

An empirical study on the way to create effective audio-visual information using telops and sound effects

金, 基弘

<https://doi.org/10.15017/459597>

出版情報 : 九州大学, 2006, 博士 (芸術工学), 課程博士
バージョン :
権利関係 :

第1章 序論

1.1 はじめに

一般に、テレビや映画などの映像/情報メディアは、映像だけでは成り立たない。必ずといっていいほど「音」が存在する。音のない映像は、もはや機械的なトラブルを連想させてしまうほどであり、音が組み合わせられることによって、はじめて映像は成り立つのである。しかし、我々は映像の中の音の存在に無意識になっているようである。映像メディアを単に「みる」といって、普段「視聴」するとはいわない。この様に、映像メディアは視覚優位になりがちであり、音の存在を意識しないことも多い。本論文は、映像メディアにおいて軽視されがちな音をもたらす効果について取り組んだものである。

心理学の大きなテーマの一つは、人が自然界の事象をどの様に知覚し、情報処理を行っているのかである。自然界では、視覚事象と聴覚事象が同時に起きることはごく一般的な現象である。例えば、物が転がると「ゴロゴロ」と音がするし、風が吹くと木々が「ザワザワ」と音を立てる。我々は、発達の過程の中で自然に「視聴覚情報統合」を身につけているのかもしれない。小さい子供は自分のおもちゃを動かしながらそれらしい音（効果音）を口にする。

一方、人工環境においても我々は自然界と擬似した感覚モダリティ間の情報統合を経験することができる。Cohen (2005) は、自然界と人工環境での多感覚間の情報統合には大きな違いがあるとした。つまり、自然界では視聴覚情報の同時発生は、物理法則に従属されるが、人工環境では映像にどんな音でも自由に組み合わせることができるという。しかし、映像にどの音を組み合わせても良くなるわけではない。映像に合う特定の音を組み合わせたときのみ効果的になり、そこには興味深い心理法則が存在する。

それでは、映像メディアのように人工的に視覚情報と聴覚情報を構成する場合、どの様に音と映像を組み合わせればより有効な視聴覚情報になるのか。この疑問を解くためには、人の心内で視覚情報と聴覚情報がどの様に統合されるのかを考慮しなければならない。近年、心理学の分野では「音と映像の主観的調和」に影響する要因を調べることでこの問題に取り組んでおり、主観的調和には心内で視覚情報と聴覚情報の統合の度合いが反映されるという知見を得た。本論文では、マルチメディア文脈におけるテキストと音果音に焦点を当て、人が視覚と聴覚から構成されるマルチモーダル情報をどの様に

統合するのかを実験心理学的な手法を用いて検討する。

1.2 先行研究について

本節では、映画音楽の機能や役割に関する理論的/実証的な考察、視覚と聴覚の感覚特性に基づく視聴覚情報構成の基礎的な研究、映像メディアにおける音と映像の相互作用や主観的調和に関する研究などを概観し、本論文の背景および位置づけを行うと同時に問題設定を行う。

1.2.1 映画における音楽の機能と役割—理論的/実証的アプローチ

文明社会を通して音楽は、詩や舞踊、演劇、儀式などで使われてきた。歴史を遡ると、音楽は古代ギリシア劇、中世時代のオペラやバレエなどで表現の様式として使われた (Cohen, 1990)。1895年12月28日に無声映画が最初に公開されたときも、音楽は写真機からの雑音を遮蔽するとともに画面内の動作 (action) を描写/説明する機能として存在した (London, 1936; Pauli, 1976)。今日の音楽は、新たに多様な映像メディアに使われており、その機能的側面が目立つようになった。本項では、映像の音として取り上げられることの多い音楽を中心に論を展開することにする。ただし、現代では音楽が示す範囲が広く、本論文で取り上げる「音楽」はミュージック・コンクレート (musique concrète: 具体音楽) のような現代音楽を除き、西洋芸術音楽 (いわゆるクラシック) の系列上のもののみ限定する。

映像メディアのはじまりは映画であり、映画の「音」において音楽は一番多く議論されている。その大部分は、映画音楽の機能と役割について理論的/実証的考察を行ったものである。実証的アプローチは、映画音楽に対する我々の心的処理 (mental processes) を明確にするための手掛かりを提供してくれる。実験より示された客観的な証拠によって、その認知的処理過程を効率的に理解することができる。一方、映画音楽に対するこうした実証的アプローチを憂慮する声もある。その多くは、統制された実験室での結果のみで映画音楽を解釈するには物足りないという。確かに、そういう見解も間違いではないと考えられるが、主観的所見による解釈のみでは曖昧すぎる点が多い。本章では、映画音楽の分析の普遍的な方法論は確立していないことを前提に、客観的な

立場から映画音楽に対する理論的/実証的な考察を行う。そうすることによって、各種の状況下における音と映像の関係の理解を総合的に深めることを可能にする。以下の理論的考察では映画理論書や概論書を中心に、実証的考察では音楽心理学の論文を中心に概観する。

1.2.1.1 理論的考察

映画監督であり理論家であった Eisenstein (1898-1948) は、恐らく映画論の中の最も重要な人である。ほとんどの映画理論書が彼を言及しており、今日の映画的思考に大きな影響を与えた。特に、戦艦ポチョムキン〈Броненосец (Потёмкин) (1925)〉を制作し、モンタージュ (montage: 映像の配列によって意味を創り出すこと) 理論に関するユニークな方法論を確立したことは有名である。さらに、発声映画 (talkies) の登場とともに、モンタージュ理論を単なる編集理論からマルチメディアの統合理論へ発展させた。映画における映像と音楽の構造的類似性を強調し、視聴覚のモンタージュにおいては映像と音楽の「共通感覚」が決定的な役割を担うと論証している。音楽は映像の単なる従属物ではなく、音楽には映像のショット (shot) のグループを結び付ける又は分割する機能があり、「垂直のモンタージュ (vertical montage)」と命名した。彼は、実践と理論の両方において映画表現の発展に大きく寄与した。

音楽学者である Lissa (1965) は、音楽映画を除き、一般的な映画音楽の機能を次のように四つに分類した。それぞれ説明的 (illustrative) 機能、表現手段の機能、観客導入の機能、映画解説の機能と呼ばれるものである。その中で、映像と関連する音楽の機能は、説明的機能と表現手段の機能である。説明的機能としての音楽は、動きを強調したり、実際のノイズを音楽的に様式化したり、空間を再現したりする。心理的経験の表現手段の機能としての音楽は、認知を表す記号であり、記憶の再現、想像の反映、感情の表現手段となる。Manvell and Huntley (1975) は、映画音楽に関与した数多くの音楽家たちの立場から映画音楽の歴史や作曲について検討した。映画音楽作曲家たちの事例を基に、映像に対する音楽の機能を吟味した。まずは、無声映画と音楽の関係について触れた後、音楽の作曲原理が発声映画 (sound film) にどの様に適用されたかを考察した。特に、映画音楽を「事實的 (realistic)」と「機能的 (functional)」に分類

し、「機能的映画音楽 (functional film music)」の効果に注目した。Boggs and Petrie (2000) は、現代アメリカ映画を用い、映画における音楽の重要性を、(i) 映像の情動的内容を強化する (ii) 運動感 (kinetic sense) と想像力を刺激する (iii) 映像のみで伝えきれない情動の表現や暗示を行うと述べた。なお、彼らは音と映像の構造的リズムと情動的反応の生成といった映画音楽の一般的機能に加え、次のような特殊機能があると述べた：(a) 映像の欠点や弱点を補完 (b) 台詞の劇的效果を強化 (c) 内面的物語を表現 (d) 時間と空間の感覚を提供 (e) ノスタルジア感を喚起 (f) 劇的テンションの形成やイベントを暗示 (g) 視覚イメージへ意味を付加 (h) 音楽を通じた性格描写 (i) 条件反射を誘発 (特定の状況ごとに様式化された音楽) (j) 「進行音楽 (*traveling music*)」(速い動作を描写する音楽) (k) 重要な場面転換を提供 (l) メインタイトルと同伴 (m) スコアの一部としての音楽的音 (音響効果との関係) (n) 内部モノローグとしての音楽 (o) 演出された動作のための音楽。特に、Chion (1985/1993) は、映画における映像と音楽の感情の相乗作用を、両者の特定の関係によって生じるものであると主張した。映像シーン (scene) の情緒的雰囲気音楽が位置づけられるには、一般に「強調」と「補足」の二通りがあると考えられるが、彼はもう少し細分化して三つに区別した。第一は、音楽が登場人物又はシーンの感情に直接に関与し、劇的雰囲気を造成して意味を強化することから、「感情移入音楽 (*musique empathique*)」と呼んだ。第二は、音楽が登場人物又はシーンの強烈な感情 (e. g., 死, 危機, 狂気, 衝撃) に対して無関係のよう (対照的) に作用し、異なる意味が加わることで劇的雰囲気を強化することから「非感情移入音楽 (*musique anempathique*)」と呼んだ。例えば、〈Hiroshima mon Amour (1959)〉で、原爆の犠牲者名簿の場面に無心に流れる音楽が挙げられる。最後に、「教育的対位法 (*contrepoint didactique*)」とは、音楽が補足的な概念や観念を表象するのに用いられた場合、要するに、音楽を読み取るべきもの、解釈すべきであるものとしてある場合を指す。例えば、〈Padre Padrone (1977)〉で、貧しいイタリアの田舎の不毛の風景に流れる *An der schönen, blauen Donau* (Strauss Johann Jr. 作曲) はある観念の表象となる。一般に、非感情移入音楽と教育的対位法は同一視されるが、前者が読み取りを仲介しない直接的な方法で感情を強化する反面、後者は音楽が感情を強化するためではなく、ある観念の理解を目的に使われる。なお、〈Le musique au cinéma (1995/2002)〉には、音楽芸術としての映画を論ずる彼の見解が集大成されている。

Giannetti (1995) は、映画における各形式要素がどのように意味を生成/伝達するのかに焦点を置き、その「言語体系 (language system)」の解明を通して映画コミュニケーションの可能性を示唆した。彼は、映画における音楽の機能について最近の映画を例に挙げながら検討した。オープニング・クレジットとはじまる音楽は、一種の序曲として映画全体の雰囲気やテーマを暗示する機能がある。また、あるジャンルの音楽は、場所、階級、種族などを暗示する。さらに、音楽は、劇的文脈において予期せぬ前兆として利用できるし、場面の感情変化の統制にも利用できる。音楽は、逆説的対照を付与したり、音楽的モチーフを利用して性格を描写したりする。最後に、しばしば音楽は台詞を強調する機能を果たす。この様に、映画音楽は多様な形式の意味伝達機能を持っており、その効果は映像と結びつくことによって一層高まる。こうした映画を美学的観点からではなく、メディアとして捉える動きは 1970 年代の映画研究からはじまった。Faulstich (1994) も、映画を音と映像が統合された視聴覚コミュニケーションとして捉え、メディアとしての映画論を展開した。今までの映画分析が不完全であったのは、「映画言語 (filmsprache)」における音楽の役割を軽視したことに原因があるとし、映画音楽の分析を軽んじる音楽学者たちを憂慮した。また、映画音楽の分析を妨げる原因として (i) 「映画言語 = 映像言語」のように、単純に音楽を映像に従属させる問題 (ii) 映画音楽のプロトコル (記録) の作成に関する問題 (iii) 映画音楽の基礎知識の不足に関する問題を指摘した。

一方、Bordwell and Thompson (1993) は、映像と音楽の形式的関係による映画美学の理解を論じている。映画における音を研究するためには、映画を「見る」のではなく「聴く」べきであると主張し、音を映像の付属物のように思う一般的風潮に対して懸念を持った。そして、映画音楽が持つ長所を次のように述べた。音楽は、映像と並んでもう一つの表現の様式を担う。例えば、無声映画時代にはオーケストラやピアノ伴奏などがこの役割を果たした。音楽は、具体的に映像を解釈する方法を担う。つまり、音楽によって同じ映像が違うように解釈されるのである。音楽は、画面内で特定の注意を誘導し、注視すべき対象を表す。音楽は、暗示を行い、期待を持たせる。無音は、劇的效果をもたらす。音楽は、豊富な創意的可能性を持つ。また、彼らは、映画における音と映像の関係を (i) リズム (ii) 忠実度 (fidelity) (iii) 空間 (iv) 時間次元に分けて検討した。(i) リズムは、映画の視覚リズムと聴覚リズムの同期、非同期、対比が関係す

る。(ii) 忠実度は、例えば犬の映像に犬の声のように、知覚される音と映像の一致又は不一致が関係する。(iii) 空間は、映像と音源の場所が関係し、画面内にある「内在音 (diegetic sound)」と画面外にある「外在音 (nondiegetic sound)」がある。(iv) 時間は、物語、プロット、画面持続 (鑑賞) の三種類の時間があり、物語と音の関係によって「同時音 (simultaneous sound)」と「非同時音 (nonsimultaneous sound)」となる。

Chion (1990/1994) は、映画における音と映像の関係について最も著名な理論家である (作曲家、監督、批評家、編集者でもある)。彼は、映画における音と映像の多様な現象を、音を中心に新たな概念や用語を用いて論じている。こうした試みは映画のみに留まらず、その範囲を他の映像メディア (テレビ、ビデオ・アート、ミュージック・ビデオなど) にまで広めた。彼によると、映画における音と映像の最も重要な関係は、「視聴覚錯覚の現象 (phenomenon of *audiovisual illusion*)」であり、それは音が映像を豊かにする情報的/表現的価値によってもたらされるという。なお、この付加価値 (*added value*) は、音と映像の「シンクレス」の状況下で作用する。シンクレス (*synchresis*: *synchronism* と *synthesis* の造語) とは、一定の聴覚現象と視覚現象の間で、それらが理性的論理とは無関係に同時に合って生じる「自発的結合」であると述べた。ランダムな聴覚事象と視覚事象の流れで、あるものはシンクレスによって一緒にゆくが、あるものはそうでない。それは完全に自動的でなく、意味に依存し、ゲシュタルト原理と文脈によって体制化される。シークエンス (*sequence*) は、音と映像の相互強調のパターンと良い形態 (*phenomena of good form*) によって自ら分節したものである。また、彼はシンクレスの効果がより顕著になったのが「視聴覚の同期点 (*point of synchronization*)」であり、一つのシークエンスの中で有意義な視聴覚の同期点が浮かび上がるには、大体にゲシュタルト原理に従うと考えた。特に、(i) 視聴覚流れの予想外の二重崩し (*double break*) (ii) 音と映像のシークエンスの終止形式 (iii) 物理的特性 (iv) 情緒的/意味論的特性によって強調されるという。

また、Chion (1985/1993) は、映像に対する音の基本的関係を調べることを目的に、図 1.1 のような「三等分の円 (*three zones of circle*)」を提案した。ここで、「フレーム外の音 (*offscreen sound*)」とは、画面の中でその音源は同時には見えないが、演じられている場面と同一の時間にあり、画面が示す空間に隣接する空間にあることには

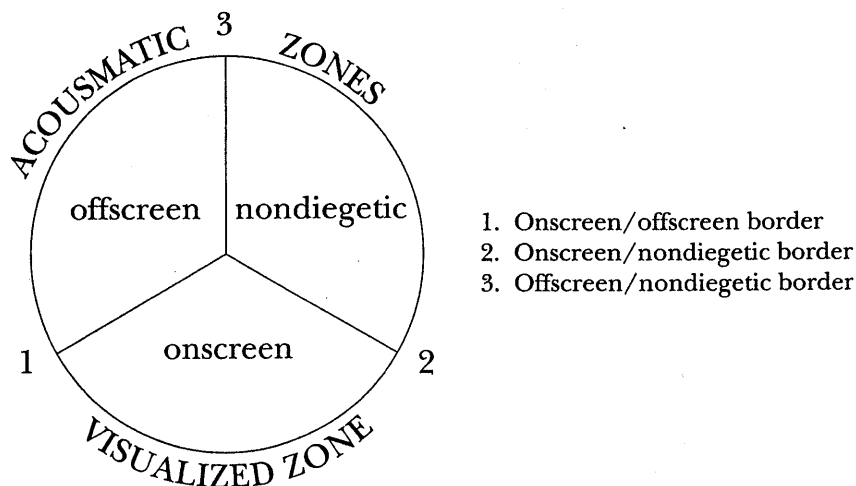


図 1.1 映像メディアにおける音の三つのゾーン (onscreen, offscreen, nondiegetic).

