

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Étude psychophysique d'une illusion visuelle induite par le son

Par
Catherine Éthier-Majcher

Département de psychologie
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de maîtrise (M.Sc.)
en psychologie

Août 2008

© Catherine Éthier-Majcher, 2008



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Étude psychophysique d'une illusion visuelle induite par le son

présenté par :

Catherine Éhier-Majcher

a été évalué(e) par un jury composé des personnes suivantes :

Isabelle Peretz
Président-rapporteur

Frédéric Gosselin
Directeur de recherche

Franco Lepore
Membre du jury

RÉSUMÉ

« Les situations dans lesquelles un individu est exposé à des signaux provenant d'une seule modalité constituent l'exception et non la règle. » (traduit de Shams, Ma & Beierholm, 2005). Récemment, Shams, Kamitani et Shimojo (2000) ont rapporté une illusion visuelle induite par le son où, lorsque plusieurs stimuli auditifs (des sons de courte durée) sont présentés simultanément à un stimulus visuel (un disque blanc), des observateurs perçoivent systématiquement plusieurs stimuli visuels. Le présent travail consiste d'abord en une réplique de l'illusion auprès de huit participants naïfs. Ensuite, dans le but d'éclaircir les paramètres temporels qui sous-tendent cet effet illusoire, une deuxième expérience fut conduite où, cette fois-ci, les participants devaient juger du nombre de stimuli visuels présentés simultanément à des stimuli auditifs dans un contexte où le délai entre la présentation d'un stimulus visuel et d'un stimulus auditif était systématiquement varié. Les résultats montrent qu'un son présenté au début et à la fin de la stimulation sonore influence l'occurrence de notre version de l'illusion, mais que le délai entre la présentation d'un son et la présentation d'un disque n'est pas un facteur influençant la perception illusoire d'un disque.

Mots-clés : Vision, Audition, Intégration multisensorielle, Illusion, Décours temporel.

ABSTRACT

“Situations in which an individual is exposed to sensory signals in only one modality are the exception rather than the rule.” (Shams, Ma & Beierholm, 2005) Recently, Shams, Kamitani and Shimojo (2000) have reported a sound-induced visual illusion in which the presentation of two brief sounds (beeps) with one white disk (a flash) induces the perception of an illusory second disk. The present work consists of two main experiments: a replication of the illusory effect with eight naïve subjects, followed by a second experiment conducted in order to investigate the temporal constraints of the illusion. In this second experiment, participants had to judge the number of flashes perceived when simultaneously presented with a number of beeps, where the delay between a flash and a beep was varied systematically. The results show that the presence of a beep at the beginning and at the end of the auditory stimulation influences the occurrence of our version of the illusion. However, the delay between a beep and a flash does not seem to be a factor driving the perception of an illusory flash.

Key words: Vision, Audition, Multisensory integration, Illusion, Temporal delays.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT.....	iv
LISTE DES FIGURES.....	vii
LISTE DES ABRÉVIATIONS.....	ix
REMERCIEMENTS.....	x
CHAPITRE PREMIER : INTRODUCTION GÉNÉRALE	1
Les théories d'intégration de l'information multisensorielle	4
L'illusion visuelle induite par le son	8
<i>La découverte de l'IVIS</i>	8
<i>Études psychophysiques</i>	9
<i>Études en électrophysiologie et en imagerie cérébrale</i>	17
Objectifs du travail	19
<i>Études préliminaires</i>	20
<i>Décours temporel de l'intégration de l'information multisensorielle</i>	25
CONTRIBUTIONS À L'ARTICLE.....	30
DÉCLARATION DES COAUTEURS DE L'ARTICLE.....	31
CHAPITRE DEUXIÈME : ARTICLE	32
Is timing everything for the sound-induced flash illusion?.....	33
Abstract	34
Introduction	35

Methods	37
<i>Experiment 1</i>	37
<i>Participants</i>	37
<i>Stimuli</i>	38
<i>Procedure</i>	38
<i>Results and discussion</i>	39
<i>Experiment 2</i>	40
<i>Participants</i>	40
<i>Stimuli</i>	40
<i>Procedure</i>	41
<i>Results and discussion</i>	41
General discussion.....	44
Références	46
Figure captions	48
CHAPITRE TROISIÈME : DISCUSSION GÉNÉRALE.....	52
Explication selon un modèle bayésien	53
Modifications possibles.....	55
Pistes d'explication de l'IVIS	57
Conclusion.....	62
RÉFÉRENCES	64

LISTE DES FIGURES

CHAPITRE PREMIER

Figure 1

a) Schéma représentant la tâche initiale de la première expérience de Berger et al. Après la présentation des deux intervalles, les participants devaient ajuster la fréquence de présentation du stimulus visuel à l'aide de la barre de défilement. La partie b) représente un schéma de la tâche que nous avons élaborée dans le but de répliquer l'expérience présentée en a). Suite à la présentation de l'intervalle, les participants devaient déterminer combien de stimuli visuels étaient présentés et appuyer sur la touche correspondant à leur choix. 24

Figure 2

Réplication des résultats de la première expérience de Berger et al. 25

Figure 3

Cette figure démontre bien l'influence de la supposition que la lumière provient toujours du haut sur notre perception. Lorsque la figure est observée à l'endroit, les cercles de la rangée de haut semblent être des bosses alors que les cercles de la rangée du bas semblent être des creux. Lorsqu'on tourne la page, les cercles qui semblaient être des bosses semblent maintenant être des creux et vice-versa. 27

CHAPITRE DEUXIÈME

Figure 4

Results of our replication of the sound-induced flash illusion as presented by Shams et al. (2000). When fewer beeps than flashes are presented, observers tend to perceive fewer flashes than presented, whereas when more beeps than flashes are presented, observers tend to perceive more flashes than presented. 49

Figure 5

For all subjects, the percentages of perceived flashes as a function of the number of flashes presented. For 4 flashes presented and more, the accuracy of the participants decreases. For all visual conditions, there is a small percentage of trials where participants perceive over 2 flashes more or less than the number presented. 50

Figure 6

The classification vectors averaged across subjects for each audio-visual condition. The X axis represents the time course of the auditory stimulation. The Y axis represents the presence (positive) or the absence (negative) of a beep in terms of Z-scores. The blue lines represent the threshold for signification. The red stars (superimposed with the blue lines) represent the position of the flashes as well as the threshold for statistically significant results. 51

CHAPITRE TROISIÈME

Figure 7

Pour toutes les conditions visuelles confondues, la probabilité qu'un disque soit présenté en fonction du temps total de la stimulation (une seconde). Les positions temporelles où il est le plus probable de percevoir un disque correspondent au tout début et à la fin d'un essai. 54

Figure 8

(a) et (b) Alternance d'un chiffre et de son opposé à une vitesse lente et à une vitesse rapide. (c) Lors de la chute, les chiffres sont présentés sur un bracelet auquel est fixé un cadran numérique. (d) Les participants sont relâchés au haut de la « Suspended Catch Air Device » et font une chute libre de 31 m. Tiré de Stetson et al. (2007). 60

LISTE DES ABRÉVIATIONS

cd/m ²	candela par mètre carré
cm	centimètre
Hz	hertz
kHz	kilohertz
m	mètres
ms	millisecondes
s	seconde

REMERCIEMENTS

Ces deux années de maîtrise furent de belles années de découvertes, autant sur le plan académique que sur le plan personnel. Je prends ici quelques lignes pour remercier tous ceux qui ont su me soutenir et m'accompagner dans ce cheminement.

Tout d'abord, j'aimerais remercier mon directeur de recherche, Frédéric Gosselin, avec qui j'ai beaucoup de plaisir à travailler. Merci pour ta confiance en mes capacités, pour ta franchise et pour ton écoute. Tu me pousses constamment à dépasser mes limites et à développer ma créativité et, conséquemment, j'apprends beaucoup avec toi.

Merci à Caroline, ma partenaire de travail, mon amie. Merci pour ta patience, ton écoute, et ta capacité à aider les autres. Je ne m'en serais pas sortie sans ton support et ta motivation dans toutes ces heures de discussion, de programmation et de remise en question. Merci à Daniel qui a su croire en moi dès le début, qui m'a fourni de sages conseils et qui m'a offert de belles opportunités. Merci à Cynthia pour son écoute, ses encouragements constants, sa confiance en moi et son amitié. Merci à Nicolas, mon ami et professeur de Matlab, qui m'a transmis son savoir avec beaucoup de patience, d'empathie et d'encouragement. Merci à tous les autres membres du laboratoire, qui font de cet environnement de travail un endroit agréable, enjoué et divertissant. C'est un plaisir de vous côtoyer quotidiennement.

Merci à mes amies, Julie, Sarah, Maude, Valérie, Vicky et les autres, qui sont toujours prêtes à s'amuser et qui me permettent de « lâcher mon fou » ou de me confier en cas de besoin. Vous êtes des filles extraordinaires!

Finalement, un merci tout spécial à mes parents et à mon frère, sans qui je me serais découragée bien avant le début de ce projet. Vous avez su être patients, vous m'avez supportée dans les bons comme dans les mauvais moments et par-dessus tout, vous avez toujours cru en moi. Merci de me suivre dans mes projets les plus fous et de me ramener à la réalité quand je dépasse les bornes. Vous êtes mon équilibre! Je vous aime très fort.

Ce projet fut réalisé avec le support d'une bourse d'excellence octroyée par le département de psychologie de l'Université de Montréal.

CHAPITRE PREMIER : INTRODUCTION GÉNÉRALE

Imaginez un instant que vous êtes assis seul dans l'une des grandes salles d'opéra de ce monde. À votre droite, une femme portant des lunettes de soleil trouve le chemin de son siège à l'aide d'une canne blanche. À votre gauche, un homme muni d'un appareil auditif s'approche, confiant, et s'assoit en vous criant « bon spectacle! ». Le rideau se lève, le spectacle s'apprête à commencer. En voyant le somptueux décor qui se présente à vos yeux et en entendant les premières notes de votre opéra favori, vous ne pouvez vous empêcher d'avoir une pensée pour vos deux voisins. Apprécieront-ils le spectacle de la même façon que vous?

Les situations dans lesquelles un individu doit traiter simultanément de l'information visuelle et auditive ne sont pas rares : elles constituent la règle et non l'exception (Shams, Ma, & Beierholm, 2005). Dans ces situations, la privation d'une modalité sensorielle entraîne, certes, une perte d'information sur l'environnement qui nous entoure, mais est-il possible que l'intégration de l'information provenant de deux modalités distinctes entraîne une mauvaise combinaison des signaux sensoriels, menant ainsi à un percept global incorrect? Il semble que plusieurs manifestations perceptives illusoire soient le résultat de l'interaction entre l'information provenant de deux modalités sensorielles. Par exemple, l'effet ventriloque (Howard & Templeton, 1966) constitue un phénomène perceptif fréquemment observé dans lequel la capacité à localiser spatialement un son est influencée par une stimulation visuelle (e.g. lorsqu'on assiste à un concert et que la voix du chanteur semble provenir directement de la bouche du chanteur et non des haut-parleurs qui projettent le son dans la salle). Cet effet illusoire a été typiquement interprété comme une illustration de l'influence dominante du système visuel sur le système auditif (Caclin,

Soto-Faraco, Kingstone & Spence, 2002 ; Mateef, Hohnsbein & Noack, 1985). Cependant, il semble que le cerveau n'utilise pas nécessairement la vision comme modalité dominante dans toutes les situations où il doit combiner de l'information multisensorielle. Dans l'effet McGurk (McGurk & MacDonald, 1976), phénomène dans lequel la vision altère la perception de la parole — de sorte que le son « ba », associé à l'image du mouvement des lèvres produisant le son « ga », tend à être perçu comme étant le son « da » — le percept global est complètement différent des stimuli auditif et visuel présentés initialement, suggérant une intégration partielle de chaque stimulation présentée. Des études plus récentes démontrent que dans certains cas, même le toucher (Bresciani, Dammeier, & Ernst, 2006; Ernst & Banks, 2002; Violentyev, Shimojo, & Shams, 2005) et l'audition (Maeda, Kanai, & Shimojo, 2004; Sekuler, Sekuler, & Lau, 1997) peuvent influencer la vision, ce qui suggère que le cerveau n'intègre pas toujours l'information multisensorielle de la même façon. En fait, Alais et Burr (2004) ont montré que, dans des conditions précises, l'effet ventriloque — cet exemple de domination de la modalité visuelle — peut lui aussi résulter d'une intégration partielle de la stimulation visuelle et de la stimulation auditive qui l'ont produit, sans être le reflet unique de l'une de ces stimulations.

Dernièrement, Shams, Kamitani, & Shimojo (2000) ont rapporté une Illusion Visuelle Induite par le Son (dorénavant nous utiliserons l'acronyme IVIS) où, lorsque plusieurs stimuli auditifs (des « beeps », c'est-à-dire des sons de courte durée) sont présentés simultanément à un stimulus visuel (un « flash », c'est-à-dire un disque blanc brièvement présenté), des observateurs humains perçoivent systématiquement plusieurs stimuli visuels. Cette illusion a suscité beaucoup

d'intérêt dans le domaine de la perception multimodale puisqu'elle constitue la première démonstration de la « capture » de la vision par l'audition (Shams, Kamitani & Shimojo, 2004). En effet, alors que d'autres études impliquant de l'information audio-visuelle ont montré une modification de la perception d'un stimulus visuel au niveau de ses caractéristiques temporelles — comme la durée (Walker & Scott, 1981), la fréquence (Gebhard & Mowbray, 1959 ; Shipley, 1964 ; Welch, Duttonhurt & Warren, 1986) et la synchronie (Fendrich & Corballis, 2001 ; Shimojo et al., 2001) — l'IVIS est le premier phénomène à montrer un changement qualitatif important du percept visuel par le son (Shams, Kamitani & Shimojo, 2002).

