Extension de KLM au Pointage du Regard et Validation Gestuelle dans HoloLens

Florent Cabric, Emmanuel Dubois, Marcos Serrano

▶ **To cite this version:**

Florent Cabric, Emmanuel Dubois, Marcos Serrano. Extension de KLM au Pointage du Regard et Validation Gestuelle dans HoloLens. 32ème Conférence Internationale Francophone sur l'Interaction Homme-Machine (IHM 2021), Apr 2021, Metz, France. pp.1-14. hal-03232391v1

HAL Id: hal-03232391

https://hal.science/hal-03232391v1

Submitted on 21 May 2021 (v1), last revised 21 Jun 2021 (v2)

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

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

Extension de KLM au Pointage du Regard et Validation Gestuelle dans HoloLens

Florent Cabric, florent.cabric@irit.fr
Université de Toulouse, IRIT, Toulouse, France

Emmanuel Dubois, emmanuel.dubois@irit.fr
Université de Toulouse, IRIT, Toulouse, France

Marcos Serrano, marcos.serrano@irit.fr
Université de Toulouse, IRIT, Toulouse, France

The advances in mixed reality head-mounted displays led researchers to investigate such systems but the design of mixed reality app could be laborious. Evaluate the effects of the arrangement of these interfaces in the physical world on user performance is complex. In this context, the use of a predictive model allows researchers to estimate, prior to the implementation, the usability issues of such applications. To this end, we propose a new KLM extension to model HMD-based interaction focusing on searching, pointing and validation tasks. First, we define new KLM operators and estimate through two user studies the estimated time for each operator. Then, we validate the model for an ecological task. This final study confirms that the model could be applied for predicting the interaction time on mixed reality interaction.

CCS CONCEPTS • Human-centered computing • Human computer interaction (HCI) • HCI design and evaluation methods • User models

Additional Keywords and Phrases: evaluation method, design method, KLM, Head-mounted display, mixed reality

Les progrès réalisés dans le domaine des casques de réalité mixte (HMD) font entrer ce domaine dans une nouvelle ère où les prototypes conçus et développés en laboratoire peuvent techniquement devenir des produits commerciaux. Toutefois, le problème réside désormais dans la conception de ces systèmes : il est difficile d'anticiper les effets de la disposition spatiale de l'interface et le développement ou l'évaluation d'applications de réalité mixte sont fastidieux. Dans ce contexte, l'utilisation de modèles de prédiction du temps de réalisation d'une tâche est une approche prometteuse pour aider les concepteurs à anticiper les problèmes d'utilisabilité dès les premières étapes du processus de conception (i.e. avant le développement de telles solutions interactives). Dans ce but, nous présentons une nouvelle extension de KLM pour modéliser l'interaction basée sur les HMDs en mettant l'accent sur la recherche spatiale d'une cible, le pointage du regard et la validation gestuelle. Nous définissons d'abord de nouveaux opérateurs KLM requis pour modéliser ces interactions, et déterminons expérimentalement l'unité de temps pour chaque opérateur. Ensuite, nous validons le modèle en comparant le temps prédit avec le temps mesuré pour une tâche écologique combinant plusieurs de ces nouveaux opérateurs. Cette étude confirme la validité de nos nouveaux opérateurs et leur conformité avec le modèle KLM.

Mots-clés additionnels : méthode d'évaluation, méthode de conception, KLM, réalité mixte

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word.

In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 Introduction

Au cours des dernières années, l'avènement des casques de réalité mixte (HMD), apportant de multiples avancées et de nouvelles applications, a fait avancer la recherche en visualisation et en interaction. Maintenant, les HMDs peuvent cartographier l'environnement spatial, pour placer des interfaces n'importe où dans notre environnement, et peuvent aussi interpréter des gestes en l'air et le suivi du regard pour effectuer un pointage spatial. Par conséquent, le développement de systèmes immersifs nécessite de prendre en compte les effets de la disposition spatiale de l'interface. En effet, certains éléments peuvent être situés en dehors du champ de vision de l'utilisateur. Évaluer le temps nécessaire pour effectuer ces tâches de recherche, de pointage et de validation est un problème pour les chercheurs et les concepteurs. Le développement et les évaluations de technologies immersives sont fastidieux, d'où l'inefficacité, dans le contexte de l'interaction immersive [1], de l'approche habituelle combinant le développement de prototypes et des études utilisateurs.

Pour faciliter l'exploration des possibilités de conception des applications immersives, nous présentons une extension du modèle prédictif Keystroke Level Modeling (KLM) [5]. KLM permet aux concepteurs de prédire les temps de réalisation d'une tâche à partir d'interactions atomiques appelées opérateurs, et d'identifier les problèmes d'utilisabilité pendant la phase de conception plutôt que pendant la phase de développement [4]. Bien que KLM ait fait l'objet de nombreuses extensions pour des domaines spécifiques (tels que l'interaction basée sur des smartphones), peu d'attention a été accordée à la modélisation des interactions dans le contexte des environnements immersifs. Pourtant, un tel contexte implique des comportements spécifiques tels que le déplacement de la tête pour rechercher l'information correcte dans l'espace immersif ou le fait d'amener la main dans le champ de vision du casque pour effectuer une validation gestuelle. L'originalité de notre travail est de se concentrer spécifiquement sur la conception d'une extension de KLM qui modélise l'interaction pour effectuer des tâches spécifiques au contexte de la réalité mixte avec un casque telles que la recherche, le pointage et la sélection d'objets numériques.

Dans cet objectif, nous identifions tout d'abord un ensemble d'opérateurs, nouveaux ou existants dans la littérature, utiles pour modéliser les interactions atomiques susmentionnées. Pour les opérateurs nouvellement identifiés, nous réalisons deux études utilisateurs afin de définir leur temps unitaire respectif, c'est-à-dire le temps nécessaire pour effectuer l'action correspondante. Dans une étude utilisateur finale, nous mesurons le temps nécessaire pour effectuer une tâche qui combine deux des opérateurs KLM nouvellement introduits, que nous comparons au temps prédit (i.e. calculé avec notre extension du modèle KLM). Toutes ces études ont été réalisées de façon à ce que les mesures sanitaires imposées par le COVID-19 soient scrupuleusement appliquées. Cette étude finale valide nos contributions puisque la différence entre le temps observé et le temps prédit est d'environ 2 %, ce qui est conforme à une approche KLM.

