映像・音声信号の相対時間差補償

正会員 澁井 栄蔵[†], 正会員 釣 譲一[†]
正会員 児玉 勝彦[†]

フレームメモリーを応用したデジタル映像機器が番組の制作から送出まで数多く利用されているが、その結果、映像に対して音声が行先するという自然界では起こり得ない現象が発生している。こうした問題、特にフレームシンクロナイザーの使用に伴って生じる映像信号と音声信号の特異なずれの発生機構とその補正について、ずれに関する主観的評価実験結果を含めて述べる。

1. ま え が き

最近の映像信号のデジタル処理技術の進歩は目ざましい。特に大容量デジタル IC メモリーを使用した 1 フレームメモリーの応用技術により、フレームシンクロナイザー (FS)、デジタル特殊効果装置 (DVE, DVP)、移動体軌跡表示装置、ノイズリデューサー、画質補正装置などの各種の装置が開発され、番組の制作から放送までの各過程で活用されている。

これらのデジタル 1 フレームメモリーの応用映像機器 (以下フレームメモリー応用映像機器という) の活用に伴い自然界では起こり得ない“映像に対して音声が進む”現象が発生し始めた。

これは、フレームメモリー応用映像機器では、一般に装置 1 台あたり映像信号がメモリー容量に相当する分だけ、すなわち 1 フレーム (33 ms) 遅延するために発生する問題である。

特に FS の場合には、メモリーへの書込みと読出しがそれぞれ独立した周波数、位相で行われるため、相互の周波数差により映像信号の遅延量は 0 フレームから 1 フレームまで変化することになり、音声の進みの補正を難しくしている。

本稿では、FS の使用による映像信号と音声信号のずれの発生機構とその補正について、ずれに関する主観的評価実験の結果を含め考察する。

2. 映像と音声のずれの主観評価

映像に対する音声のずれを補正する第 1 ステップとして、ずれに対する許容限を把握することが必要であり、日常放送している多種多様な番組の中から代表的なシーン (写真 1~写真 8) を選び出し、0 フレームから 1 フレーム間隔で最大 4 フレームまで映像に対する音声の進み、および遅れについて図 1 に示す視聴条件により評価実験を行った。

この評価実験の結果を図 2~図 5 に示す。図の縦軸は評価尺度を、横軸は映像に対する音声のずれ量を示

映 像	表 示 装 置 周 囲 光 照 度 視 距 離	20 インチカラーモニター 50:1 75lx (画面の垂直照度) 6~7H (H: 画面の高さ)
	再 生 装 置 聴 取 距 離	モニター用スピーカー 2~2.5m

図 1 視聴条件

表 1 評価尺度

評価尺度	内 容
5	ずれがわからない
4	ずれがわかるが気にならない
3	ずれが気になるが我慢できる
2	ずれが我慢できない
1	ずれが全く我慢できない

[†] NHK 運用技術局

“Compensation of Relative Time Difference between Picture and Sound” by Eizo Shibui, Joichi Tsuru and Katsuhiko Kodama (NHK Operation Center, Tokyo)

写真 1 アナウンサーアップ (会話)

写真 5 ゴルフ (スポーツ)

写真 2 落語家 (会話)

写真 6 野球 (スポーツ)

写真 3 歌手アップ (歌謡)

写真 7 琴 (楽器演奏)

写真 4 コーラス (歌謡)

写真 8 オーケストラ (楽器演奏)

図2 評価の分布〔会話〕

図3 評価の分布〔歌謡〕

図4 評価の分布〔スポーツ〕

図5 評価の分布〔楽器演奏〕

す。この結果をまとめると、

- ① 映像に対する音声の進みについての評価は、音声の遅れに比べ評価曲線の傾斜が急であり、音声の進みに対して違和感が大きいことを示している。
- ② 音声の遅れに対しては、〔会話〕がやや厳しいものの全般的に似かよった甘い評価となっている。
- ③ 〔楽器演奏〕と〔会話〕、ロングサイズとアップサイズを比べると、〔会話〕およびアップサイズの方が評価がやや厳しく、絵柄により評価が変わる。