Les théories d'intégration de l'information multisensorielle

De nombreuses hypothèses ont été émises concernant la façon dont l'information provenant de deux modalités distinctes est intégrée par le cerveau. L'une des premières théories proposées fut la capture visuelle (Rock & Victor, 1964), où la modalité visuelle, vue par certains comme le principal sens, domine presque complètement l'autre modalité. Cette théorie a trouvé son parallèle auditif tout récemment dans l'IVIS, où l'audition semble capturer la vision. Cependant, l'hypothèse de la capture d'une modalité par une autre modalité manque de pouvoir prédictif puisqu'elle ne permet pas de dire à l'avance quelle modalité sensorielle sera dominante. De plus, cette hypothèse ne permet pas d'expliquer l'effet McGurk, où le percept global constitue un amalgame des modalités visuelle et auditive sans domination complète de la vision ou de l'audition. En 2004, Andersen, Tiippana et

Sams¹ ont tenté d'expliquer l'IVIS en faisant appel à quatre hypothèses liées à l'intégration de l'information multisensorielle, soit l'hypothèse de discontinuité, l'hypothèse de la modalité la plus pertinente à la tâche, l'hypothèse de l'information la plus fiable pour la tâche et l'hypothèse de l'attention dirigée. Selon les auteurs, ces hypothèses permettraient de déterminer l'influence relative de chaque modalité sur le percept multisensoriel. L'hypothèse de discontinuité de l'information (Shams et al., 2002) propose que la modalité dans laquelle le stimulus est discontinu domine; l'hypothèse de la modalité la plus pertinente à la tâche (Welch & Warren, 1980) propose que la modalité qui domine est celle qui est la plus adéquate pour effectuer une tâche donnée (e.g. l'audition est plus adéquate dans une tâche de discrimination temporelle alors que la vision est plus adéquate dans une tâche de discrimination spatiale); l'hypothèse de l'information la plus fiable propose que la modalité qui offre l'information la plus fiable domine (Schwartz, Robert-Ribes & Escudier, 1998); et l'hypothèse de l'attention dirigée propose que la façon dont les directives d'une tâche sont données est susceptible d'affecter l'attention qui est dirigée vers chacune des modalités (Welch & Warren, 1980). Une version alternative de l'hypothèse de la modalité la plus pertinente à la tâche fut aussi proposée par Guttman, Gilroy et Blake (2005), où les auteurs proposent que l'encodage des inputs sensoriels se ferait dans la modalité appropriée à l'information (e.g. l'aspect temporel de la vision serait traité suivant un code auditif). Ces hypothèses ont été peu explicitées dans la littérature, et cela pourrait s'expliquer par le fait qu'elles reflètent une réalité confinée aux expériences qui les sous-tendent et que, bien que l'ensemble des hypothèses semble converger vers des idées très semblables, peu d'efforts ont été faits pour les

¹ À ne pas confondre avec « Shams ».

généraliser à d'autres situations. Récemment, plusieurs auteurs ont proposé une idée parcimonieuse qui englobe toutes ces hypothèses (Alais & Burr, 2004 ; Andersen, Tiippana & Sams, 2005 ; Bresciani et al., 2006 ; Burr & Alais, 2006 ; Ernst & Banks, 2002 ; Ernst & Bühlhoff, 2004 ; Shams, et al., 2005). Selon eux, l'intégration multisensorielle se ferait selon un modèle de combinaison optimale de l'information provenant de chacune des modalités. Selon Ernst et Bühlhoff, l'être humain intègre l'information en utilisant deux stratégies : la combinaison sensorielle (qui consiste à maximiser l'information qui provient des différentes modalités sensorielles) et l'intégration sensorielle (qui réduit la variance dans l'estimé sensoriel pour augmenter sa fiabilité). Dans une situation où l'information provient de deux modalités sensorielles distinctes, le cerveau amassera le plus d'information possible à propos de son environnement afin de résoudre l'ambiguïté qui se présente. Le principal but de l'intégration sensorielle étant de diminuer la variance de chacun des estimés sensoriels pour augmenter leur fiabilité, l'ambiguïté ne sera pas résolue par l'utilisation de l'information provenant de la modalité la plus appropriée à la tâche, mais par l'utilisation de l'information qui présente le plus de précision (Ernst & Banks, 2002). Selon Ernst et Bühlhoff, la dominance de la vision pour certaines tâches s'explique par le fait qu'un poids relatif, inversement proportionnel à la variance de l'estimé sensoriel donné par une modalité, est alloué au traitement de l'information en provenance de chacune des deux modalités sensorielles.

Une version mathématique de cette hypothèse a été formulée par Ernst et Banks (2002), dans le contexte d'une expérience sur l'intégration d'information haptique et visuelle. Ce modèle fut appliqué, peu après, par Alais et Burr (2004),

dans le contexte de l'effet ventriloque et fut utilisé pour expliquer d'autres situations de perception multisensorielle (Ernst & Bühlhoff, 2004) pour devenir une théorie de l'intégration de l'information multisensorielle. Mathématiquement, cette théorie se traduit par un modèle de *Maximum Likelihood Estimation* (MLE) dans lequel il est proposé que la méthode d'estimation optimale (i.e. qui produit le moins de variance dans l'estimé) provient de l'addition des estimés sensoriels pondérés selon leur variance réciproque normalisée (Ernst & Banks, 2002). Donc, l'estimé sensoriel intégré \hat{S} est la somme pondérée de chaque estimé individuel (Équation 1) pour lequel cette pondération est proportionnelle à l'inverse de la variance de l'estimé (Équation 2).

$$\hat{S} = \sum_i w_i \hat{S}_i \quad (1)$$

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \quad (2)$$

Un modèle semblable comportant une famille de facteurs de plus — les a priori ou « priors », en anglais — fut proposé par Shams et al. en 2005. Ce modèle, qui sera mieux décrit dans la section *Études psychophysiques*, permet de présenter l'IVIS comme le résultat d'une stratégie optimale du système nerveux.

Dans le but de mieux comprendre les processus d'intégration multisensorielle, le présent travail s'intéressera à la façon dont l'information audiovisuelle est intégrée dans le contexte de l'IVIS. Plus spécifiquement, ce mémoire aura pour objectif de clarifier les paramètres (e.g. la fréquence de présentation des sons, la fréquence de présentation des disques, la position temporelle des sons par

rapport à la position temporelle des disques, etc.) qui mèneraient à une influence maximale du son sur la perception visuelle et donc à l'amplification du phénomène illusoire. Cette clarification des paramètres de l'IVIS est essentielle puisqu'elle permettra de circonscrire le phénomène illusoire à une définition précise et à des mesures psychophysiques fiables caractérisant l'étendue du phénomène. D'abord l'illusion, telle qu'observée à l'origine sera rapportée, suite à quoi des études plus récentes procurant de nouvelles informations à propos du phénomène d'intérêt seront exposées. Enfin, les principaux objectifs du présent travail seront approfondis et l'expérimentation sous-jacente sera présentée en détails.

L'illusion visuelle induite par le son

La découverte de l'IVIS

L'IVIS fut présentée pour la première fois dans *Nature* en 2000 par Shams et al. Dans cet article, les auteurs rapportent que, dans une tâche où l'on présente simultanément un seul stimulus en périphérie du champ visuel (disque blanc présenté à 5 degrés d'excentricité) et plusieurs stimuli auditifs (2, 3 ou 4 courts sons présentés à 57 ms d'intervalle) des observateurs ($N = 8$) perçoivent systématiquement de multiples stimuli visuels (effet de fission), mais l'effet inverse — soit, par exemple, un disque perçu alors que deux disques sont accompagnés d'un son (effet de fusion) — n'est pas observé. L'illusion semble présente autant chez les observateurs naïfs que chez les observateurs qui savent pertinemment qu'un seul disque est présenté, mais un délai maximal de 100 ms entre la présentation d'un disque et la présentation d'un son serait nécessaire pour percevoir un disque illusoire. Les auteurs proposent que l'IVIS résulterait d'une altération de la perception visuelle par les stimuli

auditifs. Depuis ce premier article, de nombreuses études portant sur l'IVIS ont fourni d'autres détails sur le phénomène illusoire ou des explications alternatives de ce phénomène. La section qui suit sera consacrée aux apports de ces études, autant au niveau psychophysique qu'au niveau de l'électrophysiologie et de l'imagerie cérébrale.

Études psychophysiques

En 2002, Shams et al. ont tenté de montrer que, dans l'IVIS, l'altération du percept visuel n'est pas due à la difficulté de la tâche ou à des artefacts cognitifs ou attentionnels, mais bien à une interaction perceptuelle multimodale. Pour éliminer l'aspect « difficulté de la tâche », ils ont montré que le nombre de disques perçus en absence de sons correspondait au nombre de disques réellement présentés (pour 1, 2, 3 et 4 disques présentés), ce qui indique que l'effet illusoire n'est pas dû uniquement à la difficulté des participants à compter le nombre de disques présentés. De plus, afin de vérifier que les participants avaient bien compris la tâche et s'étaient bien concentrés sur le nombre de disques présentés et non le nombre de sons, les résultats aux essais où un son était présenté simultanément à un, deux, trois ou quatre disques ont été observés. Le raisonnement des auteurs était que si les participants avaient répondu en fonction du nombre de sons, on aurait pu s'attendre à ce que les observateurs aient toujours répondu qu'un seul disque était présenté dans ces conditions, ce qui ne fut pas observé. Cependant, cet argument est plutôt faible puisqu'il met entièrement à l'écart l'influence que peut avoir le son sur la vision. Dans l'IVIS, si le son permet de « voir plus », pourquoi ne permettrait-il pas de « voir moins »? Des études subséquentes ont montré que « voir moins » (phénomène

de fusion visuelle) était effectivement possible dans une situation expérimentale où moins de signaux sonores que de disques sont présentés (Andersen et al., 2004 ; Berger, Martelli & Pelli, 2003 ; Shams, Ma & Beierholm, 2005). Sachant cela, les tentatives de Shams et al. (2002) à démontrer que l'IVIS n'est pas due à des artefacts cognitifs ou attentionnels semblent peu convaincantes et elles n'éliminent pas toutes les possibilités. Shams, Kamitani et Shimojo (2002) se sont aussi demandé si la perception d'un disque illusoire était comparable à la perception d'un disque réel. Pour répondre à cette question, les auteurs ont comparé le nombre de disques rapportés pour la condition où un disque et deux sons sont présentés et pour la condition où deux disques et un son sont présentés. Comme le nombre de disques rapportés est le même pour les deux conditions, les auteurs concluent que, d'un point de vue perceptuel, le disque illusoire est équivalent au disque réel, ce qui fut aussi démontré à l'aide de potentiels évoqués visuels (Shams, Kamitani, Thompson & Shimojo, 2001). Cependant, la présence d'un disque illusoire ne se limite pas à la condition où un disque est accompagné de deux sons. Berger et al. ont montré qu'avec quatre stimuli visuels et plusieurs stimuli auditifs (un plafond apparaît à 10 sons), il est possible de percevoir jusqu'à six stimuli visuels. Ceci suggère que la relation entre le nombre de disques perçus et les conditions de présentation audiovisuelle (nombre de disques et nombre de sons présentés) n'est pas linéaire et, par conséquent, il est fort probable que le disque illusoire n'ait pas les mêmes caractéristiques d'une condition à l'autre.

Dans une deuxième expérience, Shams, Kamitani et Shimojo (2002) ont étudié la façon dont les délais entre la stimulation visuelle et la stimulation auditive

font varier l'occurrence de l'illusion. Leurs résultats révèlent que l'effet illusoire commence à se dégrader lorsque le son précède ou devance le disque d'environ 70 ms, mais l'effet est encore présent (environ 23 à 33% des essais) jusqu'à 115 ms de délai. Ces résultats sont intéressants, mais, dû à des limites expérimentales, ils révèlent peu d'information sur les contraintes temporelles de l'IVIS. Par exemple, les délais utilisés entre les sons et les disques sont pré-déterminés, ce qui entraîne des limites dans les délais critiques observables. De plus, seulement la condition où un disque et deux sons sont présentés fut étudiée.

Les résultats des deux expériences de Shams, Kamitani et Shimojo (2002) sont expliqués à l'aide de l'hypothèse de discontinuité, qui prédit que le stimulus discontinu (ici le son) d'une modalité altère la perception du stimulus continu de l'autre modalité (le disque), mais sans que l'effet inverse soit possible. Toutefois, cette hypothèse procure peu d'information sur l'intégration de l'information multisensorielle et elle est peu généralisable à d'autres phénomènes multisensoriels, sans compter que tel que mentionné plus haut, « l'effet inverse » (effet de fusion) est possible dans le cas de l'IVIS. Aussi, les auteurs rejettent l'hypothèse de la modalité la plus pertinente qui prédirait correctement que l'audition est la modalité dominante dans la tâche (l'audition est connue pour être plus appropriée que la vision pour des tâches de discrimination temporelle), parce qu'elle ne permettrait pas d'expliquer pourquoi on observe un phénomène de fission, mais pas de phénomène de fusion. Cependant, cette explication fut invalidée, tel que mentionné plus haut, par des études subséquentes qui ont montré que lorsque plusieurs disques sont présentés simultanément à un seul son, par exemple, les observateurs ont tendance à percevoir

moins de disque que ce qui leur a été présenté. Il est toutefois fort étonnant que cet effet ne soit pas apparu clairement dans l'expérimentation de Shams et de ses collaborateurs, qui le mentionnent comme étant un effet présent, mais tout à fait négligeable.