Chion (1990/1994) より転載

変わりがないと想像される音を指す。「オフ (*nondiegetic*) の音」とは、画面で示される場面とは別の時間と場所にある不可視の音源が発する音を指す。これらの二種類の音は、「アコースマティック (*acousmatic*: 音源が見られないまま聞くこと)」に属する。そして、「インの音 (*onscreen sound*)」とは、その場面でその音源が見える音 (*visualized sound*) を指す。これらの三つの概念は、それぞれ一つの円を三等分した領域を占めており、他の二つの領域とコミュニケーションを行う。実際の映像メディアにおいて、これらの三つの領域で説明不可能な音も存在する (e. g., <Look who's Talking (1989)> の赤ん坊の台詞など) が、映像メディアにおける音と映像の関係を考慮する際に非常に有用であり、その価値が落ちることはない。

Pauli (1976) は、映画における映像と音楽が結びつく場合 (i) 注釈的 (ii) 両極的 (iii) 対立的関係があると考えた。(i) 注釈的 (*paraphrasierend*) 関係とは、音楽の特定のキャラクターが映像の特定の内容と対応する、付加的な効果をいう。つまり、映像の内容から直接に誘導される音楽は注釈的關係になる。(ii) 両極的 (*polarisierend*) 関係とは、音楽の特定のキャラクターが映像の曖昧な内容を音楽の方へ移動させることをいう。(iii) 対立的 (*kontrapunktierend*) 関係とは、音楽の特定のキャラクターと映像の特定の内容が矛盾することをいう。さらに、音楽学者である Cook (1998) は、はじめて音楽、ことば、映像などの異なるメディアからマルチメディ

アを構成するための一般的な理論を立てた。彼は、音と映像の関係について、構成メディアが外延的 (denotative) と内包的 (connotative) 側面で果たす役割を考慮すべきであると主張した。また、異種のメディアが結び付く場合、図 1.2 に示すように (i) 一致 (conformance) (ii) 相補 (complementation) (iii) 競争 (contest) の三つの基本モデルがあると述べた。彼の説明によると、これらは構成メディア間の類似 (similarities) と相違 (differences) の二段階処理に基づくものであるという。最初の段階では、構成メディアが互いに一致しているかどうかの「類似性の判断 (similarity test)」が行われる。言い換えると、視覚モダリティと聴覚モダリティからの情報が同一であるかが問われるのである。そして、類似性がなくて両者の凝集性 (coherent) を決定するようになった場合は、次の段階の「相違性の判断 (difference test)」が行われる。その結果、構成メディア間の意味が矛盾すると競争となる反面、そうでない場合は相補となる。彼の理論的モデルは、異種のメディア間の相互関係を論じる際に非常に有効である。

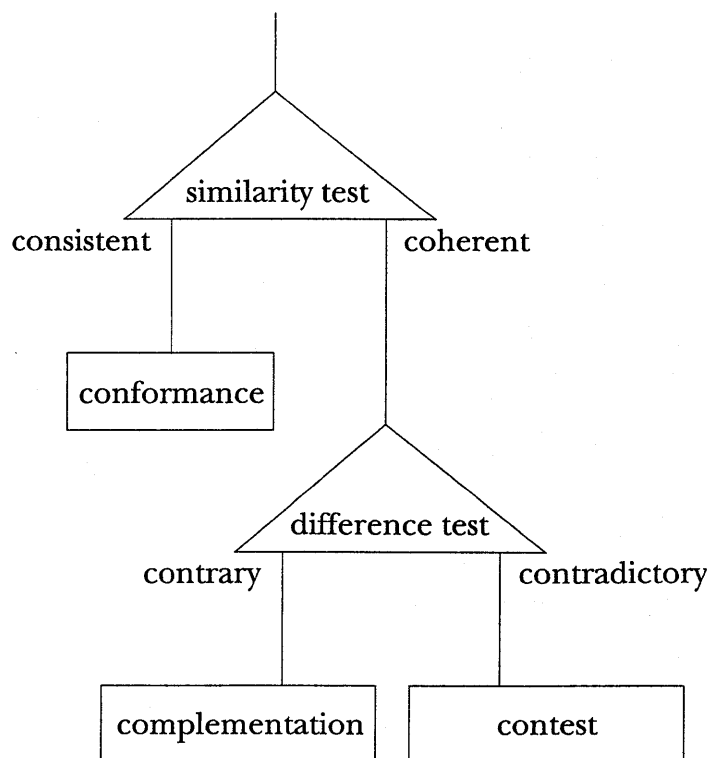


図 1.2 マルチメディア構成における三つの基本モデル。Cook (1998) より転載

以上、映画理論家、制作者、音楽学者などの、映像メディアにおける音と映像の関係や機能について検討した。古典的な映画論では音楽の機能を、主に (i) 意味を付加する (ii) 記憶を助ける (iii) 不信感を無くすことに重点を置いた。Manvell and Huntley (1975) によると、最良の映画音楽とは、観客が音楽の存在を意識することなく、多様な機能を果たすことであるという。従って、あまり素晴らしい音楽は、観客の注意を逸らすために望ましくないと思われた。こうした保守的/伝統的な見解に対し、Boggs and Petrie (2000) の現代的な見解は、音楽への観客の意識的注意を認めるとともに、音楽が映像の時間的、物語的、リズム的要素と不可分の関係を持っているかぎり、音楽が映像を支配することも可能であるという立場である。映画音楽に対する伝統的な立場であれ現代的立場であれ、要は度を越した注意を傾けさせて映画全体の効果を崩す音楽は良好な存在ではない。良い映画音楽とは、その自体が目的でなく、映画全体において意味のある構成要素として、それが「効果的」か「適切」かによると考えられる。

また、音楽は映像と同等に映画の意味を構成しており、決して映像に従属されたものではない。映画音楽は映像と有機的関係を持っており、音楽自体の分析よりは映像との相互作用の中でその効果を分析すべきであると考えられる。もちろん、本論文で映画音楽に関するすべてを網羅することは不可能である。それでも、映画における映像と音楽の間には、様々な心理的現象が働いており、心理学と深い関連を持つことがうかがえた。特に、音楽は我々の映画経験に対する情動の強化、記憶や解釈に影響する「力」を持つと考えられる。しかし、あくまでも推論的な立場からの解釈が多く、体系的で厳密な実験を通して実証されたわけではない。次の 1.2.1.2 項では、映画における音と映像の関係や機能に関する心理学者たちの見解を述べることにする。

1.2.1.2 実証的考察

Münsterberg (1916/2001) は、最初に映画論と心理学を関連付けて論じた心理学者である。彼は、映画は「心の言語 (language of the mind)」を話すユニークなメディアであると考えた。また、映画における音楽の役割について (i) テンションの解決 (ii) 興味の維持 (注意の持続) (iii) 安心感 (comfort) の提供 (iv) 情動の強化 (v) 美的経験への貢献であると主張した。心理学的手法を用いて実証はしていないが、発声映画が登場するにも前に映画における音楽の重要性を唱えたことは驚嘆に値する。

音楽心理学や音楽学の分野において映画音楽の研究は、軽視されてきた。しかし、1980年代後半から映像やデジタル技術の発展とともに、映画音楽の効果を調べた研究が増えてきた。特に、〈*Psychomusicology, 13, The psychology of film music* (1994)〉は、多方面からの映画音楽の心理学的研究の現状を概観している。こうした心理実験からの実証的な証拠は、音楽がどの様に映画の知覚/認知に役に立つかを教えてくれる。

Cohen (1999, 2001) は、映画やマルチメディア文脈における音楽の機能について最も活発に取り組んできた心理学者である。彼女は、映画音楽の機能を、(i) 不要な雑音を遮蔽する (ii) ショットとショットの間の連続性を与える (iii) 構造または連合(意味)的調和によって画面の重要部に注意の誘導する (iv) 特定の注意が働かない場合に雰囲気誘発する (v) 特に、多義的な状況において特定の意味や物語を伝える (vi) 記憶の連想性を用いて特定の対象を象徴化する (i. e., *leitmotiv*) (vii) 現実性や覚醒の水準を向上する (viii) 芸術形式として美的効果を与えると概観した。この8項目の中で、(iii) に関しては, Bolivar, Cohen, and Fentress (1994), Marshall and Cohen (1988) の研究などが、(v) に関しては, Bullerjahn and Guldenring (1994), Cohen (1993) の研究などが、(vi) に関しては, Boltz, Schulkind, and Kantra (1991), Boltz (2001) の研究などが深い関連を持つ。さらに、彼女 (Cohen, 2005) は、映画における映像と音楽の相互作用に関する実験心理学のゴールは、(i) 刺激として目と耳にエネルギーを送る映像素材と音楽素材の定量化 (ii) 単独と組み合わせ呈示における刺激の心的効果の定量化 (iii) 予想モデルを通して映像と音楽の組み合わせの効果を描写することでであると強調した。

岩宮 (2000) は、映像作品における映像と音楽の組み合わせの手法を言及した。彼は、音楽がかもしだす雰囲気や象徴的な意味、歌詞の内容などを利用して映像表現の効果を高める方策を例示している。そして、最も典型的な手法として映像の動きに合わせて音楽のメロディ・ラインやリズム・パターンを付ける「ミッキー・マウシング (Micky-mousing)」を挙げた。ミッキー・マウシングは、ウォルト・ディズニーのアニメーションのキャラクターから名づけられており、特にアニメーションにリアリティを与えるのに効果的である。ミッキー・マウシングは、古くはオペラ音楽 <e. g., Richard Wagner の *Das Rheingold*> やバレエ音楽 <e. g., Moris Ravel の *Daphnis et Chloe*> に由来するが、この手法を初期の発声映画からアニメーションまでに適用した作曲家は

Steiner (1888-1971) である。この様な音と映像の同期 (e. g., Lipscomb, 1996, 1999 ; Sugano & Iwamiya, 1998; 菅野, 岩宮 1999) は, 音楽が映像にアクセントを与えたり, 映像の動きが音楽の構造を鮮明にしたりする効果がある。

Lipscomb and Tolchinsky (2004, 2005) は, 映画における音楽コミュニケーションの役割について認知心理学的な立場から論じている。彼らは, 映画音楽の役割に関する古典的理論 (i. e., 物語の情動的側面の向上や補完) から離れ, 意味を伝達するのに役に立つ映画音楽の構成法を多角的に検討した。そして, 制作者や作曲家によって意図された音楽は, 情報の受け手に (i) 映画の一般的な雰囲気 (e. g., 恐怖, 不安など) (ii) 登場人物の描写 (e. g., 感情, 思考など) (iii) 明確な物語の構造を伝えると述べた。特に, (i) に関しては, 映画音楽の情動に関する Cohen (2001) の (iv) (v) (viii) と深い関係を持つ。この様に, 心理学の観点からも, 音楽は映像の知覚/認知に関する興味のある手掛かりを提供する。しかし, 未だに音と映像の相互作用について心理学的手法を用いた研究が少ないのが現状である。

1.2.2 視聴覚間の各種の相互作用—感覚モダリティから映像メディアまで

1.1.1 項では, 映画音楽の機能や役割に関する理論的/実証的な考察を行い, 映画理論家も心理学者も映像の音に対する共通した見解を持っていることを探った。本項では, 実験心理学的な手法を用いた視聴覚間の各種の相互作用について概観する。

1.2.2.1 視覚と聴覚の様相間の相互作用

従来, 心理学の分野では感覚系の情報処理の基礎として視覚と聴覚の相互作用の研究が行われた。特に, 映像メディア経験の観点から見ると, 視覚と聴覚の間の空間と時間の知覚に関する基礎的な知見は重要である。しかし, その詳細のすべてを述べることは本論文の範囲を超えることであり, ここでは簡略に紹介するに留める。

一般に, 視覚は聴覚に比べて空間情報の処理に優れているといわれており, 視覚刺激と聴覚刺激が同時に呈示された場合には, 視覚を基準系とした空間定位が行われるのがよく知られている。こうした感覚系の情報処理における視覚優位の現象は「腹話術効果 (ventriloquism)」ないしは「視覚的捕獲 (visual capture)」と呼ばれている。腹話

術効果（視覚的捕獲）は、「不整合検出 (discordance detection) 課題」(Bertelson & Radeau, 1981; Choe, Welch, Gilford, & Juola, 1975; Jack & Thurlow, 1973; Thurlow & Jack, 1973) と「選択的定位 (selective localization) 課題」(Bermant & Welch, 1976; Bertelson & Aschersleben, 1998; Klemm, 1909; Pick, Warren, & Hay, 1969; Radeau, 1992; Thurlow & Jack, 1973) で調べることができる。これらの研究は、聴覚と比較して空間的な領域での視覚優位の効果を示すものである。これらに対し、聴覚刺激によって視覚の空間定位が影響を受けるという研究もあるが、その例は非常に少なく効果も僅かである (Bertelson & Radeau, 1981; Radeau & Bertelson, 1987)。

Bertelson (1999) は、精神物理学的な手続きに基づき、腹話術効果に関する新たなパラダイムを提起した。彼は、被験者が視聴覚間のずれ (discrepancy) を意識していないときでも、音源定位に対する視覚バイアス（視覚の影響で音の方向がずれること）が生じることを示した。また、腹話術効果が意識的な空間的注意の移動によって説明できないことを証明した。従って、腹話術効果は無意識（純粹に知覚的）の様相間の組み合わせ (crossmodal pairing) の現象を反映すると結論づけた。つまり、様相間の知覚単位の構成は前意識処理 (pre-conscious processing) の段階で行われており、「意識的知覚融合 (conscious perceptual fusion)」とは明確な違いがあるという。彼は、Radeau and Bertelson (1977), Welch and Warren (1980) と同様に、様相間の相互作用 (cross-modal interaction) を感覚と認知のカテゴリーに分けて考察している。腹話術効果が情報処理のどのレベルで生じているのかに関してはいくつかの意見があるが、彼の研究はそれが感覚レベルで生じる現象であることを後押しするものである。彼の研究の意義や問題点については、Welch (1999a) の論評で詳細に述べられている。

