

Université de Poitiers

Faculté Lettres et Langues

Département IME



Mémoire

Master 2



**LA CONTIGUÏTE TEMPORELLE DANS LES PRESENTATIONS
MULTIMEDIA POUR L'APPRENTISSAGE**

Les effets de la désynchronisation entre le texte et la narration

Stéphane Muller

Directeur de mémoire : Ludovic Le Bigot

Membres du jury : Jean-François Cerisier, Ludovic Le Bigot, Jacques Viens

22 septembre 2011

RESUME

L'objectif de cette étude est de déterminer les effets d'un décalage temporel entre la présentation d'un texte écrit et son équivalent audio sur la mémorisation et la compréhension, dans un objet d'apprentissage multimédia. Nous avons émis l'hypothèse qu'une faible avance du texte par rapport aux commentaires audio produirait de meilleurs résultats lors des tests. Les 30 participants ayant pris part à notre expérimentation ont visionné une série de cinq vidéos présentant des procédures de premiers secours accompagnées d'un texte écrit et d'un commentaire audio équivalent. Les vidéos étaient déclinées en cinq synchronisations différentes : 0, 500, 1000, 1500 et 2000 ms (avance du texte sur la narration). Il a ensuite été demandé aux participants de répondre à des questionnaires de mémorisation et de compréhension. L'analyse des résultats n'a révélé aucun effet significatif de la désynchronisation.

Mots-clés : apprentissage multimédia, contiguïté temporelle, partage attentionnel.

ABSTRACT

The objective of this study was to determine the effects of a time shift between written text and its audio equivalent on retention and comprehension. We hypothesized that a small text leading would show better results in tests. The 30 participants who took part in our experiment had to watch a series of five videos showing first-aid procedures. These videos were combined with written text and narration which were synchronized in five different ways : 0, 500, 1000, 1500 and 2000 ms (text leading). The participants were then asked to answer retention and comprehension tests. The result analysis did not reveal any significant effect of the time shift on comprehension and retention.

Keywords : multimedia learning, temporal contiguity, split attention effect.

SOMMAIRE

| | |
|---|-----------|
| RESUME | 3 |
| ABSTRACT | 4 |
| SOMMAIRE | 5 |
| REMERCIEMENTS | 7 |
| INTRODUCTION | 8 |
| PARTIE THEORIQUE | 11 |
| CADRE THEORIQUE | 12 |
| I. LES BASES DES APPRENTISSAGES MULTIMEDIA | 12 |
| 1. LA MEMOIRE DE TRAVAIL..... | 12 |
| a. <i>La boucle phonologique</i> | 13 |
| b. <i>Le calepin visuo-spatial</i> | 14 |
| c. <i>L'administrateur central</i> | 15 |
| 2. LE PARTAGE ATTENTIONNEL..... | 15 |
| II. LES PRINCIPES MULTIMEDIA | 17 |
| 1. L'EFFET DE MODALITE..... | 17 |
| 2. LE PRINCIPE DE REDONDANCE..... | 18 |
| 3. UTILISATION DE LA VIDEO ET DES ANIMATIONS (PRINCIPE MULTIMEDIA)..... | 20 |
| III. LE PRINCIPE DE CONTIGUÏTE | 21 |
| 1. LA CONTIGUÏTE SPATIALE..... | 22 |
| 2. LA CONTIGUÏTE TEMPORELLE..... | 23 |
| IV. LA COMPOSANTE TEMPORELLE | 25 |
| 1. FENETRE TEMPORELLE D'INTEGRATION AUDIOVISUELLE..... | 25 |
| 2. PERCEPTION DE LA SIMULTANEITE..... | 26 |
| 3. L'EFFET COLAVITA..... | 27 |
| V. SYNTHESE ET HYPOTHESES | 27 |
| PARTIE EXPERIMENTALE | 29 |
| METHODE | 30 |
| I. PARTICIPANTS | 30 |
| II. MATERIEL | 30 |
| 1. LES VIDEOS..... | 30 |
| 2. LES PROCEDURES..... | 32 |

| | |
|--|-----------|
| 3. LES COMMENTAIRES AUDIO | 32 |
| 4. LES QUESTIONNAIRES | 33 |
| <i>a. Questionnaire de pré-test</i> | 33 |
| <i>b. Questionnaire de mémorisation (variable dépendante 1)</i> | 33 |
| <i>c. Questionnaire de compréhension (variable dépendante 2)</i> | 34 |
| <i>d. Questionnaire d'informations générales</i> | 34 |
| III. CONTRE-BALANCEMENT | 34 |
| IV. PROCEDURE | 35 |
| 1. ACCUEIL DES PARTICIPANTS | 35 |
| 2. QUESTIONNAIRES | 36 |
| 3. FIN DE L'EXPERIMENTATION | 36 |
| V. VARIABLES DEPENDANTES | 36 |
| 4. MEMORISATION | 36 |
| 5. COMPREHENSION | 36 |
| RESULTATS | 37 |
| DISCUSSION | 41 |
| BIBLIOGRAPHIE | 45 |
| ANNEXES | 49 |

REMERCIEMENTS

J'aimerais avant tout remercier Ludovic Le Bigot, mon directeur de mémoire, pour ses nombreux conseils, ses explications et son aide précieuse.

Un grand merci à Chloé pour ses encouragements et ses bons conseils.

Ma reconnaissance va également à Laurent pour nos longues discussions, nos échanges et son aide dans la réalisation des vidéos.

Merci à Didier d'avoir sauvé Laurent cinq fois d'une mort certaine et devant la caméra en plus.

Merci à Caroline d'avoir prêté sa belle voix à la narration des vidéos.

Merci à Emilie, ma fidèle relectrice à qui rien (ou presque) n'échappe.

Merci à Hélène d'avoir imprimé ce mémoire et de l'avoir livré au jury.

Merci aussi à Fanny pour son écoute et sa patience.

Et enfin, je tiens à remercier tous mes cobayes de leur participation (même s'ils n'ont pas produit les résultats escomptés).

INTRODUCTION

Les enseignants ont actuellement de plus en plus recours aux présentations multimédia pendant leurs cours. Le plus souvent, ces présentations prennent la forme de diaporamas PowerPoint et sont en général très appréciées des étudiants (Savoy *et al.*, 2008). Malgré l'attrait très net pour ce genre d'outils dans un cadre éducatif, les bénéfices pour l'apprentissage sont moins évidents. Dans leur étude, Savoy *et al.* ont comparé les effets de deux modalités de présentation d'un même cours (de contenu identique) donné par un même professeur. Dans un cas, le professeur donnait son cours de manière traditionnelle, sur un tableau noir, avec de la craie (« *chalk-and-talk* »), tandis que dans l'autre, il avait recours à une présentation PowerPoint. Bien que les étudiants aient dans l'ensemble préféré la seconde modalité, ils ont retenu 15% de moins des informations orales présentées par le professeur par rapport au cours traditionnel.

En plus d'être motivant, le support informatique offre bien de nouvelles possibilités. Il est maintenant possible, et même relativement aisé, de combiner images, vidéos, sons et textes pour illustrer une présentation. Que faire alors ? Que faut-il privilégier ? La motivation des étudiants ou ce qui leur permet de mieux apprendre ?

La motivation intrinsèque est un facteur bénéfique très important dans l'apprentissage (Lin *et al.*, 2003). Ce serait donc une erreur de ne pas en tenir compte dans la construction d'un cours. C'est pourquoi un début de solution consiste à utiliser des diaporamas qui soient efficaces pour l'apprentissage. Pour cela, il est notamment possible de se référer aux principes de base de l'apprentissage multimédia (Mayer, 2001) ou de suivre les préconisations de Jamet (2006) sur la conception de diaporamas plus efficaces.

C'est malheureusement rarement le cas. Les professeurs ont tendance à faire appel à leur intuition quand ils créent leurs présentations, ce qui, comme l'étude de Savoy *et al.* (2008) nous l'a montré précédemment, n'aboutit pas toujours à un résultat optimal. La facilité avec laquelle il est actuellement possible de réaliser un diaporama ne fait que contribuer à ce manque d'efficacité.

Le travers le plus courant que l'on retrouve dans les présentations est l'affichage de la totalité du cours à l'écrit sur l'écran. En toute logique cette pratique a ses avantages. Elle rassure l'orateur, qui n'a alors plus qu'à lire ce qui est affiché, et les étudiants ont accès à l'information de deux manières : à l'oral et à l'écrit. Après tout, la répétition est connue depuis longtemps comme étant bénéfique à l'apprentissage. Dans cette situation ce n'est malheureusement pas le cas et ceci peut s'expliquer à l'aide du modèle de la mémoire de travail et du principe de redondance, qui font l'objet de la première partie de cette étude.

Avec le développement de l'enseignement à distance, on retrouve le même genre de problèmes au niveau de la conception. Seulement dans ce cas, l'apprenant se retrouve seul face au support d'apprentissage. Afin d'éviter cela, il est de plus en plus commun de mettre à disposition des apprenants des enregistrements vidéo de conférences ou de cours, accompagnés de leur présentation PowerPoint. Ce type de présentation, de part la manière dont elles sont construites, soulève un certain nombre de questions. La simple juxtaposition de la vidéo du présentateur (à gauche généralement) et de son diaporama (à droite) est-elle vraiment idéale ? Quels sont les paramètres sur lesquels il est possible d'agir pour favoriser l'apprentissage à l'aide de ce type de contenu ? Ne serait-il pas efficace par exemple de désynchroniser le discours du présentateur avec l'affichage des éléments du diaporama ? C'est ce à quoi nous nous sommes intéressés dans cette étude. Et plus particulièrement encore : « Que se passerait-il si nous faisons varier la synchronisation entre l'affichage du texte et la narration ? ». C'est à cette question que nous avons tenté de répondre par le biais de l'expérimentation décrite dans la seconde partie.

Nous avons émis l'hypothèse qu'une légère avance de la présentation du texte par rapport à celle de la narration aurait un effet positif sur la rétention d'information et la compréhension du sujet. La troisième partie présente les résultats des participants à ces tests. La dernière partie est consacrée à l'interprétation des résultats et leurs implications par rapport à notre hypothèse de départ.

PARTIE THEORIQUE

CADRE THEORIQUE

I. Les bases des apprentissages multimédia

Lorsqu'un enseignant explique le phénomène de formation des éclairs à sa classe, ou qu'un élève ouvre son livre et lit une explication similaire, l'information n'est alors présentée que d'une seule manière, sous forme de mots. Si l'enseignant utilise un support visuel, tel que le tableau ou une présentation via un support électronique, en complément de son exposé oral, la situation d'apprentissage devient alors multimédia car l'apprenant traite à la fois des mots et des images.

Mayer (2003) définit un objet d'apprentissage multimédia comme étant une présentation contenant des mots et des images, et dont le but est de favoriser l'apprentissage. Les mots peuvent être présentés à l'écrit ou à l'oral, et les images peuvent être de plusieurs natures : illustrations statiques, cartes, photos, animations et/ou vidéos. Pour comprendre comment se fait l'apprentissage au travers de ces deux médias, nous allons nous intéresser aux mécanismes cognitifs qui interviennent lorsque les apprenants y sont exposés. Nous décrirons particulièrement deux d'entre eux : la mémoire de travail et le partage attentionnel. Le premier est le modèle cognitif sur lequel se base l'ensemble de notre réflexion : il décrit comment sont traitées les informations, selon qu'elles soient de nature verbales ou visuelle. Le second permet de comprendre la manière dont les apprenants traitent le flux d'informations à leur disposition et ce sur quoi leur attention va se porter.

