられる。しかしながら、これは朗読音声には顕著な特徴であるが、日常会話音声においてはしばしば大きな話速変化が観測される。このことから会話などにおいて、局所話速変化の存在する個所には話者の意図や強調などの表現が含まれていると考えられる。我々は、そのような個所を捉えるために、局所話速変化抽出の研究を進めている [1]。

一方、人間の口の動きは音声に深い関わり合いがあると考えられる。これまでに、口の動きのみから話している言葉を推測しようとする読唇など、口唇画像を音声処理に応用する研究が進められている [2]。筆者らは、口唇画像情報を話速分析へ利用することを考え、視覚的情報である口唇形状の変化を、話速変化を捉えるための補助情報として利用することを目的とした解析手法を検討している。

本稿では、3名の話者について口唇開口面積の変化を口唇画像を用いて測定し、話速変化との関連や話者による相違について比較を行った結果から、口唇開口面積の変化からの、話速変化抽出に利用可能な特徴の抽出法について検討する。

## 2 処理の流れ・測定方法

口唇開口面積の変化を求める手順を述べる。まず、口唇全体を正面から家庭用ビデオカメラで撮影し、同時に DAT で音声を録音する。この際、口唇内は暗くなるように照明をあてる。カメラ・マイク・話者・照明の位置関係を図に示す。ビデオカメラで撮影した映像から画像サイズ 640 × 480 (pixel)、フルカラー (1677 万色)、フレームレート 30 (fps) で口唇画像を取り込み口唇の面積測定を行い、DAT により 48kHz サンプリングで録音した音声を 12kHz にダウンサンプリングしたものを用いて、音素区切りおよび文節区切りごとに求めた平均話速の時間変化を求める。ビデオカメラの音声を基にしてビデオカメラの画像と DAT の音声を同期させる。

### 2.1 顔画像からの口唇開口面積の測定方法

本稿では開口面積変化と話速との関係の有無についての基礎的な検討を目的とするため、画像からの開口部抽出はシンプルな手法を用いる。

\* A basic study for detection of local speech rate variation using lip shape variation. By Takahiro Kakumaru, Makoto Hiroshige, Kenji Araki and Koji Tochima (Hokkaido Univ.)

フルカラーで取り込んだ画像をグレースケールに変換した後、いくつかのキーフレームにおいて取り込んだ画像から処理対象となる口唇周辺領域を手動で抽出する。残りのフレームについてはキーフレームにおいて定めた領域をあてはめて抽出する。

閾値を設定して画像を2値化し、主に暗い部分を開口領域として抽出するが、開口領域に歯が含まれて明るくなる可能性を考慮し、以下の2段階からなる方法を用いる。

#### 2.1.1 暗い開口領域の抽出法

暗くなっている口唇内を抽出できるように、閾値を手動で設定して画像を2値化する。閾値を少しずつ調整し、キーフレームにおいて口唇原画像と2値化画像とを見比べて最適となる閾値を用いて、暗い開口領域を抽出する。歯が出ていない場合はここで抽出した領域が開口領域となる。ほとんどの場合、キーフレームで抽出した閾値は他のフレームにそのまま適用可能であった。暗い開口領域を抽出した画像の例を図1に示す。

#### 2.1.2 開口領域内の明るい領域(歯)の抽出法

暗くなっている口唇内の中の明るい部分(主に歯)を抽出できるように閾値を2.1.1と同様にして手動で設定して画像を2値化する。この領域と2.1.1に述べた方法で抽出した領域を論理和により合わせたものを口唇開口領域としこの画素数を口唇開口面積とする。

## 3 実験と考察

### 3.1 実験と結果

実験に用いたデータは3名の男子大学院生によって発話された数文の音声と口唇画像で、話者には自然な口形で発話することと、1つの文につき以下に示す4種類の異なる速さで発話することの2つを指示した。

1. 自然な速さで発話する。
2. 全体的に速く発話する。

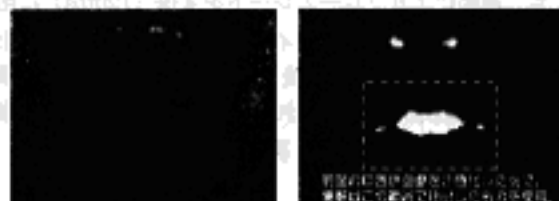


図1: 原画像(左)と2値化画像(右)

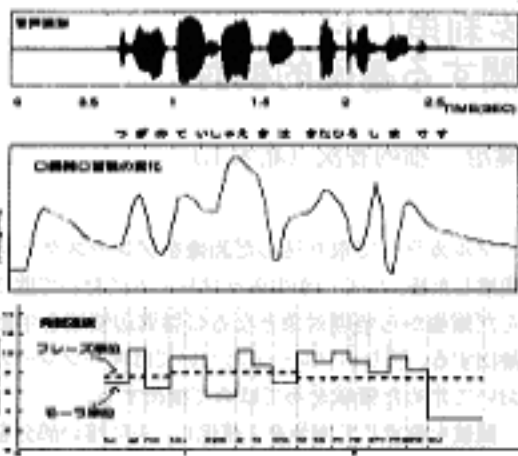


図2: 自然な速さで発話時の口唇開口面積の変化例 (話者S)

