

L'attention auditive dans la mise en scène des plans-séquences au cinéma

Transcrire les modulations attentionnelles
du personnage dans le mixage

Directeur interne : Sylvain Lambinet

Directeur externe : Éric Tisserand

Rapporteur : Laurent Millot

REMERCIEMENTS

Je remercie chaleureusement Sylvain Lambinet et Éric Tisserand, mes directeurs interne et externe, pour m'avoir dirigé et accompagné sur ce projet.

Je remercie Liza Lamy pour le soutien indéfectible et l'aide apportée dans la rédaction de mon mémoire.

Je remercie Victor Barbier, Guillaume Pradel, Hugo Corbel, Malo Thouément et Liza Lamy de m'avoir aidé à réaliser ma partie pratique de mémoire.

Je remercie ma famille pour le soutien dans mon travail, et pour les relectures et corrections apportées.

Je remercie plus largement la promotion 2020 avec qui j'ai partagé ces trois années d'étude semées de plaisir et d'embuches.

Je remercie toutes les personnes du métier rencontrées ou aperçues qui m'ont permis de plus larges réflexions sur les métiers du son au cinéma ; à commencer par Séverin Favriau, qui a eu la gentillesse de me parrainer, mais aussi Benjamin Viau, Valérie Deloof, Raphaël Sohier, Pascal Villard, Alexis Meynet et d'autres que j'oublie.

Je remercie enfin tous les professeurs, intervenants, responsables administratifs ; et l'école Louis-Lumière dans sa globalité pour m'avoir permis d'étudier dans un milieu épanouissant.

RÉSUMÉ

Les mécanismes de l'attention auditive façonnent notre perception du monde sonore. Les facultés de sélection et d'inhibition des données sensorielles constituent des outils utiles pour la bonne compréhension de notre environnement. Ces modulations attentionnelles sont le reflet de notre conscience au monde, et témoignent de notre état d'esprit immergé dans un contexte.

Le dispositif cinématographique s'empare alors de nos aptitudes perceptives et cognitives pour mettre en scène le sonore. La compréhension de notre attention auditive fournit des possibilités multiples dans ces choix de mise en scène. Au-delà des précautions liées au dispositif cinématographique, les pratiques techniques relèvent d'une écriture expressive du champ attentionnel. Notamment, l'inscription des modulations attentionnelles du personnage dans le mixage apparaît avoir des incidences sémantiques et narratives. Elles semblent influencer sur notre perception de l'état d'esprit du personnage et de l'environnement extérieur.

Pour répondre à cette hypothèse, nous établirons un test psychoacoustique au cours duquel des sujets, profanes comme professionnels, seront soumis à la diffusion d'un plan-séquence présentant différents mixages motivés par différents parcours attentionnels. Ils devront répondre par la suite à plusieurs questions concernant la perception du personnage et de son environnement fictionnel. Nous analyserons enfin les réponses pour déceler les différences significatives.

Mots clefs : attention auditive, sciences cognitives, psychoacoustique, cinéma, mixage, mise en scène sonore

ABSTRACT

The mechanisms of aural attention modulate our perception of the sound environment. The abilities to select and inhibit sensory data constitute useful tools for a thorough understanding of our surroundings. These attentional modulations are the reflection of our awareness in the world and testify to our state of mind when immersed in a particular context.

The cinematic apparatus then seizes our perceptive and cognitive skills to stage the sound field. The understanding of aural attention provides plenty of possibilities in these staging choices. Beyond the technical precautions of the cinematic apparatus, the technical practices fall within an expressive writing of the attentional field. Therefore the inclusion of attentional modulations of the character in the mixing appears to have semantic and narrative implications. They seem to affect our perception of the character's state of mind and of his external environment.

To support this assumption, we will establish a psychoacoustic test in which layman and professional subjects experience the broadcast of a single sequence. The sequence would be attributed to different mixings driven by different attentional paths. Afterwards, the participants will have to answer several questions about the character's perception and his surroundings. Eventually we will analyse the answers to reveal the significant differences.

Keywords : aural attention, cognition, psychoacoustic, cinema, mixing, sound direction

TABLE DES MATIÈRES

REMERCIEMENTS	2
RÉSUMÉ	3
ABSTRACT	4
TABLE DES MATIÈRES	5
INTRODUCTION	7
PARTIE I. L'ATTENTION AUDITIVE	9
A. Le démasquage binaural	13
B. Les mouvements de l'attention auditive	15
1. Les mécanismes ascendants	18
2. Les mécanismes descendants	21
C. L'intentionnalité de l'écoute	24
D. La saillance auditive	27
1. La saillance physique	27
2. La saillance cognitive	30
Conclusion	34
PARTIE II. LE CHAMP ATTENTIONNEL AU CINÉMA	36
Introduction	36
A. Modification du champ attentionnel face au dispositif cinématographique	37
1. Le champ frontal et le champ ambiant	37
2. L'étagement des sons	38
3. Le cas du plan-séquence	40

B. Expressivité du champ attentionnel en plan-séquence	44
1. Le point d'écoute	45
2. Les modulations attentionnelles	59
PARTIE III. PARTIE PRATIQUE DU MÉMOIRE	69
A. Protocole du test	69
1. Les stimuli	70
2. Le questionnaire	71
B. Mise en place du test	72
1. L'écriture du plan	72
2. Le tournage	73
3. La post-production	74
C. Analyse des résultats	78
1. Les conditions d'écoute	78
2. Les parcours attentionnels	80
3. La perception du champ organisme/environnement	83
4. La perception de l'état d'esprit du personnage	88
CONCLUSION	91
ANNEXES	94
A. L'analyse de la scène auditive	94
B. L'effet de masque	98
C. Valeurs de l'effet de démasquage binaural	100
D. Les dispositifs de diffusion multicanale	101
E. Les résultats du test psychoacoustique	104
F. Liens internet des stimuli	110
BIBLIOGRAPHIE	111
FILMOGRAPHIE	115

INTRODUCTION

Mon mémoire s'inscrit dans la rencontre entre les sciences psychoacoustiques et cognitives d'une part, et le domaine cinématographique d'autre part. La compréhension de notre perception du monde sensible, et notamment du champ visuel et sonore questionne évidemment le dispositif cinéma, fait d'images et de sons. Au delà du développement des dispositifs techniques pour essayer d'immerger le spectateur dans une scène fictive, la mise en scène se nourrit de l'expérience du réel pour alimenter son champ des possibles. Hors, dans le domaine sonore, la perception nous permet d'ouïr le monde sensible de manière commune (sans possibilité d'y échapper contrairement à la vue) tandis que l'attention semble façonner la perception de notre environnement en fonction d'une individualité immergée dans celui-ci. Ce caractère singulier, à l'inverse de la simple perception auditive, peut alors être l'initiateur d'une démarche cinématographique ; comme le souligne Brian de Palma :

« On ne raconte jamais une histoire par l'intermédiaire de la caméra. C'est toujours à travers des gens, à travers ce qu'ils voient et ce qu'ils ressentent. »¹

De Palma, Brian

Je compte alors étudier les mécanismes de l'attention auditive pour déceler les incidences et les possibilités de mise en scène du champ sonore.

Si le dispositif cinématographique n'a pas pour essence de transmettre une illusion du réel et serait d'ailleurs bien en retard en comparaison à d'autres technologies immersives comme la réalité virtuelle, il essaie tout de même, dans sa forme majoritaire, de s'approcher d'une expérience sensorielle convaincante pour porter la narration. L'essor de nouvelles technologies comme la diffusion sonore en Dolby Atmos vont dans ce sens. Mais la forme classique du cinéma porte toujours des biais perceptifs considérables dans cette intention, comme, par exemple, par le découpage du champ image qui apporte des ruptures inhabituelles dans la perception des champs visuel et sonore. Le plan-séquence apparaît alors comme l'unique procédé de mise en scène au plus proche de la perception

¹ DE PALMA, Brian. Cité dans DURAFOUR, Jean-Michel. *Épanchements : sang, perception, théorie*. L'Harmattan, 2013.

du réel, en conservant la continuité spatiale et temporelle. Je circonscrirai alors mon étude au plan-séquence, tout d'abord pour ne pas amalgamer des modulations sonores s'inscrivant dans un parcours attentionnel à des modulations motivées par le découpage image. De plus, le plan-séquence peut permettre au spectateur de se projeter en un « personnage invisible » par la personnification d'un point de vue, à la fois continu et à hauteur d'homme² (contrairement au spectateur omnipotent ayant accès à tous les angles de vue). Alors, il m'apparaît pertinent d'introduire le concept d'attention auditive, propre à l'individu, dans la réflexion autour de la mise en scène du plan-séquence.

Dans un premier temps, l'étude des mécanismes attentionnels nous permettra de comprendre notre appréhension du champ sonore. Puis, l'analyse filmique de plans-séquences nous conduira à comprendre l'intérêt expressif des modulations attentionnelles pour la mise en scène. On questionnera alors la subjectivité de l'écoute des spectateurs, entre une diffusion d'un champ sonore naturaliste à la reconstitution d'un parcours attentionnel inscrit dans le processus de fabrication de la bande sonore. Pour cela, je confronterai différents mixages (motivés par l'état de l'art) d'un même plan-séquence à un large panel d'individus, afin d'y déceler les éventuelles différences de réception du plan (perception du personnage et de son environnement).

² À l'image du plan-séquence d'introduction du film *The Revenant* réalisé par Alejandro González Iñárritu en 2015.

PARTIE I. L'ATTENTION AUDITIVE

L'attention auditive est un mécanisme de sélection d'une cible sonore et, dans le même temps, d'inhibition du champ ambiant. Il s'inscrit dans un processus plus global d'analyse de la scène auditive (Bregman, 1994), comprenant des mécanismes perceptifs et cognitifs. Le système auditif nous donne tout d'abord à percevoir les ondes acoustiques du spectre audible (de 20Hz à 20kHz). Les signaux nerveux sont ensuite décomposés et organisés en éléments ou flux sonores, via des processus de groupement simultané et séquentiel³, pour aboutir à une description mentale du champ audible. C'est la scission ou la fusion des signaux nerveux en divers représentations mentales (Bregman, 1994). Enfin, l'attention auditive discrimine les éléments au sein de la description de l'environnement sonore préalablement établie.

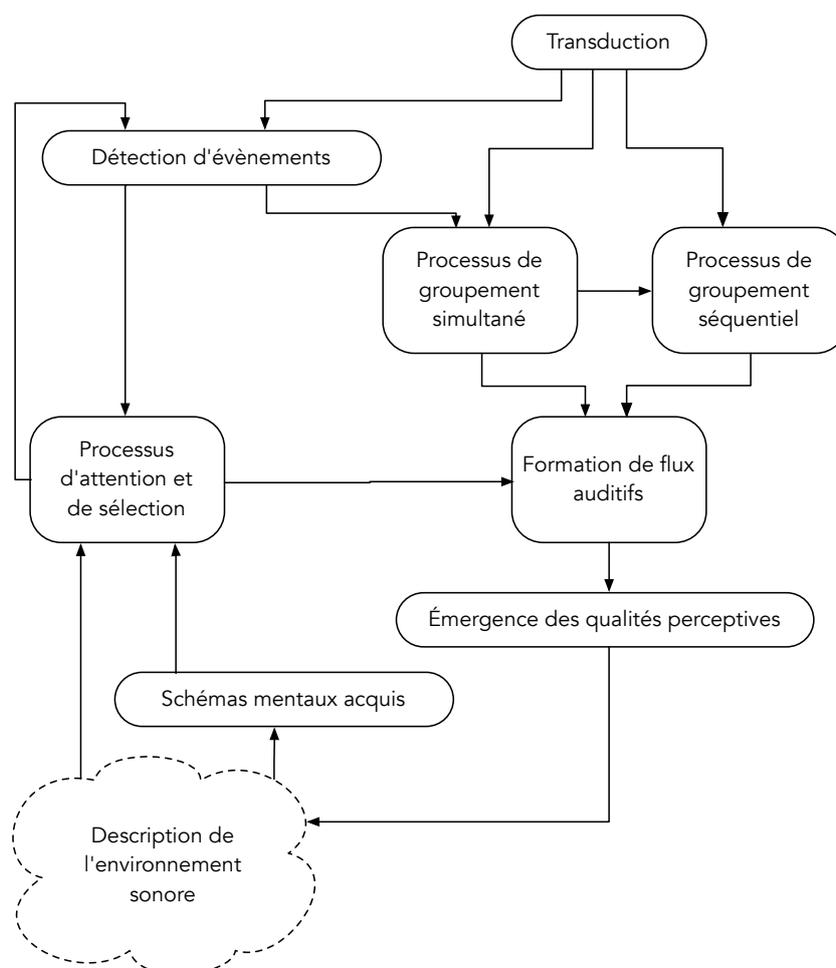


Figure 7 - Schéma de représentation de l'analyse de la scène auditive (McAdams, 1997)

³ Une description détaillée des processus de groupement est disponible en annexe.

La localisation des éléments sonores, via l'écoute binaurale et la fonction de transfert relative à la tête (HRTF, Head-Related Transfer Function), est nécessaire au fonctionnement de ces processus.

L'analyse de la scène auditive permet la bonne compréhension de notre environnement sonore. En formant des représentations mentales des différents objets sonores, nous pouvons ainsi juger de l'identité, du comportement et de la disposition des objets rencontrés ; et de ce fait nous organiser au sein de cet environnement. La bonne connaissance de notre environnement permet une meilleure intégration des différents objets dans ces représentations mentales. De plus, nos systèmes perceptifs se sont ajustés à la structure du monde physique et en particulier aux objets et événements qui nous sont biologiquement significatifs (Shepard, 1981). L'oreille humaine présente par exemple un maximum de sensibilité autour des fréquences caractéristiques du cri du bébé.

Dès l'Antiquité, les philosophes ont remarqué que l'esprit permet de voir les choses plus clairement ou au contraire de ne plus les remarquer. De même, Descartes note qu'une idée apparaît plus claire et distincte quand l'attention se dirige vers elle.

En XVII^e, le philosophe allemand Gottfried Wilhelm Leibniz fut le premier à donner un statut scientifique à la notion d'attention, en proposant le concept d'« aperception » (Meilleur, 2020).

En 1890, le psychologue et philosophe américain William James proposa sa définition du concept d'attention qui fera et fait encore aujourd'hui consensus dans le monde scientifique. Il définira l'attention comme :

« la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui semblent possibles [...] Elle implique le retrait de certains objets afin de traiter plus efficacement les autres ».

(James, 1890, p.403)

Un siècle plus tard, les progrès en imagerie cérébrale fonctionnelle (IRM) permettent des avancées majeures dans la compréhension des mécanismes attentionnels, en confrontant les différents modèles proposés jusqu'alors avec l'expérimentation scientifique en neuroimagerie (Meilleur, 2020).

Aujourd'hui, l'attention - bien que perçue comme sensation unique - est abordée sous plusieurs angles, à la fois comme une fonction de sélection, de distribution de ressources, de régulation du comportement et de contrôle du comportement (Possamaï & Bonnel & Requin, 1993).

L'expérience du monde sensible s'opère d'abord par la perception que l'on a de celui-ci. Elle inclut l'expérience de sa propre existence au sein de cet environnement. Cette perception s'effectue de manière passive et constante. En cela, elle diffère de l'attention qui régit la sélection des informations sensorielles, impulsée par une volonté interne (ciblage) ou par un événement externe (distraction). L'attention se dirige vers une représentation caractérisée par un endroit particulier, un objet particulier ou un moment précis d'apparition (Camus, 1996, p.54). L'attention auditive, comme l'attention visuelle, permet de discriminer dans le champ objet un point, une cible, tout en rejetant le bruit ambiant. L'auditeur focalise alors son écoute de manière plus ou moins vive sur une zone, un objet sonore, qui "se voit" rehaussé en netteté, en présence. L'apport de la vision et la connaissance de l'objet augmente l'efficacité de la focalisation, de la lisibilité (Bailblé, 1999, p.258).

Pierre Schaeffer décrit dans son *Traité des objets musicaux* (Schaeffer, 1966) quatre différents niveaux d'écoute, qu'il résume par cette phrase :

« Je vous ai ouï malgré moi, bien que je n'aie pas écouté à la porte, mais je n'ai pas compris ce que j'ai entendu. »

Ouïr, c'est percevoir l'environnement sonore dans lequel nous baignons. Nous ne cessons jamais d'ouïr tant que le monde ne cesse d'être sonore.

Écouter, « c'est dans le même mouvement détecter, localiser, rehausser la zone sélectionnée, mais aussi atténuer le non-sélectionner. » (Bailblé, 1999). C'est un niveau d'écoute actif qui nécessite l'attention de l'auditeur.

Entendre, c'est, après avoir écouté un signal sonore, le qualifier à l'aune de ses propres connaissances.

Comprendre, c'est déceler du sens dans le flux sonore, en fonction de connaissances linguistiques, bruitistes, musicales... On retrouve ici l'écoute sémantique.

Dans le domaine sonore, la différence entre perception et attention apparaît être la même différence qu'entre ouïr et écouter. Ouïr, c'est percevoir par l'oreille, écouter, c'est proprement faire attention. L'attention auditive nous permet donc de mieux entendre et de mieux comprendre les sons qui nous attirent. Nous pouvons ainsi rester concentré sur une discussion dans un milieu bruyant, c'est l'*effet cocktail party*⁴. Nous pouvons aussi contempler un paysage sonore, en nous baladant dans l'émergence aléatoire des sons, ou bien nous focaliser sur un élément qui attire davantage notre attention du moment. Enfin, bien que dans une attitude de relâchement attentionnel, il nous arrive d'être subitement attiré par un son, par peur, par curiosité, par familiarité entretenue avec ce son... Il semble que le contexte, l'expérience de l'auditeur ainsi que son état émotionnel guident en partie l'attention auditive portée sur le champ sonore, volontairement ou non. Ceci rejoint un principe fondateur de la Gestalt-thérapie :

« A tout moment, tout homme fait partie d'un champ, et son comportement est toujours la résultante d'un champ global incluant lui-même et ce qui l'entoure. »

(Perls, 2003, p.33)

A. Le démasquage binaural

L'effet de masque (dont les principales caractéristiques sont à trouver en annexe) est un phénomène acoustique expliquant le masquage d'un son masquant sur un son cible. Il est apparu que cet effet dépend notamment de l'alignement temporel du son masquant sur le son cible (antérieur, simultané, postérieur) et de la relation binaurale entretenue entre eux, autrement dit de leurs localisations respectives. Les mécanismes binauraux agissent donc dans la sélectivité spatiale des sons, c'est le phénomène de **démasquage binaural**⁵.

Cette notion provient de différentes études menées indépendamment dans les années 1940. Des travaux (Langmuir et coll., 1944 ; Hirsh, 1948) basés sur les sons purs ont

⁴ On appelle effet *cocktail party* la capacité à diriger son attention pour suivre un discours ou une conversation dans une ambiance bruyante, par exemple lors d'un cocktail, tout en restant conscient de l'environnement sonore.

⁵ Expliquant notamment nos capacités dans l'*effet cocktail party*.

amené l'appellation de cette notion la plus courante de nos jours : **BMLD** (Binaural masking level differences). D'autres travaux axés sur la parole comme signaux d'étude, mené par Licklider (1948), présentent le terme de **BILD** (Binaural intelligibility level differences). Le démasquage binaural se mesure en dB. C'est la différence entre le seuil masqué en écoute monaurale, et le seuil masqué en écoute binaurale.

Un protocole d'expérimentation proposé par Green en 1976 détaille différents cas de figure de relation binaurale entre son masquant (N) et son masqué (S) (Botte, 1989, p. 113). Il utilise différents sons purs pour la cible et du bruit blanc pour le masque. Le test se fait au casque. Il diffuse des signaux (masqué et masquant) en phase vers les deux oreilles, en opposition de phase entre les deux oreilles, non-corrélés entre les deux oreilles et enfin diffusés vers une seule oreille. Les résultats montrent tout d'abord que le démasquage est de 0dB pour N et S en phase, en opposition de phase, et pour N non-corrélé et S sur une des deux oreilles ; ce qui reste logique avec la définition du démasquage binaural. Les résultats les plus probants atteignent **un démasquage de 13dB à 15dB** pour les signaux N **en phase** et S **en opposition de phase**, et inversement. On observe aussi selon une autre étude (Durlach et Colburn, 1978) que **le démasquage dépend de la fréquence**, et qu'il est plus efficace pour les fréquences inférieures à 500Hz (10 à 13dB en moyenne) et se stabilise après les 2000Hz (3dB en moyenne).

Une autre étude portant sur le démasquage et l'intelligibilité de la parole en fonction de l'azimut du bruit masquant démontre que le démasquage - qui est de 0dB à 0°, servant de référence - peut atteindre 10dB de 50° à 135° environ pour un niveau de bruit masquant s'élevant à 60dB et qu'il redescend à environ 4dB pour un azimut à 180°, soit derrière le sujet (Botte, 1989, p.113). L'ensemble des résultats de ces études sont à retrouver en annexe.

Le démasquage binaural dépend de plusieurs paramètres (Bailblé, 1999, p.257) :

- l'écart entre la bande de bruit masquant et la bande fréquentielle écoutée,
- l'azimut du bruit masquant par rapport à la direction du son écouté,
- la position du son écouté par rapport au plan frontal ; il est plus aisé de se focaliser sur une source frontale que sur une source oblique. Enfin le rayonnement de la source vers l'auditeur augmente nettement son intelligibilité.

L'écoute binaurale nous offre la capacité de démasquer une zone de l'espace, augmentant de fait la lisibilité des sons s'y contenant. On parle de cône de présence. À l'instar de la vision, un détachement figure sur fond s'applique aussi à l'ouïe. Le reste de l'environnement sonore est atténué, éloigné dans « la perspective attentionnelle ». Mais les éléments sonores « non-sélectionnés » bénéficient toujours d'une perception auditive et peuvent à tout moment ressurgir dans le giron attentionnel. Mais il est impossible de porter son attention simultanément sur l'ensemble des événements sonores car nous disposons de capacité attentionnelle limitée (Weil-Barais, 2011, p.134). Il peut être d'ailleurs très fatigant de se focaliser sur un seul flux sonore, quand celui-ci demeure trop enfoui dans un brouhaha. Par ailleurs, on expérimente tous les jours différents niveaux d'attention, selon notre concentration sur un objet cible et selon le contexte donné. La réalité sonore devient subjective pour chaque individu, qui se focalise ou se relâche au gré de ses désirs et de ses attentes.

B. Les mouvements de l'attention auditive

Nous sommes constamment sujets à de nombreuses sollicitations sensorielles perçues par nos systèmes récepteurs mais toutes ne parviennent pas jusqu'à notre conscience. L'attention est définie par notre capacité à discriminer toutes ces informations afin de se focaliser sur l'une d'entre-elles (ou sur plusieurs d'entre-elles). Cette capacité est alors à questionner pour en déceler ses potentielles limites. Exercer deux tâches différentes semble faisable si elles ne mettent pas à contribution les mêmes capacités mentales, comme suivre une discussion en faisant la vaisselle. En revanche, suivre deux discussions à

la fois paraît beaucoup plus difficile. Des études sur la détection et la recherche proposent la notion de ressources conscientes - ou ressources attentionnelles - à disposition pour l'exécution d'une tâche (Weil-Barais, 2011, p.135). Par exemple, l'écoute de plusieurs stimuli sollicite les mêmes ressources, l'attention auditive se retrouve alors partagée et devient donc inévitablement moins efficace⁶. Autrement dit, on doit nécessairement porter notre attention vers une seule représentation pour en optimiser son efficacité. Mais, plusieurs messages de différentes origines sensorielles coexistent (comme la vue et l'ouïe, primordial dans l'extéroception⁷). Dans ce cas, bien que les ressources attentionnelles soient différentes, l'apport de la vue peut améliorer la compréhension de la cible. La focalisation s'accroît⁸ (Bailblé, 1999, p.258).

L'attention auditive, par définition, ne peut donc pas se centrer sur l'ensemble des événements sonores. Elle bénéficie de ressources attentionnelles limitées qui nous contraignent à la sélection. Cette sélection peut s'effectuer volontairement, c'est notamment le cas lorsque nous écoutons un interlocuteur. Elle peut aussi surgir involontairement, comme la sirène de police qui vient rompre avec le continuum sonore urbain. Il apparaît donc une ambivalence entre deux modes d'attention, que résume parfaitement Jean-François Camus :

« Il semble exister une dynamique complexe entre, d'une part, cette direction automatique de l'attention qui permet de surveiller continuellement l'environnement, mais qui peut aussi entraîner une continuelle distraction, et, d'autre part, une orientation volontaire qui organise de manière cohérente et soutenue notre conduite mais qui, en résistant trop aux distractions, peut manquer parfois de flexibilité. »

(Camus, 1996, p.54)

⁶ Colin Cherry expérimente le premier en 1953, avec la technique de *poursuite-répétition* (*shadowing*), la capacité d'orientation de l'attention sur un message parmi un ensemble de messages, et la difficulté à intégrer l'ensemble des messages.

⁷ L'extéroception, par opposition à la proprioception, regroupe les sensations causées par des stimuli extérieurs.

⁸ Par exemple, la focalisation de l'attention à l'écoute d'une voix s'améliore à la vue du mouvement des lèvres.

L'auteur met ici en évidence la cohabitation entre un processus attentionnel automatique et un processus attentionnel conscient, qui, à eux deux, motivent les réorientations ou les fixations de notre attention ; ce sont **les mouvements de l'attention auditive**. Il y a en quelque sorte une concurrence entre l'attraction des sons écoutés et la distraction de l'écoute. Nous allons donc essayer de comprendre les différents mécanismes cognitifs qui participent, en plus du démasquage binaural, aux différents états de l'attention auditive.

Je précise avant tout qu'il semble exister dans la littérature scientifique différents termes pour désigner ces modes d'attention, mais qu'il y a, à ma connaissance, un manque de consensus sémantique. En effet, Jean-François Camus définit *l'attention sélective* comme « *l'attention capturée par un événement surprenant* » en opposition à *l'attention préparatoire* définie comme « *l'attention engagée de manière volontaire* ». L'opposition est ici sur le caractère volontaire de l'attention. Or, Annick Weil-Barais écrit que « *Si [...] votre intérêt porte sur une seule information, on parlera d'attention sélective. Mais si plusieurs informations sont concernées, on parlera d'attention partagée.* ». L'opposition est définie ici par le nombre d'informations sur lesquelles on porte notre attention. Pour la cohérence de la suite de l'étude, j'utiliserai les termes définis par Posner (Posner & Raichle, 1998) :

L'attention exogène : c'est l'attention capturée par un stimulus extérieur de manière involontaire (distraction). Le processus est « bottom-up », ou ascendant, du stimulus à notre conscience.

L'attention endogène : c'est l'attention délibérée, causée par la seule intention du sujet. Elle peut être sélective ou partagée (selon A. Weil-Barais), soutenue ou relâchée. Le processus est « top-down », ou descendant, de notre conscience vers le stimulus.

Je précise aussi que l'on peut tout à fait passer d'une attention exogène (la surprise d'un ami qui vous appelle au loin) à une attention endogène (suite à cette rencontre, la discussion se met en place).

1. Les mécanismes ascendants

La **négativité de discordance**, plus connu sous le nom de **MMN** dans la littérature scientifique, est une onde différentielle générée par la détection de variations des caractéristiques physiques du son. Selon le modèle de Rist Näätänen (1990), l'apparition de ce mécanisme survient d'abord par l'apparition d'une trace sonore régulière au cours du temps. Puis, lorsqu'un son déviant intervient, il est comparé à la trace sonore stockée dans la mémoire sensorielle. Si celui-ci est différent, le mécanisme est activé, l'onde différentielle est envoyée dans le cortex auditif primaire. Plus récemment, la MMN fût décrite par Winkler (2007) comme une réponse à une rupture au sein d'une régularité auditive, comme un changement de caractéristique physique du son (différence d'intensité, de fréquence, de localisation et de durée). Cette déflexion négative est générée de 100 à 250 millisecondes après le début du son déviant (Moreau, 2013, p.4). Selon l'ampleur de l'onde, l'information peut atteindre notre conscience et entraîner une réaction d'orientation de l'attention. Ce mécanisme fonctionne continuellement, de manière préconsciente, sur l'ensemble du champ sonore (assimilable à du traitement d'informations en parallèle). Il fonctionne même pour des variations - d'intensité, de hauteur et de durée - inférieures aux seuils différentiels, mais n'est alors pas suffisamment ample pour atteindre la conscience (Camus, 1996, p.118). Cependant, la MMN peut voir son ampleur diminuée si le sujet est focalisé, en attention soutenue, sur un autre objet, allant même jusqu'à disparaître (Bailblé, 1999, p.271).

