

歌唱時における顔画像の伝送制御

7B-3

竹内英人 橋本周司 大照完
早稲田大学

動画伝送をリアルタイムで実現する一手法として、送・受信の両端で知識を共有し、送信側の分析結果をもとに、受信側で画像合成する知的符号化が広く検討されている。筆者らは、顔動画の分析の一例として歌唱時の口唇運動の分析を試みた[1]。最近これらをさらに改良し、分析合成手法による歌唱時の顔画像リアルタイム伝送に応用したので、その概要を報告する。

1. システムの概略

送信側及び受信側ではあらかじめ、歌詞、楽譜等の情報が与えられているものとする。また、伝送開始前に、基準顔画像を受信側に与えておく。

まず、送信側では、口唇の動きをダイナミックにとらえ、運動分析部で歌詞の各音節の発音時刻を求め、歌詞の各発音時刻の間隔から歌唱のテンポを求め、予め与えられているその曲の標準テンポとの比を受信側に伝送

する。受信側では、歌声再生部で、コンピュータ内に保存されている歌声を伝送されたテンポで再生し、同時に、顔画像再生部で、その歌声に対応した口形の顔画像の再生を行う。

2. 送信システム

入力部では、白黒CCDカメラからの画像(256×256×8bit)を対象物位置検出装置を用いて1/30秒毎に、口唇の上、下、左、右4組のXY座標を実時間で検出する。この時、口唇にある程度以上輝度を持たせるため口紅又はマーカーを付着させる。

運動分析部では、口唇の上下のY座標(Ay, By)及び、左右のX座標(Cx, Dx)から、口唇運動情報(縦に開いた長さVL、横に開いた長さHL)

$$VL(t) = By(t) - Ay(t)$$

$$HL(t) = Dx(t) - Cx(t) \quad (1)$$

を算出し、このうちVL(t)に対して次の学習機械型のフィルタを用いて発音時刻の特定を行なう。

$$Y = U \left(\sum_{k=0}^n Fil(k) \cdot VL(t-k) - \theta \right) \quad (2)$$

(但しU()はステップ関数)

上式のFil(k)及びしきい値θは、開口時にY=1、それ以外でY=0になるように、事前に得られた口唇運動情報をもとに学習により決定する。同様に、閉口時を検出するフィルタも作成する。両フィルタから求められた口唇の開閉開始時刻を発音時刻とし、歌唱のテンポを求め受信側に伝送する。

なお、同じ母音を含む音節が続き、その間に閉口形を伴わない場合は、一つの変化で二つの音が発音されたとみなす。

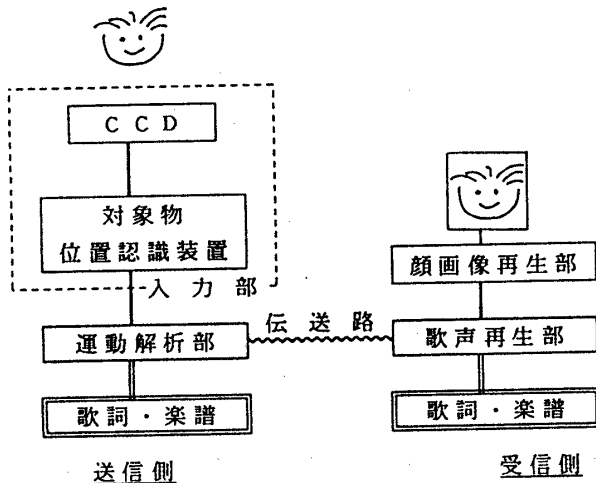


図1 システムの概略

3. 受信システム

歌声再生部では、実時間音声速度制御システム[2]を用いて、伝送されたテンポをもとに歌声を再生する。これは、あらかじめ標準テンポで歌った歌声をDSP(デジタル・シグナル・プロセッサ)で処理し、ピッチ変化をさせずに再生速度を制御する。

顔画像再生部では、SFM(スプリング・フレーム・モデル)[3]を用いて、合成した顔画像を表示する。ここでは、声にあわせてリアルタイムで顔画像を再生するために、あらかじめ基準顔画像からいくつかの口形の顔画像を作成しておき、その顔画像の切り替えで再生を行った。

この受信側の顔画像の口形の種類を増やすことで、より滑らかな顔画像の再生が可能になる。

4. 伝送制御実験

上記のシステムを用いて、歌唱時の顔画像の伝送制御の実験を行った。対象曲として、母音の変化が多く、テンポのゆっくりした童謡「きらきら星」を用いた。

その結果、曲の標準テンポから大きくずれることなく歌唱が行われれば、受信側の歌声及び顔画像は、送信側で検出されたテンポに合わせて再生できることが判った。しかし、運動分析部において発音時刻の検出に失敗すると、その後の対応が取れず、受信側と異なる再生が行われる恐れがある。

なお、本手法では、受信側で既に保存された歌声を再生するので、あらためて音声の伝送を行なう必要はない。

5. むすび

送・受信側で歌詞、楽譜等の知識を共有させることにより、受信側の運動分析で検出されたテンポ情報を送るだけで、受信側にある歌声をテンポに合わせて歌わせ、さらにその歌声に合った顔画像を制御することができた。

今後は、顔の表情などの情報も伝送できるように、送信側の分析能力を高めていく予定

である。このため、マーカー等を使用しない、カラー画像を用いた口唇の実時間自動抽出を検討中である。また、運動解析部の発音検出ミスについては、特定の母音の認識を平行して行い歌詞の進行を監視する方法を検討している。

なお、本手法で伝送距離を短くし、送・受信を同一場所で行った場合、受信側の歌声及び顔画像を他の人物にすることにより、仮想的な二人の合唱を行わせることができる。

参考文献

- [1]大照, 橋本他, "歌唱時における口唇の運動分析", 情報処理学会春季全国大会, 2-323(1991)
- [2]大照, 橋本他, "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", 情報処理学会春季全国大会, 1-315(1991)
- [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", 電通学会春期全国大会, 1188(1987)

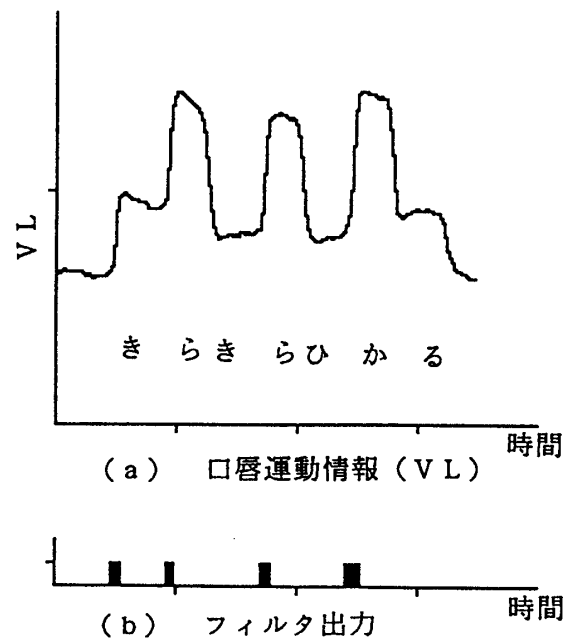


図2 口唇運動情報とフィルタ出力(開口時)