Nos contributions sont 1) des opérateurs KLM pertinents pour décrire les interactions avec un casque de réalité mixte, 2) la définition de temps unitaires appropriés pour ces opérateurs par le biais de deux études utilisateurs et 3) la validation expérimentale de nos contributions.

2 État de l'art

L'évaluation dans le domaine de l'IHM repose traditionnellement sur des études utilisateurs, qui sont parfois difficiles à mettre en oeuvre par la complexité des tâches, la multiplicité des modalités d'interaction ou des contextes d'utilisation particuliers. L'évaluation basée sur un modèle [15] permet aux concepteurs d'évaluer un système ou une technique d'interaction avant son développement. Dans cette section, nous synthétisons les travaux antérieurs concernant le modèle de prédiction KLM et ses extensions.

2.1 KLM

Le modèle KLM (Keystroke-Level Model) [5] fait partie de la famille GOMS (l'un des premiers modèles prédictifs). Le modèle GOMS (pour Goals, Operators, Methods, Selection) [6] modélise à la fois les actions de l'utilisateur sur le système et ses actions cognitives. Contrairement à d'autres modèles (par exemple, le *cognitive walkthrough* [24]), la

famille GOMS modélise une tâche routinière (c'est-à-dire une tâche exécutée par l'utilisateur sans erreur) pour obtenir une prédiction du temps de réalisation de la tâche.

KLM a été conçu pour prédire le temps d'exécution selon la séquence des opérateurs pour réaliser une tâche. Par rapport au modèle GOMS initial, KLM n'utilise que les concepts d'opérateurs (O) et de méthodes (M), ignorant ainsi les objectifs (G) et les sélections (S). Avec KLM, la modélisation d'une tâche nécessite de décomposer la tâche en un ensemble d'opérateurs tels que l'appui sur une touche, le pointage, les actions cognitives, etc. Chaque opérateur est défini par une unité de temps : il correspond à l'estimation de la durée de ces actions lorsqu'elles sont effectuées par l'utilisateur. Par conséquent, le temps pour réaliser la tâche est égal à la somme de la durée de chaque opérateur modélisant la tâche. Le modèle KLM était initialement limité aux environnements de bureau, mais l'émergence de nouveaux dispositifs ou de nouveaux besoins des utilisateurs a conduit les chercheurs à concevoir des extensions du modèle KLM.

2.2 Extensions de KLM

L'une des motivations principales pour étendre le modèle KLM est d'évaluer de nouvelles technologies ou applications qui n'étaient pas prises en compte dans le modèle KLM original [1]. L'extension d'un KLM peut prendre différentes formes : nouveaux opérateurs, nouvelles heuristiques ou adaptation d'opérateurs existants.

L'avènement des téléphones mobiles a conduit les chercheurs à proposer de multiples extensions pour l'interaction avec un téléphone mobile [2,17,25]. Par exemple, Holleis et al. [12] ont étendu le KLM original en ajoutant et en adaptant des opérateurs pour l'interaction avec des téléphones mobiles (par exemple, Macro et Micro Attention Shift). Leur modèle étendu peut prédire un temps de réalisation d'une tâche avec une marge d'erreur comprise entre -15% et +8%. Plusieurs travaux ont exploré les extensions de KLM pour l'interaction avec les smartphones soit en ajoutant de nouveaux opérateurs (par exemple Tilt [25] ou Zoom [2]), soit pour un autre contexte d'utilisation (par exemple les applications de jeux [17]).

Les systèmes tactiles embarqués dans des véhicules ont également été largement étudiés [9,19,23,26]. Par exemple, Schneegaß et al. [26] ont modélisé une nouvelle formule pour intégrer la séparation entre la tâche de conduite principale et les tâches d'interaction secondaires, en plus des nouveaux opérateurs. Green [9] a ajouté de nouveaux opérateurs pour évaluer les tâches de navigation selon la règle des "15 secondes" (c'est-à-dire que la tâche de navigation ne doit pas prendre plus de 15 secondes). En outre, plusieurs travaux ont exploré les extensions de KLM pour de nouveaux dispositifs d'entrée sur les systèmes embarqués dans un véhicule (par exemple, le type de commande et la structure du menu [19] ou un contrôle à distance [23]).

Plus pertinentes pour les environnements immersifs, d'autres extensions ont exploré les interactions naturelles de l'utilisateur (Natural User Interfaces ou NUI), c'est-à-dire l'interaction avec les gestes de la main ou du bras uniquement. Par exemple, les travaux d'Erazo et al. [8] visent à décrire l'interaction avec les interfaces gestuelles de la main. Les gestes sont détectés à l'aide d'une Kinect. Ils utilisent l'unité G (G-unit [20]) permettant de décrire la période de temps entre le début et la fin d'un geste et adaptent un modèle KLM pour estimer le temps nécessaire pour réaliser chaque geste. Par exemple, ils définissent un opérateur de pointage (P) pour modéliser le temps nécessaire pour déplacer la main d'un point de départ à un point d'arrivée. Ils ont également conçu de nouvelles heuristiques (basées sur les heuristiques de Kieras [14]) pour placer leurs opérateurs cognitifs dans leur modélisation. Pour les lecteurs intéressés, Al-Megren et al. [1] proposent un état de l'art complet sur les extensions de KLM. Cependant, à notre connaissance et selon l'état de l'art d'Al-Megren, il y a toujours un manque d'extensions KLM pour les interactions en réalité mixte.

3 Notre extension KLM pour la realité mixte

Dans cette section, nous présentons notre extension KLM pour la réalité mixte. En particulier, nous visons à modéliser la recherche spatiale, le pointage du regard et la validation gestuelle d'objets numériques affichés dans l'environnement immersif. La pertinence de ces tâches est justifiée par le fait qu'elles font partie des tâches expliquées dans le tutoriel de prise en main de HoloLens (e.g. tutoriel Learn Gestures) et sont les premières à être abordées dans les tutoriels de développement [21]. De plus, ces tâches constituent des composantes fondamentales d'interactions plus complexes ; leur définition et caractérisation sont donc essentielles et incontournables avant de tendre vers un modèle plus complet. Nous

présentons d'abord les opérateurs hérités (que nous réutilisons de la littérature), avant d'introduire nos nouveaux opérateurs.

