4J−3

協調型ドラマシーン理解システムによるシーン・カット・音声の対応付け（理論）

柳沼良知　影山誠　坂内正夫
東京大学生産技術研究所

1　はじめに

映像、音声といったマルチメディア情報がワークステーション上で容易に扱えるようになるにつれて、マルチメディア情報の認識、加工、利用に対するニーズが高まりつつある。そのような試みの1つとして、現在、映像・シナリオ・音声といった、対応が必ずしも明確でないメディア情報を統合し、ドラマシーンのより高度な認識を行なうシステムの開発を行なっている[1]。本システムの特徴としては、メディアの種類や数に依存しない協調システムを実現するため、メディアに依存しない「共通概念」を用いて協調を行なう点が挙げられる。本稿では、「共通概念」によって、映像・シナリオ・音声の対応づけ（同期）を行なう方法について述べる。

2　映像・シナリオ・音声の対応づけ

メディア同士を協調させ、認識を行なう場合、その前段階として、あるメディアの解析しようとする部分が、他のメディアのどの部分に対応するかを、まず、知る必要がある（図1）。例えば、ある映像を解析する場合、その映像と対応するのは、シナリオ上でどの部分かを知ることによって、初めて映像とシナリオを用いた認識が可能になる。映像は、時間情報を持つメディアであり、1つの画面が決まれば、その時間は、一意に定まる。一方、シナリオについては、シナリオのある部分を指定してもその時間は定まらない。そのため、そのままでは、映像とシナリオの対応をとることは難しい。このような問題を解決するため、提案システムでは、各メディアからメディアの種類に依存しない「共通概念」を抽出し、この「共通概念」を用いて協調を行なう。各メディアには、概念を抽出するためのアナライザが取り付けられており、「時間」、「人数」等の概念が抽出される。抽出された概念はスーパバイザによって協調が行なわれる。

ここでは、「共通概念」を映像、音声、シナリオから抽出し、その協調により、3つのメディアの対応づけ（同期）を実現する方法について述べる。今回、対応付けに用いた4つの共通概念は、

- 時間
- 人数
- ABAB（カメラ切替えによる2人会話）
- 女性の存在

である。

これらの「共通概念」を用いた対応づけは、以下のように行なう。（ただし、映像、音声は、時間情報を持っており、対応づけが容易であるため、以下では、「映像・時間」とシナリオの対応づけについて述べる。）

時間の対応

映像・音声）　絶対時間が分かるため、そのカットの開始時間、終了時間を抽出する。

シナリオ）　まず、シーン中の文字数に比例して、時間を比例配分し、そのシーンのおおまかな開始時間、終了時間を求める。この時間には、誤差が含まれているため、誤差の許容範囲を定め、開始時間-誤差時間、終了時間+誤差時間を、改めて、それぞれ、開始時間、終了時間とする。

スーパバイザでは、上記の時間により、大まかな映像・音声とシナリオの対応づけを行なう。

人数の対応

映像・音声）　映像から人間らしいオブジェクトを抽出し、その人数を求める。

シナリオ）　シナリオの配役表には、出演者がどのシーンに登場するかが記述されている。これをもと

Matching between scene, cuts and sound using multimedia cooperative drama scene recognition system (theory)
Yoshitomo Yaginuma, Makoto Kageyama, Masao Sakauchi
Institute of Industrial Science, University of Tokyo

に、そのシーンに登場する人数を求める。
スーパバイザでは、両者から得られた人数が矛盾のないように、映像・音声とシナリオの対応づけを行なう。(シナリオに名前があっても実際の映像上に映っていない人物がいるため、今回は、「映像中の人数≤シナリオ中の人物」という仮定をし、人数の対応を行なっている。)

ABAB（カメラ切替えによる2人会話）
ABABとは、ドラマシーンに特有な、カメラ切替えによる2人の会話シーンのパターンである。まず、1人で映っているAさんが話し、次に、カメラが切替えられ、Bさんが話す。これが、2回繰り返された場合、ABABとする。この会話パターンは、映像、シナリオから容易に抽出することができる。
映像・音声)　連続するカットにおいて、1番目のカットと3番目のカットが似ていて、2番目と4番目のカットが似ていれば、その部分をABABとする。
シナリオ)　シナリオ中で、台詞の前の名前が1番目と3番目が等しく2番目と4番目が等しい部分があれば、その部分をABABとする。
スーパバイザでは、ABABの位置が合うように、映像・音声とシナリオの対応づけを行なう。

女性の存在
映像・音声)　音声から女性が存在する部分を抽出する。(今回は、基本ピッチの抽出を行ない、基本ピッチが高い部分を女性が存在する部分としている。)
シナリオ)　登場人物の中に女性の名前があれば、女性が存在する部分の候補とする。
スーパバイザでは、女性が存在する位置が合うように、映像・音声とシナリオの対応づけを行なう。

3　おわりに

メディアに依存しない「共通概念」によって、映像・シナリオ・音声の対応づけ（同期）を行なう方法について述べた。今後は、主人公の抽出等、より高度なドラマシーンの認識、理解について検討する予定である。

参考文献

[1] 影山 誠, 柳沼 良知, 坂内 正夫: “マルチメディア協調型ドラマシーン理解システムによるカットとシーンの対応付け”, 1992年電子情報通信学会秋期大会, (1992)

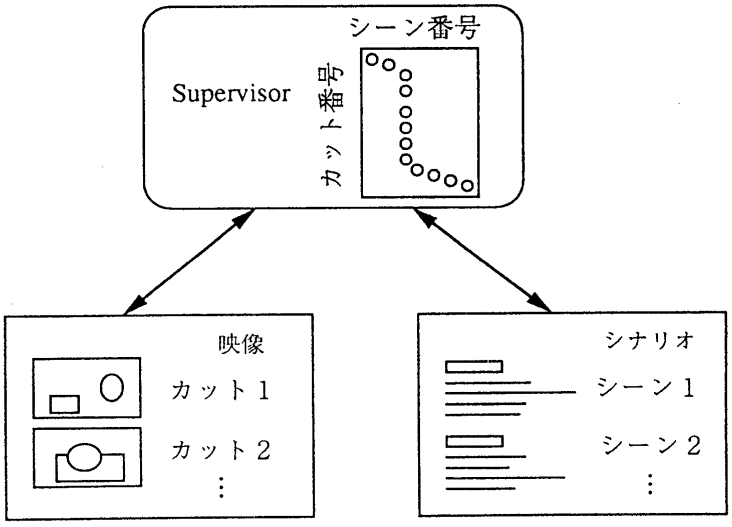

図1　メディア同士の対応付け