視覚情報によって聴覚情報が影響を受けるもう一つの代表的な現象に、マガーク効果 (McGurk effect: McGurk & MacDonald, 1976) がある。例えば、/ba/ という音声と同時に /ga/ を発話する人の映像を呈示すると /da/ に聞こえてしまう。人は、視覚情報と聴覚情報を統合しながら言語コミュニケーションを行うのである。また、視覚情報によって触覚定位 (tactile location) が影響を受けるという研究もある (Rock & Victor, 1964)。

一方、聴覚は視覚に比べて時間情報の処理に優れているといわれている。このことから、視覚刺激と聴覚刺激が同時に呈示された場合に、聴覚を基準系とした時間情報の

処理が行われることが予想される。例えば, Welch, DuttonHurt, and Warren (1986) は, 反復的な系列を持つ聴覚刺激と視覚刺激の間で自然に生じるずれ感に関する一連の実験を行った。ユニモーダル条件では, 単独で呈示した 4 Hz の視覚フリッカーが 6 Hz くらいに知覚されたが, 単独で呈示した 4 Hz の聴覚フラッター (聴覚的ちらつき) は比較的正確にその反復周期が知覚できた。しかし, これらの刺激を対にしたバイモーダル条件の場合, 被験者の ME (magnitude estimates) 判断は一定の周期で呈示されているはずの視覚刺激を聴覚刺激のみの条件と同じ周期に誘引した。

菅野 (2005) は, 時間的なずれを含み一定の間隔で反復的に呈示される視聴覚刺激に対し, どちらかの一方のモダリティに合わせる同期タッピング課題を行った。実験の結果, 視聴覚刺激がほぼ同期した条件 ($SOA < 100$ ms) で視覚刺激に対する聴覚刺激の大きな誘引効果が認められた。さらに, その効果は視覚刺激が聴覚刺激に大きく先行する条件 ($SOA > 200$ ms) でもみられた。

下條, シャイア, ニジャワン, シャムズ, 神谷, 渡辺 (2001) は, 聴覚刺激による視知覚の変化に関する三つの新しい発見を報告した。第一に, 視覚の時間分解能は付随する音によって向上あるいは低下し, それは聴覚刺激と視覚刺激の系列と各刺激間の遅延に依存した。第二に, 単一の視覚フラッシュが, 複数のビーブ音が伴うと複数のフラッシュとして知覚されることがあった (*illusory flash effect*: Shams, Kamitani, & Shimojo, 2000)。第三に, 相互に近づくように動く多義的な運動パターンは, それと同期していない音 (あるいは音がなくても) が鳴っても二つの物体が交差しているように知覚されるが, 同期する音が伴えば衝突して跳ね返るように知覚された。聴覚情報が視覚の時間的な側面のみならず, その見え方や解釈にも影響を与えられるということは意義のある発見であるといえる。他に, 聴覚刺激によって知覚される視覚の強さ (i. e., 明るさ) が向上されるという感覚の感受性 (閾値) の変化も興味深い (Odgaard, Arieh, & Marks, 2003; Stein, London, Wilkinson, & Price, 1996)。

これらの研究は, 空間定位の場合とは対照的に, 時間的処理に関しては視覚と比較して聴覚の影響を受けやすいことを示唆するものである。この様な感覚系の情報処理における聴覚優位の現象は「聴覚的駆動 (*auditory driving*)」ないしは「聴覚的捕獲 (*auditory capture*)」と呼ばれている。Aschersleben and Bertelson (2003), Gebhard and Mowbray (1959), Myers, Cotton, and Hilp (1981), 長嶋 (2003), Shipley (1964)

などの研究でも視聴覚間の相互作用における聴覚優位の効果が示されている。

視覚と聴覚の様相間の相互作用に関する古典的な見解は、「人間は、視覚優位の動物である (Rock, 1966)」と思われているように、視覚を中心に他の感覚モダリティからの情報が統合されるというものであった。前述した視覚情報によって音源定位が影響を受ける腹話術効果や、視覚情報によって音韻知覚が影響を受けるマガーク効果はその代表的な例といえる。しかし、最近、聴覚的駆動 (聴覚的捕獲) といった視知覚に及ぼす聴覚の効果を扱った研究も新しく現れてきている。要するに、視聴覚間の相互作用において視覚は空間情報の処理に、聴覚は時間情報の処理により有効であるといえる。この様な観点から、Welch (1999b), Welch and Warren (1980) は「モダリティ適切性 (modality appropriateness) 仮説」ないしは「モダリティ精度 (modality precision) 仮説」を立てた。彼らは、マルチモーダル情報に係わる感覚モダリティの相対的な適切さ/信頼性によって様相間の相互作用の主体が決まるといえる。Shimojo and Shams (2001) も、精神物理学的研究の証拠を基に、様相間の相互作用に関する新たな提案をした。彼らは、従来信じられてきた視覚優位性に反対し、感覚モダリティ間の統合において入力モダリティにかかわらず、過渡的 (transient/discontinuous) 信号の存在が様相間の相互作用の方向を決めると述べた。人は、感覚モダリティが何であれ、最も強く正確な (多義性のない) 信号を伝達するモダリティを信頼して複数の感覚モダリティからの情報を統合するという。

以上の概観より、感覚モダリティ間の相互作用における視覚と聴覚の情報処理の特徴を理解することができた。しかし、その相互作用が人の情報処理経路のどのレベルで生じているのかは明らかではない。最近の脳イメージング研究では、感覚間の情報統合に貢献する要因 (e.g., 時間, 空間, 意味など) を解明しつつある。例えば, Bushara, Grafman, Hallett (2001) は PET (positron emission tomography) を用い、視覚刺激と聴覚刺激の間の「開始の非同期 (onset asynchrony)」の知覚が島皮質 (insular cortex) と、局所的に後頭頂葉 (posterior parietal), 前頭前野 (prefrontal), 小脳 (cerebellar) の領域の活動と深い関連を持つことを示した。また, Calvert, Hansen, Iversen, Brammer (2000) の fMRI (functional magnetic resonance imaging) を用いた研究では、時間的に同期した視聴覚刺激と同期していない視聴覚刺激の呈示の間に島皮質と上丘 (superior colliculus) に反応の増加と減衰が有意であった。これらの結果は、時

間的に特徴付けられた様相間の相互作用が比較的初期の処理レベルを介することを示唆するものと考えられる。さらに, Calvert, Brammer, Bullmore, Campbell, Iversen, and David (1999) は, 音声知覚においてユニモーダル条件と比較してバイモーダル条件の暴露後に視覚野 (V5) と聴覚野 (BA 41/42) の活動が相対的に増加し, その増加は視覚信号と聴覚信号との調和に依存することを明らかにした。この様に, 心理学の分野だけでなく生理学的な観点からも様相間の情報統合に関する考察が行われているが, 未だに不明な点も多い。

1.2.2.2 視覚と聴覚の様相間のマッチングと類似性

人は, 身の回りの情報を複数の感覚器から受容しており, 感覚モダリティ間の主観的バランスが崩れた場合には違和感が生じる。こうした感覚モダリティ間の不整合や非類似は, 知覚的/認知的な負荷として作用する可能性が高い。そのため, 感覚モダリティ間の知覚的/認知的な等価や類似性を調べることは有意義である。特に, 映像メディアのように人工的に感覚モダリティ (視覚と聴覚) を組み合わせる必要がある場合は, その重要性を増す。本項では, 感覚特性に基づくマルチモーダル情報構成の基礎として, 視覚と聴覚の「様相間マッチング (cross-modal matching)」と「様相間類似性 (cross-modal similarity)」について概観する。

様相間マッチングは, 感覚モダリティ間の知覚的等価 (perceptual equivalence) を調べるために行われる。例えば, 音の大きさを光の明るさにマッチさせるように, 一つの感覚領域における感覚の強さを他の感覚領域における他の感覚の強さにマッチさせる。こうした様相間マッチングによる感覚間の等価は, 感覚システム機能の本質的な特性を反映すると考えられる。視覚と聴覚の様相間マッチングに関する先行研究からは, 両者の間に「強さ (intensity)」という共通する知覚的要因があることを見いだした (J. C. Stevens & Marks, 1965; J. C. Stevens, Mack, & Stevens, 1960; Marks & Stevens, 1966; Mashhour & Hosman, 1968; Root & Ross, 1965; S. S. Stevens, 1959)。すなわち, 視覚のブライトネスが増加すればするほどマッチする音のラウドネスも増加するようになる。こうした視覚と聴覚の強さを正確にマッチングする能力は, 生後1ヶ月の乳児にも現れており (Lewkowicz & Turkewitz, 1980), ブライトネスとラウドネスは非常に整合性の高い感覚属性であると解釈できる。さらに, ブライトネスはピッチ (音の高さ)

との整合性も良いことが知られている (Marks, 1974, 1978)。

Marks (1982a) は、人の言語活動における視覚と聴覚の「共感覚的メタファー」に関する興味深い実験を行った。共感覚的メタファー (synesthetic metaphors) とは、例えば「the dawn comes up like thunder (Edgar Allan Poe)」のように、ある感覚モダリティが持つ意味が他の感覚モダリティへ転移されることである。彼は、共感覚的メタファーも基本的には視覚と聴覚の強さなどの知覚的要因が関連するのではないかという疑問を抱いた。そこで、被験者に視覚と聴覚のメタファーを含んだいくつかの詩句を読ませた後、視覚と聴覚に含意されたメタファーのレベルを白色光と 1 kHz の純音を用いてセットするように教示した。実験の結果、被験者はラウドネスとブライトネスを互いに比例してセットする傾向があり、全般的に詩より生じたラウドネスの感覚を光のブライトネスとマッチングした。この研究により、言語を媒介とした人の認知レベルでの共感覚的メタファーが様相間マッチングの影響を強く受けることが分かった。さらに、Marks (1982b) は、視覚的明るさを表す単語 (e. g., sunlight, moonlight) を音の高さと大きさの次元とメタファー的にマッチさせる実験を行ったところ、音高の方がラウドネスより共通性が高かった。

感覚モダリティ間の関係を規定するもう一つの方法に、様相間類似性がある。様相間類似性からは、異なる感覚モダリティにおける知覚的経験の質や次元を調べることができる。Marks (1989) は、どの様に聴覚のラウドネスと音高が視覚のブライトネスと対応するかについて詳しく調べた。彼は、四つの条件 (低音高/小ラウドネス, 高音高/小ラウドネス, 低音高/大ラウドネス, 高音高/大ラウドネス) ごとに合成した 16 種類 (4 水準の音高×4 水準のラウドネス) の純音を被験者に呈示し、常に呈示されている暗い光と明るい光のどちらに音刺激が類似しているかを判断させた。実験の結果、音高とラウドネスは両方とも様相間類似性に貢献したが、音高の方がラウドネスよりブライトネスとよく対応することが分かった。なお、音高とラウドネスの様相間類似性に対する貢献度には個人差が存在したり、刺激の文脈 (*simulus context*) に大きく依存したりした。さらに、先の実験で制限されていた視覚のブライトネスを暗いから明るいまで系統的に操作し (5 水準), 合成した 9 種類の純音 (3 水準の音高×3 水準のラウドネス) を加えた 14 種類の刺激に対して、すべて可能な対 196 通り (14×14) を作成して被験者に呈示し、ブライトネスとラウドネス/音高との対応の類似/非類似の度合いを評定さ

せた。多次元尺度構成法 (MDS) により、聴覚のラウドネスと音高が視覚のブライトネスと両方ともに共通した感覚属性であることが分かった。この結果は、感覚モダリティ間の「非一元的関係 (多義性)」を示唆するものと考えられる。

さらに, Marks (1987) は, 視覚と聴覚の感覚属性間の類似性を, 被験者のキー押しの反応時間を指標として検討した。視覚と聴覚の様相間の対応関係があると反応時間は速くなり, その逆の場合は反応時間が遅くなることが示されている (e. g., Bernstein & Edelman, 1971)。視覚刺激と聴覚刺激はそれぞれ異なる水準の感覚量を持っており, 様相間の対応関係を想定して「音高-ブライトネス」「音高-ライトネス」「ラウドネス-ブライトネス」「ラウドネス-ライトネス」「音高-形」の五つの視聴覚ペアに分類した。これらの視聴覚ペアごとに感覚量の類似した条件 (matching) と類似していない条件 (mismatching) を設けて選択反応課題を行った。その結果, (a) 暗い/明るい光と色は, 低い/高い音と対応する (b) 低い/高い音は, 暗い/明るい光と色に対応する (c) 暗い/明るい光 (色は例外) は, 小さい/大きい音と対応する。 (d) 丸い/鋭い形は, 低い/高い音と対応することが示された。これらの結果は, 視覚と聴覚の感覚属性間の類似性が知覚的処理を促進させる効果があることを示唆するものである。なお, Walker and Smith (1984) は, 50 Hz と 5500 Hz の聴覚刺激を画面の中央に現れる 4 単語の中でふさわしい 1 語に頼ってキーを押す実験を行った。その結果, 音の高さと呈示された課題語との感覚的属性が類似していないときは, 被験者の反応時間が遅くなり, 言語を媒介とした意味的処理レベルにおいても感覚属性間の類似性が認知的処理を促進させる効果があることが示唆された。

以上の視覚と聴覚の様相間のマッチングと類似性に関する研究から, 視覚次元のブライトネス, ライトネス, 形と聴覚次元の音高, ラウドネスの間に対応関係が存在することが示された。特に, 視覚のブライトネスは聴覚の音高とラウドネスの二つの次元とマッピングすることが裏付けられた。こうした感覚モダリティ間の対応関係が, 先天的なもの (J. C. Stevens & Marks, 1980; Zwillocki & Goodman, 1980) であるか, 後天的なもの (Mellers & Birnbaum, 1982) であるかは学者たちの間に議論がある。面白いことに, Marks, Hammeal, & Bornstein (1987) は, 約 500 名の子供と 100 名以上の大人を対象に視覚と聴覚の対応関係に関する実験を行った。その結果, 子供群も大人群も音高をブライトネスに, ラウドネスをブライトネスにマッチングでき, 非常に幼い子供さ

えも視覚と聴覚の間の知覚的類似性が認知できた。しかし、約 11 歳までの子供たちには、一貫して音高とサイズの間の知覚的類似性が認知できなかった。彼らはこの結果を、ブライトネスとラウドネス/音高との対応関係は本質的な知覚的特性によるものであり、形と音高の対応関係は学習によるものと解釈した。視覚と聴覚の様相間の対応関係には、絶対的/相対的な関係、またその対応に多様性があると考えられる。