3.1 Opérateur hérité : Mental act (M)

Dans un premier temps, Card et al. [6] modélisent, avec l'opérateur **M**, le temps nécessaire pour réaliser des tâches cognitives entre deux actions utilisateurs. Cet opérateur peut modéliser différentes tâches cognitives (c'est-à-dire la recherche visuelle d'éléments parmi une liste, la planification d'actions futures, etc.).

Proposed Mnemonic	Task	Execution Time (ms)	
		Card et al.	Figure 2-28 & Figure 2-30
M_S	Simple Reaction	240 [105 – 470]	277 [±44]
M_P	Physical Matching	310 [130 – 640]	510 [±59]
M_N	Name Matching	380 [155 – 810]	485 [±52]
M_L	Class Matching	450 [180 – 980]	566 [±96]
M_C	Choice Reaction	200 + 150 $\log_2(N + 1)$	
M_V	Visual Search		498 + 41 N

Figure 1 : Les opérateurs mentaux de MacKenzie et leur unités de temps (Figure 7.33 tirée de [18:272])

MacKenzie [18] a affiné cet opérateur en le divisant en 5 opérateurs mentaux différents : réaction simple (**Ms**), correspondance physique (**Mp**), correspondance de nom (**Mn**), correspondance de classe (**Ml**) et recherche visuelle (**Mv**) (cf. Figure 1). L'unité de temps pour chaque opérateur a été mesurée expérimentalement, et nous les utilisons dans notre extension de KLM. Toutes les unités de temps pour les opérateurs M sont indiquées dans la figure 7.33 de [18:272].

3.2 Nouveaux opérateurs pour l'interaction avec un casque de réalité mixte

Nous présentons quatre nouveaux opérateurs pour modéliser la recherche spatiale, le pointage du regard et la validation gestuelle. Le dernier est divisé en deux opérateurs, l'un pour lever la main et l'autre pour effectuer un geste de validation.

3.2.1 Pointage du regard (Gp)

Le pointage du regard est la solution par défaut proposée par HoloLens pour pointer et sélectionner un objet numérique. L'opérateur de pointage du regard (Gaze Pointing : **Gp**) modélise le temps nécessaire à un utilisateur pour pointer un objet déjà visible dans le champ de vision du casque de réalité mixte à l'aide d'un curseur. Cet opérateur implique de légers mouvements de tête pour amener le curseur d'un point de départ à un point d'arrivée (c'est-à-dire la cible). Le pointage du regard dans la plupart des casques de réalité mixte est « *viewport-based* » [7] : un curseur est placé au centre du champ de vision et reste au centre lorsque l'utilisateur bouge la tête.

3.2.2 Recherche spatiale (Ss)

Les nouveaux appareils de réalité mixte tels que l'HoloLens ont un champ visuel restreint par rapport au champ visuel humain. La recherche d'une cible en dehors du champ de vision du casque de réalité mixte est donc une tâche très courante dans les applications de réalité mixte. L'opérateur de recherche spatiale (Spatial Search : **Ss**) décrit le temps nécessaire à un utilisateur pour rechercher un objet en dehors du champ de vision du casque de réalité mixte. Cette action implique des mouvements de tête importants et multidirectionnels.

3.2.3 Lever la main (Rh)

La première étape pour effectuer tout geste en l'air est de lever la main afin que le casque de réalité mixte puisse la détecter et reconnaître les gestes. L'opérateur "Lever la main" (Raise Hand : **Rh**) modélise le temps nécessaire à l'utilisateur pour faire passer sa main d'une position de repos à une position détectée par le casque.

3.2.4 Air Tap (At)

Sélectionner ou cliquer sur un bouton sont deux actions majeures pour préparer la manipulation d'un objet ou pour activer une commande. Avec l'HoloLens, cette action s'effectue en exécutant un geste d'Air Tap lorsque la main est

détectée par le casque. Ce geste consiste à abaisser l'index et à le relever à nouveau. Ce dernier opérateur (Air Tap : **At**) modélise le temps nécessaire à l'exécution de ce geste.

4 ÉTUDES UTILISATEURS : DÉFINITION D'UNE UNITÉ DE TEMPS POUR LES NOUVEAUX OPÉRATEURS

Après avoir introduit les nouveaux opérateurs de notre extension KLM, nous définissons l'unité de temps pour chacun d'eux, c'est-à-dire la durée prévue de l'action par l'utilisateur. Nous avons réalisé deux études utilisateurs : la première pour déterminer le temps des deux gestes en l'air (Air Tap et Lever la main) et la seconde portant sur les deux actions de la tête (Recherche Spatiale et Pointage du regard). Pour ces deux études, nous avons utilisé un casque HoloLens v1. Toutes les mesures présentées dans ces expérimentations sont captées par l'HoloLens et sont liées à des évènements systèmes gérés par la MixedReality Toolkit (MRTK) [22].

4.1 Étude 1 : Gestes en l'air

4.1.1 Tâche

Pour mesurer le temps nécessaire pour Lever la main (**Rh**) et effectuer le geste Air Tap (**At**), nous avons demandé aux participants d'amener leur main dominante dans le champ de vision de l'HoloLens, et d'effectuer quatre gestes Air Tap successifs. Avant de commencer chaque essai, les participants devaient placer leur main le long de leur cuisse. Pour limiter les actions cognitives ou physiques des participants, ils devaient regarder un minuteur affiché sur un cube devant leurs yeux et attendre le compte à rebours des 3s pour commencer l'essai. En effet, effectuer d'autres actions (cognitives ou physiques) qu'un Air Tap ou un Lever la main pouvait entraîner des erreurs dans l'unité de temps mesurée.

4.1.2 Participants, équipement et protocole

12 participants (3 femmes et 9 hommes âgés de 23 à 32 ans, avec une moyenne de 27,4) ont réalisé cette étude. 10 sont des étudiants et 2 sont des ingénieurs.

En préambule à l'expérimentation, une vidéo illustrant la tâche a été présentée aux participants. Étant donné que l'expérimentateur ne peut pas voir ce que le participant voit dans le casque, la vidéo sert à expliquer la tâche. Les participants pouvaient poser n'importe quelle question pendant la lecture de la vidéo. Suite à cette vidéo, les participants ont effectué un tutoriel officiel de Microsoft intégré à l'HoloLens appelé "Learn Gestures ", dans lequel ils ont appris à effectuer les gestes de base pour les interactions de l'HoloLens : Lever la main, Air Tap et Pointer du regard.

