



**HAL**  
open science

## Interactions image-son : autour de la notion de congruence

Rémi Adjiman

► **To cite this version:**

Rémi Adjiman. Interactions image-son : autour de la notion de congruence. Le son au cinéma, Artois  
Presse Université, 2010, 978-2-84832-112-7. hal-01638142

**HAL Id: hal-01638142**

**<https://amu.hal.science/hal-01638142v1>**

Submitted on 19 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# INTERACTIONS IMAGE-SON : ATOUR DE LA NOTION DE CONGRUENCE

Rémi Adjiman

*Université de Provence*

**L**e thème des interactions entre les images et les sons n'est plus original principalement depuis que Michel Chion, précurseur en France mais aussi dans le monde, y a consacré de nombreux ouvrages. D'autres auteurs comme Claude Bailblé ou Daniel Deshays, avec des approches différentes et complémentaires, ont fait et font également progresser les réflexions sur ce thème.

Aujourd'hui, la question des relations entretenues par les images et les sons pourrait sembler plus actuelle dans les autres arts audiovisuels, les installations vidéo, le VJing, les scénographies contemporaines... Les dispositifs mis en œuvre sont souvent plus complexes et moins contraints que celui de la traditionnelle salle de projection. Il est pourtant encore prometteur de s'intéresser aux relations qu'entretiennent les sons et les images dans un cadre purement cinématographique. Même si, dans la chaîne de production filmique, les rapports image-son se figent dès que les sons sont positionnés dans la fenêtre d'édition du montage – ce qui n'est pas le cas des œuvres ou des performances où le spectateur peut interagir avec le dispositif – il existe encore de nombreuses voies de réflexions à investir pour saisir la richesse des perceptions que ces rapports génèrent.

De plus, la dialectique des rapports image-son reste en évolution. D'une part, le champ des possibles continue sans cesse de s'ouvrir en même temps que le cinéma croise le « monde » du film expérimental et s'enrichit de ses pratiques. D'autre part, les techniques cinématographiques se transforment et l'évolution des

outils de la production numérique offre des possibilités renouvelées d'interaction<sup>1</sup>.

Finalement, même si de nombreux théoriciens l'ont critiqué pour mieux en préciser les limites ou les incohérences<sup>2</sup>, la notion de contrepoint orchestral appliquée au cinéma parlant<sup>3</sup> a ouvert en son temps une problématique encore pertinente aujourd'hui. Les associations de l'image et du son méritent toujours d'être mieux perçues, comprises et expliquées.

Pour définir les relations qu'entretiennent les images et les sons dans les films, il est souvent question de considérer deux principes fondamentaux : la coïncidence temporelle, c'est-à-dire le synchronisme, et la cohérence sémantique, c'est-à-dire la relation de sens, qui s'établit entre les images et les sons (empathie, anempathie, redondance, contrepoint...).

Ces deux types de mises en relation constituent les principaux fondements de l'articulation audiovisuelle. Il est intéressant d'essayer d'en préciser un troisième, moins usuel, qui permettrait de décrire les situations où les images et les sons ne seraient ni exactement synchrones, ni associés par un lien exclusivement sémantique, mais où ils s'associeraient par congruence. Du point de vue perceptif, il s'agit d'une relation qui se perçoit de façon moins immédiatement évidente et immédiate (« à la longue »), parfois de façon irrégulière et instable. La congruence<sup>4</sup> – au sens de convenance, d'être en accord, de « qui va avec » – pourrait alors être considérée comme un critère d'association de l'image et du son.

Nous allons essayer de le clarifier au travers de l'analyse de différentes séquences, sélectionnées pour les perceptions de jonction et de disjonction audiovisuelle qu'elles proposent.

---

<sup>1</sup> Les systèmes numériques ne cessent d'évoluer et de permettre de manière paradoxale d'accroître la productivité de tâches sans cesse plus complexes et simultanément d'autoriser, voire de favoriser des possibilités nouvelles de création.

<sup>2</sup> Jean Mitry et Michel Chion, en particulier, ont remis en cause l'idée selon laquelle la logique du contrepoint, qui consiste à écrire plusieurs mélodies superposées, destinées à être entendues simultanément, pourrait s'appliquer à l'image et au son.

Jean Mitry, *Esthétique et psychologie du cinéma*, Paris, Editions du Cerf, 2001

Michel Chion, *La Musique au cinéma*, Paris, Fayard, 1995

<sup>3</sup> Le manifeste du contrepoint orchestral a été rédigé en 1928 par S. Eisenstein, V. Poudovkine et G. Alexandroff

<sup>4</sup> Appliquée à l'audiovisuel et au cinéma, c'est Jacques Jouhaneau, professeur titulaire de la chaire d'acoustique du CNAM et enseignant à l'école Louis Lumière, qui le premier m'a parlé de l'effet de congruence.

Nous notons que cette démarche, qui va du phénoménologique au texte, est parfois « coincée dans l'infiniment petit » du cours d'action de la projection et de la description des données filmiques. Notre approche ne conduit pas à appréhender le film dans une perspective globale et contextuelle, plus propice à comprendre comment le film prend sens dans sa relation à la situation sociale, culturelle, historique ou comment les données filmiques rentrent en résonance avec les intentions manifestes, déposées par l'auteur.

Ici, nous nous intéressons, tout en restant à un niveau premier d'analyse du récit, à décortiquer la structure filmique et les relations entretenues par les images et les sons.