1.2.2.3 映像メディアにおける音と映像の相互作用

映像作品における音楽の影響を心理学的手法で捉えようとした最初の試みは、Tannenbaum (1956) の研究である。彼は、Osgood, Suci, and Tannenbaum (1957) の三つの普遍因子に属する形容詞対を用い、35 分のテレビ・ドラマに音楽を付加した場合と付加しない場合の印象の相違を SD (semantic differential) 法によって調べた。実験の結果、音楽はドラマの力動性と活動性を強化する効果があった。ただし、ドラマの総合的評価に関しては、有意な変化はみられなかった。彼の研究は、音と映像の相互作用の研究において重要な位置を占めるものであるが、実験過程にいくつかの問題点もあった。例えば、音刺激が映像刺激のために専用作曲した音楽でなかったため、特定の場面と合わない可能性があった。また、蓄音機から音刺激を呈示しており、正確に映像刺激との同期がとれたとは言い難い。最後に、手動で音量を操作し、映像シーンとの劇的効果 (dramatic impact) を合わせたことも再現性の問題になる。

Marshall and Cohen (1988) は、2 分の抽象的なアニメーションに弱い印象と強い印象を持つように作曲したいくつかの単純なピアノ曲を組み合わせ、音楽が映画の「登場人物 (characters)」として呈示した三つの幾何学的な図形の心理的属性にどのような影響を与えるかを SD 法によって調べた。印象評定尺度は、Tannenbaum (1956) と同様に、Osgood et al. (1957) の三つの普遍因子に属する形容詞対を用いた。実験後、五つの条件 (映像のみ、弱い印象の音楽のみ、強い印象の音楽のみ、弱い印象の音楽と映像の組み合わせ、強い印象の音楽と映像の組み合わせ) を比較したところ、Tannenbaum (1956) の実験結果と同様の傾向が得られた。すなわち、音楽は映像の力動性と活動性には直接に影響を与えたが、評価に関しては対呈示された映像と音楽の間に明瞭な関係はなかった。その原因の一つは、呈示した音刺激が極端に単純/反復的な構造を持つからである。単純な刺激を用いることで基本的な現象を明らかにすることは容易になるが、芸術的な

表現の様式が極度に制限されることにもなる。従って、彼女らの実験結果を実際の映像メディアまでに一般化することには少し問題があると考えられる。それでも、長い間に音楽学者たちに軽視されてきた映画音楽の分析 (Faulstich, 1994) を、積極的に取り組んだことは高く評価される。その上、彼女らの論文は音楽知覚認知の分野において、音と映像の相互作用の研究を発展させるきっかけをつくった。

Sirius and Clarke (1994) は、映像と音楽の間に効果的な相互作用が存在するという事例証拠 (anecdotal evidence) を実証するために、Marshall and Cohen (1988) の実験の一部を修正して、幾何学な図形から構成した 4 種類の 3D 映像と、それに専用に作曲した 4 種類の音楽を用いて SD 法による印象評定実験を行った。印象評定尺度は、Tannenbaum (1956)、Marshall and Cohen (1988) の先行研究と同様に、Osgood et al. (1957) の三つの普遍因子に属する形容詞対を用いた。実験データを分析したところ、映像と音楽の組み合わせの印象に対する音楽の効果は単純に付加的なものであり、そして特定の映像と音楽形式の間に相互作用は存在しないという予想外の結果が得られた。一つの例として、視聴覚刺激の評価において評価得点の高い音楽は映像刺激と関係なく、直ちに視聴覚刺激の評価得点を高くした。彼らは、その原因を特定の印象が生じなかった単純な映像刺激にあると指摘した。よって、この結果から音と映像の相互作用はないと断言することは早急であると考えている。

Lipscomb and Kendall (1994) は、はじめてメジャー制作の映画 (*Star Trek IV: The Voyage Home*, 1986) の映像と音楽を実験刺激として用い、映像に付加する音楽によって映画の意味が変化するかを印象評定実験を通して検討した。印象評定尺度は、先行研究 (Marshall & Cohen, 1988; Sirius & Clarke, 1994; Tannenbaum, 1956) と同様に、Osgood et al. (1957) の三つの普遍因子に属する形容詞対を用いた。実験の結果、音と映像を組み合わせた視聴覚刺激と多様な印象評定尺度の間に交互作用が存在した。また、クラスター分析を行い、被験者の反応を評価因子と活動性/力動性の複合因子の二つのグループに分類した。視聴覚刺激の活動性/力動性に対する反応曲線は、映像に関係なく音楽による影響が顕著であった。これは、映画における音楽が映像の意味を変えることができるという先行研究の結果を裏付けるものである。また、評価因子に関しては Marshall and Cohen (1988) の報告のように、視聴覚刺激の文脈における音と映像の組み合わせの「適切さ (appropriateness)」によって決定されると主張した。しか

し、Lipscomb and Kendall の実験では、映画クリップのみと音楽のみの印象に関しては調べていない。

以上の Lipscomb and Kendall (1994), Marshall and Cohen (1988), Sirius and Clarke (1994), Tannenbaum (1956) の研究は、音楽の意味が映像の情緒的な側面を強化したり弱化したりする「力」があることを実証したものである。これらの研究では、音楽の意味として概念的には Meyer (1956) の参照的意味 (referential meaning) を、方法的には Osgood et al. (1957), Osgood (1962) の情緒的意味 (affective meaning: いわゆる印象) を取り上げている。さらに、音楽の意味が映像の情緒的な側面に与える影響に関する興味深い研究がある。Thayer and Levenson (1983) は、映画音楽の違いが緊張感 (stressfulness) を生じさせる映像にどのような影響を及ぼすかを生理評価と主観評価を用いて検討した。被験者が三つの事故シーンから構成された 12 分の白黒の産業安全映画を視聴している間、五つの生理反応 (interbeat interval of the heart [IBI], general somatic activity [GSA], skin conductance level [SCL], pulse transmission times to the finger and the ear [PTF and PTE], finger pulse amplitude [FPA]) を測定した。それと同時に、被験者は調整ダイヤルを用いて実時間で心理反応 (落ち着いた-緊張した) を報告した。実験は、音楽無し (映像のみ) の統制条件と、それにドキュメンタリー音楽 (major 7th コード進行) とホラー音楽 (diminish 7th コード進行) を組み合わせた二つの実験条件を設けた。実験の結果、皮膚コンダクタンス・レベル (SCL) のみにおいて、三つの条件間で統計的に有意な差が認められた。つまり、映像に付加する音楽が緊張感を緩和させたり、増加させたりする効果を持っていることが示された。この実験より、映画音楽が人の体内興奮 (internal arousal) レベルを変化させる効果があることが示された。しかしながら、被験者の心理反応に関しては三つの条件とも統計的に有意な差はなかった。

音楽表現を規定する各構成要素が、映像作品の印象に及ぼす影響を検討した研究がある。Cohen (1993) は、コンピュータの画面にボールがバウンドする非常に単純な CG (computer graphics) 映像とメロディ (分散構造の長/短 3 度の反復音) を用い、映画の意味に与える音楽の影響を調べた。映像刺激においては、バウンドの高さ (低, 中, 高) と速さ (遅, 中, 速) を操作した。音刺激においては、音高 (低, 中, 高) とテンポ (遅, 中, 速) を操作した。被験者は、映像のみと音のみのユニモーダル条件の印象

と、それらを組み合わせたバイモーダル条件の印象について、5段階の形容詞対尺度（悲しい—嬉しい）を用いて評定を行った。ユニモーダル条件では、それぞれの映像と音楽の悲しさ/嬉しさについて、バイモーダル条件では音楽の結合によって映像の悲しさ/嬉しさがもとの評定より変化したかを調べた。実験の結果、音楽の印象が映像の印象に影響を与えることが示された。ユニモーダル条件の印象の場合、映像の高さと速さが増すにつれ、音楽の音高とテンポが増すにつれて（Hevner, 1936; Riggs, 1964）ほぼ線形的に「嬉しさ」が増加した。そして、ユニモーダル条件と比較したバイモーダル条件の印象の場合、嬉しい音楽（高い音高と速いテンポ）は悲しい映像（低くて遅いバウンド）の悲しさを低下させるなど、視覚と聴覚の意味が加算的に働き、視聴覚刺激の全体的な意味（悲しさ/嬉しさ）を生み出すことが明らかになった。彼女は、このような現象を「連合（association）」と呼んだ。しかし、この実験では単一の印象条件しか調べておらず、違う印象条件下で研究を進める必要があるといえる。例えば、Bolivar et al. (1994) は、狼の社会的行動を表す実写の映像と歌詞のない市販の放送用 CD の音楽を用い、音楽が映像の印象に与える影響について「友好的—攻撃的」であるか印象評定実験を行った。その結果、背景音楽（BGM）の意味（友好的/攻撃的）が系統的に映画クリップに表現された狼の社会的行動の意味に影響を及ぼした。

岩宮、佐野（1997）は、Cohen（1993）の研究はごく単純な映像刺激と音刺激を用いていたことを指摘した。そして、約 25 秒の物語の展開が曖昧なアニメーション（*The Snowman*）と飛行チームの実写映像（*Rolling in the sky*）に、同一のメロディ（Abalon）に対して伴奏形態（ロック、ジャズ、クラシック）、テンポ（遅い、普通、速い）、調性（短調、長調）を系統的に操作した音刺激を組み合わせ、映像作品の印象が組み合わせる音楽によって変化するのかを検討した。実験の結果、アニメーションの場合も実写映画の場合も、クラシックの伴奏形態を持つ音楽は、「拡がった」印象を生じさせ、ロックの伴奏形態は拡がり感を減少させた。ジャズは、その中間であった。また、速いテンポの音楽は、映像を「白熱した」印象に、遅いテンポの音楽は映像を「淡々とした」印象のものとした。さらに、長調の音楽は映像の印象を明るく、短調の音楽は映像の印象を暗くした。

音楽が映像の物語の展開や解釈に及ぼす影響を検討した研究がある。Bullerjahn and Guldenring（1994）は、Landesmedienstelle Hannover が制作した約 10 分の多義的で

終止感のない実写映画と3人のプロの映画音楽作曲家 (Rainer Kühan, Eugen Thomass, & Peer Raben) がシンセサイザーで作曲した三つのジャンル (犯罪, メロドラマ, 漠然) の五つのバージョンの音楽を用い, 映画音楽の効果について定量的 (quantitative) 分析と定性的 (qualitative) 分析を試みた。両方の分析より, 音楽は映画の情動的雰囲気とプロットの理解に影響を与えることが分かった。すなわち, 映画音楽は, 情動的カテゴリーの誘導, ジャンルの選択, 主人公の動作の理由, 映画の完了に関する期待感に効果があった。特に, 映画音楽が物語の展開に影響を及ぼすという結果は興味深い。例えば, 議論の余地はあるが, メロドラマ・ジャンルの音楽のうち, 長和音 (major chord) の終止が物語の解決の理由を説明するという。

Thompson, Russo, and Sinclair (1994) は, 映画の主観的終止 (closure) の評定における背景音楽の影響を調べるため, 三つの印象評定実験を行った。実験1では, 12秒の3D映像にシンセサイザーで作曲した2種類の伴奏形態 (強い終止感と弱い終止感) の音楽を組み合わせた視聴覚刺激を用いた。実験2では, より現実に近い素材として, 著者のSinclairが制作した30秒と34秒の2種類の実写映像とプロの映画音楽作曲家 (Garnet Willis) が作曲した強い終止感と弱い終止感をもたらす2種類の音楽をそれぞれ組み合わせた4種類の視聴覚刺激を用いた。実験3では, 公開されたハリウッド映画 (*Clue*, 1985) より抜粋した12種類の映像とシンセサイザーで作曲した12種類の音楽を用い, 映像のみの条件と音のみの条件, そしてそれらを互いに組み合わせた12種類の組み合わせ条件を設けた。実験1の結果, 背景音楽の終止感の違いによる短いアニメーションの主観的終止感の差が統計的に有意であった。しかし, 被験者たちの実験後の内観報告では, 視聴覚刺激の終止感に対して音楽より映像に基づいて反応したとしており, 音楽の影響は暗黙 (implicit) 的であったといえる。また, 実験2の結果, すべてとはいえないが, 実験1より現実的な映像クリップの場合も背景音楽の終止感の違いが映画の主観的終止感に影響を与えた。実験3の場合も, 音楽が映画の主観的終止感に影響を与えたが, その効果は音楽より映像によって強く左右された。Thompson et al. の一連の実験は, 映画の主観的終止感に対する被験者の反応が, 音楽による効果もあったが, 映像の影響が大きいことを示した。しかし, 音楽経験の豊富な被験者や映像の情報が貧弱で魅力的でない場合は, 映像より音楽による影響がもっと強まる可能性が十分にある (Marshall & Cohen, 1988)。

実際に, Cohen (1993) は, 映像の意味が多義的である場合は音楽が映画の意味に強く作用する反面, 映像の意味が絶対的な場合は音楽の影響はみられないことを示唆する実験を行った。彼女は, 女性が男性から走る多義的なシーン (F1) と二人の男性が喧嘩する絶対的なシーン (F2) の2種類の映画クリップに *Conflict* (M1) と *Say Hello to Love* (M2) と名づけた対照的な形式の2種類の映画音楽をそれぞれ組み合わせた視聴覚刺激を作成し, 音楽が映画の情緒的 (affective) 意味と外延的 (denotative) 意味に与える影響を調べた。情緒的意味に関しては, 評価, 力動性, 活動性について印象評定実験を行った。外延的意味に関しては, 被験者に様々なタイトルの中で適切なものを評定するようにした。実験の結果, F1 の場合, 映画の情緒的意味と外延的意味の両方とも, M1 からは直接に影響を受けたが, M2 からは受けていない。F2 の場合は, M1 と M2 の意味が映画の意味に影響しなかった。この様に, 男性の喧嘩シーンは男女の追跡シーンよりも決然としており, 視聴者の映画の解釈に対する音楽の影響は映像の曖昧さの度合いに大きく依存することが明らかである。こうした映像の多義的な状況を解決する音楽の機能は, Bolivar et al. (1994) の実験でも確認された。狼の社会的行動を表す映像が喧嘩しているか遊んでいるか不明な場合, 音楽はその意味の解釈に直接に影響を与えた。映画音楽は, 映像の意味を定義するのに役に立つと同時に, その効果は多義的な映像においてより効果的である。

背景音楽が映像の記憶に及ぼす影響を検討した研究がある。Boltz et al. (1991) は, 映像シーンと関連した音楽の雰囲気と挿入位置を操作し, 映画イベントの記憶に影響する音楽 (メロディ) の効果を調べた。実験刺激は, 予備実験を通して, 否定的 (negative) /肯定的 (positive) なエンディングであると決められた16種類の映画クリップに, 同じく予備実験から選ばれた否定的/肯定的な音楽を組み合わせた。被験者を三つの群に分けて, 第一群には, 統制群として音楽なしの映像刺激のみを呈示した。実験群として第二群と第三群には, 映像と音楽を同時に呈示する (accompanying) 条件と音楽が映像より先行する (foreshadowing) 条件を設けた。それぞれ, 映画の情緒的意味を強調する条件と映画イベントの進行を期待する条件ともいえる。その上, 音楽の印象を映像エピソードの結果と調和する条件と調和していない条件にした。三つの群の被験者は, 視聴実験後に16種類の映画クリップをできる限り多く再生 (recall) する課題を行った。分析の結果, 音楽の雰囲気との調和 (mood congruent) と挿入位置の間に交互作用が存在