1. La mémoire de travail

L'acquisition de connaissances est directement dépendante du système permettant de retenir et de traiter les informations : la mémoire. Les premiers travaux sur l'apprentissage reposent sur la distinction entre mémoire à court terme et mémoire à long terme (Atkinson & Shiffrin, 1968). La mémoire à court terme permet la rétention d'un nombre minimal d'informations sur une période brève. La mémoire à long terme est quant à elle supposée ne pas être limitée en termes de capacité et de durée de rétention. A cela vient également s'ajouter la mémoire sensorielle, qui stocke les

informations provenant de nos sens. Cette mémoire est effective pendant un temps très court, de l'ordre de quelques millisecondes. Dans le « modèle modal » d'Atkinson et Shiffrin, les informations sont manipulées en mémoire à court terme et récupérées en mémoire à long terme, le cas échéant. Baddeley et Hitch (1974) ont proposé un modèle de mémoire qui intègre et adapte les modèles de mémoire à court terme et de mémoire à long terme. Selon ce modèle, la mémoire de travail fonctionne comme un système capable à la fois de retenir et de manipuler les informations en exécutant des tâches complexes comme le raisonnement, la compréhension ou l'apprentissage.

Le modèle de mémoire de travail décrit une mémoire à capacité limitée, composée de deux sous-systèmes (ou systèmes esclaves) et d'un administrateur central (**Figure 1**). Le premier sous-système est la boucle phonologique, qui traite les informations verbales. Le second sous-système est le calepin visuo-spatial, qui traite les informations visuelles et spatiales. Ces deux sous-systèmes sont gérés par l'administrateur central, qui répartit les ressources attentionnelles disponibles et assure la liaison avec la mémoire à long terme.

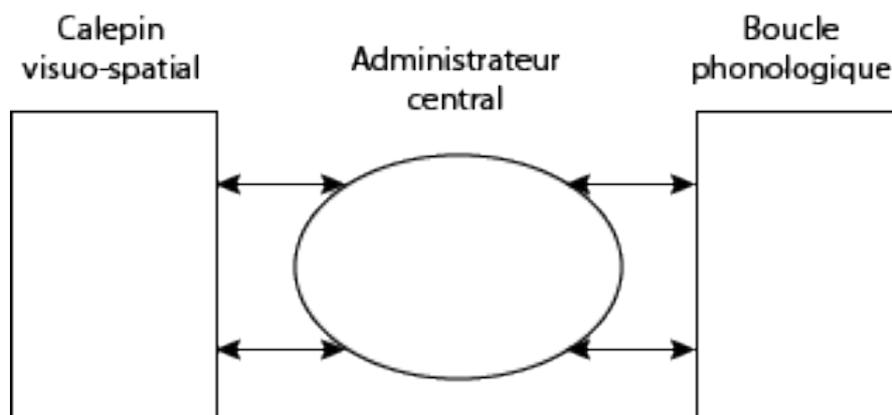


Figure 1 - Schéma de la mémoire de travail (Baddeley & Hitch, 1974)

a. La boucle phonologique

La boucle phonologique traite toutes les informations de natures phonologique et auditive. Elle est composée de deux éléments : une mémoire phonologique tampon et une boucle articulatoire.

La mémoire phonologique a été mise en évidence grâce à l'effet de similarité acoustique par Conrad (1964) en demandant à des participants de rappeler une suite de lettres présentées visuellement. Les erreurs relevées avaient tendance à présenter des sonorités similaires, par exemple

rappeler un « v » à la place d'un « b ». Il interpréta ces résultats comme un recodage phonologique de l'information visuelle.

D'autres expériences portant sur la sonorité et le sens des mots ont confirmé cette partie du modèle de la mémoire de travail. Ainsi dans l'une d'elles, les participants devaient se rappeler des suites de mots de même sonorité (blanc, paon, banc, banque, rang...) ou des mots de sens similaire (grand, large, gros, immense...), tous présentés auditivement. Le taux de rappel des mots de même sonorité était de 20 %, tandis que les mots de même sens (mais de sonorités différentes) étaient rappelés à 70-80 %. Le faible taux de mémorisation des mots de sonorité similaire montre que le codage qui a lieu en mémoire est de type sonore et non sémantique (lié au sens des mots). En effet, le délai de rappel entraîne d'abord une détérioration des informations stockées. C'est ensuite la similarité de ces informations qui sème la confusion chez les participants lors du rappel. Ceci n'est pas le cas lorsque les mots sont de sonorités différentes. Si le codage avait été d'ordre sémantique, on aurait observé un taux de rappel équivalent dans les deux conditions.

L'existence de la boucle articulatoire a été démontrée en faisant varier la longueur des mots de listes à mémoriser (Cowan *et al.*, 1992). Plus les mots étaient longs, plus il était difficile aux participants de les rappeler. Ceci peut s'expliquer par le fait que, pour être retenus après avoir été lus (ou entendus), les mots doivent être répétés intérieurement dans la boucle articulatoire. Les longs mots sont répétés plus lentement et mettent plus de temps à être rappelés, ce qui entraîne un plus grand nombre d'erreurs. Les performances à ce type de rappel déclinent significativement lorsque l'on empêche les participants de réaliser cette répétition intérieure en leur demandant d'articuler « bla bla bla » en même temps que la liste leur est présentée, et ce, quelque soit la longueur des mots. La boucle articulatoire permet donc de maintenir une certaine quantité d'informations verbales en mémoire de travail.

b. Le calepin visuo-spatial

Le calepin visuo-spatial traite toutes les informations visuelles et spatiales. Logie (1995) décrit le système comme étant composé de deux sous-systèmes : le cache visuel et le scribe interne. Le cache visuel se détériore rapidement et est sensible aux interférences provenant des nouvelles informations entrantes. Il permet de stocker provisoirement les informations de nature visuelle, telles que les formes et les couleurs. Le scribe interne quant à lui permet de traiter les informations dynamiques et spatiales, de planifier des mouvements. Il peut également être utilisé pour répéter des

représentations activées dans le cache visuel, sur le même principe que la boucle articulatoire. C'est-à-dire qu'il est possible de répéter des informations visuelles à l'aide du scribe interne pour les maintenir en mémoire de travail.

c. L'administrateur central

L'administrateur central agit comme un système de contrôle. Il assure la coordination entre les systèmes esclaves, que sont la boucle phonologique et le calepin visuo-spatial, mais aussi le changement de tâches ou le retrait d'informations, ainsi que l'attention à porter, ou non, à chaque stimulus. L'administrateur central fait également le lien avec la mémoire à long terme et établit les connexions entre les informations stockées dans les deux autres sous-systèmes. Il va par exemple permettre de lier un mot à sa représentation visuelle.

2. *Le partage attentionnel*

La mémoire de travail et chacun de ses composants ont des réservoirs de capacité limitée. Il est ainsi logique que, quand un nombre trop important d'informations est présenté à un apprenant, le traitement de celui-ci soit plus difficile, voire impossible. Lorsque l'un des systèmes est saturé, plus aucune information supplémentaire ne peut y être stockée et l'apprenant peut ainsi passer à côté d'un grand nombre d'informations. C'est le cas par exemple lorsque plusieurs informations de type visuel sont affichées à l'écran. L'apprenant doit alors analyser l'ensemble puis focaliser son attention sur l'un ou sur l'autre élément. Ce phénomène a été baptisé partage attentionnel par Mayer.

Mayer et Moreno (1998) ont fait l'hypothèse que la présentation d'une animation accompagnée de texte sous forme orale produirait de meilleurs résultats que la même animation accompagnée de texte écrit. Ils ont demandé à des étudiants de visionner une animation portant sur la formation des éclairs. Dans le premier cas, l'animation était accompagnée d'un texte décrivant les étapes importantes du phénomène et dans le second, le texte était lu par une voix off. Les participants étaient ensuite soumis à trois tests. Pour le test de rétention ils devaient restituer le contenu de l'animation en expliquant le phénomène de formation des éclairs. Pour le test de reconnaissance ils devaient replacer un certain nombre de termes sur quatre images issues de l'animation. Enfin, pour le test de transfert il leur était demandé de répondre à quatre questions de type résolution de problème. Les résultats des trois tests ont montré un net avantage de l'animation associée aux

commentaires audio. Dans leur analyse, les auteurs attribuent cet effet négatif du texte écrit à une surcharge du canal visuel. Le texte étant affiché en même temps que l'animation, l'attention ne peut se porter que sur l'un ou sur l'autre. Lorsque l'attention est portée sur le texte, l'apprenant risque de manquer des moments clés de l'animation. Au contraire, lorsque l'apprenant se concentre sur l'animation il est susceptible de ne pas voir le texte. Dans les deux cas, il a des chances de passer à côté d'informations importantes.

Cette surcharge du canal visuel s'explique notamment par une limitation physiologique. Les yeux ont un angle de vision très réduit de 1 à 2 °, ce qui correspond à une surface circulaire de 2,5 cm de diamètre pour un écran d'ordinateur placé à 80 cm. La question devient alors : entre une animation et du texte, vers quoi l'attention va-t-elle se diriger ? Il serait logique de penser que ce sont les images en mouvement d'une animation qui capteraient le plus le regard, mais en réalité c'est tout le contraire.

Ce phénomène est démontré, entre autres, par une étude de Schmidt-Weigand *et al.* (2010). Le matériel utilisé expliquait la formation des éclairs à l'aide d'une animation. Dans une première expérience les participants étaient divisés en deux groupes : les uns visionnaient l'animation complétée par des commentaires audio, tandis que les autres avaient droit aux mêmes commentaires écrits. Chaque groupe était alors encore une fois partagé en trois sous-groupes correspondant chacun à une vitesse de lecture différente (lente, modérée, rapide). L'étude des mouvements oculaires des participants a permis de conclure que, dans la modalité « animation et texte », les participants passaient plus de temps à lire le texte qu'à regarder l'animation. De plus, à chaque changement de scène, c'est le texte qui était visualisé en premier, au détriment de l'animation. Une fois le texte lu et s'ils en avaient le temps, les participants alternaient leur attention entre le texte et l'animation afin de construire des liens entre les deux. Dans la seconde expérience les participants avaient le contrôle de la vitesse de lecture. Dans cette condition également les auteurs constatent un guidage de l'attention par le texte. Au niveau des résultats aux tests de rétention, compréhension et transfert, la mémorisation est supérieure lorsque les participants ont le contrôle de la vitesse. Cette expérience montre que lorsque le canal visuel est confronté à plusieurs types d'informations, son attention est largement guidée par le texte.

L'effet de partage attentionnel indique qu'il est difficile pour un apprenant de traiter simultanément plusieurs informations de natures différentes, en l'occurrence des images en mouvement et du texte écrit, sur un même canal. Les deux informations se font concurrence et

l'attention visuelle est alors partagée, entraînant une perte potentielle des informations selon la source sur laquelle elle se porte. En cas de présence de texte, la perte d'information se fera principalement au détriment de l'image.