En 2003, Berger et al. ont choisi de reprendre le concept de base de l'illusion de Shams et al. (2000) et d'en modifier certains paramètres (le disque est remplacé par une tache de Gabor — une grille sinusoïdale présentée à travers une lucarne gaussienne — de même grosseur, la fréquence de présentation des sons est modifiée, etc.) dans le but de trouver une trace de l'effet dans une tâche psychophysique de bas niveau (sur laquelle nous reviendrons lorsque leur seconde expérience sera décrite). Dans une première expérience, malgré la modification de plusieurs paramètres, les auteurs ont observé l'IVIS, soit un plus grand nombre de stimuli visuels perçus lorsque le nombre de sons était supérieur au nombre de stimuli visuels présentés. De plus, ils ont observé un effet de fusion des stimulations visuelles, présent lorsque plus de stimuli visuels que de stimuli auditifs sont présentés. Enfin, en explorant les paramètres sonores de l'illusion, Berger et al. (2003) ont découvert que l'effet semble obéir à une loi plus ou moins logarithmique qui plafonne à une fréquence de présentation des sons d'environ 10 Hz. Dans une seconde expérience, les auteurs se sont demandé si le fait de percevoir plus de stimuli visuels pouvait influencer la performance à une tâche psychophysique de bas niveau. Leur raisonnement était basé sur des études de Verghese et Stone (1995, 1996, 1997) qui ont montré que la performance à une tâche était déterminée par le nombre d'événements perçus (le nombre d'événements perçus étant toujours un sous-ensemble du nombre total

d'événements qui peuvent être présentés). Une stimulation visuelle pouvant être interprétée comme un événement, le fait de percevoir illusoirement un plus grand nombre de stimulations visuelles pourrait influencer la performance à une tâche donnée. Pour tester cette hypothèse, Berger et al. (2003) ont présenté des sons associés à des taches de Gabor qui pouvaient, de façon équiprobable, être orientées à la verticale ou être légèrement inclinées (1 degré) par rapport à la verticale. Leur paradigme était sensiblement le même que celui ayant servi à reproduire l'IVIS, mais deux séries de stimulations audio-visuelles étaient présentées et la tâche demandée aux participants était de discriminer l'orientation de la grille sinusoïdale. Les résultats montrent que la présence de l'illusion « fissionnelle » améliore la performance des participants à la tâche². De plus, lorsque le nombre de stimulations visuelles est supérieur au nombre de stimulations auditives, l'effet de fusion est présent et la performance à la tâche est détériorée.

En 2004, Andersen et al. ont eux aussi montré qu'un effet de fusion était présent dans l'IVIS. De plus, leur étude leur a permis de montrer que l'effet illusoire pouvait être inversé lorsqu'on utilise des stimuli auditifs de qualité dégradée accompagnés des mêmes stimuli visuels utilisés par Shams et al. (2002). En effet, dans une tâche où l'on doit compter le nombre de sons présentés alors qu'ils sont accompagnés de disques, il est possible que le nombre de disques présentés altère la réponse du participant, mais seulement lorsque le volume des sons est très faible (10dB). Les auteurs ont tenté d'expliquer l'IVIS en s'appuyant sur diverses hypothèses concernant l'intégration de l'information provenant de deux modalités

² La fiabilité de ces résultats sera discutée un peu plus loin dans la section *Objectifs du travail*.

distinctes. Leurs résultats permettent de rejeter l'idée qu'une condition nécessaire à l'apparition de l'illusion est que la modalité auditive est plus appropriée à la tâche, puisqu'il semble que la vision puisse influencer l'audition dans une tâche temporelle. Par contre, ces résultats supportent l'idée que l'intégration multisensorielle se baserait sur la fiabilité de l'information disponible. De plus, les auteurs ont démontré que dans les mêmes conditions que celles de Shams et al. (2000), non seulement il est possible de répliquer l'effet de fission de l'illusion, mais il est aussi possible de retrouver un effet de fusion considérable. Sur la base de ces nouvelles informations, les auteurs ont postulé que, dans une situation où de l'information bimodale est présentée, le degré de discontinuité d'un stimulus est un facteur qui augmente les chances qu'une modalité domine, mais que cette discontinuité n'est pas nécessaire à la présence de l'illusion. Les auteurs concluent que les sons ne sont pas automatiquement intégrés aux disques pour former un percept global uniforme et que plusieurs hypothèses d'intégration multisensorielle devraient être considérées conjointement comme des facteurs contribuant à la dominance relative de chaque modalité.

En 2005, Shams et al. ont réalisé une première expérience semblable à la tâche de Shams et al. (2000) qui avait pour but d'évaluer, d'une part, la sensibilité visuelle et auditive (unimodale) des observateurs humains — les « likelihoods » dans la terminologie bayésienne — et, d'autre part, la tendance des observateurs à percevoir plus ou moins de disques et de sons en présence de plus ou moins de sons et de disques — les « priors » dans la terminologie bayésienne. Cette première expérience a permis de vérifier la capacité des participants à faire une tâche

unimodale, soit compter le nombre de stimuli présentés auditivement ou visuellement. Par la suite, les résultats furent utilisés pour construire un modèle semblable à celui présenté par Alais et Burr (2004), mais dans ce cas-ci, non seulement les auteurs tiennent compte des « likelihoods », mais ils déterminent aussi des « priors ». Pour mieux comprendre les différences des modèles d'Alais et Burr (2004) et de Shams et al. (2005), expliquons plus en détails la terminologie bayésienne dans le contexte d'intégration d'information multisensorielle. Les « likelihoods » représentent la fiabilité de l'information provenant de chaque modalité individuellement, alors que les « priors » représentent la probabilité d'intégrer l'information provenant des modalités auditive et visuelle. Par exemple, quand deux disques sont présentés, un observateur aura une certaine probabilité de percevoir un disque, une certaine probabilité de percevoir deux disques, etc., peu importe le nombre de sons présentés. Si les deux modalités permettent d'exécuter la tâche, comme c'est habituellement le cas avec la position, dans l'espace, d'un objet qui émet un son (i.e., l'effet ventriloque), l'effet combiné des « likelihoods » est suffisant pour expliquer l'illusion ; c'est pourquoi le modèle d'Alais et Burr (2004) tient compte uniquement des « likelihoods ». Toutefois, ce n'est pas le cas de l'IVIS puisque les sujets doivent porter attention seulement à la vision et l'audition n'est pas informative. C'est ici que les « priors » entrent en jeu. Ces derniers permettent de tenir compte d'une tendance qu'ont les sujets à combiner les deux modalités inconsciemment. Le théorème de Bayes permet de combiner « likelihoods » et « priors » en un « posterior » — un percept ici — très simplement. L'équation 3 illustre ce modèle, avec la « likelihood » du signal visuel représentée par $P(V|Z_V)$ et la « likelihood » du signal auditif représentée par $P(A|Z_A)$ — la

probabilité de recevoir le signal A sachant Z_A — et les « priors » représentés par $P(Z_A, Z_V)$.

$$P(Z_A, Z_V | A, V) = \frac{P(A | Z_A)P(V | Z_V)P(Z_A, Z_V)}{P(A, V)} \quad (3)$$

Shams et al. (2005) ont donc intégré les données de leur première expérience dans ce modèle bayésien. Ce modèle sans paramètre libre a ensuite été testé dans une seconde expérience réalisée sur un autre ensemble de sujets. La variance expliquée par ce modèle est excellente, mais est-ce réellement étonnant? Non, parce que les « priors » font tout le travail. Ils dictent la force des effets multimodaux dans chacune des conditions expérimentales. La deuxième expérience montre simplement que les « priors » mesurés dans la première expérience sont robustes — si un sujet voit deux disques quand un seul disque est accompagné de deux sons, la probabilité est grande qu'il en sera de même pour un autre sujet. En effet, dans le contexte de l'IVIS, il semble que lorsque la distance entre le nombre de sons et le nombre de disques présentés est grande (e.g. 1 disque + 4 sons ou 1 son + 4 disques), le degré d'intégration des deux signaux est moins grand et le degré de ségrégation des signaux est conséquemment plus fort. Lorsque le nombre de sons présentés est très proche du nombre de disques présentés (e.g. 1 disque + 2 sons ou 1 son + 2 disques), l'intégration des percepts audio-visuels a lieu. Selon Shams et al. (2005), le percept global est alors le résultat de l'influence du percept auditif sur le percept visuel, puisque la variance du signal auditif seul est plus petite que la variance du signal visuel seul. En d'autres mots, l'information provenant de la modalité auditive est plus fiable que l'information provenant de la modalité visuelle, ce qui est consistant

avec les modèles d'intégration optimale de l'information multisensorielle proposés par d'autres études (Alais & Burr, 2004 ; Ernst & Banks, 2002 ; Ernst & Bühlhoff, 2004).

Le modèle proposé par Shams et al. (2005) tient compte des situations de ségrégation et d'intégration partielle ou complète (tout dépendant des conditions audio-visuelles, l'illusion apparaît ou non et l'essai en question entre dans une des conditions) et il permet de postuler que l'illusion pourrait être le résultat d'une stratégie « optimale » du système nerveux. Comment cette stratégie est-elle implémentée dans le cerveau? Où résident les « priors »? Où sont les « likelihoods »? Et où, quand et comment ces informations sont-elles combinées?

Études en électrophysiologie et en imagerie cérébrale

Des études ont été conduites dans le but de mieux identifier les composantes électrophysiologiques ainsi que les substrats neurologiques associés à l'IVIS. Les potentiels évoqués visuels associés à l'illusion ont été mesurés (Shams, Kamitani, Thompson & Shimojo, 2001 ; Arden, Wolf & Messiter, 2003). Ces études ont rapporté une modulation de l'activité dans les aires visuelles lorsque les disques sont accompagnés de sons comparativement à la condition où les disques sont présentés en absence de sons. De plus, ces études rapportent que l'activité électrophysiologique associée au disque illusoire est très similaire à celle associée au disque réel. Ceci suggère que les « likelihoods » et les « posteriors » seraient dans le cortex visuel. Cependant, d'autres études ont plutôt montré qu'il existe une activité spécifique à l'illusion. En effet, Bhattacharya, Shams et Shimojo (2002) ont montré que les ondes

gamma sont corrélées avec la perception du disque illusoire. Par la suite, Mishra, Martinez, Sejnowski et Hillyard (2007) ont montré, à l'aide de potentiels évoqués, que le patron d'activité corticale associée au disque illusoire est différent du patron d'activité corticale associée à la présentation d'un disque réel. Bien qu'il n'y ait pas de consensus sur la nature des mécanismes sous-tendant l'illusion, une modulation générale du cortex visuel primaire est présente lorsqu'il y a occurrence de l'illusion et ce, autant dans les études d'électroencéphalographie (Shams, Kamitani, Thompson & Shimojo, 2001 ; Bhattacharya, Shams & Shimojo, 2002 ; Arden, Wolf & Messiter, 2003 ; Mishra, Martinez, Sejnowski & Hillyard, 2007 ; Mishra, Martinez & Hillyard, 2008) que dans les études d'imagerie par résonance magnétique fonctionnelle (Watkins, Shams, Tanaka, Haynes & Rees, 2006 ; Watkins, Shams, Josephs & Rees, 2007) ou de magnétoencéphalographie (Shams, Iwaki, Chawla & Bhattacharya, 2005). De plus, Watkins, Shams, Josephs et Rees (2007) ont montré que cette modulation de l'activité du cortex visuel varie selon la présence de l'effet de fusion (qui génère une diminution de l'activité du cortex visuel primaire) ou de l'effet de fission (qui génère une augmentation de l'activité du cortex visuel primaire).

Ces résultats sont fort intéressants puisqu'ils suggèrent qu'une modulation de l'activité cérébrale pourrait être engendrée spécifiquement par la présence d'une stimulation visuelle illusoire. Cependant, il est aussi possible que les résultats obtenus soient liées aux spécificités des paramètres utilisés lors des différentes tâches, et non au percept illusoire comme tel. Mishra et al. (2007) ont montré que l'occurrence de l'IVIS — telle que décrite par Shams, Kamitani et Shimojo (2000) — était très variable d'un participant à l'autre. En effet, il semble qu'un disque

illusoire serait perçu dans 3 à 86% des essais où un disque et deux sons sont présentés. Cette variation dans l'occurrence de l'IVIS est probablement due à une mauvaise connaissance des paramètres qui mènent à l'illusion. Aussi, la définition de l'IVIS en termes psychophysiques reste vague. L'IVIS est-elle définie comme l'impression de voir plus de disques en présence de sons ? Si oui, comment est-ce quantifié ? En termes psychophysiques, que signifie « voir plus » ? Est-ce défini selon une relation linéaire entre le nombre de disques présentés et le nombre de disques perçus, ou est-ce plutôt défini comme une relation logarithmique entre le nombre de disques présentés et le nombre de disques perçus ? Quels sont les paramètres qui permettent de « voir plus » ? Au niveau temporel, quels délais sont nécessaires pour que l'IVIS soit présente ? Le manque de réponses à ces questions reflète qu'aucune limite fut clairement établie concernant les paramètres psychophysiques de l'IVIS. Puisque ces bornes n'ont pas été établies et compte tenu du manque de précision en ce qui a trait à la définition du phénomène illusoire observé, il semble prématuré de tenter d'identifier les mécanismes neuronaux sous-tendant l'IVIS.

Objectifs du travail

Tel que mentionné plus haut, les paramètres psychophysiques de L'IVIS méritent une grande attention puisqu'ils permettront de déterminer ce qui qualifie réellement l'illusion. Dans une tentative d'élargir les connaissances sur le phénomène, Berger, Martelli et Pelli (2003) ont montré qu'il était possible d'utiliser l'IVIS en tant que mesure psychophysique de bas niveau de l'apport de l'intégration multisensorielle soit, dans ce cas-ci, une amélioration de la performance à une tâche

de discrimination de l'orientation d'une grille sinusoïdale. Nous avons voulu vérifier si cet effet était dû à l'induction d'un état de vigilance, proportionnel à la fréquence de l'enveloppe sonore. Notre hypothèse était que lorsque l'enveloppe sonore contient plus d'énergie, l'état de vigilance des participants est accru, ce qui entraîne la perception d'un disque illusoire. De plus, le fait de percevoir un stimulus visuel supplémentaire nous permettrait d'avoir une meilleure performance dans une tâche connexe (ici discriminer l'orientation du stimulus visuel). Pour vérifier cette hypothèse, notre premier objectif était de répliquer l'effet de sensibilité accrue de Berger et al. (2003) — qui, à notre connaissance, n'avait jamais été répliqué. Suite à plusieurs tentatives infructueuses dont nous parlerons plus loin, nous avons décidé de tenter de répliquer l'effet de Shams et al. (2000) — reproduit par plusieurs autres — pour nous assurer du bon fonctionnement de notre matériel expérimental. Suite à cette réplification, nous avons tenté de répliquer la première expérience de Berger et al. — qui n'avait jamais été répliquée par d'autres. Ces réplifications — tantôt infructueuses, tantôt fructueuses — constituent les études préliminaires de ce mémoire et seront présentées dans la prochaine section. L'expérience principale de ce travail sera ensuite décrite en détails.