した。統制条件と比較して、映像より音楽が先行する条件では映像と音楽の雰囲気調和していない条件で記憶成績が良かったが、映像と音楽を同時に呈示する条件では映像と音楽の雰囲気が一致する条件で記憶成績が良かった。さらに、再認 (recognition) 課題では、再生課題で記憶できなかつたシーンを再認することができた。彼女らはこれらの結果を、「選択的参加処理 (selective-attending processes)」によるものと解釈している。

さらに、Boltz (2001) は、音楽が映画に与える事後記憶 (subsequent remembering) のみならず、物語の理解や登場人物の動作や動機、個性の解釈に与える影響についても検討した。肯定的とも否定的とも解釈可能な3種類の短くて多義的な映画クリップ (*Cat people*, *Vertigo*, *Split Decisions*) にそれぞれ肯定的な音楽、否定的な音楽、音楽なし (映像のみ) の条件を設けて一連の実験を行った。被験者は、各映画クリップを視聴してからすぐに、映画のエンディングと主人公の動機や個人性に関する質問を文章で答える再生課題を行った。併せて、映画の主人公の動作や感情を4種類の形容詞対を用いて評定した。さらに、異なる被験者に、予告せず1週間後に各映画に出たオブジェクトの記憶に関する再認課題を行わせた。実験の結果、音楽なしの統制条件と比較して、肯定的な音楽と否定的な音楽は映像と雰囲気が調和すると、映画に対する被験者の解釈や事後記憶を有意に変化させた。すなわち、映画音楽が映像の認知に大きく影響を与えることが明らかになった。例えば、*Vertigo* (Hitchcock, 1958) の男性が女性を追いかけるシーンの場合、音楽無しの条件ではその男性が探偵に、否定的な音楽の条件では殺し屋に、肯定的な音楽の条件では恋人として再生された。また、被験者たちは肯定的な音楽の場合に「黒いセダン」「さびれた路地」より「晴れた日」「花束」など肯定的なアイテムを記憶しやすかったが、否定的な音楽の場合は反対の結果が得られた。

マスコミ研究の分野で、テレビの構成要素の操作が人の認知過程に与える影響を調べた研究がある。ここで認知過程とは、視聴者が注意を向けることによって、視覚情報を目から、聴覚刺激を耳から受容し、作業記憶 (working memory) で情報処理を行い、長期記憶 (long-term memory) に貯蔵する一連の過程をいう。特に、1970年代後半から映像情報と音声情報の相互作用に関するテレビ・ニュースの理解や記憶に関する研究が多く行われた (Edwardson, Grooms, & Pringle, 1976; Gunter, 1979; Katz, Adoni, & Parness, 1977; McDaniel, 1973)。これらの研究は、テレビ・ニュースに映像を挿入す

ることによって、映像と音声の不調和が生じたり、注意が逸脱したりする否定的な「散漫効果 (distraction effect)」と、詳細な映像を付加することによる肯定的な「学習効果 (理解や記憶が向上すること)」を調べたものである。Grimes (1990a) は、4種類のテレビ・ニュースに、視覚情報と音声情報の対応を意味的に非常に一致する条件、中間程度に一致する条件、まったく一致しない条件の三つの条件を設け、それぞれ異なる被験者群に呈示して注意と記憶に関する視聴覚情報チャンネルの対応の効果を検討した。実験の結果、意味的に非常に一致する条件では選択的注意配分の効果と映像とナレーションに関する再認成績が最も顕著であった。また、すべての条件を通して注意の容量限界を超えるときには、全体的なニュースの記憶が低下する傾向がみられた。つまり、映像情報と音声情報が意味的に一致している場合には、その重複している情報が視聴者によりよく記憶されて学習効果が得られるが、逆に映像情報と音声情報が意味的に一致していない場合は散漫効果が生じるのである (Reese, 1984; Son, Reese, & Davie, 1987)。また、Grimes (1990b) は、映像と音声の意味内容が一致していないときには、映像情報の記憶と音声情報の記憶が相互に干渉し合い、もとの情報から得た記憶が変容してしまう「転移現象 (translation phenomenon)」が生じると考えた。彼は、実験用のテレビ・ニュースを制作し、その干渉効果が映像記憶から言語記憶へ、言語記憶から映像記憶への両方において生じることを実証した。例えば、「“男性” 大学生のショット」に「最近の研究によれば、多くの“人” がデート中にストレスを感じる」というナレーションを組み合わせると、視覚情報から聴覚情報への干渉によって「多くの“男性” がデート中にストレスを感じる」という応答が多くなった。さらに、Grimes (1991) は、映像と音声といった異なる形態の刺激が一つの刺激として「一体化 (belonging hypothesis)」することによって互いの処理を容易にすると考えた。映像と音声の意味の一致度が高いほど、この一体化が容易になり、注意の容量を超えることなく処理することができるという。

マルチメディア学習の分野で、音と映像の相互作用を検討した研究もある。Mayer and Moreno (1998) は、アニメーションを用いたマルチメディアの学習効果について一連の実験を行った。実験1では稲妻の発生課程を描写したアニメーションに、その内容と一致するテキスト (AT群) と男性のナレーション (AN群) を付加して学生たちに呈示し、異なる情報源による学習効果を調べた。実験2でも、車のブレーキシステムの操作に関

するアニメーションを用い、実験1と同様にAT群とAN群に分けて学習効果を調べた。実験1と実験2で行った再生 (retention) 検査, マッチング検査, 転移 (transfer) 検査の結果, AT群の学生よりAN群の学生の学習効果がより向上された。すなわち, マルチメディア学習において学生たちは, 言語情報が視覚的に提示されるより聴覚的に提示されたとき, アニメーションとことばを簡単に統合することができた (i. e., *modality principle*)。彼らは, この結果を視覚と聴覚それぞれのチャンネルから構成されている作業記憶の「二重処理モデル (dual processing model)」で説明している。例えば, AT群の場合はアニメーションとテキストの情報が視覚作業記憶の使用容量の限界に達して過負荷となり, 結局認知的処理を妨げる結果になったという。また, Moreno and Mayer (2000)は, アニメーションの内容と一致するナレーションに (AN群) に, 背景音楽を付加した群 (ANM), 効果音を付加した群 (ANS), 背景音楽と効果音を付加した群 (ANMS) に分けて群間の学習効果を比較した。その結果, 音楽無しの群と効果音無しの群の方がより学習効果が良かった。これらの結果から Moreno and Mayer は, マルチメディア・メッセージに対して過度な/不適切な音を付加してしまうと, 人の聴覚作業記憶の負荷を増加させて認知作業を妨げる「干渉結果 (coherence effect)」が生じることを示した。しかし, 映像に付加する聴覚情報が, 選択的注意を誘導して視聴者の認知作用を助ける「覚醒効果 (arousal effect)」があることを示した研究も多くある。この様な覚醒の理論 (arousal theory) によると, 覚醒は注意と深い関係を持っており, 注意のレベルが増加すると覚醒の水準も増加するという。特に, Alwitt, Anderson, Lorch, & Levin (1980), Anderson and Levin (1976), Calvert and Gersh (1987), Calvert, Huston, Watkins, & Wright (1982), Calvert and Scott (1989) の研究は, 子供たちがテレビを視聴している間, 効果音が注意レベルの増加に寄与することを示した。例えば, Calvert and Gersh は, 子供用の白黒のテレビ番組 (Mama's Little Pirates) を用い, 効果音 (口笛) や映像 (主人公がベッドで寝ている場面) の付加が番組への注意と内容の理解に及ぼす影響について調べた。5歳児と11歳児の男女の群に, (a) 四つのプロットのみを構成した条件, (b) 四つのプロットの間効果音のみを挿入した条件, (c) 四つのプロットの間映像のみを挿入した条件, (d) 四つのプロットの間効果音と映像を挿入した条件を設けて実験を行った。実験室内には漫画やおもちゃなどを置いて, 子供が普段通りに遊びながらテレビを視聴するようにし, その様子をビデオに録

画して注意の分析に用いた。併せて、視聴実験後に 22 問から成る多重選択再認検査 (multiple-choice recognition test) を行い、番組の理解の分析を行った。その結果、幼い子供 (5 歳児群) の場合、効果音の付加が選択的注意と記憶成績の向上に貢献することが示された。Alwitt et al. は、Anderson and Levin の実験の結果を踏まえ、就学前の子供たちに約 1 時間分量のカラーの子供番組 (CM を含む) を 3 編視聴させ、37 種類の子供番組の音と映像の構成要素と注意との関係を調べた。その結果、「音響的变化」「奇妙な声」「動き」「カット (cuts)」「効果音」「笑い声」「拍手の音」は注意レベルを向上させたが、「男性の声」「ズーム (zooms) とパン (pans)」「アイ・コンタクト (eye contact)」「静止画 (still shots)」は注意レベルを低下させた。以上の結果より、人が効率的にマルチモーダル情報を認知するためには、適切な聴覚情報の付加が不可欠であることが分かる。

さらに、Mayer and Anderson (1991) は、自転車の空気入れの構造に関するアニメーションとその内容と一致するナレーションを用い、映像よりナレーションが先行する条件と映像とナレーションを同時に呈示する条件を設け、音と映像の同期が学習効果に与える影響を調べた。実験の結果、学生たちは視覚情報と聴覚情報が時間的に同期している条件で、より大きな学習効果が得られた (i. e., *temporal contiguity principle*)。Mayer の一連の研究は、マルチメディア学習における視覚情報と音声言語情報の構成に関する有益な知見を提供する。しかし、実験において因果関係について説明した視聴覚素材のみを用いており、事実関係を描写したような他の学習文脈においては違う結果が得られる可能性もある。

ここまで、心理学の諸分野における音と映像の相互作用に関する研究を概観した。そこには、「意味」「解釈」「注意」「記憶」「学習」など、人の情報処理における様々な音と映像の相互作用が存在した。こうした人の情報処理の過程には階層構造 (hierarchy) が存在しており、音と映像の相互作用も当然そのレベルに応じた反応として現れる可能性が高い。岩宮 (1992, 1993, 1996, 2000, 2002) は図 1.3 に示すように、人の視聴覚の情報処理レベルに応じた多様な音と映像の相互作用の存在を示した。彼は、市販の映像作品の一部を用い、映像あるいは音楽のみを呈示した場合と、両者を一緒に呈示した場合の評価値の違いをもとに音と映像がもたらす視覚と聴覚の相互作用を印象評定実験より検討した。その結果、最も下位のレベルで生ずる相互作用は、「感覚の感受性の

変化」である。例えば、再生音の音質の劣化が、映像と同時に呈示されることによって感じられ難くなる現象である。あたかも、映像情報が音質劣化を補償するように機能する。このような感覚間の直接的な相互作用を、Ryan (1940) は「*dynamogenic interaction*」と呼んだ。もう少し上位のレベルでは、視覚と聴覚に共通する心理的属性を通して生ずる相互作用がある。こうした視覚にも聴覚にも存在する心理的属性 (e. g., 明るさ) は

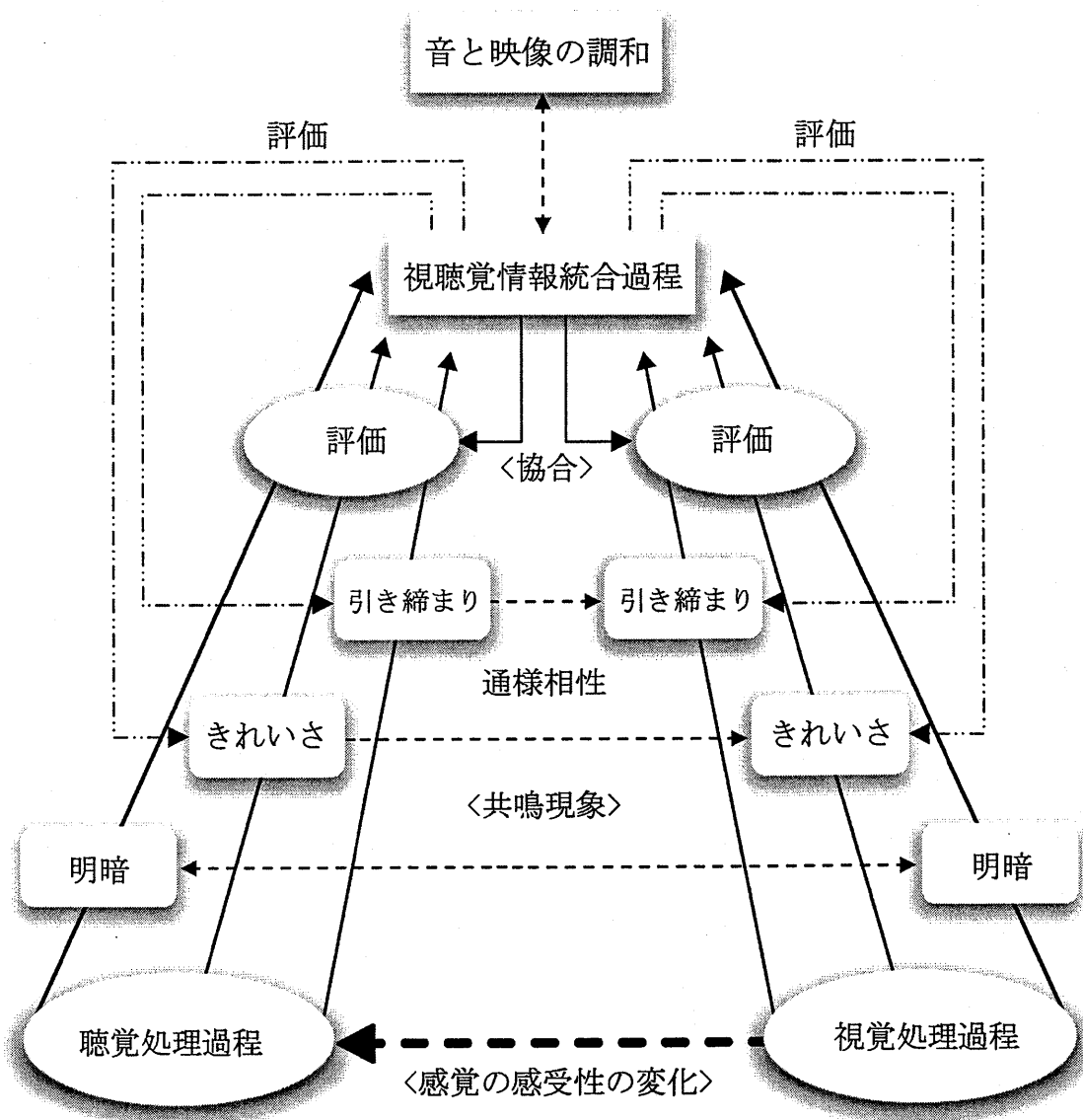


図 1.3 AV メディアを通しての情報伝達における視覚と聴覚の相互作用のモデル化. 岩宮 (2000) より転載

「通様相性 (盛永, 野口, 1969)」と呼ばれている。そして, この通様相性が同方向に変化すると「共鳴現象 (*consonance phenomenon*)」が生じる。ただし, 通様相性は, その心理的属性に応じて処理レベルが異なり, 共鳴現象の生じ方にも差がある。例えば, 「明るさ」の場合, 視覚においても聴覚においても, 比較的到低次の処理レベルで生じる心理的属性であり, 視聴覚情報統合機構の処理とは関係なく共鳴現象が生じる。すなわち, 明るさに関しては音と映像の組み合わせとは関係なく, 共鳴現象が観測されるのである。一方, 「きれいさ」「引き締まり」感は, もう少し高次の処理レベルで生ずる心理的属性であり, 視聴覚情報統合機構を媒介とする。従って, 任意に音と映像を組み合わせた場合は, 共鳴現象は生じない。また, 明るさにおいては双方向の効果が認められるが, きれいさと引き締まり感においては聴覚から視覚への効果が大きい。1.2.2.1項で述べた音源定位とは対照的に, 情緒的な情報のコミュニケーションにおいては聴覚優位な面がある。さらに, 一番高次のレベルで生ずる「協合現象 (*cooperative interaction*)」がある。協合現象は, 視聴覚情報統合機構で適切な音と映像が組み合わせられたかの判断によって生じ, 映像作品の総合的評価に反映される。協合現象により, 音と映像を単独で呈示した場合より, 音と映像を同時に呈示した方が, 評価値が上昇するのである。なお, 協合現象は音と映像の主観的調和と相関が高い。次の1.2.3項では, 音と映像を組み合わせる際に生ずる主観的調和について検討する。