3. 一部分(下線で指定)をゆっくりと発話する。
4. 全体的にゆっくりと発話する。

このデータに対して前述の方法で画像から口唇開口面積変化を測定し、話速変化との比較、および話者の違いによる変化の検出を行う。

例として、「次の停車駅は北広島です。」という文に対して、上記1,3の2つの話速パターンについて、面積変化と話速変化の例(話者S)を図2,3に示す。下線部はゆっくりと発話するように指定した部分である。

これらの結果から、同一話者が同じ文を話していても話速が変化すると口唇開口面積変化も異なってくる。話者が異なると同一文をほぼ同じ話速で発話していても口唇開口面積変化のグラフ形状は異なっていることがわかる。しかし、口を閉じているとみられる顕著に下に凸な箇所はすべてにおいてほぼ同じ回数であることが観測される。

### 3.2 考察

上述のように、口を閉じているとみられる顕著に下に凸な箇所は各条件において安定に表出していることからことから、口唇音を発話している箇所は、現時点においてもっとも安定に画像から抽出できると考えられる。口唇音強度と話速とに一定の関係があれば、画像からの情報も話速抽出に有用となりうる。開口面積変化グラフから口唇音点を検出するには、文献[3]などを参考にした以下の手法が考えられる。すなわち、補間によりフレームレートを4倍(120fps)に増加させた後、ローパスフィルタをかけ、細部の変化を取り除く(図4)。この後、閾値処理などによって顕著に下に凸な点を検出すれば、話者に依存しない特徴抽出が可能と考えられ、現在検討している。

### 4 おわりに

本稿では、局所的な話速変化検出の補助情報として口唇開口面積変化を利用することに着目し、口唇開口面

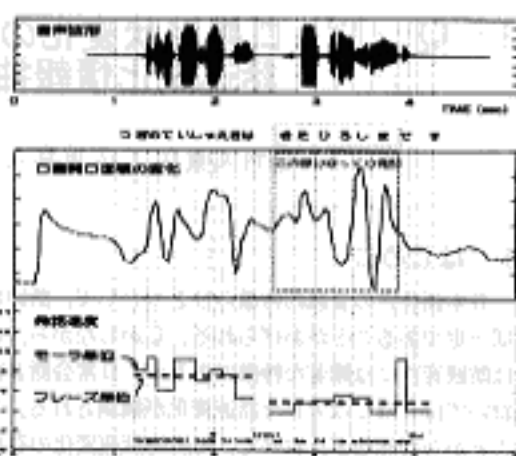


図3: 一部をゆっくり発話時の口唇開口面積の変化例 (話者S)

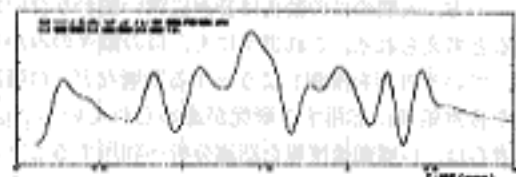


図4: 口唇開口面積の変化を補間しローパスフィルタを適用した例 (話者S, 自然な速さでの発話)

積変化と話速との比較を行ない、さらに話者による相違を調査することにより口唇画像情報の可能性について検討した。

話者の違いによる口唇開口面積変化の共通点や相違点の検討が必要であり、さらには面積変化からの話速変化パラメータの抽出、音声パワーなど他の音響パラメータとの関連、口唇開口面積の測定方法の改良などを今後進めていく予定である。

### 参考文献

- [1] 佐々木洋一, 広重真人, 宮永喜一, 柳内香次: "局所的な話速変化検出を指向した簡易セグメンテーション手法と実環境音声への適応について", 信学技報, SP98-142, pp.47-54(1999-02).
- [2] 積山薫, 城和貴, 梅田三千雄: "読唇による単音節の認識", 信学技報, IE87-108, pp.33-40(1988).
- [3] 奥村晃弘, 濱口佳孝, 岡野純治, 宮崎敏彦: "顔画像情報と音声情報の統合による発話認識", 情報処理学会論文誌, Vol.39, No.12, pp.3232-3241(1998).
- [4] 古井貞照: "デジタル音声処理", 東海大学出版会, (1985).
- [5] 角丸貴洋, 広重真人, 荒木健治, 柳内香次: "話速変化情報抽出のための口唇形状変化分析に関する予備的検討", 電気関係学会北海道支部連合大会講演論文集, p.300(1999).