Études préliminaires

Tout d'abord, nous avons tenté de répliquer la deuxième expérience de Berger et al. Cette expérience, telle que décrite dans l'article, était constituée de deux intervalles : l'un dans lequel la grille sinusoïdale était présentée à la verticale et l'autre dans lequel la grille sinusoïdale était inclinée d'un degré par rapport à la verticale. Les grilles sinusoïdales étaient présentées à une fréquence fixe de 4 Hz et

elles étaient toujours accompagnées de sons présentés à des fréquences variables (les mêmes fréquences que dans la première expérience). La tâche des participants était de déterminer dans lequel des deux intervalles la grille sinusoïdale était inclinée. Les auteurs ont obtenu pour résultat que lorsque la fréquence de présentation des sons était supérieure à la fréquence de présentation de la grille sinusoïdale (et donc que l'illusion était présente), les observateurs étaient meilleurs pour discriminer l'orientation de la grille sinusoïdale. En contraste, lorsque la fréquence de présentation des sons était inférieure à la fréquence de présentation de la grille sinusoïdale (et donc que l'illusion n'était pas présente, ou même qu'il y avait présence d'un effet de fusion), les observateurs étaient moins bons pour discriminer l'orientation de la grille sinusoïdale.

Nos tentatives de répliquer cet effet furent vaines. Malgré la modification de certains paramètres (degré d'excentricité de la grille sinusoïdale, variation de la fréquence de présentation de la grille sinusoïdale), nous n'avons jamais réussi à répliquer l'effet obtenu par Berger et al. (2003). L'ambiguïté dans les paramètres rapportés par les auteurs pourrait constituer un facteur important menant à notre incapacité à répliquer l'expérience. Un autre facteur qui aurait pu être en cause est le mauvais fonctionnement de notre matériel expérimental. Pour nous assurer que le tout fonctionnait correctement, nous avons tenté de répliquer l'expérience de Shams, Kamitani et Shimojo (2000) qui avait déjà été répliquée à plusieurs reprises par d'autres groupes de chercheurs. Nous avons réussi à répliquer ces résultats sans difficulté (voir la première expérience de l'article présenté plus bas pour plus de détails sur les paramètres de la réplification).

Puisque nous étions maintenant assurés du bon fonctionnement de notre matériel expérimental, nous avons tenté de répliquer la première expérience de Berger, Martelli et Pelli (2003), expérience qui n'avait jamais été répliquée auparavant. L'expérience constitue une tâche subjective qui consiste à juger de la fréquence à laquelle sont présentées des taches de Gabor (en périphérie du champ visuel) lorsqu'elles sont accompagnées de sons purs. Les sons sont présentés à des fréquences variables (0, 2, 2.5, 3, 3.5, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 et 33.3 Hz), mais les grilles sinusoïdales sont toujours présentées à la même fréquence (4 Hz). Le paradigme expérimental est construit de sorte que deux intervalles sont présentés aux observateurs : un premier intervalle comprend une grille sinusoïdale présentée à 4 Hz et associée à un son d'une fréquence variable et un deuxième intervalle comprend uniquement la grille sinusoïdale. Chaque intervalle a une durée d'une seconde. Chacun des sons est présenté pendant 5 essais consécutifs. La tâche du participant est d'ajuster la fréquence de la grille sinusoïdale du deuxième intervalle (à l'aide d'une barre de défilement) en fonction de ce qu'il a perçu lors du premier intervalle. Cette fréquence ajustée est ensuite conservée pour le prochain essai, où l'observateur doit réajuster son tir (jusqu'à concurrence de 5 essais, où la fréquence sonore est alors modifiée). Suivant le principe de l'illusion visuelle induite par le son, lorsque la fréquence sonore est plus élevée que la fréquence visuelle, l'observateur devrait avoir l'impression que la fréquence visuelle est plus grande que ce qui est réellement présenté, ce qui fut obtenu par Berger et al. (2003).

Lorsque nous avons tenté de répliquer cette tâche, plusieurs problèmes sont apparus et nous avons dû effectuer de nombreuses modifications pour arriver aux résultats obtenus par Berger et al. (2003). Tout d'abord, l'article ne mentionne pas à quelle fréquence est présentée la grille sinusoïdale du deuxième intervalle. Nous avons essayé de répliquer l'expérience avec un intervalle fixé à 4 Hz, mais les résultats obtenus ne correspondaient pas du tout à ce que l'on retrouve dans l'article de Berger et al. (2003). Nous avons alors fait varier la fréquence de la grille sinusoïdale du deuxième intervalle entre 2 et 6 Hz, mais les résultats obtenus n'étaient pas différents de ceux obtenus précédemment. Un autre problème important de l'expérience était le paradigme expérimental comme tel. Le fait de présenter deux intervalles différents et de demander aux participants de juger de la fréquence de présentation de la grille sinusoïdale en ajustant leur tir sur 5 essais consécutifs apportait très peu de variabilité dans les réponses des participants, surtout compte tenu du fait que la grille sinusoïdale du premier intervalle était toujours présentée à 4 Hz. Nous avons tenté de faire varier la fréquence de la grille sinusoïdale du premier intervalle, mais cela n'a rien changé aux résultats obtenus, qui étaient tout aussi aberrants. Nous avons donc laissé tomber l'idée d'ajuster le jugement de l'observateur sur 5 essais consécutifs. Nous avons aussi laissé tomber l'idée des deux intervalles, puisque nous ne pouvions pas imaginer comment cela pouvait fonctionner. Pour palier à ce problème, nous avons décidé d'utiliser un seul intervalle d'une seconde pendant lequel la fréquence de présentation de la grille sinusoïdale était variable (entre 2 et 6 Hz), autant que la fréquence de présentation des sons (mêmes fréquences que celles présentées dans l'article de Berger et al., 2003, voir la Figure 1 pour une représentation schématique de la tâche).

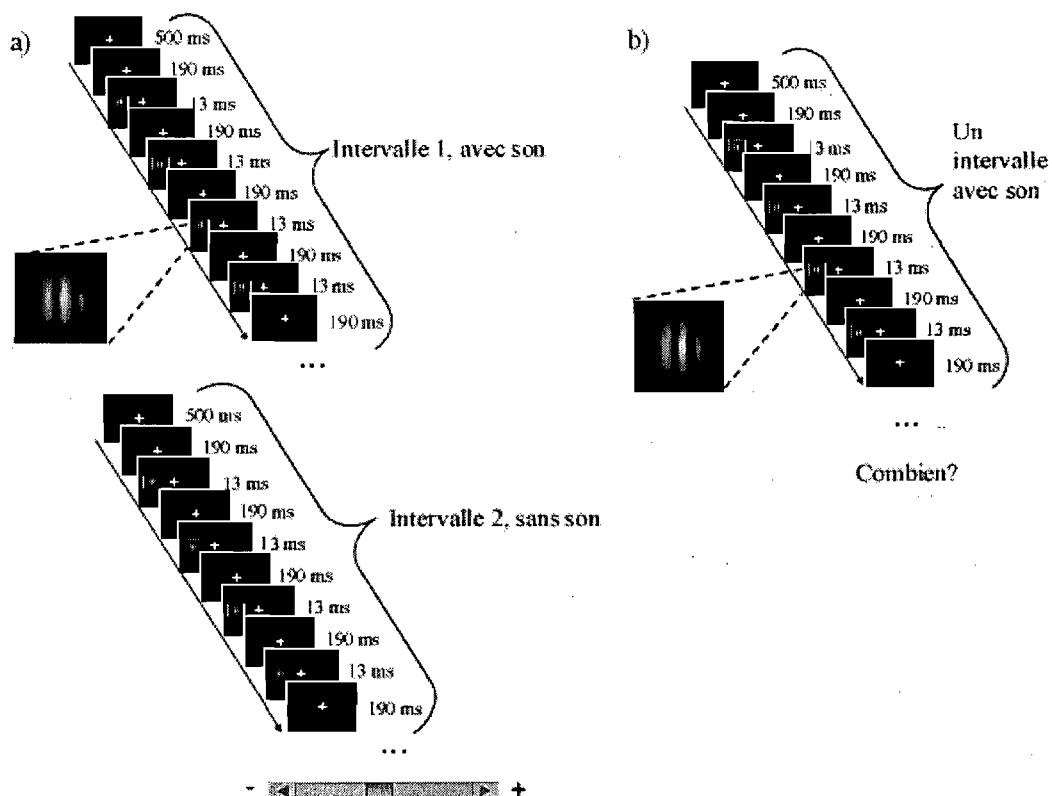


Figure 1 a) Schéma représentant la tâche initiale de la première expérience de Berger et al. Après la présentation des deux intervalles, les participants devaient ajuster la fréquence de présentation du stimulus visuel à l'aide de la barre de défilement. La partie b) représente un schéma de la tâche que nous avons élaborée dans le but de répliquer l'expérience présentée en a). Suite à la présentation de l'intervalle, les participants devaient déterminer combien de stimuli visuels étaient présentés et appuyer sur la touche correspondant à leur choix.

Cette nouvelle procédure a contribué à améliorer grandement l'efficacité de la tâche et nous obtenions déjà une plus grande proportion d'essais présentant l'illusion telle que décrite par Shams et al. (2000). Cependant, un autre problème fondamental de l'expérience restait, soit le très mince degré d'excentricité de la grille sinusoïdale (rapporté à 1.14 dans l'article). Dans l'étude de Shams et al. (2000), le stimulus

visuel était présenté à 5 degrés d'excentricité. En suivant les données de ce dernier article, en plus des modifications précédemment décrites, nous avons finalement réussi à répliquer les données obtenues dans la première expérience de Berger et al. (2003). Les résultats moyens des 7 participants sont présentés à la Figure 2.

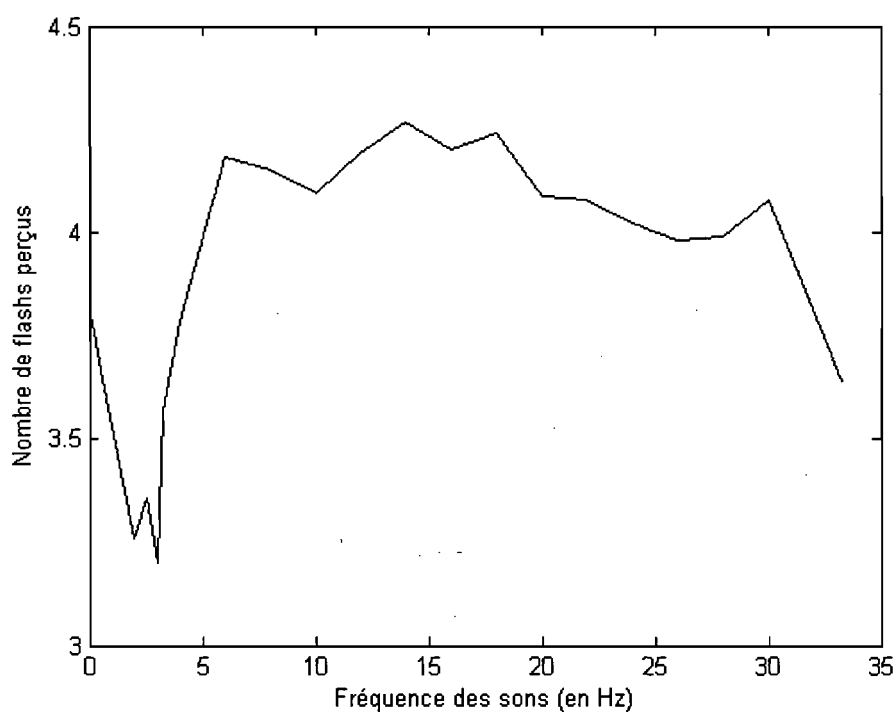


Figure 2. Réplication des résultats de la première expérience de Berger et al.

Décours temporel de l'intégration de l'information multisensorielle

Puisque le projet initial reposait entièrement sur la capacité à répliquer l'effet objectif décrit par Berger et al. et que nous n'avons pas réussi à le reproduire, le projet a dû prendre une autre tangente. En tentant de répliquer l'effet de Berger et al., nous nous sommes questionnés à plusieurs niveaux, mais entres autres, nous nous

sommes demandé si les paramètres temporels du paradigme expérimental étaient optimaux à l'apparition de l'effet illusoire. Bien qu'il ait été montré qu'une certaine fenêtre temporelle (environ 100 ms) est nécessaire pour que l'illusion soit présente (Shams et al., 2002), la littérature à ce sujet est peu explicite. Il nous semble essentiel de vérifier les conditions temporelles dans lesquelles l'illusion apparaît puisque l'information apportée par des contraintes potentielles permettra certainement de nous renseigner sur les conditions dans lesquelles le son et la vision s'intègrent.

Il semble évident que certaines associations perceptuelles sont apprises très tôt, par simple exposition à une situation. Par exemple, le fait que la lumière provienne du soleil, qui lui se situe au-dessus de nos têtes, a tendance à influencer notre perception du relief de cercles auxquels on induit une exposition à la lumière provenant de différentes directions. Si la lumière provient du haut, l'observateur a tendance à percevoir un cercle comme étant une bosse alors que si la lumière provient du bas, l'observateur a tendance à percevoir un cercle comme étant un creux (voir Figure 3).