1.2.3 音と映像の主観的調和

1.2.2.3項では, 主に映像への音の付加的機能について検討したが, 本項では音と映像の視聴覚情報統合の認知的メカニズムについて概観する。

一般に, 映像メディアの制作者は, 音と映像を組み合わせる際に両者の間の主観的調和を図っていると考えられる。Iwamiya (1994), Lipscomb and Kendall (1994) は, 市販の映像作品よりまとまりの良い部分の映像と音楽を用い, 元の組み合わせのオリジナル条件と別の映像の音楽を任意に組み合わせた条件を設け, 音と映像の調和に関する心理実験を行った。その結果, 制作者が意図した音楽との組み合わせであるオリジナル条件が最も調和することが示された。この様に, 映像作品における音と映像の組み合わせは任意なものではなく, そこには心理的現象を反映したルールが存在する。

近年, 音楽心理学の分野では, 音と映像を組み合わせる際の認知的なまとまりの良さ,

つまり音と映像の主観的調和（「調和感」と記述することもある）について「構造的調和」と「意味的調和」の二つの側面から検討することが多い。ここで、構造的調和（formal or structural congruency）とは、音と映像との時間や変化パターンの類似のことである。意味的調和（semantic congruency）とは、様々なレベルでの音と映像との意味の類似のことである。こうした音と映像の構造的調和と意味的調和は、Bolivar et al. (1994) の実験によってはじめて実証された。構造的調和は、主にゲシュタルト心理学の原理を反映して視覚情報と聴覚情報を一つの知覚的処理単位（*crossmodal unit*: Bertelson, 1999; Welch, 1999b）としてまとめる役割を果たすと考えられる（Marshall & Cohen, 1988; Bolivar et al. 1994; Iwamiya, 2002）。ゲシュタルト心理学は、20世紀の初頭に Wertheimer (1880-1943), Köhler (1887-1967), Koffka (1886-1941) のベルリン学派（Die Berline Schule）によって理論化され、視知覚におけるまとまりや群化を決定するゲシュタルト要因（Gestaltfaktoren）を明らかにした。一方、聴覚の分野では、視覚研究に後れて1970年代から Bregman 一派により聴覚研究においてもゲシュタルト心理学の原理が導入された。なお、音楽の分野においては、Deutsch (1982/1987) が音階旋律の錯覚（scale illusion）の研究においてゲシュタルト心理学の考え方を導入している。この様に、ゲシュタルト心理学の原理は、一般に視覚と聴覚に共通してみられるもので、その基本的な考え方は視覚研究において見いだされたものである（Moore, 1989/1994）。最近では、近接（proximity）の要因や共通運命（common fate）の要因のようなゲシュタルト要因が視覚と聴覚のような異種の感覚モダリティ間の情報統合に働くことを示唆する研究もある（e.g., Bertelson, 1993; Bertelson & Radeau, 1981; Welch & Warren, 1980）。一方、意味的調和は視聴覚コンテンツの意味を視聴者へ伝達するのに役に立つと考えられる（Bolivar et al., 1994; Iwamiya, 2002; Lipscomb & Kendall, 1994）。

もちろん、映像メディアにおける音と映像の関係は、調和するものばかりでなない。ときには、音と映像が対立的/競争的な関係（e.g., Cook, 1998; Pauli, 1976）となり、調和的關係よりも格段に効果がある場合がある。黒澤明（1910-1998）のように、制作者の中にもわざと調和しない音楽を映像と組み合わせて印象的な作品を制作する人もいる。この様な映画表現の手法は「対位法（counterpoint）」と呼ばれている。Whitaker (1970/1983) は、対位法的音楽の機能を、コントラストによって視覚の限界を強調す

るか、又は映像と音楽の対立から新しい風刺的メッセージを作り上げると述べた。しかし、Chion (1985/1993) が述べたように、一般に対位法の方が興味深いとは考えられているものの、この手法が映画で示されることはずっと少ないし、本当に説得力のある例となるととりわけ少ない。むしろ音と映像が調和的關係のもの例ならほとんどのあらゆる映像メディアで無数に見られる。よって、本研究では音と映像の調和的關係に焦点を当てることにする。

1.2.3.1 構造と意味の概念

Welch and Watt (1982) は、情報の認知的な影響を考察する際に、その情報を「形式 (form)」と「内容 (content)」に区別する必要があるとした。人の認知過程の階層構造を考慮すれば、情報の「形式」は低次の知覚/感覚的 (sensory) レベルに効果をもたらす、情報の「内容」はより高次の認識/意味的 (semantic) レベルに効果をもたらすという。本項では、音と映像の構造的調和と意味的調和を検討する前に、映像メディアにおける「構造」と「意味」の概念を考察する。

Cohen (2005) は、心理学的見解による映画知覚における音楽の影響を、「構造 (structure)」と「意味 (meaning)」の側面で考察した。音楽の構造は、時代や文化の中で発生した音楽の文法や様式によって特徴付けられた音の組織的な関係を意味する。例えば、西洋音楽の場合は音の物理的属性を基にした音程、3和音、音階、調性、リズムが構造に当たる。音楽の意味は、心によってもたらされる連想的側面と情緒的側面がある。一般に、音楽聴取においては構造より意味の方が優位になる。同様に、映像の構造は視覚イメージの構成要素の特徴を表し、意味は心によってもたらされる連想、感情や理解を表す。要するに、構造とは全体を成り立たせているもので形式の下部概念であり、意味とは了解された内容を表すと考えられる。

映像メディアの全体を通して構造と意味が完全に二元論的に分離できるとは考え難い。何故ならば、Whitaker (1970/1983) が述べたように「形式」と「内容」は相互補足的な関係であるからである。従って、本論文での「構造」と「意味」の区分は、重大な意味を持たない。単に、研究の立場から分類しただけであることを明視する。映像メディアの意味の側面は、構造と結び付いている。映像メディアの構造の側面は、意味を表す視覚的/聴覚的構成要素が関係すると考える。

1.2.3.2 音と映像の構造的調和

1.2.3.2.1 音と映像の時間構造の類似

時間的な同期 (temporal synchrony) は、異種の感覚モダリティ間の情報統合において貢献度が高く、本項では時間的側面での音と映像の統合について検討する。Marshall and Cohen (1988) は、音と映像の時間的な調和 (temporal congruency) によって選択的注意が働き、音楽の情緒的/含蓄的意味が映像と連合 (association) すると解釈しており、音と映像の時間構造の類似による構造的調和の効果について言及した。しかし、彼女らの研究では、音と映像の構造的調和の効果を実証していない。

Lipscomb (1996, 1999) は、Marshall and Cohen (1988) などの音楽の参照的 (referential) 機能を強調した既存の音と映像の相互作用の研究に対し、人の視覚と聴覚の知覚的統合における視覚領域と聴覚領域のアクセント構造の同期に関する一連の研究を行った。彼は、映画とアニメーションの映像的/音楽的なアクセント構造の整列レベルを図 1.4 に示すように (a) 協和 (consonant) (b) 位相ずれ (out-of-phase) (c) 不協和 (dissonant) の三つの条件を考えた。(a) 協和条件は、視覚と聴覚のアクセントが完全に同じ周期で起こる。(b) 位相ずれ条件は、アクセントは同じ周期で起こるが、

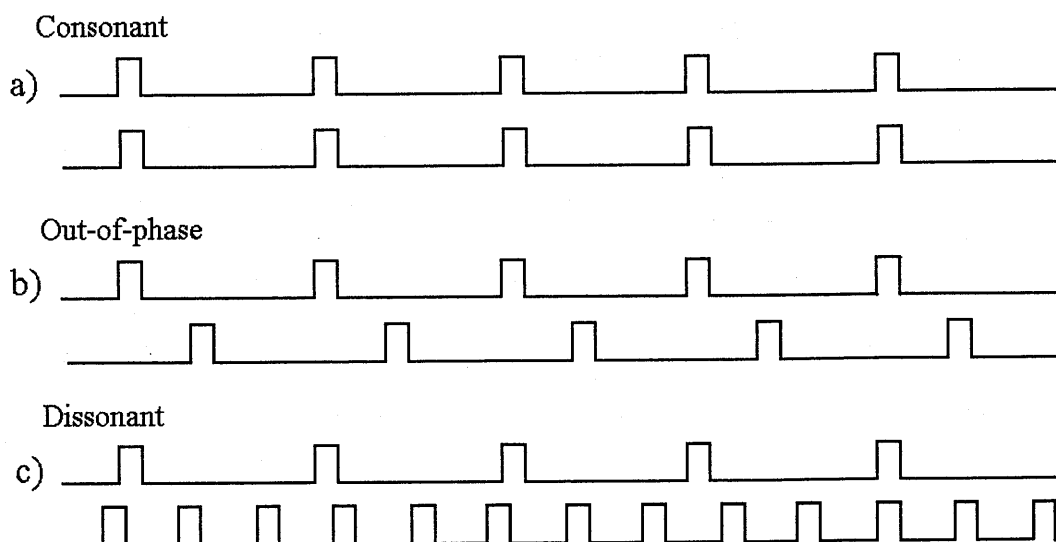


図1.4 視覚事象と聴覚事象の間のアクセント構造の整列の条件 (after Yeston, 1976).

Lipscomb (1996) より転載

その整列がずれている。(c) 不協和条件は、完全に違う周期でアクセントが起こる。被験者は二つの群に分かれ、次の二つの課題を行った。印象評定課題では「効果のないー効果のある」「同期していないー同期している」の2対の VAME 尺度を用い、比較評定課題では「同じでないー同じである」の VAME 尺度を用いた。VAME (Verbal Attribute Magnitude Estimation: Kendall & Carterette, 1993) 尺度とは、両極の形容詞対の替わりに反対語対を用い、段階を用いずに連続線状で構成したものである。実験刺激は、自作した単純なアニメーションと MIDI (Musical Instrument Digital Interface) 音源 (実験 1), Norman McLaren のアニメーションと音楽 (実験 2), 実写映画 (*Obsession*) とオーケストラ風の映画音楽 (実験 3) である。この様に、各実験では映像刺激と音刺激の複雑さ (complexity) が異なる。実験 1 の印象評定の結果、協和条件で「同期」と「効果」の評定値が最も顕著であり、位相ずれ条件はやや悪く、不協和条件では最も悪かった。また、類似性判断の結果、多次元尺度構成法 (MDS) とクラスター分析により容易に三つの次元 (視覚要素, 聴覚要素, 整列状態) に解釈できた。一方、実験 2 の印象評定と類似性判断では、刺激の複雑さの増加とともに同期の効果が弱くなる結果となった。さらに、実験 3 の結果からは、被験者の同期評定と効果評定との関係が分離された。クラスター分析からも、アクセント構造の整列状態による影響はみられなかった。この新しい 3 番目の次元は視聴覚組み合わせの「適切さ」と解釈された。Lipscomb は以上のことを踏まえ、Lipscomb and Kendall の「映画音楽知覚パラダイム (Film Music Perception Paradigm, 1994)」を図 1.5 のように改良して結果の解釈を試みた。

図 1.5 の改良されたモデルは、視聴者の注意的焦点が音楽的構成要素に移動するか、それとも潜在意識の中に留まるかを基にした (i) 連合判定 (association judgment) と (ii) アクセント構造の整列 (accent structure alignment) の二つの暗黙の処理が行われるという。(i) 連合判定は、与えられた文脈と音楽が調和するかの基準として過去の経験が強く作用する。(ii) アクセント構造の整列は、音楽と映像の重要な事象が同時に発生することによって決められる。単純な刺激の場合は、音と映像のアクセント構造の役割が顕著であるが、刺激が複雑となるにつれてアクセント構造の役割は減衰し、その代わりに音楽の意味の連合判定が顕著になる。Lipscomb の研究は、音と映像のアクセント構造の同期が視聴覚情報の主観的調和に大きく影響を与えることを実験的に示した点で有意義なものである。

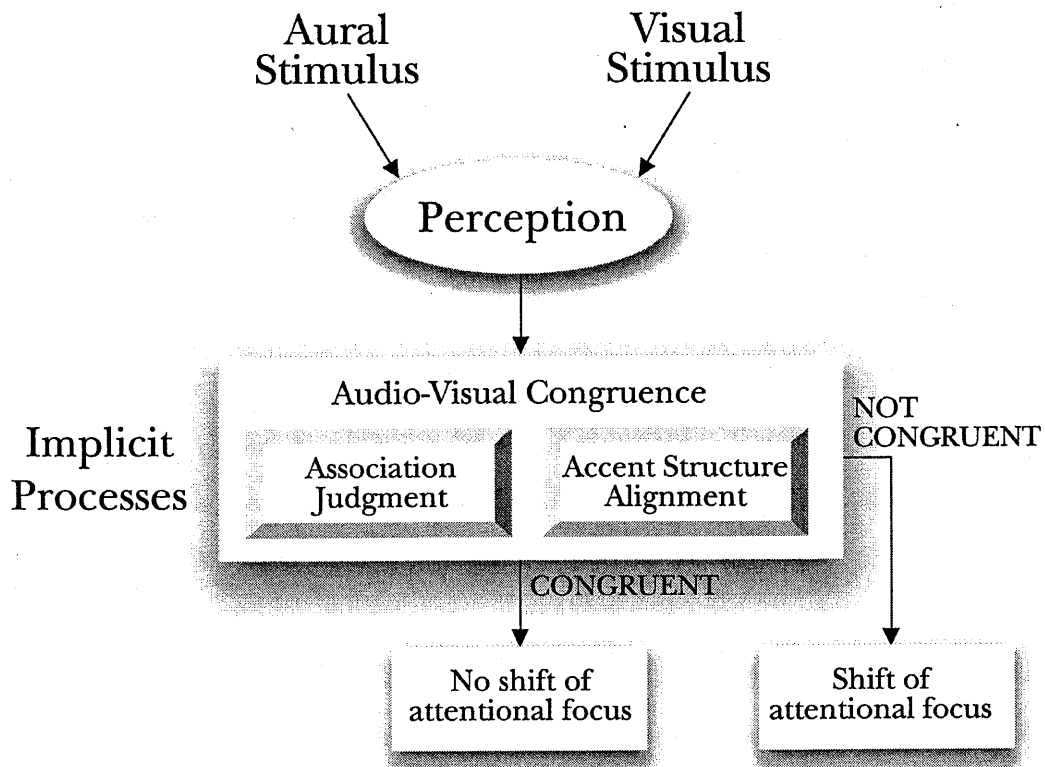


図 1.5 Lipscomb and Kendall (1994) の Film Music Perception Paradigm の修正版.
Lipscomb and Tolchinsky (2005) より転載