II. Les principes multimédia

Du modèle de la mémoire de travail et de l'effet de partage attentionnel, que nous venons d'évoquer, découle un certain nombre de principes qui vont nous guider dans la conception de présentations multimédia favorisant l'apprentissage. Dans leur article, Mayer et Moreno (2002) en citent cinq. Ces principes permettent de comprendre comment et de quelle manière arranger les médias (images, animations, textes, sons...) entre eux, ainsi que les effets que ces différentes présentations produisent sur les apprenants. Sur ces cinq principes, quatre sont pertinents à notre étude : l'effet de modalité, le principe de redondance, le principe multimédia et, plus particulièrement encore, le principe de contiguïté (spatiale et temporelle). Ce sont eux qui ont servi de base à l'élaboration de notre protocole expérimental.

1. L'effet de modalité

Le modèle de la mémoire de travail postule que les informations qu'un individu reçoit sont traitées par deux sous-composants : le calepin visuo-spatial et la boucle phonologique. Ces deux sous-systèmes étant indépendants, la quantité d'informations pouvant être traitée en mémoire de travail est déterminée, au moins en partie, par la nature de la présentation. La modalité fait référence au canal de traitement de l'information auquel l'apprenant fait appel : le canal visuel ou auditif. Lorsqu'un texte est présenté à l'écrit il sera d'abord traité par le canal visuel, tandis que s'il est présenté à l'oral il sera pris en charge par le canal auditif. Il est donc possible d'augmenter la capacité de la mémoire de travail en combinant les deux modalités. Ceci rejoint également la notion de partage attentionnel décrite précédemment (Mayer & Moreno, 1998). Il est plus facile de répartir son attention entre deux informations de natures différentes qu'entre deux de même nature.

Il est important de bien faire la distinction entre le canal de traitement sensoriel et le type de mémoire dans lequel l'information sera stockée. Un texte écrit est d'abord traité par le canal visuel, mais est ensuite stocké dans la boucle phonologique puisqu'il s'agit d'information verbale.

L'avantage du mélange des deux modalités ne se limite pas à une meilleure répartition de l'attention, il génère également une meilleure compréhension en permettant le traitement en parallèle d'informations graphiques et textuelles. C'est le principe multimédia (Moreno & Mayer, 1999). L'une des expériences de cette étude consiste à présenter séquentiellement le texte, puis l'animation (formation des éclairs). Dans un premier cas le texte est présenté à l'écrit, dans le deuxième il est présenté à l'oral. Les participants ayant eu droit à la version orale du texte ont obtenu de meilleurs résultats que l'autre groupe, suggérant ainsi qu'il est plus facile de garder en mémoire des informations verbales auditives puis de les combiner avec des informations visuelles. On peut donc conclure de cette expérience que la combinaison d'informations verbales audio avec du matériel non verbal visuel est plus efficace que la combinaison d'informations verbales visuelles avec du matériel non verbal visuel. C'est le principe de modalité.

En conclusion, les apprenants apprennent mieux lorsque sont combinées narration et animation plutôt que animation et texte à l'écran. Plus généralement encore : l'apprentissage est plus efficace lorsque les informations verbales sont présentées à l'oral plutôt qu'à l'écrit.

2. Le principe de redondance

L'apprentissage multimédia, par définition, permet de combiner plusieurs types de média mais aussi plusieurs types de modalités (visuelle ou auditive). Jusqu'à présent les études que nous avons citées faisaient appel à deux types d'informations complémentaires, à savoir : une animation et du texte. Que se passe-t-il lorsque les informations présentées sont identiques ? Lorsqu'une même information est présentée deux fois sur un ou plusieurs canaux à la fois, il s'agit de redondance.

La redondance est généralement associée à un effet négatif sur l'apprentissage (Mayer & Moreno, 2003). Dans leur expérience, Mayer et Moreno montrent que la présentation simultanée d'une même information à l'écrit et à l'oral donne de moins bons résultats aux tests de mémorisation et de transfert. Les auteurs expliquent cela par une saturation des ressources disponibles due au traitement de deux informations verbales identiques. En effet, même si les deux informations sont tout d'abord traitées dans un canal différent de la mémoire de travail (le commentaire audio dans la boucle phonologique et le texte dans le calepin visuo-spatial), les mots présentés à l'écrit sont traités dans un second temps par la boucle phonologique, ce qui a pour effet de surcharger cette mémoire.

La présentation d'information redondante va également à l'encontre du principe de cohérence qui stipule qu'il faudrait éliminer toute information superflue (Mayer, 2001). Cet effet négatif de la redondance sur l'apprentissage a été reproduit plusieurs fois, notamment par Jamet et Le Bohec (2007). Dans leur expérience, leur document principal était constitué d'une série de diagrammes commentés sur le développement des modèles de la mémoire. Dans la première condition, les diagrammes étaient présentés tels quels, tandis que dans les deux autres conditions, les diagrammes étaient complétés par les phrases écrites du commentaire audio : dans un cas les phrases étaient affichées de manière séquentielle ; dans l'autre de manière statique. Quel que soit le type d'affichage, l'ajout de texte redondant aux diagrammes donnait toujours des résultats significativement inférieurs aux tests de rétention et de transfert.

Cependant, un certain nombre d'expériences a également montré que, dans des conditions particulières, la redondance pouvait avoir un effet bénéfique sur l'apprentissage. Dans le cadre d'une étude sur les effets de présentations auditive et visuelle sur la rétention d'informations verbales, Penney (1989) démontre que des mots présentés à la fois sous formes auditive et visuelle étaient mieux retenus que des mots présentés dans une seule de ces deux modalités.

La redondance verbale s'est également révélée bénéfique pour la compréhension dans le cas de mauvais lecteurs (Montali & Lewandowski, 1996). Cette étude comparait la compréhension de textes scientifiques présentés sous plusieurs formes (visuelle, auditive et audiovisuelle) entre des élèves (âgés de 13 à 15 ans) bons et mauvais lecteurs. Lorsque le texte était présenté de manière audiovisuelle (texte écrit et lu), les mauvais lecteurs obtenaient des résultats comparables aux bons lecteurs en condition « texte écrit ».

Des effets bénéfiques pour la rétention, mais pas pour le transfert, ont également été démontrés lors de la présentation de courtes phrases écrites résumant un commentaire audio (Mayer & Johnson, 2008). Le matériel de l'expérience était constitué d'une présentation PowerPoint en 16 diapositives sur la formation des éclairs accompagnée d'une narration. Dans une condition les diapositives ne contenaient qu'une image, dans l'autre elles contenaient une image ainsi que deux ou trois mots-clés issus de la narration. Dans leur analyse, les auteurs attribuent cet effet positif de la redondance sur la rétention à un guidage attentionnel du texte. Ils émettent l'hypothèse que les phrases de quelques mots résumant le commentaire audio guident l'attention auditive des apprenants sur les concepts clés de la narration, sans pour autant ajouter un coût de traitement supplémentaire. Cette redondance faciliterait l'étape de sélection des mots-clés, très importante lors des tests de rétention.

Pour résumer, la redondance a un effet positif sur l'apprentissage lorsque le texte est court et guide l'attention des apprenants sur les éléments clés à retenir. Elle a cependant un effet négatif lorsque la totalité de l'information est dupliquée à l'écrit.

3. *Utilisation de la vidéo et des animations (principe multimédia)*

Jusqu'à présent nous avons vu que l'utilisation d'animations était optimale en combinaison avec une narration, mais quel est son avantage par rapport à des images statiques ? Par définition une animation est une rapide succession d'images, ce qui en fait le support idéal pour représenter des mouvements ou des changements dans le temps. Une animation a également la caractéristique d'être continue et ainsi de véhiculer plus d'informations que quelques images statiques. L'apprenant n'a alors pas besoin d'inférer les changements entre deux étapes (Bétrancourt & Tversky, 2000). Nous pourrions aussi émettre l'hypothèse qu'elles sont plus attractives pour les apprenants, mais comme nous l'avons vu précédemment pour les présentations PowerPoint, une préférence pour une certaine modalité n'est pas forcément synonyme de meilleurs résultats (Savoy *et al.*, 2008).

Les résultats des expériences qui ont été menées sur ce sujet ont tendance à se contredire, c'est pourquoi nous nous baserons sur la méta-analyse¹ de Höffler et Leutner (2007). Cette étude montre tout d'abord que, de manière générale, les animations et les vidéos sont supérieures aux images statiques. Cependant, comme pour toutes les méta-analyses, un certain nombre de facteurs varient entre les études sélectionnées et c'est pourquoi il est intéressant de se pencher sur des comparaisons plus spécifiques. Les auteurs ont comparé ainsi le type d'utilisation des animations, à savoir à des fins de représentation ou de décoration. Une animation représentative illustre un phénomène dynamique, tandis qu'une animation décorative n'a aucune fonction explicative. Il s'avère que les animations représentatives sont supérieures aux images statiques et ce de manière plus importante que les animations décoratives. Ceci confirmerait les conclusions de Bétrancourt et Tversky (2000) qui préconisent l'utilisation d'animations dans le cas de l'explication de phénomènes dynamiques. Le recours à des animations décoratives augmenterait alors inutilement le coût de traitement et pourrait même, le cas échéant, être préjudiciable.

¹ Une méta-analyse est une analyse statistique sur les résultats de plusieurs expériences avec le même objet d'étude et des méthodes proches.

L'expérience d'Arguel et Jamet (2009) apporte une nouvelle dimension intéressante à l'utilisation de vidéos et d'images statiques. Leur hypothèse de départ était que l'utilisation d'une vidéo combinée à l'affichage d'images statiques clés, et issues de la vidéo, faciliterait l'apprentissage procédural. Les participants étaient divisés en trois conditions : vidéo seule, images statiques seules et vidéos et images combinées. Les vidéos utilisées montraient des procédures de premiers secours et les images statiques étaient des images clés tirées de ces vidéos. Les résultats de cette expérience montrent une supériorité de la vidéo sur les images statiques, mais de meilleurs résultats encore pour la combinaison de la vidéo avec les images statiques. Les images seules ne sont pas bénéfiques, mais associées à la vidéo elles la complètent. D'après les auteurs cela s'explique par le fait que les images compensent la nature fugace de la vidéo. Une vidéo est une rapide suite d'images, lorsque les informations véhiculées par une image sont passées il est impossible à l'apprenant de revenir dessus s'il ne l'a pas retenu en mémoire de travail. Les images permettent ainsi à l'apprenant de rappeler les informations clés de la vidéo qui sont passées.

Les vidéos et les animations sont donc particulièrement indiquées pour illustrer un phénomène dynamique, mais préjudiciables dans le cas de concepts statiques. Les images fixes quant à elles sont à utiliser en combinaison avec de la vidéo (ou de l'animation) et peuvent ainsi apporter une aide supplémentaire à l'apprenant en insistant sur les étapes clés.

Les principes multimédia que nous venons de voir permettent de choisir les types de médias qui sont le plus indiqués selon la situation, ainsi que la manière de les combiner entre eux. Le principe de contiguïté donne quant à lui des précisions sur la meilleure façon de les arranger entre eux dans l'espace et le temps.

III. Le principe de contiguïté

La contiguïté fait référence à la proximité spatiale ou temporelle des informations. Nous avons établi l'utilité des vidéos et des animations dans certains contextes et dans certaines conditions. Mais comment les utiliser en combinaison avec du texte (oral ou écrit) ?

