

マルチメディアにおける同期



池田 宏明 由雄 淳一

マルチメディア, 同期, リップシンク, AV 品質, 国際標準化

1. はじめに

マルチメディアにおけるメディア間同期の問題は、古くはアナログメディアであったテレビジョンの音声と映像の同期が検討され^{(1)~(3)}、国際電気通信諮問委員会の勧告が発行され⁽⁴⁾、関連機器に対する勧告もある⁽⁵⁾。インターネットに構築された仮想情報基盤である World Wide Web で交換されるデジタル形式のマルチメディアの同期に関しては、XML^{*1}に準拠したタグ言語 SMIL^{*2}が同時期に開発され、その維持が行われている⁽⁶⁾。また、インターネットプロトコル (IP:^{*3}) によるオーディオ・ビジュアル情報のウェブキャストにおける同期の研究も開始された⁽⁷⁾。このような背景を踏まえ、本稿では最近話題になっているデジタルオーディオ・ビジュアル機器に焦点を絞

り問題点や最近の活動を概観する。

オーディオ、ビジュアル機器、放送技術、そしてメディアは従来アナログ技術で構成されていたが、現在では多くがデジタル技術によって構成され、またデジタル化によって新たな技術が導入された。例えば地上デジタル放送やすでに一般的な DVD^{*4} ビデオ、そしてフラットパネルディスプレイやマルチチャンネルオーディオなどは、近年身近になってきたものである。これらによってユーザはしばしば放送やビデオの視聴時に、その環境によっては画と音の同期が合っていない、いわゆるリップシンクが合っていない状況を経験するようになってきた。本稿ではこの状況を解決すべく開始された国際レベルでの活動^{(8)~(10)}と我が国を含む業界での活動について概説する。

2. マルチメディアの同期の問題

まず、この問題は大きく分類すればオーディオやビデオ、そのほかの情報のいわゆるマルチメディアの再生の質に関するものと認識されている。これらマルチメディアが本来持つ能力を引き出し、ユーザに十分な再生の品質を提供するためには、元のコンテンツから最終ユーザまでを総合的に見て存在する問題点を解決する必要がある、そのための活動の一部がこのマルチメディアの同期問題の解決と言える。

次に、この問題には以下の要素がある。図1はコンテンツが家庭のユーザに使われるまでを示す全体のシステムを示してしているが、それぞれ以下の部分から成る。

(1) コンテンツ

提供されるコンテンツが制作的内容を含めて意図した同期は、そのままユーザ側で再生されてこそ楽しむことができる。制作意図による同期とは、例えば花火大会を収録した放送を例に取ると分かりやすい。遠景からの収録であれば音はその距離分遅れて来るのが自然であるが、映像が望遠によって花火をクローズアップしている場合は、その自然な音の遅れはかえって不自然に感じてしまう、そこで制作意図として音の遅れを調整することがある。また元のコ

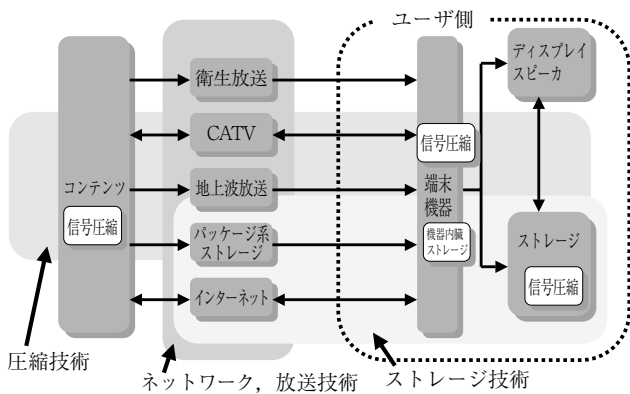


図1 全体システム図

いけだ・ひろあき 1966年千葉大学工学部電気工学科卒業、1968年同大学大学院工学研究科修了。工学博士(東京工業大学)。現職は千葉大学工学部教授。マルチメディアシステム、カラーマネジメント、感性のモデリングの教育研究に従事。1999年にIECロードケルビン賞受賞。