Ensuite, les participants ont effectué 5 essais d'entrainement, suivis de 15 essais d'étude. Nous avons enregistré le temps correspondant au lever de la main (c'est-à-dire entre le moment où la main quitte la cuisse et celui où la main est détectée par l'HoloLens), et le temps requis pour chaque Air Tap. Le temps enregistré pour un Air Tap correspond au temps nécessaire pour abaisser l'index et le relever à nouveau. Nous avons collecté 15 gestes de lever la main et 4 Air Taps x 15 essais = 60 Air Taps par participant.

4.1.3 Résultats

Pour analyser les résultats, nous avons décidé de calculer la moyenne géométrique plutôt que la moyenne arithmétique, car elle convient mieux aux données ayant une distribution à longue traîne comme le temps de complétion, avec des intervalles de confiance (IC) à 95%. Le temps moyen pour lever la main est de 899 ms (IC [861 938]). L'intervalle de confiance de taille très restreinte confère un degré de confiance élevé dans cette unité de temps.

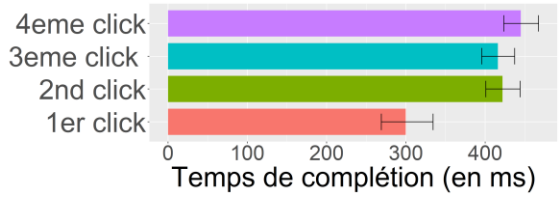

Figure 2 : Le temps de complétion des quatre Air Tap (en ms) avec des IC à 95%

Nous étudions à présent le temps pour réaliser un Air Tap en considérant chacun des quatre Air Taps consécutifs indépendamment. Comme l'illustre la Figure 2, nous observons que le premier Air Tap prend nettement moins de temps que les trois suivants (300 ms, IC [269 334]). Nous pensons que cette différence est due soit au fait que les participants ont commencé le premier Air Tap en finissant de lever leur main, soit au fait que nous calculons le temps du premier Air Tap depuis le temps de détection de la main qui peut être postérieur au temps de démarrage du geste Air Tap. Nous avons donc décidé d'exclure le premier Air Tap de notre calcul. En conséquence, le temps moyen pour réaliser un Air Tap est de 427ms IC [408 448]. Comme pour l'opérateur Lever la main (**Rh**), le petit intervalle de confiance confère un degré de confiance élevé à cette unité de temps.

4.2 Étude 2 : Actions basées sur des mouvements avec la tête

Pour définir l'unité de temps correspondant aux opérateurs de recherche spatiale et de pointage du regard, notre étude a été divisée en deux parties consécutives, une pour l'opérateur de recherche spatiale (**Ss**) et une pour l'opérateur de pointage du regard (**Gp**). Dans ces deux parties, les participants étaient les mêmes que dans l'étude 1 et nous avons utilisé le même équipement. Pour analyser les résultats, nous avons utilisé la moyenne géométrique avec un IC à 95% comme dans l'étude précédente.

4.2.1 Opérateur de recherche spatiale (Ss)

4.2.1.1 Tâche et protocole

La recherche d'une cible dans un environnement immersif consiste à déplacer (tourner) la tête jusqu'à ce que la cible entre dans le champ de vision (Field of View : FoV) du casque de réalité mixte. Le temps nécessaire pour amener le FoV sur la cible dépend de sa distance angulaire par rapport à la position initiale du centre du FoV. Selon [27;29–30], la rotation maximale confortable de la tête est de 60° (sans la rotation des épaules et l'inclinaison du corps). Ainsi, nous avons défini trois distances angulaires par rapport à la cible (45°, 60° et 75°) en dehors du champ de vision de l'HoloLens afin de tester des conditions aux limites, en deçà et au-delà. De plus, comme souligné dans [27], il est important de distinguer les mouvements verticaux et horizontaux de la tête. Dans notre étude, nous avons donc positionné les cibles dans 8 directions correspondant aux combinaisons possibles de mouvements de tête horizontaux et verticaux : deux mouvements verticaux (Nord, Sud), deux mouvements horizontaux (Est, Ouest), et 4 mouvements hybrides (Nord-Est, Nord-Ouest, Sud-Est et Sud-Ouest, cf. Figure 3). En conséquence, et conformément aux directives de l'HoloLens [29], des cibles de 7,5 cm sont placées sur une sphère de 200 cm de rayon centrée sur les yeux de l'utilisateur (soit une taille angulaire de 2.14 degrés). Chaque cible est placée dans une des huit directions. En combinant les 3 distances angulaires avec les 8 directions, on obtient 24 positions possibles de la cible autour du champ de vision.

Pour commencer un essai, les participants devaient attendre un compte à rebours de 3s placé sur un objet de départ affiché devant leurs yeux en regardant droit devant eux. Ensuite, une flèche indiquait la direction à suivre pour trouver la cible, et donc amener le champ de vision sur la cible. Pour éliminer tout effet dû à la différence entre la largeur et la hauteur du champ de vision, les participants devaient placer la cible à l'intérieur du « cercle de vision », i.e. un cercle ayant pour centre le centre du FoV et tangent aux bords les plus proches du FoV. Le rayon de ce cercle a une distance angulaire de 8,75° (la hauteur du champ de vision de l'HoloLens est de 17,5°). L'essai se termine automatiquement lorsque le centre de la cible est entré dans ce cercle et y est resté sans en ressortir pendant 500 ms. Les cibles s’affichent

les unes à la suite des autres. Notons enfin que pour que l'œil humain perçoive une cible correctement il n'est pas nécessaire de la placer au centre exact du champ de vision mais dans une zone autour de ce centre : selon [16,27], l'œil humain peut percevoir correctement du texte dans une zone comprise en 5 et 10° de part et d'autre d'un point placé devant l'utilisateur. L'utilisation d'un cercle inscrit de 8,75° dans notre étude est donc bien conforme à l'intervalle recommandé. Par conséquent, l'utilisation d'un dispositif offrant un champ de vision plus large ne remettrait pas en cause la validité du protocole adopté puisque la taille du champ de vision n'affecte pas la zone de perception d'un texte par l'utilisateur.