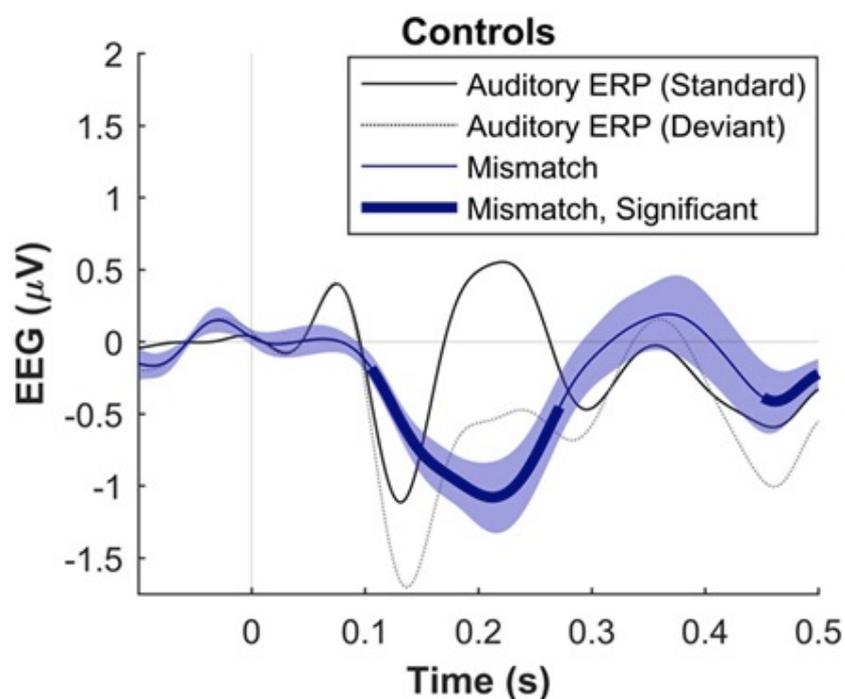


Figure 9 -Graphique présentant un signal standard, un signal déviant, et la MMN résultante
 Iyer, Parameswaran Mahadeva, et al. « Mismatch Negativity as an Indicator of Cognitive Sub-Domain Dysfunction in Amyotrophic Lateral Sclerosis ». *Frontiers in Neurology* 8 (2017)

De plus, nous disposons de **neurones phasiques** sensibles à l'émergence de transitoires dans le continuum sonore. L'apparition d'un nouveau stimuli entraîne une réponse des neurones phasiques vers le cortex auditif. Si le stimuli persiste, la réponse s'amointrit jusqu'à s'éteindre (Sauleau, 2017) . En ce sens, les neurones phasiques apportent des informations de vitesse et de fréquence de la stimulation sonore. Ce mécanisme réagit automatiquement aux variations des niveaux d'énergie et peut entraîner momentanément une réorientation de l'attention auditive (Bailblé, 1999, p.347).

Dans la pratique cinématographique, ces échelles de temps de réaction de l'attention (de l'ordre de la dizaine de millisecondes pour la perception de premières qualités auditives (Bailblé, 1999, p.290) à la centaine de millisecondes pour la réaction à une modulation ascendante) sont difficilement assimilables. En effet, un mouvement de potentiomètre trop rapide peut se faire sentir, et donc porter l'attention du spectateur, non

pas sur l'élément sonore qui capte l'attention, mais bien sur le dispositif technique lui-même. Cela devient contre-productif. Alors, comme le suggère Bailblé (1999, p.333) : « Le mixage y simule le parcours attentionnel selon un mode d'appréhension *sériel ralenti*, respectueux des apparences sonores. S'interdisant le coup de potentiomètre rapide, il développe les champs attentionnels dans une succession. ». La transcription des modulations attentionnelles dans le mixage doit alors s'adapter, et ne pas traduire strictement les modulations attentionnelles et leur constante temporelle.

L'attention auditive exogène est régie par la négativité de discordance (la MMN) et les neurones phasiques. En cela, et au vu des différentes caractéristiques de ce mécanisme, il apparaît que ces mécanismes fonctionnent comme un système de surveillance du champ sonore complet nous entourant. La MMN travaille en boucle courte par comparaison d'éléments successifs, usant pour cela de la mémoire sensorielle (mémoire échoïque), de manière préconsciente. Bien qu'ils travaillent continuellement (ne serait-ce que pour « rafraîchir » notre présence au monde), ces mécanismes attentionnels automatiques n'utilisent ni les ressources attentionnelles dédiées aux fonctions descendantes (sinon nous serions dans la plus grande difficulté à porter notre attention dans un environnement foisonnant d'événements sonores), ni la mémoire de travail. Aussi, l'ensemble des variations enregistrées n'est pas forcément assez significatif pour atteindre d'une part notre conscience, et d'autre part entraîner une réorientation spontanée de notre attention auditive. Autrement, nous serions constamment dans le « zapping » attentionnel, dans l'impossibilité de faire preuve d'attention soutenue étant contraint à réagir à une infinie succession d'événements.

Il est alors nécessaire de comprendre ce qui fait « contrepoids » à cette attention exogène, qui, à elle seule, nous plongerait dans la distraction éternelle. D'expérience, la bonne connaissance du milieu dans lequel on évolue et l'expérience personnelle nous permettent de trier dans l'ensemble des sons distrayants ceux qui semblent « mériter » notre attention de ceux qui demeurent inintéressants dans le contexte. En voici un exemple. En France, les klaxons sont synonymes pour moi (et pour une majorité de la population) d'un potentiel danger, et sont là pour prévenir les accidents de la route entre usagers. D'ailleurs, le code de la route stipule qu' « en agglomération, l'usage de

l'avertisseur sonore n'est autorisé qu'en cas de danger immédiat. Les signaux émis ne doivent pas se prolonger plus qu'il n'est nécessaire. »⁹, cela pour éviter les nuisances sonores. Dans ce contexte, un klaxon perçu assez proche entraîne spontanément chez moi une réorientation réflexe de l'attention. Or, lors d'un voyage à Mumbai en Inde, mon attention fût immédiatement portée sur les innombrables sons de klaxon qui rythmaient les rues. Me voilà transformé en « girouette » comme décrit par Claude Bailblé (2004). La nuisance sonore était fatigante. J'ai compris par la suite que l'usage du klaxon en Inde signifiait plutôt la présence d'un automobiliste sur la route, qui voulait doubler, prévenir qu'il est là, faire accélérer la voiture devant... Bien loin de l'usage restrictif que l'on en a en France. Ayant compris la subtilité, les klaxons m'apparaissaient alors comme attendus dans le décor, moins signifiants dans ce contexte indien. Ils ont donc été petit à petit filtrés de mon champ attentionnel, par apprentissage d'un nouveau milieu.

2. Les mécanismes descendants

L'attention auditive se structure par définition dans un double mouvement de focalisation sur la cible et de rejet des distracteurs. Mais si les mécanismes binauraux permettent de détacher dans un premier temps notre cible du fond masquant, ils ne peuvent empêcher les irrptions soudaines de sons distrayants d'attirer notre attention (exogène donc). Il existe alors un mécanisme de niveau supérieur qui intervient après la simple perception du champ sonore pour encoder différemment les informations selon leur pertinence. Cette caractéristique s'acquiert en partie par l'expérience et anime notre attention. Ce mécanisme régit notre attention sélective et nous permet notamment de rester concentré sur un flux sonore, de manière volontaire. Il a été étudié sous le nom de filtrage sensoriel¹⁰.

Une première théorie de filtrage développée par Broadbent (Broadbent, 1958) suggère qu'il existe un canal unique de traitement qui interviendrait très tôt dans le processus perceptif, et que de surcroit nous ne pouvons traiter qu'une seule « entrée

⁹ Article R416-1 du code de la route

¹⁰ Une première expérience menée par Hillyard et al. (1973) évoque des mécanismes de filtrage et de gain sensoriel.

perceptive » à la fois. Par la suite, des expériences¹¹ (Moray, 1959) ont démontré que l'ensemble du champ sonore était traité de manière préconsciente. Treisman (Treisman, 1960) propose alors une modification de cette première théorie en théorie de filtrage par atténuation. Le filtre ne bloque pas le traitement de l'information non-pertinente mais l'atténue seulement. Finalement, le modèle de **filtrage et de gain sensoriel** de Hillyard (Hillyard et al., 1973) est celui qui fera date. Il reste cohérent aujourd'hui et corrobore avec différentes mesures en neuroimagerie, bien que certains débats scientifiques semblent demeurer¹².

Le filtrage sensoriel permet d'effectuer la pondération des informations sonores perçues (selon leur intérêt, leur pertinence à notre égard) après leurs traitements sensoriels et sémantiques, via un mécanisme cognitif de haut niveau (Bailblé, 1999, p.268). Cette pondération est alors nécessaire pour éviter une surcharge des ressources attentionnelles, dans le « goulet d'étranglement » décrit par Annick Weil-Barais (Weil-Barais, 2011, p.134). La sélection des informations semble intervenir plus tardivement, après que l'ensemble des événements sonores ait reçu une signification, selon la théorie de la sélection tardive (Bidet-Caulet, 2007, p.60) . L'attention sélective s'établit en fait par le processus d'accès à la conscience et à la mémoire de travail d'une information sensorielle. Elle se structure en deux processus distincts. D'une part, un filtrage (supra-auditif) accordé aux différentes zones formantiques du son cible (assimilable à un filtre multi-bande) participe au détachement de la source (jusqu'à 18dB) (Botte et al., 1994). Ce mécanisme de facilitation attentionnelle est observable dès 100ms après l'apparition d'un flux pertinent (Bidet-Caulet, 2007, p.65). D'autre part, un filtrage négatif (supra-auditif) permet de rejeter les flux non-pertinents (jusqu'à 18dB aussi) (Botte et al., 1994) dans un temps de réponse minimum d'environ 250ms (Bidet-Caulet, 2007, p.68). L'inhibition est donc plus tardive que la facilitation. Par contre, il peut y avoir interférence entre un son cible et un son perturbateur

¹¹ L'expérience consistait à faire écouter deux flux sonores en situation d'écoute dichotique. Les sujets devaient alors porter leur attention sur un seul flux et en décrire le contenu en fin d'expérience. Mais ils étaient aussi capable de retranscrire quelques mots entendus dans l'autre oreille « inattentive », notamment lorsque leur prénom fût prononcé dans le flux « non-pertinent ».

¹² Débat sur l'emplacement de la sélection d'informations sensorielles dans la chaîne attentionnelle.

si ils partagent des mêmes zones formantiques¹³. Par ailleurs, ce processus est à mettre en parallèle du phénomène d'adaptation auditive, qui rejette les sons constants de faible niveau (en dessous de 20dB SPL) en s'appuyant sur leurs bandes de fréquences caractéristiques (Botte, 1989, p.33). Cette adaptation s'installe en une minute et se résorbe en moins de deux minutes après l'arrêt de l'exposition sonore.

Ces mécanismes de filtrage sensoriel, ou filtrage attentionnel, régissent **l'attention endogène**. Ils procèdent davantage en série (contrairement à la MMN et aux neurones phasiques) et utilisent ainsi ses propres ressources pour cibler une entité (attention sélective) ou un groupe d'entités (attention partagée). L'intention de l'auditeur, dans l'acte d'écoute, module ces filtres qui permettent donc une meilleure stabilité dans les mouvements de l'attention auditive. Nous sommes concentrés sur une source cible, qui se voit démasquée dans un premier niveau perceptif par le processus BMLD puis détaché dans un second niveau (dans le cortex auditif) par les filtres attentionnels. Dans le même temps, les sons distracteurs sont filtrés - rejetés - selon leurs saillances physiques et cognitives¹⁴ (cf. Intentionnalité de l'écoute) par le filtrage sensoriel. Aussi, une attention soutenue, ou préparatoire, permet d'optimiser la focalisation sur un flux sonore tout en résistant efficacement aux distracteurs, donc à l'attention exogène. Elle est plus lente à s'installer et nécessite une bonne concentration¹⁵ (Weil-Barais, 2011). On peut toutefois, dans ce cas, passer à côté d'éléments contextuellement pertinents (si la MMN est trop faible à cause d'une concentration trop importante). Finalement, notre niveau d'attention module entre sa part endogène voulue et sa part exogène résultante.

« Chaque contexte fait entendre des sons ordinaires, des bruits familiers, une ambiance connue. Pour ne pas céder à la curiosité automatique, l'auditeur recrute un certain nombre d'images catégorielles prévisibles qui vont filtrer - inhiber - l'attention réflexe. Ces contres-calques

¹³ C'est d'ailleurs pour cela que le casting de personnages fictionnels doit diversifier les timbres de voix, pour pouvoir identifier et porter attention sur chacun d'entre eux sans interférences. La remarque est d'autant plus importante pour le cas de la fiction radio qui ne présente pas d'image pour repérer les différents personnages.

¹⁴ Les sons habituels à un contexte donné sont généralement ignorés de l'attention exogène, qui reste néanmoins poreuse à certains types de sons saillants (physiquement ou cognitivement).

¹⁵ Selon la loi de *Yerkes-Dodson*, un niveau d'éveil trop important fait baisser la performance attentionnelle. (Weil-Barais, 2011, p.134)

(profils, timbres, formants), venus de la mémoire auditive, s'intercalent dans la chaîne attentionnelle et court-circuitent la distraction. »

(Bailblé, 1999, p.278)

Il y a en fait un aller-retour permanent entre le contexte et nos attentes qui détermine et revoit continuellement la saillance des objets sonores perçus ou attendus. Ce lien s'équilibre entre la volonté de porter attention pour comprendre (une situation, une discussion, une oeuvre,...) et la nécessité de surveiller l'environnement. Cette interconnexion s'alimente à très court terme - ne serait ce que pour reconnaître dans l'instant un son inattendu et mystérieux - comme elle peut perdurer dans le long terme, par exemple lorsqu'on scrute attentivement le silence de notre maison après avoir cru entendre des bruits de pas inexplicables (la dualité contexte/affect ne sera résolue qu'une fois le doute levé). Finalement, le cheminement de l'écoute, régi par les mécanismes précédemment expliqués, s'opère en lien étroit avec le contexte extérieur et les affects intérieurs, il relève d'une intention.

C. L'intentionnalité de l'écoute

« Le mot intentionnalité ne signifie rien d'autre que cette particularité foncière et générale qu'à la conscience d'être conscience de quelque chose, de porter, en sa qualité de cogito, son cogitatum en elle-même. »

(Husserl, 1947)

Les mouvements de l'attention auditive proviennent de mécanismes descendants et ascendants comme nous venons de l'étudier. Ces deux processus, par essence, tendent à se neutraliser ; l'un par distraction excessive et l'autre par un engagement attentionnel trop important. La cohérence de notre écoute réside alors en la compensation de ces mécanismes attentionnels selon notre intention. L'usage du filtrage sensoriel positif ou négatif (Claude Bailblé développe la notion de calque et de contre-calque (Bailblé, 1999, p. 278)) est alors le seul outil pour imposer une volonté, un état d'esprit dans la direction du faisceau attentionnel (puisque la MMN et les neurones phasiques relèvent de processus automatiques et préconscients). Cette intention est alors guidée par notre connaissance

personnelle du contexte et par nos motivations d'action. En effet, les recherches en psychologie soulignent l'importance du « set » ou de la préparation contextuelle dans nos choix attentionnels (Bailblé, 2004). Finalement, **l'intentionnalité de l'écoute** relève, par essence, de **mécanismes attentionnels descendants**.

L'intention se concrétise par la catégorisation des différents sons selon leur saillance, leur pertinence, leur intérêt, les émotions qu'ils nous provoquent, etc. On se focalise sur le flux sonore voulu tout en rejetant les sons contextuellement non-pertinents. La bonne connaissance de notre environnement nous permet alors de recruter en mémoire des formes sonores prévisibles pour mieux en faire abstraction (par atténuation sensoriel ou contre-calque) (Bailblé, 2004). Cependant, le système attentionnel reste ouvert à l'émergence de nouvelles formes sonores, ou simplement à des sons inhabituels relativement au contexte. Il y a alors prise de conscience, et nous pouvons momentanément interrompre l'activité attentionnelle pour se concentrer sur l'intrus, ne serait-ce que pour l'identifier ou pour comprendre la situation. De plus, nous pouvons être dans l'attente d'un son et esquisser alors un calque même en l'absence du son guetté. Plus le son nous est familier et plus le calque peut être précis. Plus généralement, la perception d'une figure auditive peut être complétée par un socle de connaissances. L'expérience de la causalité du système physique, les savoirs culturels, les compétences linguistiques, la mémorisation de formes canoniques du sonore nous aident à appréhender le champ attentionnel, à détacher la figure du fond. La reconnaissance sera alors plus efficace¹⁶. Enfin, l'état de vigilance (sur un événement dont on ne connaît pas la teneur) apparaît comme particulier. En effet, on se soumet au seul processus attentionnel ascendant pour pouvoir relever les moindres variations sonores susceptibles de représenter un intérêt. Les ressources attentionnelles ne sont pas focalisées sur une source prédéfinie mais sont disposées à traiter successivement les sons émergents. Une fois la cible repérée, nous pouvons alors essayer de l'identifier sans pour autant baisser notre vigilance.

Le champ sonore, perçu par l'individu, apparaît alors comme une perception subjective et motivée ; contrairement au monde sonore, modelé par les seuls lois de

¹⁶ D'expérience, il m'est facile de reconnaître le son de la voiture de mon père quand il rentre à la maison, bien que la route soit cachée par la haie.

propagation du son. Les éléments de ce champ sont à présence variable, dans un rapport figure sur fond modelé par notre écoute focalisante. Cet échelonnement attentionnel dénote d'une rencontre entre les particularités individuelles et un contexte momentané. C'est la notion de champ organisme/environnement développée par la théorie de la Gestalt (Lewin, 1951). Alors, la pratique du cinéma sonore doit être questionnée. Que donne-t-on à entendre? La bande sonore doit-elle être la rencontre entre le monde scénique et le spectateur ou entre le monde scénique et le personnage de fiction ? Les implications sémantiques et donc narratives semblent considérables.

De manière générale, la focalisation sur une figure auditive dépend de la situation et de nos motivations personnelles. L'intention dans l'écoute est le fait de processus endogènes, modelés par une volonté de hiérarchisation attentionnelle des éléments sonores constitutifs de notre environnement. Cependant, les situations ne sont pas figées et ne répondent pas nécessairement au modèle préconçu que l'on avait établi. Autrement dit, les différentes situations ont leur part d'imprévu, de surprises qui viennent heurter et influencer nos motivations et donc modifier notre écoute intentionnelle. En somme, si celle-ci est le fait de modulations attentionnelles descendantes, elle interagit tout de même avec l'évolution du contexte et les différentes modulations attentionnelles ascendantes qui peuvent venir détourner l'attention de l'auditeur. Ces éléments distracteurs sont alors caractérisés par leur saillance.

D. La saillance auditive

La saillance d'un objet sonore est, par définition, la qualité de celui-ci à attirer notre attention. Cette notion quantifie la capacité de distraction des modulations ascendantes. Elle dénote du degré d'influence sur l'attention du sujet par les objets perçus. Voyons alors de quoi peut dépendre cette notion.

Dans un travail de recherche, Frédéric Landragin (2011) propose une classification des différents critères de saillances visuelle et linguistique. Il ajoute ceci :

« Avec l'hypothèse que l'émergence d'une figure sur un fond est un mécanisme cognitif dépassant le langage oral, écrit ou visuel, il y a fort à parier que rapprocher et comparer les critères de saillance identifiés dans ces différents domaines permettra de mieux appréhender ce concept. De plus, considérer la saillance comme un mécanisme cognitif général [...] permet de mieux préciser la portée de la notion, et par conséquent la nature des critères de saillance. »

Il m'apparaît alors pertinent de m'inscrire dans cette démarche, en considérant la saillance comme un mécanisme cognitif général. Ainsi, par analogie au domaine visuel et linguistique, on peut essayer de rapporter les résultats de sa classification généraliste au domaine sonore. La première distinction sépare la notion de saillance en critères physiques et en critères cognitifs.

1. La saillance physique

« Du point de vue physique, être saillant, c'est ressortir en premier de la perception d'un message visuel ou linguistique compte tenu de la forme que prend ce message, c'est-à-dire des propriétés de ses composants (et compte tenu des caractéristiques physiologiques de l'organe de perception). »

(Landragin, 2004)

Critère intrinsèque aux unités (*facteur lié à l'objet en question*)

Le premier critère est lié à la saillance intrinsèque à l'objet. La forte luminosité du soleil est un exemple. Dans le domaine sonore, cela peut se traduire par un fort niveau acoustique d'émission.

La simplicité (*facteur lié à l'objet en question*)

Le deuxième critère lié aux caractéristiques physiques de l'objet est la simplicité. La forme ronde d'un objet est un exemple, car elle requiert un minimum d'information sensorielle à traiter. Dans le domaine sonore, cela peut se traduire par un son simple à identifier.

Unicité ou singleton (*facteur lié à l'objet en contexte*)

Dans un ensemble d'objets, un objet est saillant s'il se distingue des autres objets par l'une de ses propriétés. Dans le domaine sonore, cela se traduit par exemple par un son à fort transitoire dans un ensemble de nappes sonores.

Isolement (*facteur lié à l'objet en contexte*)

Dans un ensemble d'objets, un objet est saillant s'il est isolé. Dans le domaine sonore, cela se traduit par une spatialisation distincte entre un objet sonore en particulier, et un groupe perceptif composé d'autres objets sonores.

Mise en évidence explicite (*facteur lié à l'objet en contexte*)

Un objet est saillant s'il est mis en évidence dans la structure physique de la scène. C'est par exemple l'éclairage d'un comédien dans une foule de figurants. Dans le domaine sonore, on peut penser au démasquage d'une source en utilisant des égalisateurs de fréquences relativement aux autres sources à « enfouir ».

Construction dédiée (*facteur lié à l'objet en contexte*)

Ce critère est très proche du précédent. Frédéric Landragin prend comme exemple la construction d'une photographie où le personnage se situerait au dessus d'un triangle. On peut imaginer aussi un personnage allongé, où son corps serait en continuité de la ligne d'horizon. Ce critère est plus implicite que le précédent. Dans le domaine du sonore, je pense à une mise en scène laissant par exemple la place centrale (spatialement) à un personnage ou bien à une construction (temporelle) laissant du temps d'expression à un personnage et du silence entre ses phrases.

Placement à un endroit stratégique (*facteur lié à l'objet en contexte*)

Ce critère est encore très proche du précédent. Dans le domaine visuel comme dans le domaine sonore, le centre est l'endroit de perception le plus efficace, en localisation notamment. Pour le domaine visuel, les notions de tiers de cadre et de lignes de force peuvent aider à désigner un endroit stratégique autre que central. Pour le domaine sonore, l'endroit stratégique peut être là aussi temporel ce qui revient à citer le précédent exemple.

La répétition et la symétrie (*facteur lié à l'objet en contexte*)

La répétition d'objets identiques les rend saillants. Dans le domaine du sonore, l'alarme est le meilleur exemple puisqu'attirer l'attention est sa fonction même. La répétition d'un quelconque élément sonore stimule notre mémoire échoïque (mémoire à court terme) qui travaille en comparaison de l'instant passé et du devenir sur une fenêtre de 80 à 300ms (Bailblé, 1999, p.271). En revanche, je ne trouve pas d'analogie pour la symétrie dans le domaine sonore.

La rupture dans une continuité (*facteur lié à l'objet en contexte*)

Une entité continue qui se rompt devient saillante à la rupture. Dans le domaine sonore, ce phénomène est expliqué notamment par la surveillance de l'environnement par la négativité de discordance étudiée plus haut. On peut prendre ici comme exemple l'arrêt de la climatisation. On a tendance à rejeter, par adaptation auditive, le bruit continu et

stable d'une climatisation participant au fond sonore. Mais quand celle-ci s'arrête, la sensation de vide ou de silence capte notre attention. Ce n'est d'ailleurs pas un objet sonore mais bien une représentation (de la fin de la climatisation) qui est saillante ici.

2. La saillance cognitive

« Du point de vue cognitif, être saillant, c'est ressortir en premier de la perception d'une scène compte tenu des représentations mentales de l'utilisateur, de ses intentions de perception, de son attention et de sa mémoire à court terme, de ses expériences personnelles et de ses émotions. »

(Landragin, 2004)

Infraction d'une règle implicite *liée à l'intention (facteur lié à l'objet en contexte compte tenu d'un code connu et objectif)*

Un objet qui n'est pas à sa place selon sa tâche, ou un objet inattendu gagne en saillance. Il y a décalage avec l'intention initiale. Une casserole est plus saillante dans une salle de bain que dans une cuisine. Dans le domaine sonore, c'est par exemple un chanteur d'une chorale qui chante soudainement faux. Sa présence n'en sera que plus remarquée.

Exploitation d'une norme *(facteur lié à l'objet en contexte compte tenu d'un code connu et objectif)*

Dans un contexte qui comprend une suite d'événements, un schéma normé peut se mettre en place. Dans ce cas, la finalité peut devenir quasiment prévisible et devient alors saillante. Dans le domaine sonore, une suite d'événements peut renvoyer à une représentation mentale codifiée. Je pense ici par exemple aux normes musicales dans la construction d'un morceau. Pour certains types de musique, il est facile de prévoir quand le refrain arrive. Il en devient plus saillant (ce qui est certainement le but recherché).

Prédisposition physiologique (*facteur lié au sujet recevant le message*)

Un objet plus proche de soi paraît plus net, plus lisible est de fait plus saillant. De même, un objet se rapprochant de notre ligne de visée gagne en saillance. Dans le domaine sonore, le démasquage binaural ainsi que nos capacités de localisation favorisent l'élément face à soi. Plus il est proche, plus les résultats sont efficaces.

Prédisposition attentionnelle (*facteur lié au sujet recevant le message*)

« L'attention du sujet sur une certaine entité peut augmenter la saillance de cette entité, et, inversement, la saillance d'une entité peut entraîner l'attention du sujet sur celle-ci. » (Landragin, 2004). En portant notre attention sur un objet sonore, nous avons déjà vu que la présence (netteté accrue) de celui-ci augmentait considérablement.

Prédisposition affective liée à la mémoire à court et à long terme (*facteur lié au sujet recevant le message*)

Pour la mémoire à court terme, nous sommes capables de retenir sept éléments (Miller, 1956), voire cinq ou quatre (Rousselet & Fabre-Thorpe, 2003) dans un vécu récent (valable pour la perception visuelle comme pour la perception auditive). La mémorisation de ceux-ci - parce qu'ils étaient saillants - les rend plus saillants lorsqu'ils réapparaissent. On peut penser ici aux sifflements de M dans *M le Maudit* de Fritz Lang (1931). On retient une première fois cet élément pour l'enjeu dramatique qu'il porte. Puis, lorsqu'on l'entend une seconde fois, ce son devient encore plus saillant car nous savons ce à quoi il renvoie.

Pour la mémoire à long terme, il y a d'abord les connaissances culturelles qui peuvent déterminer la saillance de l'élément sonore. Par exemple, une personne de culture francophone sera plus attirée par la langue française, de même pour les cultures musicales, les cultures liées à différentes époques, etc. Ensuite, il y a l'expérience et les connaissances personnelles qui pondèrent les niveaux de saillance liés à des éléments familiers (Bailblé, 1999). En effet, si j'entends par exemple la voix de ma mère, au loin, dans le brouhaha d'une soirée, mon attention se portera vraisemblablement sur celle-ci. Plus un son nous est

familier, plus il nous apparaît saillant (par reconnaissance du profil sonore, du timbre, des formants...).

Cette classification (Landragin, 2004) est une proposition des différents critères de saillance communs aux domaines visuel et linguistique (le domaine linguistique est d'ailleurs une intersection du domaine sonore lorsqu'il s'agit de la parole). L'auteur précise qu'elle n'a pas valeur à être exhaustive. Il émet aussi une critique quant à la structuration de cette classification, dont certains critères se recoupent pour certains exemples (Landragin, 2004, p.93). Par ailleurs, aucune hiérarchisation de ces facteurs n'est établie et l'auteur s'en garde bien tant le contexte et l'état personnel peuvent pondérer tous ces différents critères. Cette classification permet néanmoins d'éclairer cette notion, aux abords de différentes disciplines, pour mieux la caractériser. J'ajoute qu'il est compliqué à certains égards de faire un lien entre la saillance visuelle et la saillance auditive à partir de critères généraux. D'après Hatwell (1994), l'imagerie visuelle renvoie majoritairement à la perception du spatial, du simultané, contrairement à l'imagerie auditive, qui par la succession des événements existe dans le temporel. De ce constat-là, un critère de rupture spatiale pour le domaine visuel doit-il être un critère de rupture temporelle pour le sonore ? J'évoque ainsi les deux possibilités, à la fois spatiale et temporelle, pour esquisser des exemples de saillance auditive.