よしお・じゅんいち 1978年日本大学理工学部卒業、パイオニア株式会社入社後CD、DVDの開発を行い同時に、IEC、JEITA、AES、1394TAなどでデジタルインタフェースなどの規格策定に従事。現在同研究所勤務。

コンテンツにも意図と異なる同期ずれが機器などによって発生し得るが、これはユーザ側では意図されたものか否かは判別ができない。また、コンテンツは現在ほとんどデジタル化され伝送路やストレージの負荷を軽減するため、ビデオでは MPEG^{*5} などオーディオでは多様な圧縮方式で圧縮されて提供される。

(2) ネットワーク、放送技術、パッケージメディア

これらはコンテンツ内容を正確に伝える必要がある。放送や伝送では送出点においてオーディオとビデオの同期は取られていることが前提であるが、放送局間での伝送やまたオーディオとビデオでそれぞれ伝送路が異なる場合においては、コンテンツとしてある程度の同期ずれの可能性もある。さらに伝送方式によって独自に信号圧縮を用いる場合があり、例えばインターネットでの配信ではそれに適した方式をも用いる。またパッケージメディア、例えば DVD ビデオなどでの供給ではコンテンツは、MPEG によってその同期情報とともにディスクに記録されているので、基本的にパッケージ内では同期はとられている。

(3) ユーザ側における再生

ここではまず伝送、メディアの受け側として各種端末機器がある。これら機器は圧縮された信号を伸張しオーディオとビデオの再生信号の同期を取って出力をするが、すべての出力において同期を取ることは簡単ではない。これら機器はアナログ再生出力のほかにデジタルインタフェースによる出力も持っており、これらすべての出力の同期を取るには設計上の制約や、また実現するためにコストも必要である。ユーザは端末機器の出力をレコーダによってストレージに記録することもできるが、このレコーダも一般に信号を圧縮、伸張して記録、再生するため同期ずれの要素が加わる。そして最後にディスプレイとアンプによってオーディオ、ビデオ信号は再生されるが、最近のフラットパネルディスプレイでは、パネルを駆動するまでに各種デジタル信号処理が入っており信号遅延がある。

また、個々の機器ではそれ自体でのオーディオとビデオの同期はおおむね取られていても、夫々を組み合わせるシステムではその接続の状況によっては無視できない同期のずれを発生してしまう可能性がある。このようにもとのコンテンツが最終的にユーザによって再生されるまでには今日のデジタル化されたシステムにおいては多くの同期ずれを生じ得る要素がある。

以上のように、最終的にユーザはオーディオとビデオ信号の同期ずれを経験することがあるが、この現象は特に人の口の動きとその音声がずれることを指してリップシンク (Lip-sync) のずれと呼ばれている。そしてこの同期ずれが生じている場合、ユーザがそれをどの程度であれば検知

できるか、また検知できる場合どのように評価するかがこの同期ずれの評価基準として必要であり、これについて検討が行われてきた。

3. これまでの経緯

アナログテレビジョン放送業界の要請に応じて、国際電気通信機構 (ITU) はリップシンクに関する研究題目を設定した。我が国 (NHK)、スイス、オーストリア主導で研究が行われ、リップシンクの主観評価の結果が ITU-R^{*6} 勧告 BT.1359-1⁽⁴⁾ となった。これを図 2 に示す。横軸は映像の遅れ時間 t (ms)、縦軸は表 1 に示した主観評価尺度の差分 d である。各直線は式 (1) と書ける。

$$d = 160 (t + 95), \quad -200 \leq t \leq -95$$

$$d = 0, \quad -95 \leq t \leq 22.5 \quad (1)$$

$$d = -145 (t - 22.5), \quad 22.5 \leq t \leq 100$$

このように、音声先行の違和感に対する感度は、音声遅れに対する感度より高く、音声先行の時間許容が小さいことが分かる。音声遅れを負として、 $[A, A'] = [-185\text{ms}, +90\text{ms}]$ を許容限界、 $[B, B'] = [-125\text{ms}, +45\text{ms}]$ を検知限界、 $[C, C'] = [-95\text{ms}, +22.5\text{ms}]$ を検知不能範囲としている。