さらに、図2～図5の評価実験結果をもとにして、映像と音声のずれの許容限を算出した。この許容限の算出例を図6に示す。

許容限(評価尺度3以下が50%の値)は評価尺度値を95%の信頼区間で区間推定して算出したもので、図6の例では許容限は②～③の範囲に分布している。

②は上限値(一番甘い値)、③は中間値、④は下限値(一番厳しい値)を示す。

図7は、この算出方法により求めた各プログラムの

図6 許容限の算出例

図7 許容限の分布

許容限の分布で、この結果をまとめると、

- ① 音声の進みは1フレームから3フレームの間に許容限がばらついているが、一番厳しい値はアナウンサーアップの1フレームである。
- ② 音声の遅れは野球の3フレームが一番厳しい値である。

以上の結果から、音声の進みに対して違和感が大きいことが現れている。そしてプログラムの内容の絵柄により許容限に幅があるものの、フレームメモリ応用映像機器の利用は1台が限度といえる。ところがFSは一般に縦続接続して使用されることが多く、その結果として音声の進み量が多くなり、この補正が必要となる。

3. フレームシンクロナイザーによる映像と音声のずれ

FSの場合、映像信号の遅延量が0フレームから1フレームまで変化し、さらに後述するメモリ上で発生する追い越され現象による急激な変化を伴うため、映像に対する音声の進みの補正には特別な工夫が必要である。

このFSで発生する映像信号遅延量変化の発生機構をFSの1フレームメモリを図8に示す円形と仮想し説明する。

図8 分布による映像信号遅延

図9 FSによる映像信号の遅延量変化説明図

局外入力映像信号のメモリへの書込みと、メモリからの読出しスタートアドレス(0番地)は局外入力映像信号および局内基準同期から発生したフレーム情報により規正される。また書込み・読出しクロックはそれぞれ局外入力映像のバースト信号および局内基準副搬送波から発生する。このためメモリへ書込むアドレスとメモリから読出すアドレスは異なることになる。

図8の例で説明すると、メモリ上の60番地に書込んでいる時、400番地から読出しが行われ、角度 ω_d だけ遅れて読出されることになる。この時の遅延時間 t_d は $33\text{ms} \times \omega_d / 360^\circ$ となる。この遅延時間 t_d は常に一定なものではなく、局外と局内の副搬送波の周波数差により徐々に変化する。

この変化の様子を局内基準副搬送波周波数が局外のそれに比べて高い場合を例にとり図9により説明すると、書込み周波数より読出し周波数の方が高いので次のようになる。

- (a) メモリへの書込みが $\triangle W$ で行われている時読出しは $\triangle R$ で行われ、この時の映像の遅延量を t_{d1} (μs)とする。
- (b) (a)から T 時間後、局内基準副搬送波周波数の方が局外のそれに比べて高いため $\triangle R$ が $\triangle W$ に徐々に近づき映像の遅延量は t_{d2} (μs)となる。
- (c) (b)からさらに T 時間後 $\triangle R$ が $\triangle W$ に追いつきメモリへ書込む直前に読出し制御が発生するため、1フレーム前に書込まれた映像を再度読出し、遅延量は0フレームから1フレームに急激に変化した後、 $\triangle R$ が $\triangle W$ から徐々に離れる。

このようなメモリ上で発生する現象は“追い越し追い越され”といわれている。

逆に局内基準副搬送波周波数が局外のそれに比べて低い時は、書込み周波数より読出し周波数が低くなるので、メモリから読出す直前に書込み制御が発生し、今まで書込まれてから1フレーム後を読出していたものが、急激に書込み直後の読出しに変わる。このため1フレーム前の映像は新しく書込まれた映像に変わり、映像の遅延量は1フ

図 10 FS 出力映像信号の遅延量変化

レーンから 0 フレームに急激に変化する。

このように FS 出力の映像信号の遅延量は時間の関数となるが、追い越し追い越され現象の発生によって映像信号の遅延量に急激な変化を伴うことになる。

この追い越し追い越され現象は、局内基準副搬送波周波数と局外のそれとの差が 10 Hz の時は 3 時間 19 分に 1 回の割合いで発生する。

4. 映像と音声のずれの補正

前述のように、FS 出力の映像信号の遅延量は 0 フレームから 1 フレームまで急激な変化を伴う特異な変化をするため、映像と音声のずれを補正するには、映像信号の遅延量の変化に追従して音声信号を遅延させる装置が必要となる。