1. *La contiguïté spatiale*

Une présentation multimédia est composée de graphismes et de texte, chacun pouvant être de natures différentes. Le principe de contiguïté spatiale va nous aider à comprendre la meilleure manière d'arranger les illustrations et les mots ensemble. Le choix se fait entre trois types de présentation : le texte et les illustrations sur des pages séparées, le texte et les illustrations séparés sur la même page ou le texte et les illustrations combinés.

En présentant sur deux pages séparées, d'abord le texte puis l'illustration, nous pourrions partir du principe que l'on expose l'apprenant deux fois à la même information sous deux formes différentes et que cela lui sera bénéfique. Le processus de répétition est depuis longtemps reconnu comme favorisant l'apprentissage. Le problème du point de vue de la théorie de l'apprentissage multimédia est que cela ne prend pas en compte la façon dont nous apprenons réellement. L'apprentissage est un processus actif au cours duquel l'apprenant essaie de donner du sens au matériel présenté (Mayer, 2001). Le fait que les informations à lier entre elles se trouvent sur deux pages différentes rend la tâche beaucoup plus difficile. L'apprenant doit tout d'abord retenir en mémoire de travail les premières informations auxquelles il est exposé, puis il doit y ajouter les informations complémentaires qui se trouvent sur la seconde page, pour finalement traiter le tout et essayer d'y trouver un sens.

La deuxième possibilité est de présenter le texte et les illustrations sur la même page mais séparément (chacun dans une zone bien distincte). Dans ce cas là nous nous heurtons à un problème familier puisqu'il s'agit du partage attentionnel. En effet, l'apprenant doit alors sans arrêt faire des allers-retours entre le texte et l'image pour lier les concepts entre eux, entraînant ainsi un traitement supplémentaire.

Enfin, la solution préconisée par la théorie de l'apprentissage multimédia est celle où l'on présente sur la même page et de manière intégrée le texte et les images. Cela veut dire que les portions du texte ou les idées clés sont présentées sur l'illustration à l'endroit auquel le texte fait référence (Mayer, 2001). L'apprenant a ainsi besoin de fournir moins d'effort pour lier le texte, où les concepts clés sont présélectionnés, aux images. Les études menées à ce sujet montrent que lorsque le texte est intégré aux illustrations les apprenants obtiennent de meilleurs résultats aux tests de rétention et de transfert (Mayer, 1989b ; Mayer *et al.*, 1995 ; Moreno & Mayer, 1999).

Le principe de contiguïté spatiale peut se résumer ainsi : dans le cas d'une présentation multimédia composée d'images et de texte écrit, les étudiants apprennent mieux lorsque le texte est placé à proximité de l'illustration (ou de la portion d'illustration) à laquelle il fait référence.

2. *La contiguïté temporelle*

La contiguïté spatiale est valable lorsque le texte est présenté à l'écrit, mais qu'en est-il lorsque celui-ci est sous forme audio ? Tout comme pour la contiguïté spatiale, il y a plusieurs possibilités. Le texte peut être présenté sous forme de narration avant l'illustration ou en même temps que celle-ci.

Dans le cas d'une présentation séquentielle nous pouvons appliquer le même raisonnement que précédemment, à savoir : si l'apprenant est exposé deux fois au même contenu ce sera plus efficace que si tout est présenté simultanément.

La théorie de l'apprentissage multimédia (Mayer, 2001) prédit cependant tout le contraire. Il est plus difficile d'établir des connexions entre deux informations qui sont séparées, que ce soit dans l'espace ou dans le temps. La raison étant que la mémoire de travail a une capacité limitée et qu'avant de pouvoir lier les deux informations, de pouvoir les traiter, l'apprenant est obligé de les retenir sur une durée plus ou moins longue. Prenons l'exemple d'une narration qui serait présentée avant une animation illustrant un phénomène scientifique. Au moment de démarrer l'animation, seule une petite partie des mots de la narration serait encore disponible en boucle phonologique, ce qui rend alors difficile la création de connexions entre les mots et les informations véhiculées par l'animation.

Dans une série d'expériences (Mayer *et al.*, 1999 ; Mayer & Anderson, 1991, 1992), Mayer et ses collègues ont obtenu des résultats mitigés en ce qui concerne les bénéfices de la contiguïté temporelle sur la mémorisation. Dans chacune de ces expériences les étudiants étaient divisés en deux groupes : dans l'un la narration et l'animation étaient présentées successivement ; dans l'autre, simultanément. Leur hypothèse était que les groupes pour lesquels la narration et l'animation étaient présentées en même temps produiraient de meilleurs résultats aux tests de rétention et de transfert. Sur cinq de ces expériences, deux montraient un avantage aux tests de rétention, tandis que pour les

autres il n'y avait pas d'effet significatif. Les auteurs expliquent cela par le fait que les apprenants aient pu se concentrer sur la narration sans autre distraction, ce qui les aurait aidés dans les tests de mémorisation. Même si la mémorisation est un processus important, le plus intéressant est d'étudier les effets de la contiguïté temporelle sur les tâches de transfert. C'est à travers ce test que l'on va pouvoir évaluer la capacité de l'apprenant à utiliser les connaissances qu'il vient de voir pour résoudre des problèmes. A ce niveau, les expériences de Mayer et de ses collègues ont montré un très net effet bénéfique de la contiguïté temporelle.

Dans deux études similaires (Mayer *et al.*, 1999 ; Moreno & Mayer, 1999), Mayer et ses collègues ont comparé la présentation de la même animation avec la même narration en simultané avec la présentation séquentielle de chaque segment de narration suivie de la portion de l'animation correspondante. Dans ce cas, les résultats n'ont montré aucune différence entre les deux types de présentation. L'explication avancée par les auteurs est que, les segments étant très courts, ceux-ci ne surchargent pas la mémoire de travail. Mais, en quoi est-ce différent de l'effet de partage attentionnel où les apprenants sont obligés de retenir les informations en mémoire de travail en faisant des allers-retours pour les relier entre elles ? Dans le cas que l'on vient de voir les informations à relier entre elles sont déjà sélectionnées pour l'apprenant. Alors que lorsque tout est présenté simultanément, il doit d'abord repérer les informations visuelles et verbales concordantes avant de pouvoir les connecter entre elles.

Pour résumer, le principe de contiguïté stipule que les informations complémentaires doivent être présentées assez proches à la fois dans l'espace et le temps. Nous avons déjà établi quelle est la proximité optimale pour l'espace. Qu'en est-il pour le temps ?

L'expérience de Baggett (1984) apporte un premier élément de réponse. Le but de cette étude était de déterminer l'effet de la contiguïté temporelle sur la formation d'associations entre du matériel visuel et auditif. Pour cela, les participants visionnaient une vidéo de 30 minutes décrivant les pièces d'un kit de construction avec leur nom et leur utilisation. La narration et les images étaient présentées de manière synchronisée ou décalée (*Tableau 1*).

| | | | | | | |
|-------|-------|------|-----|------|-------|-------|
| -21 s | -14 s | -7 s | 0 s | +7 s | +14 s | +21 s |
|-------|-------|------|-----|------|-------|-------|

Tableau 1 - Synchronisations entre narration et images

Dans une condition le visionnage était immédiatement suivi d'un test de mémorisation et dans l'autre le test avait lieu 7 jours plus tard. Dans les deux conditions, les deux groupes ayant obtenus les meilleurs scores (autour de 45% de rappels corrects pour le test immédiat et 30% pour l'autre) étaient ceux de la synchronisation parfaite et ceux où l'image avait une avance de 7 secondes sur la narration, avec un léger avantage pour la synchronisation parfaite. Il est important de noter que les trois groupes pour lesquels la narration avait une avance sur la présentation des images présentaient des résultats aux tests significativement plus mauvais que dans la condition inverse.

A partir de ces données nous pouvons conclure que la synchronisation optimale entre image et texte est de 0 seconde avec une courbe de résultats qui baisse significativement lorsque l'image a un retard de 7 secondes mais qui ne baisse que légèrement dans le cas contraire. Qu'en est-il lorsque l'on fait varier plus finement encore le décalage entre l'image et le son ? Quelle perception avons-nous de la simultanéité ?

IV. La composante temporelle

La plupart des articles scientifiques traitant de la perception sensorielle décrivent chaque sens (la vue, l'ouïe, le toucher, l'odorat, le goût) séparément, comme s'ils représentaient chacun une entité autonome. Pourtant, dans beaucoup de situations nos sens reçoivent un grand nombre d'informations concordantes concernant un objet ou un événement extérieur qui est ensuite combiné dans notre cerveau pour ne créer qu'une seule et même sensation. Ceci est notamment le cas en ce qui concerne notre perception de la simultanéité de un ou plusieurs stimuli externes.

1. Fenêtre temporelle d'intégration audiovisuelle

L'intégration audiovisuelle est un des exemples de l'intégration multi-sensorielle humaine, c'est-à-dire la manière dont le système nerveux intègre les informations provenant de l'ensemble de nos sens. La perception audiovisuelle d'un discours en est un très bon exemple. Ce phénomène est automatique lorsque les deux informations auditive (le son de la voix) et visuelle (le mouvement des lèvres) sont disponibles simultanément. Cette intégration peut même produire une illusion appelée effet McGurk (McGurk & McDonald, 1976) : l'observateur entend le son /da/ lorsque le son /ba/ est présenté en association avec le mouvement de lèvres décrivant le son /ga/. On constate que

cette illusion persiste (van Wassenhove *et al.*, 2007) malgré un décalage allant de -30 ms (avance de l'audio) à +170 ms (avance de l'image) entre le son et l'image. Cette fenêtre temporelle où le son et l'image sont fusionnés en une seule expérience perceptuelle est appelée fenêtre temporelle d'intégration audiovisuelle. Cette capacité à gérer les faibles asynchronies suggère une certaine flexibilité du système d'intégration multi-sensorielle humain.

2. Perception de la simultanéité

Comme on l'a vu avec l'effet McGurk, le système d'intégration multi-sensorielle est relativement flexible, c'est pourquoi il n'est pas surprenant que la perception de la simultanéité de stimulus visuel et auditif ne soit pas forcément directement liée au fait que les causes de ces derniers soient synchrones. Le seuil de discrimination temporelle entre deux stimulus audio et visuel se situe entre -50 ms (avance de l'audio) et +150 ms (avance de l'image) (van Eijk *et al.*, 2009). On constate d'ailleurs que ce seuil, qui correspond en réalité au seuil de perception de la simultanéité, s'étend sur une plage similaire à celle de la fenêtre temporelle d'intégration audiovisuelle. Il est intéressant de remarquer que les résultats de discrimination sont répartis symétriquement autour du point de simultanéité subjective (+50 ms), c'est-à-dire que plus on se rapproche des limites de l'intervalle et plus il devient facile de déterminer quel stimulus est présenté en premier.

La perception de la simultanéité est vraiment toute relative, même lorsque deux stimulus sont présentés simultanément, ceux-ci ne sont pas forcément perçus comme synchrones. C'est le cas notamment lorsqu'un stimulus est présenté sur le canal sensoriel sur lequel l'attention est dirigée. Ce stimulus est alors considéré comme étant antérieur à un autre stimulus qui pourtant aurait pu avoir lieu simultanément, c'est l'effet de « *prior entry*² » (Spence *et al.*, 2003).

