

2P-12 歌唱時における口唇の運動分析

竹内英人* 大照完* 橋本周司** 田中章喜***

*早稲田大学 **東邦大学 ***松下技研株式会社

顔の画像処理の中で、特徴的な口形あるいは口唇輪郭を抽出する研究が多く行われている^{[1][2]}。ここでは、口の動き、特に唇が比較的大きく明確に変わる歌唱時を対象として、その動特性の分析を行った。その結果、歌詞中の母音の変化から、歌詞の進行を実時間で追跡可能になったばかりでなく、分析合成手法を用いた歌唱中の顔画像伝送に有効であることが判明した。

1. システム

まず、図1に示す口唇の上、下、左、右4点を白黒CCDカメラでとらえる。この4点は、あらかじめ小さいランプ、または白い口紅をぬっておく。得られたビデオ信号から、当研究室で開発した対象物移動検出装置により4組のXY座標を実時間で検出しコンピュータに転送する。この検出された座標を用いて口唇の動き分析を行う。分析には上下、左右の座標の差分を用いるため、顔の動きの影響を受けずに分析を行うことができる。

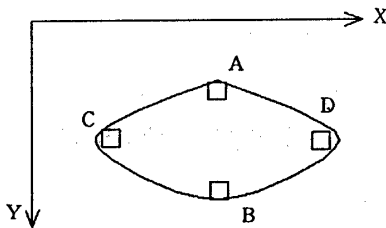


図1 口唇の特徴点

2. 対象物移動検出装置

対象物移動検出装置は1/30秒毎に、CCDカメラからの画像(256×256×8bit)を2値化し、対象物の座標を検出する。2値化画像データはシフトレジスタに入力され、左上から水平に4×4画素を走査し、4×4画素内がすべて白の領域をコンパレータにより検出す

る。点Aは最初に検出された白領域の座標、Bは1フィールド分のスキャンが終了したときの最後に検出された座標、またC、Dは、検出された座標の中でX座標が最小および最大の白領域の座標に対応する。

3. 口唇の運動分析

上記のシステムを用いて、歌唱時の口唇の運動分析を行った。歌唱は普通の会話の時よりも大きくはっきりと口を動かす。ここでは、曲、歌詞のわかっている簡単な童謡で分析を行った。単音節中の母音の認識は読唇でも比較的精度高くなるので、母音の変化を中心に歌詞の進行を実時間で追跡した。

4. 口形変化の検出

点A、BのY座標(A_y, B_y)及び、C、DのX座標(C_x, D_x)から、

$$\begin{aligned} VL(i) &= B_y(i) - A_y(i) \\ HL(i) &= D_x(i) - C_x(i) \end{aligned} \quad (1)$$

を求める。ここで、VLは口唇の縦に開いた長さ、HLは横に開いた長さに対応する。特に、 $i=0$ では口唇を閉じた時の長さ(VL_0, HL_0)となる。

同じ歌詞を歌った場合、毎回同様の口形変化が生じる。そこで、VL、 VL_0 及びVLの時間変化 ΔVL

$$\Delta VL(i) = VL(i) - VL(i-1) \quad (2)$$

の符号変化から、口唇を開く、閉じる、小口形を大口形にする、大口形を小口形にするなどの動きをとらえ、口形の変化の時刻を求める。変化と次の変化の間を母音の発音されている期間とし、歌詞の進行を追跡する。

なお図2(b)のように、同じ母音を含む音節が続く場合、母音と母音の間の子音が閉口形

Analysis of lip motion on singing

Hideto TAKEUCHI*, Sadamu OHTERU*, Shuji HASHIMOTO**, Akiyoshi TANAKA***

*Waseda Univ., **Toho Univ., ***Matushita Research Institute Tokyo, Inc.