現在、音声遅延装置は劇場やホールの音場補正やエコー効果用などに使用されているが、この種の装置では固定的な映像の遅延には対応できても、FS を使用した場合のように遅延量の変化が特異なものには対応が難しい。

また FS は一般に 1 台だけ使用されることは少なく、複数台の FS が多段接続して使用されるため映像の遅延量は、より複雑な変化をすることになる。

このため FS を使用している系統全体で映像と音声のずれを補正することは困難であり、個々に FS と組み合わせて補正することが必要となる。

4.1 デジタル自動追従音声遅延装置

映像信号の遅延量の変化に追従して音声信号を遅延させる方法として、FS 入出力映像と音声遅延装置の書込み・読出しアドレスからずれの量を検出し、自動的に音声遅延させるデジタル自動追従音声遅延装

図 11 FS 縦続接続の一例

図 12 デジタル自動追従音声遅延装置の構成

図 13 読出しクロック周波数制御方式

図 14 読出しアドレス制御方式

置の構成を図 12 に示す。

音声の遅延量を変化させる手法として各種あるが、一例として次の方法がある。

① 読出しクロック周波数制御方式

② 読出しアドレス制御方式

(1) 読出しクロック周波数制御方式

書込み・読出しクロック発生器をそれぞれ別にし、遅延量検出部で検出したずれ量により読出しクロック周波数を制御し、FS で発生する映像の遅延量変化に追従させる。

(2) 読出しアドレス制御方式

書込み・読出しクロックを同一にし、遅延量検出部で検出したずれ量により、メモリーに書込まれた PCM データを読出す時に読出しアドレスを制御して、1 ワードのデータの 2 度読みあるいは間引きを行い、映像の遅延量変化に追従させる。

読出しアドレス制御方式の概念図を図 15 に示す。

これらの方法により、映像の遅延量の変化に音声を追従させると、FS のメモリー上で発生する追い越し

図 15 読出しアドレス制御概念図

追い越され現象による遅延量の急激な変化にも音声を追従することになってしまい、異常音として知覚される。

4.2 音声異常音の防止

異常音の発生を防止するには、遅延量の急激な変化の発生を予測して音声遅延量を制御する。

遅延量の急激な変化を予測する方法として、

- ① メモリー書込み・読出しアドレスの比較
- ② メモリー書込み・読出しアドレスカウンタで発生したウインドーパルスの比較
- ③ 単位時間あたりの位相変化率からの予測

などがある。これらの方法による予測から、遅延量の急激な変化の発生前に、主観評価結果の映像に対する音声の遅れの許容限が甘いことを利用して、読出しクロック周波数または読出しアドレスを制御し、図 16 に示す遅延量の制御をする。

局内基準副搬送波周波数が FS 入力局外映像信号の副搬送波周波数より高い場合についての制御を図 16 により説明する。

A：映像信号の遅延量変化に追従させる期間
B：音声信号の遅延量を増加させる期間

図 16 局内基準副搬送波周波数が高い時の音声遅延量制御

遅延量の急激な変化の予測により x 点に達した時、音声の遅延量を徐々に増加させ (図 16 の B の期間) 映像の遅延量の急激な変化が終った時点で映像に対して音声が一一致するように制御する。

予測方法により B の期間の遅延量は可変あるいは一定のいずれかとなる。予測方法の ① ② による場合は前者、また予測方法 ③ による場合には後者となる。

予測方法 ① ② は、時間軸上の x 点が常に一定となるため局内基準副搬送波周波数と局外入力映像信号の副搬送波との周波数差により B の期間の遅延量を変える (周波数差が大きい時には位相変化率が大きいため遅延量を増し、逆に周波数差が小さい時には遅延量を少なくする) 必要がある。

これに対して予測方法 ③ では、副搬送波周波数の差により時間軸上の x 点が右または左に (周波数差の大きい時は左へ、逆の場合は右へ) 移動し、B の期間の遅延量は常に一定とすることができる。