Nous avons une perception flottante de la simultanéité et lorsque les intervalles sont très courts (de l'ordre de la centaine de millisecondes) il devient très difficile de déterminer avec précision l'antériorité d'un stimulus par rapport à un autre.

² « Entrée préalable » en français.

3. *L'effet Colavita*

L'effet de « *prior entry* » nous montre que l'on peut accorder plus d'importance à un sens qu'à un autre pour la simple raison que notre attention est focalisée dessus. Ce que l'effet Colavita va nous apporter est une preuve supplémentaire de la prédominance de la vue sur l'ouïe. Dans son expérience, Colavita (1974) présentait une série de stimuli unimodaux visuels et unimodaux auditifs, mais aussi bimodaux audiovisuels (ce dont les participants n'étaient pas au courant). Les participants devaient simplement préciser à quel type de stimulus ils venaient d'être confrontés. Pour 98% des stimuli bimodaux les participants n'identifiaient que le stimulus visuel. Dans une autre expérience (Koppen & Spence, 2007), les stimuli bimodaux ont été désynchronisés et il apparaît que l'effet Colavita disparaît au moment où les participants arrivent distinctement à repérer l'un des stimuli comme arrivant en premier.

Cela rejoint ce que nous avons vu avec la fenêtre temporelle d'intégration audiovisuelle, à savoir que notre cerveau fusionne les deux stimuli présentés en une seule et unique perception. L'effet Colavita nous montre que de cette perception nous ne sommes parfois conscient que de la composante visuelle.

V. Synthèse et hypothèses

Les différentes expériences sur la contiguïté temporelle nous indiquent qu'une synchronisation parfaite entre les contenus visuel et auditif est la plus bénéfique pour l'apprentissage. Toutefois, les décalages temporels auxquels la synchronisation parfaite était comparée étaient relativement importants. Dans les expériences de Mayer et Anderson (1991, 1992) le décalage était au minimum de la durée totale de la narration, soit plusieurs secondes (vu la longueur du texte). Dans celles de Baggett (1984) il était au minimum de 7 secondes. Ce qu'il est intéressant de remarquer c'est que l'effet de contiguïté disparaît lorsque la narration est découpée en segments plus courts et immédiatement suivie de la partie correspondante de l'animation (Mayer *et al.*, 1999 ; Moreno & Mayer, 1999). Dans ce dernier cas on peut considérer que le décalage est réduit à un maximum de 2 ou 3 secondes (le temps de lire une phrase). De même dans l'expérience de Baggett (1984), où nous constatons seulement une faible baisse des résultats lorsque l'image a une avance de 7 secondes par rapport au texte. Ce rôle prépondérant que semble avoir la vue sur l'apprentissage semble concorder

avec la perception que le cerveau a de la simultanéité et l'intégration que celui-ci fait de deux informations sensorielles rapprochées dans le temps.

Hypothèse 1 : Un faible décalage (de l'ordre de quelques secondes) entre une information visuelle et une information auditive pourrait avoir un effet bénéfique sur l'apprentissage.

Hypothèse 2 : Cet effet positif sera particulièrement prononcé lorsque l'information visuelle aura une avance sur la narration.

Dans notre expérience nous allons tester l'effet d'un décalage temporel entre l'affichage d'une phrase et de son équivalent audio. C'est pourquoi nous ajouterons que la seconde hypothèse repose également sur la notion d'attention guidée par le texte et sur le fait que la redondance d'information visuelle et auditive peut être bénéfique lorsque le texte ne reprend que les informations importantes de la narration. Le regard sera donc tout d'abord attiré par le texte et ce texte guidera l'attention auditive de l'apprenant sur les notions clés de la narration sans rajouter de coût de traitement supplémentaire.

Hypothèse 3 : La courbe de résultat prendra la forme d'une courbe en cloche symétrique dont le centre sera du côté de l'avance du texte sur la narration.

PARTIE EXPERIMENTALE

METHODE

L'étude décrite ici a pour but de déterminer la synchronisation optimale pour l'apprentissage (mémorisation et compréhension) entre l'affichage d'un texte et de son équivalent audio. Dans cette première expérience nous avons commencé par étudier le retard de l'audio sur le texte, c'est-à-dire que le texte s'affiche avant qu'il ne soit énoncé par la narratrice.

I. Participants

Trente quatre volontaires adultes (treize femmes et vingt-et-un hommes), âgés de 19 à 62 ans, ont participé à l'expérimentation. Tous étaient des stagiaires ou des salariés d'Essilor Dijon.

Nous avons dû éliminer trois participants en raison d'un score trop important au pré-test et un participant dont la langue maternelle n'était pas le français, ce qui nous en a laissé trente. Les participants restants ont été répartis au hasard en cinq groupes de six.

II. Matériel

1. Les vidéos

Nous avons réalisé une série de cinq vidéos présentant chacune une procédure de premier secours telles qu'elles sont décrites dans le référentiel national de prévention et secours civiques de niveau 1 (PSC1) édité par le ministère de l'intérieur.

Toutes les vidéos suivaient la même trame et étaient présentées de la même manière. Elles étaient chacune divisées en deux parties. La première présentait un plan large de la procédure réalisée par un secouriste sur une victime. Le film de la procédure apparaissait sur la gauche de l'écran, tandis que quatre phrases clés du commentaire audio étaient affichées sur quatre lignes à droite de l'écran. La phrase énoncée par la narratrice était d'abord affichée en noir, puis légèrement

grisée lorsque l'on passait à la suivante, afin de guider l'attention du participant. Un total de huit phrases étaient affichées dans cette partie pour chaque procédure.

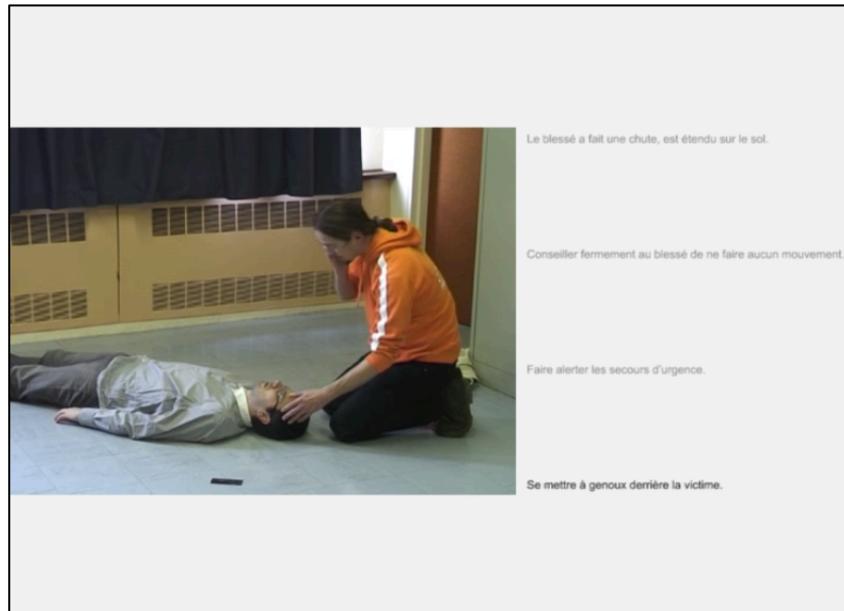


Figure 2 - Présentation de la première partie

La seconde partie de la vidéo présentait un diagramme récapitulant les étapes importantes de la procédure, avec les cases et les flèches apparaissant au fil de la narration. La phrase inscrite dans chaque case correspondait mot pour mot au commentaire audio. Un total de huit cases (donc de huit phrases) et de huit flèches a été respecté pour chaque procédure, de manière à ce que les résultats soient homogènes³.

³ Voir matériel expérimental en annexe.

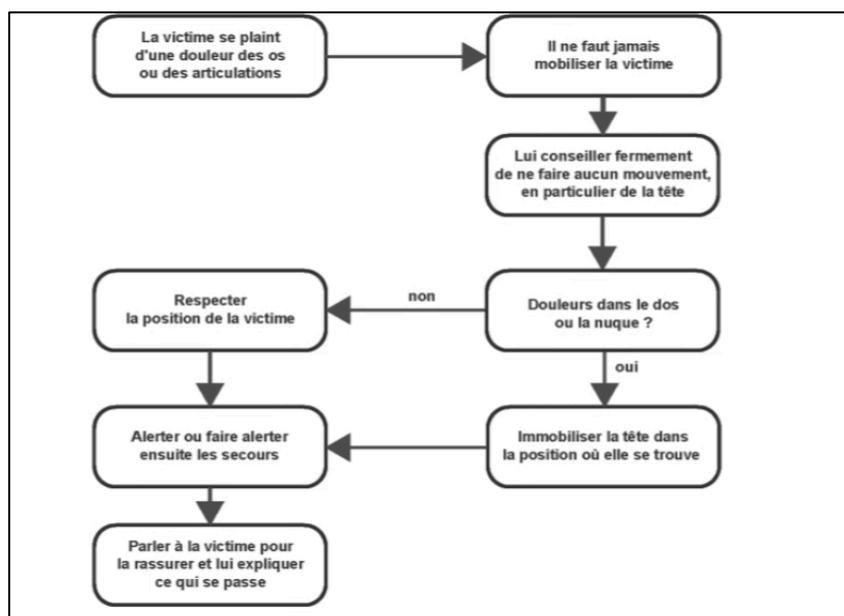


Figure 3 - Présentation de la deuxième partie

Les vidéos étaient au format 720x576 à 24 images par secondes. Elles commençaient chacune par un écran blanc sur lequel figurait le titre de la procédure, elles duraient en moyenne deux minutes et étaient systématiquement suivies d'un blanc de dix secondes.

2. Les procédures

Les 5 procédures de premiers secours présentées étaient toujours diffusées dans l'ordre suivant :

La chute (A), la brûlure (B), la libération des voies aériennes (C), le malaise (D) et l'hémorragie (E).

3. Les commentaires audio

Les commentaires audio ont été enregistrés en 44,1 KHz stéréo et ont été diffusés au même volume pour tous les participants.

Dans la première partie, les commentaires audio ont été parfaitement synchronisés avec les actions du secouriste. Seule la synchronisation avec l'apparition du texte a été modifiée. Dans la seconde partie, c'est la synchronisation avec l'apparition des cases et des flèches qui a été modifiée.