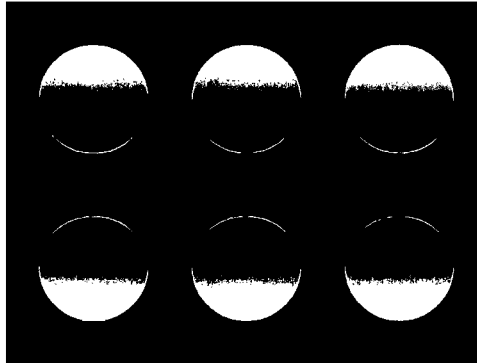


Figure 3. Cette figure démontre bien l'influence de la supposition que la lumière provient toujours du haut sur notre perception. Lorsque la figure est observée à l'endroit, les cercles de la rangée de haut semblent être des bosses alors que les cercles de la rangée du bas semblent être des creux. Lorsqu'on tourne la page, les cercles qui semblaient être des bosses semblent maintenant être des creux et vice-versa.

Dans ce cas-ci, l'association entre la localisation de la lumière et la perception de profondeur constitue un facteur influençant notre perception du monde. Dans le cas de l'illusion, il est possible que le délai entre les stimuli visuels et auditifs soit un facteur influençant la perception d'une façon semblable à ce qui a été montré avec la perception de la lumière. En effet, lorsqu'on est face à une situation multimodale, le cerveau doit décider si les signaux provenant des différentes modalités proviennent aussi de différentes sources ou d'une même source. Sur quoi se base-t-il pour prendre sa décision ? Une hypothèse serait qu'il se base sur les éléments de synchronie appris par expérience. On pourrait penser que lorsqu'un son et une image arrivent « en même temps » au cerveau, il y a de fortes chances que ces deux événements proviennent d'une même source et que le cerveau les interprète comme étant un seul événement. Lorsque la source est près de soi (distance de moins de 10 mètres), les événements provenant des modalités auditive et visuelle arrivent avec un court délai et il est plus facile de déterminer s'ils proviennent de la même source, mais plus la

source est éloignée, plus le délai entre l'arrivée des événements est grand et plus le cerveau a de la difficulté à interpréter la provenance de ces événements (Sugita & Suzuki, 2003). De plus, pour arriver à déterminer si deux événements proviennent de la même source, le cerveau a certainement recours à une fenêtre temporelle implicite et inconsciente de l'intégration de l'information multisensorielle. Cette fenêtre doit, de toute évidence, être définie par des bornes bien établies, sans quoi une suite d'événements qui n'ont pas de lien entre eux seraient incorrectement intégrés. Par exemple, si le cerveau possédait une fenêtre temporelle de l'intégration de l'information multisensorielle trop large, on pourrait imaginer que le beau temps qui suit un orage pourrait être associé aux grondements du tonnerre et ainsi, le beau temps et le tonnerre seraient perçus comme étant le même événement. Dans le cas de l'illusion, il est possible que la fenêtre temporelle d'approximativement 100 ms rapportée par Shams et al. (2002) corresponde à une fenêtre pendant laquelle le cerveau peut interpréter l'apparition d'un son et d'un disque comme étant deux événements provenant d'une même source. Le cerveau associerait donc ces deux signaux pour former un percept global audio-visuel. En s'appuyant sur le modèle proposé par Ernst et Bühlhoff (2004), ce percept global pourrait être activé même si une seule des deux composantes du percept est disponible. Par exemple, si le cerveau interprète l'occurrence d'un son et d'un disque comme étant le produit d'un seul et même événement, la présentation subséquente d'un son clairement perceptible en l'absence de disque pourrait tout de même activer la représentation d'un disque chez l'observateur.

Afin d'approfondir les caractéristiques temporelles de l'illusion, nous avons créé un paradigme expérimental qui amalgame les paramètres de la tâche subjective de Berger et al. à ceux de la tâche présentée par Shams et al. (2000). Les stimuli visuels utilisés sont les disques blancs de la tâche de Shams et al. (2000), mais ils sont présentés à une fréquence de 2, 3, 4, 5, ou 6 Hz. Les stimuli auditifs ont les mêmes caractéristiques que dans la tâche de Berger et al., soit des sons purs d'une intensité de 3 kHz et d'une durée de 30 ms chacun. Cependant, dix sons sont présentés aléatoirement dans chacun des essais, ce qui signifie que le délai entre les sons et les disques varie à chaque présentation, nous permettant de vérifier l'impact des différents délais disque-son sur l'illusion. L'étude complète sera présentée en détails dans l'article qui suit.

CONTRIBUTIONS À L'ARTICLE

L'idée originale revient à Frédéric Gosselin, de même que la méthode utilisée. La conception des programmes d'expérimentation ainsi que les analyses de la première expérience ont été effectuées conjointement par Catherine Éthier-Majcher et Caroline Blais. Le recrutement des participants et la cueillette des données a été faite par Catherine Éthier-Majcher. Les données ont été analysées par Frédéric Gosselin, en collaboration avec les deux co-auteurs. Catherine Éthier-Majcher a écrit les premières versions de l'article et tous les co-auteurs ont participé à l'écriture des versions subséquentes. Tous les co-auteurs consentent à ce que l'article soit inclus dans le présent mémoire, comme l'atteste le formulaire suivant.

CHAPITRE DEUXIÈME : ARTICLE


Éthier-Majcher, C., Blais, C., Gosselin, F. (en préparation). Is timing everything for
the sound-induced flash illusion?

Is timing everything for the sound-induced flash illusion?

Catherine Éthier-Majcher, Caroline Blais & Frédéric Gosselin

Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département
de psychologie, Université de Montréal

Address correspondence to

Frédéric Gosselin, Ph.D.
Département de psychologie
Université de Montréal
C.P. 6128, Succ. Centre-ville
Montréal, Qc
H3C 3J7
Fax: 514-343-5787
Phone: 514-343-7550
Email: 

Abstract

It has been shown that timing is an important aspect for multisensory illusory effects — such as the McGurk effect (Munhall, Gribble, Sacco & Ward, 1996) and the ventriloquist effect (Slutsky & Recanzone, 2001) — to occur. In the sound-induced flash illusion, when two beeps are presented simultaneously with a flash, a second illusory flash is perceived (Shams, Kamitani & Shimojo, 2000). In a first experiment, we replicated this effect with slightly different parameters. In a second experiment, we investigated the importance of timing in the sound-induced flash illusion. Therefore, we modified the original paradigm of the sound-induced flash illusion considerably, so that the delay between a flash and a beep was systematically varied by using a classification image technique. Our results show that beeps presented at the beginning and at the end of the audio-visual stimulation are important in the occurrence of the illusory flash. However, no specific delay between a flash and a beep enhances the occurrence of the illusion.

Introduction

In any multisensory stimulation, timing is crucial, for it helps our system decide whether sensory signals need to be integrated or segregated. For example, when an auditory track and a matching visual track are presented within a very short delay (about 250 ms), our system usually perceives them as part of the same event (Dixon & Spitz, 1980). It has been shown that distance, as well as expectations resulting from previous experiences, are taken into account when judging the acceptable delay (known as the temporal window) for two stimulations to be integrated (King, 2005 ; Sugita & Suzuki, 2003). For example, in audio-visual stimulations, light travels faster than sound, which causes a disparity in the perception of auditory and visual signals. If the delay is within the temporal window for integration of information, our system is able to integrate the two signals accurately into a global multisensory percept. According to the discrepancy between speed of light and sound, the temporal window for audio-visual integration of information is often asymmetrical, because our system expects a delay of a variable duration – depending on distance – between the perception of the two stimulations. For example, our degree of tolerance to audio-visual speech asynchrony reaches a disparity of 250 ms between the two signals when they are presented at a proper distance, with auditory signal lagging behind visual signal. However, when sound precedes sight, our degree of tolerance to the asynchrony of the signals is smaller and the temporal window of integration is approximately 130 ms (Dixon & Spitz, 1980).

In speech perception, association between the auditory and the visual modalities is obvious: in everyday life, language usually comes from the mouth of

someone and moving lips usually come with hearing language. The ecological character of this association drives our system to “naturally” integrate both signals together. In special conditions, this “natural” integration of image and sound leads to illusory effects such as the McGurk effect – where a visual “ga” combined with the sound “ba” results in hearing “da” – or the ventriloquist effect – where a sound source is mislocalized based on the influence of a visual stimulation (for example, when we attend a concert, the singer’s voice seems to emanate from his lips rather than from the speakers). In these effects, both modalities interact, which generates a global percept different from the physical reality. It has been shown that, for these multisensory illusory effects to occur, timing is crucial. For example, the McGurk effect fails to occur with an audio-visual asynchrony exceeding 200-300 ms (Munhall, Gribble, Sacco & Ward, 1996) and the ventriloquist effect fails to occur with an audio-visual asynchrony exceeding 300 ms (Slutsky & Recanzone, 2001).

In the sound-induced flash illusion, the presentation of one flash accompanied by two beeps tends to lead observers to perceive more than one flash (Shams, Kamitani & Shimojo, 2000). In this effect, the conditions leading to audio-visual integration or segregation are not clear. Could timing be as important for this audio-visual illusion as for other multisensory effects? The sound-induced flash illusion is more abstract than the multisensory effects mentioned previously inasmuch as there is no ecological association between the auditory and the visual stimulations. Why abstract auditory and visual stimuli such as beeps and flashes would be associated besides any logical context? In fact, even though the psychophysics of this phenomenon has been studied by a number of authors (Andersen, Tiippana & Sams,

2004 ; Berger, Martelli & Pelli, 2003 ; McCormick & Mamassian, 2008 ; Shams, Kamitani & Shimojo, 2002), little is known about the exact parameters driving the illusion. Recent studies have shown that the effect is not as robust as it was previously thought (Mishra, Martinez, Sejnowski & Hillyard, 2007), and there is a large disparity in the occurrence of the illusion from a subject to another. On average, only 37% of the trials where two beeps are presented with one flash generate the perception of an illusory second flash. A number of reasons might explain this variable occurrence of the illusion, and one of them could be that the parameters used to generate the illusion are non-optimal.

Since timing is an important parameter to the ecological multisensory interaction effects, if these effects are comparable to the sound-induced flash illusion, one of the critical parameters of the illusion might be the delay between the presentation of a beep and a flash. Here, we report two experiments. First, we replicate Shams et al. (2000) with our equipment and with slightly different parameters. Second, we systematically vary timing between beeps and flashes by using a classification image technique.

Methods

Experiment 1

Participants

Eight students from the Université de Montréal volunteered for this experiment. All had normal or corrected to normal visual acuity. All had normal hearing. Participants gave their informed consent before inclusion in the study.

Stimuli

The visual stimuli (henceforth “flashes”) were uniform white disks (119 cd/m²) presented for 12 ms on a black background (0.79 cd/m²). The disks were subtending 2 degrees of visual angle and were presented binocularly at 8 degrees eccentricity left to a centered fixation cross. The flashes were displayed at a resolution of 1024 x 768 pixels at a refresh rate of 85 Hz. The viewing distance was maintained constant at 50 cm using a chinrest. The auditory stimuli (henceforth “beeps”) were 3 kHz pure sounds played binaurally at 81 dB. Either 1, 2, 3, or 4 flashes were accompanied by either 0, 1, 2, 3, or 4 beeps.

Procedure

The experiment ran on a HP Compaq Pentium 4. Stimuli were created using Matlab (MathworksTM) and functions from the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). Flashes were displayed on a HP 7750 high-resolution computer monitor and beeps were presented with *Cyber Acoustics* headphones. The experiment consisted of a total of 100 trials (5 auditory conditions x 4 visual conditions, repeated 5 times). The beep-flash conditions were randomized. Each trial started with a fixation cross appearing in the center of the computer monitor for 500 ms. Then, the first beep was presented and 27 ms later, the first flash was presented. Beeps were spaced by 57 ms and flashes by 47 ms. The total duration of one trial depended on the beep-flash condition. For example, for a trial where two flashes were presented with three beeps, the total duration was 635 ms. Participants were asked to report the number of disks (1, 2, 3, or 4) by depressing the appropriate keyboard key.

Results and discussion

Overall, our results replicate Shams et al. (2000), as we find the same fission effect (see Figure 4). For instance, when one flash is presented with two beeps, observers tend to perceive more flashes. Furthermore, this fission effect is observed in all conditions where more beeps than flashes are presented. In addition, our results show a fusion effect observed when more flashes than beeps are presented. The occurrence of this fusion effect is congruent with previous replications of the sound-induced flash illusion (Andersen et al., 2004). Interestingly, when a number of flashes are presented without sound, observers do not necessarily perceive the exact amount of flashes presented. This might be resulting from the parameters of the task itself. In our experimental paradigm, every beep-flash combination is presented the same amount of times. Hence, since the probability that less than one flash or more than four flashes were presented was 0, it is not surprising that the average number of flashes reported was biased toward 2.5 flashes, the average number of flashes presented. This could explain why the presentation of one or two flashes without sound is perceived as more, whereas the presentation of three or four flashes is perceived as less.

These results show that our equipment, combined with the use of slightly different parameters (i.e. presenting the flashes in the periphery of the left visual field instead of underneath the fixation cross), replicated the sound-induced flash illusion.

Insert Figure 4 near here

Experiment 2

Participants

Six students from the Université de Montréal (not the same as in experiment 1) volunteered for this experiment. All had normal or corrected to normal visual acuity. All had normal hearing. Participants gave their informed consent before inclusion in the study.