### **Premiers exemples de congruence**

La congruence image-son se matérialise déjà dans certains des tous premiers films sonores, parmi les plus inventifs il est vrai ! Deux d'entre eux, sortis en 1931, peuvent servir à mettre en évidence le phénomène : *M. Le Maudit*<sup>5</sup> et *Enthousiasme : la Symphonie du Donbass*<sup>6</sup>.

Pour le film de Fritz Lang, la congruence est présente dès la séquence d'introduction. Le générique se termine sur une longue séquence, avec un écran noir et le silence pendant plus de 10 secondes. Une voix scandée, d'un niveau faible d'abord, fait graduellement irruption avant que le cadre ne se dévoile et nous situe dans une cour, au pied d'un immeuble modeste. Des enfants jouent. La scène est tournée en plan-séquence, filmée en plongée. Une fillette placée au centre d'un cercle d'enfants chante ou plutôt égrène en cadence les paroles découpées d'une comptine. Les mots tombent à espace régulier, notes blanches sur une partition. Simultanément, elle pivote sur elle-même pour pointer du doigt d'un geste franc, un à un ses camarades. Elle s'interrompt brièvement à la fin de chaque phrase, reprend son souffle, ménage peut-être le suspense du jeu. Lorsque le texte court de la chanson se termine, la fillette fait sortir du cercle l'enfant désigné.

Ici, le principe même de ce jeu – c'est-à-dire l'étroite articulation entre le rythme de la diction et celui du geste de désignation – crée les conditions de la congruence. Les mouvements répétés et saccadés de la fillette viennent s'associer de manière congrue avec le rythme cadencé de ses paroles. Les uns et les autres se renforcent dans la perception du spectateur. Une force d'entraînement et d'implication dans le récit se dégage de cette

---

<sup>5</sup> M, Fritz Lang, 1931

<sup>6</sup> *Entuziazm : Simfonija Donbassa*, Dziga Vertov, 1931

première scène, certainement renforcée par cette relation étroite entre l'image et le son. La mise en scène, en privilégiant le plan-séquence, a soigneusement mis en évidence cette sensation en nous donnant à voir le geste, dans la continuité du déroulement de l'action, au sein du dispositif du jeu et avec les personnages. Le jeu joyeux et entraînant dénote avec les paroles de cette comptine macabre, mais l'effet de congruence contribue à les associer, crée un battement prégnant qui – au-delà du narratif – mobilise et implique le spectateur. En une scène, le spectateur est saisi, les événements à venir sont introduits...

Le premier film sonore de Dziga Vertov, *Enthousiasme : la Symphonie du Donbass*, est un film de commande réalisé entre 1929 et 1931. Il est incroyablement précurseur et inventif, en particulier si l'on considère les variations opérées sur les interactions image-son<sup>7</sup>. On y trouve des sons synchrones, des sons désynchronisés, des sons superposés à l'image, des sons inversés (par rapport au sens de diffusion de l'image), des sons ralentis et accélérés... De très nombreuses séquences, en particulier celles de la mine et des aciéries mettent en scène des situations où s'observe la congruence, en utilisant des sons industriels produits par l'action des mineurs et des ouvriers métallurgistes ou par les mécaniques des machines-outils.

Les sons contribuent souvent à symboliser l'immense courage et le dévouement des hommes prêts à tenir des objectifs de production élevés pour la cause socialiste. Dziga Vertov semble alors chercher, en particulier dans la deuxième partie du film, un procédé audio-visuel qui permette de célébrer la force de l'action collective en valorisant la rigueur de l'organisation des tâches, en glorifiant la vitesse et la précision d'exécution des ouvriers, en magnifiant l'outil de travail.

Un exemple est celui des jets de vapeur produits par les machines<sup>8</sup>. Formes mouvantes qui se détachent sur le ciel, ils sont animés d'une force largement amplifiée par la pulsation rapide du son qui vient se coupler sur les volutes propulsées par la pression. Le plan fixe dure près de 15 secondes, mais la congruence se met en place rapidement. Le son – mais est-ce manifeste ? – n'est pas celui du jet d'eau vaporisé, mais celui de la machine elle-même (ou celui d'une autre machine). « Mécaniquement » les mouvements dans l'image – une image saccadée probablement tournée à 16 ou 18

---

<sup>7</sup> Charlie Chaplin lui-même adressa un message de félicitations à Vertov après avoir assisté à la projection de son film à Londres.

<sup>8</sup> Nous retiendrons en particulier celui de la 39<sup>e</sup> minute du film.

images par seconde<sup>9</sup> – se couplent avec les variations cycliques courtes du son.

Plusieurs effets en découlent. D'abord en effet phénoménologiquement perceptible de renforcement de la puissance de ce jet de vapeur qui va jusqu'à illusionner les perceptions et faire apparaître dans l'image elle-même une fluctuation de la pression qui suit le rythme que le son nous fait entendre. Ensuite, survient un effet d'amplification des significations, de puissance de la force de travail symbolisée par la puissance de ce jet d'eau vaporisée ; au-delà encore, dans le contexte encore plus général du film, apparaît la grandeur du collectivisme et du projet socialiste.