Chaque vidéo était déclinée en cinq versions, chacune correspondant à une synchronisation différente du commentaire audio avec l'affichage du texte et des cases (Tableau 2).

| | | | | |
|----------|----------|----------|---------|------|
| -2000 ms | -1500 ms | -1000 ms | -500 ms | 0 ms |
|----------|----------|----------|---------|------|

Tableau 2 - Synchronisations

4. Les questionnaires

Tous les questionnaires étaient distribués au format papier⁴.

a. Questionnaire de pré-test

Le pré-test était composé de six questions. La première permettait de savoir si le participant avait déjà suivi une quelconque formation en premier secours (BNS, AFPS, ...), tandis que les cinq autres concernaient leurs connaissances sur des procédures autres que celles présentées dans les vidéos. Ce questionnaire a permis d'éliminer trois participants dont les connaissances préalables auraient pu biaiser les résultats.

b. Questionnaire de mémorisation (variable dépendante 1)

Celui-ci se présentait sous la forme d'un questionnaire à choix double. Pour chaque phrase donnée dont on précisait le contexte (c.-à-d. la procédure associée), les participants devaient dire s'ils y avaient été confrontés, mot pour mot, ou non. Sur les 160 phrases proposées, 80 étaient issues des vidéos (8 pour le film, 8 pour le diagramme, soit 16x5), tandis que les autres étaient fausses. Par fausses il faut comprendre que les phrases avaient soit été radicalement modifiées, soit ne correspondaient pas au contexte. Les phrases étaient présentées dans un ordre aléatoire et dans un tableau tel que celui-ci.

| Procédure associée | Avez-vous été confronté aux phrases suivantes ? | Oui | Non |
|--------------------|--|-----|-----|
| La chute | Le blessé a fait une chute, est étendu sur le sol. | | |
| La brûlure | Une rougeur de la peau chez l'adulte. | | |

⁴ Les fichiers des questionnaires sont disponibles en annexe.

c. Questionnaire de compréhension (variable dépendante 2)

Pour le questionnaire de compréhension, les participants devaient reconstituer les diagrammes de chaque procédure. Pour cela ils devaient replacer les huit phrases de chaque diagramme dans les cases correspondantes (en y inscrivant quelques mots clés) puis relier les cases entre elles en redessinant les flèches.

d. Questionnaire d'informations générales

Ce questionnaire avait à la fois pour but d'interroger les participants sur leur âge, leur sexe, leur ville de naissance et leur niveau d'étude, mais aussi de réinitialiser leur mémoire de travail.

III. Contre-balancement

Afin d'avoir recours à moins de participants et de compenser les différences de difficultés entre les procédures, nous avons répartis les différentes modalités de chaque procédure selon le carré gréco-latin suivant :

| | | | | |
|----|----|----|----|----|
| A1 | B4 | C2 | D5 | E3 |
| B2 | C5 | D3 | E1 | A4 |
| C3 | D1 | E4 | A2 | B5 |
| D4 | E2 | A5 | B3 | C1 |
| E5 | A3 | B1 | C4 | D2 |

Les lettres désignant les procédures et les chiffres les synchronisations, comme suit :

| | |
|---|-----------------------------------|
| A | La chute |
| B | La brûlure |
| C | La libération des voies aériennes |
| D | Le malaise |
| E | L'hémorragie |

| | |
|---|----------|
| 1 | -2000 ms |
| 2 | -1500 ms |
| 3 | -1000 ms |
| 4 | -500 ms |
| 5 | 0 ms |

IV. Procédure

L'expérience a eu lieu en alternance dans deux salles de réunion de tailles et de configurations similaires. Les participants passaient en semi-collectif, par groupes de trois à six personnes. La vidéo était projetée sur un écran mural à l'aide d'un vidéoprojecteur et le son diffusé par deux petites enceintes. La durée moyenne d'une session était d'une heure et quinze minutes.

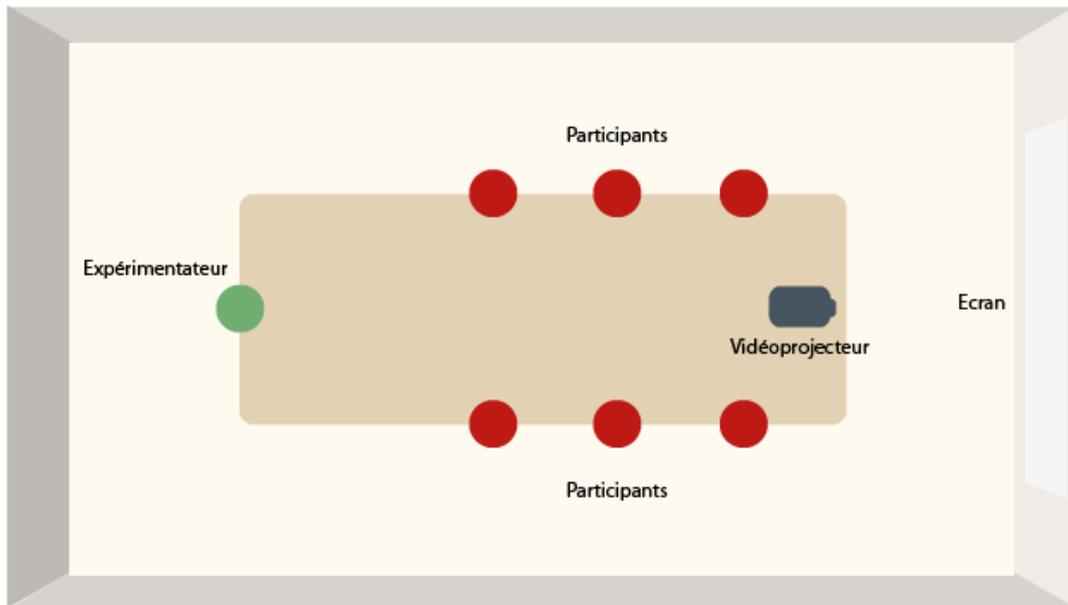


Figure 4 - Salle d'expérimentation

1. Accueil des participants

En arrivant, les participants étaient priés de prendre place autour de la table, le plus proche de l'écran possible. Les participants étaient remerciés de leur participation et on leur présentait le déroulement de l'expérimentation. Nous leur demandions ensuite de signer les formulaires de consentement en deux exemplaires. Après une courte explication sur le thème des vidéos ils devaient répondre au questionnaire de pré-test. Cela fait, nous leur expliquions le contenu des vidéos, puis on leur montrait la vidéo de familiarisation. Lorsqu'il n'y avait plus de questions et que tout le monde était prêt, la projection démarrait. Les vidéos étaient visionnées en une seule fois (avec dix secondes de blanc entre chaque procédure), sans prise de note et sans contrôle sur la lecture.

2. *Questionnaires*

Une fois les vidéos visionnées, les questionnaires d'informations générales étaient distribués, suivis du questionnaire de mémorisation. Les participants avaient alors environ 25 minutes pour dire si oui ou non ils avaient lu ou entendu chacune des 160 phrases proposées. Enfin, ils recevaient le questionnaire de compréhension et disposaient de 25 minutes pour reconstituer les diagrammes de toutes les procédures.

3. *Fin de l'expérimentation*

Une fois tous les questionnaires remplis, les participants étaient encore une fois remerciés et, le cas échéant, une discussion sur le sujet de l'étude avait lieu.

Il est à noter que lors des trois premières sessions nous demandions aux participants de nous indiquer la modalité qu'ils avaient préférée en leur faisant visionner un court extrait de dix secondes dans les cinq versions (une fois de -2000 ms à 0 ms, puis une deuxième fois dans le sens inverse). Cependant, aucun n'arrivait à percevoir la différence de synchronisation entre deux versions contigües et aucun n'était capable de nous indiquer une quelconque préférence. Cette démarche a donc été abandonnée au profit d'une discussion plus longue sur les impressions des participants.

V. Variables dépendantes

4. *Mémorisation*

Les résultats au questionnaire de mémorisation ont été obtenus en comptabilisant les phrases justes correctement reconnues. Pour chaque procédure il y avait 16 phrases issues mot pour mot de la vidéo, 8 sur la partie « vidéo » et 8 sur la partie « diagramme ». A raison d'un point par phrase nous avons obtenu des notes sur 16.

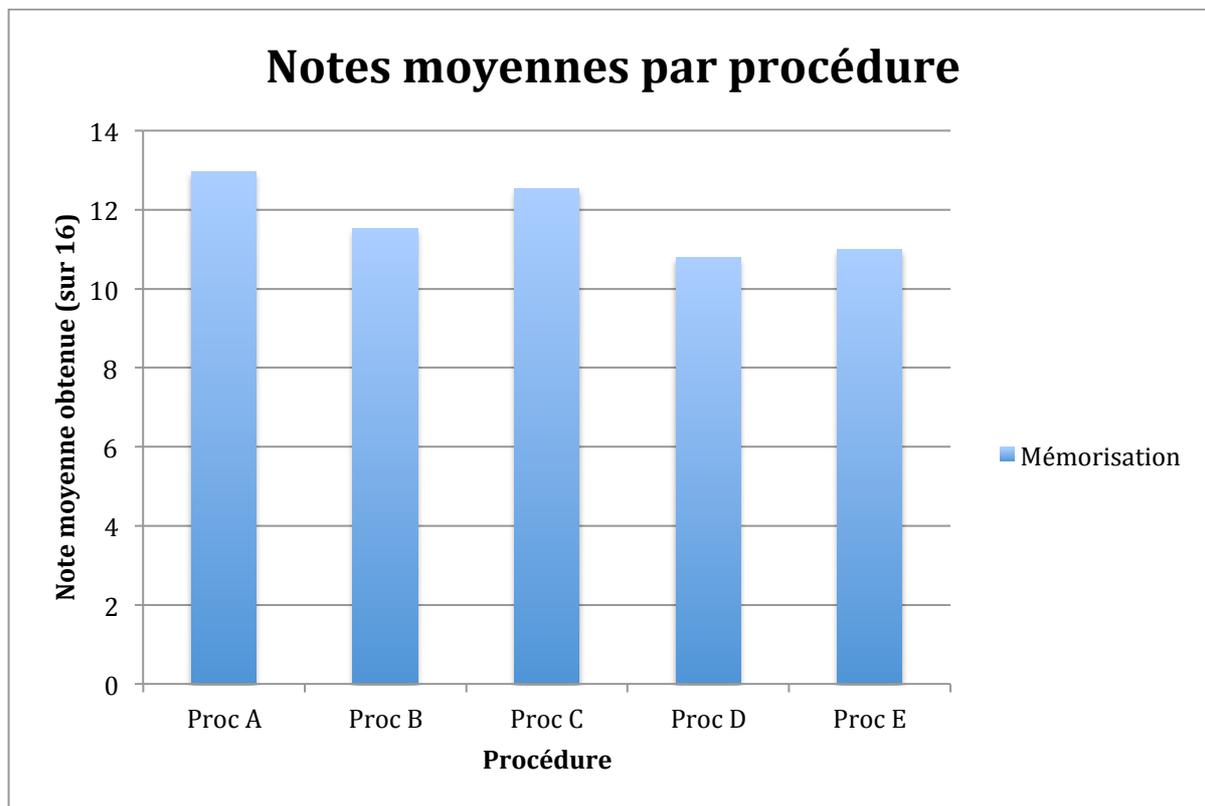
5. *Compréhension*

Les résultats au questionnaire de compréhension ont été obtenus en comptabilisant les phrases et les flèches correctement (et strictement) replacées sur les diagrammes. Chaque diagramme était composé de 8 cases et de 8 flèches, nous avons donc obtenu des notes sur 16.