La généralisation de cette classification permet tout de même d'éclairer le concept de saillance vers d'autres domaines, dont celui du sonore. Elle constitue alors un appui dans la construction de la bande son au cinéma. Les niveaux de saillance des différents éléments constituant la mise en scène sonore peuvent être travaillés (dans leurs agencements spatial et temporel) pour essayer de construire un parcours attentionnel guidant le spectateur. Alors, le montage son et le mixage peuvent relever d'une écriture expressive, au service de la narration. À l'image de l'effet koulechov¹⁷, la mise en relation d'un son captant notre attention avec l'image projetée, au fil de la narration, peut nourrir celle-ci. De plus, nous avons vu que les modulations attentionnelles d'un individu traduisent au final son champ d'intérêt au sein d'un environnement, et par conséquent offrent une possibilité de lecture

¹⁷ Procédé de montage par lequel les spectateurs tirent plus de sens de l'interaction d'un plan (prise de vue) avec un autre plan auquel il est associé, que d'un plan isolé.

de son comportement, de son état d'esprit. La compréhension du personnage de cinéma semble alors pouvoir être étoffée par une hiérarchisation attentionnelle inscrite dans le mixage (en s'écartant *a fortiori* d'un traitement naturaliste de la bande son). Cela suppose dans un premier temps que le spectateur amalgame, consciemment ou inconsciemment, les modulations de mixage avec les modulations attentionnelles supposées du personnage de fiction ; puis que la perception de ces modulations entraînent la sensation, la compréhension de son état d'esprit¹⁸. Pour la suite de l'étude, je vais me concentrer sur les possibles apports expressifs de ces modulations, rattachées à une individualité (que ce soit pour transmettre des informations, des sensations sur le personnage de fiction ou bien transmettre une sensation au spectateur, quelquefois positionné en tant que « personnage invisible » par le cinéaste¹⁹).

Enfin, si la notion de saillance se subdivise pour les différents champs perceptifs, elle résulte *in fine* de la simultanéité des sensations extéroceptives. Prenons l'exemple d'une explosion, dont la saillance n'est pas à démontrer. Le son de la détonation s'accompagne généralement d'un flash lumineux suivi par l'éclat des débris. Aussi, elle engendre certainement une vibration du sol, une sensation de déplacement d'air et d'une vague de chaleur. Alors, la saillance d'un tel événement ne peut être réduite aux seuls sons, aux seules images produites ; et de manière générale aux seules perceptions isolées. Elle forme un tout perceptif, *l'engramme*²⁰, décrit comme une « boule » de sensations par Michel Chion (2005, p.99). L'étude de la saillance en tant que mécanisme cognitif général prend alors ici tout son sens. La perception de l'objet passe par différents canaux sensoriels mais aboutit au final à une seule représentation mentale (que chacun s'approprie), supérieure à la simple addition des différentes perceptions isolées. La saillance semble donc dépendre aussi de l'interaction entre les différentes perceptions extéroceptives (vision, audition, olfaction, gustation, toucher). De cet état de fait, nous pouvons alors questionner le dispositif cinéma qui doit se contenter des seuls canaux audio-visuels pour retransmettre les événements du réel. La simple reproduction des images et des sons réels

¹⁸ Dans l'expérience du réel, l'état intérieur de l'individu motive son parcours attentionnel. Nous questionnons ici finalement la possibilité du « chemin inverse » dans le contexte cinématographique.

¹⁹ C'est notamment le cas dans le film *Climax* (2018) réalisé par Gaspard Noé, où la caméra se balade à hauteur d'Homme d'un personnage à un autre comme si nous prenions part à la fête.

²⁰ Terme scientifique désignant la trace laissée dans le cerveau par un événement individuel.

n'est donc pas une finalité pour retransmettre la saillance d'un événement. Les ingénieurs du son s'exercent à dépasser la seule trace du son émis dans la réalité, et essayent plutôt de retransmettre les sensations associées à l'événement causal (la « boule » de sensations). C'est la différence entre la reproduction sonore et le rendu sonore (Chion, 2005, p.97)²¹.

CONCLUSION

Les mécanismes perceptifs et cognitifs impliquent une subjectivité de l'écoute. Elle dépend du contexte extérieur mais aussi de nos motivations, notre humeur, notre tempérament. Nous pouvons porter consciemment notre attention endogène, avec plus ou moins de focalisation, vers une représentation cible. L'attention exogène, utile à la survie en milieu hostile, travaille, elle, de manière préconsciente comme un outil de surveillance des alentours. Ces différents mécanismes s'équilibrent alors suivant les situations pour maintenir une cohérence d'action.

Dans le contexte cinématographique, le spectateur use de son attention pour découvrir le champ sonore. Mais celle-ci peut se voir attirée vers un parcours attentionnel agencé lors de la production de la bande-son, et motivé par la narration. La compréhension de ces mécanismes apparaît alors nécessaire pour établir une stratégie attentionnelle fonctionnelle. De plus, nous avons compris que les modulations attentionnelles sont notre conscience, et donc notre réaction, face au monde sensible. Alors, les techniciens peuvent adapter ces modulations au mixage des différents éléments sonores, comme une écoute « pré-focalisée », pour essayer de retranscrire la subjectivité de l'écoute du personnage de cinéma, en portant avec elle ce qu'elle dit de lui. Un personnage attiré par un élément sonore n'aura pas la même écoute qu'un personnage vigilant aux moindres événements sonores, ou qu'un personnage perdu dans ses pensées. Mais cette divergence attentionnelle, traduite en mouvements de mixage, nous donne-t-elle la sensation d'un caractère plus déterminé, vigilant ou pensif à notre personnage ?

²¹ Par exemple, les sons de coup de poing sont généralement accentués (par le monteur son et/ou le bruiteur) pour rendre la violence de l'acte, la douleur ressentie, le déplacement d'air du mouvement, etc. On peut citer comme exemple *A prayer before dawn* de Jean-Stéphane Sauvaire (2017), où le monteur son, Séverin Favriau, s'est attaché à rendre aux impacts de coups de poing une valeur sensorielle et narrative durant les combats de boxe en prison.

Aussi, le spectateur reste sujet à sa propre subjectivité. Lorsqu'il découvre l'œuvre, son propre vécu, ses expériences mais aussi son état momentanée peuvent induire chez lui des mouvements de l'attention l'entraînant en dehors du chemin attentionnel proposé par la mise en scène. Alors, il n'est pas certain que l'effet escompté advienne.

PARTIE II. LE CHAMP ATTENTIONNEL AU CINÉMA

INTRODUCTION

Le cinéma est une projection d'images en mouvement, d'un montage alternant les différents axes et valeurs de plans, apportant un matériel riche à réceptionner pour le spectateur. Le passage du cinéma muet au cinéma parlant a ajouté de nouvelles informations sensorielles venant compléter le schéma visuel du film. Mais le dispositif cinéma n'offre pas en soi l'impression de réalité (histoire du cinéma, écran rectangulaire, son monophonique. Encore aujourd'hui avec les évolutions techniques, on accepte la salle noire, les spectateurs présents dans la salle). Il offre simplement les matériaux nécessaires pour permettre au spectateur la schématisation d'une scène de cinéma. On établit la reconstitution de la scène dans l'image mentale - au contraire de l'image frontale découpée, montée, altérée et qui nous donne les indices qu'elle veut bien livrer - subjectif à chacun, en fonction du vécu, de l'expérience mais aussi de nos craintes et de nos désirs. Il y a interaction entre ce qu'on nous donne à voir, le perceptif, et ce que l'on reconstruit, le cognitif. L'un est prépondérant à l'autre. Le cinéma s'empare alors de cela pour induire au spectateur une narration visuelle et sonore. Au-delà de la construction de l'enchaînement des dialogues, la succession des différents points de vue et points d'écoute nous mène vers une schématisation mentale particulière des différentes séquences filmiques. La compréhension des mécanismes perceptifs et cognitifs de l'attention auditive permet alors un autre regard sur la mise en scène sonore au cinéma. Au-delà d'une volonté de reconstitution des apparences, la subjectivité des modulations attentionnelles semble offrir un précieux outil d'expression aux cinéastes. Alors, nous allons essayer d'approcher les pratiques techniques, sous le prisme de la rencontre entre les sciences de la perception et de la cognition et le cinéma, afin d'étayer leurs implications scénographiques et leurs incidences à la réception du film.

A. Modification du champ attentionnel face au dispositif cinématographique

1. Le champ frontal et le champ ambiant

Le dispositif cinématographique et ses pratiques techniques se sont adaptés à notre perception de l'espace et à nos expériences attentionnelles. En effet, notre perception du champ sonore se polarise en deux niveaux : le champ frontal et le champ ambiant (Bailblé, 1999, p.299). Face à nous, l'écoute et le regard peuvent agir de pair pour guider notre attention la plus sélective au sein du cône de présence. Les événements nous parviennent avec clarté et peuvent être finement scrutés et compris. *A contrario*, l'espace périphérique est atténué et ne bénéficie que d'une surveillance attentionnelle à bas niveau. La vision n'apporte plus ses précieuses informations et il est fatiguant d'écouter de côté. D'ailleurs, le premier réflexe face à l'intrusion d'un élément inconnu est de se réorienter pour le positionner dans le champ frontal. Ainsi, le cinéma résulte de cette polarisation. Il s'est construit sur nos aptitudes perceptives et attentionnelles pour nous proposer des histoires composées d'images et de sons. Le champ frontal est restitué par la projection visuelle et par la diffusion sonore (les bonimenteurs²², puis la diffusion monophonique et l'apparition d'image stéréophonique). Le champ ambiant, inexistant au début du cinéma, apparaît avec le Dolby Stéréo (et son canal *surround*) et se généralise avec les formats usuelles 5.1 et 7.1 (offrant plusieurs canaux *surround*) jusqu'au Dolby Atmos. Cependant, l'immobilité du dispositif implique l'immobilité des spectateurs, et fixe donc le rapport frontal / ambiant dans une référence spatiale fixe. La scène fictionnelle ne suit pas nos mouvements corporels (contrairement à la réalité virtuelle). Le dispositif cinématographique impose donc aux ingénieurs du son, de par sa structure et son histoire, certaines précautions et conventions qui tendent à distordre l'expérience du réel, pour mieux intégrer la fiction. Cependant, les pratiques instrumentales (de la captation, au mixage en passant par le montage son) se sont développées parallèlement aux outils et rendent possible aujourd'hui la reconstitution de champs sonores vraisemblables, de bandes son sensoriellement convaincantes.

²² Le bonimenteur de cinéma était chargé de commenter les films à l'époque du cinéma muet.

2. L'étagement des sons

Avant l'arrivée du Dolby Stéréo, les bandes sonores monophoniques étaient confrontées au problème d'intermasquage. En effet, si nous sommes capables de reconstruire cognitivement une localisation image / son incohérente grâce à l'effet ventriloque²³, l'usage d'un seul haut-parleur posait le problème de l'intermasquage sur l'ensemble des sons diffusés simultanément. L'environnement sonore était compacté en un seul point de diffusion, en un « hublot » comme le décrit Claude Bailblé (2004). Alors, n'ayant pas la place de tout faire écouter, les ingénieurs du son procédaient à différentes techniques. Les sons étaient hiérarchisés pour leur importance vis-à-vis de la mise en scène, de la compréhension du film et des enjeux narratifs. Ainsi, les voix occupaient une place importante dans le « hublot » sonore, en dépit des autres matières sonores. Si la qualité de « réalisme » apparent n'était pas convaincante, la contrainte de la monophonie entraînait les mixeurs vers des choix signifiants pour la narration. Les mouvements des potentiomètres nous guidaient vers les seuls sons à entendre et à comprendre, amenant des premières tentatives - sous la contrainte - de modulations des mouvements attentionnels propres à notre écoute. La nécessité d'intelligibilité impliquait des choix de mise en scène forts aiguillant ainsi le spectateur dans une relation audio-visuelle quasiment dépourvue de sons insignifiants. Le langage cinématographique s'est alors emparé de l'apport du sonore dans sa relation avec l'image, dépassant la simple contrainte.

Alfred Hitchcock nous présentait, dans son film *Rear window* (1954), Jeff, un photographe, la jambe dans le plâtre, qui passe son temps à observer par la fenêtre l'immeuble d'en face. L'un des premiers plans du film, adoptant le point de vue de Jeff, nous balade de fenêtres en fenêtres (ouvertes pour cet après-midi de forte chaleur) esquissant les portraits du voisinage. On nous donnait alors à entendre pour chaque petite scènette (ici les différentes fenêtres ouvertes) les activités ou les discussions des différents voisins. Outre le caractère introductif de cette séquence, la succession des événements sonores, en passant de fenêtre en fenêtre, nous renvoie au caractère « voyeuriste » -

²³ Lorsqu'un stimuli visuel et un stimuli sonore sont temporellement coïncidents mais disjoints spatialement, le stimuli sonore est souvent entendu à l'endroit du stimuli visuel ; c'est l'effet ventriloque.

devrais-je dire « écouteuriste » - de Jeff, qui balaye finement son faisceau attentionnel sur chaque fenêtre pour en scruter tous les détails (les sons émanant de la fenêtre sont très présents, les sons en dehors de la fenêtre sont lointains ou inexistant). Une fois l'observation passée, le son s'évanouit au profit d'une autre scène sonore encadrée par la nouvelle fenêtre focalisée. Ainsi, au-delà d'une simple contrainte liée à la diffusion monophonique, l'enchaînement des événements sonores dans la relation audio-visuelle nous donne un point de vue sur Jeff, souligne son intention d'écoute et nourrit les enjeux du film (l'observation pour l'enquête, mais aussi le voyeurisme et l'intrusion chez autrui). On retrouve aujourd'hui des films, comme *Gosford Park* (2001) de Robert Altman, où la mise en scène sonore semble s'inspirer de ces grands mouvements de points d'écoute (ou de mouvements de l'attention), comparables aux pratiques du mixage monophonique, auquel on ajoute la spatialisation. Les éléments se succèdent plus qu'ils ne coexistent.

Aujourd'hui, le problème d'intermasquage ne semble plus être une contrainte trop importante²⁴. Les systèmes de diffusion multicanaux permettent une répartition du champ objet et donc une diminution de l'effet de masque (cf. L'effet de masque). Le champ sonore peut être enrichi. De plus, la qualité d'enregistrement et de restitution offrent la possibilité de répartir les éléments sonores sur l'ensemble du spectre audible.

Finalement, avec quelques précautions techniques, l'étagement des sons s'opère sans difficultés majeures dans le processus de fabrication d'une scène sonore. De plus, les pratiques techniques d'autrefois semblent nous indiquer des possibilités d'expression que la bande sonore doit saisir pour nourrir la relation audio-visuelle. Alors, la pratique des systèmes de diffusion actuels, aussi efficaces soient-ils dans l'apparent résultat de réalisme, ne doit pas nous détourner de cet enseignement.

Outre le bénéfice du démasquage, la diffusion d'ambiance débordant l'écran (diffusion par les haut-parleurs latéraux derrière l'écran mais aussi par les haut-parleurs *surround*) semble participer à l'effet de vraisemblance du champ sonore (Chion, 2005, pp.

²⁴ Le mixeur Éric Tisserand me confiait ne plus avoir à traiter systématiquement ce problème, bien qu'il puisse encore exister des cas particuliers de films où l'intermasquage demeure une contrainte. Je pense notamment à des films où l'environnement scénique veut être rendu à fort niveau, ce qui complique l'émergence des voix.

93-128), simulant notre faculté à percevoir le son à 360°²⁵. Michel Chion (2005, p.129) évoque même l'idée que le champ sonore enveloppant (du point de vue sensoriel) et contextualisant (du point de vue narratif) a en partie remplacé le rôle du plan large dans le découpage classique au cinéma. La spatialisation des divers objets sonores nous permet de retrouver notre faculté de ciblage attentionnel au sein d'un espace. Nous écoutons le champ sonore, selon nos attentes, nos désirs vis-à-vis des enjeux du film. Dans un même temps, un élément sonore peut être rehaussé cognitivement et l'ambiance alentour atténuée (sans qu'elle soit nécessairement baissée physiquement). Le potentiomètre du mixeur est substitué par « le potentiomètre attentionnel » du spectateur (Bailblé, 1999, p. 351). Dans ce cas, on le place dans une posture plus active face au film.

3. Le cas du plan-séquence

Le plan-séquence semble être un procédé de mise en scène plus conciliant avec la localisation des éléments sonores, cohérente dans le champ image et hors du champ image. D'abord, la continuité temporelle et spatiale permettent une liberté de spatialisation, sans craindre la rupture de localisation due à un changement d'axe au montage. De plus, cette continuité permet d'introduire des éléments sonores dans le champ image et prépare le spectateur à d'éventuels mouvements du référentiel spatial et de ces éléments sonores hors de l'écran (*Victoria* (2015) de Sebastian Schipper, *Birdman* (2015) d'Alejandro Iñárritu, *Climax* (2018) de Gaspard Noé). Plus généralement, la continuité perceptive nous rapproche de notre expérience du réel. La préparation contextuelle au sein d'un champ sonore continu nous permet d'intégrer davantage les éléments constitutifs, sans qu'ils ne subissent de rupture ou d'incohérence de localisation. Notre écoute, intentionnelle, fait preuve de discernement. On peut ressentir, surveiller les éléments hors-champ sans qu'on se sente obliger de tourner la tête (contrairement à la réorientation réflexe qui peut être amenée par la surprise de la rupture spatiale, sollicitant les neurones phasiques). Ainsi, dans le film *The revenant* (2015) d'Alejandro Iñárritu, le plan-

²⁵ Toutefois, l'utilisation de l'ensemble des canaux de diffusion sonore n'est pas une obligation. Par exemple, Eric Tisserand n'utilise pas nécessairement les haut-parleurs *surround* pour les séquences en intérieur, et amène ainsi du contraste en largeur d'image sonore avec les séquences en extérieur, plébiscitant ici l'ensemble des canaux.

séquence introductif use d'une spatialisation qui déborde largement du champ frontal. L'image nous montre un cours d'eau dans lequel marche le trappeur Hugh Glass, le personnage principal. Alors, surviennent progressivement des pas, supposément dans ce même cours d'eau, diffusés derrière nous. Ces éléments, d'abord dorsaux, avancent lentement sur le côté pour finalement entrer dans l'image frontale. Ce sont d'autres trappeurs accompagnant Hugh Glass. Ici, le mouvement de localisation est une entrée de champ. La préparation est alors contextuelle. La compréhension et la cohérence des sons face au contexte, dans une continuité perceptive, nous permet de conserver notre engagement attentionnel principal vers le champ frontal, bien que l'attention auditive ait été sollicitée par une modulation ascendante. Je remarque tout de même que ce procédé n'a pas pour vocation de laisser éternellement ces effets derrière nous. D'ailleurs, Alejandro Iñárritu utilise aussi cette pratique de mise en scène dans *Birdman* (2015), sauf que ce n'est pas le son qui rejoint l'image frontale, mais l'image qui se déporte davantage sur le son entendu dans le hors champ²⁶. Le mouvement de la caméra agit comme une réaction d'orientation due à l'attention exogène capturée, une sorte de personnification du point de vue qui réagit selon nos propres mécanismes attentionnels.

En somme, la forme du plan-séquence semble être le procédé cinématographique qui se rapproche le plus de notre perception du monde sonore, en confondant la temporalité du plan avec celle de diffusion, et en restituant une continuité spatiale. Les pratiques techniques ont alors la possibilité de recomposer un champ sonore perceptivement plus convaincant que la bande sonore contrainte par le découpage. Le spectateur retrouve davantage d'aptitudes perceptives et attentionnelles.

Cependant, les possibilités techniques de captation et de diffusion du son au cinéma, nous redonnant la capacité d'ouïr et d'écouter dans un même temps, ne sont que les outils d'une intention de mise en scène. En effet, la recomposition d'un champ sonore suppose généralement un point d'écoute. Au même titre que l'on décide de l'endroit où poser la caméra pour définir le donné-à-regarder et le manque-à-voir, le point d'écoute fixe

²⁶ Je pense notamment à la sortie du héros Riggan Thomson d'une petite épicerie, où un SDF l'interpelle dans le hors champ. La caméra se retourne alors sur l'intrus auditif, avant de s'en désintéresser et de continuer son chemin. Après être momentanément entré dans le champ frontal, le son repart en dehors de l'écran et s'évanouit dans les enceintes *surround*.

l'étalement du proche et du lointain en un endroit de captation. C'est la rencontre entre la source, un espace et un microphone sous le geste de l'ingénieur du son. Le monteur son intègre ensuite cette matière dans un étage de sons plus vaste, apportant profondeur, largeur et relief de l'image sonore. L'équilibre des différentes couches parfait le point d'écoute, dans un univers sonore reconstitué. Il peut être fixe ou mouvant, figé au point de vue ou errant dans le lieu scénique. Ces pratiques instrumentales organisent finalement la mise en scène du sonore, dans un lien étroit avec l'image frontale, motivées par une intention de réalisation. La juxtaposition d'un son sur une image mène alors le spectateur dans une reconstruction mentale, cognitive, des champs perçus. Au-delà de la qualité perceptivement vraisemblable, voire immersive, l'interférence sensorielle semble apporter une valeur ajoutée, valeur expressive. Le son « interprète le sens de l'image » nous dit Chion (2005, p.33). Gérard Leblanc évoque la notion de double scénarisation (Leblanc, 1991), un aller-retour entre la perception des champs sensoriels et la projection cognitive.

Aussi, la stratification de l'ensemble sonore, avec une captation en multimicrophonie (au tournage) et une reconstruction d'un univers sonore (en post-production), offre un large champ des possibles. La structuration des éléments sonores disjoints peut répondre à des intentions de mise en scène explorant d'autres voies que le réalisme supposé du point d'écoute. Le respect de la perspective, de l'équilibre naturelle du monde sonore peut être distordu pour une mise en relation narrativement plus féconde des différents éléments. Du point de vue du champ attentionnel, la narration semble quelquefois privilégier la saillance d'un objet sonore à sa véracité spatio-temporelle. La compréhension des mécanismes attentionnels nous permet alors de modeler le champ sonore pour induire discrètement au spectateur un chemin d'écoute. Certains metteurs en scène s'affranchissent même d'un naturalisme apparent, au profit d'une mise en scène des seuls sons qui semblent utiles au film²⁷, définissant leur propre esthétique. L'écoute du spectateur devient alors majoritairement guidée, par la succession d'indices sonores²⁸.

²⁷ Les films *Jusqu'au bout de la nuit* (1995) et *Ainsi soit-il* (2000) de Gérard Blain, mixé par Éric Tisserand, nous font entendre des intentions de mise en scène sonore, dépourvue par exemple d'ambiance, loin d'une esthétique naturaliste.

²⁸ À l'image de certains films construits en monophonie, contraint par la technique à la succession plus qu'à la simultanéité du son.

Enfin, les différentes possibilités qu'offre le processus de fabrication de la bande sonore, et sa relation qu'elle entretient avec l'image, sous-tendent la qualité expressive de celle-ci pour le médium cinématographique. Les mécanismes de l'attention auditive, expérimentés par tous, peuvent alors être une base de réflexion dans la mise en scène sonore. En quoi le champ attentionnel proposé aux spectateurs peut-il influencer sur la perception de l'œuvre ? Comment s'articule-t-il avec le champ image et pour quel résultat ? Pour la suite de l'étude, et au vu des qualités perceptives propres au plan-séquence, nous essaierons d'apporter des réponses à ces questions, en se concentrant sur cette pratique spécifique de mise en scène.

B. Expressivité du champ attentionnel en plan-séquence

La perception d'un champ sonore est modulée par les mécanismes attentionnels (cf. L'attention auditive). Ces modulations sont le produit de la rencontre entre une personnalité et le champ environnant. Elles s'inscrivent dans le champ organisme/environnement. Elles témoignent de son attitude d'écoute, qui émane de l'état d'esprit, de l'état émotionnel momentané de l'individu. Chacun semble donc percevoir sa propre réalité, qui le renvoie à ses particularités individuelles. Alors, la subjectivité de l'écoute questionne le statut de la bande sonore au cinéma. Reconstituer un champ sonore complet, dans un naturalisme apparent, c'est faire le choix de laisser le spectateur se balader au sein de celui-ci, de se l'approprier ; ou au contraire, de l'ennuyer par manque de repères attentionnels. L'apport de la qualité immersive me semble alors devoir être contre-balancé avec le risque d'égarement du spectateur des sentiers de la narration du film. La mise en scène joue ici un rôle capital²⁹. *A contrario*, en personnifiant un point d'écoute et en adaptant les mouvements de mixage aux modulations attentionnelles supposées du personnage, j'émet l'hypothèse que le champ attentionnel semble pouvoir être le reflet d'une humeur, d'un tempérament, d'une personnalité, d'une motivation de ce personnage.

Alors, à l'aide d'analyses filmiques et sous le prisme des mécanismes attentionnels, nous allons essayer de comprendre les perspectives expressives des pratiques techniques (prise de son, montage son, mixage). Les modulations de l'attention auditive, résultantes de l'évolution du champ organisme/environnement, se réalisent à différentes temporalités. En effet, les évolutions attentionnelles d'un individu peuvent s'opérer sur un temps court (par le surgissement d'un nouvel élément) ou sur un temps plus long (adaptation ou changement progressif du contexte ou du personnage). Les analyses se feront donc à la fois sur l'échelle du plan-séquence et sur l'échelle du film global (qui se confondent quasiment pour certains films). Mais avant tout, posons quelques définitions, proposées par Laurent Jullier (1995), pour préciser l'étude.

²⁹ Les mouvements de caméra et les points de synchronisme audiovisuels captent l'attention du spectateur, comme le suggère Michel Chion dans *L'audiovision*. La mise en scène doit donc davantage soutenir la narration, puisque la bande son se veut « transparente ».

Naturalisme : « Façon de présenter les données de l'écran la plus proche possible, compte tenu des seules limites technologiques du dispositif, de la perception telle qu'elle est définie par les Gestaltistes (= le « témoin invisible », défini par l'ocularisation, se trouve au centre d'une idéoscène contenant les sources de tous les sons qui, soumis à la loi proxémique, parviennent à ses oreilles). »

Auricularisation : « Construction par des sons filmiques d'un « point d'écoute » ou « témoin auditif invisible » qui peut, ou non, se confondre spatialement avec le « témoin visuel invisible » donné par la prise de vue (ocularisation) - c'est le cas lorsque l'auteur fait le choix de la perspective naturaliste - ou se confondre avec un personnage »

Réalisme sonore psychologique : « refus de la perspective naturaliste au profit soit de l'intelligibilité à tout prix, soit d'un parti-pris d' « attentions » sonores particulières non conditionnées par les espaces visuels qui se montrent sur l'écran. Met en jeu des mirages en majorité audio-visuel »

Le plan-séquence : « il s'agit d'un plan assez long et articulé pour représenter l'équivalent d'une séquence. » (Aumont, 2001, p.158)

1. Le point d'écoute

Le point d'écoute au cinéma est façonné par la perspective sonore, la réverbération, l'éloignement des sources, mais aussi par le dispositif de diffusion qui permet la spatialisation du champ sonore. Il est engendré par une multitude de choix, du tournage à la post-production. En ce sens, le point d'écoute doit être avant tout le résultat d'un travail de mise en scène, pensé et mené tout au long du processus de fabrication du film. C'est une ouverture sur la scène fictionnelle proposée aux spectateurs. Walter Murch, mixeur son entre autres pratiques, résume :

« Si je sors pour enregistrer le claquement d'une porte, je ne pense pas que j'enregistre une porte qui claque. J'enregistre en fait l'espace dans lequel le claquement se produit. »

(Murch, 2004)³⁰

³⁰ Cité par ONDAATJE, Micheal. The conversations : Walter Murch and the Art of film editing. New-York, Random House, Trad.Frédéric Dallaire, 2004. p.234

a. Adéquation au point de vue

Victoria, l'auricularisation au plus près du personnage

Victoria (2015) est un *thriller* réalisé par Sebastian Schipper. Il est constitué d'un unique et réel plan séquence (sans coupure dissimulée) de 2h14 tournée dans les rues de Berlin. Trois tentatives ont été accomplies pour finalement choisir la deuxième.

Synopsis du film : 5h42. Berlin. Sortie de boîte de nuit, Victoria, espagnole fraîchement débarquée, rencontre Sonne et son groupe de potes. Emportée par la fête et l'alcool, elle décide de les suivre dans leur virée nocturne. Lorsqu'elle apprend que le groupe projette de braquer une banque au lever du jour, il est déjà trop tard. La soirée dérape définitivement...

Le film nous plonge dans une fin de soirée aux côtés du personnage de Victoria. La rencontre d'un groupe de jeunes berlinois, à la sortie d'une boîte de nuit, entraîne Victoria vers une virée nocturne improvisée. Le rythme du film s'accélère alors quand elle se voit obligée d'accompagner les jeunes hommes à un braquage. L'errance se transforme vite en un tunnel infernal, que l'on subit à côté, et même avec Victoria. En effet, la caméra vient chercher la jeune espagnole dansant sur la piste de danse au début du film pour ne la lâcher que 2h14 plus tard. Entre ces deux points, nous restons à ses côtés. On la suit où qu'elle aille : à vélo, en voiture, dans l'ascenseur, sur les toits de Berlin. On vit les événements avec elle. Puis on devient inquiet de la situation. On a peur avec elle. On cherche des solutions. On scrute les potentiels dangers et les possibilités d'actions, de fuites. La sensation d'immersion règne. La continuité temporelle et spatiale nous assène les mêmes informations narratives auxquelles Victoria est confrontée. Il y a simultanéité de savoir entre Victoria et le spectateur. On reste juste impuissant, à l'image du personnage principal.