L'origine étymologique du mot congruence est latine<sup>10</sup>, mais l'un des usages de ce terme se situe dans le champ des mathématiques. Deux nombres entiers  $a$  et  $b$  sont dits congruents modulo  $n$ , si leur différence est divisible par  $n$ , c'est-à-dire s'il existe un entier  $k$  tel que  $a - b = kn$ .<sup>11</sup>

Ce qui nous concerne ici est d'une part l'existence d'un lien arithmétique qui unit ces deux nombres et d'autre part le fait que le coefficient multiplicateur  $k$ , entier quelconque, autorise une certaine souplesse quant aux conditions de résolution de cette équation. Si «  $a$  » et «  $b$  » sont congruents modulo  $n$ , le lien qui unit «  $a$  » et «  $b$  » n'est pas fixe et unique, il est proportionnel à  $n$ .

Jusqu'à un certain point, l'analogie mathématique peut devenir instructive lorsqu'on l'applique à l'image et au son. Si le «  $a$  » et le «  $b$  » représentent l'image et le son, la congruence se concrétise alors en un lien entre les éléments de la matière filmique. Comme pour la congruence arithmétique, la liaison qui les unit est moins contrainte qu'un lien figé, comme pourrait l'être celui qui caractérise le synchronisme parfait.

La congruence autorise la mise en œuvre d'une relation plus élastique, plus souple et moins contrainte que la mise en relation temporelle de deux événements instantanés.

Ainsi, dans les deux exemples de plans filmiques précédents, la connexion qui unit l'image et le son est bien réelle, mais ne nécessite pas la même exigence temporelle que celle qui unirait deux

---

<sup>9</sup> La symphonie du Donbass n'est pas un film parlant – qui nécessite une caméra avec un défilement à 24 images par seconde – il s'agit plutôt d'un film sonorisé.

<sup>10</sup> Congruere : se rencontrer, être en accord.

<sup>11</sup> [http://fr.wikipedia.org/wiki/Congruence\\_sur\\_les\\_entiers](http://fr.wikipedia.org/wiki/Congruence_sur_les_entiers)

événements synchrones et qui conduirait à l'effet de synchrèse<sup>12</sup>. La congruence constitue un lien moins rigide, plus malléable.

Néanmoins, l'analogie mathématique a des limites, les nombres entiers ne contiennent pas comme les images ou les sons un pouvoir d'expression figuratif. Alors, comment décrire davantage la congruence entre les images et les sons ? Comment expliciter ce lien spécifique qui pourrait ne pas s'appuyer pas sur une relation de parfait synchronisme, ni sur une correspondance sémantique ?

## **Partie 2 : Une relation qui s'instaure dans la durée**

Pour poursuivre notre démarche, nous décrivons deux nouvelles situations de congruence filmique.

Dans les contes de la lune vague après la pluie<sup>13</sup>, Genjuro – l'un des deux personnages principaux – veut gagner beaucoup d'argent en vendant ses pots à la ville.

Une séquence du début du film crée un moment de très forte congruence entre les mouvements dans l'image et la musique. Au noir, la musique traditionnelle japonaise s'élançe et nous attire vers elle, nous propose d'adopter son rythme frénétique. L'image survient en fondu, une seconde plus tard. Les mouvements, circulaires, sont ceux des bras de Miyagi, la femme de Genjuro, durant la fabrication des pots sur son tour de potier. Les gestes rapides d'entraînement du volant d'inertie du tour viennent s'associer sur le tempo rapide des percussions japonaises. Très vite, un régime de jonction s'instaure entre les images et les sons. Les deuxièmes viennent chronologiquement se fondre sur les premières. L'aimantation s'instaure. La mécanique de la perception bimodale, véritable embrayage se met à relier le moteur – ici probablement le son – et la transmission, les images. La congruence s'établit, dynamise le geste, renforce l'impression de productivité, rend crédibles les perspectives de gain financier.

Un bol est déjà tourné, Miyagi s'interrompt, mais Genjuro l'exhorte à poursuivre. A ce moment, le jeune fils du couple s'immisce et vient demander à manger à sa mère, ralentissant ainsi momentanément le mouvement de ses bras. Ces changements du rythme de la gestuelle créent une authentique fracture. Un interstice

---

<sup>12</sup> Notion introduite avec grand succès par Michel Chion et qui caractérise l'impression de très forte interdépendance produite par la synchronisation de deux événements l'un audio et l'autre visuel.

Michel Chion, *L'audiovision*, Paris, Nathan, 1990, page 55

<sup>13</sup> *Ugetsu monogatari*, Kenji Mizoguchi, 1953

perceptif apparaît et s'accroît entre le rythme visuel ralenti et celui de la musique qui, elle, maintient son tempo entraînant.

La même séquence visuelle avec seulement le son direct conduirait certainement à une perception bien différente. La scène n'aurait probablement pas été emballée de la sorte et la sensation de décalage qui se crée au moment de la désaccélération n'aurait pas été comparable, au mieux perçue comme un simple ralentissement. La rupture de la relation de congruence crée un interstice, une cassure bien plus intense que si la congruence audio-visuelle n'avait pas eu lieu.