菅野, 岩宮 (1999) は, Lipscomb (1996, 1999) の一連の実験では, 映像と音楽の同期のさせ方が限定された点を指摘し, 調整法による映像の動きの速さと音楽のテンポのマッチング実験を行った。映像刺激は, 球体が一定の速さで正方形を描いて運動するアクセント有りの条件と, 球体が一定の速さで円を描いて運動するアクセント無しの条件から成る。音刺激は, Abalon (Vincent Rose 作曲) を音楽的アクセント有りの条件と音楽的アクセント無しの条件を設け, それぞれ五つのテンポ条件 (100, 150, 200, 250, 300 bpm) を設定して MIDI で編曲した。被験者の課題は, 同時に流れる音楽と調和するように画面に呈示された球体の動きの速さを調整することである。実験の結果, 映像と音楽のアクセント構造の同期が音と映像の主観的調和に対して大きな影響を及ぼしていることが示された。特に, 映像の動きにアクセントがある条件で, 音と映像を同期させようとする被験者の意識が明確となった。また, 速いテンポの音楽には速い動きの映像が調和し, 遅いテンポの音楽には遅い動きの映像が調和することが示された。しかし,

調整法による実験では、映像と音楽のアクセント構造が同期しているときに、実際に音と映像の間に調和が感じられたかどうかは明白ではない。この点に関して、菅野、岩宮（1998, 2000）は単純な刺激とより複雑な刺激を用いて印象評定実験を行い、映像と音楽のアクセント構造の同期が音と映像の主観的調和に対して影響することを確認した。また、狼の社会的行動を撮影した映画クリップと音楽を用いて音と映像の「動き」の間の時間的調和を調べた Bolivar et al. (1944) の実験からも、時間的に同期する視聴覚刺激と時間的に同期していない視聴覚刺激の間の調和度に統計的に有意な差が認められた。Iwamiya (1994) も、被験者に映像と音楽が同期している条件とその同期を 500 ms ずらした同期していない条件の調和度を評定させた結果、同期している条件で調和度が高いことを確認している。

以上の Bolivar et al. (1994), Iwamiya (1994), Lipscomb (1996, 1999), 菅野, 岩宮 (1998, 1999, 2000) の研究は、映像メディアにおける音と映像の時間構造の役割について述べた Chion (1990/1994), Marshall and Cohen (1988) の主張を強く裏付けるものである。この様に、音と映像の「同期 (synchronization)」は、視覚情報と聴覚情報を統合する重要な知覚的要因である。映像メディアにおける音と映像の同期の手法は、ミッキー・マウシングと呼ばれている。この手法は、少し昔風の方法でもあるが、映像メディアの制作において有効に使われている (Burt, 1996; Weis & Belton, 1985)。

1.2.3.2.2 音と映像の変化パターンの類似

これまでに行われてきた音と映像の構造的調和に関する研究は、1.2.3.2.1 項で述べたようにそのほとんどが音と映像の時間構造の類似に関するものであった。しかし、近年、音と映像の特定の変化パターンの類似によっても構造的調和が成立しうることを示唆する研究がある。

岩宮、関、吉川、高田 (2003) は、テレビ番組や映画などで用いられている場面切り替えパターンと効果音に着目し、10 種類の効果音と 14 種類の切り替えパターンを実験刺激として用い、印象評定実験を行った。その結果、とりわけ上昇形の音列と拡大形の切り替えパターン、下降形の音列と縮小形の切り替えパターンの調和度が高かった。彼らは、こうした音と映像の主観的調和を、音と映像の変化パターンの一致に基づく構造的調和によるものと解釈した。しかし、明るい印象を持つ音列 (4 種類の上昇形と 1 種

類の下降形)は、どの切り替えパターンとも調和しており、効果音の印象が調和感に影響を及ぼした可能性が高い。また、音刺激の変化パターンも非常に限られており、系統的に操作したものではなかった。この点に関して、有田、岩宮(2004)は追加実験を行っており、系統的に操作した音列を用いた場合も、拡大する映像と上昇形の音列、縮小する映像と下降形の音列が調和することを確認した。

岩宮、尾崎、高田(2005)は、拡大あるいは縮小する3種類の抽象的な図形に音の高さと大きさを系統的に操作した7種類の音列パターンを組み合わせ、音と映像の構造的調和と視聴覚刺激の印象の影響を印象評定実験より検討した。実験の結果、一般に図形が拡大/縮小するような映像には、音高が上昇/下降する音列が調和した。なお、拡大する映像と音量の増大との対応関係も良いことが分かった。また、音と映像の主観的調和に関して、各視聴覚刺激の「調和感」と「活動感」「力動感」との間の相関係数は低く、実験の結果が図形と音列の変化パターンによってもたらされる構造的調和に基づくものであることを示した。一方、調和主成分と評価主成分との相関係数は高く、音と映像の特定の変化の類似による構造的調和と視聴覚刺激の評価が深い関係を持つことが分かった。

Lipscomb and Kim(2004)は、「色」「サイズ」「形」を小、中、大に操作した図形が飛び飛びに縦方向を狭、中、広に上がって下がる映像と「音高」「音色」「ラウドネス」「音価」を小、中、大に操作した音を用い、視聴覚構成要素間の対応関係を調べた。実験の結果、音と映像の構成要素の間に調和した対は、「音高と映像の縦位置」「ラウドネスとサイズ」「音色と形」であった。音価は、すべての映像パラメータと調和しなかった。また、色は音高とラウドネスに同じく調和した。この結果は、視覚と聴覚の「非一元的關係(non-unitary relationships)」を示す例であるといえる。すなわち、一つの映像パラメータに複数の音楽パラメータが合う。なお、Walker(1987)の報告と違って、Lipscomb and Kimの実験では音楽教育の差による実験結果の違いはみられなかった。

有田、蘇、川上、上田、岩宮(2005)は、8種類の映像の切り替わり方向(上方向、下方向、右方向、左方向、右上方向、右下方向、左上方向、左下方向)と下降と上昇の変化方向を持つ音刺激との構造的調和に関して評定尺度法と一対比較法を用いて検討した。彼らは、映像の変化方向とピアノ鍵盤の音高の変化方向に有意な関係があるという仮説を立てた。つまり、仮説が成り立つとすれば、ピアノ経験者群は未経験者群と比

較して、右方向に変化する映像と上昇音、左方向と下降音が調和すると評定するはずである。実験の結果、絶対判断でも相対判断でもピアノ経験者もピアノ未経験者も「右方向と上昇音」「左方向と下降音」が調和すると判断した。彼らの研究は、音と映像の方向変化の類似によって形成される構造的調和がピアノ経験によるものではないことを示唆している。

Eitan and Granot (2004, 2006) は、身体の動きと音楽の物理的空間パラメータとの連合の関係について検討した。まず、被験者は、1人のキャラクターのアニメーションを見て、その視覚的イメージを思い浮かぶ。そして、16種類の簡単な音楽(メロディ)を聴いてから視覚的イメージと音楽的イメージとの対応関係を強制選択質問法(動きの形態、方向、速度の変化など)で答えた。音楽パラメータとしては、ダイナミックス、音高の輪郭、音高の音程、アタックレート、アーティキュレーションをそれぞれ「強化」と「弱化」(e.g., クレッシェンド vs. デイミヌエンド, アツェレランド vs. リタルダンド)に操作した。実験の結果、大部分の音楽パラメータはいくつかのイメージした動きの次元と有意に効果があった。例えば、音高は、イメージした動きの縦の次元のみならず、横と奥行き次元とも関係した。なお、音高は、イメージした動きの速さと「エネルギー」とも関係することが示唆された。また、音楽と空間類似性(musical-spatial analogies)は、しばしば「非対称的(asymmetrical)」であった。例えば、音高の下降は空間的に下方向と強く結び付けられたが、音高の上昇は空間的に上方向とは結び付かなかった。一般に、「音楽的弱化」は空間的下降と強く関連付けられたが、「音楽的強化」は空間的上昇より「加速」と強く結び付いた。しかし、彼らの実験では被験者の内省による判断を求めており、実際に映像刺激と音刺激を組み合わせた実験は行っていない。

1.2.3.3 音と映像の意味的調和

Boltz et al. (1991) は、音と映像の雰囲気(調和(mood congruity))が記憶に及ぼす影響について実験を行ったが、Bolivar et al. (1994) は、音と映像の印象の類似による意味的調和について検討した。映像刺激は、狼の社会的行動を表す20種類の実写の映像より予備実験を通して選定した最も攻撃的な映像4種類と最も友好的な映像4種類である。音刺激としては、市販の放送用の音楽CDより選定した最も攻撃的な音楽5種類と最も友好的な音楽5種類を用いた。視聴覚刺激は、編集を経て8種類の映像に各

4通りの音楽（意味的に一致/不一致，時間的に一致/不一致）を組み合わせた計32種類である。被験者は呈示された視聴覚刺激を視聴した後，狼の社会的行動の雰囲気と音楽の雰囲気がどの程度調和（match）するのかを評定した。実験の結果，映像と音楽の印象が類似する場合（e. g. , 攻撃的な映像と攻撃的な音楽），そうでない場合よりも高い調和度が得られた。さらに，視聴覚刺激の調和評定は，その映像刺激のみと音刺激のみの意味の評定値から予測できることが示された。Bolivar et al.の研究により，音と映像の主観的調和といった心理量が科学的に測定できるものであることが分かった。しかし，音と映像の意味的調和を「攻撃的-友好的」の単一の印象条件しか調べておらず，違う印象条件下で研究を進める必要があるといえる。

岩宮，林（1999）は，色彩と音楽の印象に関する印象評定実験を行った。映像刺激は，「赤」「黄」「黄緑」「緑」「シアン」「青」「青紫」「紫」の照明条件下でピアノ演奏風景を模擬した8種類のCGである。音刺激としては，8種類のピアノ曲を用いた。実験の結果，長調でテンポの速い「明るい」印象の音楽は，「緑」「黄緑」「黄」「シアン」と相性が良く，「青」「青紫」「紫」「赤」とは合わなかった。テンポの遅い音楽は，「青」「紫」「青紫」「シアン」と調和し，「緑」「黄緑」「黄」「赤」とは調和しなかった。彼らは，一般に音楽が本来持つ特徴を助長する機能を持つ色彩が，その音楽と調和していると判断されていると考え，実験の結果を音と映像の意味的調和に基づくものであると解釈した。

岩宮，上月，菅野，高田（2002）は，映像の密度と速度，音楽のテンポと調性が映像作品の印象に及ぼす影響を検討した。映像刺激としては，一定の速度で画面の右から左へ移動する人を模擬した人形状の3D映像を用い，速い/遅い/速度条件と密/疎の密度条件を設けた。音刺激としては，テンポや調性を変えても不自然にならない楽曲をMIDI音源で長調と短調に編曲したものをを用い，それぞれのテンポを6条件（60, 90, 120, 180, 240, 300 bpm）に設定した。視聴覚刺激は，これらの4種類の映像と12種類の音楽を組み合わせた48種類である。印象評定実験の結果，「速い映像と速いテンポ」「遅い映像と遅いテンポ」「速い映像と長調」「遅い映像と短調」「高密度の映像と速いテンポ」「低密度の映像と遅いテンポ」の組み合わせが調和する傾向があった。彼らは，音と映像の調和に影響する視聴覚構成要素間の諸要因が映像作品の印象に及ぼす影響を解明した。そして，音と映像の調和が映像作品の印象に同様の効果をもたらす要因の組み合

わせによるもので、意味的調和に基づくものであると解釈した。ただし、岩宮、林 (1999) の研究と同様に、音と映像の印象を独立に操作して系統的に意味的調和の効果を実証してはいない。

音と映像の意味的調和に関する研究は、「印象」といった実験変量の操作が難しく、構造的調和に比べて研究例も非常に少ない。この様な意味で、本項の Bolivar et al. (1994), 岩宮、林 (1999), 岩宮ら (2002) の取り組みは有意義なものであるといえる。

1.3 本研究について

人は、古くから文字をコミュニケーションのメディアとして使ってきた。今日の映像メディアにおける文字は、「視聴覚言語 (audio-visual language)」的な傾向が強く、動きや変化を持つようになった文字 (moving type or kinetic typography: Woolman & Bellantoni, 2001; 上北, 合原, 三上, 田中, 1998) に音響的要素を付加して有効な視聴覚コミュニケーションを意図している。しかし、あくまでも経験に基づいて文字と音を組み合わせしており、その心的作用について客観的な検証が行われていない。本研究は、我々の日常と関わり深い映像/情報メディアであるテレビのテロップ (1.3.2.1 項参照) と効果音に着目し、異なる感覚モダリティからの情報をどの様に構成すれば調和感がありかつ効果的な視聴覚情報になるのかを、実験心理学的な手法を用いて検討したものである。

1.3.1 Cohen のマルチメディアにおける音と映像の「調和-連合の枠組み (Congruence-Associationist Framework)」

Cohen は、図 1.6 に示すように、マルチメディアにおける音楽の役割を理解するための「調和-連合の枠組み (Congruence-Associationist Framework, 2005)」を提案した。レベル A では、各感覚器で処理された情報を Level B へ送るボトムアップ処理 (bottom up processing) が行われる。レベル B では、レベル A から送られた情報を構造と意味の二つに分析し、各情報チャンネルの様相間の相互作用 (調和の決定) が行われる。レベル D では、長期記憶に保留中の過去の記憶や経験に基づいて Level C へのトップダウン処理 (top-down processing) が行われる。そして、レベル B から上った情

報とレベルDから下った情報は、レベルCの短期記憶で照合される。

このモデルは、音と映像の「連合 (association)」と「構造的調和 (structural congruence)」の二つの現象によって説明されている。構造的調和は、視聴覚の群化の現象を通して特定の視覚情報への注意に影響を及ぼす。レベルBで視覚と聴覚の構造情報が一致すると、音楽が視覚的注意点を決めるようになる。一方、連合は映画の雰囲気やプロットの設定などへ音楽の意味を直接に伝達する要因である。レベルBの音楽の意味情報はレベルCの視覚の短期記憶に伝達されて連合する。この意味の連合は、視覚と聴覚の構造情報の一致によって影響され、構造的調和で制御された注意点到適用されるという。なお、映像的物語 (visual narrative) は、レベルCの音声とレベルBの音楽から情報を得ることによって形成される。