Figure 10.1 - *Victoria* à Berlin, 5h42.



Berlin, 7h56.

L'esthétique naturaliste nous accompagne tout au long du film, hormis à deux instants où la musique remplace les sons diégétiques³¹. Les perspectives sonores paraissent cohérentes. Les voix ne sont d'ailleurs pas obligatoirement intelligibles. Elles sont, comme l'ensemble du champ sonore, entièrement spatialisées. Ceci-dit, la mise en scène essaie tant bien que mal de placer l'ensemble de l'action au sein du champ frontal. L'image se recentre souvent sur le personnage parlant. La caméra accompagne les dialogues et la voix se fixe alors sur l'enceinte centrale. De fait, il n'y a qu'à de brefs moments une forte latéralisation des voix. Au-delà de l'aspect technique (diffusion homogène pour l'ensemble de la salle), cette mise en scène confère une volonté d'intention dans l'écoute du « témoin invisible », et de Victoria par extension. On implique le spectateur. Le point d'écoute, en adéquation avec le point de vue, nous offre un champ sonore vraisemblable, immersif. On nous montre et on nous fait entendre - toujours dans une esthétique naturaliste - ce vers quoi l'attention se porterait instinctivement, vers le lieu de l'action, vers les informations portant un intérêt pour Victoria. Mais, il n'y a jamais de modulation attentionnelle inscrite dans le mixage. L'auricularisation reste cohérente tout le long du film, et fusionne avec le point de vue. Le choix du plan-séquence, allié à une esthétique naturaliste, semble alors impliquer davantage le spectateur durant les deux heures du film. Et c'est éprouvant.

L'esprit naturaliste du film nous transporte comme « témoin invisible », au côté de Victoria, dans une perception proche de la réalité, sans artifice ni distorsion du champ sonore. Le spectateur peut user de mécanismes cognitifs pour focaliser son attention dans le champ objet. Le réalisateur fait d'ailleurs le choix de ne mettre aucun son diégétique sur le peu de musiques extradiégétiques présentes dans le film, comme une volonté de ne pas superposer deux matières n'appartenant pas au même espace, de ne pas insérer d'interprétation de l'une sur l'autre. La perception reste brute. Si les débuts du film peuvent s'apparenter à de la contemplation urbaine, la suite des événements ne laisse pas de place aux errances de l'esprit. On est amené à subir le réel. La réalisation ne nous donne alors aucune porte de sortie, aucune avance narrative par rapport au personnage de Victoria. On est vigilant pour tenter de discerner l'instant d'après. On avance à petits pas. Le traitement

³¹ Ces instants de respiration semblent nécessaires, une fois pour pallier à la longueur d'un déplacement entre deux lieux et une seconde fois pour ponctuer l'aboutissement d'une première successions d'événements intenses.

naturaliste du champ sonore et sa proximité avec Victoria nous pousse alors à nous impliquer dans l'attention portée au sein du lieu scénique, par empathie pour elle.

Pour le braquage, Victoria doit amener les garçons en voiture à la banque pour l'heure d'ouverture. Pendant qu'ils forcent l'entrée, arme au poing ; Victoria les attend juste devant, prête à repartir.



Figure 10.2 - Victoria dépose le groupe de garçons qui partent braquer la banque.

On reste alors avec Victoria à attendre dans la voiture. Les jeunes berlinois entrent dans la banque, les portes se referment et aucun son ne semble alors provenir de l'établissement. Mais un problème survient : le moteur de la voiture, volée précédemment et démarrée habilement sans les clés, s'éteint brusquement. Il est alors impossible pour Victoria de redémarrer la voiture. On remarque d'ailleurs que le son du moteur capte notre attention que lorsqu'il s'éteint. C'est ici un exemple pertinent de saillance, non pas d'un son, mais d'une représentation d'un événement (par rupture dans une continuité) ancrée dans un contexte. Il est narrativement fort de sens. Le silence devient pesant et l'attente aussi longue que ce que perçoit Victoria. On tend alors l'oreille pour essayer de déceler où en est le braquage, et le moment où les garçons vont ressortir. On fait travailler nos aptitudes de vigilance, mais rien ne provient jusqu'à nous, jusqu'à Victoria. On reste sur le procédé d'auricularisation, et c'est au spectateur de faire travailler son attention auditive. Soudain, le bruit des garçons rompt l'attente et l'un d'entre-eux redémarre la voiture.



Figure 10.3 - Boxer, l'un des garçon, sort de la banque et redémarrera la voiture.

Après le braquage réussi, le groupe décide de retourner dans une boîte de nuit. La tension redescend. Trop agités, ils se font virer de la boîte de nuit quelques minutes seulement après être rentrés. L'un des garçons est particulièrement drogué, et devient très peu lucide. À la sortie, on l'aperçoit au loin en train de tituber avec un seau sur la tête. Il se rapproche alors d'une entrée pour voiture donnant sur une petite cour, l'endroit où la voiture volée a été laissée. Les autres suivent derrière.



Figure 10.4 - Le garçon avance sans remarquer la voiture de police au fond de l'allée.

Une voiture de police se trouve dans la cour intérieure, que ne remarque pas le garçon au seau sur la tête. À cette distance, aucun son des policiers ne s'entend. On ne s'aperçoit de leur présence que lorsque Victoria arrive au niveau de l'allée. Ils font alors immédiatement demi-tour, en appelant leur ami pour qu'il ressorte de l'allée. Ils se mettent ensuite à marcher, en ayant conscience qu'ils ont certainement attiré l'attention des policiers. Ils se débarrassent de leurs affaires et accélèrent le pas. Jusqu'ici, la présence des policiers ne se fait pas sentir dans le champ sonore. Il est impossible pour nous comme pour eux de savoir s'ils sont suivis. L'intérêt du traitement naturaliste de la bande sonore est ici saisissant. On implique encore le spectateur dans une écoute sélective (on instaure un filtrage sensoriel, supra-auditif, dans l'attente d'un signe). On essaie de débusquer le moindre indice sonore pour savoir si les policiers les suivent. Plus ils avancent et plus ils accélèrent le rythme. Leur réaction nous donne l'impression d'une présence fantôme, qu'on ne voit pas, qu'on n'entend pas, mais que l'on sent. On est dans l'attente d'une confirmation, par le son, de la présence - ou non - d'une filature.



Figure 10.5 - Le groupe accélère la marche puis se met à courir.

Finalement, l'apparition d'une voiture banalisée roulant à vive allure apporte la confirmation. Le groupe fuit.

Dans un plan-séquence interminable, *Victoria* nous plonge dans une expérience intense. L'esthétique naturaliste, avec une adéquation entre le point de vue et le point d'écoute, nous fait partager une partie de la nuit de la jeune espagnole. Le partage est poussé jusqu'à l'appropriation des enjeux, et donc des attitudes attentionnelles du personnage. La continuité et la proximité avec Victoria nous aide à la connaître, à l'apprécier et à éprouver de l'empathie à son égard. Cette continuité nous montre aussi l'omniprésence du danger, on ne rate pas un instant. La mise en scène de l'image et du son, dans un respect de l'auricularisation du champ sonore aux côtés de Victoria, implique alors le spectateur dans une démarche attentionnelle active. La continuité temporelle et spatiale aident, évidemment, d'une part, à une spatialisation cohérente (dans la limite du dispositif de diffusion) et d'autre part, à une personnification des champs sensoriels (le témoin invisible). L'arrivée de la lumière matinale rend palpable l'idée du temps écoulé, de cette nuit infernale passée avec Victoria et un groupe de berlinois. On éprouve alors l'escalade de tension engendrée par les différents événements, dans toute sa longueur. La perception du champ sonore naturaliste passe chez le spectateur d'une certaine contemplation (le groupe d'amis s'installe sur les toits de Berlin, Victoria joue du piano avec virtuosité pendant un long moment au premier tiers du film) à une vigilance accrue. Nos aptitudes attentionnelles sont particulièrement sollicitées. Le film parvient finalement à l'impression de vécu³².

³² Le cadre et l'esthétique de l'image convergent avec cette conclusion (Bonnani, 2016).

b. Le point d'écoute dissocié du point de vue

Gosford Park, le point d'écoute autonome mène l'enquête

Gosford Park (2001) est un film réalisé par Robert Altman. Il mélange le genre du drame et du film à énigme. Bien qu'il ne présente pas de réel plan séquence, certains plans contiennent des changements de points d'écoute pertinents pour l'étude.

Synopsis du film : Au début des années trente, dans une Angleterre fortement marquée par les inégalités de classe, une famille d'aristocrates avec à sa tête la maîtresse de maison, Lady Sylvia McCordle, organise une partie de chasse au cours de laquelle son mari Sir William McCordle est retrouvé poignardé. Cet assassinat va bouleverser l'ordre établi et révéler la complexité des liens entre les maîtres et leurs serviteurs. L'inspecteur Thompson mène l'enquête.

Robert Altman nous présente dans son film un univers foisonnant de personnages. Il y a les aristocrates, accompagnés de leurs valets respectifs. Puis il y a les servants qui s'occupent de la tenue de la demeure et du bon déroulé des activités. Dans ce somptueux manoir anglais, les liens entre les individus s'articulent autour d'espaces dédiés. Selon leur statut, les personnages occupent différents étages, différentes salles. La haute société dîne dans de luxueux salons. Les valets se retrouvent autour de tables plus modestes. Les servants occupent les cuisines, les buanderies. Ils existent bien sur des connexions. Les valets s'occupent de leur aristocrate pour lequel ils travaillent. Ils font aussi le lien entre les étages supérieurs, pour l'aristocratie, et les sous-sols, laissés aux servants. Les relations sont extrêmement hiérarchisées et codifiées. À l'arrivée de tout ce monde, les petites mains s'activent pendant que la classe supérieure fait les présentations. C'est alors l'émulsion de voix, de va-et-vient. On est confronté à tous ces personnages. La bonne distinction des identités et la compréhension des échanges sont difficiles. Les voix s'entremêlent et, à peine le temps de s'accrocher à quelques mots qu'un nouveau son vient rompre le dialogue en place. La cloche, tintant à chaque ouverture de la porte d'entrée, rythme la première séquence à l'intérieur du manoir. Elle décompte en quelque sorte le nombre impressionnant d'invités, multipliant ainsi les potentialités relationnelles durant le séjour. C'est au début du film un véritable sac de noeuds pour le spectateur.



Figure 11.1 - *L'arrivée des multiples invités dans le manoir, accueillis par Lady Sylvia McCordle.*

On remarque tout d'abord que le point d'écoute semble nous rapprocher des individus. Les voix sont brèves mais intelligibles, en proximité. Aussi, elles restent localisées au centre. Cependant, je note qu'on ne voit que très rarement les bouches lors des premiers mots prononcés par les différents personnages. Ils sont en mouvement, à l'image des valets qui ne s'arrêtent qu'un instant pour savoir où ils doivent déposer les affaires. Cette proximité sonore, associée à la mise en scène dynamique des premières séquences, nous présente un tableau général. Les données sont là, disponibles. On nous facilite même leur bonne compréhension par un point d'écoute proche des personnages parlant, qui fixe les paroles dans le canal central.

Par la suite, le rythme ralentit. Les séquences mettent en jeu plus que deux ou trois personnages à la fois. Les relations se précisent. On dévoile les caractères des uns et des autres, dans leur intimité, loin des représentations. De plus, la mise en scène apporte toujours un liant entre ces différents passages. Le point d'écoute se dissocie du point de vue pour nous placer au plus proche des discussions. Dans un même plan large, on passe alors d'un dialogue en arrière plan à l'image, mais en gros plan au son, à un début de dialogue au premier plan. La liberté du point d'écoute fait se succéder les dialogues, plus qu'ils ne cohabitent à égal perspective. Aussi, on nous fait entendre, dans cette même logique de succession, des bribes de voix provenant d'autres pièces. Bien qu'ils soient mis en acoustique, ils demeurent plus présent que des discussions provenant de la pièce dans laquelle nous sommes. Il s'agit généralement de transitions entre séquences. Cette mise en scène lie les différents personnages, et les différents lieux dans une même unité, un

gigantesque tableau. La bande sonore s'enchevêtre dans le découpage plus qu'elle ne le suit. Si on s'écarte un peu de l'analyse des plans-séquences, je remarque tout de même qu'à l'échelle d'un même plan, le point d'écoute est mouvant.



Figure 11.2 - La comtesse de Trentham, interprété par Maggie Smith, joue aux cartes à la table en arrière plan. Un valet arrive alors dans la pièce.

Pendant l'instant du premier photogramme, le point d'écoute se situe au niveau de la table, en arrière plan. Les voix, discrètes dans la locution, sont fortement présentes dans le champ sonore. On entend parfaitement la comtesse de Threntham. Quand soudain arrive le valet que l'on aperçoit en premier plan dans le second photogramme (arrivée silencieuse mais indiquée par le regard de la femme à la robe verte). Alors, le point d'écoute se déplace avec cette femme et revient au premier plan sonore, tandis que la voix de la comtesse paraît beaucoup plus lointaine, noyée dans l'acoustique. Ce mouvement de point d'écoute s'effectue dans un même plan.

Au deux tiers du film, Sir William McCordle est poignardé. Un inspecteur est dépêché pour trouver le coupable. « *A shooting party in a country house. A murder in the middle of the night. A lot of guests for the week-end. Every one's a suspect.* »³³ décrit Moris Weissman, producteur de cinéma, à l'ensemble de la tablée pour expliquer le scénario de son prochain film ; cela avant même le meurtre du maître de maison. C'est une parfaite mise en abyme de ce qui surviendra. On se charge alors de présenter l'ensemble des personnages à l'inspecteur. La mise en scène qui introduit subtilement, tout le long du film, l'ensemble des relations entre les personnages, prend tout son sens. Le tableau, présenté dans sa largeur puis dans ses finesses, nous positionne en tant qu'inspecteur avant l'inspecteur lui-même. On en sait davantage. On connaît certains liens privés, certaines

³³ « Une partie de chasse dans une maison de campagne. Un meurtre au milieu de la nuit. Beaucoup d'invités pour le week-end. Chacun est suspect. »

discussions secrètes. On théorise sur les doubles postures de chacun. On confronte les dialogues, les états d'esprit, les informations. Alors, ce point d'écoute mouvant dans la scène soutient parfaitement cette idée. On ne nous présente pas que le plan large, publique, vu de tous ; mais aussi les zones d'ombre. L'interférence entre ces plans d'ensemble à l'image et un point d'écoute beaucoup plus précis amène cette nouvelle lecture. « Le son interprète le sens de l'image » nous rappelle Chion (2005, p.33). Il confère au spectateur un savoir grandissant, normalement inaccessible, au fur et à mesure du film. Ici, l'implication n'est pas attentionnelle (comme dans le film *Victoria* et son esthétique naturaliste) puisque l'on donne les informations par le mouvement et la proximité du point d'écoute aux différents personnages ; mais le spectateur mène l'enquête, recolle les morceaux du puzzle.

« s'il faut faire travailler le sonore au cinéma, ce n'est pas en donnant plus de sons aux films - au contraire, ils en ont trop -, mais peut-être en leur trouvant une plus juste place, une importance plus distinctive c'est-à-dire mieux choisie, mieux adaptée à la construction en regard de laquelle ils s'inscrivent »

(Deshays, 2010)

La mise en scène du champ sonore dans *Gosford Park*, au sein du plan et plus généralement dans sa structure globale, met en pratique cette remarque, en s'attardant davantage sur la construction de l'enchaînement des dialogues et de leur mise en perspective dans l'image frontale, plutôt que de retranscrire un champ sonore complet mais inadéquat pour le film. On s'affranchit du point d'écoute fixé à la caméra pour venir insérer un angle de lecture autre, pertinent pour un film à intrigue, en rupture avec un présupposé naturalisme. L'apport expressif semble finalement se contenir dans le décalage entre le champ image, large ou dissimulé derrière une fenêtre, et le champ sonore, en proximité, dévoilant les secrets.

c. Le point d'écoute démultiplié

Gerry, la dissociation des personnages de leur environnement

Gerry (2002) est un film réalisé par Gus Van Sant. Il s'inscrit dans une trilogie, dites de la « survivance », avec les films *Elephant* (2003) et *Last days* (2005) réalisés par le même Gus Van Sant.

Synopsis du film : Deux hommes, nommés tous deux Gerry, traversent en voiture le désert californien vers une destination qui n'est connue que d'eux seuls. Persuadés d'atteindre bientôt leur but, les deux amis décident de terminer leur périple à pied. Mais Gerry et Gerry ne trouvent pas ce qu'ils sont venus chercher ; ils ne sont même plus capables de retrouver l'emplacement de leur voiture. C'est donc sans eau et sans nourriture qu'ils vont s'enfoncer plus profondément encore dans la brûlante Vallée de la Mort. Leur amitié sera mise à rude épreuve.

Le film met en scène deux amis dans un décor vaste, le désert californien. On y perçoit alors leur parcours au sein de cet espace, à la fois unique et changeant. S'ils semblent confiants au début de leur marche, le film nous montre une lente dégradation de l'état des personnages au fur et à mesure des heures et des jours passés. Cette transition se fait alors ressentir dans la bande sonore, façonnée par le concepteur sonore Leslie Shatz.

Tout d'abord, inscrire deux corps dans un espace nous renvoie à la notion du champ organisme/environnement. La bande sonore peut alors retranscrire les interférences entre le monde intérieur et l'univers extérieur. Les sons corporels participent au champ sonore. Le bruit des pas, celui des respirations sont autant de zones de contact entre ces deux univers, qui semblent pouvoir nous éclairer sur leurs interactions. Westerkamp (1974) nous renseigne alors sur ce qu'il qualifie de « déambulation sonore » (« soundwalk ») :

« Commencez par écouter les sons de votre corps qui se déplace. Leur proximité permet d'établir le dialogue entre vous et l'environnement. Si vous entendez ces sons infimes, vous vous déplacez dans un environnement aux proportions humaines. En d'autres termes, avec votre voix ou vos pas, vous «parlez» à votre environnement, qui à son tour répond en donnant à vos sons une qualité acoustique spécifique. »³⁴

Leslie Shatz semble alors vouloir s'approprier ces enseignements pour inscrire dans la bande sonore ce fameux « dialogue » entre l'individu et l'environnement, incluant l'individu. Il y a avant toute chose un travail de prise de son, de matière sonore chère à Shatz dans une double volonté. Il veut pouvoir à la fois transmettre le réel capté par le microphone, dépendant de sa sensibilité ; mais aussi avoir l'opportunité de retravailler ces prises de son pour faire rencontrer les matières, les points d'écoute (Dallaire, 2017, p.165). Le concepteur sonore témoignait d'ailleurs sur la prise de son de ce film :

« Nous avons conservé tous les sons enregistrés au moment du tournage. Il y a une tendance aujourd'hui à remplacer le son original parce qu'il contient ce que l'on considère comme des défauts - tous ces bruits qui ne sont pas liés au film lui-même. Gus était vraiment en amour avec l'immédiateté du son original. De plus, j'ai eu l'idée d'enregistrer en stéréo, ce qui est généralement une technique utilisée pour la musique ou les effets sonores, mais pas pour la prise de son lors du tournage d'un film. La stéréophonie crée une ambiance très singulière. »

(Shatz, 2005)

Avant toute chose, on remarque ici une pensée contraire, dans le principe, à celle de Daniel Deshays, qui appelait de ses vœux à « retirer des sons » et à « introduire l'idée de

³⁴ « Start by listening to the sounds of your body while moving. They are closest to you and establish the first dialogue between you and the environment. If you can hear the quietest of these sounds you are moving through an environment that is scaled to human proportions. In other words, with your voice or your footsteps you are 'talking' to your environment, which in turn responds by giving your sounds a specific acoustic quality. »

démontage son » (2010, p.84). Cependant, le contexte de la déambulation se prête davantage à la restitution d'un point d'écoute et de tous les bruits contacts entre l'individu et l'environnement. L'écoute microphonique, par son décalage avec nos mécanismes perceptifs, « illumine, hallucine notre rapport au monde » (Deshays, 2006, p.83). Leslie Shatz a utilisé un micro stéréophonique MS, proche de la caméra, et des « micros-cravate » pour la prise de son de *Gerry*. Il propose alors une restitution brute de sa captation, en adaptant les différentes matières pour recréer un point d'écoute cohérent lors du début de la déambulation. La sensation naturaliste domine. Puis au cours du film, la confiance des deux hommes pour retrouver leur chemin diminue. Ne voulant pas nécessairement se l'admettre, l'errance devient survie. Les corps fatiguent, l'environnement paraît de plus en plus hostile. La mise en scène et le choix des décors vont d'ailleurs de pair. La vitalité des corps et la vitalité des sols décroissent parallèlement.



Figure 12 - La lente dévitalisation des corps et des sols.

Les sons captés en stéréophonie, à une certaine distance des personnages, ont tendance à fusionner avec l'environnement par delà la réverbération (réflexions sur le sols, sur les rochers). Cette fusion se dégrade alors au fur et à mesure de la déambulation pour scinder finalement le point d'écoute. Sur le désert de sel, on ne retrouve plus qu'une ambiance désertique, distordue par des effets de mixage, et de brefs bruits de respirations

en proximité. Le dialogue, évoqué par Westerkamp, est rompu entre les deux Gerry et l'environnement. Les différentes matières, captées lors du tournage, sont modulées progressivement pour désincarner le point d'écoute. Les respirations peuvent même nous rapprocher d'une perception intériorisée des personnages³⁵. Le désert, troisième personnage de l'histoire, nous fait part de sa rudesse. Il recouvre par intermittence les faibles sons que l'on pouvait encore percevoir des deux hommes. Notre attention guète les quelques signes de vie, mais les pieds se font de plus en plus lourds jusqu'à ne plus pouvoir avancer pour l'un des personnages, et les respirations s'enfouissent dans le paysage sonore. L'environnement n'est finalement plus à leur hauteur (dans le sens du *soundwalk* de Westerkamp). Finalement, le mélange de ces deux points d'écoute, disjoints, tord l'espace durant le long plan-séquence sur le désert de sel. Il y a cohabitation de ces deux perceptions, de ces deux matières qui nous racontent l'affrontement entre le corps et l'environnement.

La transmission d'un point d'écoute, au vu de ces trois films, apparaît donc comme ayant un pouvoir expressif, vers des possibilités diverses. Il s'inscrit évidemment dans un processus de mise en scène plus large, mais semble être déterminant dans la sensation éprouvée par les spectateurs face au film. Dans ces exemples, le point d'écoute est manipulé de différentes manières, à des fins sensorielles, expressives, narratives. Ces différents traitements, confrontés au champ image, proposent un angle de lecture particulier, une valeur ajoutée dans la réception du film. On place alors le spectateur dans une posture attentionnelle active, dans le sens où il parcourt le champ sonore de son écoute, qu'il ait recours à une attention sélective hautement sollicitée (*Victoria*) ou qu'il n'ait pas nécessairement besoin de produire cet effort attentionnel, grâce à un point d'écoute se rapprochant des données à entendre (*Gosford Park*). Dans tous les cas, il apparaît que la seule reconstitution d'un point d'écoute, respectant la loi proxémique, ne témoigne que de l'environnement extérieur. Pour le cas de *Gerry*, la mise en place de points d'écoute disjoints, en surimpression sonore, exprime le rapport entre l'environnement et les

³⁵ Je pense ici au film *Le scaphandre et le papillon* (2007) de Julian Schnabel. Il met en scène un personnage qui, après un long coma, perd toutes ses capacités motrices, excepté le clignement de l'oeil. Pascal Villard, le monteur son, a alors reconstruit la perception intérieure du personnage, à l'aide de son de battement de coeur et d'écoulement sanguin dans le corps. Les sons ne prennent donc pas l'espace, ne dialoguent plus avec l'extérieur, et restent enfouis dans le seul corps de l'individu.

personnages et donc reflète aussi l'état des individus. Il me semble cependant que le(s) point(s) d'écoute, aussi distordu(s) soi(en)t-il(s), ne reflète(nt) pas en son sein des modulations attentionnelles telles que décrites dans la partie I. J'entends par ici une esthétique s'inspirant du réalisme sonore psychologique, inscrivant dans le mixage des mouvements de l'attention probable, « réaliste », d'un personnage ; témoignant d'un trait d'esprit particulier. Cependant, il est vrai que le traitement du point d'écoute dans *Gerry* nous témoigne autant de l'hostilité grandissante de l'environnement que de l'incapacité pour les personnages à avancer.

Alors, analysons maintenant les possibilités expressives des modulations attentionnelles inscrites, par le travail du monteur son et du mixeur, dans le champ attentionnel proposé au spectateur.

2. Les modulations attentionnelles

« Il y a signification lorsque les données du monde sont par nous soumises à une déformation cohérente. »

(Merleau-Ponty, 1960)

La phénoménologie pense le rapport entre l'expérience vécue et le sujet. Ce courant philosophique cherche le sens de l'expérience au travers de la perception de celle-ci par le sujet. La citation de Merleau-Ponty, qui s'inscrit dans cette démarche, nous éclaire alors sur l'existence d'une sémantique contenue dans notre subjectivité perceptive.

L'Homme, immergé dans le monde sonore, hiérarchise les données auditives sous une intention d'écoute, modelée d'abord par l'instinct de survie et évoluant face au contexte. Les mécanismes attentionnels sont les outils de la conscience, mais aussi du préconscient, pour opérer cette structuration sensorielle. Elle porte alors une signification en elle dans le rapport organisme/environnement.

Alors, le champ sonore au cinéma peut-il, par des déformations cohérentes, porter une signification sur l'individu face à l'univers fictionnel ?

a. Le preset contextuel

Bad Boy Bubby, de la perception primitive à la confrontation à un monde dense

Bad Boy Bubby (1993) est une comédie dramatique réalisé par Rolf De Heer.

Synopsis du film : Séquestré depuis sa naissance par sa mère, Bubby ignore tout du monde extérieur qu'il croit empoisonné. L'arrivée de son père, dont il était tenu éloigné, va bouleverser sa vie. Le jour de ses 35 ans, Bubby va enfin sortir. Il découvre un monde à la fois étrange, terrible et merveilleux où il y a des gens, de la pizza, de la musique ou encore des arbres.

Bubby, un homme de 30 ans n'a jamais vu le monde extérieur. Sa mère l'a contraint psychologiquement à l'enfermement. Mais un jour, il découvre l'ampleur de l'escroquerie, tue sa mère, et découvre les beautés du monde. Face à un contexte riche, comment Bubby, vierge de tout repère, réagit-il?

À l'image de l'allégorie de la caverne de Platon, Bubby n'a pas connaissance du monde extérieur. Il ne connaît que la pièce dans laquelle il vit, avec sa mère. En faisant croire que le monde extérieur est empoisonné, elle le maintient alors chez elle, dans une autre réalité. Pour appuyer ce mensonge, la mère sort de chez elle constamment masquée, et répète sans cesse qu'il doit rester ici.



Figure 13.1 - Bubby, enfermé par sa mère, seul chez lui.

Les seuls sons qu'il entend, au début du film, sont donc des sons provenant de l'intérieur de la pièce (les sons d'interaction avec sa mère, le frigo, les couverts, etc...). Puis, des sons métalliques, provenant de l'extérieur, attirent finalement son attention. Ces sons

sont difficilement interprétables. Ils ne font écho à aucune connaissance que pourrait avoir Bubby. On le sent tout de même curieux vis-à-vis de ces sons émanant d'un monde qu'il ne connaît pas. Il réoriente son regard, sa tête, pour tenter de mieux les discerner, de comprendre d'où ils viennent (nous avons déjà vu que les mouvements de tête sont précieux pour localiser un son). On peut comprendre ici les premières envies de Bubby d'aller au-delà de ses quatre murs. C'est alors son écoute sélective qui va transpercer les cloisons, à défaut de ne pouvoir le faire physiquement. Puis, un jour, son père toque à la porte. Sa mère, hésitante, finit par ouvrir la porte. Bubby découvre pour la première fois une nouvelle personne, extérieure à son monde. Le père apparaît sans masque, et plutôt en bonne santé.



Figure 13.2 - *L'arrivée du père.*

L'arrivée du père constitue le déclic chez Bubby. Quelques jours après, il tue ses parents et va explorer le monde extérieur. Comme un moyen d'apprentissage, il va imiter le comportement d'individus, croisés ici ou là. Mais il va être confronté à de nouveaux stimuli qui vont attirer son attention.