これに関連して、音声や映像信号処理装置に依存する固有のスループット遅延時間を明示すべしとの勧告が同時になされた⁽⁵⁾。

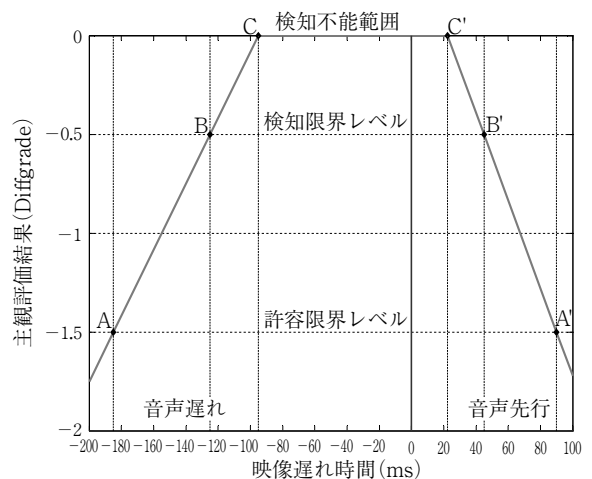


図 2 音声遅延許容度の勧告 ITU-R 1359-1

表 1 リップシンクの主観評価尺度

5	同期ずれが検知できない
4	同期ずれが検知できるが、気にするほどではない
3	同期ずれが多少気になる
2	同期ずれが気になる
1	同期ずれが大いに気になる

その後、図1に示したように音声や映像がデジタル信号となり、プログラム交換や再編集技術の変化、および最終的にそれを視聴者に物理的に提示する機器の技術の変化を踏まえ、図2の再検討が再度ITUの研究課題となっており⁽⁸⁾、これに対応してアメリカ⁽¹¹⁾やイギリス⁽⁹⁾で活動を開始した。また、IPネットワークを使ったプログラム配布における要求仕様が、リップシンクを含めて新しい研究課題となった⁽⁷⁾。

同時に、国際電気標準会議 (IEC: *¹³) においても、オーディオ・ビデオ・マルチメディア分野の国際標準開発を担う TC 100*⁷ における長期戦略グループ (AGS: *⁸) においても、マルチメディア品質保持の観点から、音声・映像遅延の問題の検討を開始した⁽¹⁰⁾。我が国でも次項に述べる活動が、IECを舞台に開始された。

4. 解決への取り組み

このオーディオとビデオの同期の問題は、多くの関係する業界で認識されてきたが、近年機器やシステムのデジタル化が進むにつれて、この問題の顕在化が起きてきている。このため各業界ではこの問題解決への動きが同時発生的に生じている。放送やプロオーディオなどの業務用を扱う団体、そして民生機器を扱う団体が連携していく取り組みも始まった。ここでは筆者らが担当しているIECでの取り組みと、その活動に対応する国内の活動である

JEITA*⁹での取り組みを紹介する。

現在、IEC/TC 100のプロジェクトIEC 62312, Guideline for synchronization of audio and videoでこの問題解決のためにプロジェクトチームPT 62312が活動している。プロジェクトIEC 62312は3部からなり、それぞれ、第1部は機器単体、第2部はシステムでの解決法、第3部は同期の測定法と評価法となっている。

この活動はほかの規格団体や組織とも連携または情報交換して活動しているが、図3にその様子を示す。IECに対応する国内規格団体のJEITAでは対応してAVリップシンク検討PG (プロジェクトグループ)が発足し、ここでIECへの対応と国内問題の解決を検討している。またプロオーディオや放送の団体とも連絡を取って進め始めており、ここに示した以外の関係する組織とも情報を交換し図1に示したシステム全体での問題の解決を図るべく進めている。