Dans une Histoire vraie<sup>14</sup>, Alvin Straight, chevauche sa tondeuse tractée à travers le centre est des USA sur plus de 380 Km entre le Wisconsin et l'Iowa. Ce périple se caractérise par la lenteur de sa progression. A plusieurs moments, en particulier lors du premier départ d'Alvin, la relation musique-image renforce cette langueur par un travail de congruence. Mais, il n'y a pas ici de réelle mise en correspondance formelle entre le tempo musical et des cycles repérables dans le mouvement de l'image (ce qui était le cas dans les exemples précédents). La musique transmet bien aux images les qualités<sup>15</sup> de son tempo lent<sup>16</sup>. Mais dans le sens inverse, la prise de vue subjective en plongée, le cadrage serré sur l'asphalte, la progression lente du travelling faisant avancer les lignes jaunes continues et pointillées au rythme des pas, interagissent subtilement avec la nonchalance du rythme de l'enchaînement des notes musicales. L'image contamine également ; mais loin d'un rapport de stricte mise en relation rythmique, elle diffuse ses caractéristiques propres sur la douce et rêveuse mélancolie de la musique country. L'ensemble se scelle dans un rapprochement progressif du visuel et du sonore. Le film se met à battre sur un rythme lent qui ne s'instaure qu'au fil du déroulement du plan-séquence, une fois les échanges entre images et musique, opérés. La musique et les images

---

<sup>14</sup> *The straight story*, David Lynch, 1999

<sup>15</sup> Lorsque nous utiliserons ce terme de qualité, il est à comprendre au sens de Peirce, c'est-à-dire de caractéristiques potentiellement présentes et perçues, dans les sons et/ou les images observés (les qualités du signe).

Sur cette question :

Nicole Everaert-Desmedt, *Le processus interprétatif. Introduction à la sémiotique de Ch. S. Peirce*, Liège, Pierre Mardaga, 1990.

<sup>16</sup> Michel Chion nomme *valeur ajoutée*, la valeur sensorielle, informative, sémantique, narrative, structurelle ou expressive qu'un son entendu dans une scène nous amène à projeter sur l'image.

Michel Chion, *un art sonore, le cinéma*, Paris, Nathan, 2003, page 191

seulement superposées, mais en aucun cas synchronisées par l'intermédiaire d'événements saillants, se confondent dans notre perception du sensible et nous transmettent les qualités de la lenteur.

Parmi les quatre exemples que nous avons décrits précédemment, certaines expériences de congruence audiovisuelle peuvent être très facilement accessibles. C'est en particulier le cas lorsque le son comporte une rythmique frappante sur laquelle les images – et plus particulièrement les événements, les mouvements présents dans les images – s'associent en symbiose. Cela est encore certainement plus vrai encore, si le son apparaît antérieurement aux images et que ces dernières viennent se poser puis rapidement, s'aimer dessus (c'est le cas pour deux de nos exemples).

Néanmoins, contrairement à la synchronèse (qui met en jeu des événements visuels et sonores brefs) et à la sensation immédiate qu'elle produit, par exemple lors de la synchronisation d'un effet<sup>17</sup> sonore comme un claquement de portière, des bruits de pas ou un coup de feu, la perception d'un moment de congruence audiovisuelle nécessite davantage de temps pour être ressentie. La congruence n'est pas accessible aussi instantanément, c'est-à-dire dès l'instant même où l'image et le son cohabitent. Il est nécessaire d'une part de laisser le temps au son de s'établir, de le laisser se faire entendre, d'exprimer et manifester ses qualités, sa pulsion ou ses variations. D'autre part, consubstantiellement, il faut aussi parvenir à saisir ou plutôt simplement à ressentir<sup>18</sup> le moment où l'image donne des signes de son échange avec le son.

La congruence naît de l'inscription au sein d'un « temps d'actualisation commun »<sup>19</sup> d'un événement sonore et d'un événement visuel de même durée. Au cours du temps où la mise en relation s'opère, chacune des expressions sonore et visuelle se renvoie l'une à l'autre. Cette durée peut être courte, mais pour que ce mouvement prenne son ampleur et sa réelle valeur, un intervalle de temps plus long que l'instant est nécessaire, dépendant

---

<sup>17</sup> Le terme d'effet provient de la classification utilisée dans la production audiovisuelle. Il qualifie un son bref généralement synchronisé avec précision à l'image.

<sup>18</sup> Il ne s'agit pas d'une mise en signification consciente, mais plutôt d'un processus préconscient. Pour le dire comme Peirce, nous restons ici au niveau de la priméité. L'interprétation se maintient à un niveau de sensation et non de fait (c'est-à-dire sans pouvoir identifier clairement l'objet de ces sensations, ni même chercher à le faire).

<sup>19</sup> Nous avons repris cette formule à Vincent Ciciliato, dans son mémoire de DEA Vincent Ciciliato, L'objet visio-sonore, Paris, Université Paris VIII, dir. Jean-Louis Boissier, 2003

intrinsèquement de la durée du cycle présent dans le son et de son couplage avec l'image.

Ainsi un simple fondu sonore progressif et régulier, qui soulignerait la disparition graduelle d'un objet bruyant présent dans l'image ne pourrait – à lui seul – établir un effet de congruence. La congruence s'opère lorsque certaines des caractéristiques intrinsèques, mouvantes et cycliques des images et des sons se mettent finalement à interagir et à rentrer en résonance. C'est cette dernière, qui qualifie la dimension phénoménologique de la congruence.

Une fois la congruence mise en évidence, l'articulation entre le sonore et le visuel génère ses effets et renforce la sensation produite par l'un et l'autre des éléments de la matière filmique. Il n'est alors plus possible de dire lequel du son ou de l'image est premier, l'un et l'autre tendent à se confondre et, pour le dire comme Deleuze, deviennent *indiscernables*<sup>20</sup>.