Figure 13.3 - *Bubby entend une chorale et court à leur rencontre.*

Il suit notamment le son d'une chorale qu'il entend au loin. La musique, forme d'expression qu'il découvre, semble l'attirer comme un aimant. En effet, le mixage du son de la chorale ne respecte pas la décroissance physique du niveau sonore. Elle résonne dans ce gigantesque espace de manière constante, en occupant tout l'espace sonore, pendant que Bubby court à la rencontre des chanteurs. Son attention auditive est entièrement occupée par cet élément sonore qui prend tout l'espace dans le mixage, loin d'un traitement naturaliste. C'est la première fois, depuis qu'il est sorti de chez lui, qu'on le voit faire preuve d'une attention endogène aussi soutenue. On comprend alors tout de suite l'intérêt, l'émerveillement, qu'il a face à l'écoute d'une chorale. Quand il arrive à leur niveau, le niveau sonore perçu est le même. La proximité est marquée simplement par une baisse de la réverbération dans le mixage. Cette séquence nous montre finalement les premières sensations que Bubby ressent, face aux beautés du monde dont il a été privé pendant 35 ans. Le plaisir, ici, est sonore, et le mixage tend à souligner d'avantage la projection attentionnelle de Bubby, et donc son caractère émerveillé par cette découverte, que la simple retranscription de la loi proxémique.

Juste après dans le film, nous retrouvons notre personnage entrer dans un bar assez bruyant. C'est la première fois qu'il est confronté à autant de contacts humains. Il découvre au fur et à mesure l'endroit, avant de s'installer à table. Il apparaît d'abord à l'image étonné, voir apeuré.



Figure 16 - Buddy entre dans le bar.

Son regard, puis les mouvements de caméra, naviguent de droite à gauche vers les différents groupes de personne. Mais l'ambiance sonore reste bruyante, massive, tout le long de la séquence. Là où les habitués sont familiers avec ces ambiances bruyantes et où il est commun d'user de l'effet *cocktail party* pour discerner certains éléments, Bubby reste distant et ne s'adapte pas au bruit ambiant. L'ensemble des voix paraît inintelligible et agressif à l'écoute. Ces différents regards pourraient suggérer qu'il scrute l'environnement en s'attardant sur des bribes de conversations. Mais le mixage de la séquence nous oriente plutôt vers une confrontation brutale entre le personnage et l'environnement. Il ne connaît pas ce contexte, et a du mal à s'y adapter. Il n'a pas intégré les formes sonores habituelles, ni les codes contextuels ; qui permettent rapidement de comprendre l'environnement et d'orienter son attention vers une cible précise. Il finira par manger une pizza, toujours plongé dans ce vacarme que la bande son nous fait entendre. Plus loin dans le film, Bubby finit en star de rock et côtoie régulièrement les bars. Alors, le mixage de ces séquences ne met pas autant en exergue les ambiances bruyantes sur tout le long des séquences, mais donne à entendre un chemin attentionnel plus orienté vers une écoute intentionnelle

« normale » (intelligibilité des dialogues, fixation de l'attention auditive vers l'endroit regardé). On comprend que Bubby s'adapte petit à petit à ce monde.

En somme, la bande son de ce film s'attache d'avantage à la perception et à l'attention du personnage principal qu'à la reproduction d'un champ sonore naturaliste. Les déformations de la perception par son attention (inscrites dans le mixage), pour reprendre les termes de la citation de Merleau-Ponty (1960) citée plus haut, sont signifiants. N'ayant aucune connaissance du monde extérieur, ces déformations cohérentes traduisent son état d'esprit, à la fois émerveillé par la musique et stupéfait par le brouhaha, face à la confrontation à de nouvelles interactions. Puis, le mixage traduit sa lente adaptation dans ce monde. Finalement, il nous donne autant à comprendre sur le personnage que sur son environnement.

b. L'attention en mouvement

The Revenant, la vigilance comme aptitude de survie

The Revenant (2015) est un film d'aventure réalisé par Alejandro González Iñárritu. Il utilise à plusieurs reprises le procédé du plan séquence pour mettre en scène le parcours du trappeur Hugh Glass.

Synopsis du film : Dans une Amérique profondément sauvage, Hugh Glass, un trappeur, est attaqué par un ours et grièvement blessé. Abandonné par ses équipiers, il est laissé pour mort. Mais Glass refuse de mourir. Seul, armé de sa volonté et porté par l'amour qu'il voue à sa femme et à leur fils, Glass entreprend un voyage de plus de 300 km dans un environnement hostile, sur la piste de l'homme qui l'a trahi. Sa soif de vengeance va se transformer en une lutte héroïque pour braver tous les obstacles, revenir chez lui et trouver la rédemption.

Je vais me concentrer pour cette partie sur le mixage du plan séquence où Hugh Glass se fait attaquer par un ours au milieu de la forêt. Dans cette séquence, le personnage arpente seul une partie de la forêt. Un sentiment d'insécurité va s'installer progressivement.



Figure 14.1 - *Hugh Glass marche seul dans la forêt.*

Le sentiment de danger est engendré par le contexte de la séquence. Le personnage, seul en forêt, entend ce qui semble être une présence animale. L'irruption du son ponctuel (qui sollicite les neurones phasiques et alerte le personnage) n'est pas accompagnée d'une visualisation de la source. Le personnage, comme le spectateur, se retrouve alors dans l'incertitude face à cette situation. Quelle est la source sonore entendue? Représente-t-elle un danger? Cette première question en entraîne nécessairement une deuxième : d'où vient la source? Le contexte de la séquence, ainsi que l'attitude du personnage, nous mènent à penser qu'il y a un danger potentiel dans les alentours. Les premiers sons entendus, n'ayant aucun support visuel, alimentent seulement notre imagerie mentale. On essaie alors de rationaliser cette inconnue dans un espace, un contexte donné. Le travail cognitif se met en place en même temps que les mécanismes attentionnels. On est à la recherche de nouveaux éléments visuels et/ou sonores pour lever les incertitudes, afin de mieux se préparer face à l'éventuel danger. L'attention auditive est alors vivement sollicitée, en parallèle de l'attention visuelle. On essaie de discriminer les éléments du champ attentionnel pour repérer au milieu de l'environnement sonore ce qui pourrait se rapporter à un danger. Tout ce qui est de l'ordre du connu, de l'identifié, n'apportant aucunes informations supplémentaires nécessaires à la survie du personnage est alors rejeté au profit de sons se rapportant au potentiel danger (il y a filtrage sensoriel). L'écoute du personnage augmente en focalisation, et en rapidité de mouvements auditifs. Il est dans un état de vigilance. Les mouvements de tête l'aide aussi dans la localisation de l'intrus.



Figure 14.2 - Hugh Glass scrute le champ sonore et visuel, pour cibler l'intrus auditif.

Par la suite, le personnage aperçoit au loin des ours, ce qui valide la présence d'un danger et justifie d'autant plus l'attitude du personnage. Les mécanismes perceptifs se concentrent alors sur le profil de l'ours. L'attention auditive est portée sur des sons pouvant émaner de l'ours (respirations, pas dans la forêt, ...).



Figure 14.3 - Hugh Glass met en joue l'ourson.

L'écoute focalisée est interrompu par l'apparition d'un son de respiration et de grognement d'ours (modulation attentionnelle exogène), d'autant plus discernable que l'on se trouvait dans l'attente de ce son, au profil formantique et temporel connu. L'appréciation de la distance de la source, plutôt courte en comparaison des autres sons entendus, signale le danger imminent. On découvre alors visuellement l'ours lorsque le personnage se retourne face à l'animal. La découverte visuelle du danger se fait donc quasiment au même moment entre le spectateur et le personnage. Il n'y a hélas plus assez de temps pour pouvoir armer et se défendre de l'animal qui fonce sur le personnage.



Figure 14.4 - Hugh Glass se retourne et aperçoit l'ours lui foncer dessus.

Durant toute la séquence, la mise en scène rapproche point d'écoute et point de vue du personnage, du point d'écoute et du point de vue du spectateur. Il n'y a pas total adéquation entre ces points d'écoute et points de vue. On ne se trouve pas dans un point de vue subjectif (par exemple lorsque l'on entend l'ours face à nous, le personnage l'entend derrière lui), mais on semble être au côté du personnage. Mais s'il n'y a pas exact adéquation des points de vue et des points d'écoute, la mise en scène semble amalgamer l'état psychologique du personnage avec le ressenti du spectateur. Plus précisément, il semble y avoir adéquation entre les modulations attentionnels du personnage lors de la séquence et le champ attentionnel proposé au spectateur. On se trouve à ses côtés et on vit le même cheminement perceptif que lui. En effet, les ambiances continues viennent décroître progressivement au profit de légers sons ponctuels pendant que la présence d'une menace se fait sentir. La rapidité des mouvements de caméra qui balaye tout l'espace rend compte de la vitesse des mouvements de l'attention visuelle et auditive. La localisation des sources semble complexe, on se trouve dans une phase de recherche, d'attente, voire d'hallucination auditive (on croit entendre l'ours), jusqu'au moment où le grognement de l'ours vient figer avec netteté les mouvements de l'écoute focalisée. Le danger est localisé. On nous donne à entendre une évolution de différentes masses sonores, qui s'apparente à l'anamorphose du champ attentionnel sonore naturaliste par les mécanismes de l'attention auditive - notamment d'attention sélective endogène pour la focalisation sur l'ourson mais aussi exogène pour la vigilance - face à un enjeu narratif, ici la survie du personnage. De plus, quand l'ours attaque Hugh Glass, le grognement vient prendre toute la place (il y a certainement utilisation de la divergence lors du mixage) sensorielle. L'effet de surprise et la peur vient saturer complètement la perception au profit des grognements de l'ours.

Les mouvements de mixage, décrits juste au-dessus, sont des faibles modulations qui permettent d'induire au spectateur le parcours attentionnel du personnage sans pour autant rompre avec la volonté d'immersion de la séquence (une modulations trop appuyée, ou trop rapide, sur les ambiances serait alors remarquée par le spectateur qui pourrait se déconcentrer du film³⁶). Ils sont surtout utile au début de la séquence pour installer le climat et l'état du personnage. Pour la suite, on accompagne le spectateur par les petites

³⁶ Claude Bailblé (1999, p.330) parle de modulations « *en mode sériel ralenti* ».

modulations d'effets sonores qui se précisent, puis par la proximité du grognement de l'ours. Alors, il semblerait qu'on éprouve d'autant plus la séquence en étant avec le personnage, mais aussi dans la tête du personnage.

En somme, les modulations attentionnelles des personnages de fiction, transcrites dans le mixage, semblent suggérer aux spectateurs l'état d'esprit d'un personnage dans un contexte extérieur, et son évolution au cours de la séquence (c'est le cas pour *The Revenant*) ou à l'échelle du film (notamment pour *Bad Boy Bubby*).

Ces modulations de mixage semblent orienter le spectateur, par une écoute « pré-focalisée », vers un parcours attentionnel établi par le metteur en scène. Elles sont cohérentes avec l'attitude attentionnelle et corporelle du personnage et le contexte de la séquence. Par exemple, c'est la simultanéité des orientations de la tête du personnage vers les éléments scéniques en parallèle des modulations de mixage qui souligne la réaction du personnage et apporte une valeur ajoutée³⁷. Cela nécessite un travail sur la saillance des événements signifiants pour le personnage (et/ou la narration). Ces mouvements de mixage nous renvoient à notre propre expérience attentionnelle, et de fait projettent la sensation, l'humeur, l'état d'esprit, que l'on imaginera ressentir dans ce contexte, sur le personnage. Les déformations cohérentes des données sensorielles dans l'univers fictionnel, portées par le mixage, semblent avoir des incidences sémantiques dans l'interaction entre cet univers et le personnage. Claude Bailblé (1999, p.330) précise que :

« La mise en scène n'ayant pas pour objet de faire voir et entendre l'entier du lieu scénique, en une sorte de captation exhaustive, elle tend et détend en permanence les ressorts de l'expectation en profitant de l'invisibilité (délibérée) et de l'intermittence (voulue ou assumée) des sources sonores. Qu'elles soient *in* ou *off*, dès lors importe peu : étant données à entendre par l'instance narrative, elles opèrent activement dans la conduite du récit. »

Le choix du parcours attentionnel, dans une volonté de sélection et d'inhibition des divers éléments scéniques par notre attention, semble donc déterminant puisqu'il interfère avec la narration, en ajoutant une valeur sémantique soutenant le récit.

³⁷ Par exemple, un son imprévu, ou au contraire recherché, attire davantage notre attention qu'un son insignifiant dans un contexte donné (cf. Les filtres attentionnels). Alors, la façon dont un son émerge du fond ambiant et attire l'attention du personnage peut nous renseigner sur le lien entre cet élément et le personnage.

PARTIE III. PARTIE PRATIQUE DU MÉMOIRE

Cette troisième partie sera consacrée à un test psychoacoustique pour infirmer ou confirmer l'hypothèse avancée au fil de l'étude, et que je rappelle ici :

Transcrire les modulations attentionnelles du personnage dans le mixage influe sur notre perception de l'état d'esprit du personnage. De plus, ces modulations de mixage, à un contexte fictionnel donné, peuvent orienter notre perception du personnage vers un état d'esprit cohérent avec ses modulations attentionnelles supposées.

Pour cela, nous allons établir un protocole de test, en définissant les stimuli à présenter, ainsi que les questions à poser aux différents sujets composant le panel.

A. Protocole du test

Pour ce test psychoacoustique, nous allons proposer à un large panel un unique plan-séquence pour trois mixages différents. Ces mixages seront motivés tout d'abord par l'étude de notre comportement attentionnel face à un contexte donné (pour l'aspect théorique), puis mis en forme sous l'inspiration des différentes bandes sonores étudiées préalablement.

Ensuite, les sujets seront confrontés à l'une des trois versions du plan-séquence puis répondront à une série de questions orientées sur leur cheminement attentionnel, leur perception de l'état d'esprit du personnage et du champ organisme (du personnage)/environnement. Voyons tout cela en détails.

1. Les stimuli

Tout d'abord, il faut définir les différents stimuli à proposer au panel. Premièrement, il m'apparaît nécessaire que la mise en scène du plan n'oriente pas trop explicitement les enjeux narratifs, auquel cas le champ des possibles se retrouve réduit. Il y a alors un fort risque pour que le mixage n'ait qu'une mince influence sur la perception du plan et du personnage. Aussi, le plan utilisé ne doit pas s'inscrire dans un contexte plus large, au milieu d'autres plans amenant leur lot de contexte et de développement du personnage.

Deuxièmement, il faut que j'aie accès à la matière première de la prise de son pour pouvoir moduler les différentes couches sonores entre elles (la prise de son et les ajouts de montage son).

Alors, le plus simple est de tourner son propre plan-séquence pour pouvoir répondre au cahier des charges. J'ai finalement opté pour mettre en scène une simple déambulation urbaine d'un personnage. Tout d'abord, c'est un milieu assez commun que tout le monde peut s'approprier. Ensuite, c'est un univers qui peut être riche en éléments sonores. Et j'ai besoin de cette matière pour construire différents chemins attentionnels. De plus, la déambulation me permet de ne pas montrer tout le temps le visage du comédien à l'image (on peut le filmer de dos) pour donner le moins d'indication possible sur l'état d'esprit du personnage, que peut traduire l'expression du visage. Enfin, cette mise en scène me permet de rapprocher au plus le point d'écoute du personnage avec mon système de captation sonore, pour pouvoir recréer son point d'écoute cohérent avec l'environnement. C'est une condition préalable pour traduire ensuite les modulations attentionnelles du personnage dans le mixage.

Ensuite, je compte proposer trois mixages différents. La première version sera appelée la version « objective ». Bien que cette notion d'objectivité soit galvaudée, ne serait-ce que par la distorsion de la perspective via la focale de la caméra, ce mixage essaiera de rendre les perspectives sonores cohérentes au point d'écoute du personnage, plongé dans l'environnement urbain. Cette version a pour but d'étalonner le ressenti des sujets, dans une esthétique sonore naturaliste, afin de pouvoir quantifier les éventuels variations de perception pour les versions suivantes.

La seconde version sera appelée la version « endogène », en référence à l'appellation de Posner (1998). Elle aura pour but de faire entendre un parcours attentionnel motivé par des modulations principalement descendantes. Le personnage restera attentif à des éléments précis, en inhibant le champ ambiant. Cette intention sera transcrite dans le mixage. On imagine alors souligner des interactions particulières entre certains éléments de l'environnement et notre personnage, et donc orienter un certain état d'esprit.

La troisième version sera appelée la version « exogène », toujours en référence à l'appellation de Posner (1998). Elle aura pour but de faire entendre un parcours attentionnel motivé par des modulations principalement ascendantes. Mon intention est que la plupart des événements sonores capteront l'attention du personnage, chassant immédiatement de sa conscience l'événement précédent. On imagine alors souligner un caractère distrait, ou vigilant (comme il est décrit dans la partie I.).

2. Le questionnaire

Le questionnaire s'articule en trois questions principales. Tout d'abord, je demande de « **lister les différents éléments qui attirent votre attention dans l'extrait** ». Les réponses me permettront par la suite d'analyser le parcours attentionnel effectivement emprunté par l'ensemble du panel. Il est en effet important de s'assurer en premier lieu que les modulations de mixage ont eu un réel impact dans la perception de l'extrait par les sujets.

Je questionne ensuite « **comment le personnage semble percevoir l'environnement dans lequel il s'inscrit** » puis « **dans quel état d'esprit semble être le personnage (répondre par mots clefs)** ». Ces deux questions vont déterminer alors le résultat de notre test. La première est plus axée sur la perception du champ organisme/ environnement, soit la relation entre le personnage et son environnement. La deuxième porte sur l'état d'esprit du personnage en soi. Les questions étant ouvertes, je m'attends à un large champ de réponses qu'il faudra organiser. Même si les deux questions peuvent être traitées séparément, les réponses sur la perception de l'environnement m'aideront certainement à préciser les réponses sur l'état d'esprit, pour mieux les catégoriser.

B. Mise en place du test

Pour que ce test soit significatif, je m'étais fixé l'objectif d'obtenir au moins trente réponses par stimuli, soit un total d'au moins quatre-vingt-dix sujets. Si j'avais initialement prévu de l'organiser dans les locaux de l'école, le contexte sanitaire de cette année 2020 en a décidé autrement. J'ai donc opté pour un test à réaliser en ligne. Les mixages seront donc réalisés en stéréophonie. Pour essayer de maintenir une rigueur de diffusion, j'ai donné quelques instructions comme la proscription des téléphones portables ou bien la nécessité de diffusion en plein écran.

1. L'écriture du plan

J'ai réfléchi pour ce plan à des événements qui surviendraient pendant la déambulation et que je pourrai maîtriser, comme des éléments à ajouter en montage son (des voix ou un vélo qui passe en hors-champ) ou le passage d'une passante proche de mon personnage (je me doutais qu'à la vue d'un dispositif de tournage, les piétons s'écarteraient, ce qui trahit le dispositif). Aussi, je voulais quand même une situation initiale à ce plan pour poser le décor. Mon personnage, en début de plan, sera donc adossé contre un mur, fumant une cigarette, pendant qu'il entend des voix de passants au loin. Puis, il commencera sa marche sur le trottoir. Je voulais aussi qu'il traverse la route pour obtenir un événement supplémentaire comme un passage de voiture. Enfin, il arrive sur une petite place et croise une fille. Pendant la déambulation, le personnage tournera quelquefois la tête pour donner la possibilité de jouer des réactions d'orientation suite à un événement hors champ (elles peuvent aussi être comprises comme une simple contemplation de l'environnement urbain).

Les différents événements devaient être ambiguës pour ne pas trop orienter les sujets vers des explications claires des agissements du personnage. Par exemple, on ne sait pas pourquoi il est adossé là et pourquoi il commence à marcher. Aussi, lorsqu'il croise la fille, il peut être éventuellement attiré ou bien totalement indifférent. On peut alors s'imaginer toute sorte d'éventualités, et donc, *in fine*, d'états d'esprit du personnage.

2. Le tournage

Pour le tournage de ce plan, j'ai opté pour une prise stéréophonique (couple ORTF) au plus près du personnage pour avoir une image sonore cohérente avec l'environnement, depuis l'écoute du personnage. En plus, j'ai équipé le comédien d'un micro cravate (SANKEN COS-11) afin de récupérer de la présence, du souffle. Enfin, un micro supercardioïde (MKH-60) sur perche suivra les pas du personnage.

Le jour du tournage, on a fait environ une dizaine de prises. Il y avait pas mal de monde et la pluie s'est invitée, ce qui nous a pas aidés. Finalement, l'une de prise était réussie. En plus des événements prévus, d'autres se sont ajoutés comme un passage de scooter quand le personnage arrive près de la grande rue, un passage de deux enfants et « leur mère » qui discutent et des voix de commerçants sur la place, en fin de plan. J'ai donc choisi ce plan. Il dure 1 minute et 1 seconde, ce qui me semble être la bonne durée pour ce test (on a le temps de poser une situation, les sujets peuvent réfléchir sur le personnage, sans pour autant avoir le temps d'oublier ou de se déconcentrer)

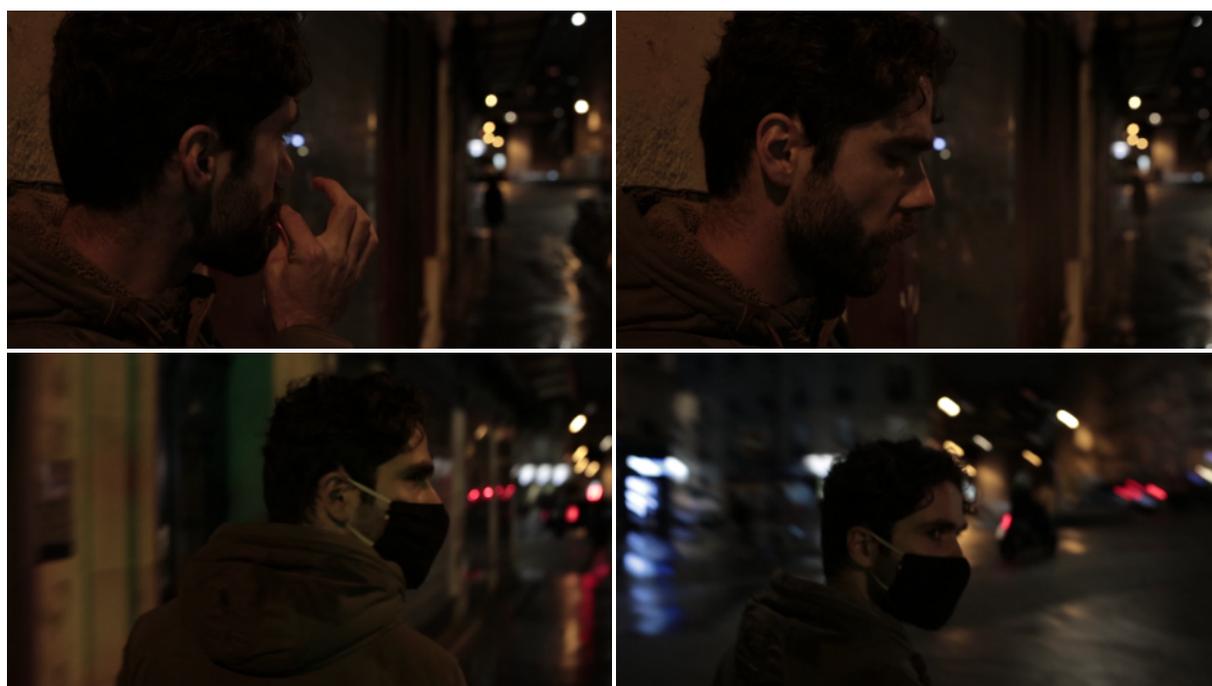




Figure 15 - Photogrammes du plan-séquence utilisé pour le test psychoacoustique

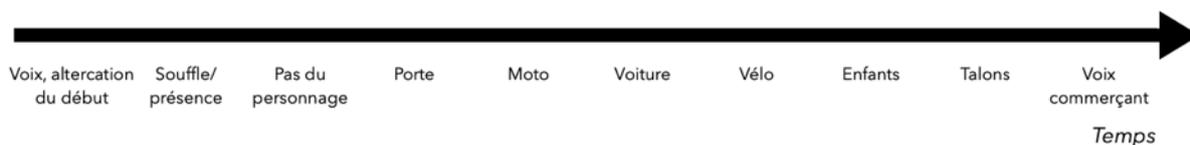
3. La post-production

a. Le montage son

Pour le montage son, la réflexion s'est portée sur le rythme de la séquence. Il me fallait un certain nombre d'événements sonores pour pouvoir ensuite construire mes différentes bandes son selon mon cahier des charges. Je devais pouvoir notamment construire un parcours attentionnel riche pour la version « exogène », où l'attention du personnage est d'avantage soumise à des modulations ascendantes.

Les sons directs me donnaient déjà, en plus des présences du comédien et de l'ambiance, les passages des voitures et de la moto, les voix des enfants, les pas des passants et notamment les pas de la fille (que l'on a aussi enregistré en son seul), ainsi que les voix des commerçants en fin de plan. J'ai ajouté à cela les voix des personnes qui semblent se crier dessus, prévues au script, ainsi que le claquement d'une porte synchronisée sur un regard à droite du comédien (pour simuler une réaction de l'attention) et enfin le passage d'un vélo en hors-champ. En dehors des présences du personnage, le montage son comporte donc huit événements sonores, soit une unité de plus que le nombre d'élément que l'on peut mémoriser dans notre mémoire à court terme (Miller,

1956). La quasi-totalité des événements, s'ils sont retenus par les sujets, peut donc être mentionnée dans le questionnaire. Néanmoins, ces éléments ne seront pas forcément tous perçus selon les différents mixages.



b. Le mixage « objectif »

Pour ce mixage, l'intention naturaliste est primordiale. Il nous servira pour la suite de référence à notre test. La bande son est donc majoritairement basée sur la captation stéréophonique, à niveau nominal tout le long de la séquence. J'ai ensuite ajouté un peu de micro perche et de micro cravate pour amener la présence et les pas qui manquaient dans notre couple stéréophonique. J'ai ensuite complété, toujours dans cette logique naturaliste, les éléments absents de la prise de son. Les voix du début de la séquence s'intégraient parfaitement, les ayant enregistrées avec le même système de captation stéréophonique dans une rue. Tout cela constitue au final mon premier stimuli.

c. Le mixage « endogène »

Pour la version « endogène », il faut établir au préalable un parcours attentionnel que le personnage peut emprunter. J'ai alors décidé d'axer celui-ci sur trois éléments principaux. Tout d'abord, le personnage entend les voix en hors-champ. Il focalise ensuite son attention sur cet élément pendant le premier tiers du plan. Cette attention est volontaire, il surveille ce flux de manière intense. Puis, le passage du scooter à proximité de lui vient rompre soudainement cette attention endogène. Il traverse ensuite la route et voit au loin la fille. Il focalise alors de nouveau son attention sur cette fille, et le bruit de ses talons jusqu'au croisement des deux personnages. Le reste des éléments, bien que

continuellement perçus et surveillés (comme nous l'avons vu dans la partie I.) n'atteignent pas sa conscience.

À l'instar du plan-séquence de *The Revenant* (2015) réalisé par Alejandro González Iñárritu, il faut réussir à transcrire les modulations attentionnelles au sein du mixage. Pour les sélections attentionnelles, il faut travailler le rapport sons cibles/champ ambiant pour faire gagner en saillance nos éléments prévus dans le parcours attentionnel. Le champ ambiant devient plus lointain, plus flou, davantage masqué par les sons cibles. Pour la réaction de l'attention au passage du scooter, la transcription des modulations doit se faire « *en mode sériel ralenti* » (Bailblé, 1999, p.330), pour ne pas se faire sentir par les spectateurs, et pour tout de même rester signifiant. L'échelle de temps de détection des qualités sonores, jusqu'à la captation de notre attention s'avère être de l'ordre de la dizaine à la centaine de millisecondes (cf. I.B.1. Les mécanismes ascendants). Les modulations de mixage doivent alors s'étendre pour ne pas commettre des ruptures brutales (cf. I.D.1. La saillance physique) et devenir plus saillantes que les éléments à rendre saillants. Donc, l'apparition du scooter et l'inhibition du champ ambiant sont modulées au mixage, dans un même temps, dans une approche ralentie comparée aux modulations attentionnelles. Par exemple, j'ai fait évoluer le rapport de niveau entre les voix du début de séquence et la prise stéréophonique d'une dizaine de décibels en environ deux secondes. Idem pour l'apparition du scooter.