遅延量の制御および次に述べる音声ピッチの変化の両者を考慮すると予測方法 ③ がベターである。

4.3 音声ピッチの変化

FS による映像信号の遅延量の変化に追従して音声の遅延量を変化させるこの種の装置では、出力の音声信号のピッチが変化することになる。このため音声ピッチの変化量が聴感上問題とならない値とすることが必要となる。

この音声ピッチの変化量を図 16 の A および B の期間について説明する。

(1) A の期間の音声ピッチの変化量

A の期間の音声ピッチの変化量は、FS 入力の局外映像信号の副搬送波周波数と局内基準副搬送波周波数との差により決まる。この副搬送波周波数の規格は、偏差が ± 10 Hz である。いま仮りに局外と局内の副搬送波周波数がそれぞれ上限および下限にずれている時、すなわち周波数差が 20 Hz の時の音声ピッチの変化は 0.00055% となり、この値は聴感上全く問題のない値である。

(2) B の期間の音声ピッチの変化量

B の期間は A と異なり、遅延量の急激な変化による異常音の発生を防ぐため 0 フレーム付近から 1 フレームまでの 33 ms 分の音声の読出しピッチを広げて映像に対する音声の遅延量を徐々に大きくしてゆく。この音声ピッチの変化に対する聴感上の検知限は、最も検知しやすい 1 kHz 付近で 0.1% 程度である。このため読出しピッチの変化を 0.1% 以下とするように、B の期間の読出しを 33 sec 以上の時間で行う。

この制御により音声ピッチの変化量は0.1%以下となり、聴感上問題のない値となる。

5. む す び

従来、映像処理の面にのみ目が向けられがちであったが、“映像に対して音声が行先する”という自然界では起こり得ない新しい問題にも目を向けなければならない。今回はこの問題と補償方法の一端を紹介したが、これらの他にも、FSの運用方法に関連して発生する問題や、この種の補償装置が多段接続された時の聴感上の問題など、多くの検討事項が残されており、こうした問題の解決に取り組んでゆく必要がある。

(昭和57年4月2日受付)

【参 考 文 献】

- 1) 西村、橋本：“系列範ちゅう法による画質評価の信頼性”，テレビ全大予（1975）
- 2) J. P. ギルフォード：“精神測定法”，培風館（1972）
- 3) 田久良久：“心理学的測定法”，東京大学出版会（1973）
- 4) 国重、福田ほか：“FS縦続接続による映像と音声のずれに対する主観的評価”，テレビ学会放送現業技術研資，8（1981）
- 5) 加納、今井ほか：“フレームシンクロナイザー”，テレビ誌，29，9（1975）
- 6) 妹川、杉本：“フレームシンクロナイザーとその運用”，放送技術，4（1976）
- 7) 特集：“テレビジョンにおけるディジタル処理技術”，テレビ誌，33，4（1979）
- 8) J. Donald Harris：“Pitch Discrimination”，J. of The Acoustical Society of America，24，11（1952）
- 9) 宮坂栄一：“過渡音の聞え方”，NHK 技研月報，5（1981）

テレビジョン学会誌 10 月号「高品位テレビジョン —— 高精細度ワイドスクリーンテレビジョン」 予定目次

【ふ お ー か す】	高品位テレビジョンと高精細度テレビジョン	シャープ	林 宏三
【特 集】	1. 高品位テレビジョンの開発とその将来	NHK	藤尾 孝
	2. 高品位テレビジョンの画質	NHK	三橋哲雄・東京芸工大 畑田豊彦
	3. 信号方式と伝送		
	3-1 信号方式	NHK	石田順一・二宮佑一
	3-2 伝 送	NHK	石田順一・二宮佑一・王丸謙治
	4. 撮像管とカメラ		
	4-1 撮像管	NHK	磯崎幸直
	4-2 カメラ	NHK	熊田純二・池上通信機 川村好永
	5. VTR	NHK	吉田 武・片山浩誠・ソニー 谷村 洋
	6. フィルム送像とフィルム録画		
	6-1 フィルム送像	NHK	石田武久
	6-2 フィルム録画	NHK	元木紀雄
	7. ディスプレイ装置	松下電器	竹田 実・鈴木有信・平 英一
	8. 内外の研究開発動向	日立	阿知葉征彦・NHK 西澤台次