を伴うものでないと二つの音節を区別することができない。この場合、歌詞からこのような時点を予測して、一つの変化で二つの母音が発音されたとする。

5. むすび

口形の変化をとらえることで、実時間で歌詞の進行を追跡を行った。今回は対象とした童謡が、母音の変化の多いゆっくりした曲であったが、今後はもっと高度な曲にも対応できるよう改良を行う。また、VL, HLから発音中の母音の判別を行い、歌詞を間違えたり、発音時刻の検出ミスが生じたときのチェックも検討中である。

さらに、図3(a)のように、送信側から、歌声と口形情報(VL, HL)を送信すれば、受信側においてSFM^[3]で合成された顔画像を実時間で制御することが可能になった。また、歌唱中の発音時刻が検出できたので、歌唱のテンポが求められ、このテンポ情報を用いることで「歌うコンピュータ」^[4]の制御を口唇の動きで行えるようになった。つまり図3(b)のように、歌詞、楽譜等を受信側に与えておけば、テンポ情報を送るだけで顔画像が制御され、予め保存された歌声をテンポにあわせて歌わせることができ、顔画像伝送の帯域圧縮にも有効である。

参考文献

- [1] 間瀬他, "オプティカルフローを用いた読唇", 信学論 D-II Vol. J73-D-II NO. 6 pp. 796-803 1990年6月
- [2] 斎田他, "口唇情報を利用した発声発語訓練器", 第9回感覚代行シンポジウム pp. 55-58 昭和58年12月
- [3] 橋本他, "スプリングフレームモデルを用いた顔画像の変形処理", 電子情報通信学会春期全国大会, 1188(1987)
- [4] 大照他, "実時間楽音制御システムによる「歌うコンピュータ」の実現", 本大会予稿

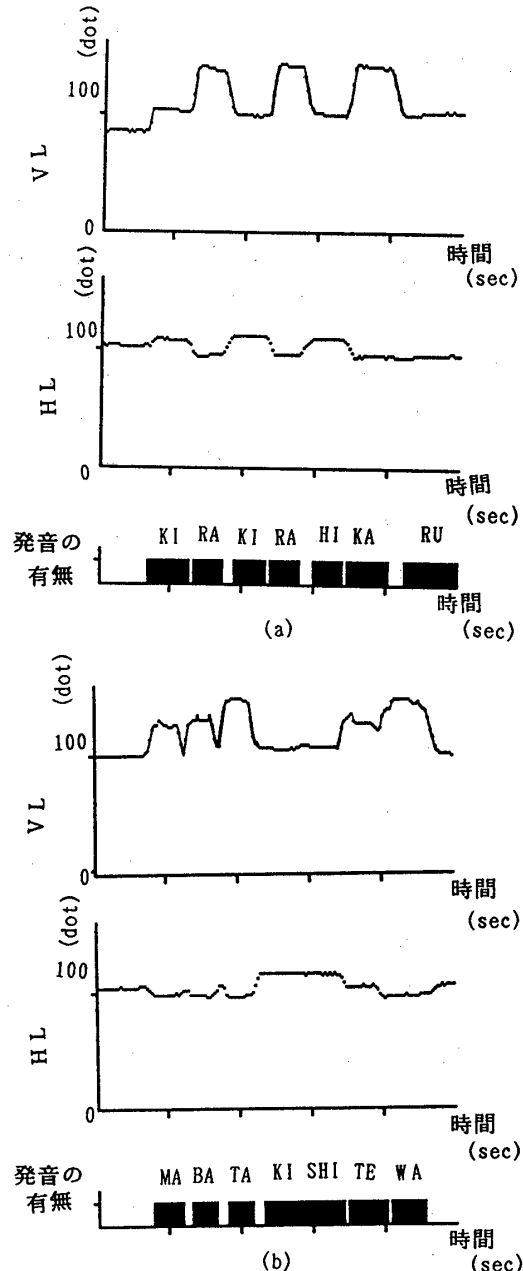


図2 口唇の動き(VL, HL)と検出された発音の有無

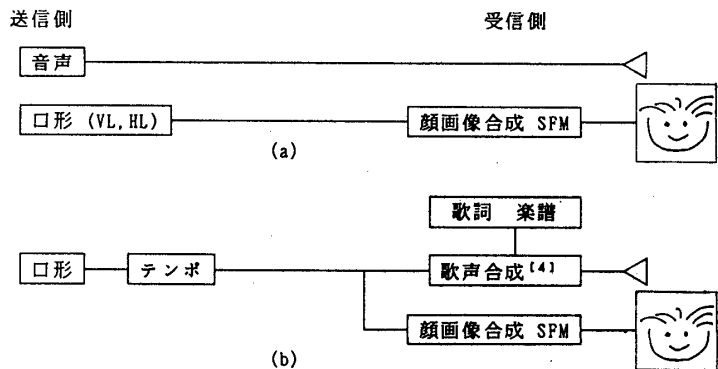


図3 受信側における歌唱時の顔画像合成