Cette partie de l'étude a suivi un plan 3x8 avec la distance angulaire (45°, 60°, 75°) et la direction (2 horizontales, 2 verticales, 4 hybrides) comme facteurs. Chaque bloc correspondait à une distance angulaire et comprenait huit cibles (une dans chaque direction) présentées au hasard au participant. Les blocs étaient contrebalancés entre les participants. Chaque bloc a été répété trois fois, le premier servant à l'apprentissage de la tâche. Au total, nous avons recueilli 3 distances angulaires x 8 directions x 2 répétitions = 48 essais par participant.

Dans cette première partie de la deuxième étude, nous avons mesuré le temps de réalisation entre la fin du compte à rebours et le moment où la cible est entrée dans le cercle de vision (c'est-à-dire que nous avons supprimé les 500 ms de validation de l'essai).

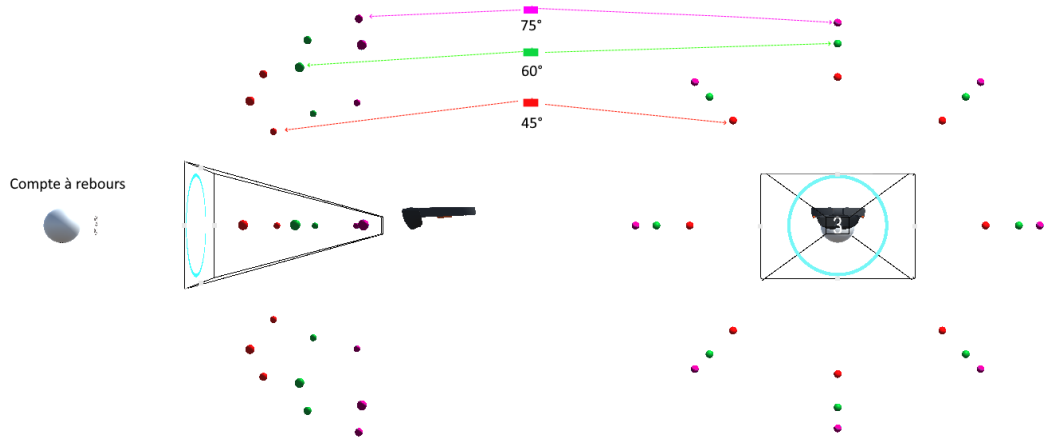

Figure 3 : Les différents composants de la tâche de Recherche Spatial avec (en rouge) les cibles à 45°, (en vert) les cibles de 60°, (en rose) les cibles à 75°, (en blanc) le compte à rebours et (en bleu) le cercle de vision.

4.2.1.2 Temps de réalisation de la tâche de recherche spatiale

Tout d'abord, nous avons observé que la distance angulaire a un impact évident sur le temps de réalisation de la recherche de la cible : le temps augmente nettement avec la distance angulaire (45° : 561 ms, IC [518 607] ; 60° : 852 ms IC [796 913] ; 75° : 1044ms, IC [969 1125]). La direction a également un impact sur ce temps de réalisation (986ms IC [907 1072] pour les mouvements verticaux, 613ms IC [560 671] pour les mouvements horizontaux et 810ms IC [757 866] pour les mouvements hybrides). La figure 4 illustre ces résultats et montre également que ces résultats restent vrais lorsque l'on considère ces deux facteurs simultanément.

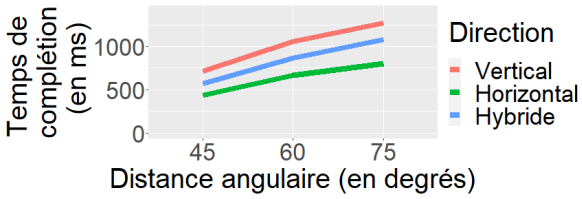

Figure 4 : Le temps de complétion pour la Recherche Spatiale en fonction de la direction (d) et de la distance angulaire (a)

De ces résultats, nous avons extrait 9 moyennes géométriques, résumées dans le Tableau 1 et qui peuvent être utilisées comme unité de temps pour l'opérateur Recherche Spatiale (**Ss**). Ces valeurs dépendent de la direction (d) et de la distance angulaire (a).

Tableau 1 : Moyennes en millisecondes du temps de complétion de la Recherche Spatiale (Ss) par rapport à la direction (*d*, colonnes) et la distance angulaire (*a*, lignes) : Ss(d,a)

	Horizontal	Hybride	Vertical
45	431 IC [365,510]	567 IC[510,629]	715 IC[617,829]
60	668 IC [595,752]	865 IC[783,957]	1054 IC[936,1187]
75	797 IC [695,914]	1083 IC[970,1209]	1270 IC[1121,1441]

4.2.2 Pointage du regard

4.2.2.1 Tâche et protocole

La deuxième partie de cette étude utilisateur vise à mesurer le temps nécessaire pour pointer du regard une cible déjà présente dans le champ de vision. Nous avons positionné huit cibles sur le « cercle de vision », comme défini pour la partie précédente de l'étude, et selon les huit directions cardinales (cf. Figure 5). Pour commencer un essai, les participants devaient attendre un compte à rebours de 3s devant eux comme dans la partie précédente de l'étude. Ensuite, ils devaient amener le curseur du FoV de l'HoloLens sur la cible et rester 500ms pour valider l'essai. Les cibles s'affichent les unes à la suite des autres.

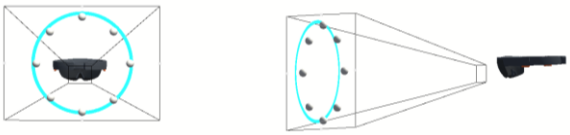

Figure 5 : Les différents composants de la tâche de Pointage du Regard avec le champ de vision, (en blanc) les cibles et (en bleu) le cercle inscrit permettant de positionner les cibles.

4.2.2.2 Temps de réalisation de la tâche de pointage du regard

Contrairement à la tâche de recherche spatiale (**Ss**), les résultats concernant la tâche de pointage du regard (**Gp**) ne permettent pas d'identifier un impact de la direction sur les temps de réalisation nécessaires pour pointer sur une cible déjà présente dans le champ de vision (Horizontal = 452ms, IC [369,554] ; Hybride = 439ms, IC [378,506] ; Vertical = 358ms, IC [299,429]) cf. figure 6).