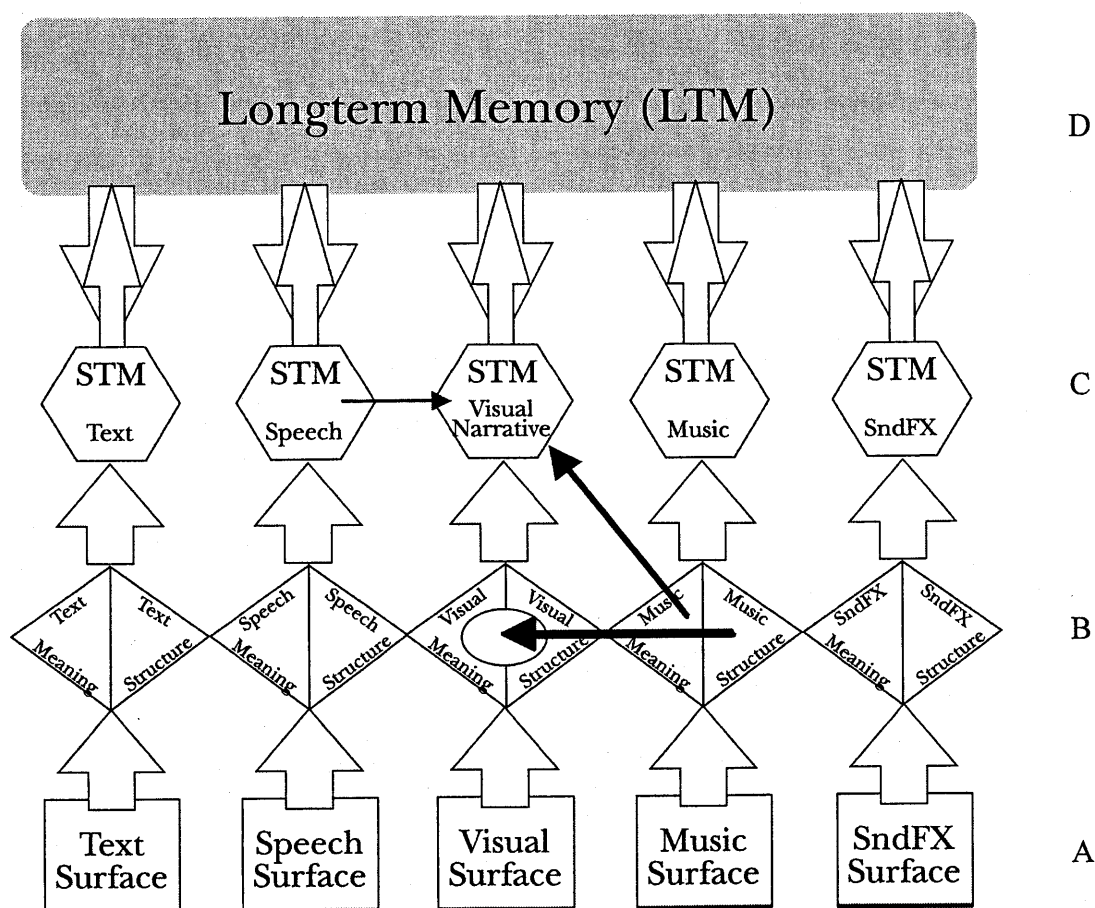


図 1.6 Cohen (2001) の Congruence-Associationist Framework の修正版。LTM, long-term memory; STM, short-term memory. Cohen (2005) より転載

図 1.6 の Cohen の調和－連合の枠組みは、彼女の「調和－連合モデル (Congruence-Associationist Model, 1988)」を基に多様な文献を参照した一番複雑なものである。しかし、その多くは一般的な仮定であり、実証されてはいない。それでも、映像への音楽の影響や意味の生成について興味深い仮説を提供してくれる。

Cohen の以前の「調和－連合の枠組み (2001)」は、音声と映像、音楽の三つのチャンネルから構成されていた。特に、効果音チャンネルが必要であることは認めたものの、研究の不足などを理由として省かれていた。多くの映画理論家 (e. g., Bordwell & Thompson, 1993; Giannetti, 1995; Manvell and Huntley, 1975) も、映画の音を「台詞」「音楽」「効果音」に区分している。図 1.6 の改正されたモデルでは、新たに効果音 (SndFX) とテキストが加わり、五つの情報チャンネルから構成されるようになった。かつて記号学者である Mets (1971) も、Cohen (2005), Stam (2000) と同じく映像メディアにおける情報のチャンネルを五つであると定義しているが、構成要素としての側面から映画の中の音を「話し言葉」「音楽」「モノ音」としている。興味深いことに、これらの過半数を占めているのは視覚的なものではなく、聴覚的なものであり、映像メディアにおける音の重要性を示唆している。その中で、1.2 節で述べたように「映像と音楽」「映像と話し言葉」「映像とモノ音」を実験刺激として用いた研究はあった。特に、映像と音楽の相互作用に関する研究が多かった。しかし、テキストやモノ音 (効果音) も視聴覚コンテンツを構成する重要な要素であるにもかかわらず、これらを実験刺激として用いた研究はなかった。本研究では、特にテキストに組み合わせる音響要素の機能的な側面に注目しており、音の機能要素としての名称である「効果音」を用いることにする。なお、音楽的な表現要素を含んだ音 (e. g., 電子音や楽器音のメロディなど) でも、それ自体で自立しない場合は音楽と呼ばずに効果音とする。

1.3.2 テロップと効果音

1.3.2.1 テロップ (Telop)

テロップ (telop) とは、Television Opaque Projector の略で、縦 10×横 12.5 (cm) の黒いテロップ・カード上に表現された白い文字や図形 (逆の場合も同様) を固定カメラで撮影し、テレビの映像と合成することをいう。スーパーインポーズ (super impose) と呼ばれることもある。最近では、デジタル技術の発展により、コンピュータ上で処理さ

れることが多く、フルカラーや動きのある文字を合成することもできるようになった。日本では、報道、バラエティ、スポーツなど、番組のジャンルを問わず、テロップが多用されている。

テロップは、言語情報としての意味と非言語情報としての意味(印象)を同時に持つ。文字の非言語情報は、種類(漢字、ひらがな、カタカナ)、書体(形、大きさ、配列、線の太さ、濃さなど)、空間使いなどで表される(石田, 2004; 白石, 工藤, 河地, 1984; 藤本, 祖父, 永井, 河原, 佐藤, 佐藤, 2002; 村越, 2002)。この様に、テロップには情報を伝達するための多様な視覚表現が含まれており、本論文ではその基本構成要素を(i) 意味内容 (ii) 動きや変化 (iii) 書体 (iv) 色と考える。その中で、文字の(i) 意味内容は主に言語的な情報伝達の機能を、(ii) 動きや変化 (iii) 書体 (iv) 色は情緒的な情報生成の役割を担っていると考える。

テロップは、文字放送用のテレビ字幕(closed caption)のように詳細な字幕の制約はないようである(テレビ字幕制作ハンドブック, 2002)。実際のテロップは、まずその番組のコンセプトを前提としたデザインからはじまる。そして、短時間で文字を確実に判読できるようにテロップを加工する。そのためには、背景映像とテロップの文字の色の違いではなく、明度差を作り出さなければならない。そこで、テロップの書体(主に、ゴシック系)の色を決定したら、図1.7に示すように明度差をつけたエッジ、シャドウ、「ザブトン」などで強調する。この様な方法で、喋り手の特定、情動の設定、ことばの意味合いなどを考慮しながら作成してゆく。そして、何より大事なことは、動きと表示時間である。その場のテンション、意味合いによってどの様に「イン」させて「アウト」させるかを決定する。また、視聴者がテロップの言語情報を十分に認識できるような提示時間を考えなければならない。よくテロップを読んでいる途中で消えてしまうことがある。視聴者の理解を助けるためには、例えば喋るときに表示されるテロップの場合、喋りと同期して提示するのみならず、喋りが終わった後でもある程度表示させておく必要がある。

読みやすくなる

岩宮研究室

(a) エッジ

読みやすくなる

読みやすくなる

(b) シャドウ

岩宮研究室

(c) 3D

岩宮研究室

岩宮研究室

岩宮研究室

(d) 「ザブトン」(角ザブ、丸ザブ、木ザブなど)

岩宮研究室

岩宮研究室

(e) エンボス



(f) ハーフエリア (高透明「ザブトン」)

図1.7 グラフィックの文字の視認性 (visibility) を高めるためのデザイン方法

1.3.2.2 効果音 (Sound effect)

効果音 (sound effect) とは、番組の中で、その場に適した、あるいはその場の雰囲気高めるために付加する音を指す (新版音響用語辞典, 2003)。効果音は音楽と違って、短くても多様な演出効果を持っている。作曲家が効果音を指定したのは、イタリアの Monteverdi (1567-1643) が最初であるといわれている。日本では、伝統芸能である歌舞伎の「ツケ」から効果音の歴史を論じることが多い。効果音の制作は、実際の環境音を録音して編集する方法, シンセサイザーなどで生成する方法, 実音と似た「生音 (スタジオで作った擬音)」を作る方法などがある。

映像の中の効果音は豊かな表現力を持っており、映像をより印象的なものにする機能を持つ (Whitaker, 1970/1983; 大木本, 1972; 松下, 2003; 大和, 2001)。大和 (2001) は、テレビなどの映像メディアで使われている効果音を (i) 現実音 (ii) 抽象音 (iii) 誇張音 (iv) 無音 (nonmodulation) に分類した。(i) 現実音は、我々の日常生活を取り巻いている音であり、主に映像に映っているシーンの場所 (e. g., 波の音は海) や状況 (e. g., ブレーキ音は事故), 時間 (e. g., 鶏の声は朝) などを設定する。また、現実音は、映像と「非同時性」を持つ画面外 (offscene) の効果に有効である。例えば、遮断機の音がすると画面にはその音源が映っていなくても、近くに踏み切りや線路があることが判断できる。Chion (1985/1993) は、このような機能を持つ音をフレーム外の音 (*offscreen sound*) と定義した。さらに、現実音は、心理描写を強調 (e. g., 心拍音) したり、映像との対位的な手法にも使われる。(ii) 抽象音は、現実音や生音を加工して用いるため、何の音かは判断し難い。そのため、使い方が難しい効果音であり、単なる描写音とは違って表現性の問題を含む音である。主に、登場人物の感情表現/深層心理、象徴的な表現、場面の急展開するところ、重要な要点の提示などに有効である。(iii) 誇張音は、例えば蝶が飛んでいる音や蟻が歩いている音など実際は音が存在しているが、日常的には聞こえてこない音である。主に、映像の要求によって作られる想像上の音であるため、ドキュメンタリーなどの現実的な作品には向いてない場合が多い。アニメーションやCGなどは、音の誇張と独創性が最大に発揮できる分野であるといえる。(iv) 無音は、映像の音を省略することを指し、音を入れた以上の効果を生み出せる場合がある。映像に音を付加するだけでなく、効果的に取り除くことも考えるべきである。特に、映像と高揚した瞬間にカットアウト (cut out: 音を瞬時に絞ること) し

た後の無音は劇的効果を高めるものである。また、無音に次に展開する音を誘起する有効なポイントでもある。

テロップに付加する効果音も種々の演出効果をもたらす。受け手にとって分かりやすく印象的な視聴覚情報を生成するのに有効であると考えられる。テロップの効果音は、主に要点の提示や表現性を高めることを目的に短いアクセント感のある音が多く用いられている。

1.3.3 本研究の目的

一般に、映像/情報メディアのコンテンツは、音と映像といった聴覚情報と視覚情報によって構成されている (Cohen, 2005; Metz, 1971; Stam, 2000)。マザーク効果 (McGurk & MacDonald, 1976) が示したように、人はこうした視覚情報と聴覚情報をそれぞれ個別の事象としてではなく、統合しながらコミュニケーションを行っている。それ故に、効果的な視聴覚情報を構成するには、感覚モダリティ間の情報統合の原理に基づいた情報デザインを行う必要がある。

1.2.3項で概観したように、視聴覚情報をより効果的にするためには、音と映像の主観的調和 (認知的まとまりの良さ) が重要な要素であり、構造的調和と意味的調和の二つの側面がある。既存の音と映像の相互作用に関する先行研究では、実験刺激として具象的又は抽象的な映像に音楽を組み合わせたものが主流であった。しかし、テキストや効果音も視聴覚コンテンツを構成する重要な要素であるにもかかわらず、これらを実験刺激として用いた研究はなかった。本研究では、最も身近な映像/情報メディアであるテレビのテロップと効果音を実験刺激として取り上げ、視覚情報と聴覚情報をどの様に構成すればより効果的な視聴覚情報になるのかについて音と映像の構造的調和と意味的調和の概念を用いて検討する (図 1.8)。



Telop (caption)



Sound effect

視覚情報と聴覚情報をどの様に構成すれば、より効果的な視聴覚情報になるのか。

図 1.8 本研究の概略

1.3.4 本論文の構成

映像/情報メディアにおけるテキスト（文字）の最大の特徴は、それが音響的要素と同じく時間的に変化する点にある。特に、テロップのような単純な映像の場合、音と映像の時間的な同期は欠かせないものと考えられる。それ故に、本論文ではテロップと効果音の構造的調和の効果を検討した後、意味的調和の効果について検討を進める。

第2章の実験1では、制作者が意図して制作したテレビ番組のテロップと効果音を実験刺激として用いて印象評定実験を行い、音と映像の主観的調和に影響する諸要因を探る。実験2では、合成したテロップ・パターンとサウンド・パターンを実験刺激として用い、実験1でテロップと効果音の調和感に影響を及ぼすと考えられた音と映像の時間構造や変化パターンといった構造的調和の要因について明らかにする。また、これらの構造的要因から生じる印象が視聴覚情報の印象に及ぼす影響と、視聴覚情報の主観的調和と印象の関係について検討する。さらに、視聴覚情報の主観的調和へのテロップと効果音の意味的調和の影響があったかどうかを検討する。実験3では、実験2から明らかになったテロップと効果音の構造的調和の要因について、対称となる視聴覚刺激を用いて印象評定実験を行い、種々の構造的調和がどのような要因によって生じているのかについて考察を加える。

第3章の実験4と実験5では、第2章の実験1からテロップと効果音の意味的調和の要因として考えられた文字の書体と効果音の印象の類似に基づく意味的調和の効果を報告する。特に、音と映像の印象の類似によってもたらされる主観的調和と総合的評価（良し悪し）の関係について明らかにする。また、実験6では印象の類似したテロップと効果音を組み合わせた視聴覚刺激の時間構造の同期を崩すことにより、音と映像の主観的調和に対する構造的調和の影響について考察を加える。

第4章の一連の実験では、第2章の実験1でテロップと効果音の主観的調和に影響を及ぼす要因として考えられたことばと効果音の印象の類似に基づく意味的調和の効果について検討する。特に、テロップの言語情報と効果音の印象の類似によってもたらされる主観的調和と総合的評価の関係について明らかにする。

第5章では、第2章から第4章までの内容を総括し、テロップと効果音の構造的調和と意味的調和が音と映像の主観的調和と総合的評価に及ぼす影響について全体的な考察を行う。最後に、本研究の意義と今後の展望について述べる。