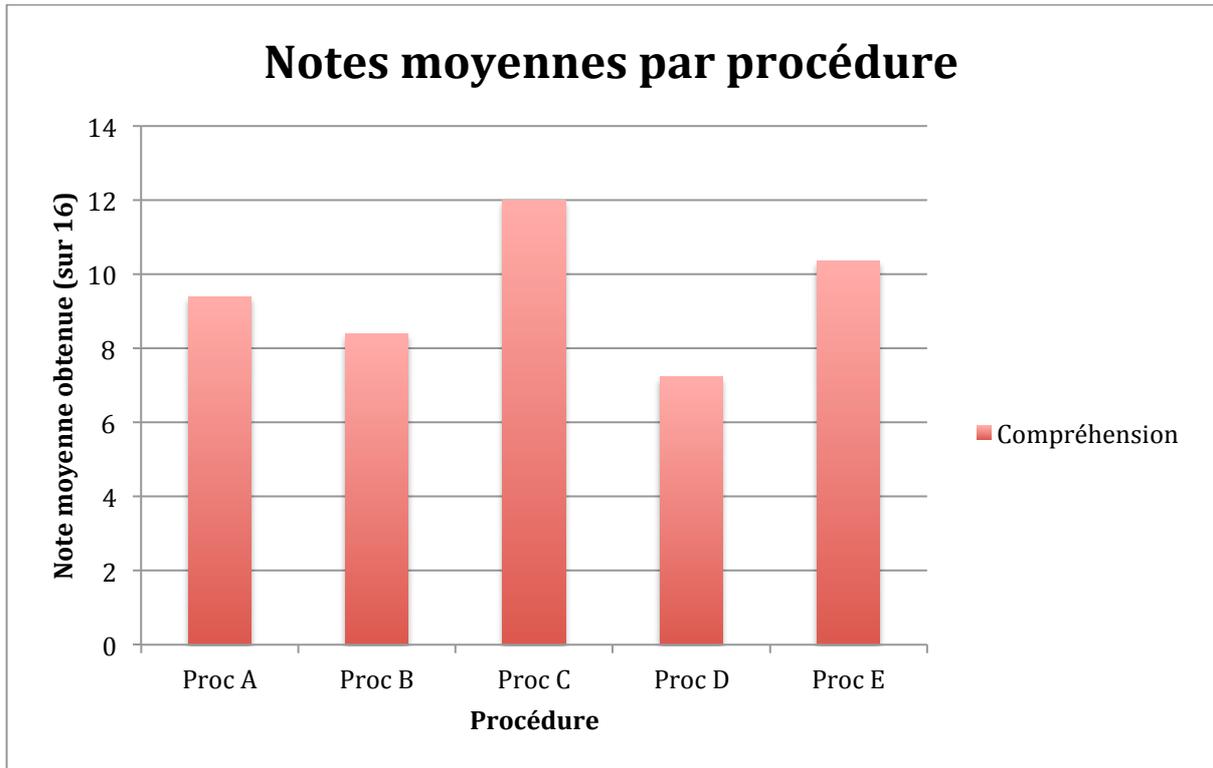
RESULTATS

Les données relatives aux procédures ont été représentées dans les graphiques 1 et 2, celles relatives aux synchronisations dans les graphiques 3 et 4. Les tableaux de résultats complets sont disponibles en annexe.

Graphique 1 : *Notes moyennes des participants aux tests de mémorisation (sur 16) en fonction des procédures.*

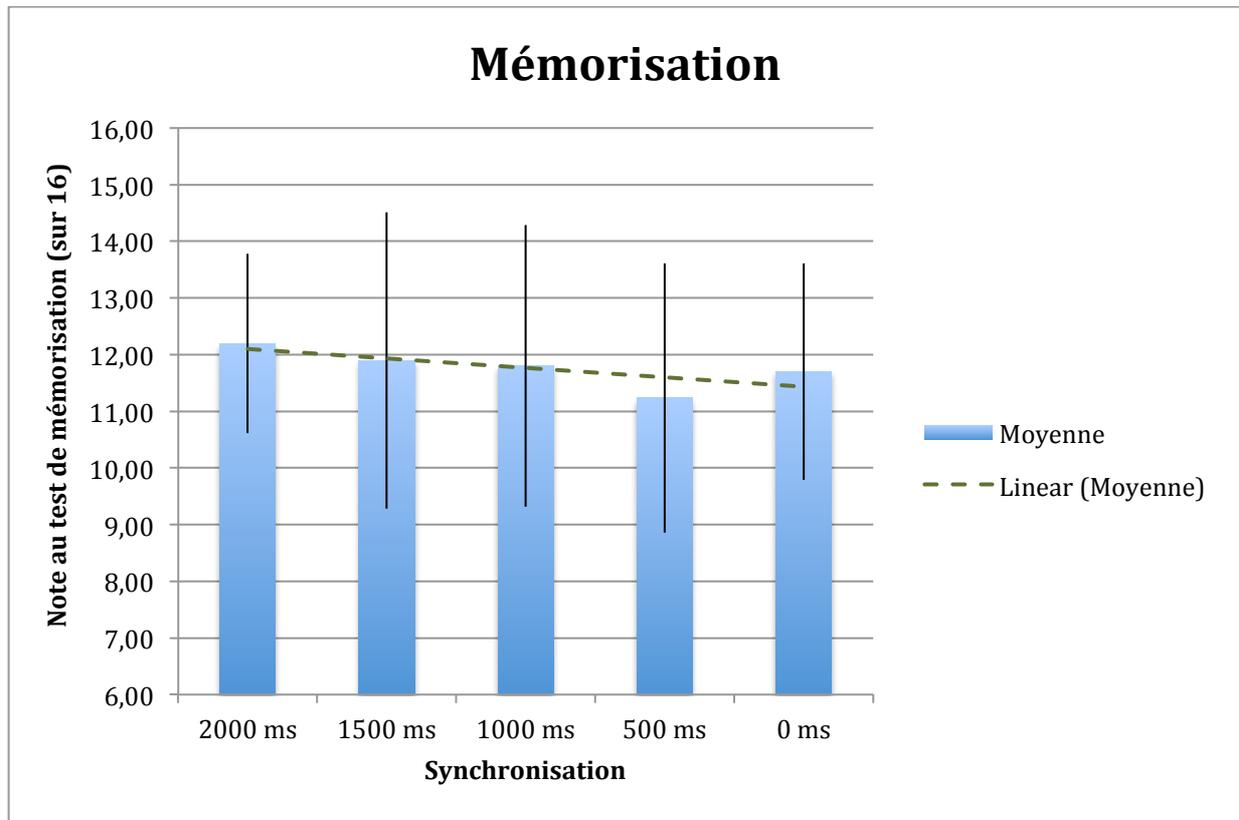


Graphique 2 : Notes moyennes des participants aux tests de compréhension (sur 16) en fonction des procédures.



Une première analyse nous montre que le niveau de difficulté des procédures n'est pas homogène. Les procédures B et D, par exemple, ont obtenu des résultats nettement moins bons que les autres au test de compréhension. Ceci n'a toutefois pas une grande importance puisque nous avons effectué un contre-balancement des procédures et des synchronisations.

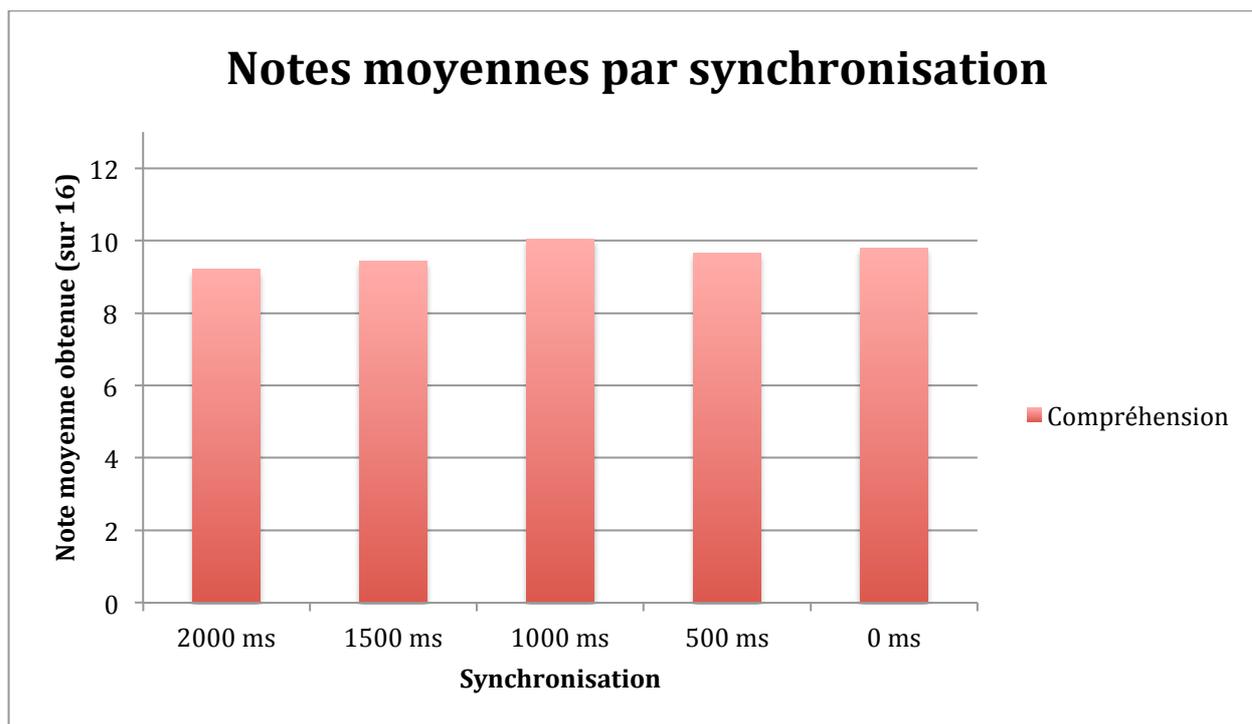
Graphique 3 : Note moyenne des participants au test de mémorisation (sur 16) avec écart-type et droite de tendance.



Les données de mémorisation et de compréhension ont fait l'objet d'analyses de la variance pour mesures répétées avec la synchronisation du texte par rapport au commentaire audio en facteur intra-participant.

L'analyse n'a pas révélé d'effet de la synchronisation des commentaires sur la mémorisation et la compréhension, respectivement $F(4, 116) = 1.089, p = .365$ et $F(4, 116) < 1, p = .888$. Bien que nous puissions constater une différence significative entre S1 et S4, il est impossible d'en tirer une conclusion en raison de la forte variation des données. Il en va de même pour la droite de tendance qui pourrait indiquer une baisse des résultats plus on se rapproche de la synchronisation parfaite.

Graphique 4 : Note moyenne des participants au test de compréhension (sur 16).



L'analyse des notes moyennes par synchronisation ne révèle aucun effet de la synchronisation sur la compréhension.

A l'issue de l'expérimentation, les participants nous ont révélé qu'ils n'avaient pas lu le texte qui s'affichait sur la droite de l'écran. Selon eux il était écrit trop petit et ne présentait aucun intérêt puisqu'il était parfaitement redondant avec la narration. Ils renchérisaient en ajoutant que la narration était même plus complète.

Les participants ont également été incapables de donner leur préférence lorsque nous leur avons présenté les cinq synchronisations les unes à la suite des autres. Ils n'arrivaient en réalité pas à percevoir la différence entre les différents décalages. Après que nous leur ayons expliqué les différences de synchronisation, la plupart nous répondaient qu'intuitivement ils préféraient la synchronisation parfaite.