Stimuli

The parameters of the visual stimuli were the same as in experiment 1, except that the disks were flashing at different frequencies (either 2, 3, 4, 5 or 6 Hz) and they were displayed at a resolution of 1024 x 768 pixels and at a refresh rate of 75 Hz. The frequency at which disks were displayed was confounded with the number of disks presented (e.g. when 3 disks were presented, they were displayed at a 3 Hz frequency). The phase of the visual stimulation remained constant across trials. The viewing distance was maintained constant at 75 cm using a chinrest. The auditory stimuli were 3 kHz pure sounds each lasting 30 ms. For each trial, 10 beeps were presented at random positions on the one-second temporal vector of the trial. The number of possible beep positions was determined in order to avoid beep clusters and respect a minimal delay of 30 ms between beeps. A total of 8656 beep positions (48% of the total temporal vector) were sampled. For each trial, 10 of these

positions were chosen randomly in order to create the sound vector, which resulted in systematic variations in the delay between a flash and a beep. For each trial, a fixed number of beeps (10) were accompanied by either 2, 3, 4, 5 or 6 flashes. Our paradigm is quite different from the way the illusion was originally presented – the number of beeps is fixed and the number of flashes goes up to six – but it serves our purpose as we aim to isolate the effect of the delay between a flash and a beep.

Procedure

The basic procedures were the same as in experiment 1. The experiment consisted of 25 blocks of 100 trials each, for a total of 2500 trials per participant. The number of flashes presented for each trial as well as the positions of the beeps was randomized. Each trial began with a small fixation dot appearing in the center of the screen for 500 ms. The audio-visual stimulation per se started immediately after, lasting for a total duration of 1s. Participants were asked to report the number of perceived disks (2, 3, 4, 5, 6, 7, or 8) by depressing the appropriate keyboard key. Participants did not know that a maximum of 6 flashes could be presented and we chose to give them the possibility to respond that 7 or 8 flashes were presented, allowing the presence of illusory flashes when 6 physical flashes were presented.

Results and discussion

The percentages of all possible answers for each audio-visual condition are presented in Figure 5. Even though six participants did the experiment, the results of only three of them are presented here, for response biases were observed in the results of the three other participants (i.e. they would always respond the same

amount of perceived flashes, which leads to very few variation in the response distribution). These biases are not inconsistent with results reported by Andersen et al. (2004). In order to identify, for each visual condition, the sound patterns leading to a specific number of perceived flashes, we summed the sound vectors leading to a particular response-stimulus category. For example, when two flashes were presented, all the sound vectors leading to a perception of three flashes were summed across subjects. Prior to the summations, we changed the sampling rate of the sound patterns from 18,000 to 600 points using the `resample` Matlab function (Mathworks Inc.). The resulting classification sounds were smoothed with a Gaussian kernel of 25 ms. We performed Bootstrap analyses (e.g., Efron & Tibirishiani, 1986) to estimate the mean and the standard deviation of the distribution of the null hypothesis at every point of the smoothed classification sounds (i.e., what would have been observed if there was no link between responses and beeps). These means and standard deviations were used to transform the points of the smoothed classification sounds into Z-scores. We applied a Pixel test to the Z-scored classification sounds ($p < .05$, two-tailed). The statistical threshold provided by this test corrects for multiple comparisons while taking the spatial correlation inherent to structured images into account (Chauvin, Worsley, Schyns, Arguin & Gosselin, 2005).

Figure 6 shows the plotted classification sounds for the conditions where: the number of flashes presented and the response are the same (e.g. 3 flashes presented and 3 perceived flashes), more flashes (1 more) than the amount presented is perceived (e.g. 3 flashes presented and 4 flashes perceived), and fewer flashes (1

less) than the amount presented is perceived (e.g. 3 flashes presented and 2 flashes perceived). Since there is a small number of trials where the difference between the number of flashes perceived and the number of flashes presented exceeds one, we do not report these conditions (see Figure 5).

Insert Figure 5 and 6 near here

In most conditions, it seems that the beginning and the end of the auditory stimulation is important. For two and three flashes presented, the presence of a beep at the beginning and at the end of the auditory stimulation drives a fission effect, where one more beep than the amount presented is perceived. Conversely, the absence of a beep at those positions drives a fusion effect, where fewer flashes than were presented are perceived. There seems to be a different strategy involved for four or five flashes presented. In these conditions, the beginning and the end of the auditory stimulation seems to be important, but in a different way. A specific pattern is hard to figure, and further analysis would be necessary to understand this different strategy. In any case, the results for the conditions where four, five or six flashes are presented have to be interpreted carefully. As we can see in Figure 5, the percentages of responses where the number of perceived flashes is coherent with the number of flashes presented are lower for 4, 5 or 6 flashes than for 2 or 3 flashes. This might be explained by a difficulty in counting the number of flashes. Hence, the results for the conditions where 4, 5 or 6 flashes are presented might not be the consequence of an

illusory perception, but more of a difficulty in executing the task. In the condition where six flashes are presented, nothing reaches significance.

General discussion

Our results show that, for two and three flashes presented, specific beep positions are important in driving the sound-induced flash illusion. Indeed, for those visual conditions, the presence of a beep at the beginning and at the end of the auditory stimulation drives a fission effect, whereas the absence of a beep at those positions drives a fusion effect. For more flashes presented, a different strategy seems to occur. When five flashes are presented, it seems that the absence of a beep at the beginning and at the end of the auditory stimulation can lead either to a fission effect or a fusion effect. Furthermore, for five flashes presented, the presentation of a beep at the beginning and at the end of the auditory stimulation leads to the perception of the correct number of flashes.

The classification sounds observed for three and fewer flashes when the observers perceived more flashes than the number actually presented are strikingly similar to the probability that a flash occurs across conditions and time (the priors in Bayesian terminology). This suggests that an association between a beep and a flash exists, and this association drives, especially when it's most likely (i.e., at the beginning and at the end of a trial), the perception of a flash even when, in reality, a beep alone was presented. This hypothesis represents only one possible explanation and other analysis of the data should be performed to provide a better understanding of the reasons why the specific auditory patterns observed here occur. It is not clear

why the beeps presented at the beginning and at the end of the auditory stimulation can play a significant role in the occurrence of a fusion effect as well as in the occurrence of a fission effect, or else, in some cases, in identifying the correct amount of flashes.

Even though the classification vectors show that specific beep positions lead to a fission or a fusion effect, the auditory patterns observed here do not show evidence that specific timing between a flash and a beep could be important. This might be because the association between a beep and a flash is arbitrary rather than logical. As opposed to more ecological effects, the sound-induced flash illusion might therefore depend on cognitive factors such as attention or other cognitive biases rather than depend on a specific temporal window for integration or segregation of information. A better understanding of the illusion could be provided by further studies presenting a variation in the psychophysical parameters of the sound-induced flash illusion or providing tools to observe the influence of cognitive biases on the phenomenon.

References

- Andersen, T., Tiippana, K., Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21, 301-308.
- Berger, T., Martelli, M., Pelli, D. (2003). Flicker flutter: Is an illusory event as good as the real thing? *Journal of Vision*, 3, 406-412.
- Brainard, D. H., (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.
- Chauvin, A., Worsley, K., Schyns, P., Arguin, M., Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5(9), 659-667.
- Dixon, N., Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9, 719-721.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54-75.
- King, A. J. (2005). Multisensory integration: Strategies for synchronization. *Current Biology*, 15(9), 339-341.

- McCormick, D., Mamassian, P. (2008). What does the illusory-flash look like? *Vision Research*, 48, 63-69.
- Mishra, J., Martinez, A., Sejnowski, T., Hillyard, S. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120-4131.
- Munhall, K.G., Gribble, P., Sacco, L., Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics*, 58, 351-362.
- Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics : Transforming numbers into movies, *Spatial Vision*, 10, 437-442.
- Shams, L., Kamitani, Y., Shimojo, S. (2002). Visual illusion induces by sound. *Cognitive Brain Research*, 14(1), 147-152.
- Slutsky, D.A., Recanzone, G.H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12, 7-10.
- Sugita, Y., Suzuki, Y. (2003). Implicit estimation of sound-arrival time. *Nature*, 421, 911.

Figure captions

Figure 4. Results of our replication of the sound-induced flash illusion as presented by Shams et al. (2000). When fewer beeps than flashes are presented, observers tend to perceive fewer flashes than presented, whereas when more beeps than flashes are presented, observers tend to perceive more flashes than presented.

Figure 5. For all subjects, the percentages of perceived flashes as a function of the number of flashes presented. For 4 flashes presented and more, the accuracy of the participants decreases. For all visual conditions, there is a small percentage of trials where participants perceive over 2 flashes more or less than the number presented.

Figure 6. The classification vectors averaged across subjects for each audio-visual condition. The X axis represents the time course of the auditory stimulation. The Y axis represents the presence (positive) or the absence (negative) of a beep in terms of Z scores. The blue lines represent the threshold for signification. The red stars (superimposed with the blue lines) represent the position of the flashes as well as the threshold for statistically significant results.