Je pense que l'apport narratif de ce mixage est plus conséquent. Les choix de sélection et d'inhibition des sons tendent à orienter les interactions entre le personnage et son environnement. Ainsi, on peut supposer que le personnage commence à marcher pour échapper aux voix qu'il perçoit menaçantes tout en continuant à les surveiller, puis est attiré par cette fille qu'il voit au loin.

d. Le mixage « exogène »

La version « exogène » est axée sur la succession des différents événements qui attirent l'attention de notre personnage. Ainsi, un élément sonore chasse le précédent. Les voix du début de séquence, qui durent pendant le premier tiers de la séquence dans notre précédente version, sont atténuées une fois que le personnage jette sa cigarette et commence à marcher. Il n'y porte plus attention. L'intention de cette mise en scène sonore amène finalement le mixage à préciser les différents éléments sonores, par le travail en niveau mais aussi par le démasquage fréquentiel des sources masquantes sur les sources cibles, de manière dynamique. L'approche « *en mode sériel ralenti* » reste de mise.

Cette attitude attentionnelle du personnage donne à celui-ci un caractère plus distrait, ou vigilant, selon notre étude de l'attention auditive exogène. Autrement dit, la MMN devient plus ample, l'information arrive plus aisément à la conscience. Les mécanismes ascendants prennent le dessus sur les mécanismes descendants.

Vous pouvez retrouver des liens internet menant aux différents stimuli en annexes.

C. Analyse des résultats

Pour la version « objective », 31 sujets ont répondu aux questionnaires mais les réponses de 5 d'entre eux ont dû être enlevées des résultats parce qu'ils avaient utilisé le son de leur téléphone ou de leur ordinateur. Les réponses de 26 sujets ont donc été exploitées.

Pour la version « endogène », 32 sujets ont répondu aux questionnaires mais les réponses de l'un d'entre eux ont dû être enlevées des résultats parce qu'il avait utilisé les hauts-parleurs de son téléphone. Les réponses de 31 sujets ont donc été exploitées.

Pour la version « exogène », 34 sujets ont répondu aux questionnaires mais les réponses de 4 d'entre eux ont dû être enlevées des résultats parce qu'ils avaient utilisé le son de leur téléphone ou de leur ordinateur. Les réponses de 30 sujets ont donc été exploitées.

1. Les conditions d'écoute

Le test étant réalisé en ligne, les différents sujets n'ont pas écouté les stimuli sur les mêmes systèmes de diffusion. En amont du test, j'ai donné des consignes sur la diffusion des stimuli. Les sujets devaient regarder et écouter les stimuli en plein écran, avec un système de diffusion sonore stéréophonique convenable. J'ai notamment proscrit les téléphones portables ainsi que la diffusion sonore sur les enceintes des ordinateurs. Ils devaient aussi notifier quel système d'écoute ils utilisaient. Voici alors la répartition des systèmes d'écoute en fonction des différents stimuli.

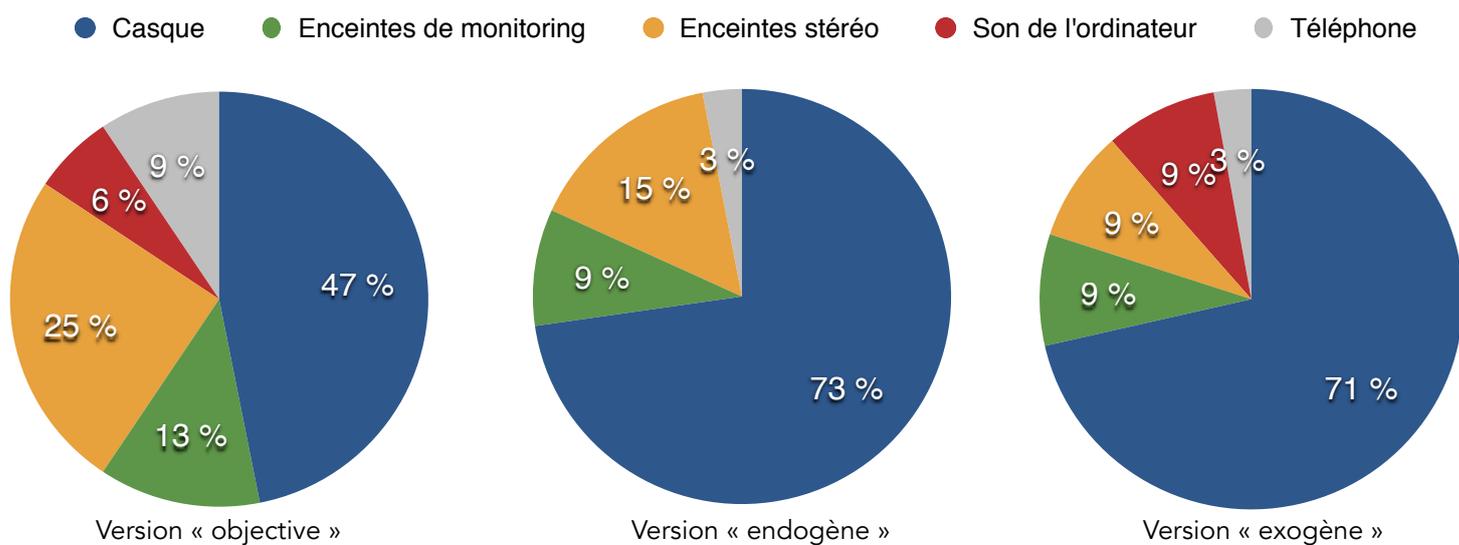


Figure 16 - Graphiques de la répartition des systèmes d'écoute

On remarque une majorité d'écoutes au casque. Les résultats pour les deux dernières versions sont d'ailleurs très proches, ce qui va dans le sens d'un lissage des résultats avec un nombre significatif de sujets. Aussi, les sujets ayant tout de même utilisés les hauts-parleurs du téléphone ou de l'ordinateur ont été enlevés des résultats.

2. Les parcours attentionnels

Après le visionnage du stimuli, le sujet devait d'abord lister les différents éléments qui attiraient son attention. L'ensemble des résultats m'a permis d'établir les différents parcours attentionnels empruntés par les sujets en fonction des différents mixages.

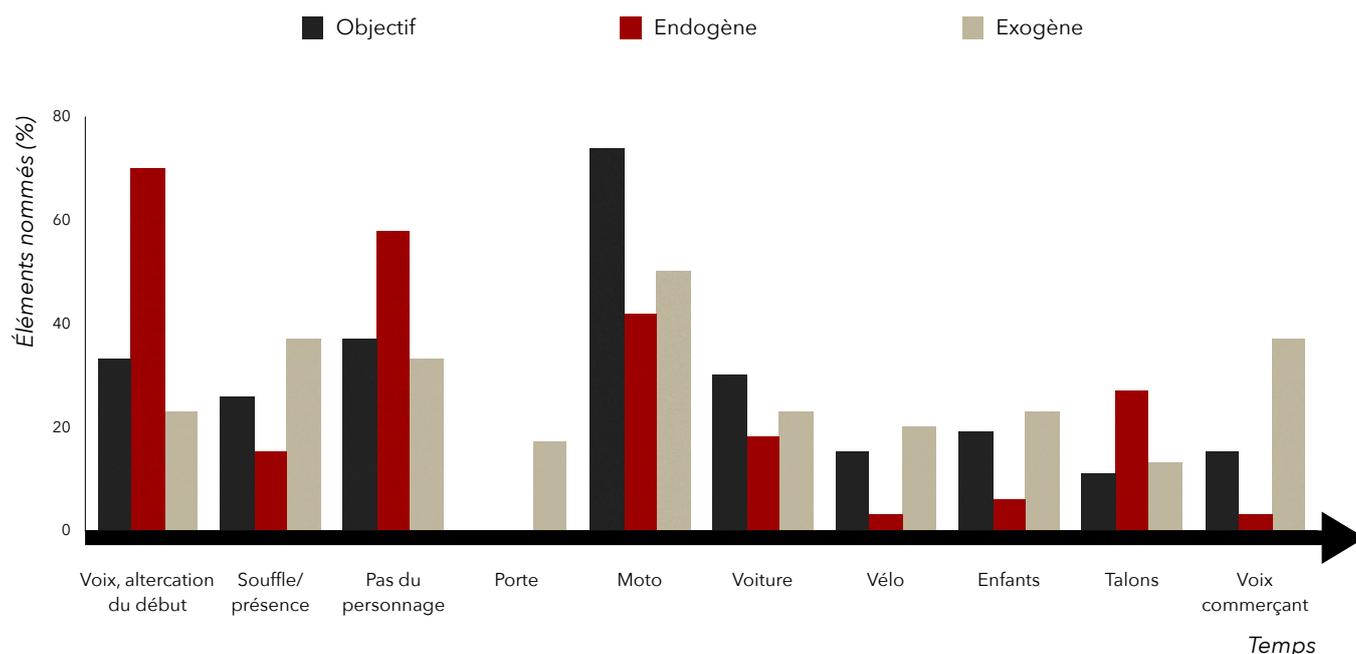


Figure 17 - Graphique du parcours attentionnel des sujets en fonction du mixage

a. Le mixage « objectif »

Ce premier stimuli nous aide à comprendre comment a réagit le panel face à un mixage d'esthétique naturaliste. On remarque d'abord une forte saillance du passage de la moto (trois sujets sur quatre l'ont cité). Cet élément sollicite fortement l'attention exogène des sujets, de par la rupture qu'il constitue dans le continuum sonore urbain. Autrement, les voix du début de séquence ainsi que les passages de voiture, le souffle et les pas du personnage ont été cités par environ un sujet sur trois. Les autres éléments ont été notifiés par moins d'un sujet sur cinq. La présence du personnage ainsi que les voix du début ont donc attiré davantage l'attention. On peut supposer que certains sujets ont notifié une interaction entre le personnage et ces voix, comparé aux autres éléments. Le regard du personnage vers ces voix hors-champ oriente aussi notre attention et souligne cette

éventuelle interaction. Une fois les qualités perceptives du flux sonore repérées, notre représentation mentale de l'élément en hors-champ est contextualisée puis questionnée (Qui sont ces personnes ? Représentent-elles une menace ? etc...). C'est la double scénarisation (Leblanc, 1991), l'aller-retour entre la perception des champs sensoriels et la projection cognitive. L'un des sujets, pour ce premier stimuli, pose même la question : « Il a l'air méfiant. Qu'est ce qui le motive à partir ? ». Néanmoins, les deux tiers des sujets n'ont pas notifié ces voix et ne semblent donc pas s'être concentrés sur cette éventualité.

Finalement, hormis la moto, l'attention des sujets semble s'être relativement « promené » dans le champ sonore, en retenant quelques éléments ici ou là. Il n'apparaît pas de parcours attentionnel affirmé par une majorité du panel.

b. Le mixage « endogène »

Pour le second stimuli, il apparaît que l'attention des sujets se soit orientée majoritairement vers les voix du début de séquence, ainsi que sur les pas du personnage. Ces voix se font entendre, dans cette version, jusqu'à l'arrivée du personnage au niveau de la grande route (soit sur un segment d'environ 25 secondes), alors qu'elles s'éteignent progressivement pour les autres versions (elles se retrouvent masquées au bout de 15 secondes dans la version « objective »). Parallèlement, le souffle du personnage (lorsqu'il fume sa cigarette) est moins notifié que pour le premier stimuli alors qu'il est au même niveau sonore. Outre l'effet de masque accru de ces voix sur le souffle du personnage, on peut supposer que l'attention sélective du panel concentrée sur les voix a inhibé de fait le souffle du personnage. Ainsi, comme nous l'avons vu, si l'attention se focalise nettement sur un flux sonore, les ressources attentionnelles ne peuvent se concentrer sur un autre élément simultané.

Une plus faible portion du panel a aussi notifié le son des talons de la passante (30% du panel). La saillance de ces éléments est donc significative, comparé à d'autres éléments (le vélo, les enfants, les voix des commerçants) qui ne sont nommés qu'entre 3 à 6% (soit un ou deux sujets). On remarque ici que pour un même champ visuel, l'élément prédominant

dans le champ sonore attire davantage l'attention (au détriment des autres éléments). Cette pratique de mixage est notamment utilisée, à des fins burlesques, dans *Playtime* (1967) réalisé par Jacques Tati, où l'on nous fait percevoir la chute d'une canne dans l'arrière plan visuel, mais au premier plan sonore. L'attention se retrouve directement happée vers l'élément.

Finalement, l'intention de mixage pour ce second stimuli est fonctionnelle. Le parcours attentionnel prévu initialement est majoritairement emprunté par les sujets. Il reste alors à analyser les potentielles incidences sémantiques et narratives du mixage sur le panel.

c. Le mixage « exogène »

Ce troisième stimuli présente un champ sonore beaucoup plus détaillé. Les éléments ponctuels ressortent davantage du continuum urbain. Mais il n'y a pas volonté, pour ce troisième mixage, de sélection de certains éléments et d'inhibition d'autres éléments (comme dans le mixage « endogène »). Le parcours attentionnel des sujets se rapproche alors de celui du mixage « objectif ». Il n'y a pas, à l'inverse du second stimuli, d'éléments nettement plus saillants en comparant avec la première version. On remarque cependant une légère hausse de l'attention portée sur différents éléments ponctuels (le vélo, les enfants, les talons de la passante, les voix des commerçants). Il y a même le son d'une porte qui se ferme, jusque là encore jamais cité, qui apparaît à hauteur de 18%.

L'attention exogène des sujets semble donc avoir été davantage sollicitée, par la succession des différents événements sonores. On peut aussi supposer que si le passage de la moto est moins cité que pour la version « objective », ce n'est pas forcément qu'elle est moins saillante en soi (son passage a d'ailleurs été augmenté en niveau lors du mixage pour simuler les processus attentionnels ascendants), mais que sa mémorisation en mémoire à court terme est parasitée par la présence d'autres éléments qui, eux, ont gagné en niveau sonore est donc en saillance. Cet élément reste d'ailleurs celui le plus cité.

On remarque finalement que, même s'il n'y a pas de parcours attentionnel précisément défini, l'attitude attentionnelle des sujets face à ce stimuli diffère de la version « objective ». Les sujets continuent à parcourir le champ attentionnel, sans grande tendance de cheminement particulier, mais leur attention exogène est davantage sollicitée, ce qui confirme notre intention de mixage. On peut supposer que, pour la suite de l'analyse, cette différence d'attitude impliquera chez les sujets une perception de l'état d'esprit du personnage davantage orientée vers le caractère vigilant, ou distrait, conformément à l'étude du comportement de notre attention auditive.

3. La perception du champ organisme/environnement

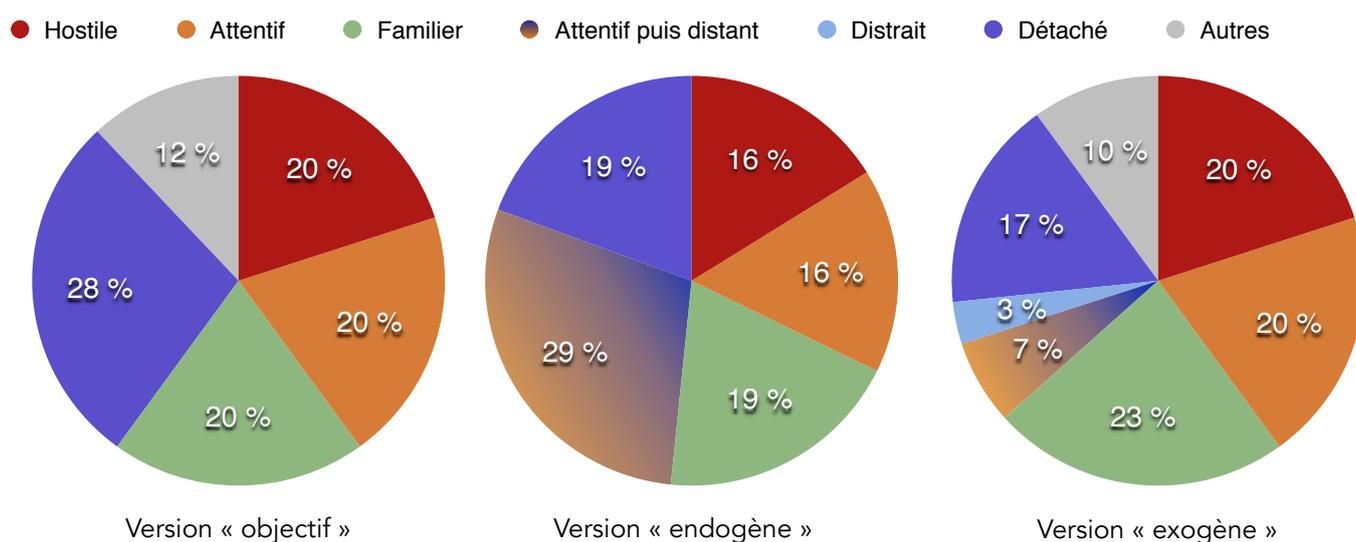


Figure 18 - Graphique sur la perception du champ organisme/environnement par les sujets en fonction du mixage

Le graphique ci-dessus représente l'ensemble des réponses, organisée par champs lexicaux, à la question « **comment le personnage semble percevoir l'environnement dans lequel il s'inscrit ?** ». Les réponses qualifiaient parfois l'environnement, parfois le personnage dans cet environnement. J'ai donc nommé l'ensemble des résultats « la perception du champ organisme/environnement » par le panel, l'organisme faisant référence au personnage de fiction et l'environnement au lieu scénique. Les réponses, détaillées et organisées dans des tableaux, sont consultables en annexes.

a. Le mixage « objectif »

L'intention de réalisation de ce plan-séquence est axée sur la neutralité, ou du moins l'ambiguïté du personnage au sein de son environnement, et a pour but de laisser une vaste place aux interprétations des enjeux de la scène. Le mixage de la première version des stimuli respecte cette direction en s'orientant vers une esthétique naturaliste. Mais l'interférence des champs sonore et visuel et de leur projection cognitive amène nécessairement une lecture particulière de la scène et de son personnage. Nous allons alors pouvoir établir une référence, avec ce premier stimuli, quant à la perception par le panel du champ organisme/environnement filmique.

Tout d'abord, nous remarquons une répartition équitable des catégories « hostile », « attentif » et « familier » qui représentent chacune 20% des réponses. La catégorie « détaché » est à 28%, soit deux réponses de plus. Ce stimuli inspire donc des réponses assez éloignées les unes des autres mais quatre champs lexicaux se forment. La catégorie « hostile » exprime une certaine hostilité de l'environnement ressentie par le personnage. La catégorie « attentif » signifie que le personnage est attentif à son environnement. L'ensemble des réponses, comme « il perçoit chaque son un par un », sous-entendent une attention vive, équilibrée entre les processus descendants et ascendants, qui focalise vers les événements successifs. La catégorie « familier » signifie que le personnage connaît son environnement, qu'il n'y prête pas nécessairement attention, qu'il est indifférent. L'inscription des filtres attentionnels négatifs pour un environnement connu libère l'attention auditive des événements prévisibles (cf. I.B.2.). La catégorie « détaché » signifie que le personnage est détaché de son environnement, qu'il n'y prête pas attention.

Ces différentes catégories impliquent diverses attitudes attentionnelles. On peut ainsi former deux familles sur cette distinction. Les catégories « hostile » et « attentif » sous-tendent une attention vive du personnage vers son environnement. Elle se regroupe derrière une attitude plutôt vigilante. Les interactions entre le personnage et son environnement sont fortes. Les catégories « familier » et « détaché » expriment, elles, une attention portée vers ses alentours moins vive, voire inexistante. Ces deux grandes distinctions sont à nuancer sur l'étendue et la progressivité des attitudes attentionnelles, mais elles restent de bons indicateurs quant au ressenti des sujets.

La répartition, presque équitable, des réponses dans ces catégories diverses nous montre les larges interprétations qu'ont eu les sujets, certainement provoquées par l'intention de réalisation. Ces premiers résultats sont à prendre comme référence pour la suite de l'étude. Nous analyserons ensuite les éventuelles différences significatives par rapport à cette version « objective », présentant un mixage naturaliste.

b. Le mixage « endogène »

Ce second stimuli porte dans son mixage un parcours attentionnel affirmé. Le personnage entend les voix au début de la séquence, et focalise alors son attention sur celles-ci jusqu'au passage du scooter qui vient détourner son attention. Puis, après avoir traversé la route, le personnage voit au loin une passante et focalise à nouveau son attention sur elle. Les autres événements sonores sont alors délaissés au « second plan attentionnel ». Ce cheminement attentionnel, transcrit dans le mixage, apparaît avoir été perçu par le panel comme nous l'avons vu précédemment. Alors, quelles implications a-t-il eu sur la perception du panel portée sur notre personnage au sein de l'environnement fictionnel ?

Nous remarquons tout d'abord l'apparition d'une nouvelle catégorie : « attentif puis distant », qui représente 29% des réponses. Elle regroupe un ensemble de réponses marquées par un ressenti évolutif du personnage au cours de la séquence. On note par exemple des réponses comme « attentif au début, indifférent ensuite », « attentif lorsqu'il fume, puis attitude passive dès qu'il marche » ou encore « impression qu'il est dans sa bulle, à part les voix du début qu'il semble écouter / qui semblent déclencher son départ ». Ces remarques, inexistantes pour la version « objective », soulignent une interaction perçue entre ces voix (perçues comme une altercation pour une majorité des sujets) et le personnage. On comprend alors ce phénomène lorsque l'on met en perspective, pour ce stimuli, d'une part, les résultats du parcours attentionnel emprunté par les sujets et, d'autre part, les 29% de réponses dans la catégorie « attentif puis distant ». Le panel a massivement porté son attention sur ces voix, l'amenant ensuite à les interpréter face à l'attitude du personnage. L'orientation de mixage, vers cette focalisation de l'attention, a donc eu un réel impact quant au ressenti du champ organisme/environnement. La

hiérarchisation des données sensorielles, motivée par une intention d'attitude attentionnelle particulière du personnage, a guidé l'interprétation des sujets.

Mais, si cette catégorie, représentant un petit tiers des réponses, souligne l'attention du personnage au début de la séquence, elle nous dit aussi que le personnage est ensuite beaucoup plus distant de son environnement, qu'il y prête moins d'attention. Alors, l'inhibition des autres éléments tels que le claquement de porte, le vélo, les enfants ou encore les voix des commerçants, transcrite dans notre mixage, a certainement influencé ce ressenti (le mixage a eu une influence sur le parcours attentionnel emprunté par le panel). Alors, en prenant en compte les résultats des autres catégories, une majorité du panel trouve notre personnage attentif à son environnement (principalement les voix du début), voire vigilant, au début de la séquence (61% en regroupant la catégorie « hostile », « attentif » et « attentif puis distant ») ; ce qui représente une hausse significative par rapport à la version « objective » (40% en regroupant la catégorie « hostile », « attentif »). Ensuite, le panel trouve notre personnage davantage détaché de son environnement pour la suite de la séquence. Presque une réponse sur deux entre dans la catégorie « détaché ou distant » de son environnement (48% en regroupant les réponses « détaché » et « attentif puis distant »). On monte à 68% des sujets, en incluant la catégorie « familier », qui trouvent le personnage inattentif ou neutre vis-à-vis des alentours (contre 48% pour la version « objective »).

En somme, la stratégie attentionnelle a fonctionné sur une majorité des sujets pour le début de la séquence. L'attention du personnage portée sur ces voix est nettement soulignée dans les réponses, en comparaison avec le mixage naturaliste. Par contre, même si le passage de la passante est plus cité que pour la version « objective » dans le parcours attentionnel du panel (mais n'excède pas les 30%), cela semble avoir eu aucun impact dans le ressenti du champ organisme/environnement. On peut essayer d'expliquer cela en deux points : à l'inverse des voix, le personnage ne tourne pas franchement la tête vers la fille (ce qui ne souligne pas l'interaction). Aussi, la courte focale utilisée lors du tournage rend la profondeur de champ assez floue, ce qui n'aide pas à se focaliser sur la passage de la passante qui arrive au loin. On peut alors conclure que, même si des stratégies attentionnelles peuvent être réalisées en post-production sonore, elles doivent être élaborées de paire avec la mise en scène globale du projet.

c. Le mixage « exogène »

Ce dernier stimuli présente un mixage où les différents événements sonores se détachent davantage du fond ambiant. Il est inspiré d'une attitude attentionnelle du personnage plutôt vigilante. Son attention auditive se focalise sur chaque événements successifs émergents du continuum sonore urbain. Il peut alors cibler volontairement un élément sonore, tout en étant suffisamment alerte pour surveiller le champ environnant. Au sein du mixage, les différents événements sonores se succèdent plus qu'ils ne cohabitent. Le niveau des ambiances est diminué par rapport à la version « objective », au profit d'un champ sonore plus détaillé. Les éléments ponctuels sont davantage démasqués. Cependant, ce mixage se rapproche plus de la première version du stimuli dans le sens où l'ensemble des événements sonores ont un niveau sonore cohérent entre eux (*a contrario* de la version « endogène »).

On remarque tout d'abord que les résultats sont proches de ceux de la version « objective ». On peut expliquer cela par la plus grande proximité des mixages. Autrement, on note l'apparition d'une nouvelle catégorie, « distrait », pour une seule réponse mais qui apparaît pertinente : « Il semble distrait. Il passe d'un élément à un autre sans y prêter attention. ». Malgré des résultats qui ne varient pas de manière significative avec la version « objective », beaucoup de réponses soulignent en effet une forte présence sonore des différents effets. On peut citer : « les sons de son environnement sont très présents », « il semble tout entendre », « très conscient de son environnement, la présence sonore du vélo renforce cette impression », « il passe d'un élément à un autre », « il semble percevoir très fort son environnement direct », « conscient de tout ce qui se passe », « il perçoit chaque son un par un ». Alors malgré ces détails remarqués par une partie du panel, il semble que ça n'ait pas eu d'influence dans les différentes qualifications proposées. Ainsi, certains rapprochent le fait de « tout bien entendre » avec la familiarité de l'endroit, comme d'autres y associent la méfiance du personnage à l'égard de son environnement, et ce, dans des proportions inchangées de la version « objective ». Il nous reste à savoir si ce mixage a eu une influence sur le ressenti du panel quant à l'état d'esprit du personnage.

4. La perception de l'état d'esprit du personnage

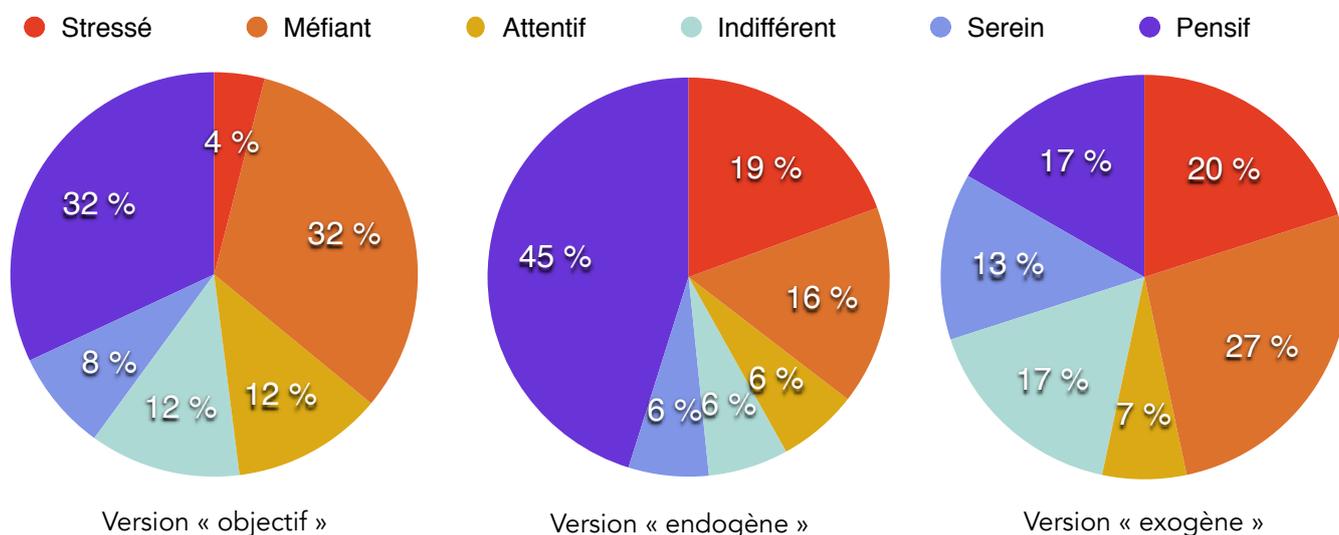


Figure 19 - Graphique sur la perception de l'état d'esprit du personnage en fonction du mixage

Le graphique ci-dessus représente l'ensemble des réponses, organisées par champs lexicaux, à la question « **dans quel état d'esprit semble être le personnage ?** ». Pour m'aider à catégoriser certaines réponses, je me suis quelquefois reporté aux réponses à la précédente question. Par exemple, la réponse « inquiet » peut faire référence à une inquiétude liée à un élément scénique (à placer dans la catégorie « méfiant »), ou bien à une inquiétude plus profonde sans rapport avec l'environnement (à placer dans la catégorie « pensif »). Mais la réponse à la question précédente était « sur ses gardes ». J'ai donc placé la réponse « inquiet » dans la catégorie « méfiant ». Les réponses, détaillées et organisées dans des tableaux, sont consultables en annexe.