DISCUSSION

L'objectif de notre étude était de déterminer les effets sur la mémorisation et la compréhension que pouvait avoir un décalage temporel entre la présentation d'un texte écrit et de son équivalent oral dans un objet d'apprentissage multimédia. Les participants visionnaient une série de vidéos présentant des procédures de premiers secours accompagnées d'un texte écrit et d'un commentaire audio équivalent. Nous leur avons ensuite demandé de répondre à des questionnaires de rétention et de compréhension.

Hypothèse 1 : Malgré une droite de tendance qui montre un léger effet positif croissant lorsque le décalage entre le texte et le son augmente, on ne peut en tirer aucune conclusion en raison de la variabilité des données. L'hypothèse est donc invalidée.

Hypothèse 2 : La première hypothèse étant invalidée nous n'avons pas procédé à la deuxième expérience qui aurait étudié l'effet de l'avance du son par rapport au texte. Nous n'avons donc aucune base de comparaison.

Hypothèse 3 : La courbe de résultats étant proche d'une droite à faible pente, cette hypothèse est également invalidée.

L'analyse des résultats aux questionnaires n'a révélé aucun effet significatif de la désynchronisation sur nos deux variables dépendantes. Même si la trop grande variabilité des données ne nous permet pas de tirer de conclusion définitive, nous avons pu constater que la tendance des résultats était à la hausse plus on s'éloignait de la synchronisation parfaite. Nous pouvons ainsi émettre l'hypothèse que les résultats seraient plus marqués avec un nombre de participants plus important. Toutefois, si cet effet n'est pas significatif avec 30 participants il risque de ne pas être très important, même avec un plus grand nombre de participants.

L'effet de contiguïté temporelle stipule que deux informations de natures différentes mais complémentaires doivent être présentées de manière rapprochées dans le temps (Mayer, 2001). Ce principe se base sur une capacité limitée à la fois dans sa taille et dans le temps de la mémoire de travail. L'explication la plus simple au fait que notre expérimentation n'a produit aucun résultat significatif est que la capacité de la mémoire de travail ne diminue pas sur une période aussi courte que les intervalles que nous avons utilisés. Ceci concorderait avec les expériences de Mayer et de ses collègues (Mayer *et al.*, 1999 ; Moreno & Mayer, 1999).

L'avance de la présentation du texte sur le commentaire audio devait permettre aux participants de prendre connaissance des éléments importants de la narration avant qu'ils ne soient présentés en guidant leur attention auditive (Mayer & Johnson, 2008). Nous avons postulé que ce texte serait lu malgré la présence d'une vidéo en raison du guidage de l'attention visuelle par le texte (Schmidt-Weigand *et al.*, 2010). L'avance du texte sur la narration aurait ainsi dû produire de meilleurs résultats aux tests de rétention et de compréhension qu'une synchronisation parfaite. Ce n'est pas ce que nous avons observé. Les discussions que nous avons eu avec les participants après l'expérimentation vont peut-être nous apporter un début de réponse.

A la question « *Avez-vous remarqué une différence de présentation entre les différentes procédures ?* », les participants nous répondaient presque invariablement que non. Nous leur expliquions alors qu'il y avait une désynchronisation entre le texte et la narration, ce sur quoi ils nous répondaient qu'ils n'avaient pas lu le texte. Les raisons avancées étaient que, d'une part, le texte était écrit trop petit et que, d'autre part, il était parfaitement redondant avec la narration. La redondance et la difficulté de lecture rendaient pour eux le texte inutile et les amenaient à se concentrer exclusivement sur la vidéo. Ceci ne signifie toutefois pas qu'ils n'aient pas lu le texte, ils ont pu en prendre connaissance inconsciemment, mais le fait de diriger volontairement leur attention sur la vidéo a dû considérablement diminué l'effet de la désynchronisation.

Un autre facteur qui a pu jouer dans cette expérimentation est la modalité technique de présentation. En effet, les vidéos étaient projetées sur un écran mural augmentant ainsi la distance entre la vidéo, sur laquelle leur attention était tournée, et le texte. Les mouvements oculaires que les participants avaient alors à faire entre les deux étaient peut-être trop importants. Ceci aurait pu augmenter l'effet de partage attentionnel et ainsi masquer celui de la désynchronisation.

Après les questionnaires et pour les trois premiers groupes de participants, nous avons montré un extrait d'une dizaine de secondes de l'une des vidéos de l'expérience. Cet extrait était décliné dans chaque version de désynchronisation. Les vidéos étaient présentées une première fois de S1 à S5, puis une deuxième de S5 à S1. Nous avons ensuite demandé aux participants de nous indiquer la version qu'ils préféraient. Ils se sont avérés incapables de répondre à la question car ils ne voyaient aucune différence entre les différentes modalités, à part entre S1 et S4. Ne produisant aucun résultat, cette question a par la suite été abandonnée mais nous pouvons tout de même en tirer un début de conclusion. Bien que la fenêtre temporelle d'intégration audiovisuelle se situe entre -30 et +170 ms, il semblerait que lorsque les participants ne savent pas ce à quoi ils doivent faire attention ils ne remarquent pas la désynchronisation en dessous d'un certain seuil (ici 1500 ms). Il est ainsi possible que les décalages temporels utilisés étaient trop faibles. Rappelons-nous de l'expérience de Baggett (1984) pour laquelle les résultats ne baissaient que faiblement entre la synchronisation parfaite et l'avance de 7 secondes de l'image par rapport au son. Le maximum de la courbe en cloche postulé dans l'hypothèse 3 se trouverait alors plus loin que les 2 secondes de notre expérience.

Le questionnaire de compréhension a également pu poser problème. Les participants devaient reconstituer les diagrammes d'actions des procédures en remplaçant les phrases dans les bonnes cases et en redessinant les flèches. Les notes étaient basées sur le nombre de phrases et de flèches bien placées. Le problème avec ce système de notation est qu'il n'évalue pas toujours de façon très précise le véritable degré de compréhension des participants. Il suffisait par exemple qu'ils mettent la mauvaise action en premier, décalant ainsi toute la suite. Ainsi, même si 6 des actions (sur 8) étaient dans le bon ordre la note pouvait être 0 (pour la partie remplacement du texte). Certaines personnes oublièrent de redessiner les flèches, d'autres les mettaient aux bons endroits mais relierent les mauvaises actions entre elles obtenant ainsi des points sans vraiment avoir compris.

Le thème des premiers secours ne se prêtait peut-être pas assez au but de l'étude. L'apprentissage des premiers secours relève plus de l'apprentissage procédural que de l'apprentissage d'informations factuelles ou de concepts. Il aurait été plus judicieux de rester dans la lignée des expériences de Mayer (2001) en utilisant une animation expliquant un phénomène naturel tel que la formation des éclairs ou le fonctionnement d'une pompe à vélo. Nous serions ainsi restés dans le même cadre que les recherches sur lesquelles nous nous sommes basées pour notre étude.

Pour toutes les raisons que nous venons de mentionner, l'absence de différences significatives entre les synchronisations ne permet pas de conclure avec certitude qu'il n'y a pas d'effet. Il serait intéressant de refaire une expérience similaire sur la base de ce que nous avons appris.

BIBLIOGRAPHIE

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* (45), pp. 598-607.

van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2009). Temporal Interval Discrimination Thresholds Depend on Perceived Synchrony for Audio-Visual Stimulus Pairs. *Journal of Experimental Psychology: Human Perception and Performance* , 35 (4), pp. 1254-1263.

Arguel, A., & Jamet, E. (2009, March). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior* (25), pp. 354-359.

Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior* (25), pp. 354-359.

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. Dans K. Spence, & J. Spence, *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Baddeley, A., & Hitch, G. (1974). Working Memory. Dans G. Bower, *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Baggett, P. (1984). Role of Temporal Overlap of Visual and Auditory Material in Forming Dual Media Associations. *Journal of Educational Psychology* , 76 (3), pp. 408-417.

Bétrancourt, M., & Tversky, B. (2000). Effect of computer animation on users' performance: a review. *Travail-Humain* (63), pp. 311-329.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in effects of word length on immediate memory. *Journal of Memory and Language* (31), pp. 1-17.

- Colavita, F. B. (1974). Human sensory dominance. *Perception and Psychophysics* (16), pp. 409-412.
- Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology* (55), pp. 75-84.
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction* (16), pp. 511-525.
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction* (17), pp. 722-738.
- Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia learning. *Computers in Human Behavior* (24), pp. 786-797.
- Jamet, E. (2008). Peut-on concevoir des documents électroniques plus efficaces ? L'exemple des diaporamas. *Revue européenne de psychologie appliquée* (58), pp. 185-198.
- Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology* (32), pp. 588-598.
- Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and Instruction* (18), pp. 135-145.
- Koppen, C., & Spence, C. (2007). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research* (1186), pp. 224-232.
- Lin, Y.-G., McKeachie, W. J., & Kim, C. Y. (2003). College student intrinsic and/or extrinsic motivation and learning. *Learning and Individual Differences* (13), pp. 251-258.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum Associates.
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology* (81), pp. 240-246.
- Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction* (13), pp. 125-139.

- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology* (83), pp. 484-490.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology* (84), pp. 444-452.
- Mayer, R. E., & Johnson, C. I. (2008). Revising the Redundancy Principle in Multimedia Learning. *Journal of Educational Psychology* , 100 (2), pp. 380-386.
- Mayer, R. E., & Moreno, R. (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. *Journal of Educational Psychology* , 90 (2), pp. 312-320.
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction* (12), pp. 107-119.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist* , 38 (1), pp. 43-52.
- Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology* (91), pp. 638-643.
- Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research & Development* (43), pp. 31-43.
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature* (264), pp. 746-747.
- Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities* , 29 (3), p. 271.
- Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology* , 91 (2), pp. 358-368.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition* , 17 (4), pp. 398-422.

Savoy, A., Proctor, R. W., & Salvendy, G. (2009). Information retention from PowerPoint and traditional lectures. *Computers & Education* (52), pp. 858-867.

Schmidt-Weigand, F., Kohnert, A., & Glowallia, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction* (20), pp. 100-110.

Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception and Psychophysics* (65), pp. 318-328.

ANNEXES

La totalité du matériel expérimental se trouve sur le CD accompagnant ce mémoire ainsi que sur l'ENT de l'université de Poitiers.

Contenu du CD :

- Questionnaire de prétest
- Questionnaire d'informations générales
- Questionnaire de mémorisation
- Questionnaire de compréhension
- Une vidéo exemple
- Tableau complet des résultats

Analyse de Variance de Mesures Répétées avec Taille des Effets et Puissance (Resultats) Paramétrisation sigma-restreint Décomposition efficace de l'hypothèse

| | SC | Degré de - Liberté | MC | F | p | Eta-deux partiel | Non- centralité | Puissance observée (alpha=0.05) |
|-----------------------|--------------|-----------------------|--------------|--------------|--------------|---------------------|--------------------|------------------------------------|
| Ord.Or ig. | 13900. 91 | 1 | 13900. 91 | 648.64 14 | 0.0000 00 | 0.957205 | 648.6414 | 1.000000 |
| Erreur | 621.49 | 29 | 21.43 | | | | | |
| R1 | 12.49 | 4 | 3.12 | 0.2832 | 0.8883 30 | 0.009673 | 1.1330 | 0.110933 |
| Erreur | 1279.1 1 | 116 | 11.03 | | | | | |