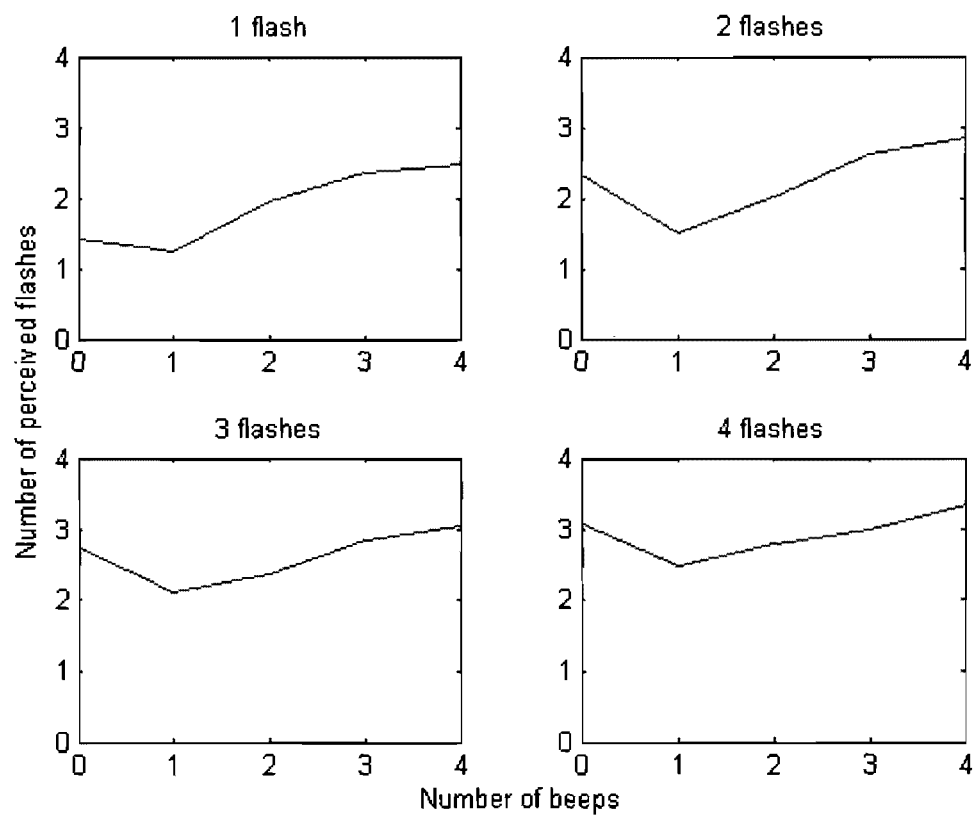


Figure 4

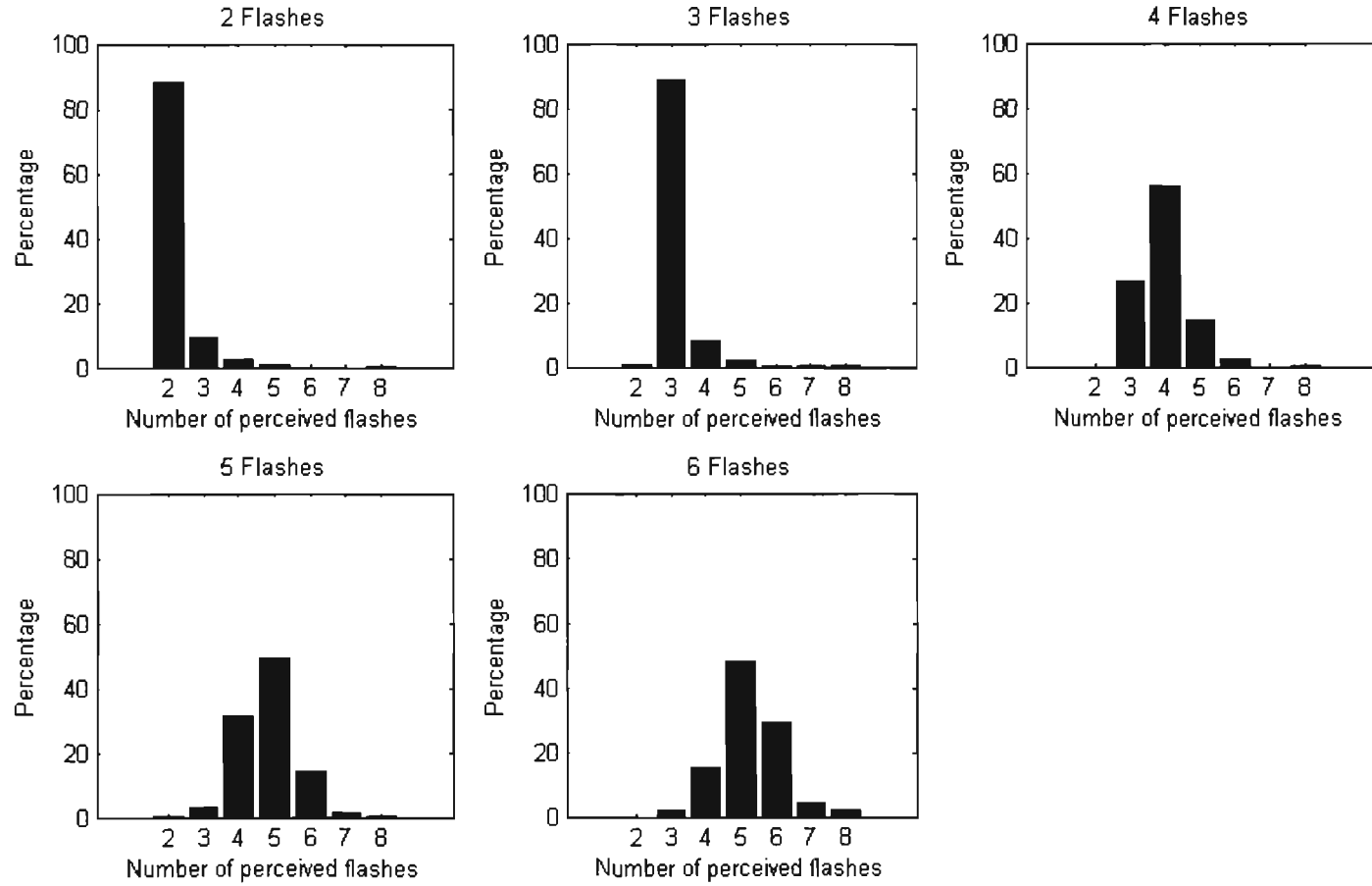


Figure 5

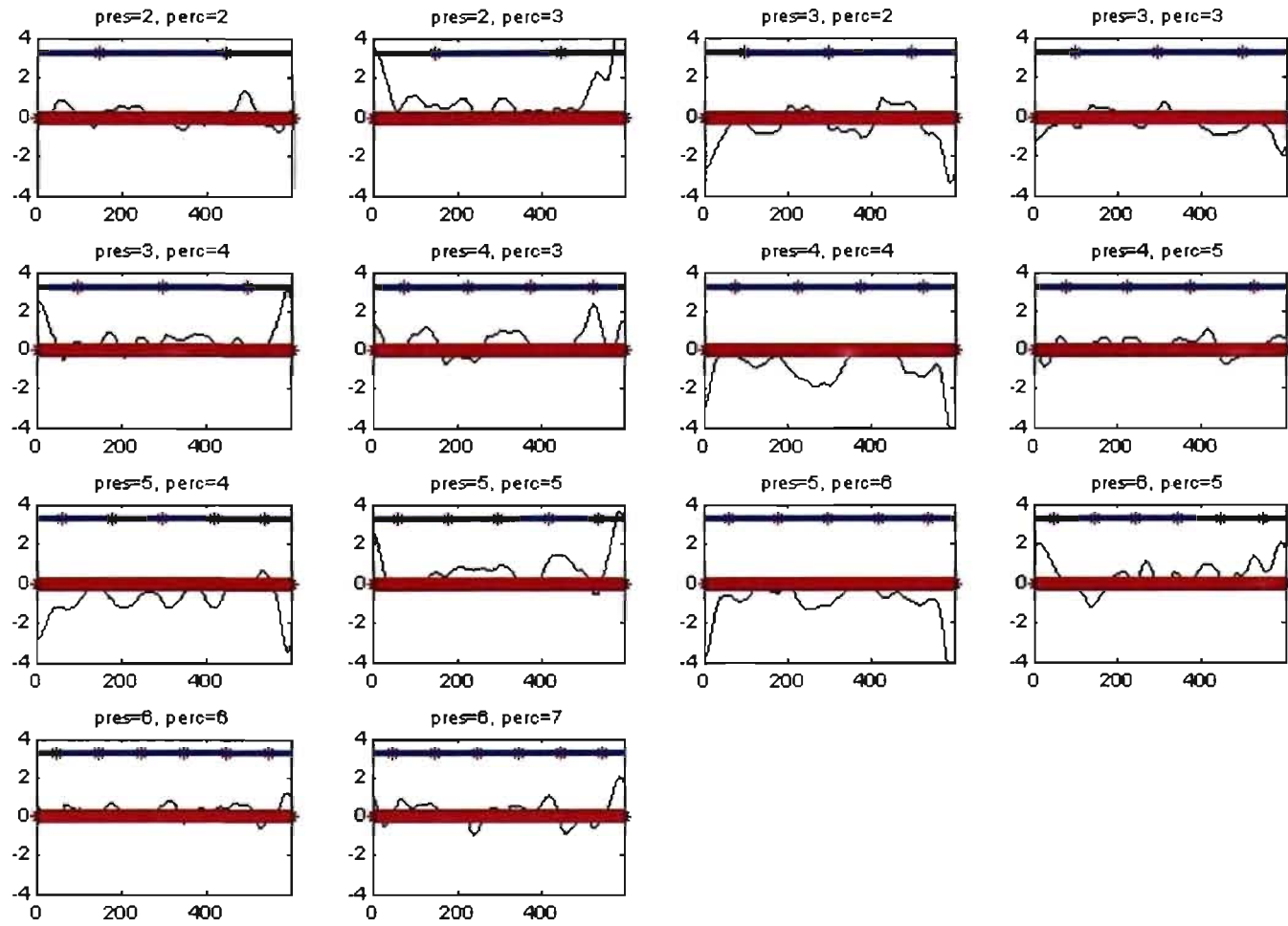


Figure 6

CHAPITRE TROISIÈME : DISCUSSION GÉNÉRALE

Les résultats présentés dans la section précédente suggèrent qu'a priori, l'IVIS ne repose pas sur des contraintes temporelles telles que le délai entre la présentation d'un son et d'un disque. Cependant, d'autres analyses de nos données pourraient encore démontrer que certains aspects temporels influencent l'IVIS. Par exemple, il n'est pas impossible qu'une analyse temps-fréquences (e.g., Tallon-Beaudry et Bertrand, 1999) révèle que certaines fréquences sonores constituent un facteur influençant l'occurrence de l'illusion à un moment ou à un autre de la stimulation auditive totale. Il serait alors possible d'observer, par exemple, qu'une prédominance de fréquences sonores élevées en début de stimulation tend à mener à une perception de plus de disques que ce qui fut réellement présenté.

Explication selon un modèle bayésien

Ce qui semble particulièrement influencer l'occurrence de l'illusion est la présence ou l'absence d'un son au début et à la fin de la stimulation sonore, et ce pour la majorité des conditions visuelles. Il est encore difficile de déterminer pour quelles raisons ces positions sonores pourraient influencer l'occurrence de l'IVIS. Une possibilité est que ces positions présentent des zones d'incertitude au niveau du signal visuel. Dans le but d'explorer cette hypothèse, nous avons calculé, pour chaque position temporelle, la probabilité qu'un disque soit présenté (voir Figure 7). En additionnant les probabilités pour toutes les conditions visuelles (2, 3, 4, 5 ou 6 flashes présentés), on obtient une plus grande probabilité que les disques soient présentés au début et à la fin de la stimulation plutôt qu'à toute autre position temporelle. De plus, les positions où il est plus probable de percevoir un disque

correspondent aux positions où les sons influencent maximalement la perception visuelle.

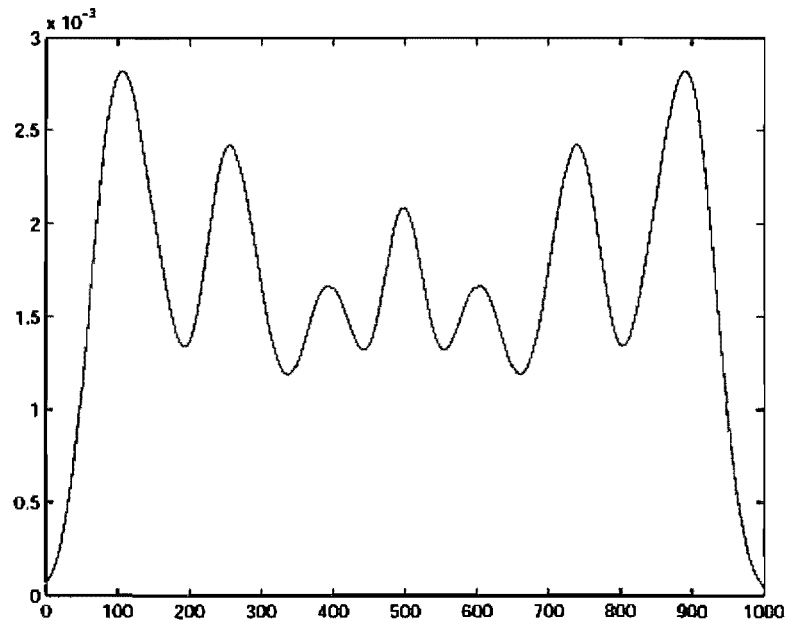


Figure 7. Pour toutes les conditions visuelles confondues, la probabilité qu'un disque soit présenté en fonction du temps total de la stimulation (une seconde). Les positions temporelles où il est le plus probable de percevoir un disque correspondent au tout début et à la fin d'un essai.

En supposant qu'il existe une association « superstitieuse » entre un disque et un son, la plus grande probabilité d'occurrence d'un disque au début et à la fin de la stimulation pourrait expliquer pourquoi ce sont ces positions qui ressortent comme étant importantes au niveau de la stimulation auditive. Cette hypothèse appuie l'idée, telle que mentionnée dans l'introduction, d'un modèle bayésien de l'intégration de l'information semblable à celui présenté par Shams et al. (2005). Dans le modèle de Shams et al., l'incertitude du signal visuel est traduite par une « likelihood »

incertaine, alors que les « priors » représentent la tendance qu'ont les sujets à combiner les deux modalités inconsciemment. Dans notre cas, ces deux éléments sont définis de façon légèrement différente. Ici, les « priors » deviennent la probabilité qu'un disque apparaisse à différents moments, toutes conditions confondues (voir Figure 8), alors que la « likelihood » représente la probabilité de détecter un flash. En supposant que la « likelihood » est toujours la même au cours de la stimulation, les « priors » expliqueraient assez bien nos résultats.

Modifications possibles

Afin d'améliorer la qualité de l'étude et dans le but d'explorer d'autres hypothèses, certaines modifications pourraient être apportées à notre paradigme expérimental. Tout d'abord, afin de s'assurer que les résultats obtenus représentent bien les conditions spécifiques de l'effet illusoire et non la conséquence d'une difficulté à effectuer la tâche demandée, il serait important de tester explicitement combien de disques peuvent être perçus par les participants. Dans le cas où ceci montrerait qu'il est difficile de compter six disques, l'absence de résultats significatifs pour cette condition pourrait s'expliquer par une simple difficulté à compter le nombre de disques et cette condition pourrait être éliminée dans des expériences futures. En faisant passer la tâche à un plus grand nombre de participants avec un plus petit nombre de conditions visuelles, il est possible que d'autres positions sonores ressortent dans les résultats et que le délai entre la présentation d'un disque et d'un son apparaisse un facteur important dans l'occurrence de l'illusion. Il est aussi possible qu'il y ait des différences individuelles dans

l'importance des délais son-disque. Ceci pourrait peut-être expliquer la variabilité d'occurrence de l'illusion d'un participant à un autre lorsque celle-ci est présentée suivant les paramètres de Shams et al. (2000). Par exemple, les paramètres originaux de l'IVIS pourraient être optimaux uniquement pour certains participants et, ainsi, provoquer l'illusion principalement chez ces participants, alors que d'autres paramètres temporels pourraient provoquer l'illusion chez d'autres participants. Il serait donc intéressant de faire passer la tâche de Shams et al. (2000) ainsi que notre tâche (où les délais entre les sons et les disques sont aléatoires) à un même groupe de participants pour vérifier si de telles différences existent. En lien avec une explication selon un modèle bayésien, il serait intéressant de vérifier si les positions sonores significativement importantes dans la perception visuelle des participants varient lorsqu'on augmente la zone d'incertitude du signal visuel, par exemple en modifiant la phase de présentation des disques. Une autre modification que l'on pourrait apporter à notre paradigme expérimental concerne la façon dont le vecteur sonore est créé. Dans l'étude présentée ci-haut, les positions sonores qu'il est possible d'échantillonner ne couvrent pas l'entièreté du vecteur temporel. Il serait bon de modifier le vecteur sonore de façon à échantillonner toutes les positions du vecteur temporel. Enfin, il serait intéressant de faire varier le délai entre la présentation des disques. Dans notre étude, la présentation de 2 disques se faisait toujours à une fréquence de 2 Hz, la présentation de 3 disques à une fréquence de 3 Hz, etc. Faire varier la position des disques de la même façon que nous avons fait varier la position des sons (i.e. aléatoirement) pourrait révéler plus d'information sur les conditions temporelles de l'IVIS.

L'expérimentation présentée plus haut ainsi que les résultats obtenus constituent un premier pas vers une meilleure compréhension de phénomènes illusoire multisensoriels tels que l'IVIS. Cependant, ces résultats ne nous permettent pas d'expliquer la façon dont le cerveau gère l'information provenant de deux modalités sensorielles distinctes, particulièrement la vision et l'audition. Dans le but de mieux comprendre ce type de phénomènes, plusieurs pistes doivent encore être explorées. Dans le cas spécifique de l'IVIS, l'étude plus approfondie des paramètres psychophysiques nous semble une étape essentielle, mais plusieurs autres pistes d'explications semblent des avenues de recherche intéressantes pouvant mener à une meilleure compréhension de l'effet illusoire. Ces pistes sont présentées dans la section qui suit.