従来のアナログの時代から、例えば放送とそれを受ける民生機器はそれぞれ独立してこの同期問題を見てきた。すなわち民生機器から見ると放送の同期は合っているはずであり、逆に放送から見れば民生機器の再生では同期は合っていないはずであった。しかし今日デジタル化によってこの前提は変わり放送と民生機器はともに問題を解決するように協力することが必要である。

4.1 プロジェクトIEC 62312

第1部の機器単体は、個々の機器内での同期についてのガイドラインである。まず同期の要素、例えばデジタルインタフェースなどを含む一般化した機器モデルを明確にする。次に特定の機器において、例えばDVDプレーヤはオーディオとビデオについてアナログ、デジタルの各種出力を持つが、それらの同期や遅延量について指標を示すことを目指す。

第2部システムでの解決法は機器単体が組み合わせられて使用されるユーザの実際の環境での解決法を示す。例えば、DVDプレーヤのオーディオ、ビデオ出力はアナログ出力またはデジタルインタフェースによるデジタル出力によりTVまたはモニタに、さらにオーディオアンプに接続される。ここでは機器間の接続方法は多様であり、ま

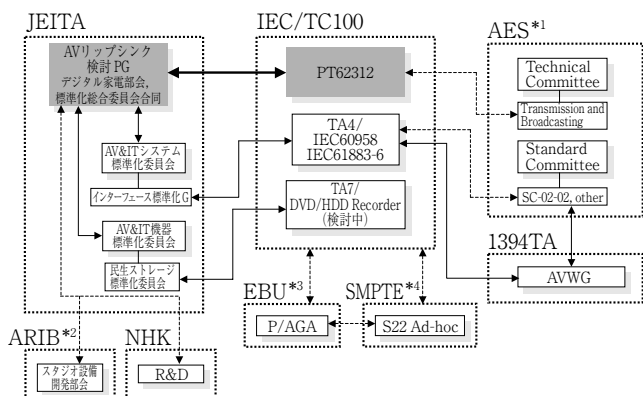


図3 IECでの取り組みと各組織との関係

*1 (Extensible Markup Language)
 *2 (Synchronized Multimedia Integration Language)
 *3 (Internet Protocol)
 *4 (Digital Versatile Discs)
 *5 (Moving Picture Experts Group)
 *6 (International Telecommunications Union - Radiocommunications)
 *7 (Technical Committee 100 on Audio, Video and Multimedia Systems and Equipment)
 *8 (Advisory Group on Strategy)

*9 (Japan Electronics and Information Technology Industries Association, 社団法人電子情報技術産業協会)
 *10 (Audio Engineering Society)
 *11 (Association of Radio Industries and Businesses, 社団法人電波産業会)
 *12 (European Broadcasting Union)
 *13 (International Electrotechnical Commission)
 *14 (Society of Motion Picture Television Engineers)
 *15 (Pulse Coded Modulation)

た接続される機器も複数ある。さらにそれぞれの機器は出力によって異なる遅延量を持ち、また受け側の機器ではユーザーへの最終出力も遅延量を持つ、これは例えばモニターに映像が表示されるまでの時間を指す。

これらを解決するために、この第2部では全体のオーディオとビデオの遅延量を合わせる方法を規定する。一つはタイムコードの伝送による方法、ほかは遅延量情報の伝送による方法である。これら情報の機器間の伝送のためには、デジタルインタフェースである IEC 規格を用いる。すなわち、リニア PCM^{*15} データ伝送のデジタルオーディオインタフェース規格の IEC 60958、そしてその圧縮オーディオデータ伝送規格の IEC 61937、また IEEE 1394 のオーディオ伝送規格である IEC 61883-6 がある。IEC 61883-6 は、また IEC 60958 をそのまま伝送可能である。これら規格は上記タイムコード、遅延量情報の伝送を行う内容を含めて現在改定中である。