Comme pour la précédente analyse, nous pouvons catégoriser les qualificatifs en deux familles, selon les attitudes attentionnelles. Les catégories « stressé », « méfiant », « attentif », associées aux couleurs chaudes, représentent une attitude attentionnelle vive, alerte, voir de vigilance. À l'inverse, les catégories « indifférent », « serein », « pensif », associées aux couleurs froides, sous-entendent une attitude attentionnelle pas particulièrement vive, où le personnage est détaché de son environnement extérieur.

a. Le mixage « objectif »

Cette première version continue à être notre référence. Les sujets ont répondu pour un tiers « méfiant », un tiers « pensif » pour les grandes tendances. 12% d'entre eux l'ont trouvé « attentif », 12% « indifférent », 8% « serein » et 4% « stressé ». On remarque aussi que la répartition des réponses selon la distinction sur l'attitude attentionnelle est quasiment équitable. 48% des sujets l'imaginent plutôt attentif, voire sur ses gardes ; 52% des sujets l'imaginent pas nécessairement alerte, voire complètement détaché de son environnement.

b. Le mixage « endogène »

On remarque pour cette seconde version une hausse des réponses le qualifiant de « pensif ». Nous avons vu précédemment que, hormis l'interaction avec les voix du début de séquence, le panel a jugé en majorité que le personnage était distant, détaché de son environnement. De plus, d'après l'analyse des différents parcours attentionnels, les sujets ont plutôt porté leur attention sur les pas du personnage (près de 60% d'entre eux contre moins de 40% pour les autres versions) et moins sur l'environnement extérieur global. On peut expliquer cela par l'inhibition des événements sonores (sauf les voix du début, le scooter et les talons de la passante) qui facilite l'émergence des bruits de pas du personnage. Il y a moins d'interactions avec l'extérieur suggérées au mixage, on l'imagine donc plutôt centré sur lui-même. Les différentes réponses aux questions apparaissent alors cohérentes entre elles. Ces choix de mixage ont donc favorisé la hausse des réponses de la catégorie « pensif ».

Plus globalement, les sujets ont ressenti notre personnage un peu moins alerte (en regroupant les catégories « attentif », « méfiant » et « stressé », on passe de 48% à 41%). Mais on remarque un déport des réponses de la catégorie « méfiant » pour le mixage « objectif » vers la catégorie « stressé » pour le mixage « endogène ». On peut supposer

que l'éventuelle focalisation de l'attention du personnage sur les voix au début (assimilées en grande partie à une altercation), largement évoquée par le panel, a influencé la perception de son état d'esprit vers la catégorie « stressé ». Ayant cet élément là particulièrement à l'esprit, les sujets semblent se projeter sur les éventualités qu'implique la présence d'une altercation à côté du personnage.

c. Le mixage « exogène »

Enfin, pour ce dernier stimuli, on remarque une baisse significative de la catégorie « pensif » (on passe de 32% pour la version « objective » à 17% pour la version « exogène »). On remarque aussi une hausse globale des catégories « stressé », « méfiant » et « attentif » regroupées (on passe de 48% pour la version « objective » à 54% pour la version « exogène »). Si l'on avait pas noté de différence significative avec la version « objective » pour la perception du champ organisme/environnement, il semble tout de même que ce mixage « exogène » ait eu une influence sur la perception de l'état d'esprit du personnage. La forte présence des divers éléments sonores dans le mixage soulignent la conscience de notre personnage face à son environnement, et, de fait, le rend moins « pensif ». Cependant, il n'y a pas de catégorie qui soit bien plus importante que toutes les autres. On peut conclure que, si ce mixage tend à inclure notre personnage dans son environnement, il n'oriente pas particulièrement la perception de l'état d'esprit du personnage. Chacun interprète à sa façon cette forte conscience qu'a le personnage de son champ environnant.

Mais ce résultat peut être pertinent dans un contexte narratif plus riche, où l'ensemble des éléments de mise en scène convergent vers une intention. Alors, la transcription des modulations attentionnelles dans le mixage peut être un outil supplémentaire dans la caractérisation d'un personnage de fiction.

CONCLUSION

En résumé, nous avons établi pour ce test différents mixages, motivés par des intentions d'attitude attentionnelle de notre personnage. Nous avons ensuite questionné si, premièrement, ces attitudes attentionnelles étaient perçues par le panel, puis, dans un deuxième temps, si elles ont un impact dans la caractérisation du personnage immergé dans le lieu scénique.

Notre première intention de mixage (pour le stimuli « endogène ») était de traduire une attitude attentionnelle motivée principalement par les processus descendants. Le personnage se focalise volontairement, dans une attention sélective soutenue, vers des éléments sonores prévus dès l'écriture de la séquence. Cette attitude attentionnelle implique le recul dans la « perspective attentionnelle » des autres éléments du champ ambiant. Autrement dit, ces éléments restent sous surveillance de son attention exogène mais n'atteignent pas sa conscience. Alors, cette intention nous mène à inscrire dans le mixage le parcours attentionnel du personnage, par modulations progressives des différentes matières sonores. Les rapports en niveau des sons sélectionnés et des sons inhibés sont travaillés pour esquisser le cheminement attentionnel. Les sons cibles gagnent en saillance tandis que les éléments supposément rejetés de l'attention du personnage deviennent partiellement masqués par le fond sonore du milieu urbain. L'écoute du spectateur est alors guidée vers le parcours attentionnel de notre personnage.

Alors, les résultats du test psychoacoustique nous confirme tout d'abord que le panel a majoritairement emprunté ce parcours. Cette différence de mixage a ensuite eu une influence significative dans la qualification de l'état d'esprit du personnage, et de son évolution au sein de son environnement. Ainsi, la focalisation attentionnelle (supposée du personnage mais effective pour le panel) sur les voix au début de la séquence a été soulignée et interprétée par le panel. Le personnage leur a semblé davantage réagir face à ces voix, de par son regard vers la source fictive puis par son départ. Cependant, la focalisation attentionnelle supposée du personnage vers la passante qu'il croise en fin de séquence n'apparaît pas majoritairement dans le parcours attentionnel emprunté par le panel (cet élément est cité dans 30% des réponses, au dessus des 5 à 10% des sons inhibés mais bien en dessous des 70% des voix du début). Cet élément de mise en scène semble

être resté insignifiant pour le panel, qui n'a pas souligné une éventuelle interaction entre le passage de la passante et notre personnage. Si j'avais prévu un échange de regard entre les deux personnages à l'écriture du plan, cela n'apparaît pas clairement à l'écran. Mais cela nous enseigne une chose : la focalisation attentionnelle des spectateurs vers un élément scénique puis son interprétation dépend aussi de la réaction du personnage face à cet élément (orientation de la tête, déplacement dans l'espace, etc...). En effet, l'attitude attentionnelle du personnage fait partie d'une attitude plus globale et cohérente avec l'attitude corporelle. Alors, notre approche attentionnelle dans la pratique du mixage doit être incluse dans un processus plus large de mise en scène, motivé par une intention définie, pour gagner en efficacité.

Notre deuxième intention de mixage était de traduire une attitude attentionnelle plus globale, axée sur les processus ascendants. On supposait que le personnage était davantage vigilant sur l'ensemble du champ sonore, sans se focaliser particulièrement sur un élément plus qu'un autre. Le mixage démasquait les événements sonores successifs du fond urbain. Alors, le panel a davantage ressenti le personnage comme conscient, attentif à son environnement, en comparant avec les résultats du mixage de référence. En revanche, ce mixage n'a pas significativement orienté le ressenti du panel vers une qualification précise de l'état d'esprit du personnage. Il n'y a pas non plus de potentielles interactions, ou réactions du personnage face à son environnement qui ont été massivement relevées. Cela peut s'expliquer par le manque de contexte et de connaissance de notre personnage, en amont de cette séquence. Les différentes catégorisations du panel restaient large. Cependant, dans un contexte narratif plus riche et détaillé, cette stratégie attentionnelle appliquée au mixage peut être davantage significative (je pense notamment au film *Bad Boy Bubby* (1993), réalisé par De Heer, où l'évolution de la bande sonore évoque la lente adaptation de Bubby à un univers qui lui était inconnu).

Finalement, l'attention auditive façonne le monde sonore, et le monde sonore façonne notre attention auditive. Ce contact organisme/environnement apporte une signification sur le contexte et sur l'individu. Alors, dans le contexte cinématographique, l'approche des pratiques instrumentales en connaissance des mécanismes de l'attention

auditive semble pertinente à plus d'un titre, dans la mise en scène sonore d'un plan-séquence.

Pour une esthétique naturaliste, le montage son et le mixage permettent de recréer un univers sonore respectant un supposé réalisme. Alors, ils construisent un champ sonore sensoriellement convaincant, apportant des informations sur le contexte extérieur. Ils peuvent ainsi façonner l'environnement pour articuler les éléments diégétiques importants pour la narration de façon à stimuler l'attention auditive du spectateur. Celui-ci se positionne alors dans une posture d'attention sélective face au champ sonore, et stimule chez lui une sensation de réalisme.

D'autre part, l'inscription dans la bande sonore des modulations attentionnelles supposées du personnage permettent d'amener au spectateur des clefs de compréhension d'un état d'un personnage dans un environnement connu. Les modulations attentionnelles et les données contextuelles permettent au spectateur d'interpréter ces états — et ses possibles évolutions —, par l'identification de mécanismes expérimentés de tous. Ce type de mise en scène, plus proche du réalisme sonore psychologique, rend possible une volonté, chez le metteur en scène, de compréhension de l'état d'esprit d'un personnage par le spectateur. Cette signification inscrite dans les modulations du mixage, n'est pas nécessairement ostentatoire, mais semble pouvoir se ressentir chez le spectateur par des mouvements progressifs des différentes matières sonores. Il est aussi louable d'expérimenter des variations plus radicales, pour souligner l'aspect extraordinaire d'un personnage (sous le choc, pris d'hallucination, etc...). Le spectateur semble alors moins solliciter son attention sélective, puisque que le mixage oriente déjà le champ sonore dans une simulation d'écoute « pré-focalisée ». On guide ainsi le spectateur dans un parcours attentionnel. Si nous nous sommes concentrés sur la caractérisation du personnage de fiction, ce parcours attentionnel suggéré aux spectateurs peut se détacher d'une attitude attentionnelle liée à un personnage, et offrir des interférences signifiantes entre les champs sensoriels et l'évolution du récit.

ANNEXES

A. L'ANALYSE DE LA SCÈNE AUDITIVE

Contrairement à la vue qui donne un champ perceptif frontal, l'ouïe opère tout azimut. Cette aptitude se révèle dès l'origine nécessaire à la survie de l'Homme dans un environnement hostile. Les dangers sont perçus dans toutes les directions et sa réaction peut être immédiate (fuite, réorientation de la tête pour voir le danger). Aujourd'hui, la nécessité de survie est moins prégnante (bien qu'il reste utile d'entendre le klaxon de la voiture fonçant derrière nous) mais l'émergence de multiples sources sonores permet néanmoins de se situer continuellement au sein d'un espace, par appréciation des distances, des différentes localisations et des acoustiques mais aussi de déterminer les comportements ou les messages émis par lesdites sources afin de réagir en fonction. Tout ceci implique une organisation perceptive et attentionnelle de l'environnement sonore constitué de diverses sources, mécanisme décrit par Bregman comme « **l'analyse des scènes auditives** » (Bregman, 1999).

Les différents processus de localisation azimutale et d'évaluation de la distance aboutissent à l'**externalisation** des objets sonores (Bregman, 1999). D'un signal acoustique capté, nous projetons le son entendu sur la source localisée, dans l'espace. Sans ces mécanismes, nous entendrions les différentes sources collées au tympan.

Aussi, le flux sonore, qui est capté initialement sur les membranes cochléaires, se compose généralement d'une multitude de sources. L'expérience nous montre que nous sommes capable de les discriminer dans ce continuum afin de former des entités propres. C'est la **scission**. Cependant, un objet sonore peut être identifié par l'assemblage de sous-entités réunies spatialement (de façon simultanée) et temporellement (reconstruction d'un flux sonore cohérent provenant d'une même source sonore). C'est la **fusion**. Cet état de fait met en évidence des processus de groupement simultané ou séquentiel qui tendent à intégrer différentes composantes du monde sonore sous une même entité, ou bien de les disjointre pour percevoir plusieurs sources sonores. Examinons alors les différentes caractéristiques acoustiques qui favorisent la fusion ou la scission (Bregman, 1999).

- **Les processus de groupement simultané**

L'harmonicité commune : une relation harmonique entre différentes composantes favorise la fusion. Les relations inharmoniques ou la juxtaposition de plusieurs séries harmoniques donnent une perception ambiguë de la hauteur tonale ou la perception de plusieurs hauteurs tonales, ce qui favorise la scission. Un écart de 3 à 8% des fréquences fondamentales est nécessaire pour marquer la scission.

La synchronicité des attaques et des chutes : si différentes entités sonores voient leur attaque et leur chute sonore synchronisées, la fusion sera favorisée. Un décalage d'attaque de 30ms fait baisser la perception de fusion. De 40 à 80ms, une impression de dédoublement des sources apparaît, actant la scission.

La cohérence de la modulation d'amplitude et de fréquence : une évolution temporelle de l'amplitude et de la fréquence similaire entre deux entités contribue à la sensation de fusion.

La corrélation binaurale : si différentes composantes sonores ont la même relation binaurale, c'est-à-dire la même localisation, la fusion est favorisée. De plus, une différence de relation binaurale entre une source et un son masquant favorise l'émergence de cette source.

Aussi, plusieurs composantes sonores peuvent cohabiter temporellement pour former un continuum sonore. Selon différentes caractéristiques acoustiques, ce groupe d'entités sonores déployé dans le temps peut fusionner pour former un seul flux provenant d'une même source ou bien être perçu comme plusieurs flux sonores émanant de différentes sources.

- **Les processus de groupement séquentiel**

La proximité fréquentielle : les composantes sonores successives présentant une hauteur tonale similaire ont tendance à être perçues comme appartenant au même flux sonore, ils fusionnent. Inversement, une différence trop importante de tonalité participe à la scission en plusieurs flux, ou à l'impression d'apparition d'un nouveau flux. Plus la cadence des enchaînements sonores est lente, plus la différence en fréquence doit être importante pour scinder un flux.

La similarité de timbre : une succession de composantes sonores de même timbre a tendance à former un flux sonore unique. Une rupture dans la continuité spectrale donne l'impression qu'une nouvelle source est apparue.

La continuité d'intensité : la succession de sons dont l'intensité ne subit pas de rupture peut être perçue comme un seul flux. Van Noorden a montré que pour une succession d'événements sonores (à même fréquence) tous les 400ms, une différence de 5dB entre deux sons suffisait pour amener la scission du flux (Van Noorden, 1977).

Le taux de répétition : la rapidité de succession des sons favorise la scission du flux. Le ralentissement du rythme favorise la fusion.

De plus, lorsque deux flux sonores ne fusionnent pas, on remarque qu'il est difficile de donner un ordre précis des événements apparus pour les deux flux simultanés mais qu'il est aisé de donner l'ordre d'apparition des événements pour chacun des flux (McAdams, 1997). Aussi, il nous est possible de porter une attention auditive sur un flux en particulier, alors qu'il paraît impossible de rehausser un élément constitutif d'un flux sonore en faisant abstraction des autres éléments du même flux (une note de piano est une entité propre au timbre complexe, mais on ne peut sélectionner une harmonique en particulier en rejetant le reste du spectre).

Cependant, la notion de flux sonore est à mettre en perspective avec notre expérience attentionnelle. Prenons l'exemple d'une voiture roulant sur des pavés. Au

premier abord, le son émis par cette voiture en mouvement s' imagine comme un flux sonore cohérent au vu des différentes caractéristiques acoustiques des processus de groupement. Mais si nous portons une attention particulière sur ce déplacement, nous pouvons clairement distinguer le son du moteur du son des roues battant le pavé. Il y a donc deux flux sonores simultanés. Et c'était sans compter sur les petits graviers qui viennent taper la carrosserie, détail que nous remarquons sans doute après un certain temps d'attention accrue sur la masse sonore globale. Dans la réalité complexe du champ auditif, certains flux sonores apparents sont vraisemblablement constitués de plusieurs flux correspondant aux critères préalablement exposés. Cependant, il n'est pas question de les percevoir séparément mais bien comme un tout, une entité sonore complexe et unique. Car c'est la singularité des objets sonores qui s'y joue, par accumulation d'indices sonores primaires indissociables. Claude Bailblé (2004) nous explique que « *le monde réel n'est pas directement audible en tant que tel. Il se reflète dans la conscience en qualia, sorte de sensations élémentaires dont le regroupement combiné aboutit à la figure auditive, à la fois image par ses qualia combinés et objet par sa reproduction dans l'espace-temps* ». Il n'y a que le monteur son qui peut dissocier et associer ces différentes qualités perceptives, par accumulation de couches sonores, afin de rendre à l'accessoire de cinéma toute sa prétendue qualité sensorielle, singulière et vraisemblable aux oreilles du spectateur.

B. L'EFFET DE MASQUE

Dans un environnement donné, le champ attentionnel peut être constitué de plusieurs sources sonores cohabitant ensemble. Les sources se trouvent de part et d'autre de l'espace, plus ou moins éloignées les unes des autres, rayonnant de façons diverses dans plusieurs directions. À nos oreilles s'effectue le mixage passif - contraint par la perspective - de toutes ces vibrations. L'individu qui s'y trouve capte alors ce signal acoustique issu des multiples sources sonores comme une entité (la vibration sur le tympan) qu'il s'agira de décomposer. C'est la bonne compréhension de la scène auditive et de ses éléments constitutifs. Cependant, nous avons tous expérimenté que dans un environnement bruyant, les sons faibles n'arrivent plus à émerger aisément, ou sont même complètement inaudibles. C'est le phénomène de masquage d'un son recouvrant un autre son, ou l'**effet de masque**.

L'effet de masque peut être total ou partiel. On le quantifie en mesurant le nombre de décibels qu'il faut ajouter au son masqué pour que celui-ci retrouve sa sonie d'origine. L'effet de masque partiel se caractérise par une diminution de la sonie du son masqué, sans que celui-ci ne soit pour autant inaudible. L'étude (Wegel & Lane, 1924) de l'effet de masque sur les sons purs nous enseigne que :

- l'effet de masque est maximal pour les fréquences voisines de celles du son masquant ;
- l'effet de masque est plus efficace sur les fréquences immédiatement supérieures au son masquant que sur les fréquences immédiatement inférieures ;
- l'effet de masque est négligeable tant que le niveau de masque est faible ;
- l'effet de masque croît beaucoup plus vite que le niveau du son masquant.

Cependant, la grande majorité des sons peuplant notre champ attentionnel ne sont pas simplement des sinusoïdes. Ils présentent des contenus spectraux complexes pouvant occuper l'ensemble du spectre, comme les chutes d'eau, assimilable à du bruit blanc. Une expérience (Fletcher, 1940) étudie alors le masquage du bruit blanc sur un son pur. Il s'aperçoit que seule une bande de fréquences située autour de la fréquence masquée participe à l'effet de masque. C'est la bande critique. Sa largeur de bande est d'environ 90Hz pour une bande critique centrée dans le bas du spectre (jusqu'à 500Hz) puis est de 10

à 15% de la fréquence centrale de la bande critique pour les fréquences aigües (Bailblé, 1999). L'effet de masque peut aussi intervenir pour deux sons non-simultanés. Le masquage temporel s'applique jusqu'à 200ms après propagation du son masquant sur le son masqué³⁸ (Favreau, 2018).

En pratique, l'enveloppe dynamique d'une source joue un rôle crucial dans l'émergence d'un son dans un environnement bruyant. L'oreille réussira à percevoir approximativement les sons à forts transitoires. Aussi, selon le contenu spectral du bruit ambiant, une source peut être partiellement masquée fréquentiellement mais demeure perceptible sur une autre bande de fréquence³⁹. Ainsi, nous ne percevons les objets du monde sonore que partiellement. Le spectre des sources sonores est en partie masqué dans le bruit pour nous faire parvenir la partie du spectre émergente, le timbre apparent (les premières réflexions en champ clos participent aussi au timbre apparent). De plus, le champ diffus qui nous donne de précieuses informations de perspective et d'acoustique peut lui aussi être masqué (ou devenir masquant selon les rapports de niveaux). La perception de la perspective sonore peut alors se voir trompée, distordue (Chatauret, 2009). Cependant, une écoute binaurale, discriminant les différentes positions des objets sonores, favorise la perception des sons recouverts s'il y a une différence de relation binaurale entre le masquant et le masqué (De Cheveigné, 2000), c'est-à-dire une différence de position. Plus généralement, l'écoute binaurale nous permet de comprendre la scène auditive.

³⁸ De 0 à 3ms, l'effet de masque reste identique ; de 3 à 200ms, il diminue progressivement.

³⁹ On arrive à comprendre un flux de parole dans une soirée bruyante même si l'on ne perçoit pas les basses fréquences de la voix, on essaie d'ailleurs de parler plus fort mais aussi plus aigüe pour se faire comprendre dans ce cas.

C. VALEURS DE L'EFFET DE DÉMASQUAGE BINAURAL

1.	$M\pi S\pi, Mo So, Mu Sm$	0dB
2.	$Mu S\pi$	3dB
3.	$Mu So$	4dB
4.	$M\pi Sm$	6dB
5.	$Mo Sm$	9dB
6.	$M\pi So$	13dB
7.	$Mo S\pi$	15dB

Tableau 1. Valeurs numériques du démasquage binaural selon les conditions de présentation du signal S et du bruit M sur les deux oreilles. (Green, 1976)

o : S et M en phase sur les deux écouteurs

π : S ou M en opposition de phase sur les écouteurs

u : bruits non corrélés sur les deux écouteurs

m : S ou M sur un seul écouteur

D. LES DISPOSITIFS DE DIFFUSION MULTICANALE

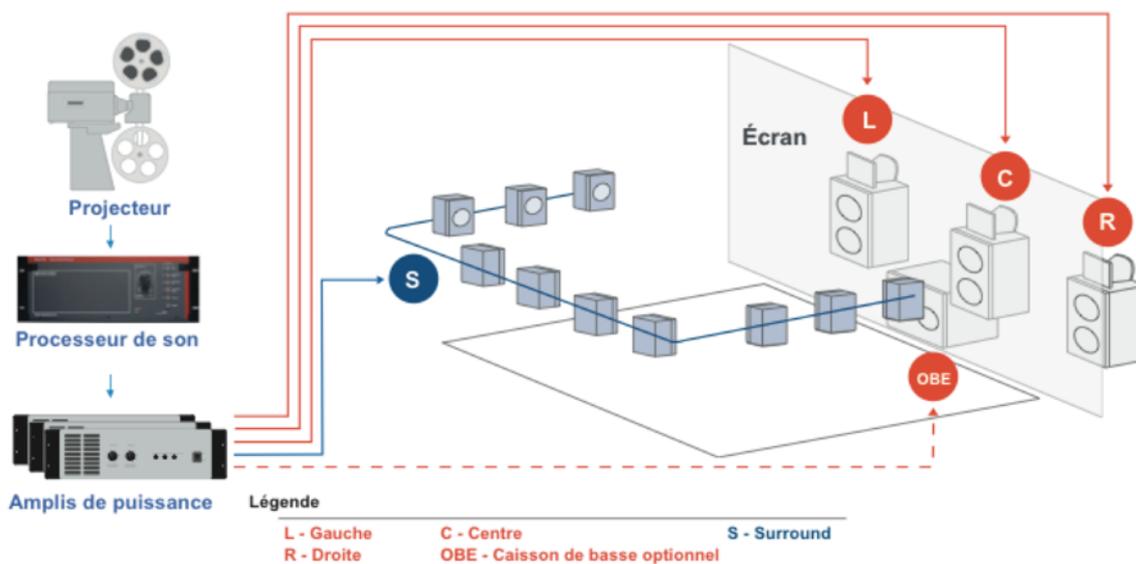


Figure 20 - Configuration pour une salle au format Dolby Stéréo (LCRS matricé en LtRt)

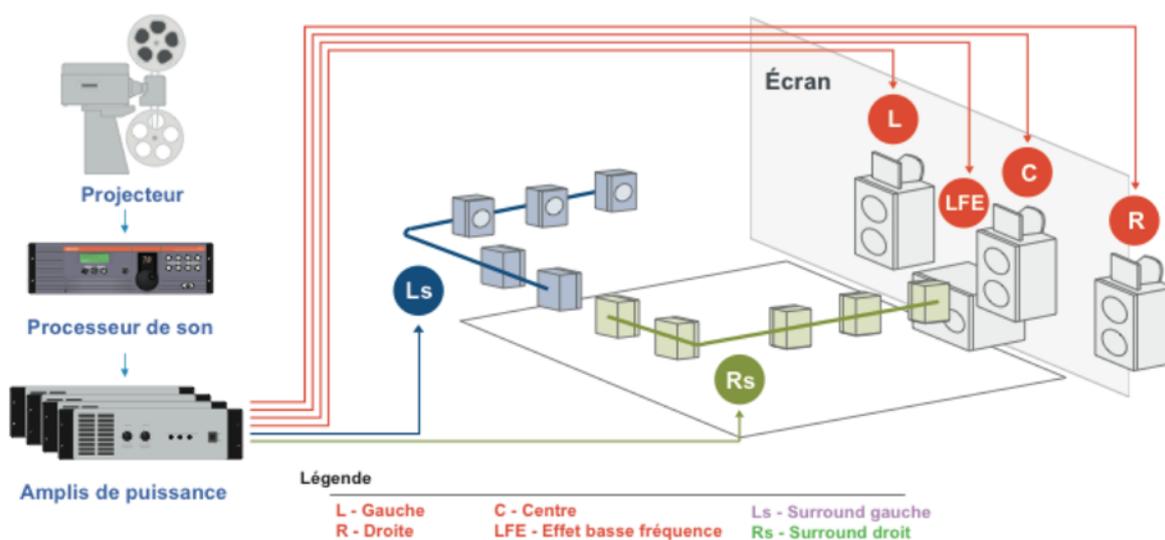


Figure 21 - Configuration pour une salle au format Dolby Surround 5.1

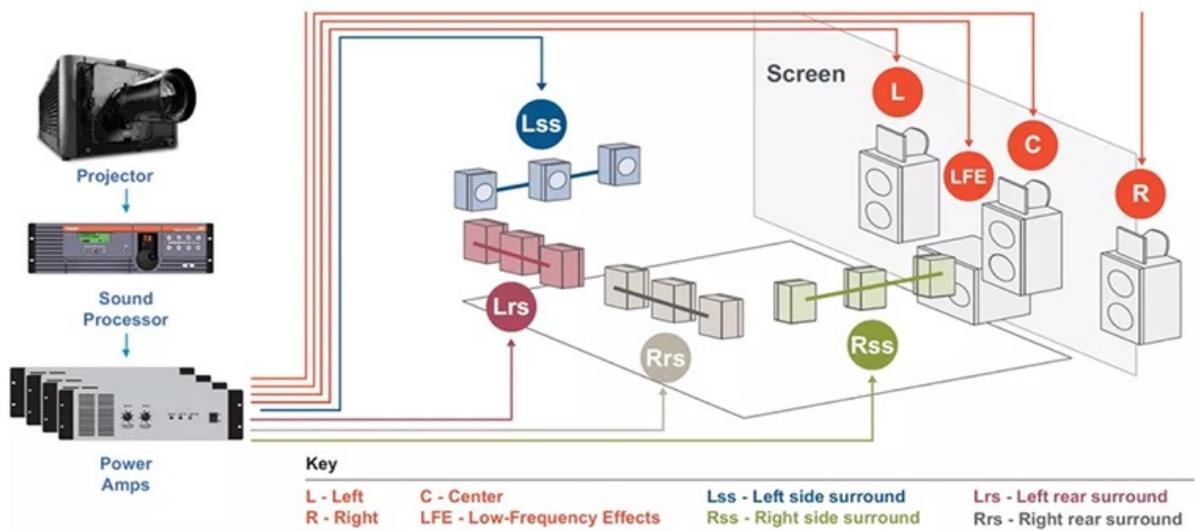


Figure 22 - Configuration pour une salle au format Dolby Surround 7.1

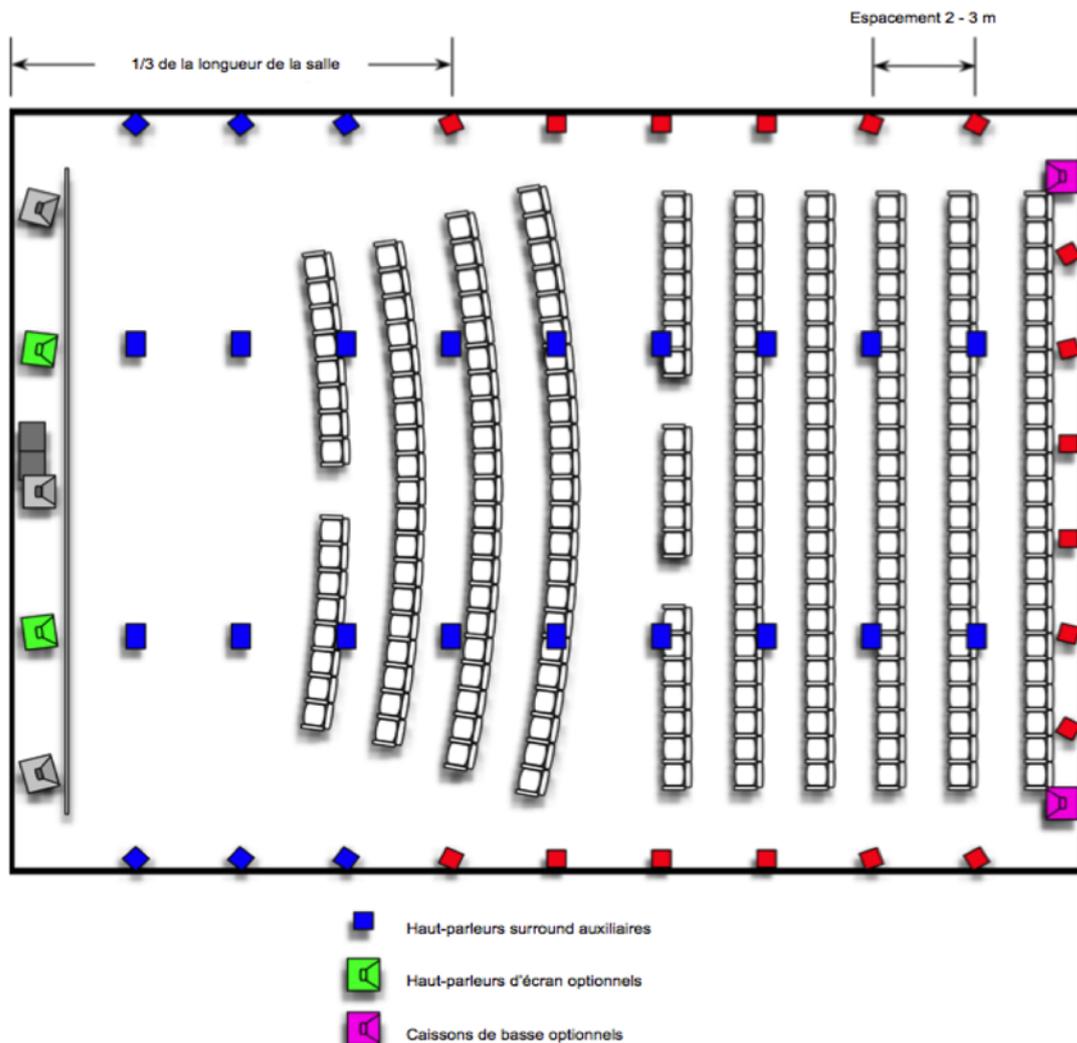


Figure 23 - Implantation d'un système Atmos dans une salle de cinéma. En rouge on voit les haut-parleurs surround déjà existant, les haut-parleurs surround auxiliaires sont représentés en bleu, tandis que les subwoofer additionnels sont en rose.