Pistes d'explication de l'IVIS

Tel que mentionné dans l'introduction, l'objectif original du présent travail était de répliquer la deuxième expérience de Berger et al. (qui rapportait que la présence de l'IVIS pouvait influencer la performance à une tâche connexe) dans le but de vérifier notre hypothèse sur les processus menant à l'illusion. Cette hypothèse proposait que l'IVIS s'expliquerait par un effet d'augmentation de la vigilance et elle se basait en partie sur l'idée selon laquelle un état de vigilance accru pourrait permettre de « percevoir plus » ou de « percevoir plus longtemps » (e.g. Broadbent, 1958). Dans le cas de l'IVIS, nous proposons que la fréquence de l'enveloppe sonore de la stimulation auditive pourrait influencer l'état de vigilance, de sorte qu'une enveloppe sonore contenant plus de sons (ou de « beeps », i.e. un plus grand nombre d'événements sonores) augmenterait l'état de vigilance alors qu'une

enveloppe sonore contenant moins de sons diminuerait l'état de vigilance. Conséquemment, un état de vigilance accru pourrait faire en sorte que l'on « perçoive plus » de disques – dans le cas de l'illusion telle que présentée par Shams et al. (2000) – ou de taches de Gabor – dans le cas de la tâche de Berger et al., où le fait de « percevoir plus » permettrait de mieux voir l'orientation de la tache de Gabor.

Dans la vie quotidienne, plusieurs indices nous font penser que la vigilance pourrait influencer notre perception. Par exemple, les gens qui vivent des situations stressantes menant sûrement à un niveau de vigilance élevée (e.g. un accident de voiture) rapportent souvent que lors de l'événement stressant, le temps leur a paru plus long. Des chercheurs ont testé empiriquement cette idée. En 2007, Stetson, Fiesta et Eagleman ont fait subir à de braves étudiants volontaires une chute libre de 31 m qui se terminait dans un filet de sûreté. Bien que les participants savaient que la chute était sécuritaire, il n'en restait pas moins que l'idée de tomber en chute libre représentait, pour tous, un événement stressant. Avant la chute, les participants devaient effectuer une tâche de perception visuelle où ils devaient identifier des chiffres présentés sur un cadran numérique. Chaque chiffre présenté alternait avec son complément visuel (voir Figure 8a et 8b) à une certaine vitesse et plus la vitesse augmentait, plus il devenait difficile d'identifier le chiffre présenté. Le but de cette tâche était d'identifier, pour chaque participant, une vitesse seuil où le participant n'arriverait plus à identifier trois chiffres aléatoires consécutifs sur le cadran. Pendant la chute libre, chaque participant était muni d'un bracelet sur lequel était fixé un cadran identique à celui qui leur avait été présenté plus tôt (voir Figure 8c) et

pour lequel la vitesse d'alternation des chiffres avec leur opposé était fixée au seuil de perception du participant. Pendant la chute, les participants avaient pour directive de regarder le cadran et de rapporter le chiffre qui y apparaissait.

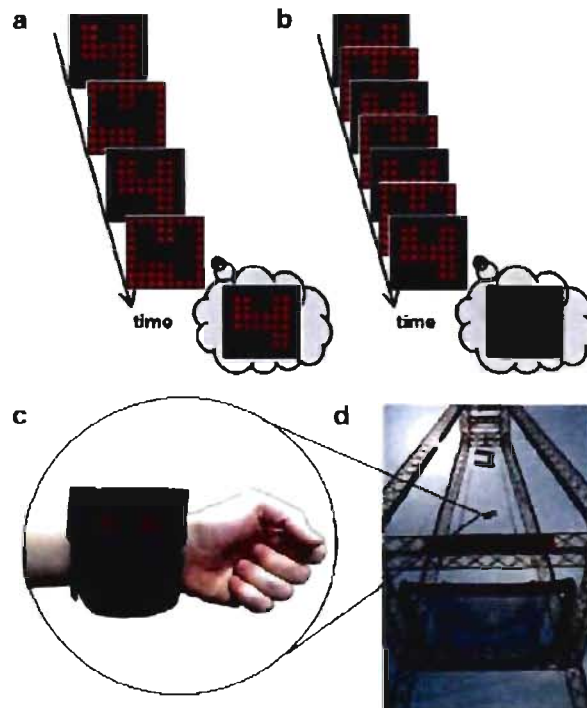


Figure 8. (a) et (b) Alternance d'un chiffre et de son complément visuel à une vitesse lente et à une vitesse rapide. (c) Lors de la chute, les chiffres sont présentés sur un bracelet auquel est fixé un cadran numérique. (d) Les participants sont relâchés au haut de la « Suspended Catch Air Device » et font une chute libre de 31 m. Tiré de Stetson et al. (2007).

L'hypothèse de Stetson et al. était que si l'événement stressant entraînait un effet de « ralentissement du temps » (i.e., un peu comme les vidéos filmés à 500 Hz qui ont permis aux télé-spectateurs de percevoir les moindres détails des mouvements des athlètes aux Jeux Olympiques de Pékin en 2008), les participants

auraient plus de facilité à percevoir le chiffre présenté sur leur cadran lors de la chute. Dans le cas de la tâche de Berger et al., une hypothèse très similaire à celle-ci était proposée, soit que la présence de l'illusion permettrait de voir la stimulation visuelle plus longtemps et ainsi en voir les moindres détails avec plus de facilité. Dans le cas de la chute, bien que lorsqu'on demandait aux participants d'estimer la durée de la chute qu'ils avaient subie, ils rapportaient systématiquement une plus grande durée que la durée réelle, les résultats au niveau de la perception visuelle ne démontraient pas d'effet d'amélioration de la perception du cadran lors de la chute. Les auteurs en ont conclu que l'idée de « ralentissement du temps » dans les moments stressants était liée au souvenir du moment stressant et non à une réelle perception temporelle accrue. Cependant, lorsque des caméras de télévision sont venues filmer les lieux de l'expérimentation et que l'expérience fut répétée, il semble que les choses se seraient passées autrement et que certains participants auraient mieux perçu les chiffres de leur cadran lors de la chute. Il semble donc que d'autres études devraient être faites pour mieux comprendre quelle est l'influence de la situation stressante de chute libre sur la perception visuelle.

Bien que nous n'ayons pas réussi à la vérifier, l'hypothèse d'une augmentation de la vigilance reste une explication possible de l'IVIS. Il semble que la question des biais attentionnels ou motivationnels fut très peu explorée par les chercheurs qui ont travaillé sur l'IVIS. Shams et al. (2002) ont mis de côté ces hypothèses, mais, tel que mentionné dans l'introduction, leurs arguments pour écarter de tels biais sont faibles et il est encore possible de croire que l'IVIS pourrait s'expliquer par une variation de la vigilance. Malgré que l'étude de Stetson et al. ne

conclue pas à une influence directe d'un événement stressant sur la perception temporelle, d'autres études suggèrent qu'un état émotionnel induit par de la musique pourrait influencer la performance à une tâche visuelle (Jefferies, Smilek, Eich & Enns, 2008 ; Olivers & Nieuwenhuis, 2005). Les conclusions de ces études suggèrent que l'interaction entre deux modalités sensorielles comme la vision et l'audition est complexe. Shams et al. (2002) ont proposé que l'IVIS est un effet perceptuel qui ne reposerait pas sur des biais cognitifs ou attentionnels. Cependant, puisque l'interaction entre la vision et l'audition semble reposer sur plusieurs aspects cognitifs, il apparaît fort probable que des aspects attentionnels ou motivationnels, non explorés à ce jour, entrent en jeu dans la perception de l'IVIS.

Une autre piste qui pourrait être explorée dans le but d'expliquer l'IVIS concerne la façon dont le cerveau prélève l'information provenant d'une ou de plusieurs modalités. Certaines études suggèrent que le cerveau échantillonnerait l'information de façon continue alors que d'autres vont dans le sens d'un échantillonnage discret de l'information provenant des différentes modalités sensorielles. Cette question fondamentale concernant l'échantillonnage de l'information a des incidences sur l'explication de tous les phénomènes perceptifs, y compris l'IVIS. Un échantillonnage discret de l'information pourrait suggérer une explication pour l'IVIS, dans la mesure où il serait possible que, lorsque deux stimulations visuelles sont présentées dans un court délai, le décalage temporel du cycle d'échantillonnage du système visuel permette de percevoir la même stimulation deux fois plutôt qu'une. De plus, il est possible que la fréquence de présentation de la stimulation sonore totale influence la vitesse d'échantillonnage

visuel, de sorte qu'une stimulation sonore de plus haute fréquence pourrait entraîner une vitesse d'échantillonnage visuel plus rapide. Une fois de plus, un échantillonnage visuel rapide pourrait permettre de percevoir la même stimulation visuelle deux fois plutôt qu'une.

Conclusion

L'étude de ces pistes d'explication de l'IVIS pourrait, certes, nous aider à mieux comprendre le phénomène illusoire, mais elle nous permettrait aussi de mieux comprendre comment interagissent les modalités auditive et visuelle. Les situations dans lesquelles nous devons intégrer l'information provenant de deux modalités sensorielles différentes ne sont pas rares. Conséquemment, la problématique de l'intégration de l'information provenant de différentes modalités sensorielles est particulièrement intéressante. Il est fascinant de réaliser à quel point le cerveau est habile pour discriminer l'information multisensorielle provenant d'une même source de l'information multisensorielle provenant de sources distinctes. Sur quoi se base-t-il pour accomplir une telle tâche? Dans le cas de l'IVIS, nous avons vérifié si des contraintes temporelles pourraient régir ce phénomène. Nos résultats montrent qu'un son présenté au début et à la fin de la stimulation sonore influence l'occurrence de l'illusion, mais que le délai entre la présentation d'un son et la présentation d'un disque n'est pas un facteur influençant la perception illusoire d'un disque. Cependant, cette hypothèse ne peut pas être définitivement exclue puisque d'autres analyses restent à faire et une modification de certains paramètres de notre paradigme expérimental pourraient mener à d'autres conclusions. Les phénomènes illusoires multisensoriels spécifiques tels que l'IVIS méritent d'être étudiés en

profondeur, autant au niveau des paramètres psychophysiques qui les régissent qu'au niveau des biais cognitifs qui pourraient les influencer, puisque ce type de phénomènes peut nous fournir de précieux renseignements sur la façon dont le cerveau gère l'information provenant de différentes modalités sensorielles.

RÉFÉRENCES

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, 14(3), 257-262.
- Andersen, T., Tiippana, K., Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21, 301-308.
- Andersen, T., Tiippana, K., Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience Letters*, 380, 155-160.
- Arden, G., Wolf, J., Messiter, C. (2003). Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. *Vision Research*, 43, 2469-2478.
- Berger, T., Martelli, M., Pelli, D. (2003). Flicker flutter: Is an illusory event as good as the real thing? *Journal of Vision*, 3, 406-412.
- Bhattacharya, J., Shams, L., Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *NeuroReport*, 13(14), 1727-1730.

- Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *J Vis*, 6(5), 554-564.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Burr, D., Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research*, 155, 243-258.
- Caclin, A., Soto-Faraco, S., Kingstone, A., Spence, C. (2002). Tactile "capture" of audition. *Perception and Psychophysics*, 64, 616-630.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci*, 8(4), 162-169.
- Fendrich, R., Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics*, 63(4), 719-725.
- Gebhard, J., Mowbray, G. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology*, 72(4), 521-529.

- Guttman, S., Gilroy, L., Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science*, 16(3), 228-235.
- Howard, I.P., Templeton, W.B. (1966) *Human Spatial Orientation*. New York: Wiley.
- Jefferies, L. N., Smilek, D., Eich, E., Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychological Science*, 19(3), 290-295.
- Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Curr Biol*, 14(23), R990-991.
- Mateef, S., Hohnsbein, J., Noack, T. (1985). Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception*, 14, 721-727.
- McCormick, D., Mamassian, P. (2008). What does the illusory-flash look like? *Vision Research*, 48, 63-69.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- Mishra, J., Martinez, A., Sejnowski, T., Hillyard, S. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120-4131.

- Olivers, C. N. L., Nieuwenhuis, S. (2005). The beneficial effects of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science*, 16, 265-269.
- Recanzone, G.H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89, 1078-1093.
- Rock, I., Victor, J. (1964) Vision and touch: an experimentally created conflict between the two senses. *Science* 143, 594-596.
- Schwartz, J.-L., Robert-Ribes, J., Escudier, P. (1998) Ten years after Summerfield: a taxonomy of models for audio-visual fusion in speech perception, in: D. Burnham (Ed.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, Psychology Press, Hove, U.K, 1998, pp. 85-108.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308.
- Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: an MEG study. *Neurosci Lett*, 378(2), 76-81.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, 408(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Res Cogn Brain Res*, 14(1), 147-152.

Shams, L., Kamitani Y., Shimojo, S. (2004) Modulations of visual perception by sound. In Gemma Calvert, Charles Spence, Barry Stein (Eds.), *The Handbook of Multisensory Processes* (pp.27-33). London: MIT Press.

Shams, L., Kamitani, Y., Thompson, S., Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *NeuroReport*, 12(17), 3849-3852.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17), 1923-1927.

Shimojo, S., Scheier, C. R., Nijhawan, R., Shams, L., Kamitani, Y. and Watanabe, K. (2001). Beyond perceptual modality: auditory effects on visual perception. *Journal of Acoustical Science and Technology*, 22 (2), 61-67.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145, 1328-1330.

- Spence, C., Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, 13, 519-521.
- Stetson, C., Fiesta, M., Eagleman, D. (2007). Does time really slow down during a frightening event? *PLoS ONE*, 2(12), e1295.
- Tallon-Baudry, C., Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4), 151-162.
- Verghese, P., Stone, L. S. (1995). Combining speed information across space. *Vision Research*, 35, 2811-2823.
- Verghese, P., Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, 381, 161-163.
- Verghese, P., Stone, L. S. (1997). Spatial layout affects speed discrimination. *Vision Research*, 37, 397-406.
- Violentyev, A., Shimojo, S., & Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, 16(10), 1107-1110.
- Walker, J. T., Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, 7(6), 1327-1339.

- Watkins, S., Shams, L., Tanaka, S., Haynes, J-D., Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31, 1247-1256.
- Welch, R. B., Duttonhurt, L. D., Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics*, 39, 294-300.
- Welch, R. B., Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, 88, 638-667.