第3部はリップシンクの測定法と評価法である⁽¹⁰⁾。可能な限り実使用に対応した各種形式での試験用標準コンテンツを規定し、それを用いてオーバーオールでの映像とそれに付随する音声の相対遅延時間を測定する方法を与える。また、一般利用者が入手可能な AV メディアでのコンテンツのリップシンクの程度を計測する方法論も含める予定である。なお、物理的に計測された映像遅延時間は、最終的には、図2に示した限界値を勧告して、視聴者の主観値に換算して評価すべきである。この観点から、アナログテレビジョンの時代に発効した ITU-R 勧告⁽⁴⁾である図2を再検討し、それも含めることになろう。これは AV メディア再生システムの品質評価につながる。

4.2 JEITA リップシンク検討 PG

JEITA は日本国内の電子機械工業を発展すべく活動を行っており、日本での規格化また IEC への日本としての対応を行っている。この IEC 62312 に対応してリップシンク検討 PG が発足し、JEITA 内の関連する委員会からの参加によって検討が開始された。

この目的は IEC 62312 規格への寄与とともに日本国内での問題を解決することである。JEITA では、この PG より先にテレビネットワーク事業委員会においてディスプレイの画と再生音声のずれであるリップシンクの測定に関するガイドラインを規定するなど、個々の分野での対応をして

きているが、JEITA の活動目的に則してシステム全体での同期問題の解決を横断的 PG によって目指す。

5. おわりに

近年ユーザーのマルチメディア環境は大きく変化しつつある。DVD ビデオはすでに普及し、大型フラットパネルによるテレビ、ディスプレイの人氣が高まり、また地上デジタル放送への移行が間近になった、さらにネットワーク配信もビジネス段階になった。このような環境でユーザーに供給されるマルチメディア再生の品質を保つことは重要であり、本稿の同期問題への取り組みはその一部分である。このためには個々の機器や放送などにとどまらず、全システムを通じた解決が求められるとともに、最終的に対象となる人間のユーザーの特性の検討も重要である。

本稿がそれに向けた関連者の活動について、何らかの情報となれば幸いである。

謝辞：本稿執筆の機会を与えていただいた、富士通研究所 村上敬一氏に感謝する。

(平成 18 年 2 月 10 日受付)

文 献

- (1) 黒住・赤井田・岡田：「テレビの映像と音声の相対時間差に関する検討」, テレビジョン学会技術報告, **20**, 50, pp.37-42 (1996-10~03)
- (2) 黒住・赤井田・岡田：「テレビの映像と音声のリップシンクに関する検討」, AES 東京コンベンション' 97 予稿集, pp. 100-103 (1997)
- (3) 黒住・赤井田：「テレビの映像と音声の相対時間差に関する検討(その3) - 刺激に白信号と純音を用いた場合 -」, 日本音響学会講演論文集 3-8-1, pp.373-374 (1998)
- (4) Recommendation ITU-R BT.1359-1, Relative Timing of sound and vision for broadcasting, Rec. ITU-R BT.1359-1 (1998)
- (5) Recommendation ITU-R BT. 1377, Labelling of video and audio apparatus throughput (processing) delay (1998)
- (6) W3C Recommendation SMIL 2.1, Synchronized Multimedia Integration Language (2005-12)
- (7) Question ITU-T 11/9 - Requirements and methods for sound and television transmission over IP networks "webcasting" (Study period: 2005-2008)
- (8) Question ITU-R 68/6 - Synchronization necessary for the satisfactory reception of sound and picture signals (2003-10)
- (9) EBU Technical Recommendation R37-2002: The relative timing of the sound and vision components of a television signal (2005)
- (10) 100/AGS (Secretariat) 193, Multimedia quality - Measurement and assessment of synchronization of audio and video, Advisory Group on Strategy of International Electrotechnical Commission (IEC) (2005-09)