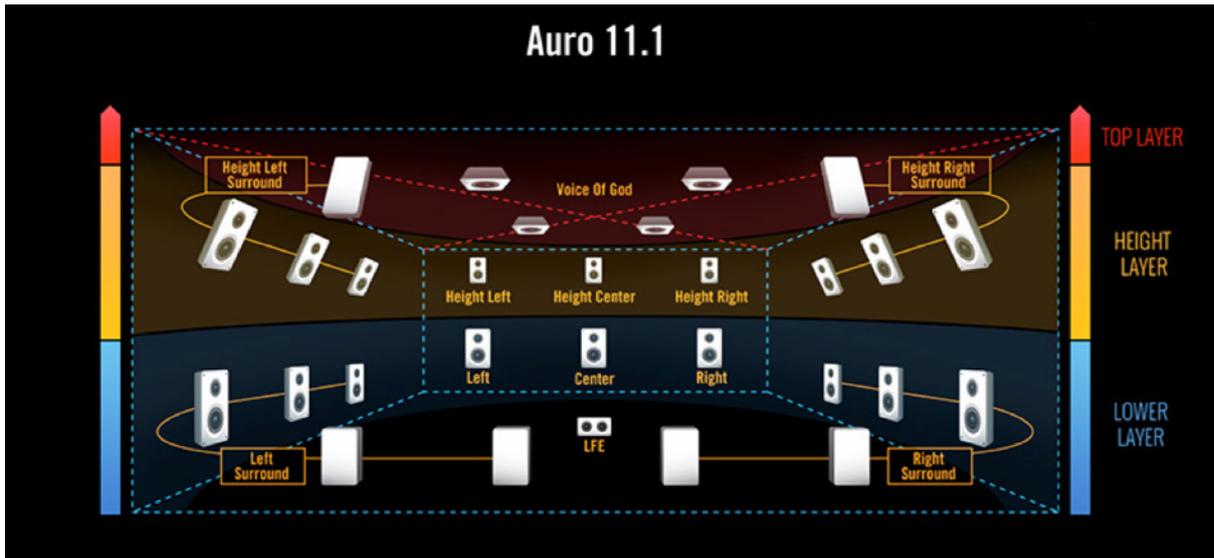


Figure 24 - Configuration de diffusion Auro 3D 11.1

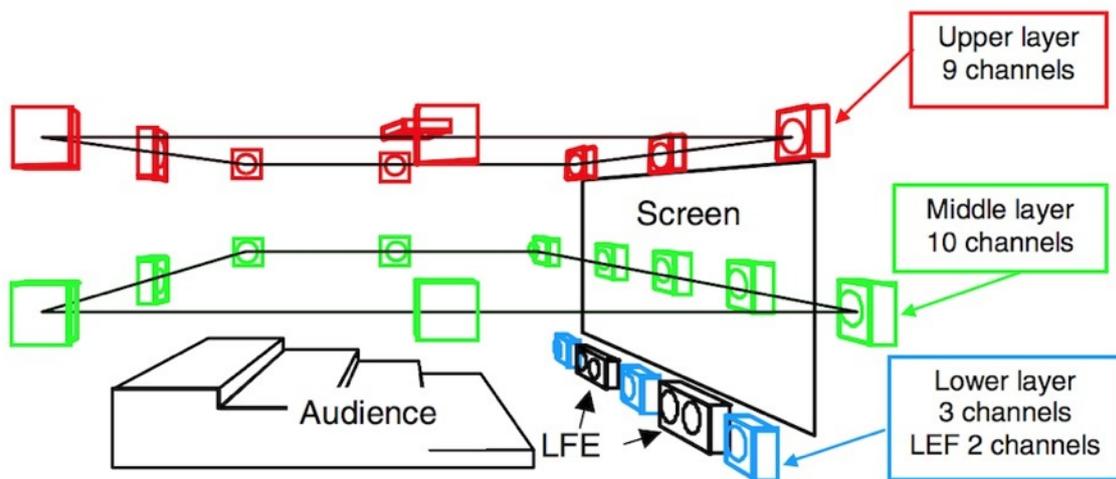


Figure 25 - Configuration de diffusion du 22.2 NHK

E. LES RÉSULTATS DU TEST PSYCHOACOUSTIQUE

• MIXAGE « OBJECTIF »

	Stressé	Méfiant	Attentif	Indifférent	Serein	Pensif
	pas serein, angoissé	inquiet, sur ses gardes	pressé, attentif, non joyeux	exaspéré, lassé, pas pressé, connaît l'endroit	familier, routine	pensif, concentré, introspection
		craintif, suspect, attentif	confiant, observateur, déterminé	indifférent, déterminé	déterminé, habitué à cet environnement	neutre, pensif
		regarde autour de lui, intrigué	attentif aux sons, mais détaché	déterminé, pressé, impassible		détaché, pensif
		aux aguets, anxieux, stressé				concentré, va à un rdv important
		craintif, déterminé				détaché, occupé
		méfiant				préoccupé, pensif, inquiet
		méfiant, inquiet, attentif				pensif, perdu, inquiet
		attentif, tendu, angoissé				pensif, blasé
Total / 25	1 / 4%	8 / 32%	3 / 12%	3 / 12%	2 / 8%	8 / 32%

Tableau 1 - Résultats, pour le mixage « objectif », à la question : « Dans quel état d'esprit semble être le personnage ? »

● Stressé ● Méfiant ● Attentif ● Indifférent
● Serein ● Pensif

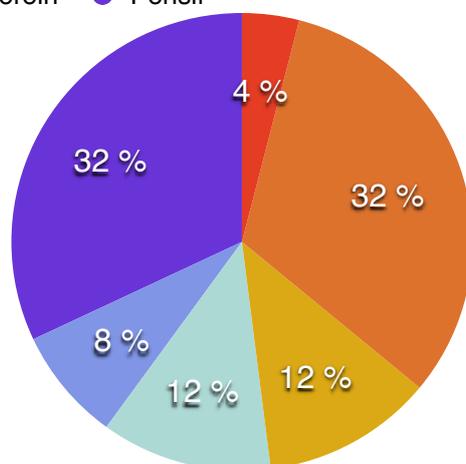


Figure 19.1 - Graphique sur l'état d'esprit du personnage perçu pour le mixage « objectif »

	Vigilant / Hostile	Attentif	Normalement / Familier	Détaché	Autres
	avec attention, méfiance	observateur, attentif	assez neutre	détaché, dans ses pensées	par la vue uniquement
	aux aguets, menacé	très attentif aux sons	son quotidien, familier	ne pas y prêter attention	pleinement
	regarde autour de lui, intrigué	observateur	environnement quotidien	ne le perçoit pas	il a l'air d'aller à un rdv important
	stressant	analytique, attentif	familier	légèrement en dehors	
	aux aguets	très attentif	Normalement, immersif	indifférent, s'en extrait	
				absent, dans ses pensées	
				n'y prête pas attention	
Total /25	5 / 20%	5 / 20%	5 / 20%	7 / 28%	3 / 12%

Tableau 2 - Résultats, pour le mixage « objectif », à la question : « Comment le personnage semble percevoir l'environnement dans lequel il s'inscrit ? »

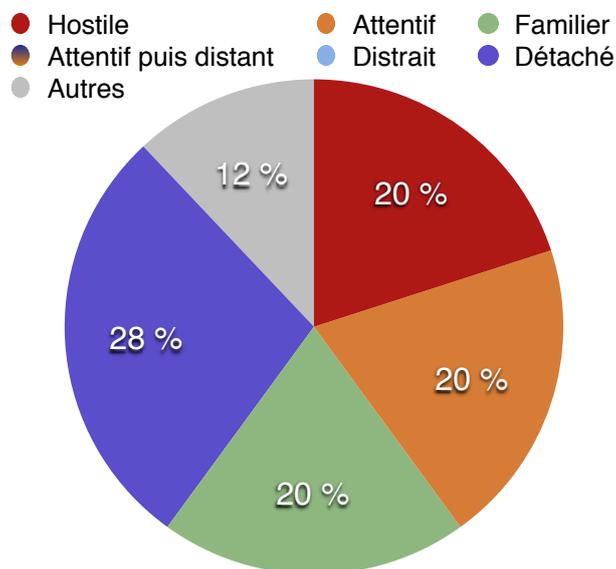


Figure 18.1 - Graphique sur la perception du champ organisme/environnement pour le mixage « objectif »

• **MIXAGE « ENDOGÈNE »**

	Stressé	Méfiant	Attentif	Indifférent	Serein	Pensif
	inquiet, agacé, sous pression, fatigué	particulièrement attentif, sur la défensive	attentif à son environnement	neutre, passif	serein, imperturbable, confiant, décidé	résolu pensif
	stressé, préoccupé	pas l'air serein	observateur, déterminé	sans crainte, pressé	savoir ou aller, détendu	pensif, stressé, mélancolique
	nerveux, déterminé	à l'affut, stressé, angoissé, en danger				essequé, préoccupé, nonchalant
	déterminé, tourmenté, stressé, sous pression	sous-tension, attentif au début PUIS calme a la fin				inquiet, pensif, indifférent
	pas à sa place, stressé, pressé	apeuré, inquiet, peu confiant				ailleurs, seul, à part
	anxieux, stressé					pensif nostalgique
Total / 31	6 / 19,3%	5 / 16,1%	2 / 6,5%	2 / 6,5%	2 / 6,5%	détaché, inquiet, perplexe

Tableau 3 - Résultats, pour le mixage « endogène », à la question : « Dans quel état d'esprit semble être le personnage ? »

● Stressé ● Méfiant ● Attentif ● Indifférent
● Serein ● Pensif

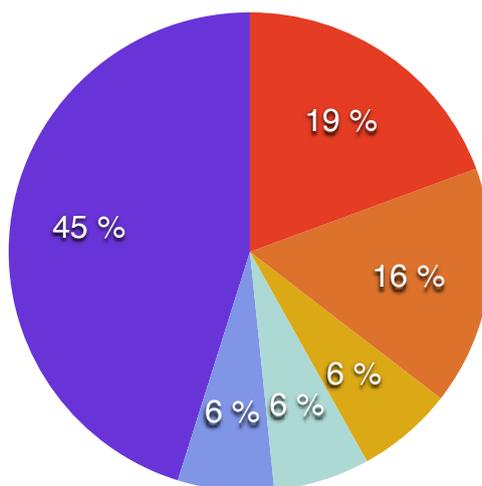


Figure 19.2 - Graphique sur l'état d'esprit du personnage perçu pour le mixage « endogène »

dans ses pensées, esprit occupé
distant, inquiet
inquiet, pensif, déboussolé, perdu
dans l'attente
normal, inquiet, attente
dans sa bulle, inquiété, apeuré, déterminé
blasé, agacé
14 / 45,1%

	Vigilant / Hostile	Attentif	Normalement / Familier	Attentif / vigilant puis distant	Détaché
	l'environnement semble hostile	particulièrement attentif	De façon naturelle	curieux mais garde ses distances	pas d'interactions avec l'environnement, inattentif, préoccupé
	Anxiogène	observateur	terre à terre	passif tout en étant attentif	en décalage avec son environnement
	à l'affut, en danger	attentif	sans crainte	il y prête attention tout en étant distant	à part
	Avec inquiétude	entend avec précision ce qui l'entoure	normalement	à la fois imprégné de l'environnement, à la fois détaché	dans son monde, détaché de l'environnement
	Environnement hostile	attentif aux éléments sonores	il connaît l'environnement	Avec crainte, il s'en détache	indifférent
			Normal	Attentif au début, indifférent ensuite	Des bruits lui échappent, ne traite pas les informations sensibles
				Attentif dès qu'il fume, puis passif en marchant	
				Attentif au début, puis dans sa bulle	
				Hostile au début puis dans sa bulle	
Total / 31	5 / 16,1%	5 / 16,1%	6 / 19,4%	9 / 29%	6 / 19,4%

Tableau 4 - Résultats, pour le mixage « endogène », à la question : « Comment le personnage semble percevoir l'environnement dans lequel il s'inscrit ? »

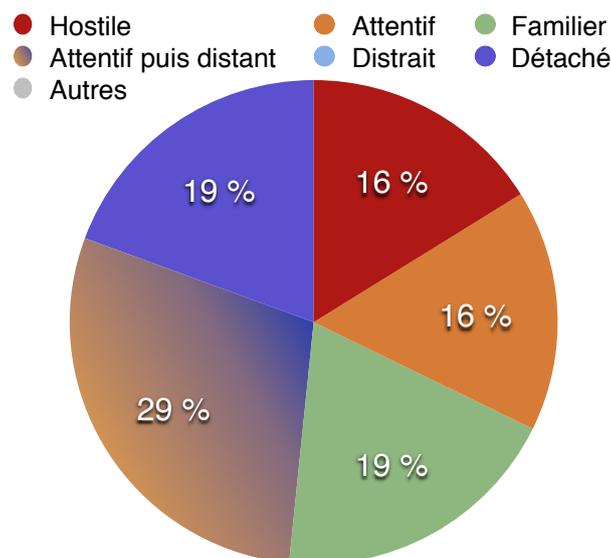


Figure 18.2 - Graphique sur la perception du champ organisme/environnement pour le mixage « endogène »

• **MIXAGE « EXOGÈNE »**

	Stressé	Méfiant	Attentif	Indifférent	Serein	Pensif
	Stressé, peur	stressé, traqué, anxieux, perturbé	déterminé, attentif à son environnement, concentré	fatigué, lassé, distant	Détendu, naturel	distrait, pensif, conscient
	un peu stressé, s'apprête à faire quelque chose d'important	méfiant, inquiet, sur la défensive	attentif, neutre, calme, reposé	neutre	calme, serein, désintéressé	Neutre, concentré, contrarié, rêveur, absent
	tendu, décidé	sur ses gardes, inquiet		neutre	serein mais déterminé	pensif, déterminé
	stressé, a vif	méfiant		neutre, indifférent	normal et détendu	blasé, envie d'ailleurs
	stressé, troublé, oppressé	sur ses gardes, stressé		indifférent, déterminé		dans ses pensées, déterminé, agacé, concentré
	inquiet, déterminé	méfiant, calme				
		inquiet, vigilant, préoccupé, pressé				
		précaution, discrétion				
Total /30	6 / 20%	8 / 26,6%	2 / 6,6%	5 / 16,6%	4 / 13,3%	5 / 16,6%

Tableau 5 - Résultats, pour le mixage « exogène », à la question : « Dans quel état d'esprit semble être le personnage ? »

● Stressé ● Méfiant ● Attentif ● Indifférent
● Serein ● Pensif

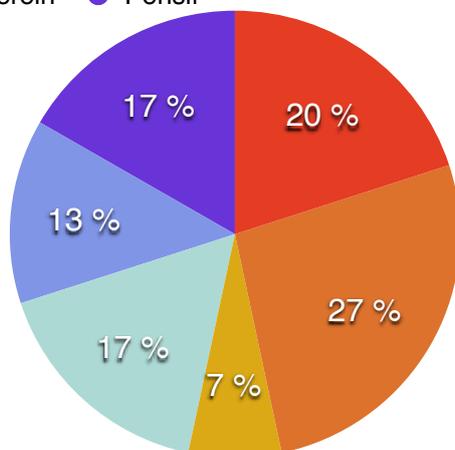


Figure 19.3 - Graphique sur l'état d'esprit du personnage perçu pour le mixage « exogène »

	Hostile	Attentif	Normalement / Familier	Attentif puis distant	Distrait	Détaché	Autres
	violent, méfiant	très à l'écoute	habituel	indifférent sauf au début	distrait, il passe d'un élément à un autre sans y prêter attention	peu attentif à l'environnement	impression d'être à sa place
	environnement hostile	attentif à tout ce qui l'entoure	familier	inquiet au début puis indifférent		détaché de son environnement	par la vitesse de ses pas
	angoissant, oppressant, agressif	très conscient de son environnement	habitué à cet environnement			n'y porte pas attention	calme et frontale
	un peu angoissant	il perçoit très fort son environnement direct	il connaît son environnement, il a déjà tout entendu			indifférent	
	légèrement hostile, inamicaux	conscient de tout ce qui se passe autour de lui	l'environnement urbain est son habitude, sa routine			ne s'y intéresse pas	
	avec insécurité	il perçoit chaque son un par un	de manière habituelle, neutre				
			parfaitement normalement				
Total / 30	6 / 20%	6 / 20%	7 / 23,3%	2 / 6,7%	1 / 3,3%	5 / 16,7%	3 / 10%

Tableau 6 - Résultats, pour le mixage « exogène », à la question : « Comment le personnage semble percevoir l'environnement dans lequel il s'inscrit ? »

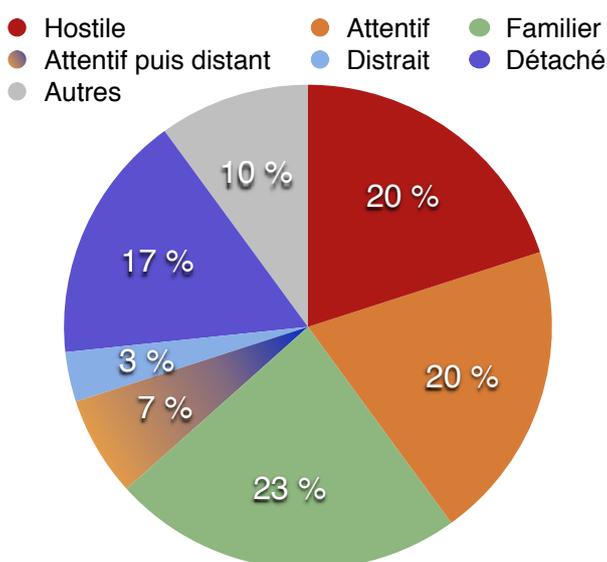


Figure 18.3 - Graphique sur la perception du champ organisme/environnement pour le mixage « exogène »

F. LIENS INTERNET DES STIMULI

Version « **objective** » : <https://youtu.be/8-OkyqdHQT0>

Version « **endogène** » : <https://youtu.be/3F1lqXoBpno>

Version « **exogène** » : <https://youtu.be/EW1N1aFXORs>

BIBLIOGRAPHIE

- AUMONT, Marie. *Dictionnaire théorique et critique du cinéma*. Paris, Nathan, 2001.
- BAILBLÉ, Claude. *La perception et l'attention modifiées par le dispositif cinéma*. Université Paris VIII, 1999
- BAILBLÉ, Claude. « Entendre, écouter, agir... ». Neurosciences, ouvrage collectif De Boeck Université. Juin 2004.
- BIDET-CAULET, Aurélie. *Mécanismes neurophysiologiques de la perception de flux sonores chez l'Homme : effets des contextes acoustiques et attentionnels*. Neurosciences [q-bio.NC]. Université Claude Bernard, Lyon I, 2007.
- BONNANI, Simon. *Le plan séquence, une esthétique de l'immersion*. Mémoire de fin d'études, dirigé par Guisy Pisano, ENS Louis-Lumières, 2016.
- BOTTE, M.-C., DRAKE, C., BROCHARD, R., McADAMS, S. et ARGENTI, A.-M. « Focalisation attentionnelle appliquée aux flux auditifs. Premières mesures ». Le Journal de Physique IV 04, n° C5 (mai 1994)
- BREGMAN, A. S. *Auditory Scene Analysis : The perceptual Organization of sound*. MIT Press paperback edition, 1994.
- BROADBENT, D. E. *Perception and communication*. Pergamon, 1958.
- BOTTE, Marie-Claire. *Psychoacoustique et perception auditive*. Paris: INSERM, 1989.
- CAMUS, Jean-François. *Psychologie cognitive de l'attention*. Armand Collin, 1996.
- CHATAURET, Jean. « La perspective sonore. Communications n°85, L'Homme a-t-il encore une perspective ? » Le seuil, 2009.
- CHION, Michel. *Le Guide des objets sonores*. Buchet/Chastel, 1983.
- CHION, Michel. *Un art sonore, le cinéma*. Cahiers du cinéma, 2003.
- CHION, Michel. *L'audiovision*. Armand Colin, 2005.
- DALLAIRE, Frédéric. *Point de vue et point d'écoute au cinéma. Chap : Le point d'écoute, du microphone à l'auditeur : déambulation dans l'espace sonore de Leslie Shatz et Gus Van Sant*. Presse universitaire de Rennes, Le spectaculaire, 2017.
- DE CHEVEIGNÉ, Alain. « Analyse de scène auditive et parole ». IRCAM, CNRS, 2000.
- DESHAYS, Daniel. *Pour une écriture du son*. Klincksieck, Paris, 2006.

- DESHAYS, Daniel. *Entendre le cinéma*. Klincksieck, Paris, 2010.
- FAVREAU, Sophia. *Psychoacoustique : étude comparative des seuils différentiels d'intensité et de hauteur en fonction du niveau de perte auditive*. Mémoire, université de Lorraine, 2018.
- FLETCHER, H. Auditory patterns. Rev. Mod. Phys. 12, 47-56, 1940.
- HATWELL, Y. *Traité de psychologie expérimentale*. Paris : Presse Universitaire de France, 1994.
- HILLYARD, S. A. HINK, R. F. SCHWENT, V. L. PICTON, T. W. Electrical signs of selective attention in the human brain. Science, vol 182, 1973.
- HUSSERL, Edmund. *Méditations cartésiennes, "Deuxième Méditation"*. Trad. G. Pfeiffer et E. Levinas, Vrin, 1947.
- JAMES, William. The Principles of Psychology. American science series, Vol. 1, Chap. 11, « Attention », 1890.
- JULLIER, Laurent. *Les sons au cinéma et à la télévision*. Armand Collin, 1995.
- JULLIER, Laurent. *Le son au cinéma*. Les Cahiers du cinéma, 2006.
- LANDRAGIN, Frédéric. *De la saillance visuelle à la saillance linguistique*. Presses Universitaires de Franche-Comté, 2011.
- LANDRAGIN, Frédéric. « Saillance physique et saillance cognitive ». Corela, n° 2-2,15 décembre 2004.
- LEBLANC, Gérard. *Le double scénario chez Fritz Lang*. Armand Collin, cinéma et audiovisuel, 1991.
- LEWIN, Kurt. *Field Theory in Social Science*. Tavistock Pub. LDT, London, 1951.
- McADAMS, Stephen. BIGAND, Emmanuel. *Introduction à la cognition auditive*. In : *Penser les sons : Psychologie cognitive de l'audition*. PUF, Paris, 1-9, 1994.
- McADAMS, Stephen. « L'organisation perceptive de l'environnement sonore ». Rencontres IPSEN en ORL, 1997.
- MERLEAU-PONTY, Maurice. *Signes*. Gallimard, Paris, 1960.
- MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, vol. 63, no 2, 1956.

- MORAY, Neville. Attention in dichotic listening: affective cues and the influence of instructions. Quarterly journal of experimental psychology, février 1959.
- MOREAU, Patricia. *Étude de la négativité de discordance dans l'amusie congénitale*. Thèse de doctorat, université de Montréal, 2013.
- NÄÄTÄNEN, R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav Brain Sci, 13, 201-288, 1990.
- ONDAATJE, Michael. *The conversations : Walter Murch and the Art of film editing*. New-York, Random House, 2004.
- PERLS, Frederick. S. *Manuel de Gestalt-thérapie: la Gestalt : un nouveau regard sur l'homme*. Paris, ESF, 2003.
- POSNER, M. I., & RAICHLE, M. E. *L'esprit en images*. Paris: De Boeck. 1998.
- POSSAMAÏ, C.-A., BONNEL, A.-M., REQUIN J. « L'attention ». Encyclopédia Universalis, Paris, 1993.
- ROUSSELET, G. A. THORPE, S. J. FABRE-THORPE, M. Processing of one, two or four natural scenes in humans: the limits of parallelism. Vision research, vol 44, issue 9, pp. 877-894, 2004.
- SCHAEFFER, Pierre. *Traité des objets musicaux*. Seuil, 1966.
- SHATZ, Leslie. « It all sounds good, if he's done his job ». Trad. Dallaire F., Susan King, Los Angeles Times, Juillet 2005.
- SHEPARD, Roger. *Psychophysical complementarity*. In : *Perceptual organization*. Hillsdale, NJ, 1981.
- TREISMAN, A. M. Contextual cues in selective listening. Quarterly journal of experimental psychology, vol 12, 4, 1960.
- VAN NOORDEN, L.P.A.S. « Minimum differences of level and frequency for perceptual fission of tone sequences ». ABAB. J. Acous. Soc. Am. 1977.
- WEGEL, R. L., LANE, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev. 23, 266, Février 1924.
- WEIL-BARAIS, Annick. *L'homme cognitif*. Quadriges Manuels, 2011.
- WESTERKAMP, Hildegard. « Soundwalk. Sound Heritage, vol. 3, n°4, 1974.
- WINKLER, I. Interpreting the mismatch negativity (MMN). J Psychophysiol, 21, 147- 163, 2007.

Sources internet

MEILLEUR, Catherine. *L'attention décryptée*. KnowledgeOne, 2020. knowledgeone.ca/lattention-decryptee

SAULEAU, P. Cours de neurophysiologie, chapitre Récepteurs somatosensoriels. google.com/site/aphysionado, 2017.

FILMOGRAPHIE

ALTMAN, Robert. Gosford Park. Royaume-Uni, 2001.

DE HEER, Rolf. Bad boy Bubby. Australie, 1993.

HITCHCOCK, Alfred, *Rear window*. États-Unis, 1954.

IÑÁRRITU, Alejandro González. The revenant. États-Unis, 2015.

SAUVAIRE, Jean-Stéphane. *A prayer before dawn*. France, Angleterre, 2017.

SCHIPPER, Sebastian. Victoria. Allemagne, 2015.

VAN SANT, Gus. Gerry. États-Unis, 2003.