Cette relation, et la fusion perceptive qu'elle engendre, est plus complexe – et potentiellement plus créative – que la seule synchronisation. Elle n'est pas binaire c'est-à-dire n'a pas comme seule perspective d'être ou de ne pas être. Elle peut être plus ou moins perçue, portée par une imbrication audio-visuelle aux effets assurés ou beaucoup plus fragile et incertaine, se matérialisant par moments et s'effaçant à d'autres. Laissant à certains spectateurs, mais pas à d'autres, la possibilité de la ressentir.

### **Partie 3 : Une relation audio-visuelle plus ou moins étroite**

Dans le jour où la terre s'arrêta<sup>21</sup>, le film commence par l'arrivée d'une soucoupe volante. Contrairement aux nombreux « films de soucoupes volantes » sortis durant ces années-là, celui-ci est le seul où l'extra-terrestre vient de l'espace pour sauver la terre plutôt que pour la menacer. L'arrivée de la soucoupe volante, à partir du moment où elle commence à se manifester jusqu'au moment où elle se pose à Washington, dure près d'une minute et trente secondes. Sa présence est d'abord révélée par le son, en hors-champ. Sur une succession de plans fixes, situés au cœur de Washington, un son continu apparaît, synthétique, avec une variation du profil de masse<sup>22</sup>

---

<sup>20</sup> La notion d'indiscernabilité fait référence à l'image-temps.

Gilles Deleuze, *L'image-temps*, Paris, Editions de minuit, 1985

<sup>21</sup> *The day the earth stood still*, Robert Wise, 1951

<sup>22</sup> Nous reprenons ici un des sept critères morphologiques de la typologie définie

rapide et marquée. Son irruption est progressive, d'abord masquée par le bruit du passage de deux automobiles et accompagnée ensuite par un très lent fondu graduel et régulier qui ne le laisse pas percevoir durant les deux ou trois premiers plans.

Dans les plans suivants, devenu pourtant présent et perceptible, ce son ne parvient pas à s'associer naturellement à la scène visuelle. Il ne fusionne pas encore avec l'image des plans visualisés, il ne semble pas encore réellement faire partie de la diégèse de ces plans courts de passants qui déambulent et commencent à se retourner en observant le ciel.

Ce son étrange, qui va se poursuivre toute la séquence durant, est l'œuvre de Bernard Hermann lui-même. Tel un *sound designer*, il détourne deux instruments atypiques : les ondes musicales Martenot et le Theremin qui produisent cette sonorité céleste et permettent de faire varier ses pulsations propres. Il utilisera par ailleurs ce dernier instrument dans toute la musique du film, pour lui communiquer son étrangeté. L'insolite, compte tenu de la rareté des usages contemporains de l'un et l'autre des instruments, subsiste encore aujourd'hui pour le spectateur.

Cette nappe sonore ne commence à prendre réellement de la place dans notre conscience, qu'à partir du moment où un point lumineux – tâche blanche elliptique – apparaît dans l'immensité du ciel derrière le dôme du Capitole. C'est le moment où les badauds et touristes, déambulant devant le Smithsonian museum, aperçoivent le vaisseau qui progresse au-dessus des toits de la ville. Presque simultanément le son s'amplifie, se met à osciller en pulsations régulières, en trémolo. S'associe alors une deuxième composante sonore, plus proche celle-là du bruit produit par la tuyère d'un avion à réaction. Cette seconde couche sonore, plus modulée encore, vient se fondre sur la première, sans néanmoins en modifier la structure d'ensemble, la régularité. C'est un premier palier d'évolution. L'OVNI traverse le ciel de gauche à droite, sans soubresaut, mais au contraire avec une grande constance ; le point lumineux glisse et se rapproche. Le son évolue, crée une dynamique, renforce le mouvement, pourtant l'OVNI n'a pas changé d'allure dans son déplacement. Le cadrage reste également le même, nous le suivons toujours au loin. Rien dans l'image elle-même ne justifie réellement cette légère variation. Ce qui change, en fait, c'est principalement

---

par Pierre Schaeffer pour décrire les objets sonores. Le profil de masse est le profil du son, dont la masse est « sculptée » par des variations internes.  
Traité des Objets Musicaux, Pierre Schaeffer, Paris, Editions du Seuil, 1966

que, dans le récit, les promeneurs, maintenant tous les promeneurs, présents dans les rues scrutent le vaisseau et suivent sa trajectoire.

Lors du contrechamp suivant – un père et sa fille, assis sur le gazon, se redressent ébahis – la première composante change de tonalité, devient plus aigue. A compter de ce plan, et pendant que les passants accompagnent du regard le déplacement de l'aéronef qui traverse par deux fois l'écran presque de bout en bout, le son devient par moments plus saccadé tout en conservant une composante continu qui maintient l'impression de fluidité du déplacement. Seule évolution dans la présence visuelle de la soucoupe elle-même : une oscillation lente, « vivante », de l'intensité lumineuse de l'ellipse blanche.

Jusqu'à ce moment et depuis les bruits de circulation automobile, dernières manifestations sonores des humains dans la ville, seul est présent le son continu de l'objet volant, s'approchant inexorablement. Aucun son direct, ni aucune musique ne vient troubler la séquence, permettant de rendre audible les subtiles variations du son du vaisseau. L'évolution légère du son du mouvement de l'OVNI évolue bien avec son déplacement, son rapprochement, mais le panoramique sur les toits de la ville n'occasionne finalement que peu de variations dans les valeurs de plan visuel. L'évolution légère de cette nappe sonore est bien davantage congruente avec la prise de conscience progressive des passants, elle renforce l'idée que peu à peu, de plus en plus d'humains observent l'intrus avec angoisse.