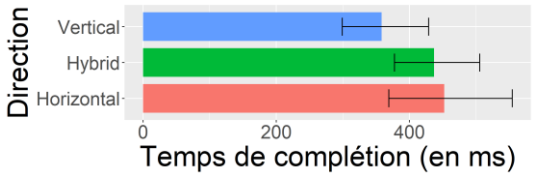

Figure 6 : Le temps de complétion pour la tâche de Pointage du regard en fonction de la direction (en ms)

De plus, nous avons collecté le nombre de fois que le curseur entrait et sortait de la cible. Sur les 192 essais, 105 ont été réalisés sans sortir de la cible (54%), 76 ont été réalisés en sortant une fois de la cible (39%) et seulement 11 (5.7%) ont nécessité plus de deux entrées sur la cible avant la sélection finale.

En conclusion, le temps moyen est de 419 ms, IC [380 463]. Nous considérons donc que ce résultat peut être utilisé comme l'unité de temps de l'opérateur de pointage de tête de notre extension de KLM.

4.3 Résumé

Dans ces deux études, nous avons défini des valeurs de temps unitaires pour les opérateurs nouvellement définis de notre extension KLM pour l'interaction avec un casque de réalité mixte : Air Tap (**At** : 427 ms), Lever la main (**Rh** : 899 ms), Recherche Spatiale (**Ss** : 9 valeurs en fonction de la distance angulaire et de la direction) et Pointage du regard (**Gp** : 419 ms). Tous les gestes et leur unité de temps associée sont résumés dans le Tableau 2.

Les unités de temps pour les opérateurs Air Tap (**At**) et Lever la main (**Rh**) sont fortement établies. Pour renforcer encore la validité des deux mouvements de la tête, nous avons réalisé une troisième étude dans laquelle ces deux opérateurs sont utilisés l'un à la suite de l'autre.

Tableau 2 : Notre extension de KLM

Nom	Abréviation	Unité de temps (in ms)	Origine de l'opérateur
Pointage du Regard	**Gp**	419 IC[369,554]	Nouveau
Recherche Spatiale	**Ss**	De 431 à 1270 cf. Figure 4	Nouveau
Lever la Main	**Rh**	899 IC[861,938]	Nouveau
Air Tap	**At**	427 IC[408,448]	Nouveau
Mental Act	**M**	De 277 à 566 cf. Figure 1	[18]

5 Etude 3 : Etude des opérateurs modélisant les mouvements de la tête

Une tâche très courante en environnement immersif consiste à rechercher et à pointer une cible qui est initialement hors du champ de vision. Cette tâche est donc composée d'une combinaison des opérateurs Pointage du regard (**Gp)** et Recherche Spatiale (**Ss)**, pour lesquels nous avons défini expérimentalement l'unité de temps dans l'étude précédente. Cette troisième étude vise à valider le temps unitaire précédemment défini pour ces nouveaux opérateurs par leur utilisation dans une tâche plus écologique les utilisant de manière séquentielle. L'objectif de cette étude est donc de vérifier que le temps d'exécution d'une tâche "pointer hors du champ de vision" mesuré expérimentalement est similaire à celui que nous prédisons grâce à notre extension KLM.

5.1 Design de l’étude et tâche

Nous avons recruté les mêmes participants pour cette étude. Nous étudions les mêmes facteurs que ceux présentés dans la tâche de recherche spatiale (cf. section 4.2.1) : distance angulaire et direction.

La tâche s'est déroulée en 2 étapes : 1) rechercher la cible dans la direction indiquée par une flèche et la placer à l'intérieur du cercle inscrit dans le « cercle de vision » (opérateur de Recherche spatiale - **Ss**) et 2) pointer avec le curseur de l'HoloLens sur la cible (opérateur de Pointage du regard - **Gp**). Nous avons utilisé le même protocole que la tâche de recherche spatiale.

5.2 Comparaison entre temps mesuré et temps prédit

La tâche est composée des deux opérateurs **Gp** et **Ss**. Le modèle KLM propose des heuristiques pour insérer des opérateurs correspondant à des actions cognitives entre deux actions physiques. Dans notre modélisation de la tâche, un opérateur mental doit être ajouté entre les opérateurs **Gp** et **Ss**. Parmi les cinq différents opérateurs mentaux présentés précédemment dans la section 3.1, seul l'opérateur de réaction simple (**Ms** = 277 ms) est approprié dans la tâche de cette troisième étude car aucune recherche visuelle parmi plusieurs objets n'est requise (c'est-à-dire qu'une seule cible est affichée en même temps dans le champ de vision). Dans une étude précédente, nous avons constaté que l’opérateur **Ss** dépend des distances angulaires et de la direction du mouvement de la tête. Nous avons modélisé **Ss(d,a)**, où d est la direction (horizontale, verticale ou hybride) et a est la distance angulaire (45°, 60° ou 75°) (cf. Tableau 1). Ainsi, pour la tâche écologique considérée dans cette étude, nous modélisons le temps prédit en fonction de la direction et de la distance angulaire : pt(d,a) = **Ss(d,a)** + **Ms** + **Gp**. Par exemple, pt(hybride, 60°) = 865 + 277 + 419 = 1561 ms.

Dans le tableau ci-dessous, nous indiquons le temps moyen d'exécution mesuré lors de l’exécution de la tâche par les participants avec un IC de 95% pour chacune des 3 directions et distances angulaires. Dans chaque cellule du tableau, nous calculons le pourcentage de correspondance des moyennes des temps mesurés par rapport à la prédiction KLM : un résultat de 100% signifie que le temps observé correspond parfaitement au temps prédit. Une valeur supérieure (resp. inférieure) signifie que le temps mesuré est supérieur (resp. inférieur) au temps prédit.

En moyenne, nous avons trouvé une correspondance de 98,7% : le temps mesuré est légèrement inférieur au temps prévu. Le pourcentage de correspondance varie de 90% (horizontal 60) à 105,7% (vertical 45) (cf Tableau 3).

Ces pourcentages de correspondance sont conformes aux travaux précédents [12] qui ont montré que leurs prédictions KLM représentaient 85 % à 108 % des valeurs observées. De plus, étant donné la taille des intervalles de confiance liés à l'unité de temps pour l'opérateur **Ms** donné par MacKenzie [18] (de 113ms à 528, soit une différence de -164ms et +251ms avec l'unité de temps de **Ms**) et ceux observés pour nos nouveaux opérateurs, la différence restante entre le temps de complétion mesuré et le temps prédit est cohérente. Par conséquent, ces résultats tendent à confirmer que nos contributions, les opérateurs identifiés et leurs temps unitaires, constituent une extension KLM valable. Tous les scripts et données utilisés pour notre analyse sont disponibles en ligne[1].