Durant cette séquence, le son tente d'être cohérent par rapport au déplacement de la soucoupe volante à l'origine du son entendu. Il est céleste, régulier, continu. Mais le véhicule de l'espace est vu par les passants, principalement en plan lointain comme un ovale lumineux sans détail, qui traverse le ciel de gauche à droite régulièrement sans fluctuation de vitesse, ni changement de direction particulier, ni rapprochement significatif qui justifierait de créer des variations notables dans le bruit émis par l'engin.

La congruence entre le son et le mouvement de la soucoupe reste donc peu marquée, assez lâche, seulement entretenue par la correspondance des trémolos sonores et des oscillations lumineuses.

Finalement, le rapport image-son s'appuie principalement sur un lien de nature sémantique, cette nappe synthétique doit s'entendre et se reconnaître comme étant « naturellement » produite par l'engin volant.

Par contre, les variations sonores mises en œuvre cherchent davantage à établir une interaction avec l'évolution de la narration, à exprimer la tension progressive produite sur ce public des rues ébahi. Le principe fondateur de la séquence et sa construction par paliers successifs, basée sur un montage qui alterne les plans sur le déplacement de la soucoupe et les autres sur la surprise, puis l'angoisse de la population, ne laisse que peu de doute sur l'intention de l'auteur. Le public et la terreur qui se répand en lui, préparent déjà la suite du récit, c'est-à-dire les réponses inappropriées des humains au message pacifique de l'occupant de l'OVNI. C'est ce message qui intéresse Robert Wise<sup>23</sup>.

Il est intéressant d'analyser – plus de 50 ans après – les usages possibles des sons associés aux engins de l'espace, pour en analyser la congruence. C'est dans l'épisode II de la guerre des étoiles<sup>24</sup> que nous avons trouvé un moment où le travail du son, dans son rapport à l'image, autorise cette analyse<sup>25</sup>.

La séquence observée rompt totalement avec la précédente. Loin du parallélisme faiblement congruent des séquences visuelles et sonores du film de Robert Wise, l'approche du *sound design* de Ben Burtt développe au contraire un travail d'élaboration des variations sonores qui se calquent au plus près sur les mouvements et déplacements présents dans l'image.

Dans la séquence choisit, une poursuite entre deux véhicules spatiaux, un premier vaisseau s'approche et grossit dans le cadre de l'image puis s'éloigne, après un mouvement panoramique de la caméra de 180°. Ce passage de l'astronef dure moins de six secondes et s'accompagne d'un travail permanent sur les variations de texture, sur l'usage des basses fréquences, sur la simulation du traitement acoustique associé à l'éloignement visuel de la fin du plan. Plusieurs couches de son très différentes s'associent pour donner corps à cette machine.

L'arrivée du vaisseau survient en deux secondes et au cours de ce moment, le vrombissement médium au grain accentué s'amplifie fortement, associé à l'action croissante d'un effet doppler. Une fois à

---

<sup>23</sup> Sans savoir si un usage congruent de la relation image-son y sera développé, il aurait été intéressant de comparer cette séquence avec celle du remake qui sortira à la toute fin de l'année 2008. Il n'est pas encore à l'affiche au moment de la rédaction de cet article.

The day the earth stood still, David Scarpa, 2008

<sup>24</sup> Star Wars: Episode II - Attack of the Clones, Georges Lucas, 2002

<sup>25</sup> Je tiens à remercier Clément, mon fils, pour avoir visionné sa collection personnelle, en quête d'une séquence adaptée.

proximité, les tuyères se font davantage entendre et voir. Une composante aigue se mêle au premier bourdonnement et juste avant que le vaisseau ne soit au plus près, elle s'estompe rapidement. Survient alors le basculement de l'axe caméra puis l'éloignement. Ce dernier est associé à un changement de sonorité. Au début du mouvement, le son du moteur à réaction domine. Mais, rapidement, il s'atténue et est remplacé par une nouvelle sonorité beaucoup plus vibrante. En s'éloignant, l'engin interstellaire conserve cette signature, mais l'apport de la réverbération l'accompagne au fur et à mesure que le point lumineux s'amenuise dans l'espace et se rapproche de la planète orange. Survient alors un souffle qui caractérise l'absence du vaisseau et la présence – toute cinématographique – du vide intersidéral. Arrive alors le poursuivant...

Au cours de ce plan-séquence de six secondes environ, pas moins de six sons différents, probablement plus, sont utilisés successivement et simultanément, calés minutieusement à des moments précis. Le mixage, plus complexe encore, a consisté à adapter de façon dynamique les réglages à chacun des instants, presque à l'image près. Le travail d'intégration des couches sonores ne s'est fait que dans son rapport instantané aux images, de même que le réglage des traitements sonores.

Curieusement, à l'exception du bruit de moteur à réaction facilement associé à la vision des fusées de propulsion de l'engin spatial, une écoute attentive des autres sons utilisés ne fournit pas de lien sémantique évident avec l'astronef. Le couplage est principalement obtenu par la concomitance des variations visuelles et sonores. La congruence est portée à son paroxysme en s'adaptant aux mouvements présents dans l'image, en les renforçant et même en les suggérant.

Le son s'adapte ici aux nouvelles possibilités des prises de vue réalisées avec des maquettes, puis intégrées dans des images de synthèse, permettant ainsi des mouvements de caméra inimaginables (même si ici nous n'avons affaire qu'à un mouvement panoramique de 180°, particulièrement rapide tout de même !). Nous nous situons très précisément au cœur du cinéma décrit par Laurent Jullier dans l'écran post-moderne<sup>26</sup>.

Au-delà du travail de *sound design*, sur le timbre, sur le grain, sur la tessiture des sons, le point crucial est que les déplacements

---

<sup>26</sup> Laurent Jullier, *l'écran post-moderne : un cinéma de l'allusion et du feu d'artifice*, Paris, L'Harmattan, Champs Visuels, 1997

présents dans l'image et les variations aménagées dans le son parviennent à se rapprocher dans une relation d'extrême finesse. Alors que nous n'avons pas – à proprement parlé – de point de synchronisation ponctuel et saillant, les rapports entretenus par les images et les sons sont, bien davantage que pour l'extrait du film de Robert Wise, principalement verticaux<sup>27</sup>. Cette dimension verticale ne s'exprime pas exactement dans l'instantanéité d'événements ponctuels synchrones, mais dans la mise en correspondance d'une variation évolutive de la trame sonore qui s'aimante et s'unit étroitement avec les images. Les images et les sons sont moins synchrones que ne le sont les mouvements ou les variations présents dans le visuel et le sonore. Le travail rendu possible sur les interactions image-son<sup>28</sup> atteint aujourd'hui son potentiel d'articulation et d'expression le plus élevé (nous ne plaçons pas notre propos sur un plan artistique).

Par l'analyse de ces deux séquences, la congruence audio-visuelle peut maintenant se qualifier davantage.

Pour la première séquence, la congruence s'établit dans la durée, elle est diachronique et ne produit pas de forte perception d'imbrication et de mise en correspondance. Elle est relayée par l'interaction de deux variations, de deux mouvements audio-visuels légers, celui du scintillement de la soucoupe et celui du trémolo. Nous pouvons la qualifier de congruence lâche.

La seconde, très verticale, crée un lien fort qui s'actualise et s'entretient en permanence par une mise en correspondance de

---

<sup>27</sup> Sur la question des rapports verticaux et horizontaux, nous nous sommes référés aux nombreuses discussions au sujet de la validité des propositions apportées par le manifeste du contrepoint orchestral.

Sur les rapports verticaux et horizontaux :

Michel Chion, Un art sonore, le cinéma, Paris, Cahiers du Cinéma, Essais, p. 192-195

<sup>28</sup> Sans que cela soit notre objectif prioritaire, il est possible de préciser par quelles évolutions ergonomiques ou techniques, l'informatisation et l'introduction des stations de travail a accompagné cette évolution :

Une généralisation de la représentation temporelle sur la *timeline*

Une introduction de références temporelles communes pour l'image et le son (time code)

La précision extrême du montage

La possibilité de contracter ou d'étirer temporellement les sons

L'évolution de l'ergonomie des interfaces homme-machine, des éditeurs et des *plug-in*

L'évolution de la chaîne de production (le *workflow*) qui permet aujourd'hui, au moment du mixage, de revenir sur certains points du montage son.

variations image-son intimes et sans cesse réactivées. A l'autre bout de l'axe, nous pouvons qualifier cette congruence d'étroite.

#### **Partie 4 : Une imbrication des deux mouvements, sonore et visuel**

Nous avons vu, pour la dernière séquence analysée, que la mise en relation image-son s'appuie sur l'élaboration d'une concordance temporelle étroite entre d'une part les variations rapides présentes dans l'image, rapprochement, proximité et éloignement, et d'autre part l'évolution du son utilisant de petits éléments travaillés par touches successives.

Faisons le point sur les différentes situations rencontrées.

Dans les séquences de *M le Maudit* et des *Contes de la lune vague*, il apparaît que la mise en relation du rythme de la diction ou du rythme musical<sup>29</sup> avec un mouvement de rotation cyclique, présent simultanément dans l'image conduit à une imbrication forte.

Dans le film de Vertov, la situation est différente. Le mouvement présent dans l'image est celui des volutes de vapeur qui sortent de la cheminée. L'impression de cycle reste présente dans l'image, mais plus du tout celle de rotation. La congruence s'exerce et le rythme de la mécanique s'aimante parfaitement avec les mouvements dans l'image.

Dans le cas d'une Histoire vraie, le déplacement dans l'image est rectiligne et la notion de cycle n'est plus produite que par la vision des lignes jaunes pointillées. Il n'est pas certain que ces dernières aident nécessairement à rendre la musique congrue. La mise en phase s'opère peut-être davantage à cause de la vitesse lente de défilement de la chaussée, dont la texture est vue en gros plan. Le mouvement de caméra, qui « panote » vers le tracteur d'Alvin Straight ne modifie en rien la perception puisque la lenteur du déplacement reste perceptible. La congruence s'atténue totalement lorsque la caméra s'élève et fixe – à l'horizon – le ciel nuageux.

La séquence du *Jour où la terre s'arrêta* est un cas de congruence beaucoup moins marquée. Une particularité tient au double mouvement : celui présent dans l'image avec le déplacement de la soucoupe volante et celui de l'image elle-même avec le panoramique de la caméra qui nous fait suivre la soucoupe au-dessus des toits de Washington. Or, le son ne cherche pas à marquer le déplacement de la caméra. Par quelques modifications subtiles, il

---

<sup>29</sup> Et par extension de toutes formes sonores périodiques comme celle que l'on pourrait par exemple utiliser en faisant rouler une bille dans un seau ou une assiette.

atteste par contre de son intervention à deux niveaux. Dans la relation avec la narration qu'il renforce à des moments clés. Dans la relation avec l'oscillation de l'ellipse lumineuse par l'intermédiaire d'un trémolo dont la fréquence varie. C'est ici que la congruence s'installe pour animer la soucoupe et la rendre vivante.