[1] https://osf.io/y85nv/?view_only=49ca125399a9418da1d9d04e93829e3d

Tableau 3 : Temps de complétion (CT) avec des IC à 95% (en ms) et le pourcentage de correspondance entre le KLM prédit et le temps mesuré.

	Horizontal		Hybrid		Vertical	
	CT	Deviation	CT	Deviation	CT	Deviation
45	1155 [1040,1283]	102.5 [92,113]	1299 [1218,1385]	102.8 [96.4,109.7]	1491 [1295,1716]	105.7 [91.8,121.6]
60	1229 [1109,1360]	90 [81,99.7]	1567 [1462,1680]	100.4 [93.7,107.6]	1601 [1449,1768]	91.4 [82.8,101]
75	1440 [1303,1592]	96.4 [87,106]	1796 [1665,1938]	101 [93.6,108.9]	1924 [1761,2102]	97.9 [89.6,106.9]

6 Discussion

6.1 Opérateurs

L'HoloLens, et plus largement la réalité mixte, permet à l'utilisateur de réaliser des interactions plus complexes que celles évoquées dans ce travail (e.g. manipulation d'objets virtuels 3D). Ces interactions nécessitent donc d'autres opérateurs pour être correctement modélisées. Cependant, ces interactions ne peuvent être réalisées qu'après une tâche de recherche, pointage puis sélection d'un objet numérique. C'est pourquoi nous nous sommes concentrés dans cette première étude sur des opérateurs fondamentaux, étude qui nécessitera de futures extensions portant sur des opérateurs (par exemple Pinch, Grab, Drag, etc.) utiles à la réalisation de tâches plus avancées.

6.2 Niveau d'expertise

L'utilisation de KLM repose sur l'identification et la combinaison d'opérateurs correspondant à des étapes d'interaction atomique. A chaque opérateur est associé une unité de temps correspondant au temps mis par un expert pour réaliser cette interaction atomique. Dans le contexte de nos études, l'utilisation du tutoriel intégré à HoloLens pour apprendre les gestes dure approximativement 10 minutes et les différentes phases d'entrainements précédent chaque tâche durent entre un tiers et un quart du temps total des expérimentations. Ce tutoriel et cet entrainement permettaient donc aux participants de se familiariser puis de maitriser les gestes demandés. De plus, les gestes effectués (lever la main, tourner la tête, baisser le doigt) sont des gestes mécaniques simples qui ne demandent pas d'expertise métier particulière. Par conséquent, il semble raisonnable de considérer que les temps établis dans ces travaux sont indépendants du degré d'expertise de l'utilisateur avec des dispositifs immersifs.

7 Conclusion et travaux futurs

Dans cet article, nous proposons une nouvelle extension de KLM [5], pour les applications immersives de réalité mixte, modélisant la recherche spatiale, le pointage du regard et les tâches de validation gestuelles. Tout d'abord, nous avons identifié et décrit les nouveaux opérateurs nécessaires pour modéliser les différentes tâches : Pointage du regard (**Gp**), Recherche spatiale (**Ss**), Lever la main (**Rh**) et Air Tap (**At**). À travers deux études utilisateurs, nous avons défini des temps unitaires pour chaque opérateur. Nous avons mené une dernière étude pour utiliser les temps unitaires des opérateurs **Gp** et **Ss** dans une tâche écologique ce qui a permis de confirmer que l'utilisation de ces opérateurs demeurent valables, dans un contexte plus élaboré et mêlés aux opérateurs mentaux existants.

Dans un avenir proche, nous prévoyons de confronter ce modèle à des tâches plus complexes. Par exemple, nous pourrions comparer une tâche de navigation dans différents menus placés dans différentes parties de l'agencement spatial

avec notre temps prédit et le temps observé. Cette tâche nous permettra d'évaluer l'effet du placement des objets dans l'agencement spatial du casque de réalité mixte sur le temps de réalisation de la tâche. Plus globalement, ce modèle sera utile pour évaluer et comparer, pendant la phase de conception, différentes solutions parmi le large ensemble de possibilités d'interaction dans les environnements immersifs.

Pour soutenir davantage l'exploration de cet espace de conception, plusieurs extensions sont envisagées. En particulier, des travaux antérieurs sur les extensions KLM ont modélisé des interactions basées sur des smartphones. La combinaison d'écrans tactiles avec des HMDs pourrait être utile pour surmonter les problèmes d'interaction en l'air en réalité mixte, tels que la fatigue [10,11,13], la précision limitée [28] et la découverte des gestes [3]. À cette fin, il serait intéressant d'étendre notre modèle KLM avec des opérateurs pour les interactions basées sur les smartphones (comme présenté dans l'état de l'art). Ce nouveau modèle pourrait être utilisé pour concevoir et évaluer de manière prédictive de nombreuses techniques d'interaction.

8 Remerciements

Ces travaux sont financés par la région Occitanie dans le cadre du projet IDPHYx et conjointement par le projet neOCampus (Université de Toulouse 3).