L'astronef de la *Guerre des Etoiles* est bien différent, la congruence y surgit dès qu'un laps de temps suffisant nous permet de ressentir dans le son, l'impression de déplacement et de rapprochement. Un laps de temps suffit et le son rentre en relation immédiate avec le mouvement rapide du vaisseau spatial, soutenu ensuite par le changement d'axe soudain et l'éloignement.

Dans l'ensemble de ces situations, le dénominateur commun est la mise en correspondance des deux mouvements, mouvements présents à la fois dans l'image et dans le son. Cette mise en correspondance peut s'opérer par la similitude des rythmes visuels et sonores, par la jonction entre la vitesse d'un mouvement dans l'image et le rythme d'un son ou l'évolution progressive et simultanée du son<sup>30</sup>.

Cette mise en relation conduit à la perception d'un mouvement unique, résultat du couplage audio-visuel<sup>31</sup>.

Dans ces situations, nous notons que voix, musique ou effets ont pu être utilisés. La congruence ne s'appuie pas seulement sur le rythme musical.

Dans tous les cas, la synchronisation d'événements ponctuels n'est pas nécessaire, même si les rapports verticaux, entretenus par une concomitance habile des mouvements, participent grandement de la fusion entre les éléments visuels et sonores.

## **La congruence est encore imprécise**

Notre objectif était de contribuer à dégager, en plus des notions de synchronisme et de cohérence sémantique, un autre critère qui pourrait décrire la mise en relation du son et de l'image.

---

<sup>30</sup> Dans d'autres séquences, nous avons pu voir que le changement dans l'image pouvait être une variation de l'intensité lumineuse sur laquelle venait s'aimer – après compréhension de la situation – le bruit difficilement identifiable d'un volet roulant.

Lost Highway, David Lynch, 1997

<sup>31</sup> Pour l'effet de synchrèse, les anglo-saxons utilisent le terme « perceived as one » c'est-à-dire perçu comme une entité unique.

La notion de congruence semble bien être capable de décrire le fait d'une mise en correspondance formelle de l'image et du son au cours d'un certain laps de temps.

Elle n'est cependant pas indépendante de la coïncidence temporelle, ni de la cohérence sémantique. Elle s'appuie sur l'un et l'autre pour exister et produire ses effets, mais favorise grandement les rapports verticaux. La congruence ne peut être considérée indépendamment de la notion de mouvement et de durée, mais elle s'intéresse davantage à mettre en correspondance des évolutions, des changements que réellement à caler de façon parfaitement synchrone des événements visuels et sonores. En aucun cas, elle ne peut être définie comme un troisième critère indépendant.

Elle exprime la façon même dont certaines articulations audiovisuelles parviennent à exister par un renforcement formel mutuel : parfois par une accumulation de mises en relation coordonnées de « petites » variations sonores et visuelles, parfois par une association rythmique véritable stimulation énergétique pour le spectateur... Une existence que l'on peut à la fois décrire lorsque l'on dissèque la matière de l'expression<sup>32</sup> filmique, mais que l'on doit également chercher à comprendre lorsque l'on considère le travail opéré par le spectateur.

Néanmoins, la formalisation de la congruence est encore incomplète. Il est encore difficile de lui donner une définition précise. De plus, il semble que différentes formes de congruence puissent exister.

Une congruence rythmique, qui parviendrait certainement à s'affranchir d'une mise en relation temporelle instantanée, mais travaillerait plutôt sur l'établissement d'un rapport entre des cycles, de mouvements visuels et sonores imbriqués.

Une autre congruence qui s'établirait davantage par une logique d'association des mouvements et de convergence des déplacements. Il semble, par exemple, envisageable d'appliquer à l'audiovisuel les lois de la Gestalt, initialement définies pour le domaine visuel seulement. La loi de destin commun – loi qui précise qu'un sujet a tendance à regrouper des éléments disparates en une forme unique lorsque ces éléments partagent les mêmes trajectoires – paraît prometteuse pour comprendre plus encore la congruence. Il faut néanmoins vérifier qu'elle peut s'appliquer à la perception multimodale et aux interactions de l'image et du son en particulier.

---

<sup>32</sup> Cette expression est empruntée à Roger Odin  
De la fiction, Roger Odin, Bruxelles, De Boeck, 2000

Au niveau de l'analyse des données filmiques, il serait intéressant de travailler encore plus finement sur les séquences en observant les caractéristiques précises de concordance et de synchronie des événements sonores et visuels. Par exemple, une extrême précision sur l'emboîtement des cycles des différents mouvements simultanés doit apporter des enseignements complémentaires.

Pour finir, même si cela n'a pas été l'objectif prioritaire de ce texte, il est essentiel de considérer le potentiel d'expression de cette association image-son. La congruence est avant tout repérable par ces effets phénoménologiques, il est donc essentiel d'interroger les rapports que peuvent entretenir des moments identifiés de congruence avec les perceptions sensibles rencontrées. Cela conduit à investir davantage le champ des possibles de l'interstice audio-visuel.