Bibliographie

<bib id="bib1"><number>[1]</number>Shiroq Al-Megren, Joharah Khabti, and Hend S. Al-Khalifa, 2018. A Systematic Review of Modifications and Validation Methods for the Extension of the Keystroke-Level Model. In *Advances in Human-Computer Interaction* 2018. DOI: https://doi.org/10.1155/2018/7528278</bib>
<bib id="bib2"><number>[2]</number>Karim El Batran and Mark D. Dunlop, 2014. Enhancing KLM (keystroke-level model) to fit touch screen mobile devices. In In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, 283–286. DOI: https://doi.org/10.1145/2628363.2628385</bib>
<bib id="bib3"><number>[3]</number>Thomas Baudel and Michel Beaudouin-Lafon, 1993. Charade: remote control of objects using free-hand gestures. In *Communications of the ACM* 36, 7: 28–35. DOI: https://doi.org/10.1145/159544.159562</bib>
<bib id="bib4"><number>[4]</number>Regina Bernhaupt, David Navarre, Philippe Palanque, and Marco Winckler, 2008. Model-Based Evaluation: A New Way to Support Usability Evaluation of Multimodal Interactive Applications. In In *Maturing usability*. 96–119. DOI: https://doi.org/10.1007/978-1-84628-941-5_5</bib>
<bib id="bib5"><number>[5]</number>Smart K Card, Thomas P Moran, and Allen Newel, 1980. The KLM for User Performance Time with Interactive Systems. In *Communications of the ACM* 23, 7: 396–410.</bib>
<bib id="bib6"><number>[6]</number>Stuart K. Card, Thomas P Moran, and Allen Newell, 1983. The psychology of human-computer interaction.</bib>
<bib id="bib7"><number>[7]</number>Yuan Chen, Keiko Katsuragawa, and Edward Lank, 2020. Understanding Viewport- and World-based Pointing with Everyday Smart Devices in Immersive Augmented Reality. In 1–13. DOI: https://doi.org/10.1145/3313831.3376592</bib>
<bib id="bib8"><number>[8]</number>Orlando Erazo and José A. Pino, 2018. Predicting user performance time for hand gesture interfaces. In *International Journal of Industrial Ergonomics* 65: 122–138. DOI: https://doi.org/10.1016/j.ergon.2017.07.010</bib>
<bib id="bib9"><number>[9]</number>Paul Green, 1999. Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43, 18: 987–991. DOI: https://doi.org/10.1177/154193129904301809</bib>
<bib id="bib10"><number>[10]</number>Jeffrey T. Hansberger, Chao Peng, Shannon L. Mathis, Vaidyanath Areyur Shanthakumar, Sarah C. Meacham, Lizhou Cao, and Victoria R. Blakely, 2017. Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions. In In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 505–520. DOI: https://doi.org/10.1007/978-3-319-57987-0_41</bib>
<bib id="bib11"><number>[11]</number>Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani, 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 1063–1072. DOI: https://doi.org/10.1145/2556288.2557130</bib>
<bib id="bib12"><number>[12]</number>Paul Holleis, Friederike Otto, Heinrich Hussmann, and Albrecht Schmidt, 2007. Keystroke-level model for advanced mobile phone interaction. In In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 1505–1514. DOI: https://doi.org/10.1145/1240624.1240851</bib>
<bib id="bib13"><number>[13]</number>Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani, 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3328–3339. DOI: https://doi.org/10.1145/3025453.3025523</bib>
<bib id="bib14"><number>[14]</number>David Kieras, 2001. Using the keystroke-level model to estimate execution times. In *University of Michigan*: 1–11. Retrieved from http://www-personal.umich.edu/~itm/688/KierasKLMTutorial2001.pdf</bib>
<bib id="bib15"><number>[15]</number>David Kieras, 2009. *Model-Based Evaluation*. In CRC Press, Mahwah, New Jersey. DOI: https://doi.org/10.1201/9781420088892</bib>
<bib id="bib16"><number>[16]</number>Bernard C. Kress, 2019. Digital optical elements and technologies (EDO19): applications to AR/VR/MR. In In *Digital Optical Technologies 2019*, 76. DOI: https://doi.org/10.1117/12.2544404</bib>
<bib id="bib17"><number>[17]</number>Ahreum Lee, Kiburm Song, Hokyoung Blake Ryu, Jieun Kim, and Gyuhyun Kwon, 2015. Fingerstroke time estimates for touchscreen-based mobile gaming interaction. In *Human Movement Science* 44: 211–224. DOI: https://doi.org/10.1016/j.humov.2015.09.003</bib>

<bib id="bib18"><number>[18]</number>I. Scott MacKenzie, 2013. Human-computer Interaction: An Empirical Research Perspective. In In *Human-computer Interaction*. Elsevier, i–iii. DOI: https://doi.org/10.1016/B978-0-12-405865-1.00009-1</bib>
<bib id="bib19"><number>[19]</number>Daniel Manes and Paul Green, 1997. Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth).</bib>
<bib id="bib20"><number>[20]</number>David McNeill, 1994. Hand and Mind: What Gestures Reveal about Thought. In *Language* 70, 2: 345. DOI: https://doi.org/10.2307/415833</bib>
<bib id="bib21"><number>[21]</number>Microsoft, . MR Input 210: Gaze. Retrieved from https://docs.microsoft.com/fr-fr/windows/mixed-reality/develop/unity/tutorials/holograms-210</bib>
<bib id="bib22"><number>[22]</number>Microsoft, . MRTK homepage. Retrieved from https://microsoft.github.io/MixedRealityToolkit-Unity/README.html</bib>
<bib id="bib23"><number>[23]</number>C Nowakowski, Y Utsui, and P Green, 2000. Navigation System Destination Entry: the Effects of Driver Workload and Input Devices, and Implications for SAE Recommended Practice. Retrieved from http://www.umich.edu/~driving/publications/UMTRI-2000-20.pdf%5Cnhttps://trid.trb.org/view/654833</bib>
<bib id="bib24"><number>[24]</number>Peter G. Polson, Clayton Lewis, John Rieman, and Cathleen Wharton, 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. In *International Journal of Man-Machine Studies*. DOI: https://doi.org/10.1016/0020-7373(92)90039-N</bib>
<bib id="bib25"><number>[25]</number>Andrew D. Rice and Jonathan W. Lartigue, 2014. Touch-Level Model (TLM): Evolving KLM-GOMS for Touchscreen and Mobile Devices. In In *Proceedings of the 2014 ACM Southeast Regional Conference on - ACM SE '14*, 1–6. DOI: https://doi.org/10.1145/2638404.2638532</bib>
<bib id="bib26"><number>[26]</number>Stefan Schneegaß, Bastian Pfleging, Dagmar Kern, and Albrecht Schmidt, 2011. Support for modeling interaction with automotive user interfaces. In In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '11*, 71. DOI: https://doi.org/10.1145/2381416.2381428</bib>
<bib id="bib27"><number>[27]</number>Alvin Tilley and Henry Dreyfuss Associates, 1994. The Measure of Man and Woman : Human Factors in Design. In *Ergonomics in Design: The Quarterly of Human Factors Applications*. DOI: https://doi.org/10.1177/106480469400200211</bib>
<bib id="bib28"><number>[28]</number>Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen, 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque and Marco Winckler (eds.). Springer International Publishing, Cham, 315–330. DOI: https://doi.org/10.1007/978-3-319-22668-2_25</bib>
<bib id="bib29"><number>[29]</number>Interactable object. Retrieved July 9, 2020 from https://docs.microsoft.com/fr-fr/windows/mixed-reality/interactable-object</bib>